Prédiction de liens fonctionnels par détection de coévolution entre familles de gènes : application aux gènes du cycle cellulaire chez les Firmicutes

Pierre Garcia

▶ **To cite this version:**

Pierre Garcia. Prédiction de liens fonctionnels par détection de coévolution entre familles de gènes : application aux gènes du cycle cellulaire chez les Firmicutes. Bio-informatique [q-bio.QM]. Université de Lyon, 2018. Français. NNT : 2018LYSE1316 . tel-02418607v1

HAL Id: tel-02418607

https://theses.hal.science/tel-02418607v1

Submitted on 19 Dec 2019 (v1), last revised 19 Dec 2019 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

N° d'ordre NNT : 2018LYSE12316

THÈSE DE DOCTORAT DE L'UNIVERSITÉ DE LYON
opérée au sein de
l'Université Claude Bernard Lyon 1

École Doctorale ED205
Ecole Doctorale Interdisciplinaire Science Santé

Spécialité de doctorat : Bioinformatique

Soutenue publiquement le 18/12/2018, par :
Pierre Simon Garcia

Prédiction de liens fonctionnels par détection de coévolution entre familles de gènes : Application aux gènes du cycle cellulaire chez les *Firmicutes*

Devant le jury composé de :

Dr. GRIBALDO Simonetta, Directrice de recherche, Institut Pasteur	Rapporteure
Pr. MIJAKOVIC Ivan, Professeur des universités, Chalmers University	Rapporteur
Dr. ABBY Sophie, Chargée de recherche, CNRS	Examinatrice
Dr. DAUBIN Vincent, Directeur de recherche, CNRS	Examinateur
Pr. FLANDROIS Jean-Pierre, Professeur des universités, UCBL	Examinateur
Dr. MORLOT Cécile, Chargée de recherche, CNRS/CEA	Examinatrice
Dr GRANGEASSE Christophe, Directeur de recherche, CNRS	Directeur de thèse
Pr. BROCHIER-ARMANET Céline, Professeur des universités, UCBL	Co-directrice de thèse

UNIVERSITÉ CLAUDE BERNARD - LYON 1

Président de l'Université	**M. le Professeur Frédéric Fleury**
Président du Conseil Académique	M. le Professeur Hamda BEN HADID
Vice-président du Conseil d'Administration	M. le Professeur Didier REVEL
Vice-président de la Commission Recherche	M. le Professeur Philippe LALLE
Vice-président de la Commission formation et Vie Universitaire	M. le Professeur Philippe CHEVALIER
Directeur Général des Services	M. Alain HELLEU

COMPOSANTES SANTÉ

Faculté de Médecine Lyon Est – Claude Bernard	Directeur : M. le Professeur G. RODE
Faculté de Médecine et de Maïeutique Lyon Sud – Charles Mérieux	Directeur : Mme la Professeure C. BURILLON
Faculté d'Odontologie	Directeur : Mme la Professeur D. SEUX
Institut des Sciences Pharmaceutiques et Biologiques	Directeur : Mme la Professeure C. VINCIGUERRA
Institut des Sciences et Techniques de la Réadaptation	Directeur : Dr X. PERROT
Département de formation et Centre de Recherche en Biologie Humaine	Directeur : Mme la Professeure A-M. SCHOTT

COMPOSANTES ET DÉPARTEMENTS DE SCIENCES ET TECHNOLOGIE

Faculté des Sciences et Technologies	Directeur : M. le Professeur F. DE MARCHI
Département Biologie	Directeur : Mme la Professeure K. GIESELER
Département Chimie Biochimie	Directeur : Mme C. FELIX
Département GEP	Directeur : Mme R. FERRIGNO
Département Informatique	Directeur : M. B. SHARIAT
Département Mathématiques	Directeur : M. I. BEN YAACOV
Département Mécanique	Directeur : M. M. BUFFAT
Département Physique	Directeur : M. J-C. PLENET
Département Sciences de la Terre	Directeur : M. G. CUNY
UFR Sciences et Techniques des Activités Physiques et Sportives	Directeur : M. Y. VANPOULLE
Observatoire des Sciences de l'Univers de Lyon	Directeur : Mme I. DANIEL
Polytech Lyon	Directeur : M. E. PERRIN
École Supérieure de Chimie Physique Electronique	Directeur : M. B. BIGOT
Institut Universitaire de Technologie de Lyon 1	Directeur : M. C. VITON
École Supérieure du Professorat et de l'Education	Directeur : M. A. MOUGNIOTTE
Institut de Science Financière et d'Assurances	Directeur : M. N. LEBOISNE

Résumé en français

Le cycle cellulaire chez les bactéries est un processus très étudié mais il apparaît que les modèles actuels ne rendent pas compte de la complexité et surtout de la diversité des machineries et des mécanismes de régulation impliqués. En fait, notre connaissance du cycle cellulaire repose sur l'étude de quelques organismes modèles. Or les analyses comparatives ont montré que certains systèmes et mécanismes décrits sont peu conservés et donc difficilement transposables d'un taxon à l'autre. Des approches évolutives telles que la phylogénomique peuvent être utilisées pour l'étude fonctionnelle de tels systèmes biologiques à l'échelle des bactéries. Ces approches permettent notamment de déterminer les événements évolutifs clés qui ont conduit à une telle diversité mais également d'identifier des liens fonctionnels potentiels entre protéines. De plus, le développement des méthodes de séquençage à très haut débit a conduit à une accumulation de données génomiques sans précédent, notamment chez les procaryotes. Dans ce contexte, j'ai réalisé une analyse phylogénomique à très large échelle des protéines impliquées dans le cycle cellulaire et sa régulation chez les *Firmicutes*. Mon objectif était de rechercher des patrons de coévolution entre familles protéiques pouvant refléter des liens fonctionnels. L'application des méthodes développées dans le cadre cette thèse aux protéines impliquées dans le cycle cellulaire chez les *Firmicutes* a permis de reconstruire l'histoire évolutive de ce processus cellulaire fondamental à l'échelle de ce *phylum* bactérien majeur. En particulier, j'ai pu mettre en évidence l'existence de quelques points chauds correspondant par exemple à l'émergence des *Bacilli* ou des *Streptococcaceae*. L'émergence de ces *taxa* s'est accompagnée de nombreuses acquisitions et/ou de pertes de gènes ainsi que de nombreux réarrangements dans l'organisation des clusters de gènes codant pour ces protéines, suggérant que des changements majeurs se sont produits au niveau du cycle cellulaire et de sa régulation. J'ai également pu mettre en évidence de possibles liens fonctionnels qui n'ont jamais été décrits jusqu'à présent entre des gènes impliqués dans différentes machineries du cycle cellulaire. L'application de ces approches à l'ensemble des protéomes de *Firmicutes* a également permis d'identifier des protéines présentant des patrons de coévolution communs avec les protéines impliquées dans la division cellulaire et sa régulation, suggérant de possibles liens fonctionnels qu'il serait nécessaire de tester expérimentalement.

Résumé en anglais

The bacterial cell cycle is a very well studied process but current models don't reflect the complexity and diversity of involved molecular machineries and associated regulation mechanisms. In fact, our knowledge of cell cycle is based on study of a few model organisms. Yet, comparative analyses showed that some described systems and mechanisms are not conserved and not transposable from a taxon to another. Evolutionary approach such as phylogenomic can be used for functional studies of such systems at the bacterial scale. Those approaches allow to determine the key evolutionary events that lead to a such diversity but also to identify potential functional links between proteins. Furthermore, the development of high throughput sequencing methods leads to a big amount of genomic data, particularly for prokaryotes. In this context, I realized a very large scale phylogenomic analysis of proteins involved in cell cycle and its regulation in *Firmicutes*. My goal was to search some coevolution patterns between protein families reflecting potentially functional links. The application of methods that I developed during my PhD to cell cycle proteins allowed to reconstruct the evolutionary history of this cell process in *Firmicutes*. Notably, I highlighted some hot-spots corresponding for example to the emergence of *Bacilli* or *Streptococcaceae*. The emergence of such *taxa* has been accompanied by many acquisitions/losses of cell cycle genes but also many genomic rearrangements in gene clusters suggesting that major changes have occurred at the level of the cell cycle and its regulation. I also highlighted some potential functional links between genes involved in different machineries of cell cycle that have never been described. The application of these approaches to the entire proteomes of *Firmicutes* allowed to identify proteins presenting same evolution patterns than cell cycle proteins suggesting potential functional links that have to be experimentally tested.

Remerciements

Tout d'abord je tiens à remercier Dr. Simonetta Gribaldo et Pr. Ivan Mijakovic pour avoir accepter de juger le travail réalisé durant cette thèse. Je remercie aussi Dr. Sophie Abby, Dr. Vincent Daubin, Pr. Jean-Pierre Flandrois et Dr. Cécile Morlot pour l'intérêt qu'ils ont porté à mon travail en participant à ce jury.

Je remercie également les deux membres de mon comité de suivi de thèse, Dr. Eduardo Rocha et Dr. Anne Galinier pour leurs conseils avisés et leur intérêt pour mon sujet.

Je tiens à remercier mes collaborateurs, notamment l'équipe de Grenoble, Diego Carriel, Irina Gutsche et Sylvie Elsen. Merci de m'avoir fait connaître les LAOdc et de votre vif intérêt pour la phylogénie qui a mené à de nombreuses discussions.

Je tiens ensuite à grandement remercier mon premier directeur de thèse, Christophe Grangeasse. Christophe, tout d'abord, c'est toi qui m'as recruté et qui m'as montré que la microbiologie était passionnante, et particulièrement la division cellulaire. C'est aussi grâce à toi que j'ai eu l'opportunité de découvrir la bioinformatique, ce qui a été pour moi une vraie révélation. Bien que j'ai été peu présent au labo, ça a toujours été un plaisir de profiter de tes conseils avisés et pragmatiques mais également de ton humour sans limites. Merci aussi de m'avoir associé à de nombreuses collaborations, ma thèse n'en est que plus riche.

Je remercie aussi vivement Céline Brochier, ma deuxième directrice de thèse. Merci de m'avoir fait découvrir la phylogénie pendant mon Master et ma thèse. Nos discussions vives et nos débats scientifiques où nous étions rarement d'accord ont été enrichissantes. De plus, j'ai apprécié de recevoir de vous cette rigueur dans l'analyse phylogénétique et dans la rédaction, même si cela demande des efforts considérables et une patience à toute épreuve.

Merci aussi à tous les membres de l'équipe B3P, Chryslène, Sébastien, JP, Marine, Adrien, Stéphanie, Anaïs mais également aux anciens, Aurore, Sylvie, Laure et Julien. Chryslène, ça a toujours été un plaisir d'être en ta compagnie durant ma thèse. On ne s'est finalement pas

beaucoup vu au labo mais plutôt dans les restos et les bars. Même si tu comprends toujours pas l'intérêt de ce que je fais, je suis sur qu'un jour tu auras besoin de la phylogénie... et tu penseras à moi ;-). Merci à Sébastien pour ton humeur constante et ton humour .... décapant. Merci à Marine pour son aide pendant le développement de GeneSpy, en mode Beta-testeuse de la gestion des proxys OKLM. Une pensée pour ceux qui ne sont plus là... enfin au labo ! Sylvie, merci pour ta réceptivité à mes élucubrations scientifiques dont la signification est souvent obscure pour la plupart des gens. Merci aussi à toi et Aurore pour l'accueil +++ à Boston pendant le congrès de la Gordon. Enfin, merci à Laure Zuchzuch ... pour tes chansons et tutti canti mais aussi pour ton soutient sans faille durant cette difficile période qu'a été la préparation du concours de l'EDISS.

Merci à tous les membres du LBBE et particulièrement de l'équipe BPGE. Un grand merci à Jean-Pierre Flandrois qui a suivi de loin et de près l'avancement de ma thèse et qui m'a conseillé sur beaucoup de problématiques, et tout ça toujours avec une pointe d'humour. Merci à Wandrille dont les conseils avisés, notamment en programmation et en modélisation m'ont permis de comprendre des tonnes de concepts et d'outils. Merci aux membres du club Célinettes sous composé des groupes Orphelines et Biomérieux. Merci aux deux orphelines, Monique et Anne pour ... beaucoup de choses ! Tout d'abord parce que l'on a traversé toutes les épreuves de la thèse ensemble, qu'on a parcouru le Japon ensemble (Aligato gosaimaaaaaas, dédicace à Anne, TMTC) et qu'on va soutenir quasiment en même temps. Merci aussi à Fred, le seul membre du groupe Biomérieux, dont la drôlerie des blagues est systématiquement proportionnelle au nombre de références qu'il utilise. Merci aussi à Thibault san qui a rendu ce congrès au Japon encore plus fou. Je remercie aussi Najwa qui m'a aidé dans beaucoup de problématiques techniques et aussi à Héloïse qui m'a fait passé de très bonnes pauses et avec qui on a organisé les activités de la fête de la science. Enfin, bon courage aux thésards en cours et à venir.

Merci aussi aux membres de l'équipe MDH, Léa, Yani, Chloé A, Chloé 1, Lolo et Céline. Cet instant d'une heure quotidienne ou la recherche cartésienne basée sur l'élaboration conceptuelle d'hypothèses mécanistiques et la succession d'expérimentations infructueuses ponctuées par de rares mais gratifiantes sérendipités laisse place aux bonnes blagues bien grasses... STVCQJVD. Yani, courage pour la suite, une fois que t'auras largué la PLD, tu passeras à autre chose. Léa,

tu vas tout déchirer après ta thèse, tu es le futur de la médecine régénerative ! Chloé A, j'ai jamais rien compris à ce que tu fais mais si ça a un rapport avec Michaelis et Mentens, c'est probablement que c'est vachement intéressant. Enfin, bon courage à Lolo et Chloé 1 qui découvrent ce que ça fait de fréquenter 4 thésards en fin de thèse, ça présage évidemment que du bonheur.

Merci aux membres de l'amicale des quartiers gentrifiés de Lyon qui m'ont assisté principalement dans les « moments non professionnels ». Tout d'abord, les membres du 1er, les psys et assimilés (Ludo, Lucie, Michel, Antoine, Adèle, Astrid, Hermine, Léo) et les autres (François, Karim, Marie, Chach, Julie, Wiliam). Aussi, les membres du 7eme (actuels ou expatriés depuis), principalement des meds et assimilés : Toto et Claire, Robin et Olivia, Paul, Claire et Paul, Charles, Wilou et Maude, Marine et Clément, Arthur et Mathilde. Un petit bigup pour les anciennes collocs Marie, Chach la rage et Marine et pour Julie qui m'a corrigé les fautes.

Je veux aussi remercier les anciens de la promo de biologie/biochimie : Damien, Hanna, Clément, Nikoleï, Vincent, Pépé mais aussi Flora. Avec vous, les débuts dans la bio, ça a été incroyable, plein de débats scientifiques sans queue ni tête et beaucoup d'amour.

Ensuite, un grand merci à la PFST, François, Simon et Thomas. C'est la team du début, et ça sera probablement celle de la fin. On a à peu près tout traversé ensemble et pourvu que ça dure !

Évidemment, je tiens à remercier ma famille. Mes parents d'abord, qui ont été un soutient sans faille durant toute ma vie, qui m'ont toujours laissé le choix dans tout ce que j'entreprenais et qui ont aiguisé ma curiosité sur les choses de la vie. Merci à ma sœur, Sophie, à son mari, Benichatte et aux petits qu'ils ont engendrés, Henri et Charlotte. De vous voir tous ensemble c'est mon petit rayon de soleil. Merci aussi à mon frère Thomas, avec qui j'ai fais « les 400 coups » et à Manon, sa compagne.

Enfin, merci infiniment à toi, Jean-Minou, qui m'a soutenu quotidiennement durant ces trois ans de thèse, pour les moments légers et plus difficiles. Une nouvelle aventure nous attend

à Paris, et j'espère qu'elle sera aussi riche que celle que l'on vient de traverser.

Cette thèse a été soutenue par une allocation doctorale de recherche de la Région Auvergne-Rhône-Alpes.

Liste des Abréviations

ABC Adenosine DiPhosphate

ADN Acide DésoxyriboNucléique

aLRT approximate Likelihood Ratio Test

ARN Acide RiboNucléique

ATP Adenosine TriPhosphate

BIC Bayesian Information criteria

BLAST Basic Local Alignment Search Tool

BLOSUM BLOcks SUbstitution Matrix

CAT CATegory substitution model

COG Clusters of Orthologous Genes

DCW Division and Cell Wall cluster

dif Site de résolution des dimères de chromosomes

ERAEG Événement de Réarrangement Avec Événement de Gènes

ERSEG Événement de Réarrangement Sans Événement de Gènes

FCC Famille de gènes du Cycle Cellulaire

FT Famille de gène Totale

GBFF GenBank Flat File

GDP Guanosine DiPhosphate

GenBank Genomic dataBank

GFF Generic Feature File

GFM GFF Minimal content

GPD Generalized Pareto Distribution

GTP Guanosine TriPhosphate

HMM Hidden Markov Model

HSP High scoring Segment Pair

JTT Jones-Taylor-Thornton

KOPS FtsK Orienting Polar Sequences

LG Le et Gascuel

LMSP Localy Maximal scoring Segment pair

MCMCMC Metropolis Coupling of Markov chain Monte-Carlo

ML Maximum Likelihood

MRL Mean Residual Life

MSP Maximum scoring Segment Pair

NCBI National Center of Biotechnology Information

NJ Neighbor Joining

NO Nucleoid Occlusion

Ori Domaine de l'origine

oriC Origine de réplication

PASTA PBP And Serine-Threonine kinase Associated

PBP Penicillin Binding Protein

PMSF Posterior Mean Site Frequency model

PP Probabilité Postérieure

PSDG PSeuDoGène

PSSM Position Specific Scoring Matrix

RefSeq Reference Sequences

SH-aLRT Shimodaira–Hasegawa approximate Likelihood Ratio Test

SH-like Shimodaira–Hasegawa like

SLC Single Likage Clustering

SPR Subtree Prune Regraft

Ter Domaine Terminal

TSMS Threshold stability for Modified Scale

TSS Threshold Stability for Shape

VT Variable Time

WAG Whelan et Goldman

XML Extensible Markup Language

+F Estimation de la fréquence de chaque acide aminé

+G4 Loi gamma avec 4 catégories discrètes

+I Estimation de la proportion de sites invariables

Table des matières

Table des figures **27**

Liste des tableaux **35**

Avant-propos **37**

1 Le cycle cellulaire bactérien **41**

1.1 Le cycle du chromosome 44

1.1.1 Structure du chromosome bactérien 44

1.1.2 La réplication du chromosome bactérien 45

1.1.3 La ségrégation des chromosomes fils 50

1.2 Le divisome 54

1.2.1 FtsZ, la protéine centrale de la division cellulaire 54

1.2.2 Régulation du positionnement de l'anneau Z 56

1.2.3 Régulation de la dynamique de l'anneau Z 61

1.2.4 Protéines du divisome associées à FtsZ 63

1.3 L'élongasome 65
1.3.1 MreB, protéine centrale de l'élongasome 65
1.3.2 Autres composants de l'élongasome 67
1.4 Le peptidoglycane, composant de la paroi cellulaire bactérienne 68
1.4.1 Rôle et structure du peptidoglycane 68
1.4.2 Étapes cytosoliques et membranaires de la biosynthèse 69
1.4.3 Étapes extracellulaires de la biosynthèse : les PBPs 73
1.4.4 Remodelage du peptidoglycane 74
1.4.5 Recyclage du peptidoglycane 76
1.5 La sporulation 77
1.5.1 Généralités 77
1.5.2 Mécanismes moléculaires de la sporulation 79
1.6 La capsule polysaccharidique 80
1.6.1 Rôle et structure de la capsule 80
1.6.2 Synthèse de la capsule 81
1.7 La coordination des machineries du cycle cellulaire 83
1.7.1 Coordination élongation-division : MreB et FtsZ 83
1.7.2 StkP, métarégulateur du cycle cellulaire 84
1.7.3 Coordination division-ségrégation chromosomique 87
1.7.4 Régulation de la synthèse du peptidoglycane 87

1.7.5 Autres coordinations entre machineries 88

1.8 Limites des modèles du cycle cellulaire bactérien 89

2 La phylogénomique : principes et applications 93

2.1 De la classification morphologique du vivant à la phylogénomique 96

2.1.1 La classification des êtres vivants et la phylogénie 96

2.1.2 La génomique 97

2.1.3 La phylogénomique 98

2.2 La construction de familles de gènes homologues et orthologues 100

2.2.1 Choix des séquences initiales et de la base de données 101

2.2.2 Construction des familles d'homologues et d'orthologues 102

2.3 Inférence de l'histoire évolutive d'un gène par la phylogénie 107

2.3.1 Alignement multiple 107

2.3.2 Inférence phylogénétique 108

2.3.3 Arbre d'espèces et identification d'événements évolutifs 114

2.4 Annotation fonctionnelle des génomes 116

2.4.1 Annotation par homologie et orthologie 116

2.4.2 Données phylogénomiques : comparaison entre familles de gènes 117

2.5 Les *Firmicutes* 128

2.5.1 Généralités 128

2.5.2 Taxonomie et Systématique des *Firmicutes* 131

3 Application de l'approche de co-évolutions aux familles de gènes du cycle cellulaire 133

3.1 Base de données des protéomes de *Firmicutes* 135

3.1.1 Construction des bases de données de séquences et d'annotation 135

3.1.2 Composition des bases de données 135

3.2 Inférence de la phylogénie des espèces de *Firmicutes* 138

3.2.1 Construction de l'arbre d'espèces daté des *Firmicutes* 138

3.2.2 Analyse de la topologie de l'arbre d'espèces des *Firmicutes* 143

3.3 Construction des familles de gènes impliqués dans le cycle cellulaire 148

3.3.1 Méthodes de construction des familles 148

3.3.2 Composition des familles 161

3.4 Événements évolutifs majeurs des familles de gènes du cycle cellulaire chez les *Firmicutes* 171

3.4.1 Reconstruction de l'histoire évolutive des familles de gènes 172

3.4.2 Événements évolutifs majeurs du cycle cellulaire chez les *Firmicutes* . . 179

3.5 Inférence de relations fonctionnelles par approche corrélative 203

3.5.1 Méthodes de corrélation 203

3.5.2 Liens fonctionnels entre familles de gènes du cycle cellulaire 220

3.6 Histoires évolutives détaillées et implications fonctionnelles 235

3.6.1 Description et interprétation fonctionnelle de l'histoire des clans 235

3.6.2 Description et interprétation fonctionnelle de liens évolutifs ponctuels . 263

3.6.3 Observations générales et conclusion 270

3.7 Identification de nouvelles familles potentiellement impliquées dans le cycle cellulaire 276

3.7.1 Méthode d'identification des familles de gènes corrélant avec les familles du cycle cellulaire 277

3.7.2 Familles corrélées évolutivement aux familles du cycle cellulaire 279

3.8 Conclusion 286

4 Étude de l'histoire évolutive des domaines PASTA chez les *Lactobacillales* 287

4.1 Le rôle des domaines PASTA 288

4.2 Matériel et méthodes 290

4.2.1 Assemblage des jeux de données 290

4.2.2 Groupement des domaines PASTA 291

4.2.3 Inférence des phylogénies 292

4.2.4 Construction des arbres d'espèces 292

4.3 Résultats/Discussion 294

4.3.1 Analyse phylogénétique des domaines PASTA chez les *Lactobacillales* . 294

4.3.2 Les différents rôles des domaines PASTA de StkP chez *S. pneumoniae* . 306

4.4 Conclusion 311

5 Développement d'un logiciel de visualisation de contextes génomiques, GeneSpy **315**

5.1 Contexte 316

5.1.1 La visualisation du contexte génomique : principes et enjeux 316

5.1.2 Les logiciels disponibles 318

5.2 GeneSpy, un outil flexible et facile d'utilisation pour l'exploration des contextes génomiques 320

5.2.1 Librairies, environnement et implémentation 320

5.2.2 Construction et gestion de la base de données 320

5.2.3 Importation des identifiants de gènes 324

5.2.4 Options de visualisation des figures 326

5.2.5 Exportation des données 327

5.3 Conclusion 328

6 Discussion générale et perspectives **331**

6.1 Limites et apports de la méthode 331

6.1.1 Méthodes de détections des homologues et choix des seuils 331

6.1.2 Le cas complexe des familles multigéniques 332

6.1.3 Méthode de reconstruction des états ancestraux et coûts d'événements 333

6.1.4 Méthodes corrélatives basées sur la reconstruction des états ancestraux par parcimonie 334

6.1.5 Méthodes de détection de potentielles nouvelles familles du cycle cellulaire 335

6.1.6 Développement d'un outil de visualisation de contextes génomiques . . 335

6.2 Inférences de liens fonctionnels et identification de nouveaux composants du cycle cellulaire 336

6.2.1 Reconstruction des familles de gènes du cycle cellulaire et de leur histoire 336

6.2.2 Identification des familles potentiellement liées fonctionnellement 337

6.2.3 Identification de nouvelles familles du cycle cellulaire 337

6.3 Perspectives 338

6.3.1 Amélioration de la méthode 338

6.3.2 Perspectives expérimentales 339

Bibliographie **341**

Annexes **397**

.1 Génomes et protéomes complets de *Firmicutes* contenus dans la base de données locale. 397

.2 Génomes et protéomes complets de *Firmicutes* après échantillonnage taxonomique. 408

.3 Protéomes utilisés pour le groupe externe. 414

.4 Nombre de protéines ribosomiques identifiées par souche. 416

.5 Nombre de séquences et de positions par alignement de protéines ribosomiques. 421

.6 Récapitulatif des familles reconstruites. 423

.7 Rôles biologiques des familles reconstruites. 428

.8 Protéomes de référence utilisés dans la recherche par profil HMM. 434

.9 Récapitulatif des phylogénies des familles de gènes. 436

.10 Nombre de gènes du cycle cellulaire analysés pour chaque souche. 442

.11 Phylogénies non racinées des familles multigéniques. 449

.12 Phylogénies non racinées des familles/sous-familles de gènes. 515

.13 Apparitions des familles de gènes pour les trois points chauds d'apparitions. . 702

.14 Nombre de FT proies par FCC. 704

.15 FT proies présentant une synténie avec des FCC. 706

.16 FT corrélant avec des FCC du clan Spo. 708

.17 FT corrélant avec des FCC du clan *Bacilli*. 710

.18 FT corrélant avec d'autres FCC. 712

.19 Comparatif des différents outils de visualisation de contextes génomiques. . . . 714

.20 Différents formats utilisés par GeneSpy. 716

.21 Publications scientifiques publiées au cours de ma thèse. 722

Table des figures

1 Quelques exemples de la diversité morphologique des bactéries. 38

1.1 Le cycle cellulaire bactérien et la fission binaire 44

1.2 Structure du chromosome bactérien 45

1.3 Modèle de réplication du chromosome chez *E. Coli* 47

1.4 Réparation de l'ADN et résolution des dimères de chromosomes 50

1.5 Modèles de ségrégation des chromosomes 52

1.6 Structure et mécanisme moléculaire de l'anneau Z 55

1.7 Les systèmes de régulation négative du positionnement de l'anneau Z 58

1.8 Les systèmes de régulation positive du positionnement de l'anneau Z 60

1.9 Distribution taxonomique des systèmes Min, NO et MapZ chez les *Lactobacillales* 60

1.10 Régulation de la dynamique de l'anneau Z chez *E. coli* et *B. subtilis* 62

1.11 Le modèle du divisome chez *E. coli* 64

1.12 Le modèle de l'élongasome chez *E. coli* 66

1.13 Représentation schématique de l'enveloppe cellulaire chez les bactéries à Gram négatif et positif 68

1.14 Structure du peptidoglycane chez *E. coli* 70

1.15 Variabilité du peptidoglycane en acides aminés chez quelques bactéries 71

1.16 Synthèse, maturation et recyclage du peptidoglycane 72

1.17 Classification des PBPs selon Sauvage *et al.*, 2008 74

1.18 Mécanismes cellulaires et moléculaires de la sporulation 78

1.19 Synthèse de la capsule chez *E. Coli* sérotype K30/K40 et *S. pneumoniae* sérotype 2 82

1.20 Le divisome chez *S. Pneumoniae* 85

1.21 Représentativité des bactéries modèles chez les l'ensemble des bactéries 92

2.1 La classification du vivant 97

2.2 Schéma général d'une étude phylogénomique 99

2.3 Quelques exemples d'événements évolutifs 100

2.4 Principe et types de BLAST 103

2.5 Homologie, orthologie, paralogie et xénologie 106

2.6 Modèle général de substitution des acides nucléiques 109

2.7 Illustration du principe des MCMC 113

2.8 Principe de la réconciliation 115

2.9 Principe de prédiction de lien fonctionnel par voisinage génomique 119

2.10 Principe de prédiction de lien fonctionnel par la méthode de la pierre de Rosette 121

2.11 Principe de prédiction de lien fonctionnel par profils phylogénétiques 123

2.12 Trois méthodes de comparaison de profils phylogénétiques 125

2.13 Principe de prédiction de lien fonctionnel par co-évolution 126

2.14 Phylogénie des *Firmicutes* 132

3.1 Composition en *phyla*/domaines de la base de données de procaryotes 136

3.2 Composition en classes/ordres/familles de la base de données des *Firmicutes* . 137

3.3 Phylogénie des *Firmicutes* inférées à partir de protéines ribosomiques 140

3.4 Algorithme de correction des arbres pseudo-ultramétriques 142

3.5 Convergence de la chaîne lors de la datation de la phylogénie des *Firmicutes* . 143

3.6 Phylogénie des familles des *Firmicutes* 144

3.7 Distribution des supports par branche de l'arbre d'espèces des *Firmicutes* . . . 145

3.8 Phylogénie des*Firmicutes* de cette étude 146

3.9 Phylogénie des *Firmicutes* de Antunes *et al.* [11] 147

3.10 Comparaison des deux topologies 149

3.11 Familles de gènes étudiées par processus cellulaire 151

3.12 Contexte génomique des principaux clusters du cycle cellulaire chez *S. pneumoniae* et *B. subtilis* 152

3.13 Pipeline de construction de familles de gènes 153

3.14 Principe de la recherche itérative par BLASTP 156

3.15 Distribution de la taille des familles de gènes 162

3.16 Familles monogéniques : cas de la famille FtsZ 163

3.17 Famille multigénique : le cas simple de la famille GidA 165

3.18 Famille multigénique : le cas modérément complexe de la famille FtsL/DivIC . 166

3.19 Famille multigénique : le cas modérément complexe de la famille FtsA/MreB/MreBH/Mbl.167

3.20 Famille multigénique : le cas complexe de la famille GlmU 168

3.21 Famille multigénique : le cas très complexe de la famille des PBPs type A . . . 169

3.22 Comparaison de deux familles construites par les COGs et dans cette étude . . 170

3.23 Différences de topologies en fonction des programmes d'inférence, le cas de la famille ZapA 174

3.24 Algorithme de reconstruction d'états ancestraux par parcimonie 177

3.25 Principe de la reconstruction des synténies ancestrales 178

3.26 Apparitions inférées par parcimonie 180

3.27 Apparitions inférées par réconciliation 182

3.28 Distances en nœuds des apparitions inférées par les deux méthodes 184

3.29 Pertes inférées par parcimonie 187

3.30 Pertes inférées par réconciliation 189

3.31 Différences d'inférence des pertes entre profils phylogénétiques et réconciliation lié aux grandes branches 190

3.32 Transferts horizontaux inférés par réconciliation 193

3.33 Remplacements homologues inférées par réconciliation 195

3.34 Duplications inférées par réconciliation 197

3.35 Événements de clusters de gènes inférées par parcimonie 200

3.36 Comparaison de la distribution l'information mutuelle et du coefficient ϕ . . . 205

3.37 Distribution des Z-score de corrélation par les profils de présence/absence . . . 214

3.38 Distribution des Z-score de similarité par les événements 215

3.39 Distribution des Z-score de similarité par les synténies 216

3.40 Sélection du seuil de score moyen 218

3.41 Adéquation des valeurs extrêmes avec la GPD 221

3.42 Réseau 1 de relations évolutives entre familles du cycle cellulaire 223

3.43 Réseau 2 de relations évolutives entre familles du cycle cellulaire 225

3.44 Arbre de distances de co-évolution des familles du cycle cellulaire 226

3.45 Histoire évolutive du clan *Bacilli* 236

3.46 Histoire évolutive du clan DCW 241

3.47 Histoire évolutive du clan Spo 244

3.48 Histoire évolutive du clan Mre/Min 247

3.49 Histoire évolutive des clans Ori1 et Ori2 250

3.50 Histoire évolutive du clan Scp/Xer/Smc 254

3.51 Histoire évolutive du clan Cps 256

3.52 Histoire évolutive du clan Fem 258

3.53 Histoire évolutive du clan StkP 259

3.54 Histoire évolutive du clan FtsH 261

3.55 Histoire évolutive du clan Wal 262

3.56 Distribution taxonomique et contexte génomique du clan Nag/mur 264

3.57 Histoire évolutive de BX1 et FtsW 265

3.58 Histoire évolutive de FtsJ et RecN 266

3.59 Histoire évolutive de FtsE, FtsX et MinJ 267

3.60 Distribution taxonomique de MapZ/MurN/M et CDP3 268

3.61 Distribution taxonomique de MurG1 et MurG2 269

3.62 Rôles des familles de gènes impliquées dans la traduction 271

3.63 Distribution taxonomique des systèmes de positionnement de l'anneau Z . . . 275

3.64 Distribution taxonomique des systèmes de résolution de dimères de chromosomes 276

3.65 Nombre de familles retrouvées dans les bases de données SiLiX par seuil d'identité 278

3.66 Distribution des fonction cellulaires générales des FT par FCC 282

4.1 Approche de groupement des domaines PASTA 293

4.2 Phylogénie des domaines PASTA chez les *Lactobacillales* 296

4.3 Phylogénie des domaines PASTA de StkP chez les *Lactobacillales* 297

4.4 Phylogénie des domaines PASTA de StkP chez les *Carnobacteraceae* 298

4.5 Phylogénie des domaines PASTA de StkP chez les *Leuconostocaceae* 299

4.6 Phylogénie des domaines PASTA de StkP chez les *Lactobacillaceae* 300

4.7 Phylogénie des domaines PASTA de StkP chez les *Streptococcaceae* 301

4.8 Phylogénie des domaines PASTA de StkP chez les *Enterococcaceae* 302

4.9 Arbre de distances patristiques moyennes entre groupes de position 304

4.10 Diversité de la composition en domaines de StkP chez *Lactobacillales* 305

4.11 Histoire évolutive des domaines PASTA de StkP chez les *Lactobacillales* 307

4.12 Phylogénie des domaines PASTA chez les *Streptococcaceae* inférée par approche bayésienne 312

4.13 Distribution des distances patristiques par paire issues des arbres bayésiens de chaque type de domaines. 312

4.14 Distribution taxonomique des protéines à domaine Glucosaminidase chez les *Streptococcaceae* 313

4.15 Distribution taxonomique de StkP, de LytB et des résidus impliqués dans l'interaction entre le domaine PASTA C et LytB chez les *Streptococcaceae* 313

5.1 Exemples de figures de contexte génomiques générées par différents programmes 319

5.2 Interface utilisateur de GeneSpy 321

5.3 Diagramme de classes de GeneSpy 322

5.4 Fonctionnement interne de GeneSpy 323

5.5 Gestion des identifiants GeneSpy 325

5.6 Exemples de figures générées par GeneSpy 329

Liste des tableaux

2.1 Taxonomie des *Firmicutes* selon le NCBI 130

3.1 Algorithme de supports composites 139

3.2 Clusters de gènes inférés à l'ancêtre des *Firmicutes* 198

3.3 Méthodes utilisées pour comparer les histoires de familles de gènes 204

3.4 Composition en familles de gènes des clans par le réseau et l'arbre 222

3.5 Liens fonctionnels inférés entre les familles de gènes 231

3.6 Familles impliquées dans la traduction ou interagissant avec les ARN 270

3.7 Correspondance entre la classification COG et les fonction cellulaires générales 280

Avant-propos

Une très grande diversité de morphologies cellulaires est retrouvée au sein du domaine bactérien. Il existe par exemple des bactéries en bâtonnet (*Escherichia coli*), en coque (*Staphylococcus aureus*) mais aussi en hélice (*Leptospira interrogans*), ou encore en forme de gingembre (*Verrucomicrobium spinosum*) (figure 1). La forme cellulaire chez les bactéries dépend en grande partie de l'étape de formation des cellules. En effet, c'est lors de cette phase que la plupart des bactéries produisent la paroi cellulaire, véritable exosquelette permettant de maintenir la forme des cellules et de résister aux contraintes chimiques et mécaniques de l'environnement. L'ensemble des étapes qui permettent la formation de cellules filles à partir d'une cellule mère constitue le cycle cellulaire. Classiquement, la cellule mère rentre d'abord dans une phase de croissance, réplique son génome et répartit une copie dans chaque future cellule fille puis se divise, c'est-à-dire que les deux cellules filles sont individualisées.

Chez les bactéries, il existe plusieurs manière de mener à bien le cycle cellulaire. Certaines espèces comme *Streptomyces* forment un filament qui se compartimentalise en une multitude de cellules filles individuelles. D'autres se divisent de façon asymétrique comme *Caulobacter crescentus/vibrioides*, c'est-à-dire que les deux cellules filles n'ont pas la même morphologie. Néanmoins, parmi les bactéries les plus étudiées, de nombreuses se divisent de facon symétrique (ou fission binaire). Ce type de division conduit à la formation de deux cellules filles identiques. C'est le cas de *Escherichia coli*, *Staphylococcus aureus* ou encore *Bacillus subtilis*. Un autre mécanisme cellulaire qui présente des similitudes avec la division cellulaire classique est retrouvé chez certaines bactéries : la sporulation. Il s'agit d'une division asymétrique qui conduit à la formation d'une structure rigide capable de résister à des conditions extrêmement défavorables : la spore.

FIGURE 1 – Quelques exemples de la diversité morphologique des bactéries. Adapté de [224].

Afin de mener à bien le cycle cellulaire, une grande variété de mécanismes moléculaires ont émergé durant la diversification des bactéries [379]. Ces mécanismes sont à l'origine de la diversité de formes cellulaires et de modes de division. Les mécanismes moléculaires impliqués dans le cycle cellulaire ont été caractérisés principalement chez cinq bactéries modèles : *Escherichia coli*, *Bacillus subtilis*, *Caulobacter crescentus/vibrioides*, *Streptococcus pneumoniae* et *Staphylococcus aureus*. Néanmoins, ces mécanismes identifiés chez ces organismes sont loin de représenter la diversité de tous les mécanismes mis en jeu lors du cycle cellulaire chez les bactéries.

Dans ce contexte, il parait intéressant de connaître l'étendue et les limites des modèles mécanistiques décrits chez ces organismes modèles mais également de déterminer les événements évolutifs majeurs ayant conduit à l'apparition de tels mécanismes. Des mutations ponctuelles, des pertes/gains de gènes ou de domaines, ou encore des réarrangements génomiques peuvent être à l'origine de l'émergence d'un nouveau mécanisme moléculaire. En utilisant une approche de phylogénomique couplant à la fois l'étude évolutive des gènes et leur organisation génomique, il est possible de faire des hypothèse sur comment et quand de tels mécanismes ont émergé [44]. Il est également possible à partir de ces données d'identifier des coévolutions entre les gènes qui peuvent mettre en évidence de potentiels liens fonctionnels [303], [123]. Enfin, les données génomiques permettent d'identifier des gènes dont la fonction n'est initialement pas connue qui pourraient potentiellement être impliquées dans ce processus cellulaire [288].

Nous nous sommes intéressés particulièrement aux *Firmicutes*, un des grands *phyla* bactériens. On y retrouve notamment *Streptococcus pneumoniae*, le modèle d'étude utilisé au sein de mon

laboratoire. Il s'agit d'une bactérie pathogène commensale dont les processus de cycle cellulaire ont été étudiés bien que de nombreux mécanismes restent à découvrir. De façon plus générale, les *Firmicutes* présentent plusieurs intérêts médicaux. Tout d'abord, en raison du fait qu'ils représentent une grande proportion du microbiote intestinal [471], ils sont directement impliqués dans la physiologie de l'être humain. Un déséquilibre dans la composition du microbiote peut ainsi favoriser des pathologies telles que l'obésité ou le diabète [464]. De plus, ce *phylum* contient d'autres pathogènes tels que *Staphylococcus aureus* ou encore *Listeria monocytogenes*. Un certain nombre de souches de *Firmicutes* comme *Enterococcus faecalis* présentent des résistances aux antibiotiques [340] induisant une inefficacité des traitements classiques chez certains patients. Ainsi, l'émergence de bactéries résistantes aux antibiotiques est devenu un problème de santé publique majeur [358]. Nombre de ces antibiotiques ciblent les enzymes de la synthèse de la paroi cellulaire et les systèmes de traduction [152]. Dans ce contexte, il devient nécessaire d'identifier de nouvelles cibles thérapeutiques, c'est-à-dire de cibler de nouvelles fonctions essentielles à la bactérie pour enrayer les infections. Le cycle cellulaire parait être un réservoir potentiel de cibles thérapeutiques puisque nombre des processus liés au cycle cellulaire sont essentiels pour les bactéries [285]. Il apparaît donc qu'une meilleure compréhension du cycle cellulaire chez les *Firmicutes* et l'identification de nouveaux liens fonctionnels au sein de ce processus sont d'un intérêt certain.

Dans cette étude nous avons donc, à travers une approche de phylogénomique, reconstruit l'histoire des gènes du cycle cellulaire chez les *Firmicutes*, inféré de nouveaux liens fonctionnels entre les protéines appartenant aux machineries associées et identifié des protéines potentiellement impliquées dans ce processus. Nous décrirons dans un premier chapitre les différentes machineries protéiques impliquées dans le cycle cellulaire. Dans le second chapitre, nous aborderons les méthodes phylogénomiques développées à ce jour afin d'annoter les génomes et d'identifier des potentiels liens fonctionnels entre gènes. Nous décrirons aussi le phyum des *Firmicutes*. Dans le troisième chapitre, nous détaillerons l'histoire évolutive d'un grand nombre de gènes impliqués dans le cycle cellulaire, les liens évolutifs observés entre ces gènes et leurs implications fonctionnelles. Nous présenterons aussi des familles de gènes potentiellement impliquées dans le cycle cellulaire, identifiées par co-évolution. Le cas spécifique de l'histoire des domaines fonctionnels de la famille StkP sera abordé en chapitre quatre. Dans le cinquième chapitre, nous détaillerons

l'implémentation du logiciel GeneSpy développé durant ma thèse. Enfin, nous discuterons des résultats globaux de ma thèse et des perspectives.

Chapitre 1

Le cycle cellulaire bactérien

Sommaire

1.1 Le cycle du chromosome 44

1.1.1 Structure du chromosome bactérien 44

1.1.2 La réplication du chromosome bactérien 45

1.1.3 La ségrégation des chromosomes fils 50

1.2 Le divisome 54

1.2.1 FtsZ, la protéine centrale de la division cellulaire 54

1.2.2 Régulation du positionnement de l'anneau Z 56

1.2.3 Régulation de la dynamique de l'anneau Z 61

1.2.4 Protéines du divisome associées à FtsZ 63

1.3 L'élongasome 65

1.3.1 MreB, protéine centrale de l'élongasome 65

1.3.2 Autres composants de l'élongasome 67

1.4 Le peptidoglycane, composant de la paroi cellulaire bactérienne . 68

1.4.1 Rôle et structure du peptidoglycane 68

1.4.2 Étapes cytosoliques et membranaires de la biosynthèse 69

1.4.3 Étapes extracellulaires de la biosynthèse : les PBPs 73

1.4.4 Remodelage du peptidoglycane 74

1.4.5 Recyclage du peptidoglycane 76
1.5 La sporulation 77
1.5.1 Généralités 77
1.5.2 Mécanismes moléculaires de la sporulation 79
1.6 La capsule polysaccharidique 80
1.6.1 Rôle et structure de la capsule 80
1.6.2 Synthèse de la capsule 81
1.7 La coordination des machineries du cycle cellulaire 83
1.7.1 Coordination élongation-division : MreB et FtsZ 83
1.7.2 StkP, métarégulateur du cycle cellulaire 84
1.7.3 Coordination division-ségrégation chromosomique 87
1.7.4 Régulation de la synthèse du peptidoglycane 87
1.7.5 Autres coordinations entre machineries 88
1.8 Limites des modèles du cycle cellulaire bactérien 89

Le cycle cellulaire chez les bactéries peut être segmenté en plusieurs étapes clés [121]. Nous décrirons ici les grandes étapes générales qui s'appliquent à la majorité des cas et particulièrement aux bactéries qui se divisent par fission binaire. Le schéma général du cycle cellulaire est présenté en figure 1.1.

Tout d'abord, la bactérie est dans une phase dite végétative. Elle commence par entamer la phase de réplication qui consiste en une duplication de l'ADN afin d'obtenir deux copies identiques du génome. La cellule met ensuite en place une structure essentielle à la division cellulaire : l'anneau de constriction Z. Cette structure est principalement composée de la protéine FtsZ et va servir d'échafaudage à un grand nombre de protéines impliquées dans la division cellulaire [305]. Chez les bactéries à division symétrique, l'anneau Z se place au milieu de la cellule. L'anneau Z et les protéines recrutées forment le divisome, une machinerie complexe dont le but principal est de mener à bien la septation ou division cellulaire [121]. La cellule assume simultanément la ségrégation chromosomique pendant laquelle les chromosomes migrent vers les futures cellules filles. La cellule s'allonge également afin que la longueur des deux futures cellules filles soit du même ordre de grandeur que la cellule mère initiale. Enfin, la septation ou division cellulaire s'enclenche et l'anneau Z se contracte séparant ainsi les deux cellules filles. Ces phases très générales ont été déduites à partir d'études chez plusieurs organismes, mais les processus varient énormément d'une bactérie à l'autre [379], [472], [329], [158]. De plus, l'ensemble de ces processus est finement régulé spatialement et temporellement par de nombreux systèmes. Par exemple, le positionnement de l'anneau Z est régulé par une très grande variété de systèmes, positivement ou négativement en fonction des organismes [328]. Autre exemple, les machineries moléculaires responsables de l'élongation et de la division dont la localisation cellulaire est distincte chez *E. coli* et *B. subtilis* peuvent co-localiser au site de division comme chez *S. pneumoniae* [426].

Nous décrirons ainsi de façon succincte les connaissances actuelles sur les processus associés au cycle cellulaire décrits chez les bactéries modèles : La réplication, la ségrégation chromosomique, la division cellulaire, l'élongation et la synthèse de peptidoglycane. Nous décrirons également les processus cellulaires reliés de façon moins ténue au cycle cellulaire que sont la sporulation et la synthèse de la capsule. Nous décrirons ensuite les moyens de coordination entre l'ensemble de ces processus, puis nous aborderons les limites des modèles moléculaires

actuellement établis.

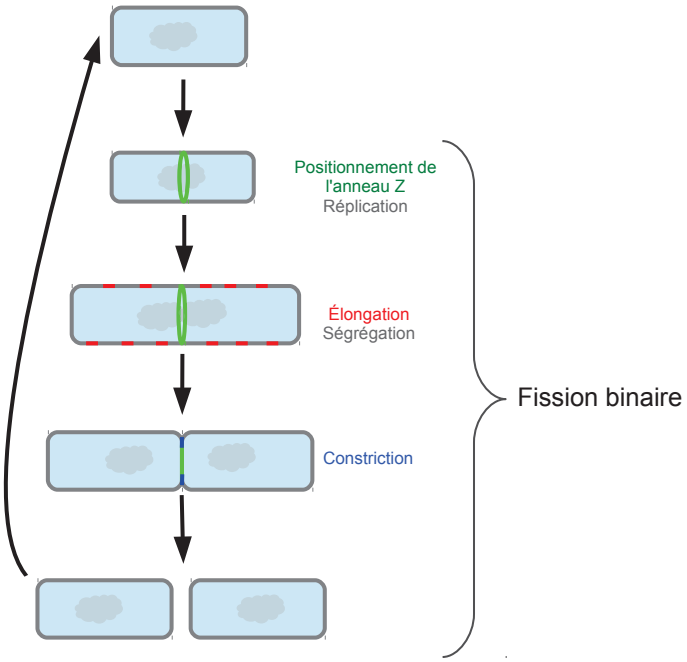


FIGURE 1.1 – Le cycle cellulaire bactérien et la fission binaire.

1.1 Le cycle du chromosome

1.1.1 Structure du chromosome bactérien

Lorsque la cellule est en phase végétative, le chromosome forme une structure opaque et irrégulière appelée le nucléoïde [403] (figure 1.2). Le chromosome est souvent unique, circulaire et organisé de façon intrinsèque en différents macro-domaines : le domaine Ori, les deux domaines droit et gauche et le domaine Ter [19] (figure 1.2). Le domaine Ori contient le site d'initiation de la réplication, la région *oriC* tandis que le domaine Ter contient les séquences d'arrêt de la réplication ainsi que le site *dif* permettant la recombinaison en cas de dimères de chromosomes (Cf 1.1.2.5).

Ces régions sont localisées dans la cellule de façon précise par une multitude de protéines et

présentent également un niveau de compactage finement régulé. Certaines protéines régulent une zone spécifique. Par exemple chez *E. coli*, la protéine MatP intervient dans le compactage spécifique de la région Ter [317]. D'autres protéines régulent l'état de compactage du chromosome de façon non spécifique. Ces protéines sont appelées NAPs (« Nucleoid-associated proteins ») [95]. Chez *E. coli*, la protéine H-NS a par exemple été décrite comme étant capable de rapprocher des régions génomiques linéairement distantes [178]. La protéine HU est quant à elle impliquée dans le compactage global de l'ADN en jouant un rôle similaire aux histones chez les eucaryotes [95]. Il est intéressant de noter que l'état de compactage a un effet drastique sur la transcription des gènes [95]. En effet, les gènes présents dans les zones décompactées du chromosome seront plus facilement transcrits car étant plus accessibles à la machinerie de transcription.

FIGURE 1.2 – Structure du chromosome bactérien. (A) Chromosome de *E. Coli* formant le nucléoïde, observé en microscopie électronique à transmission. Adapté de http ://bioloevelyn.blogspot.com. (B) Représentation schématique du chromosome de *E. Coli*.

1.1.2 La réplication du chromosome bactérien

1.1.2.1 Généralités

La réplication correspond à la duplication du chromosome de la cellule mère afin de fournir une copie pour chacune des deux futures cellules filles. Nous décrirons ici principalement les mécanismes moléculaires de la réplication observés chez *E. coli* qui sont les plus étudiés et compris. On peut distinguer principalement trois grandes phases en figure 1.3 : l'initia-

tion, l'élongation et la terminaison. Il est à noter que d'importants travaux ont également été effectués chez *Bacillus subtilis* laissant apparaître certaines différences [218]. Des études ont également été réalisées chez d'autres bactéries contenant non pas un mais deux chromosomes comme *Vibrio cholerae* et dans lesquelles des mécanismes particuliers, que nous ne détaillerons pas ici, ont été rapportés.

1.1.2.2 Phase d'initiation de la réplication

La réplication est initiée par la fixation de la protéine DnaA sur les « Dna-box », séquences situées dans la région *oriC* du chromosome (figure 1.2, 1.3A). DnaA se lie ensuite à un ATP enclenchant ainsi la séparation des deux brins d'ADN [232]. Le complexe DnaA-ADN recrute ensuite les protéines DnaB (hélicase) et DnaC (chargeur d'hélicase) qui se fixent alors sur chaque brin [232]. Le complexe DnaA-DnaB-DnaC forme la fourche de réplication en déconcatenant l'ADN. Il s'agit du point de rupture entre la région ou l'ADN est encore en double hélice et celle où les brins sont dissociés.
Afin d'éviter qu'une seconde initiation ne s'enclenche, il existe des systèmes régulant la fixation de DnaA sur l'origine de réplication. Le premier implique la protéine SeqA qui séquestre DnaA et inhibe ainsi sa fixation sur le chromosome par compétition [452]. Le deuxième système met en jeu la protéine Hda, un homologue de DnaA qui favorise l'hydrolyse de l'ATP lié à DnaA via l'interaction avec DnaN [237], [57]. DnaA ne peut ainsi plus séparer les deux brins d'ADN. Il est intéressant de noter que chez *B. subtilis*, DnaI joue le rôle de DnaC et que l'initiation nécessite deux protéines additionnelles qui sont absentes chez *E. coli*, DnaB et DnaD. Néanmoins, leur rôle respectif n'est pas encore élucidé [218]. Aussi, SeqA et Hda sont absents chez *B. subtilis* mais la protéine YabA semble remplir une fonction similaire à Hda en interagissant avec DnaN bien que le mécanisme précis ne soit pas encore élucidé [66],[191].

1.1.2.3 Phase d'élongation

Une fois l'initiation de la réplication effectuée, la fourche progresse le long du chromosome de façon bidirectionnelle [387] (figure 1.3B). La fourche de réplication est asymétrique, c'est-

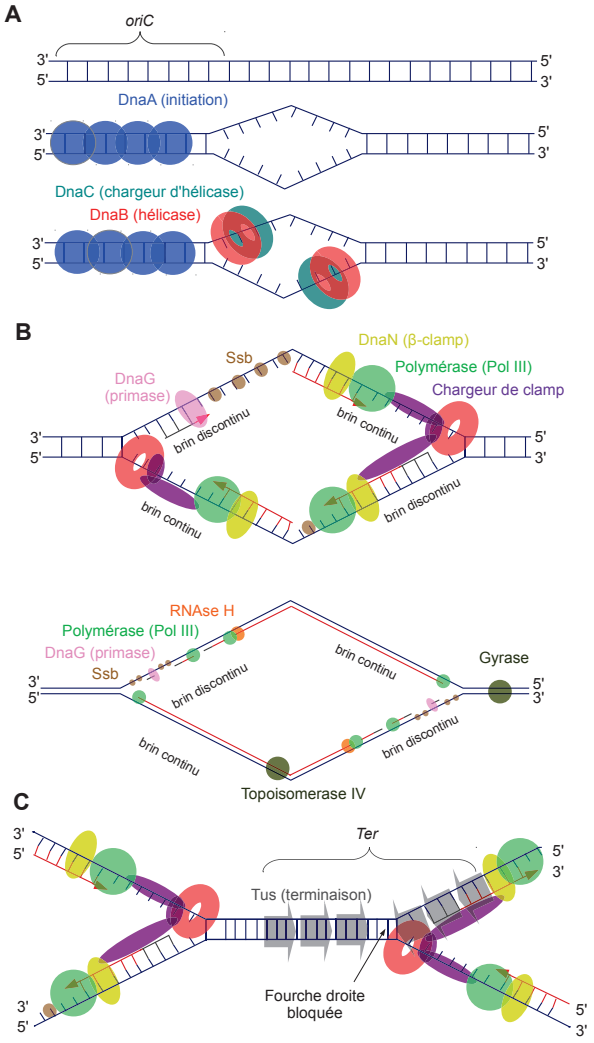


FIGURE 1.3 – Modèle de réplication du chromosome chez *E. Coli*. (A) Initiation de la réplication. DnaA se fixe sur *oriC* puis recrute DnaB et DnaC. (B) Phase d'élongation de la réplication. Les deux fourches de réplication progressent. L'hélicase DnaB sépare les brins d'ADN, le chargeur de clamp recrute DnaN et la polymérase III. Pour les deux brins discontinus, DnaG synthétise des amorces ARN, Pol III synthétise l'ADN à partir de l'amorce puis la RNAse H digère l'amorce suivante afin que la polymérase III synthétise l'ADN à la place. (C) Phase de terminaison. La fourche de droite est bloquée par les protéines Tus de gauche tandis que la fourche de gauche continue à progresser. L'arrêt s'effectue à la rencontre des deux fourches.

à-dire qu'un brin est répliqué de façon continue (brin continu) tandis que l'autre est répliqué de façon discontinue (brin discontinu). Ceci est lié au fait que la polymérisation de l'ADN se fait uniquement dans le sens $5'->3'$.
Cette phase de la réplication nécessite le recrutement de plusieurs protéines [399]. Tout d'abord, l'hélicase DnaB permet de séparer les deux brins d'ADN et mène donc la progression de la fourche de réplication. Ensuite un échafaudage moléculaire du réplisome, appelé « chargeur de clamp » est constitué. Il permet le recrutement de la polymérase III et de son facteur de processivité, DnaN ou « β-clamp ». Ce complexe peut ainsi synthétiser le brin continu.
Pour le brin discontinu, plusieurs étapes et protéines supplémentaires sont nécessaires. Tout d'abord, il est nécessaire de synthétiser une courte amorce d'ARN qui va permettre d'amorcer la synthèse de courts fragments d'ADN appelés fragments d'Okasaki [356]. Les amorces d'ARN sont synthétisées par la primase DnaG et les brins d'ADN par la polymérase III. Plusieurs fragments d'Okasaki sont synthétisés le long du brin discontinu (figure 1.3B). Les amorces sont ensuite digérées par la RNAse H puis remplacés par de l'ADN par la polymérase III.
La réplication induit une instabilité structurale de l'ADN ; non seulement parce que certaines portions sont désappariées tant que la polymérase n'a pas synthétisé le brin complémentaire mais aussi parce que la progression de la fourche de réplication induit des tensions physiques dues à un sur-enroulement. La protéine Ssb se place ainsi le long des portions d'ADN simple brin afin de les stabiliser (figure 1.3B). Pour diminuer les tensions dues au sur-enroulement, deux enzymes sont mises en jeu : la topoisomérase IV en amont de la fourche et l'ADN gyrase en aval [383].
Il arrive que la fourche de réplication soit stoppée ou bloquée. Il est alors nécessaire de réamorcer la réplication à l'aide de la protéine PriA en association avec DnaB et DnaC [195].

1.1.2.4 Phase de terminaison

La réplication est stoppée lorsque les fourches de réplication ont parcouru l'intégralité du chromosome et se rejoignent à la région Ter du chromosome (la zone opposée à l'*oriC*) (figure 1.3). Cet arrêt est rendu possible par la présence au niveau de la région Ter de séquences particulières permettant de recruter la protéine Tus chez *E. coli* qui agit comme bloqueur de la

progression de la fourche de réplication [346]. Les séquences de la région Ter mises en jeu sont polaires, c'est-à-dire qu'elles permettent de laisser progresser la fourche de réplication dans un sens mais la bloquent dans l'autre. Ce mécanisme permet à la machinerie de réplication de parcourir l'ensemble du chromosome et que la rencontre des deux fourches s'effectue systématiquement dans la région Ter [201]. Il est intéressant de noter que chez *B. subtilis*, Tus est absent mais qu'une autre protéine, la protéine Rtp, semble jouer un rôle comparable [433].

1.1.2.5 Réparation de l'ADN

L'existence d'un dommage à l'ADN lors de la réplication est très probable à cause de différents stress pouvant être rencontrés par la cellule au cours de sa croissance (rayons UV, agression chimique, ..). Il est donc nécessaire de réparer l'ADN par recombinaison homologue, c'est-à-dire en utilisant le brin complémentaire pour reformer le brin cassé. Ce processus est géré par le complexe de recombinaison contenant notamment la protéine initiatrice RecA, ainsi qu'une myriade de protéines (RecBCDEFGJNORQTU, RuvABC, SbcCD, ...) [250].
Après réparation, il est nécessaire de résoudre les jonctions entre les deux brins d'ADN. RuvABC et RecGU résolvent ces jonctions mais produisent des monomères (sans chevauchement) et des dimères de chromosomes (avec chevauchement) [202] (figure 1.3). Les dimères de chromosome sont létaux pour la cellule et doivent absolument être résolus. Chez *E. coli*, les recombinases XerC et XerD en association avec la protéine de ségrégation FtsK permettent de résoudre ces dimères durant la ségrégation de la région terminale du chromosome [71],[177] (Cf 1.1.3). Tout d'abord, XerCD sont activées par FtsK puis se fixent sur le site *dif* localisé dans la région Ter. Elles clivent ensuite la liaison phosphodiester de deux brins pour les rabouter puis catalysent une réaction similaire pour les deux brins restants [61] (figure 1.3B). De façon intéressante, on retrouve XerCD chez *B. subtilis* [419] mais pas chez textitS. pneumoniae qui ne possède qu'une seule recombinase : XerS [265].

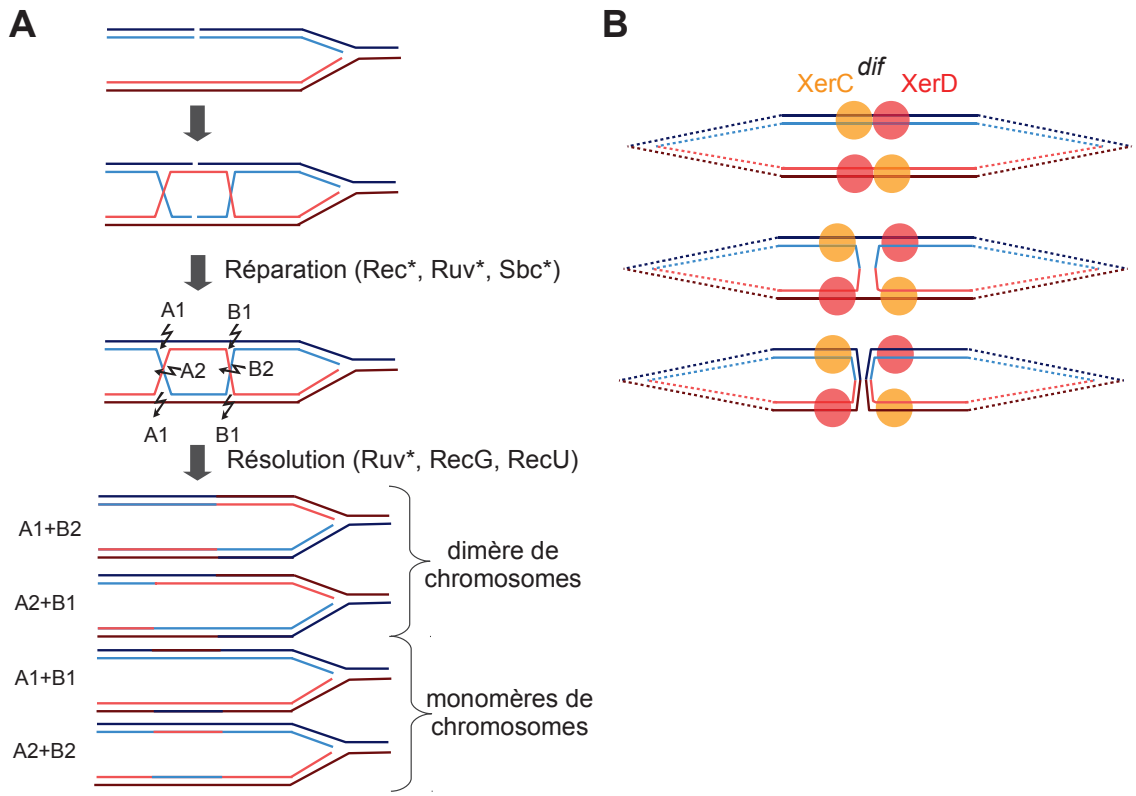


FIGURE 1.4 – Réparation de l'ADN et résolution des dimères de chromosomes. (A) Réparation des cassures double brin. Après cassure double brin de l'ADN, les brins d'ADN néoformés sont échangés au niveau du site de coupure puis la cassure est réparée. Afin de résoudre cette configuration, quatre possibilités de coupures sont possibles, A1, A2, B1, B2. La combinaison de ces coupures peut induire la formation de dimères de chromosomes. (B) Résolution des dimères de chromosomes par les recombinases XerC et XerD. XerC et XerD sont d'abord recrutés au site *dif* puis le dimère de chromosomes est séparé en deux monomères.

1.1.3 La ségrégation des chromosomes fils

1.1.3.1 Généralités

De façon concomitante à la réplication, les chromosomes nécessitent d'être ségrégés dans les deux cellules filles respectives. Contrairement aux eucaryotes, le chromosome n'est pas guidé par un fuseau mitotique [485]. Il a tout d'abord été postulé que le chromosome migrait de façon

passive grâce à son ancrage à la paroi cellulaire de façon concomitante à l'allongement cellulaire [215]. Des travaux plus récents ont alors démontré que le chromosome migrait plus vite que l'allongement cellulaire, indiquant ainsi qu'il existait une machinerie moléculaire dédiée [485],[496]. De façon générale, plusieurs systèmes ont émergé chez les bactéries pour permettre la ségrégation des chromosomes. Il apparaît qu'au sein d'une même cellule, ce processus implique aussi bien des facteurs protéiques que des phénomènes purement mécaniques [19].

1.1.3.2 Ségrégation de l'origine : système ParABS

Ce système a été décrit initialement comme étant impliqué dans la ségrégation des plasmides [16]. Pour la ségrégation chromosomique, il est composé de deux protéines ParA (Soj) et ParB (Spo0J) et de sites *parS* localisés sur le chromosome au niveau de la région *oriC* (figure 1.5). ParB est une protéine à motif hélice-tour-hélice qui se fixe en dimère de façon spécifique sur le site *parS*. Un grand nombre de dimère se fixe sur chaque site *parS* formant ainsi un imposant complexe protéine-ADN [176]. L'ATPase ParA est présente dans le cytoplasme et se fixe à l'ADN de façon non spécifique [198]. L'activité ATPase de ParA permettrait alors la séparation des deux *oriC* dupliquées.
Ce mécanisme reste encore peu compris mais chez certaines bactéries, comme *Vibrio cholerae*, il a été proposé que ParA forme des filaments dynamiques qui se fixent à ParB et permettent la ségrégation de ParB et de la région *oriC* du chromosome vers les pôles par phénomène de traction [151]. ParB, en se fixant sur les filaments de ParA, stimulerait l'activité ATPase ce qui aurait pour conséquence de rétracter le filament et de donner ainsi une force motrice directionnelle au complexe. Une hypothèse alternative propose que le mouvement directionnel soit mené par l'élasticité de l'ADN notamment chez *Caulobacter crescentus*/vibrioides [281].
La ségrégation chromosomique par le système ParABS a été décrit initialement chez *Vibrio cholerae* [151] puis chez un grand nombre d'espèces comme *Mycobacterium tuberculosis* [23], *Myxococcus xanthus* [214] ou *Bacillus subtilis* [341] . Néanmoins, ce système n'est pas présent chez *E. coli* et n'a été montré comme essentiel que chez *C. crescentus/vibrioides* [467]. Chez *Bacillus subtilis*, ParB (Spo0J) est aussi impliqué dans la sporulation [512].

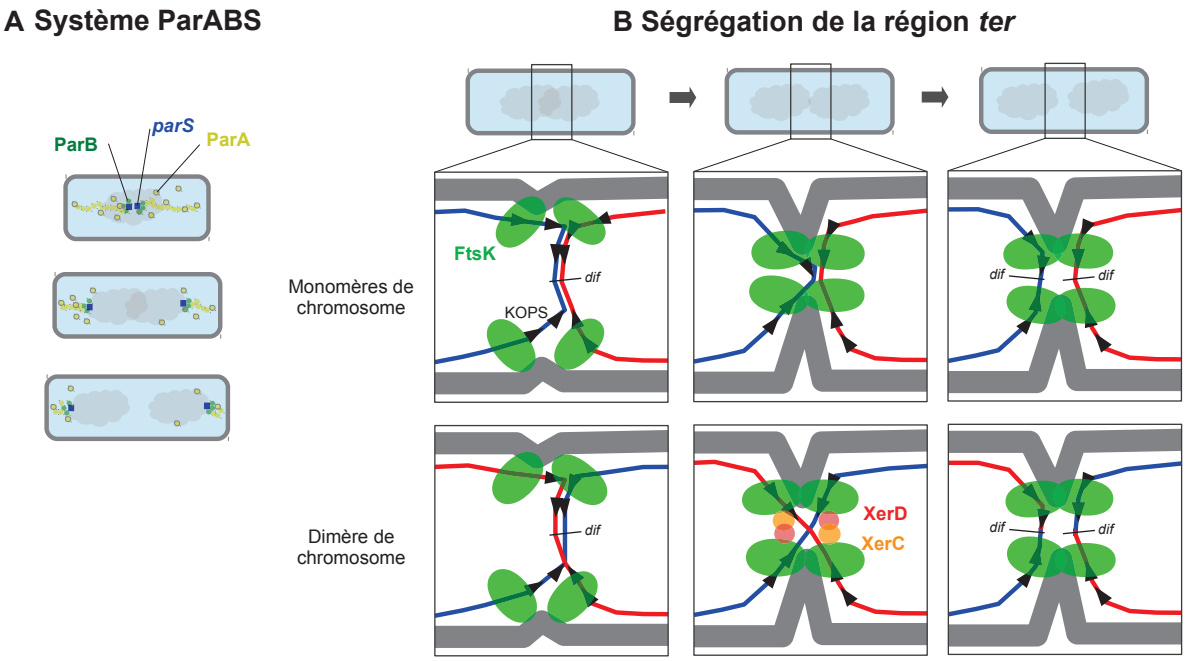


FIGURE 1.5 – Modèles de ségrégation des chromosomes. (A) Modèle mécanistique du système ParABS proposé chez *Vibrio cholerae*. ParB se fixe aux séquences *parS* puis ParA entraîne ParB aux pôles. (B) Ségrégation de la région Ter. FtsK se fixe aux séquences KOPS et pompe le chromosome vers les deux futures cellules filles. Les dimères de chromosomes sont résolus par les recombinases XerC et XerD.

1.1.3.3 Ségrégation de l'hétérochromatine

Après l'origine de réplication, le reste du chromosome (hétérochromatine ou « bulk chromosome ») nécessite aussi d'être ségrégué. Ainsi, deux systèmes ont été démontrés. Chez *E. coli* ainsi que d'autres *Proteobacteria*, ce système est composé des protéines MukBEF. Bien que le mécanisme exact de contribution de ce système à la ségrégation ne soit pas encore élucidé, plusieurs études ont montré le rôle de MukBEF dans ce processus. Tout d'abord, la délétion de l'ATPase MukB induit la désorganisation de la structure habituelle du chromosome durant la ségrégation résultant en un nombre anormal de cellules anuclées [501]. De plus, le complexe MukBEF est recruté au niveau de la région *Ori* [78]. Aussi, il a été montré que MukBEF sti-

mule l'activité de la topoisomérase IV et contribue ainsi à la séparation des chromatides sœurs [275]. Le système Muk semble néanmoins restreint aux *Proteoacteria* [351].
Chez *B. subtilis*, un système qui serait similaire au système MukBEF a été décrit. Il s'agit du système Smc/ScpAB dans lequel l'ATPase Smc jouerait un rôle identique à celui de MukB. La délétion de Smc entraîne également des défauts de ségrégation chromosomique et Smc est aussi recruté au site d'origine de réplication [50],[181]. Smc semble promouvoir la condensation du chromosome et contribuer ainsi à augmenter l'efficacité de la ségrégation chez *B. subtilis* [180]. La délétion de ScpA ou ScpB chez *Bacillus subtilis* induit un phénotype similaire à celui d'un mutant Smc mais leur rôle respectif reste encore peu compris [435]. Smc est aussi retrouvé chez *S. pneumoniae* et semble impliqué dans la ségrégation chromosomique bien qu'il ne soit pas essentiel [324]. De façon intéressante, Smc a été démontré comme étant recruté par ParB chez cet organisme [324]. Le système Smc-ScpA-ScpB est absent chez les *Gammaproteobacteria* comme *E. coli* [351].

1.1.3.4 Ségrégation de la région terminale

La région terminale subit également une ségrégation active, principalement menée par la protéine FtsK (figure 1.5). Cette protéine a été étudiée principalement chez *E. coli*. FtsK est une translocase d'ADN qui comporte plusieurs activités et qui est localisée au site de division (milieu de la cellule) [36],[269]. Tout d'abord, elle stimule l'activité de la topoisomérase IV ce qui permet de faciliter la déconcaténation des deux copies de chromosome lors de la réplication [133]. Elle agit aussi comme une pompe à ADN pour ségréguer définitivement les régions Ter des deux chromatides sœurs et active le complexe XerCD qui résout les dimères de chromosomes [270], [445]. Pour jouer son rôle de pompe, FtsK fixe spécifiquement des séquences particulières appelées KOPS (« FtsK Orienting Polar Sequences ») situées tout le long du chromosome (mais surtout à la région Ter) et orientées vers le site *dif* permettant ainsi un déplacement directionnel des deux chromosomes vers leur cellule fille respective [35]. Chez *B. subtilis*, FtsK est nommé SpoIIIE et n'est pas impliqué dans la ségrégation chromosomique classique. Elle permet la ségrégation chromosomique lors de la sporulation en association avec la protéine RacA [511]. Néanmoins, un paralogue de SpoIIIE, SftA, est impliqué dans la ségrégation du chromosome

lors des phases végétatives [37].

1.1.3.5 Phénomène physique de la ségrégation chromosomique

Certaines études proposent qu'à la multitude de systèmes protéiques impliqués dans la ségrégation chromosomique s'ajoutent des phénomènes purement physiques. Un des phénomènes mécaniques qui pourrait être à l'origine de la ségrégation des chromosomes chez *E. coli* a été proposé par Kleckner et ses collaborateurs [245]. L'hypothèse consiste à expliquer la ségrégation par le fait que la séparation des deux chromatides sœurs induit une réorganisation de la structure du chromosome. Au début de la réplication, les origines de réplication des deux chromatides sœurs se condensent et restent cohésives. L'ADN condensé avec les protéines associées à l'ADN forment des agrégats. L'accumulation de ces corps agrégés augmente les tensions internes entre les deux origines à cause de l'encombrement stérique. A partir d'une certaine taille seuil, les forces internes sont tellement élevées que les corps agrégés se séparent violemment, dans deux directions opposées. C'est cette séparation brutale qui causerait en partie la migration des origines de réplication vers les pôles.

1.2 Le divisome

1.2.1 FtsZ, la protéine centrale de la division cellulaire

1.2.1.1 Généralités

La division cellulaire bactérienne qui consiste en la séparation des deux cellules filles est exécutée par le divisome. La protéine centrale du divisome est FtsZ (« filamentous temperature sensitive Z »), une protéine similaire structuralement aux protéines de type tubuline des eucaryotes [350]. FtsZ polymérise de façon GTP-dependante [336] pour former des protofilaments [128] qui s'assemblent en un anneau au site de division, l'anneau Z (figure 1.6A) [276]. Néanmoins la structure annulaire est actuellement remise en question [220]. Historiquement, c'est la

première protéine cytosquelettique procayotique a avoir été découverte [85] et probablement la plus étudiée. FtsZ joue le rôle d'échafaudage pour les autres protéines constituant le divisome en recrutant de nombreuses protéines et est de ce fait une des premières protéines à localiser au site de division [472]. La polymérisation de FtsZ et la formation de l'anneau Z sont régulées par un grand nombre d'interactants protéiques.

FIGURE 1.6 – Structure et mécanisme moléculaire de l'anneau Z. (A) Structure de l'anneau Z chez *E. coli* par microscopie à fluorescence. Tiré de [164]. (B) Structure de la protéine FtsZ. (C) Modèle du protofilament de la protéine FtsZ à partir de la structure de FtsZ de *P. aeruginosa*. Les protéines FtsZ s'assemblent de façon « head-to-tail ». Domaine globulaire : bleu et cyan, domaine C-terminal : magenta. Adapté de [127]. (D) Modèle du tapis roulant. FtsZ est polymérisé à une extrémité des protofilaments ancrés à la membrane par FtsA et dépolymérisé de l'autre. La machinerie de synthèse du peptidoglycane est entraînée par ce mouvement et forme des couches concentriques.

1.2.1.2 Structure de FtsZ et dynamique d'assemblage de l'anneau Z

FtsZ est organisé en 3 domaines fonctionnels : un domaine globulaire, un linker et un domaine C-terminal (figure 1.6B). L'activité GTPase est portée par le domaine globulaire [357] et la régulation de la polymérisation de FtsZ par ses partenaires est effectuée par le linker et le domaine C-terminal [294],[126]. La fixation du GTP au domaine globulaire induit la polymérisation de monomères de façon « head-to-tail » [357],[336] et lorsque le GTP est hydrolysé en GDP, les protofilaments sont désassemblés [337] (figure 1.6C). Cet équilibre entre les formes sans GTP ou GDP, avec GTP et/ou GDP induit une grande plasticité de la structure

de l'anneau et un « turnover » des monomères élevé [7].

La structure en anneau, d'abord présentée comme une structure continue [495] a été décrite par de récentes études comme étant discontinue et formant des patchs [220],[446]. Plusieurs modèles ont été proposés pour expliquer la constriction de l'anneau Z. Initialement, il a été postulé que les protofilaments glisseraient les uns sur les autres et créeraient ainsi un mouvement de constriction [205],[206]. L'anneau Z a ensuite été décrit comme discontinu et une nouvelle hypothèse sur le mécanisme de sa construction a été proposée. Les protofilaments formant des patchs oscilleraient entre un état droit et courbé de par l'alternance entre fixation du GTP et GDP, induisant ainsi des constrictions locales [276],[360]. La somme de l'ensemble des constrictions générerait la constriction de l'anneau. Néanmoins, le modèle faisant autorité actuellement est celui du tapis roulant (« treadmilling ») [39], [517] (figure 1.6D). Dans ce modèle, les protofilaments de FtsZ sont polymérisés à une extrémité et dépolymérisés de l'autre. Cette dynamique induit un déplacement le long des filaments de la protéine FtsA (cf 1.2.3) et entraîne ainsi la machinerie de synthèse du peptidoglycane. Le peptidoglycane est donc synthétisé en couches concentriques ce qui a pour effet de diminuer le diamètre du septum induisant ainsi la septation. De façon intéressante, l'origine de la force motrice semble différer selon les organismes. Chez *B. subtilis*, le déplacement de FtsZ semble générer la force motrice [39] tandis que chez *S. aureus*, c'est la machinerie de synthèse du peptidoglycane qui fourni majoritairement la force nécessaire à la contraction de l'anneau [331].

1.2.2 Régulation du positionnement de l'anneau Z

Le positionnement de l'anneau Z et du divisome est considéré comme la première étape de la division cellulaire [305]. Il permet de définir un plan de division ou va s'effectuer la séparation finale entre les deux cellules filles. Il existe plusieurs systèmes moléculaires qui permettent de définir la localisation du site de division. Ces systèmes sont très variés en composition et en fonctionnement selon les bactéries. En effet, certains agissent de façon négative sur l'anneau Z c'est-à-dire en inhibant la polymérisation de FtsZ tandis que d'autres régulent de façon positive en jouant le rôle de balise moléculaire au site de division [158]. Nous allons décrire ici les grands

systèmes identifiés à ce jour.

1.2.2.1 Régulation négative

Chez *E. coli*, deux systèmes ont été décrit comme régulant le positionnement de FtsZ, tous deux de façon négative (figure 1.7). Le premier de ces systèmes est le système Min (« mini-cells ») composé des protéines MinCDE [397]. La protéine MinD, d'abord sous sa forme libre, fixe l'ATP et dimérise. La forme dimerisée de MinD se localise aux pôles puis recrute MinC [207], un inhibiteur de la polymérisation de FtsZ [75]. Le dernier élément du système est la protéine MinE qui stimule l'activité ATPase de MinD. Elle forme un anneau à proximité du septum qui se rapproche progressivement du pôle [255],[207],[188]. Le contact avec l'anneau MinE détache le complexe MinCD de la membrane qui va alors se reformer à l'autre pôle [188]. L'anneau E se reforme ensuite de l'autre coté de la cellule. Cette oscillation crée un gradient de MinC au sein de la cellule, le plan médian étant l'endroit ou la probabilité de présence de MinC est la plus faible [397]. L'anneau Z ne peut alors que se former au niveau du plan médian. De façon concomitante, l'anneau Z est régulé par un deuxième système de régulation négative : le système NO pour "nucleoid occlusion" [338]. Chez *E. coli*, il est composé de la protéine SlmA, un inhibiteur de la polymérisation de FtsZ qui se fixe de manière non spécifique sur le chromosome [33]. La septation est inhibée tant que les chromosomes sont au milieu de la cellule laissant ainsi la possibilité de fermer le septum uniquement lorsque les chromosomes ont migré aux pôles. Ce système permet d'éviter l'effet "guillotine" de l'ADN, phénomène qui consiste en la fermeture prématurée du septum coupant ainsi le nucléoïde en deux [230].
Chez *B. subtilis*, deux systèmes similaires ont été décrit mais présentent des composants protéiques différents (figure 1.7). Le système Min est composé de MinCD mais MinE est remplacé par les protéines DivIVA et MinJ [120],[366]. DivIVA est tout d'abord recruté aux pôles par sa capacité à reconnaître les courbures de la membrane [63],[268]. MinJ fait ensuite l'intermédiaire entre DivIVA et MinD. MinD se fixe alors à la membrane de façon ATP-independante puis recrute MinC [48],[306]. Aucune oscillation n'est décrite chez *B. subtilis*. Néanmoins, MinCD semble être recruté au niveau du septum afin d'empêcher la formation d'un potentiel deuxième anneau Z [179]. Le système NO chez *B. subtilis* est similaire à celui de *E. coli* mais ne fait pas

intervenir la même protéine. Il s'agit de Noc, un homologue de ParB qui ne présente aucune similarité de séquence avec SlmA [513],[230]. Une autre différence avec SlmA est que Noc ne semble pas interagir avec FtsZ directement, suggérant ainsi un autre mode d'action [230].

Un autre système de régulation négative de l'assemblage de l'anneau Z a été décrit chez *Caulobacter crescentus/vibrioides*. Dans ce système, la protéine MipZ qui est homologue à ParA et MinD joue un rôle d'inhibiteur de la polymerisation de FtsZ [462],[243] (figure 1.7). MipZ interagit avec ParB interagissant lui même avec l'origine de réplication. Au contact de ParB, MipZ fixe de l'ATP et dimérise, ce qui confère sa propriété d'inhibition de l'assemblage de l'anneau Z. L'hydrolyse de l'ATP provoque la dissociation des dimères de MipZ pouvant ainsi de nouveau interagir avec ParB. Lors de la ségrégation des chromosomes, ParB se localise majoritairement aux pôles ce qui crée un gradient décroissant de MipZ dimérisé des pôles vers le centre de la cellule.

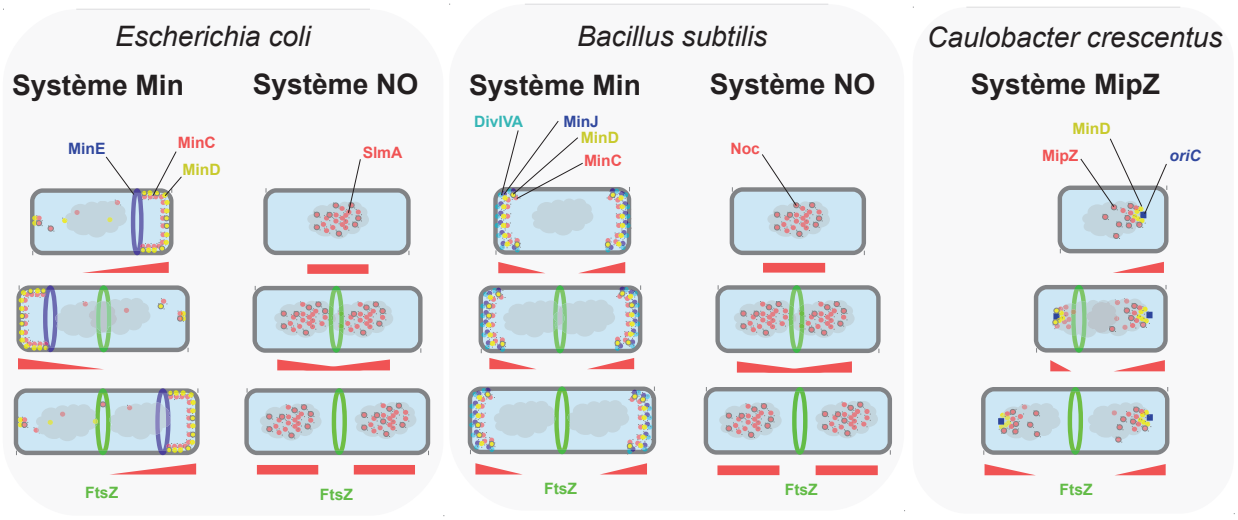


FIGURE 1.7 – Les systèmes de régulation négative du positionnement de l'anneau Z chez *E. coli*, *B. subtilis* et *C. crescentus/vibrioides*. Les inhibiteurs de la formation/contraction de l'anneau Z sont représentés en rouge. Le gradient d'inhibition est représenté par des triangles/rectangles rouge.

1.2.2.2 Régulation positive

Il existe aussi des systèmes de régulation positive du positionnement de l'anneau Z chez différentes bactéries. Ces systèmes sont composés de protéines agissant comme de véritables balises moléculaires pour la protéine FtsZ et indiquant le site de division.
Chez *Streptomyces coelicolor*, le système SsgA/SsgB permet de localiser l'anneau Z (figure 1.8). SsgA est une protéine qui se localise au niveau des lieux de remodelage du peptidoglycane puis recrute la protéine SsgB [469],[507]. SsgB sert alors de balise moléculaire à FtsZ mais également de stimulateur de l'assemblage de l'anneau Z [507].
La protéine PomZ chez *Myxococcus xanthus* permet via une régulation positive de réguler la position de l'anneau Z (figure 1.8). Il s'agit d'une protéine de la même famille que MipZ, ParA et MinD qui localise au niveau du futur site de division et permettrait de recruter FtsZ [470]. Néanmoins, aucun mécanisme permettant d'expliquer comment PomZ reconnaît le site de division n'a été décrit à ce jour.
Enfin, la protéine MapZ chez *Streptococcus pneumoniae*, joue le rôle de véritable balise moléculaire afin de recruter FtsZ [149],[301] (figure 1.8). MapZ est une protéine transmembranaire qui interagit avec le peptidoglycane et FtsZ. Lors de l'élongation, MapZ est poussé par la synthèse de la paroi cellulaire de par son interaction avec le peptidoglycane. Dès l'arrêt de l'élongation, MapZ se place au milieu de chaque cellule fille, ce qui correspond au futur site de division. Lorsque les cellules filles rentrent en division, FtsZ interagit avec le domaine cytoplasmique de MapZ, ce qui lui permet de se localiser au milieu de la cellule et d'enclencher à nouveau la division cellulaire.

1.2.2.3 Variabilité des systèmes de positionnement

Il existe ainsi une grande diversité mécanismes de régulation du positionnement de l'anneau Z. Non seulement les protéines mises en jeu ne sont pas les mêmes d'un organisme à un autre, mais les moyens de localiser l'anneau Z sont parfois complètement différents. Aussi, il est intéressant de soulever qu'un certain nombre d'organismes n'ont aucun des systèmes décrits au sein de leur génome [158]. Par exemple, les *Leuconostocaceae* ne présentent aucun

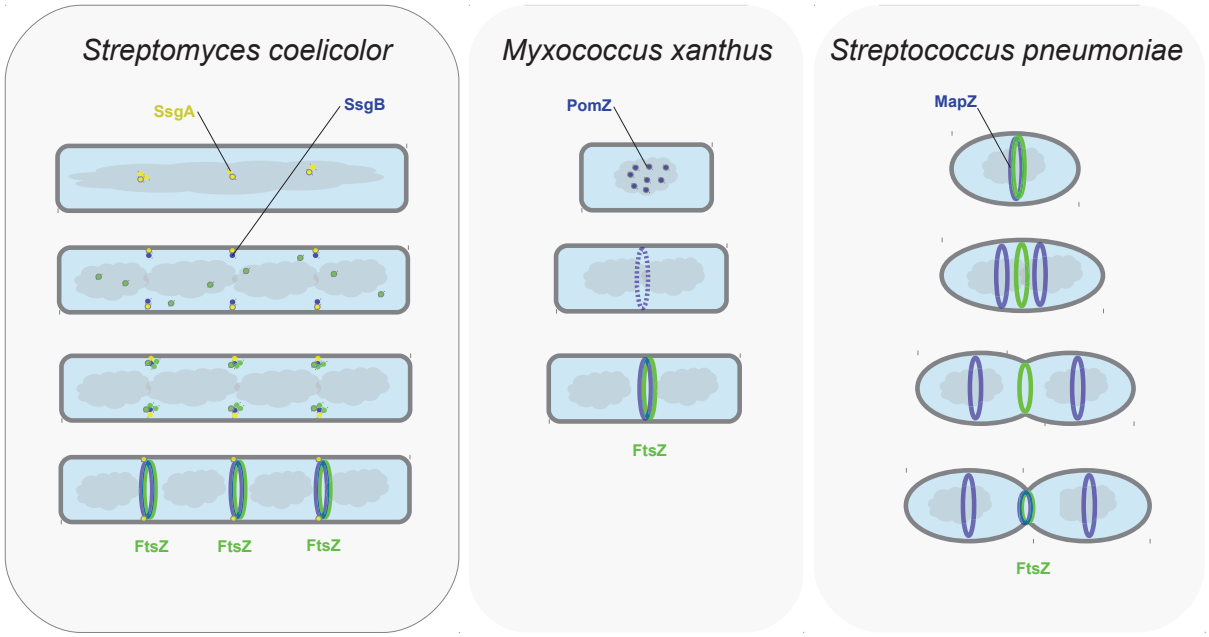


FIGURE 1.8 – Les systèmes de régulation positive du positionnement de l'anneau Z chez *S. coelicolor*, *M. xanthus* et *S. pneumoniae*. Les protéines jouant le rôle de balise moléculaire de l'anneau Z sont représentées en bleu.

de systèmes NO, Min et MapZ (figure 1.9). Cela suggère ainsi une diversité des mécanismes de positionnement de l'anneau Z encore largement sous estimée .

Image non disponible

FIGURE 1.9 – Distribution taxonomique des systèmes Min, NO et MapZ chez les *Lactobacillales*. Les *Aerococcaceae* et les *Leuconostocaceae* ne possèdent aucun de ces trois systèmes. Tiré de [158].

1.2.3 Régulation de la dynamique de l'anneau Z

L'état de polymérisation de l'anneau Z est régulé de façon très fine par une multitude de protéines (figure 1.10). Celles-ci sont recrutées au niveau du site de division pour agir sur la protéine FtsZ en régulant soit l'ancrage à la membrane de FtsZ, soit directement l'assemblage des protofilaments.

1.2.3.1 Ancrage à la membrane de FtsZ

Tout d'abord, FtsZ étant une protéine cytosolique, elle nécessite d'être ancrée à la membrane pour permettre non seulement de former l'anneau Z mais également pour transmettre les forces de constrictions à la membrane [126],[47]. Plusieurs protéines sont responsables de l'ancrage membranaire de FtsZ.

La protéine FtsA est une probable ATPase de type actine [410] qui permet l'ancrage de FtsZ à la membrane via son hélice amphipatique [376] (figure 1.10). FtsA serait aussi non seulement impliqué dans la stabilisation de l'anneau Z chez *E. coli* [376] mais également dans le recrutement de composants du divisome [401]. FtsA est très conservée chez les bactéries indiquant probablement un rôle important dans la physiologie de la cellule. Néanmoins, une délétion de FtsA chez *B. subtilis* n'est pas létale, ce qui suggère que d'autres mécanismes régulent l'ancrage de l'anneau à la membrane [26].

La protéine SepF joue justement ce rôle redondant et a été proposée d'ancrer FtsZ à la membrane chez *B. subtilis* via une hélice amphipatique similaire à celle de FtsA [108] (figure 1.10B1). Ces observations sont cohérentes puisque SepF devient essentiel chez *B. subtilis* en l'absence de FtsA, suggérant une redondance entre les deux protéines [190]. SepF interagit avec le domaine C-terminal de FtsZ et semble également inhiber l'activité GTPase de FtsZ et ainsi de stabiliser les protofilaments [430].

Chez *E. coli*, la protéine ZipA de façon similaire à SepF chez *B. subtilis* semble également avoir un rôle redondant avec celui de FtsA [375] (figure 1.10A1). ZipA s'ancre à la membrane via son domaine transmembranaire et interagit avec le domaine C-terminal de FtsZ par son domaine cytoplasmique [283].

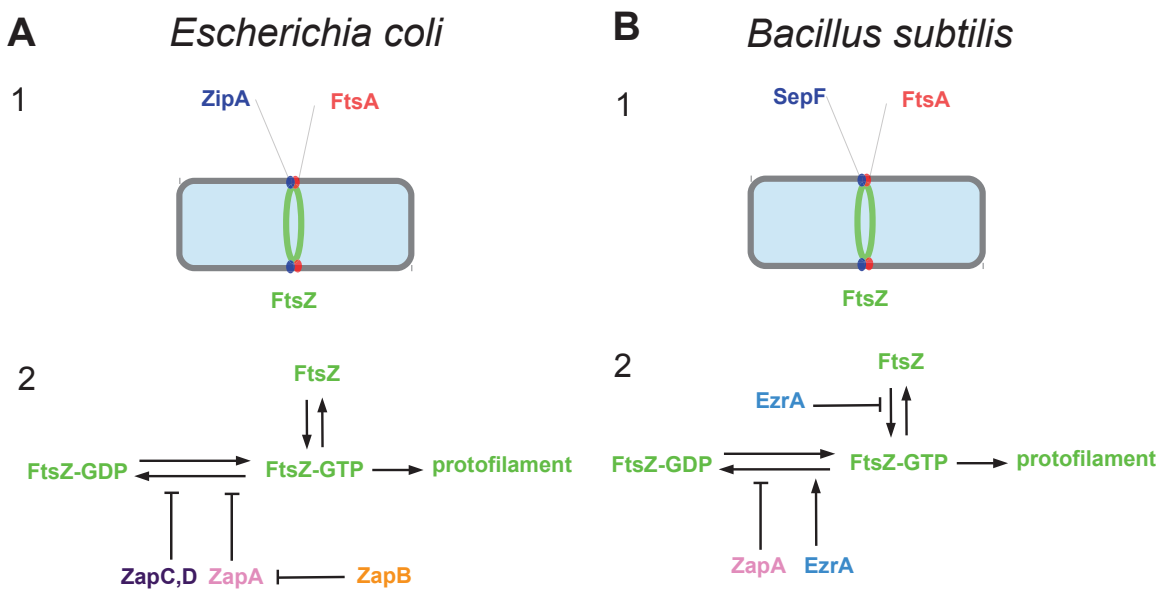


FIGURE 1.10 – Régulation de la dynamique de l'anneau Z chez *E. coli* et *B. subtilis*. (A1) Ancrage de l'anneau chez *E. coli*. (A2) Régulation de l'assemblage de l'anneau chez *E. coli*. (B1) Ancrage de l'anneau chez *B. subtilis*. (B2) Régulation de l'assemblage de l'anneau chez *B. subtilis*.

1.2.3.2 Régulation de l'assemblage de l'anneau Z

Chez *E. coli* et *B. subtilis*, la protéine ZapA régule l'assemblage de l'anneau Z en stabilisant les polymères de FtsZ [182] (figure 1.10A2,B2). ZapA s'organise en tétramère et relie les protofilaments entre eux [286]. Cette stabilisation s'explique en partie car l'interaction entre ZapA et FtsZ défavoriserait l'hydrolyse du GTP [432]. Chez *E. coli*, ZapA forme un système de régulation avec les protéines ZapBCD [116],[112],[111]. ZapB séquestre ZapA et induit ainsi une inhibition de la stabilisation des polymères FtsZ [153]. ZapC et ZapD ont été montré comme étant des inhibiteur directs de l'activité GTPase de FtsZ ayant ainsi un rôle de stabilisateurs [112],[111]. ZapA est relativement conservé chez les bactéries mais de façon surprenante, les protéines ZapBCD ne sont présentes que chez les *Proteobacteria* et donc absente chez *B. subti-*

lis. Cela suggère que la régulation de ZapA chez *B. subtilis* s'effectue grâce à d'autres protéines encore non caractérisées.
Un autre exemple de régulation de la formation de l'anneau Z est accompli par la protéine EzrA présente chez chez *B. subtilis* [272]. EzrA inhibe la polymérisation de FtsZ en empêchant la fixation du GTP et en accélérant son hydrolyse [186],[67]. La protéine ClpX chez *E. coli* inhibe aussi la formation de l'anneau Z non pas en régulant l'activité GTPase mais en dégradant les protéines FtsZ [58]. En effet, cette protéine fait partie du complexe de remodelage des protéines ClpXP qui permet de dégrader activement les protéines pour assurer le renouvellement du pool de protéines [319]. Il existe également des protéines régulatrice de l'anneau Z qui s'expriment uniquement en cas de stress cellulaire comme les protéines SulA et ZapE chez *E. coli* qui jouent un rôle d'inhibiteur respectivement en cas de dommage à l'ADN [209],[34] ou en condition d'anaérobie [307].

1.2.4 Protéines du divisome associées à FtsZ

Lorsque l'anneau Z est assemblé au site de division, il recrute un certain nombre de protéines, formant le divisome, dont le rôle est de mener à bien la septation et effectuer la constriction de la cellule. Lors de la constriction, il est nécessaire de synthétiser du peptidoglycane afin de former la paroi cellulaire au site de division. Nous traitons de cette voie métabolique en détail dans la section 1.4. Nous décrirons principalement ici le recrutement des protéines nécessaires à la septation en utilisant *E. coli* comme modèle car c'est chez cet organisme que le divisome a été le mieux décrit.

Chez *E. coli*, les protéines du divisome sont recrutées dans un ordre qui semble linéaire [169], [293] (figure 1.11B) mais peut différer chez d'autres espèces. Après le positionnement de FtsZ et de ses régulateurs (Cf 1.2.2), le complexe FtsEX est recruté par FtsZ [70]. FtsE est une protéine qui lie l'ATP et FtsX est une protéine membranaire. Les deux composants forment un transporteur ABC (« ATP-Binding Cassette ») [87]. FtsX recrute ensuite EnvC, un activateur des amidases AmiA et AmiB qui seront recrutées tardivement au site de division

FIGURE 1.11 – Le modèle du divisome chez *E. Coli*. (A) Modèle schématique du divisome chez *E. coli*. Tiré de [472]. (B) Ordre chronologique du recrutement des protéines du divisome au site de division. Adapté de [293].

[516],[475]. Ces amidases permettent notamment d'hydrolyser l'ancien peptidoglycane afin de modifier ses propriétés mécaniques lors de la synthèse du nouveau peptidoglycane [391]. FtsK, protéine également impliquée dans la ségrégation chromosomique, est ensuite recrutée au septum et recrute la protéine membranaire FtsQ [169].

FtsQ recrute à son tour 4 protéines au site de division : FtsL, FtsB , FtsW et FtsI [169]. Le role de FtsL et FtsB n'est pas clairement défini. FtsW est une protéine membranaire qui intervient dans la synthèse du peptidoglycane. Il a été proposé qu'elle jouerait le rôle de Lipide II flippase [327] et permettrait ainsi de mener à bien la synthèse septale du peptidoglycane. Néanmoins, le rôle de FtsW a récemment été remis en cause [312]. En effet, son homologue RodA qui avait également été suggéré comme possédant une activité Lipide II flippase lors de l'élongation a été récemment montré comme ayant une activité de polymérisation du peptidoglycane [312]. Il a donc été inféré que FtsW présenterait également une activité peptidoglycane polymérase. De plus, la protéine MurJ a été démontrée comme jouant le rôle de Lipide II flippase [425]. Enfin, FtsI (PBP3) est une PBP (« Pencillin-binding protein ») qui permet de polymériser le peptidoglycane. Elle est probablement recrutée par FtsW [327]. Une autre PBP est d'ailleurs recrutée au site de division chez *E. coli* : PBP1b. Cette dernière est régulée par la lipoprotéine LpoB [473]. Cette protéine recrutée à la membrane externe est un activateur de l'activité transpeptidase de PBP1b. Ce complexe est donc principalement impliqué dans la synthèse septale de la paroi lors de la constriction de l'anneau Z.

La protéine FtsN est l'une des dernières protéines du divisome à être recrutée [2]. Elle est

présente principalement chez les *Proteobacteria* et absente chez les *Firmicutes* [401]. Elle a été montrée comme étant impliquée dans la stabilisation de l'anneau Z mais également des interactions entre les composants du divisome [170]. Il a aussi été observé qu'une surexpression de FtsN supprimait l'essentialité de FtsA, ZipA, FtsQ et FtsI [74],[91],[161] indiquant ainsi probablement une fonction redondante avec ces protéines. L'invagination de la membrane externe lors de la constriction chez *E. coli* est menée par un complexe composé des protéines TolAQR et Pal recruté tardivement [163],[92]. Enfin, la protéine FtsH est une protéase recrutée au site de division mais dont le rôle et le recrutement au site de division n'est pas encore clarifié [498].

De façon intéressante, la plupart des protéines du divisome sont conservée chez les bactéries indiquant, malgré le peu de données sur leur rôle respectif, une grande importance dans le processus de division. Il a été néanmoins observé quelques différences chez certains organismes. Par exemple, chez *Streptococcus pneumoniae*, il a été proposé que FtsEX recrute l'hydrolase PcsB [424] de manière similaire à EnvC comme chez *E. coli*. Par ailleurs, la protéine LpoB n'est présente que chez les *Gammaproteobacteria*, ce qui suggère que les PBPs du divisome sont régulées différemment chez les autres espèces [472]. Cela est d'ailleurs confirmé pour certaines bactéries comme *S.pneumoniae* chez qui les PBPs du divisome PBP1a et PBP2b sont régulées respectivement par CozE et MacP [142], [141]. Aussi, le sytème Tol-Pal semble spécifique des bactéries à Gram négatif en raison du fait qu'il est impliqué dans l'invagination de la membrane externe.

1.3 L'élongasome

1.3.1 MreB, protéine centrale de l'élongasome

Lors de la division cellulaire, les bactéries ont d'abord besoin de s'allonger afin que les cellules filles aient une taille du même ordre que la cellule mère. Cette élongation est opérée par un complexe protéique appelé l'élongasome qui va conduire la production du peptidoglycane

FIGURE 1.12 – Le modèle de l'élongasome chez *E. coli*. (A) Structure en hélice de MreB par microscopie à fluorescence. Tiré de [129]. (B) Les deux modèles de la structure de l'élongasome. Le modèle des patchs est maintenant admis. Tiré de [503]. (C) Modèle de l'élongasome chez *E. coli*. Tiré de [472].

tout le long de la paroi cellulaire [342]. Nous traitons de cette voie métabolique en détail dans la section 1.4.

Tout comme pour le divisome, l'élongasome nécessite la présence d'une protéine échafaudage. Cette protéine est la protéine MreB et est considérée historiquement comme la protéine responsable de la forme en bâtonnet chez les bactéries [252]. MreB est en général absent des génomes de bactéries en forme de coque [379] et a été décrit principalement chez *E. coli* et *B. subtilis*. Il s'agit d'une protéine de type actine de la même famille que FtsA [478] qui s'organise de manière hélicoïdale le long de la bactérie, au contact de la membrane [228] (figure 1.12A). Tout comme pour l'anneau Z, la continuité de l'hélice MreB a été controversée. En effet, plusieurs études récentes utilisant des techniques de pointe de microscopie ont montré que MreB formait en fait des patchs discontinus et mobiles tout le long de la cellule [159], [98], [480] (figure 1.12B).

Il existe plusieurs copies (ou isoformes au niveau protéique) chez la plupart des bactéries à Gram positif. Chez *B. subtilis*, trois isoformes sont retrouvés : MreB, Mbl et MreBH. Ces trois isoformes colocalisent au sein de la cellule en formant une structure discontinue pseudo-hélicoidale [239]. Il semble que le rôle de ces protéines soit redondant avec néanmoins certaines fonctions spécifiques [239],[129]. Mbl a été montré comme étant essentiel dans le processus d'élongation en dirigeant l'insertion hélicoïdale du peptidoglycane néoformé dans la paroi [239]. MreBH participerait au remodelage du peptidoglycane par l'interaction avec LytE, une peptidoglycane hydrolase [59].

1.3.2 Autres composants de l'élongasome

MreB tout comme FtsZ recrute les protéines impliquées dans la synthèse du peptidoglycane (figure 1.12C). Le rôle de la plupart de ces protéines est encore peu compris bien qu'elles s'avèrent essentielles au bon fonctionnement de la machinerie d'élongation. La plupart sont conservées chez les bactéries et leur rôle respectif a été démontré chez divers organismes incluant principalement *C. crescentus/vibrioides*, *E. coli* et *B. subtilis*.

Tout d'abord, les protéines MreC et MreD, deux protéines membranaires, semblent être impliquées dans le positionnement de l'élongasome chez *C. crescentus/vibrioides* mais leur fonction précise n'est pas encore connue [96],[504]. De façon intéressante, MreC forme des hélices qui ne colocalisent pas avec celles de MreB chez *C. crescentus/vibrioides* [114]. Une autre protéine impliquée dans l'élongation est la protéine RodZ. Il s'agit d'une protéine membranaire qui permettrait de promouvoir l'ancrage de MreB à la membrane via son domaine cytoplasmique [30],[479]. Son rôle a été montré chez *Thermotoga maritima* et *E. coli*. On retrouve aussi la protéine RodA, un homologue de FtsW comme faisant également partie de l'élongasome [474]. Elle a été montrée comme ayant une activité de polymérisation du peptidoglycane chez *B. subtilis* [312], [124]. Les protéines MraY et MurG, deux protéines impliquées dans les étapes précoces de la synthèse du peptidoglycane sont également recrutées aux sites d'élongation chez *E. coli* [326]. Enfin, deux PBP sont recrutées spécifiquement par l'élongasome : PBP2 et PBP1A chez *E. coli* ou PBP1 chez *B. subtilis* [258],[437],[240]. Chez *E. coli*, l'activité transpeptidase de PBP1a est régulée par LpoA [473].

Comme pour le divisome, certaines bactéries diffèrent fortement de ce modèle canonique. Chez *S. pneumoniae* par exemple, MreB est absent et l'élongasome est fusionné avec le divisome au site de division [426]. De façon intéressante, MreC et MreD sont présents et essentiels chez *S. pneumoniae* indiquant que leur rôle peut être indépendant de MreB [258]. Aussi, la régulation de l'activité des PBPs au sein de l'élongasome varie en fonction des espèces. En effet, LpoA n'est présente que chez les *Gammaproteobacteria* [472]. Chez *S. pneumoniae* par exemple, PBP1a (différente de PBP1a chez *E. coli*) est régulée par CozE qui interagit également avec MreC et MreD [142] tandis que PBP2b (PBP2a chez *B. subtilis*) est régulée par MacP [141].

1.4 Le peptidoglycane, composant de la paroi cellulaire bactérienne

1.4.1 Rôle et structure du peptidoglycane

La grande majorité des bactéries produisent une paroi cellulaire, matrice les protégeant des agressions mécaniques et chimiques. Cette structure est synthétisée tout au long du cycle cellulaire, durant l'élongation et la septation mais également lors de la sporulation [472]. Chez les bactéries à Gram négatif, la paroi cellulaire est située entre deux membranes cellulaires tandis que chez les bactéries à Gram positif, la paroi est plus épaisse et directement à l'extérieur de la cellule (figure 1.13).

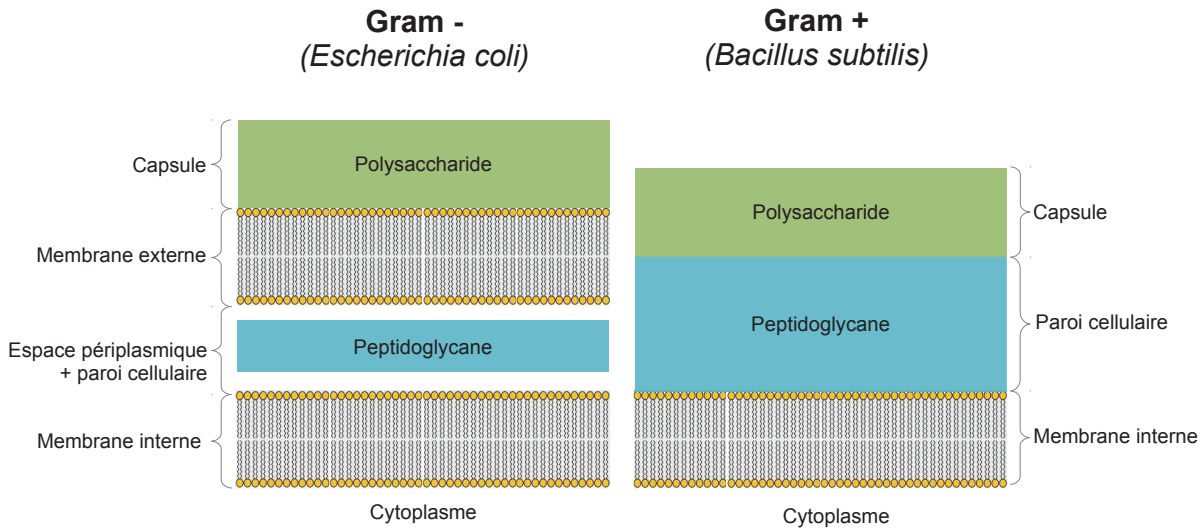


FIGURE 1.13 – Représentation schématique de l'enveloppe cellulaire chez les bactéries à Gram négatif et positif.

Cette structure est constituée principalement de peptidoglycane, un polymère complexe d'oses et d'acides aminés organisé en réseau réticulé [486]. Il existe également un deuxième composant, les acides teichoïques. Nous ne détaillerons pas ces molécules dans cet exposé et nous concentrerons uniquement sur le peptidoglycane.

D'un point de vue biochimique, le peptidoglycane est composé d'un empilement d'un polymère osidique linéaire de N-acétylglucosamine-N-acétylmuramate liés en $\beta 1- > 4$ et réticulé par des ponts peptidiques (tetra/pentapeptidique) (figure 1.14). La longueur des chaînes polyosidiques varie de façon conséquente au sein d'une bactérie mais également entre les différentes espèces [465],[497],[392]. Les ponts peptidiques sont liés de façon covalente par une liaison amide au carboxylate du dérivé du lactate situé sur le carbone 3 du N-acétylmuramate. Ils forment entre eux une liaison amide via un pont interpeptidique (ou branchement), ce qui a pour conséquence la réticulation du peptidoglycane (figure 1.14).
La composition en acides aminés du peptidoglycane varie en fonction des espèces (figure 1.15). On retrouve notamment la plus grande variabilité au niveau du résidu 3 du pentapeptide [486]. Ainsi, le résidu à la position 3 est un acide méso-Diaminopimelique chez *E. coli* et *B. subtilis* et une L-Lysine chez *S. aureus* ou *S. pneumoniae* [486]. On observe aussi la présence d'un certain nombre de résidus amidés chez certaines espèces. Par exemple, chez *B. subtilis*, l' acide méso-Diaminopimelique est présent sous forme amidée [13] et chez *S. aureus*, le D-Glutamate 2 est converti en iso-Glutamine [251]. Enfin, le branchement entre les peptides peut être de différentes natures. Chez *E. coli* et *B. subtilis*, le pont interpeptidique est une simple liaison amide entre le méso-Diaminopimelate 3 et la D-Alanine 4 alors que chez d'autres bactéries, un ou plusieurs résidus d'acides aminés permettent de faire la liaison [486]. Chez *S. aureus*, 5 Glycines lient la L-Lysine 3 à la D-Alanine 4 alors que l'on peut retrouver 2 Alanines ou 1 Alanine/1 Sérine chez *S. pneumoniae*. L'ensemble de ces différences s'expliquent par la présence d'enzymes spécifiques à certaines espèces (Cf 1.4.2).

1.4.2 Étapes cytosoliques et membranaires de la biosynthèse

La synthèse du peptidoglycane est initiée par une succession d'étapes cytoplasmiques qui vont conduire à la production du précurseur membranaire undécaprenoyl-pyrrophosphate-N-acétylmuramoyl-N-acétylglucosamine-pentapeptide aussi appelé lipide II [472]. Nous allons ici principalement développer le modèle de synthèse décrit chez *E. coli*, bien que certaines voies aient été caractérisées chez d'autres organismes [24] (figure 1.16).

FIGURE 1.14 – Structure du peptidoglycane chez *E. coli*. Le peptidoglycane forme un réseau complexe de chaînes polyosidiques. Les oses sont représentés en vert, les dérivés d'acides aminés en jaune, rouge, violet et bleu. La réticulation s'effectue par les peptides via un pont interpeptidique (en gris). Adapté de [487].

Le premier métabolite de la synthèse du peptidoglycane est le fructose-6 phosphate provenant de la deuxième étape de la glycolyse. Le fructose-6 Phosphate est converti par GlmS en glucosamine-6-Phosphate [18]. La réaction inverse de désamination est catalysée par NagB [6]. La glucosamine-6-P est convertie de façon réversible en glucosamine-1-P par la mutase GlmM [227] puis en UDP-N-acétylglucosamine par GlmU [316]. Ce dernier métabolite est alors converti en UDP-N-acétylglucosamine-enopyruvate par MurA [52], puis en UDP-N-acétylmuramate par MurB [32]. Il est intéressant de noter que chez les bactéries à Gram positif, il existe deux copies de MurA : MurA et MurZ [104].
Il existe une voie alternative pour produire l' UDP-N-acétylmuramate à partir de la N-acétylglucosamine. Tout d'abord, le N-acétylglucosamine est phosphorylé par MurK, produisant ainsi le N-acétylglucosamine-6P. Ce dernier peut alors prendre la voie classique de la synthèse des précurseurs par l'action de NagA qui élimine l'acétate du N-acétylglucosamine-6P pour donner la glucosamine-6-P. Il est ensuite converti en N-acétylmuramate-6-P par MurQ [185] puis dephophorylé par MupP [43]. L' UDP-N-acétylmuramate est alors synthétisé par l'action successive de AmgK et MurU [167]. Cette voie alternative est particulièrement utilisée dans le recyclage des produits de dégradation du peptidoglycane [226]. Néanmoins, le recyclage du peptidoglycane n'est pas retrouvé chez toutes les bactéries et notamment pas chez *S. pneumoniae* [42].
Le pentapeptide est alors construit par addition successive d'acides aminés sur le dérivé lactoyl du UDP-N-acétylmuramate [472]. La protéine MurC ajoute la L-Alanine 1, MurD le

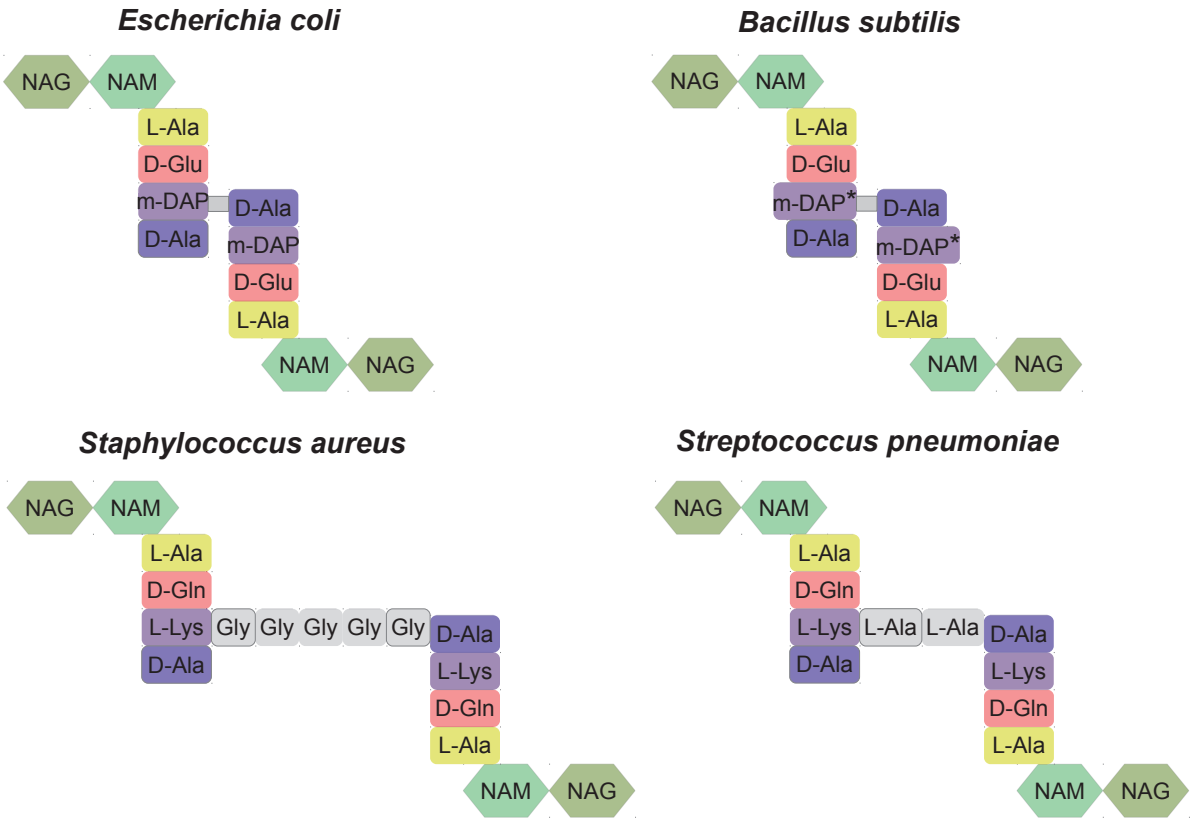


FIGURE 1.15 – Variabilité du peptidoglycane en acides aminés chez quelques bactéries. La structure varie principalement au niveau du résidu 3 (violet) et du pont interpeptidique(gris). L'astérisque correspond à l'amidation du résidu.

D-Glutamate 2, MurE la L-Lysine ou le meso-Diaminopimélate 3 puis MurF le dipeptide D-Alanyl-D-Alanine précédemment constitué par deux ligases DdlA et B [280], [386], [314], [109],[522]. Deux types de racémases sont mises en jeu afin de produire les formes D d'acides aminés : Alr et DadX, qui permettent la conversion de la L-Alanine en D-Alanine [506] et MurI celle du L-Glutamate en D-Glutamate [99]. L'undécaprenoyl-Phosphate est alors greffé au composé en remplaçant l'UDP par MraY pour donner le Lipide I [213]. Enfin, MurG catalyse la réaction d'addition du N-acétylglucosamine en $\beta 1- > 4$ pour donner le Lipide II [315]. Les ponts interpeptidiques sont ajoutés ensuite chez les espèces qui nécessitent des résidus supplémentaires. Chez *S. aureus*, FemX, Fmh, FemA et FemB ajoutent successivement 5 Glycines au résidu Glutaminyl après avoir modifié le glutamyl en glutaminyl par GatD et MurT

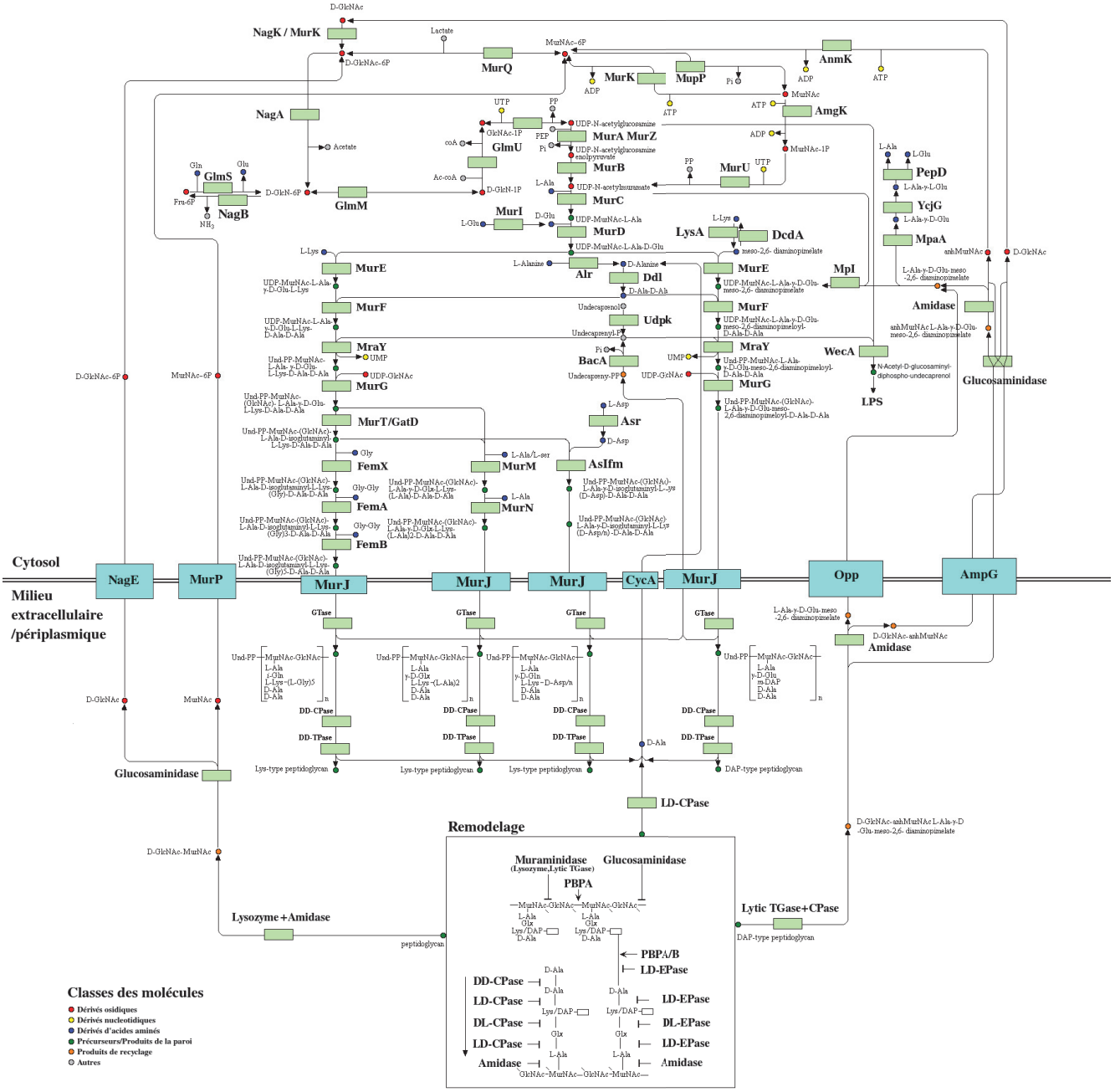


FIGURE 1.16 – Synthèse, maturation et recyclage du peptidoglycane. Les enzymes sont représentées en rectangles vert clair, les transporteurs en rectangles cyan et les métabolites en ronds.

[486],[248],[339], [139]. Chez *S. pneumoniae*, MurM et MurN catalysent l'addition successive d'Alanines ou de Sérines sur le résidu Glutaminyl/Glutamyl [144],[145]. Chez *Enterococcus faecium*, un résidu D-Aspartate/Asparagine est greffé au résidu Glutaminyl par Aslfm [29]. Le D-Aspartate est synthétisé par la racémase Asr à partir du L-Aspartate.
Les métabolites ainsi formés sont pris en charge par la Lipide II flippases MurJ afin de les faire passer du coté externe de la membrane [425]. Les protéines FtsW et RodA ont été initialement proposées d'être les Lipide II flippases [327],[408] mais comme expliqué dans les sections 1.2 et 1.3, leur rôle est débattu et le modèle maintenant admis est que seule la protéine MurJ aurait une activité Lipide II flippase [312],[425].

1.4.3 Étapes extracellulaires de la biosynthèse : les PBPs

Une fois les précurseurs synthétisés, le peptidoglycane est réticulé et polymérisé à la surface externe de la membrane plasmique (figure 1.16). Les enzymes de polymérisation/réticulation du peptidoglycane sont appelées PBPs pour « Penicillin-Binding Proteins » [414]. Elles ont été tout d'abord découvertes par leur capacité à lier les molécules de la famille de la pénicilline [41] et par leur propriété à conférer des résistances à ces antibiotiques en accumulant certaines mutations [436],[192]. Les PBPs sont classiquement organisées en trois domaines : un petit domaine intracellulaire, un domaine transmembranaire puis un large domaine extracellulaire qui porte la fonction catalytique. Il existe trois types de PBPs [414]. Le type A possède deux activités catalytiques : glycosyltransferase et transpeptidase. Le type B porte uniquement l'activité transpeptidase. Le dernier groupe correspond aux PBPs de faible poids moléculaire dans lequel on retrouve notamment les carboxypeptidases et les endopeptidases. Il existe également des PBPs ayant uniquement une activité glycosyltransferases comme MGT chez *S. aureus* [494] . Une classification des PBPs a été proposée par Sauvage et ses collègues (figure 1.17) [414]. C'est d'ailleurs sur la base de cette classification que nous avons étudié les PBPs dans cette thèse. Enfin, les protéines RodA et FtsW initialement décrites comme présentant des activités Lipide II flippase semblent être impliquées dans la polymérisation du peptidoglycane [312],[425]. RodA a été montré comme présentant une activité glycosyltransferase [124]. Cependant, l'activité de

FtsW n'a pas été démontrée à ce jour.

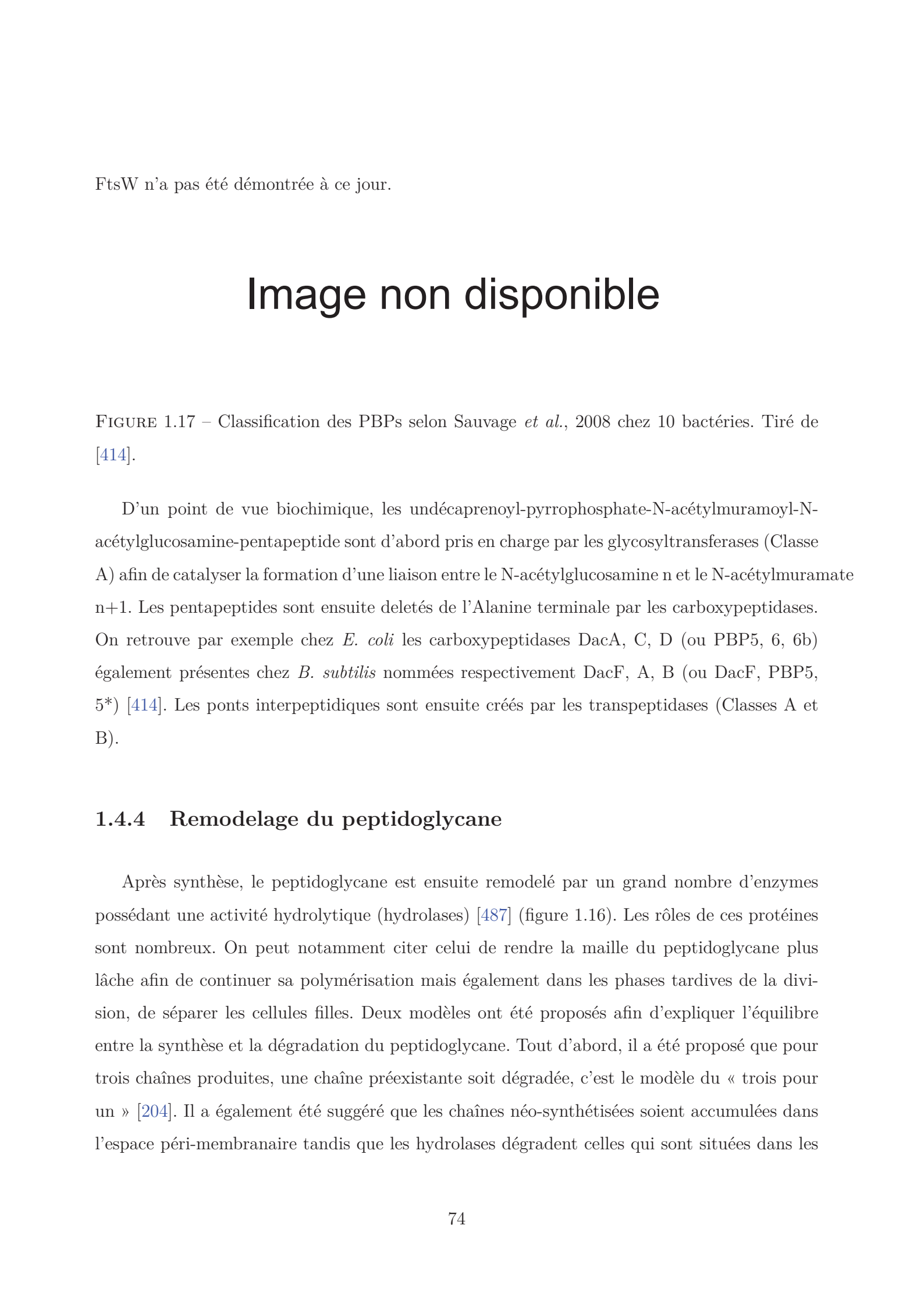


FIGURE 1.17 – Classification des PBPs selon Sauvage *et al.*, 2008 chez 10 bactéries. Tiré de [414].

D'un point de vue biochimique, les undécaprenoyl-pyrrophosphate-N-acétylmuramoyl-N-acétylglucosamine-pentapeptide sont d'abord pris en charge par les glycosyltransferases (Classe A) afin de catalyser la formation d'une liaison entre le N-acétylglucosamine n et le N-acétylmuramate n+1. Les pentapeptides sont ensuite deletés de l'Alanine terminale par les carboxypeptidases. On retrouve par exemple chez *E. coli* les carboxypeptidases DacA, C, D (ou PBP5, 6, 6b) également présentes chez *B. subtilis* nommées respectivement DacF, A, B (ou DacF, PBP5, 5*) [414]. Les ponts interpeptidiques sont ensuite créés par les transpeptidases (Classes A et B).

1.4.4 Remodelage du peptidoglycane

Après synthèse, le peptidoglycane est ensuite remodelé par un grand nombre d'enzymes possédant une activité hydrolytique (hydrolases) [487] (figure 1.16). Les rôles de ces protéines sont nombreux. On peut notamment citer celui de rendre la maille du peptidoglycane plus lâche afin de continuer sa polymérisation mais également dans les phases tardives de la division, de séparer les cellules filles. Deux modèles ont été proposés afin d'expliquer l'équilibre entre la synthèse et la dégradation du peptidoglycane. Tout d'abord, il a été proposé que pour trois chaînes produites, une chaîne préexistante soit dégradée, c'est le modèle du « trois pour un » [204]. Il a également été suggéré que les chaînes néo-synthétisées soient accumulées dans l'espace péri-membranaire tandis que les hydrolases dégradent celles qui sont situées dans les

couches externes [246]. Les hydrolases ont notamment été très étudiées chez *S. pneumoniae* et *E. coli*. C'est pourquoi nous prendrons particulièrement des exemples de ces bactéries, bien qu'il existe une très large diversité d'hydrolases au sein du domaine bactérien. Les hydrolases peuvent être classées en plusieurs catégories en fonction de leur activité [487].

Tout d'abord, on retrouve les hydrolases qui clivent les liaisons au sein des chaînes osidiques, classées en deux catégories : les muraminidases et les glucosaminidases (figure 1.16). Les muraminidases permettent de cliver la liaison $\beta 1- > 4$ entre le N-acétylmuramate et le N-acétylglucosamine. Il existe deux types de muraminidases. Le premier type correspond à la famille du lysozyme comme LytC chez *S. pneumoniae* [156] qui se caractérisent par un repliement particulier par rapport aux autres muraminidases [368] et libèrent du muropeptide lors de l'hydrolyse du peptidoglycane. Les deuxièmes types de muraminidases sont les transglycosylases lytiques comme les Mlt chez *E. coli* [487]. La particularité des transglycosylase lytiques est la production d'anhydro-muropeptide lors de l'hydrolyse du peptidoglycane. Pour ce qui est des glucosaminidases, elles clivent la liaison $\beta 1- > 4$ entre le N-acétylglucosamine et le N-acétylmuramate, comme par exemple LytB chez *S. pneumoniae* [402].

D'autres hydrolases hydrolysent le clivage de liaisons au sein des ponts peptidiques et permettent ainsi de faire baisser le taux de réticulation du peptidoglycane (figure 1.16). Ces hydrolases sont classées en trois types en fonction de leur activité catalytique. Tout d'abord, les carboxypeptidases permettent de libérer l'acide aminé terminal du peptide, peu importe la longueur de ce dernier. On peut citer par exemple PBP5 (ou DacA) chez *E. coli* nommé PBP3 chez *S. pneumoniae* [414], [333], [83]. Ces hydrolases sont aussi impliquées dans la polymérisation puisqu'elles jouent un role dans la formation des pont interpeptidiques (Cf 1.4.3). Ensuite, les hydrolases de type amidase dans lesquelles on retrouve AmiA, B et C chez *E. coli* [194] permettent de cliver la liaison amide basale du peptide. Enfin, les endopeptidases, comme par exemple PBP4 (ou DacB) chez *E. coli* [244], clivent de façon spécifique les liaisons peptidiques.

1.4.5 Recyclage du peptidoglycane

Les métabolites issus de la dégradation du peptidoglycane chez un grand nombre de bactéries sont recyclés par la cellule, que ce soit chez les Gram positif ou négatif [226]. Il est tout de même important de noter que le recyclage du peptidoglycane n'est pas efficient chez toutes les bactéries, comme chez *S. pneumoniae* [42]. La majorité des enzymes que nous décrivons ici ont été caractérisées chez *E. coli* et sont représentées figure 1.16.
Il existe principalement trois voies de recyclage. La première voie est celle de la D-Alanine. Lors de l'action des carboxypeptidases, la D-Alanine est libérée puis est internalisée via le transporteur CycA [417]. La deuxième voie de recyclage concerne les produits issus de l'action des lysozymes, des glucosaminadases et des amidases générant des N-acétylosamines. Ainsi, le N-acétylmuramate et la N-acétylglucosamine sont produits puis internalisés et phosphorylés respectivement par MurP et NagE [226]. Ces composés sont directement réutilisables par la voie de recyclage MurQ-MurK explicitée dans la section 1.4.2. Enfin, il existe une dernière voie de recyclage des produits issus de la dégradation du peptidoglycane par les transglycosylases lytiques. Plus précisément, l'action conjointe des transglycosylases lytiques, des carboxypeptidases et des amidases génère deux produits : Le tripeptide Alanyl-Glutamyl-meso-diaminopimelate et le diose N-acétylglucosamine-anhydro-N-acétylmuramate. Ces derniers sont internalisés respectivement par Opp et AmpG [226]. Le tripeptide peut être réincorporé au précurseur du peptidoglycane par l'action de Mpl ou dégradé par l'action consécutive de MpaA, YcjG et PepD. Le diose subit d'abord la lyse de la liaison osidique par une glucosaminidase produisant ainsi la N-acétylglucosamine et l'anhydro-N-acétylmuramate. Ce dernier est alors converti en N-acétylmuramate-6-P par l'action de AnmK qui peut ensuite être réintroduit dans la voie de synthèse du peptidoglycane [476].

1.5 La sporulation

1.5.1 Généralités

La sporulation correspond à une division cellulaire asymétrique à partir d'une cellule végétative qui aboutit à la formation d'une spore, structure rigide contenant le chromosome de la bactérie. La spore est dotée d'une grande résistance mécanique et chimique et est particulièrement adaptée à des milieux hostiles et peu propices à la prolifération. Lorsque les conditions de vie sont plus favorables, la spore rentre en germination et donne à nouveau une cellule végétative. Nous allons ici nous concentrer sur le modèles de la sporulation chez *Bacillus subtilis*. La formation de la spore s'effectue en plusieurs étapes décrite en figure 1.18A [86]. Certains mécanismes et protéines sont en commun avec la fission binaire classique mais de nombreux points différent entre les deux processus [113].

Tout d'abord, la cellule met en place un anneau Z vers un pôle, induisant ainsi une asymétrie au sein de la cellule. Le chromosome est alors répliqué et en partie ségrégué. La cellule forme ensuite la préspore puis le reste du chromosome est ségrégué. Par la suite, la préspore est englouti par la cellule (« engulfment ») et une couche de peptidoglycane est parallèlement produite entre la membrane de la préspore et celle de la cellule mère (le cortex). Une multitudes de protéines s'agrègent ensuite à la surface de la préspore formant ainsi le manteau. La cellule mère est enfin lysée, relarguant ainsi la spore dans le milieu extracellulaire.

D'un point de vue moléculaire, la sporulation est régulée par de nombreux mécanismes. Un grand nombre de protéines impliquées dans la cytokinèse sont recrutées pendant la sporulation mais aussi des protéines spécifiques à ce processus : les protéines Spo. On retrouve également dans les régulateurs de la sporulation la classe des facteurs sigma (SigEFGH). Il s'agit de facteurs de transcription qui coordonnent à un niveau temporel et spatial les événements moléculaires qui surviennent durant ce processus. Enfin, il est important de noter que les étapes moléculaires au sein du préspore et de la cellule mère sont différentes mais nous n'allons pas décrire ces notions de localisation ici. L'ensemble des protéines impliquées dans la sporulation constituant une véritable cascade de signalisation est présenté figure 1.14B [364], [3], [378], [199], [113], [122],[457].

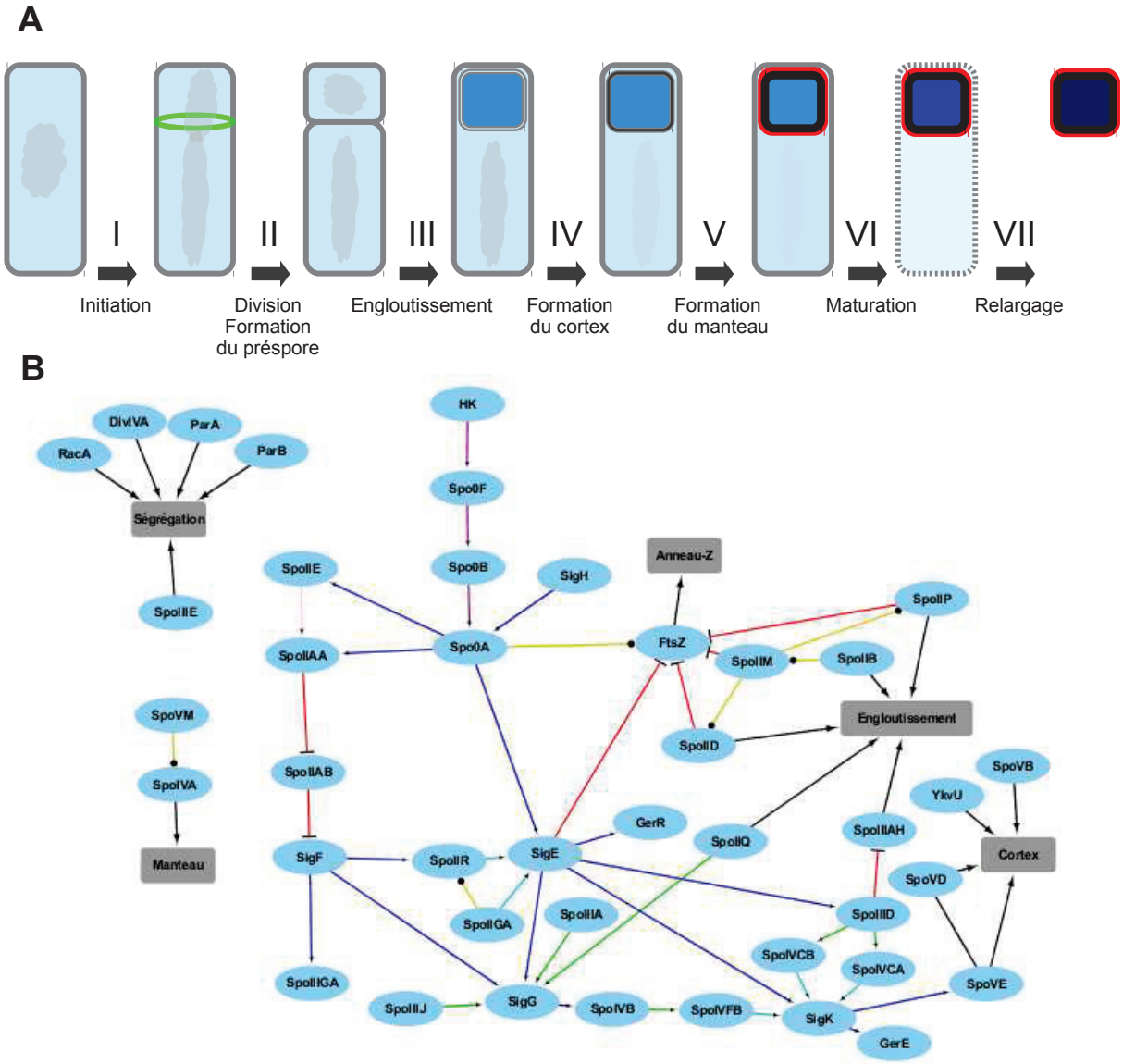


FIGURE 1.18 – Mécanismes cellulaires et moléculaires de la sporulation. (A) Phases cellulaires de la sporulation. L'anneau Z est représenté en vert, la membrane en gris, la spore en bleu, le cortex en noir, le manteau en rouge. (B) Cascade de signalisation de la sporulation. Les protéines sont représentées en ovales bleu et les mécanismes cellulaires en rectangles gris. La couleur des flèches représente le type de lien : violet : phosphorylation, rose : déphosphorylation, bleu : transcription, rouge : inhibition, vert : activation, cyan : maturation, jaune : localisation , noir : déclenchement de processus.

1.5.2 Mécanismes moléculaires de la sporulation

La sporulation est enclenchée par un stimulus extérieur qui va tout d'abord activer une cascade de phosphorylation. Spo0F, Spo0B et enfin Spo0A sont séquentiellement phosphorylés et constituent le système de phospho-relais nécessaire à l'enclenchement de la sporulation [378]. Spo0A, la dernière protéine du phospho-relais à être phosphorylée va alors déclencher une cascade de signalisation en activant directement ou indirectement la transcription de nombreux facteurs de transcription . Une myriade de gènes impliqués dans le processus de sporulation sont alors transcrits initiant ainsi le processus de sporulation [378], [110], [236], [173], [287], [94].

La division asymétrique est alors enclenchée afin de compartimenter le préspore. Pour cela, un anneau Z se place non pas au milieu cellulaire mais près d'un pôle par l'action de Spo0A [273]. Celui-ci va délimiter la spore du reste de la cellule. La formation d'un second anneau est inhibée par un système composé des protéines SigE, SpoIID, SpoIIM et SpoIIP [122]. Il est ensuite nécessaire de ségréguer les chromosomes pour qu'une des deux chromatides sœurs localise dans le futur spore. Pour cela, deux systèmes sont mis enjeu. Le premier est constitué des protéines RacA, ParA (Soj), ParB (Spo0J) et DivIVA qui vont diriger la ségrégation de l'origine de réplication [463], [512]. Le deuxième système consiste en la ségrégation du reste du chromosome et est composé principalement de la protéine SpoIIIE, un homologue de FtsK [511].

L'engloutissement de la préspore a ensuite lieu après la compartimentation. La membrane cellulaire de la cellule mère est invaginée autour de la préspore permettant ainsi la séparation totale entre le cytosol de la préspore et du reste de la cellule. Pour cela, il est nécessaire d'assouplir le peptidoglycane pour faciliter le mouvement de la membrane autour de la préspore. La protéine SpoIIM recrute donc deux hydrolases, SpoIIP et SpoIID [15]. Également, les protéines SpoIIQ et SpoIIIAH sont impliquées dans le remodelage du peptidoglycane [51].

Une fois la spore engloutie, la couche de peptidoglycane est épaissie pour former le cortex par des protéines comme SpoVE, un homologue de FtsW et RodA [483]. SpoVE interagit avec SpoVD, une PBP spécialisée dans la synthèse de peptidoglycane durant la sporulation [136]. On retrouve aussi deux protéines nécessaires à la synthèse tardive du peptidoglycane, SpoVB

et YkvU qui, de par leur homologie avec MurJ, seraient probablement impliquées dans le transport membranaire des précurseurs du peptidoglycane [135].
Enfin, le manteau est formé par le recrutement d'une couche complexe de protéines autour de l'endospore. La protéine SpoIVA par l'action de SpoVM recrute les protéines à adresser au manteau en formant des structures fibrillaires autour de la préspore de façon ATP-dépendante [388], [394]. Environ 70 protéines ont été identifiées comme étant impliquées directement dans la formation du manteau [380].
La spore termine ensuite sa maturation puis la cellule mère est lysée, libérant ainsi la spore dans le milieu extracellulaire.

1.6 La capsule polysaccharidique

1.6.1 Rôle et structure de la capsule

La capsule chez les bactéries correspond à une couche polysaccharidique qui joue un rôle important dans de nombreux mécanismes comme la virulence, l'attachement cellulaire et la formation de biofilms du fait de sa position à la surface de la cellule. Elle est donc directement exposée au milieu extérieur. De façon intéressante, la synthèse de la capsule est directement liée à la division cellulaire chez *S. pneumoniae* puisque qu'elle est produite en même temps et au même endroit que la synthèse du peptidoglycane (au centre de la cellule) [197], [256], [311].
La capsule est un polymère complexe osidique. Il s'agit d'une répétition d'une unité saccharidique de 2 à 8 résidus O-acétylés et ramifiés de façon très variable. En terme de sucres, on retrouve aussi une grande diversité (Galactose, Glucose, Rhamnose, Fucose, *etc*...). Cette grande variabilité de composition chez les bactéries est retrouvée entre les différentes espèces et au sein d'une même espèce. Pour une même espèce, les différences de compositions de la capsule sont responsable de la diversité des sérotypes. Nous allons ici décrire très succinctement les étapes de synthèse et de régulation de la capsule chez *E. coli* et *S. pneumoniae* puisque de nombreuses données expérimentales sont disponibles pour ces espèces. Ces deux exemples ne couvrent néanmoins pas la diversité de composition et de synthèse de la capsule bactérienne.

1.6.2 Synthèse de la capsule

De façon générale et pour une espèce donnée, l'ensemble des gènes impliqués dans la synthèse de la capsule peuvent être séparés en deux groupes : les gènes communs à tous les sérotypes et les gènes sérotype spécifiques. Tout comme pour le peptidoglycane, la synthèse de la capsule est composée d'une phase cytoplasmique et d'une phase extracytosolique [505], [162] (figure 1.19). Classiquement, les répétitions osidiques sont d'abord assemblées et fixées à la membrane via un lipide (LPS pour « lipid polysaccharide »). Les répétitions sont ensuite exportées à la membrane externe, polymérisées puis ancrées à la surface cellulaire.

Concernant *E. coli*, nous prendrons l'exemple de la capsule des groupes 1/4 (sérotype K30/K40) (figure 1.19) [505]. Tout d'abord, un N-acétylglucosamine-1-P est transféré à un undecaprenol-P par l'action de l'enzyme WecA pour le sérotype K40 [390]. Chez le sérotype K30, le sucre ajouté à l'undécaprenyl-P est le Galactose-P par l'enzyme WbaP [101]. La suite des étapes sont similaires chez les deux sérotypes [505]. Une série de sucres est ensuite transférée pour constituer la répétitions osidique. La polymérase Wzy associée à la flippase Wzx, la tyrosine-kinase Wzc et la phosphatase Wzb sont ensuite responsables du transport dans le milieu périplasmique et de la polymérisation des chaînes osidiques [393]. Enfin, les chaînes polyosidiques sont exportées à travers la membrane externe via la protéine Wza puis attachés à la membrane de façon directe ou indirecte par Wzi [505], [102]. Le mécanisme d' ancrage à la membrane n'est pas encore élucidé.

Chez *S. pneumoniae*, il existe également de nombreux sérotypes avec différents types de capsule. Nous allons décrire la synthèse de la capsule chez le sérotype 2 qui est le plus étudié (figure 1.19) [162]. La protéine CpsE transfère d'abord un Glucose-P à l'undecaprenol-P pour former de l'undecaprenol-PP-glucose [217]. L'unité saccharidique est ensuite séquentiellement assemblée par l'action d'autres glycosyltransferases (CpsT, CpsF, CpsG, CpsI) [217]. Des O-acétyltransferases acétylent ensuite les oses de façon très variable. Il a été proposé que CpsJ jouerait le rôle de flippase en prenant le précurseur en charge et en le faisant passer de l'autre

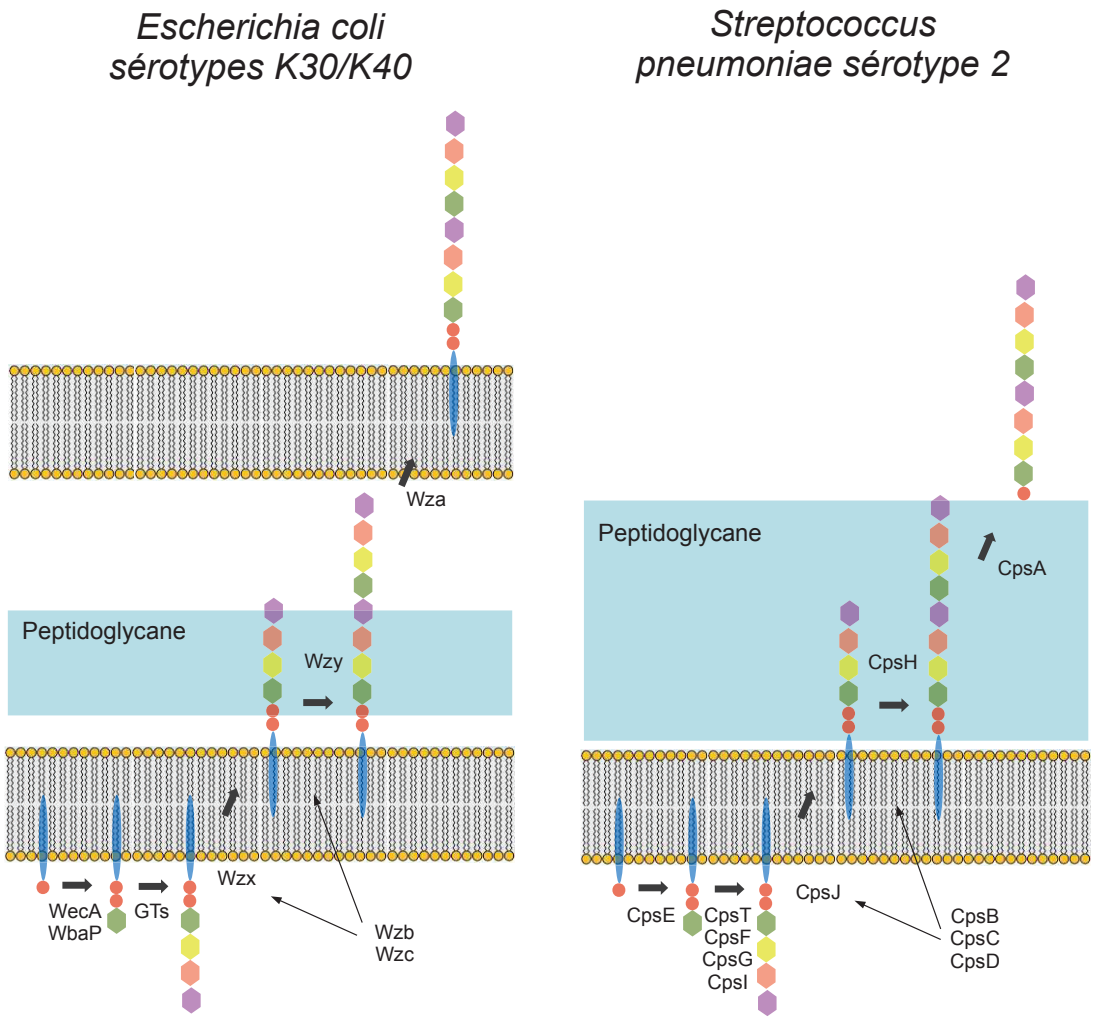


FIGURE 1.19 – Synthèse de la capsule chez *E. Coli* sérotype K30/K40 et *S. pneumoniae* sérotype 2. Oval bleu : undécaprenyl, rond rouge : phosphate, hexagone : ose. Les protéines régulatrices (kinase et phosphatases) sont Wzb/Wzc pour *E. Coli* et CpsB/CpsC/CpsD pour *S. pneumoniae*. Adapté de [505] et [352].

coté de la membrane [514]. La protéine CpsH a été suggérée comme étant la polymérase des polysaccharides [514]. Ces derniers sont ensuite fixés directement aux couches externes de la paroi. Il a été proposé que CpsA était responsable de l'ancrage au peptidoglycane bien que cela n'ait jamais été formellement montré [115]. Ce système est régulé par 3 protéines, CpsB, CpsC, et CpsD [352], [335]. CpsD, un homologue de ParA, est une tyrosine-kinase qui s'autophosphoryle, CpsC ancre CpsD à la membrane et CpsB est la phosphatase associée [197], [334]. Il est intéressant de noter que les enzymes chez *E. coli* et chez *S. pneumoniae* bien que présentant des activités similaires ne sont pas homologues, excepté les kinases CpsC/CpsD et Wzc [335].

1.7 La coordination des machineries du cycle cellulaire

Pour accomplir son cycle cellulaire, la bactérie nécessite de réguler finement toutes les différentes étapes, d'une façon temporelle et spatiale. Pour cela, toutes les machineries décrites précédemment doivent communiquer et se coordonner. La façon dont la cellule coordonne ses machineries du cycle cellulaire varie énormément en fonction des espèces. Également, ces machineries sont reliées à de nombreux autres processus cellulaires comme le métabolisme [189],[493], l'assemblage et le positionnement du flagelle [12] ou encore la compétence [14]. Nous présentons ici de façon non exhaustive des liens établis entre les différents modules moléculaires présentés précédemment.

1.7.1 Coordination élongation-division : MreB et FtsZ

Chez un certain nombre de bactéries la protéine MreB, protéine centrale de l'élongation et FtsZ, protéine échafaudage essentielle à la septation, interagissent. Chez *E. coli*, une étude a montré une interaction directe entre FtsZ et MreB [140]. De plus, FtsZ pourrait aussi intervenir dans la synthèse périphérique de peptidoglycane et ainsi participer à l'élongation [481]. Chez *C. crescentus/vibrioides*, MreB colocalise avec FtsZ au niveau du site de division de façon FtsZ-dépendante [143]. Étant donné le rôle prépondérant de ces deux protéines dans les pro-

cessus d'élongation et de division cellulaire, ces observations montrent qu'il existe une véritable coordination entre l'élongasome et le divisome.

1.7.2 StkP, métarégulateur du cycle cellulaire

StkP est une protéine qui fait partie de la famille des kinases de type eucaryote et qui possède une activité Serine/Thréonine kinase [359]. Chez les bactéries, elle est typiquement composée d'un domaine catalytique intracellulaire, d'un domaine transmembranaire et d'une répétition de domaines PASTA (« PBP-Associated and Serine-Threonine Associated ») extracellulaires [520]. Les domaines PASTA ont été montrés comme interagissant avec certains fragments de peptidoglycane (muropeptide) et avec les β-lactames [325], [298], [438]. Le rôle et l'histoire évolutive des domaines PASTA chez les *Streptococcaceae* et les *Lactobacillales* sont discutés chapitre 4. La phosphatase associée à StkP est PhpP et permet de réguler le degré de phosphorylation des substrats de StkP .
StkP est impliquée dans la régulation de nombreux processus cellulaires tels que la division cellulaire, la synthèse du peptidoglycane, la virulence, ou encore le métabolisme [300], [223],[377], [444], [372], [321]. Il a été récemment montré que son appellation « de type eucaryote » ne faisait plus de sens au vue de nos connaissances actuelles [444]. En effet, la grande majorité des organismes vivants présentent au moins une copie de cette famille de gène. Cependant, le rôle joué dans les différents processus cellulaires chez les bactéries dépend énormément des espèces. Nous allons ici décrire les rôles dans le cycle cellulaire de StkP chez deux *Firmicutes*, *S. pneumoniae* et *B. subtilis*.

1.7.2.1 Rôles de StkP chez *S. pneumoniae*

Chez *S. pneumoniae*, StkP phosphoryle plusieurs protéines substrats impliquées dans la division cellulaire [148], [150], [149], [141] [477].
Tout d'abord, il a été montré que StkP localise au site de division et interagit *in vitro* avec FtsZ [166]. StkP fait ainsi partie du divisome chez *S. pneumoniae*.

FIGURE 1.20 – Le divisome chez *S. pneumoniae*. (A) Représentation schématique du divisome chez *S. Pneumoniae*. Les protéines impliquées dans l'élongation co-localisent avec le divisome. 1a, 2b, 2x : PBP ; Q, L, B, W, X, E, K, Z, A : Fts. Tiré de [311]. (B) Modèle de commutation entre division cellulaire et élongation. La phosphorylation de DivIVA par GpsB induit la septation. Adapté de [150].

Ensuite, il a été montré que la protéine StkP coordonne la septation et l'élongation cellulaire [150]. En effet, chez *S. pneumoniae*, ces deux processus sont intimement liés puisque les deux machineries associées sont fusionnées et colocalisent au niveau du septum [200] (figure 1.20A). Également, ces deux machineries sont coordonnées par FtsZ et pas MreB puisque cette dernière est absente chez S pneumonie [379]. La division se passe alors en deux temps, d'abord une phase d'élongation puis une phase de septation. La protéine StkP joue ainsi le rôle de commutateur moléculaire entre les processus d'élongation et de septation [150] (figure 1.20B). Au début de la division cellulaire, *S. pneumoniae* synthétise du peptidoglycane périphérique. La protéine StkP est ensuite activée au niveau du site de division par la protéine GpsB, un homologue de DivIVA. La protéine EzrA qui interagit directement avec FtsZ permet de recruter le complexe StkP-GpsB-DivIVA au niveau de l'anneau Z. La phosphorylation de DivIVA par StkP oriente la production de peptidoglycane septale pour permettre la septation. Il est intéressant de noter que chez *B. subtilis*, EzrA et GpsB ont été montrés comme étant impliqués dans la régulation du transport de PBP1 des sites de synthèse périphériques au site de division [68]. Ces deux protéines seraient donc d'une importance certaine dans la communication entre ces deux machineries mais de façon foncièrement différentes.

La protéine StkP phosphoryle d'autres substrats impliqués dans la division cellulaire. En effet, MapZ, une protéine impliquée dans le positionnement de l'anneau Z chez *S. pneumoniae* est

également phosphorylée par StkP [149]. Sa phosphorylation régule principalement la constriction de l'anneau Z. La protéine Jag (EloR) est aussi un substrat de StkP [524]. Il s'agit d'une protéine fixant l'ARN qui semble impliquée dans la régulation de la production de nombreuses protéines de synthèse du peptidoglycane et de la division cellulaire comme FtsA. Il a aussi été montré que Jag promouvait l'élongation sous sa forme phosphorylée [443]. Récemment, un nouveau substrat de StkP chez *S. pneumoniae* a été identifié, MacP [141]. Cette protéine est impliquée directement dans la régulation de PBP2A en formant un complexe avec cette dernière. Ainsi, StkP chez *S. pneumoniae* agit comme un véritable méta-régulateur du cycle cellulaire en modulant notamment les protéines impliquées dans la division cellulaire et dans la synthèse du peptidoglycane.

1.7.2.2 Rôles de PrkC chez *B. subtilis*

Chez *B. subtilis*, StkP nommée PrkC semble réguler de nombreux processus cellulaires tels que le métabolisme, la germination des spores et la synthèse du peptidoglycane [377], [423], [277], [381]. Bien que sa délétion chez *B. subtilis* n'impacte pas la division cellulaire, PrkC a été montrée comme étant recrutée au septum, comme chez *S. pneumoniae* [381]. De plus, EzrA, GpsB et DivIVA semblent activer la kinase PrkC et cette dernière phosphoryle GpsB. Une autre étude a montré que PrkC régule l'activité de la kinase WalK [277]. Cette kinase est impliquée dans un système à deux composants en association avec le régulateur WalR [453]. Ces deux protéines régulent la synthèse du peptidogycane. Enfin, PrkC a été démontrée comme impliquée dans le déclenchement de la germination [423]. Il a été postulé que le contact avec des résidus de peptidoglycane dans le milieu externe des spores active PrkC par interaction avec les domaines PASTA. La protéine PrkC semble donc liée à la germination, à la synthèse du peptidoglycane et dans une moindre mesure à la division cellulaire chez *B. subtilis*.

1.7.3 Coordination division-ségrégation chromosomique

La septation doit aussi être finement coordonnée avec la ségrégation chromosomique notamment afin d'éviter l'effet guillotine de l'ADN. Plusieurs éléments relient ainsi ces deux processus cellulaires. Tout d'abord, nous avons vu que le système NO permet d'inhiber la septation tant que le chromosome n'est pas ségrégué [230]. Également, la protéine FtsK chez *E. coli* joue un rôle non seulement dans la ségrégation du chromosome mais aussi dans le phénomène de septation, ce qui en fait un connecteur majeur entre ces deux processus [169],[269]. On retrouve aussi des interactions entre des protéines des deux machineries. Par exemple chez *E. coli*, MatP qui est impliqué dans la compaction de la région *ter* durant la ségrégation chromosomique et ZapB, un régulateur de la polymérisation de l'anneau Z interagissent et permettent de coordonner la division cellulaire et la ségrégation chromosomique [132]. Chez *C. crescentus/vibrioides*, il a été démontré que ParB et MipZ interagissaientt faisant également le lien entre ségrégation et division [461]. Chez *B. subtilis*, DivIB (FtsQ) initialement impliqué dans les étapes tardives de la septation semble réguler la ségrégation via ParA/ParB [398].

1.7.4 Régulation de la synthèse du peptidoglycane

De par son rôle central dans le cycle cellulaire, la synthèse du peptidoglycane est régulée par de nombreux mécanismes et coordonnée avec notamment l'élongation et la septation.
Tout d'abord, une part conséquente des gènes codant pour la synthèse du peptidoglycane sont retrouvés dans des clusters de gènes impliqués dans la division cellulaire et dans l'élongation tels que le cluster DCW (« Division and Cell-Wall ») ou le cluster d'élongation [310], [134], [456], [274]. Cette organisation en clusters de gènes suggère une co-régulation transcriptionnelle.
Au niveau protéique, certaines protéines impliquées dans les phases cytosoliques du peptidoglycane sont recrutées directement aux sites de division et d'élongation. Chez *E. coli*, MurG a été montré comme localisant au site de division mais également comme étant recruté aux sites d'élongation, de façon MreC/MreD dépendante [326]. Chez *C. crescentus/vibrioides*, les protéines de l'élongasome MreB et MreD ont été montrées comme responsables du recrute-

ment des protéines MurB, MurC, MurE, MurF et MraY, toutes impliquées dans la synthèse des précurseurs du peptidoglycane [504].

Également, de nombreuses enzymes impliquées dans les étapes extracellulaires de la synthèse du peptidoglycane telles que les PBPs font partie intégrante du divisome et de l'élongasome. Chez *E. coli* par exemple, PBP1b et FtsI sont deux PBPs du divisome tandis que PBP2 et PBP1A sont recrutées aux sites d'élongation [327], [473], [258],[437]. Chez *B. subtilis*, on retrouve PBP1 et PBP2b localisées au septum tandis que PBP3 et PBP4a semblent plutôt faire partie de l'élongasome [416], [240]. Chez *S. pneumoniae*, la machinerie d'élongation et de septation étant co-localisées, les PBPs, notamment PBP2b et PBP2x, sont recrutées au septum [150].

Enfin, le métabolisme du peptidoglycane est régulé par le système à deux composants WalRK [105]. Chez *S. aureus*, il a en effet été montré que WalR et WalK étaient impliqués dans la régulation d'hydrolases responsables de la dégradation du peptidoglycane [105]. Chez *S. pneumoniae*, les protéines WalR et WalK sembleraient impliquées également dans l'expression d'une hydrolase, PcsB [347]. Chez *B. subtilis*, il a été montré que WalRK régulaient l'expression de CwlO et LytE impliquées respectivement dans la synthèse et le remodelage du peptidoglycane [38]. Chez *B. subtilis*, trois autres protéines liées fonctionnellement au système WalRK ont été identifiées : WalI , WalH et WalJ [451]. WalI et WalJ semblent réguler l'activité kinase de WalK mais le role de WalJ n'a pas encore été élucidé.

1.7.5 Autres coordinations entre machineries

D'autres liens entre certaines machineries sont encore peu décrits. Néanmoins, certaines études apportent quelques éléments de réponse. Par exemple, un lien entre la ségrégation chromosomique et la capsule faisant intervenir ParB et CpsD a été montré chez *S. pneumoniae* [352]. Plus précisément, l'autophosphorylation de la protéine CpsD impliquée dans la production de la capsule régule non seulement la constriction du septum mais aussi la mobilité de ParB lors de la ségrégation chromosomique. De plus, il a été récemment montré que RocS (nommée CDP3 dans le chapitre 3) agissait comme une véritable pierre angulaire au sein de ce

système [318]. En effet, RocS interagit avec ParB et CspD et semble être un connecteur direct entre la constriction, la ségrégation chromosomique et la production de la capsule.
Également, un lien entre la réplication et la ségrégation chromosomique a été établi. ParA semble réguler l'activité de DnaA en contrôlant sa localisation cellulaire [341]. Il parait intéressant de noter que les familles ParB-Noc et ParA-CpsD-MipZ-MinD-PomZ semblent être d'une grande importance dans le cycle cellulaire puisqu'elles sont impliquées dans un grand nombre de systèmes et qu'elle servent de charnière entre la division, l'élongation, la ségrégation et la synthèse de la capsule. Une étude a d'ailleurs suggéré que la famille ParA-CpsD-MipZ-MinD-PomZ serait spécialisée dans la localisation intracellulaire de protéines [292].

1.8 Limites des modèles du cycle cellulaire bactérien

Notre connaissance sur les processus du cycle cellulaire chez les bactéries a considérablement été enrichie depuis ces 30 dernières années. Néanmoins, il reste de nombreux points insaisissables sur la physiologie de la division bactérienne.

1.8.0.1 Rôles peu connus, non concordants ou redondants de certaines protéines

Premièrement, de nombreux points au sein des machineries ne sont pas résolus ou sont contradictoires, que ce soit au niveau structural ou fonctionnel. L'exemple le plus parlant est la structure des cytosquelettes bactériens, composés de MreB et FtsZ. Comme expliqué précédemment, la continuité de ces structures protéiques a été remise en question et les mécanismes sous-jacents induisant leur mobilité/constriction ne sont encore pas clarifiés [98],[220]. Un autre exemple est celui des protéines du divisome comme FtsL, FtsQ ou encore FtsB dont les rôles ne sont pas connus mais qui apparaissent comme étant essentielles à la division [169]. Aussi, l'implication de FtsW dans la synthèse du peptidoglycane en tant que Lipide II flippase est remise en question, puisque cette dernière semble avoir une activité glycosyltransferases [312].

De nombreuses protéines possèdent des rôles différents en fonction des espèces. Par exemple, FtsK (SpoIIIE) chez *B. subtilis* ne participe qu'à la sporulation tandis que chez *E. coli*, elle participe activement à la fission binaire [511]. Autre exemple, les Sérine/Thréonine kinases telles que StkP sont impliquées dans les processus cellulaires très variés en fonction des espèces [223],[377], [444], [321]. Il est possible de proposer l'hypothèse selon laquelle le rôle des protéines varie en fonction non seulement de leur séquence mais également du reste des protéines exprimées dans la cellule. Les déterminants moléculaires sont néanmoins, pour la plupart des cas inconnus.

Il apparaît aussi que de nombreux mécanismes moléculaires semblent *a priori* redondants [131], [413], [1]. Ainsi, on retrouve chez *B. subtilis* le système ScpA/B-Smc et le système ParA/B qui contribuent tous deux à la ségrégation des chromosomes [399]. Autre exemple, chez *E. coli* et chez *B. subtilis*, deux systèmes de positionnement de l'anneau Z existent : le système NO et le système Min [472]. Chez *B. subtilis*, on dénombre aussi trois isoformes de MreB qui semblent jouer des rôles très similaires [239],[129]. Aussi, de nombreuses espèces bactériennes produisent un grand nombre de PBPs, ce qui suggère une redondance fonctionnelle [413]. De nombreuses espèces possèdent ainsi les gènes codant pour les protéines de systèmes *a priori* redondants [131], [413], [1]. Cette apparente redondance semble répandue dans le monde bactérien, ce qui suggère qu'elle a été conservée durant la diversification des bactéries. Il semblerait donc que la redondance de certains systèmes ait une réelle utilité au sein de la cellule. Cela peut être expliqué par le fait qu'une pression de sélection s'exerce sur les bactéries pour maintenir cette redondance afin de s'adapter au mieux dans de divers environnements [515]. Il est aussi probable que la redondance observée ne soit qu'un pâle reflet de la réalité moléculaire et que nous n'ayons pas suffisamment décrit les mécanismes sous-jacents. Les raisons de cette apparente redondance ne sont ainsi pas encore clairement élucidées.

1.8.0.2 De nouveaux mécanismes sont potentiellement à découvrir

A l'opposé, il a été montré que certaines espèces ne possèdent aucun des systèmes connus pour un processus comme par exemple les *Leuconostocaceae* qui ne présentent aucun gène connu de régulation du positionnement de l'anneau Z au sein de leur génomes (figure 1.9) [158]. Ces bactéries possèdent certainement des systèmes similaires et l'absence de systèmes connus détectés est liée principalement à deux éléments. Premièrement, les bactéries ont adopté des mécanismes moléculaires très divers pour mener à bien le même processus cellulaire. Deuxièmement, la plupart des études de la division cellulaire menées chez les bactéries ont été faites chez les grandes bactéries modèles, principalement *E. coli*, *C. crescentus/vibrioides*, *B. subtilis*, *S. aureus* et *S. pneumoniae* (figure 1.21). Il existe donc un véritable sous-échantillonnage puisque 5 bactéries sont loin de représenter la diversité des millions d'espèces de bactéries existantes. Cela suggère que nous n'avons pas encore apprécié toute la diversité des mécanismes moléculaires mis en jeu lors du cycle cellulaire chez les bactéries.

Par ailleurs, il est admis qu'un grand nombre de gènes au sein des génomes bactériens ne possèdent pas de fonction identifiée, leur valant l'appellation de protéines hypothétiques [10]. On retrouve un certain nombre de ces gènes au sein des clusters génomiques du cycle cellulaire comme le cluster DCW [310], [134], [456] . Chez *S. pneumoniae*, le cluster DCW présente en effet 4 gènes dont la fonction n'est pas connue (Spr1506, Spr1507, Spr1509, Spr1512). De plus, un certain nombre de gènes présentant des fonctions *a priori* non reliées au cycle cellulaire sont présents dans le DCW comme IleS, une Isoleucyl-ARNt synthétase. L'ensemble de ces gènes pourrait représenter un pool conséquent de gènes potentiellement impliqués dans le cycle cellulaire et constituer des systèmes moléculaires inédits.

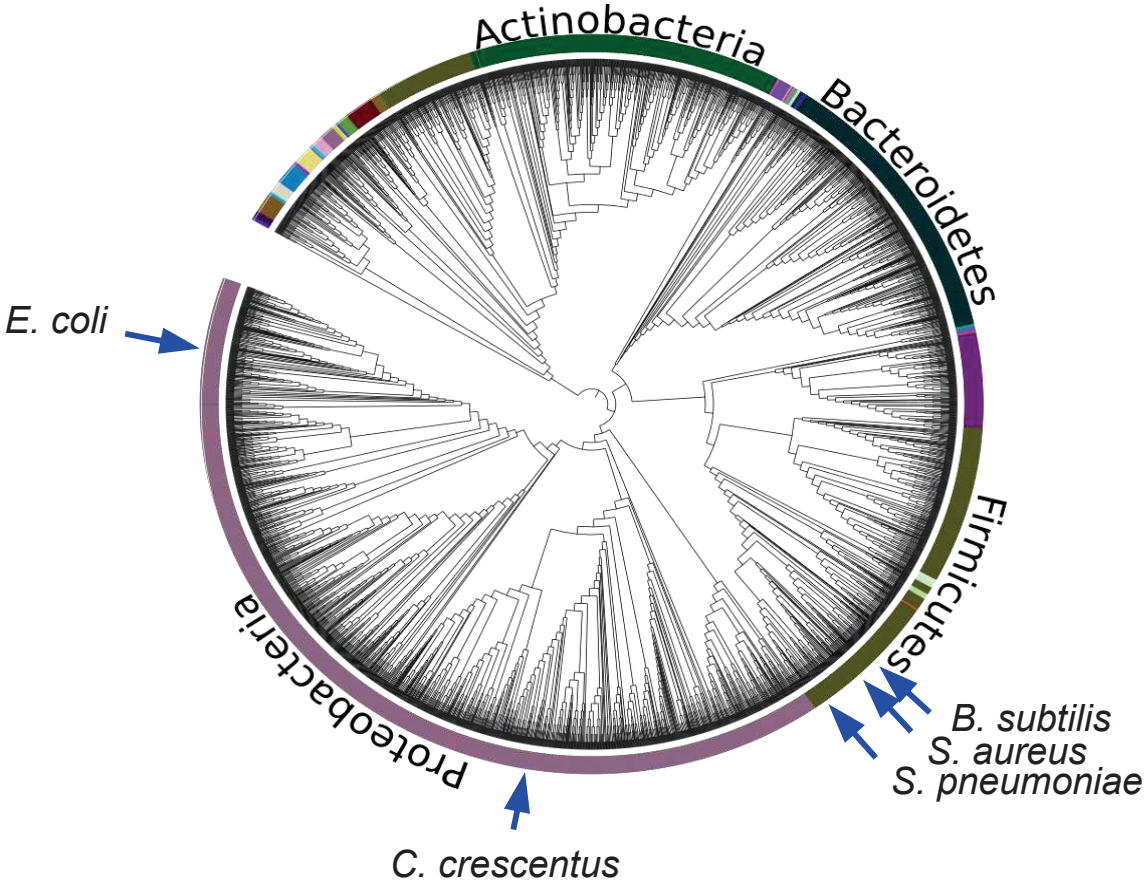


FIGURE 1.21 – Représentativité des bactéries modèles chez les l'ensemble des bactéries. Les flèches bleues correspondent aux organismes modèles dont le processus du cycle cellulaire a été étudié. L'arbre a été inféré à partir des séquences d'ARN 16S de 2975 espèces de bactéries et construit avec Fasttree.

Chapitre 2

La phylogénomique : principes et applications

Sommaire

2.1 De la classification morphologique du vivant à la phylogénomique 96

2.1.1 La classification des êtres vivants et la phylogénie 96

2.1.2 La génomique 97

2.1.3 La phylogénomique 98

2.2 La construction de familles de gènes homologues et orthologues . 100

2.2.1 Choix des séquences initiales et de la base de données 101

2.2.2 Construction des familles d'homologues et d'orthologues 102

2.3 Inférence de l'histoire évolutive d'un gène par la phylogénie . . . 107

2.3.1 Alignement multiple 107

2.3.2 Inférence phylogénétique 108

2.3.3 Arbre d'espèces et identification d'événements évolutifs 114

2.4 Annotation fonctionnelle des génomes 116

2.4.1 Annotation par homologie et orthologie 116

2.4.2 Données phylogénomiques : comparaison entre familles de gènes . . . 117

2.5 Les *Firmicutes* 128

2.5.1 Généralités 128
2.5.2 Taxonomie et Systématique des *Firmicutes* 131

Afin de clarifier les liens fonctionnels entre protéines du cycle cellulaire et d'identifier de potentiels nouveaux mécanismes, plusieurs approches expérimentales sont envisageables. Il est par exemple possible d'identifier une interaction entre deux protéines par Résonnance plasmonique de Surface ou par la méthode du double hybride. Néanmoins, ces méthodes nécessitent des *a priori*, c'est-à-dire des hypothèses déjà formulées. D'autres approches haut débit et sans *a priori* on été développées. Au niveau expérimental, on peut citer le Tn-seq qui permet d'identifier des protéines reliées génétiquement et fonctionnellement. Néanmoins, ces méthodes sont en général coûteuses, longues à mettre en oeuvre et difficilement applicables à un large ensemble d'organismes. En parallèle, des approches d'analyses *in silico* basées sur des données génomiques ont été de plus en plus utilisées pour faire des prédictions fonctionnelles, par exemple dans le cadre de l'annotation des génomes. Plus généralement, elles sont également utilisées pour étudier les liens fonctionnels entre protéines et l'étude évolutive des grands systèmes. Dans cette thèse, nous avons à travers une approche de phylogénomique, utilisé l'information évolutive des séquences déjà caractérisées afin d'inférer des liens fonctionnels entre les protéines du cycle cellulaire mais également d'identifier de potentiels nouveaux composants de ce grand processus cellulaire.

Pour cela, il est d'abord nécessaire d'identifier l'ensemble des séquences homologues de toutes les protéines du cycle cellulaire, de reconstruire leur histoire évolutive puis de comparer ces dernières afin d'identifier des co-évolutions entre familles de protéines. Dans ce chapitre, nous allons d'abord présenter de façon très générale les fondements et l'histoire de la phylogénie et de la phylogénomique, puis nous détaillerons les méthodes de construction des familles de gènes ou de protéines. Nous aborderons ensuite les méthodes de reconstruction de l'histoire évolutive de familles de gènes à travers les arbres phylogénétiques puis nous détaillerons les approches évolutives utilisées afin de détecter des liens fonctionnels entre gènes et d'annoter fonctionnellement des gènes. Enfin, nous décrirons le *phylum* bactérien que nous avons étudié dans cette étude : les *Firmicutes*.

2.1 De la classification morphologique du vivant à la phylogénomique

2.1.1 La classification des êtres vivants et la phylogénie

La classification des êtres vivants est une science ancienne qui remonte à l'antiquité. Tout d'abord, les espèces ont été classées sur la base de leurs caractères morphologiques et sans notion évolutive. Par exemple, Aristote classait la matière "vivante" en quatre catégories : minérale, végétale, animale et humaine. Durant des siècles, de nombreux scientifiques ont tenté de classer les organismes vivants mais ce n'est qu'en 1735 que Carl von Linné propose un système de classification uniformisé, basé également sur les critères de similarité de morphologie, dans l'ouvrage *Systema Naturae* [488]. Trois niveaux furent rapidement ajoutés pour donner le système de classification : Règne, Embranchement, Classe, Ordre, Famille, Genre et Espèce (figure 2.1A). De cette classification seule la nomenclature est encore d'actualité bien qu'elle se soit encore complexifiée avec l'ajout de certains niveaux comme les sous-classes ou les sous-ordres. Néanmoins, Carl von Linné était fixiste, il pensait donc qu'aucune transformation des espèces n'avait lieu et que les espèces étaient immuables. C'est Jean-Baptiste Lamarck qui introduit pour la première fois la notion d'évolution et de transformisme dans son ouvrage Philosophie Zoologique [257]. Il émet l'hypothèse que les organismes évoluent spontanément et transmettent leurs caractères aux générations ultérieures. Cette théorie est ensuite complétée par la théorie de l'évolution proposée par Charles Darwin en 1859 dans *On the origin of Species by Means of Natural Selection* [79]. Il introduit la notion de sélection naturelle qui conduit aux lignées les plus adaptées à l'environnement au sein d'une population. Également, il pose les bases de la phylogénie en introduisant la représentation sous forme d'arbre des relations de parenté. Néanmoins, l'arbre qu'il représente est basé sur des caractères morphologiques et ne prend pas en compte les micro-organismes découverts au XVIème siecle par Antoni von Leewenhooek.

Au milieu du XXeme siècle, Willi Hennig introduit la méthodologie de la systématique phylogénétique. Il s'agit de regrouper les organismes par leurs caractères hérités de façon verticale

[196]. De façon concomitante, les premières séquences protéiques sont séquencées puis dans les années 70, les premières séquences nucléiques. On comprend alors que les gènes sont soumis à l'évolution et que la classification des organismes peut être construite à partir de ces séquences biologiques [528]. En 1977, Carl Woese définit trois domaines du vivant basé sur les séquences d'ARN 16S séquencés par la méthode de Sanger [508] puis reviens sur la nomenclature en 1990 pour définir les domaines actuellement utilisés [509] : Les Archées, les Bactéries et les Eucaryotes (figure 2.1B) .

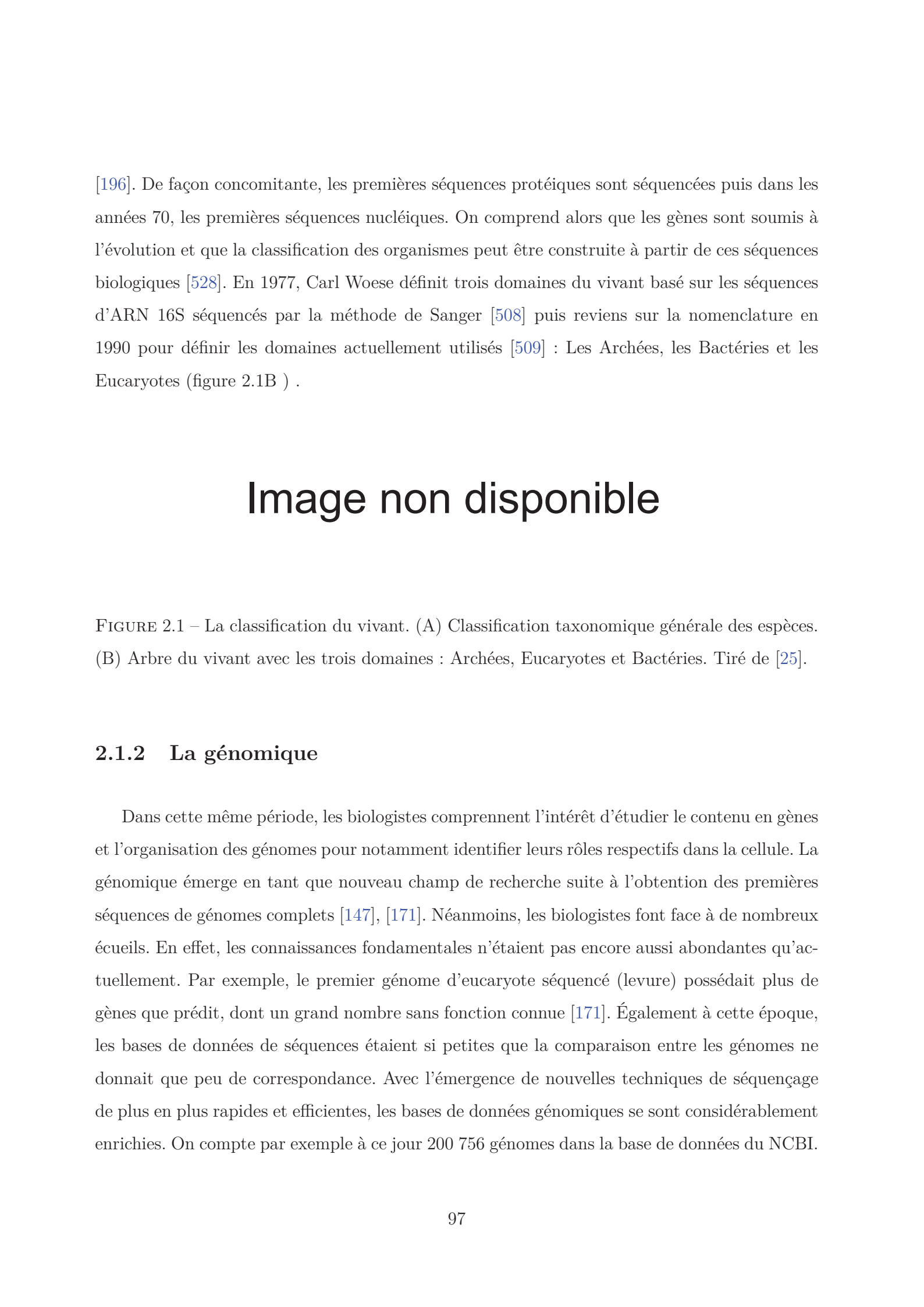


FIGURE 2.1 – La classification du vivant. (A) Classification taxonomique générale des espèces. (B) Arbre du vivant avec les trois domaines : Archées, Eucaryotes et Bactéries. Tiré de [25].

2.1.2 La génomique

Dans cette même période, les biologistes comprennent l'intérêt d'étudier le contenu en gènes et l'organisation des génomes pour notamment identifier leurs rôles respectifs dans la cellule. La génomique émerge en tant que nouveau champ de recherche suite à l'obtention des premières séquences de génomes complets [147], [171]. Néanmoins, les biologistes font face à de nombreux écueils. En effet, les connaissances fondamentales n'étaient pas encore aussi abondantes qu'actuellement. Par exemple, le premier génome d'eucaryote séquencé (levure) possédait plus de gènes que prédit, dont un grand nombre sans fonction connue [171]. Également à cette époque, les bases de données de séquences étaient si petites que la comparaison entre les génomes ne donnait que peu de correspondance. Avec l'émergence de nouvelles techniques de séquençage de plus en plus rapides et efficientes, les bases de données génomiques se sont considérablement enrichies. On compte par exemple à ce jour 200 756 génomes dans la base de données du NCBI.

Cette augmentation massive du nombre de génome s'est accompagnée du développement de méthodes rapides pour traiter l'information contenue dans les génomes.
Parallèlement, les techniques de biologie moléculaire ont permis de caractériser fonctionnellement un grand nombre de protéines chez certaines espèces. Cette augmentation fulgurante du nombre de protéines caractérisées biochimiquement/biologiquement a fourni des informations précieuses pouvant être utilisées pour l'annotation des génomes et notamment par des approches comparatives. En effet, en partant d'une séquence dont la fonction est connue, une approche classique consiste à faire l'hypothèse que ses homologues et plus précisément ses orthologues au sein d'autres génomes rempliraient la même fonction.

2.1.3 La phylogénomique

Les méthodes d'annotation fonctionnelle des génomes se sont sophistiquées en intégrant, en plus de l'information évolutive contenue dans les séquences, des informations telles que l'organisation des gènes au sein des génomes. Concrètement, une analyse phylogénomique d'un gène consiste en la reconstruction de son histoire évolutive au sens large. Cela intègre la reconstruction de sa phylogénie et l'élucidation des liens de parenté avec ses homologues, mais aussi l'étude de l'évolution des contextes génomiques, de sa composition en domaines fonctionnels, de sa distribution taxonomique, ... (figure 2.2). Ces informations permettent d'identifier les événements évolutifs majeurs ayant impacté la famille de gène. Ces événements peuvent être de plusieurs natures et concerner soit le gène lui même, soit l'environnement du gène dans les génomes (figure 2.3). En comparant les histoires évolutives des gènes à l'échelle des génomes, il est alors possible d'établir des liens entre familles de gènes et en particulier d'identifier des familles de gènes présentant des histoires évolutives similaires qui traduisent potentiellement des liens fonctionnels.

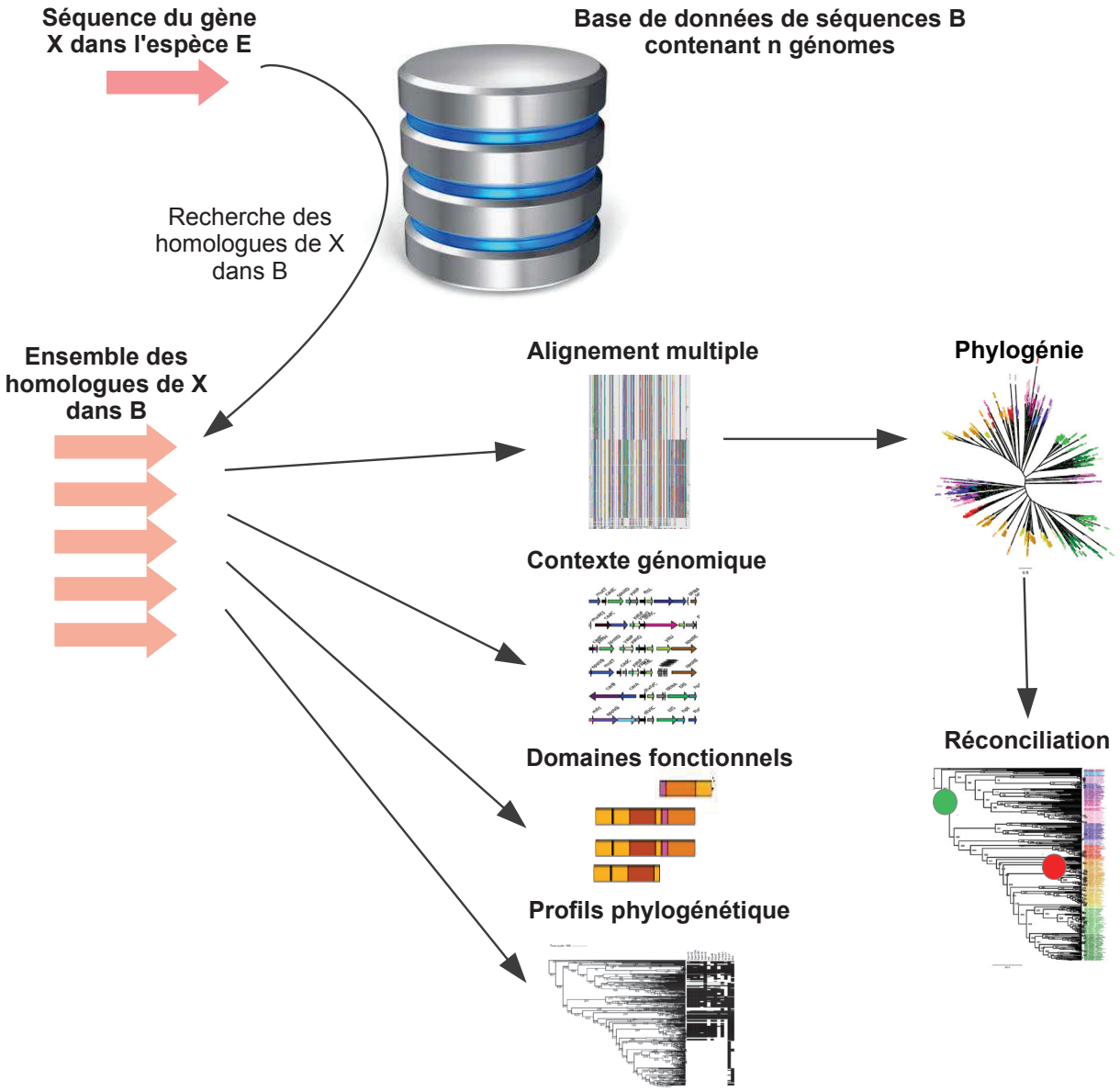


FIGURE 2.2 – Schéma général d'une étude phylogénomique. Une séquence X (ou un ensemble de séquences) est utilisée pour construire un jeu de données d'homologues à partir d'une base de données B. De nombreuses informations sont extraites de ce jeu de données : les séquences peuvent être alignées pour générer la phylogénie de gènes. Cette dernière peut être réconciliée avec un arbre d'espèces afin d'identifier les événements évolutifs ayant affecté la famille de gènes. D'autres informations peuvent être analysées comme la composition en domaines fonctionnels des séquences, le contexte génomique ou encore les profils de présence/absence.

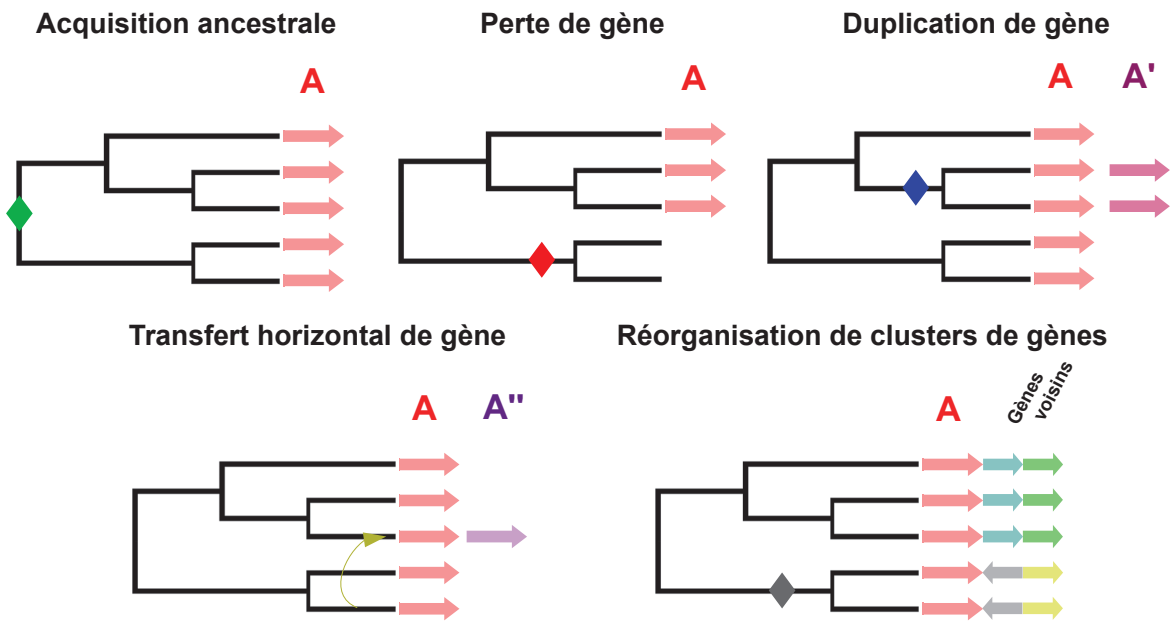


FIGURE 2.3 – Quelques exemples d'événements évolutifs. Les événements sont représentés par des losanges ou des flèches. A : gène d'intérêt, A' : paralogue du gène A, A" : xénologue du gène A.

2.2 La construction de familles de gènes homologues et orthologues

La première étape dans une analyse phylogénomique est la construction des familles de gènes homologues. Il est pour cela nécessaire de partir de séquences caractérisées expérimentalement et de rechercher toutes les séquences homologues dans une base de données de séquences.

2.2.1 Choix des séquences initiales et de la base de données

2.2.1.1 Séquences initiales

Tout d'abord, il est crucial d'établir les bornes de l'analyses, c'est-à-dire quels *taxa* et quels gènes analyser. Ainsi, il est nécessaire de constituer une liste de gènes/protéines d'intérêt à étudier. Cette liste varie en fonction de la nature de la question qui est posée. Pour reconstruire la phylogénie des espèces, on peut par exemple utiliser les séquences des protéines ribosomiques ou encore toutes les familles de gènes présentes de façon ubiquitaire et en une seule copie dans le taxon d'intérêt (« single gene core ») [396], [259]. Pour des études fonctionnelles qui consistent en la description de l'évolution d'un processus cellulaire comme celle décrite dans cette thèse, il est nécessaire de sélectionner uniquement les gènes impliqués dans ce processus.

2.2.1.2 Bases de données

Pour mener à bien la recherche d'homologues, il est nécessaire d'assembler une base de données de séquences contenant des génomes du taxon d'intérêt sur lesquels sera effectué la recherche. Cette base de données peut encore une fois être de nature protéique ou nucléique (protéique à grande échelle évolutive et nucléique à petite échelle évolutive). Il existe de nombreuses bases de données publiques de séquences génomiques. Parmi elles, celle du NCBI contient notamment les séquences nucléiques des génomes, des gènes et les séquences des protéines associées (`https://www.ncbi.nlm.nih.gov/`). La base de données du NCBI contient également des informations taxonomiques cohérentes unifiées et à jour [137]. La base de données Uniprot est également très utilisée [45]. Elle fournit un grand nombre de données sur les protéines telles que la structure, la composition en domaines fonctionnels, le rôle cellulaire ou encore les données d'interaction protéine-protéine.
La plupart des génomes séquencés à ce jour ne sont pas complets. En effet, certaines régions génomiques sont particulièrement difficiles à séquencer. De plus, l'assemblage des génomes est une tâche complexe qui ne conduit que rarement à la constitution complète du génome. Afin d'être exhaustif dans la recherche des homologues, il convient d'utiliser uniquement des

génomes complets, surtout lorsqu'il est nécessaire d'identifier des liens entre familles de gène.

2.2.2 Construction des familles d'homologues et d'orthologues

La construction des familles protéiques consiste pour chaque séquence initiale à rechercher l'ensemble des séquences présentes dans la base de données présentant une similarité de séquence. L'hypothèse sous-jacente est que des séquences présentant une similarité significative partagent un ancêtre commun. Cette recherche peut être effectuée par différentes approches que nous allons décrire.

2.2.2.1 Recherche par BLAST

BLAST ou Basic Local Alignment Search Tool est un programme basé sur un algorithme permettant de rechercher les séquences similaires d'une séquence initiale dans une base de données [4]. L'algorithme de BLAST peut être décomposé en quatre étapes (figure 2.4A).

- Le découpage de la séquence initiale (requête) en mots d'une longueur fixe.
- La recherche d'une similarité entre ces mots au sein de l'intégralité des séquences présentes dans la base de données (séquences cibles).
- L'extension des mots similaires identifiés en amont et en aval avec le calcul d'un score de similarité. Les alignements locaux correspondent à des LMSP (« Localy Maximal scoring Segment pair »).
- La terminaison de l'extension lorsque l'extrémité d'une des deux séquences est atteinte ou lorsque le score de similarité de la séquence alignée est inférieur à un seuil défini. Les alignements locaux sélectionnés correspondent à des HSP (« High scoring Segment Pair »). La HSP présentant le score le plus élevé correspond au MSP (« Maximum scoring Segment Pair ») et sera considéré comme le meilleur hit.

L'algorithme a ensuite été complexifié, notamment par le fait qu'il a la possibilité d'insérer des gaps lors de l'alignement.

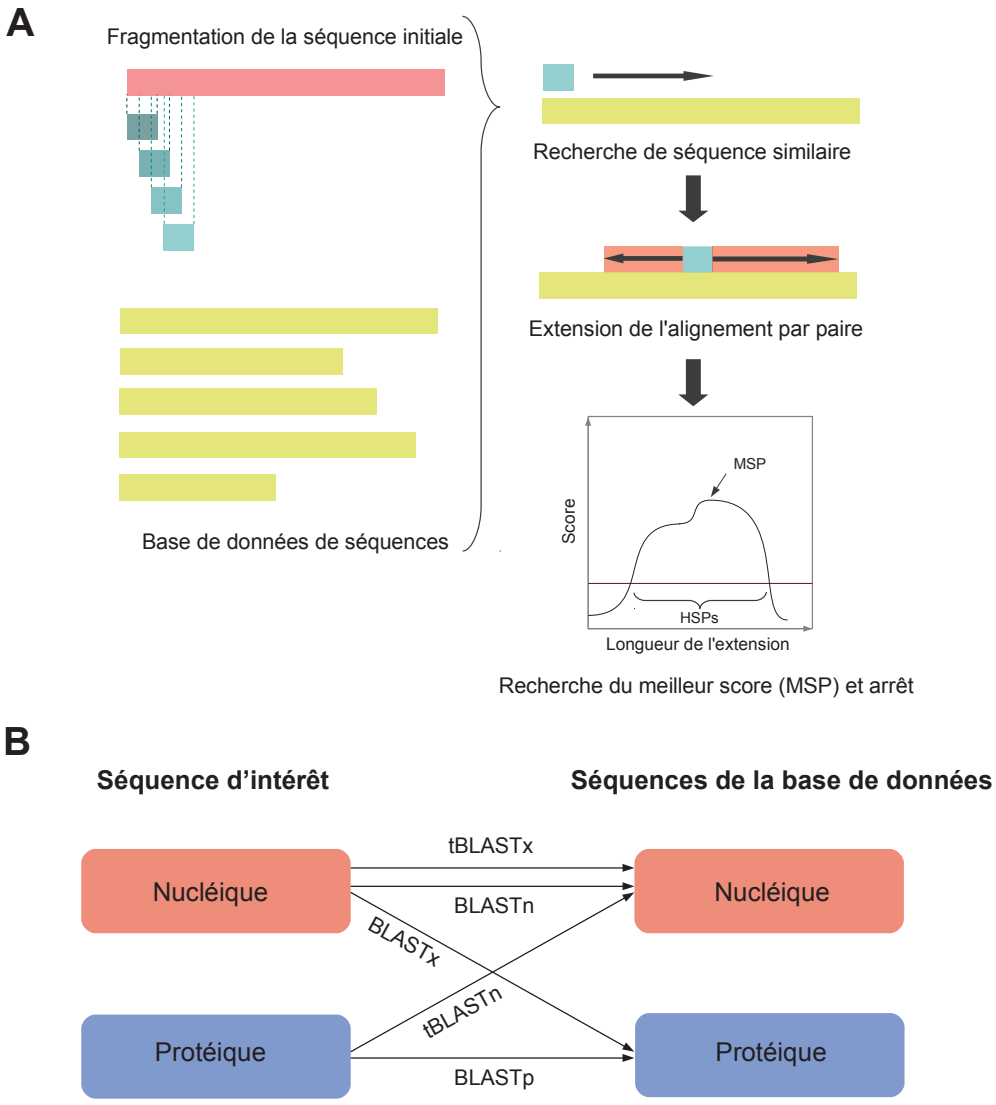


FIGURE 2.4 – Principe et types de BLAST. (A) Principe de l'algorithme de BLAST. (B) Différents types de BLAST en fonction de la nature des séquences.

Le calcul du score d'un alignement repose sur une fonction qui permet de pondérer chacun des événements modélisés sur l'alignement à savoir : l'identité de deux résidus, la substitution d'un résidu et l'insertion d'un gap. Le coût associé à une substitution ou une identité repose sur des matrices de substitution comme par exemple la matrice BLOSUM62 [118]. Ces matrices empiriques ont été construites sur la base de des jeux de données de référence. Les pénalités liées à l'insertion d'un gap se décompose en une pénalité pour l'ouverture et une pénalité pour

l'extension.

Afin d'évaluer la significativité de la similarité obtenue entre une séquence initiale et une séquence cible, la E-value a été introduite [235]. Cette valeur correspond à l'espérance du nombre de MSPs de score supérieur ou égal au score du MSP d'intérêt, lors de la comparaison de deux séquences aléatoires de longueur et de composition identique. Ainsi, plus la similarité est élevée, plus le score est élevé et plus la E-value diminue. Néanmoins, il est nécessaire d'établir un seuil au dessus duquel les hits ne sont plus considérés comme significatifs. Pour cela, la majorité des auteurs dans la littérature utilisent des seuils arbitraires, tels que 10^{-4} lorsque les séquences sont protéiques.

Enfin, il est à noter qu'il existe maintenant une multitude de types de BLAST permettant par exemple de requêter une base de données nucléique avec une séquence protéique ou inversement (tBLASTn et BLASTx). L'ensemble de ces méthodes est présenté en figure 2.4B.

La recherche par BLAST est très efficace pour identifier les homologues d'une séquence. Néanmoins, la sensibilité de BLAST est souvent insuffisante. En effet, il est fréquent de ne pas identifier l'ensemble des homologues sur la base d'une seule recherche par BLAST, surtout quand la séquence initiale présente des homologues lointains. Il est donc souvent nécessaire de réaliser plusieurs recherches en utilisant des séquences requêtes différentes (approche itérative). Les séquences requêtes peuvent être définies *a priori* ou être issues de la première recherche.

2.2.2.2 Recherche par profil

La principale limitation de la recherche par BLAST est d'etre basée sur l'information contenue dans une seule séquence (séquence requête) et d'utiliser des matrices de substitution génériques. Ainsi, quelques soient les séquences analysées, les substitutions seront considérées de la même manière à une position variable ou au contraire très conservée. Dans le but d'optimiser la recherche d'homologues, le PSI-BLAST a été développé par Altschul et collègues en 1997 [5]. La première étape de PSI-BLAST est une recherche par BLAST classique. L'ensemble des séquences présentant une E-value inférieure à un seuil défini par l'utilisateur, est sélectionné puis aligné. L'alignement permet de construire un profil : la PSSM (« Position Specific Scoring Matrix »). Cette matrice contient typiquement pour chaque colonne de l'alignement un score

égal au logarithme du rapport de la fréquence observée de chaque acide aminé/acide nucléique sur la fréquence théorique. Un poids est attribué à chaque séquence en fonction de la quantité d'information qu'elle contient (les séquences redondantes ou très proches auront un poids plus faible). La matrice est ensuite utilisée pour requêter la base de données. Les séquences présentant une E-value inférieure au seuil sont sélectionnées, ajoutées aux précédentes, alignées et utilisées pour faire une nouvelle PSSM. Le nombre d'itérations est fixé par l'utilisateur.
L'approche du PSI-BLAST est très performante mais présente quelques inconvénients. Notamment, la recherche par PSSM est fortement dépendante du jeu de données initial construit après le premier BLAST. Ce jeu de données est donc également dépendant du seuil choisi arbitrairement par l'utilisateur. Une alternative possible à la recherche par PSSM est l'utilisation de modèles de Markov cachés [117]. Il s'agit d'une approche qui se rapproche de la recherche par PSSM mais diffère notamment dans l'estimation des probabilités de transitions, basée sur un modèle de Markov. Le programme le plus utilisé pour effectuer des recherches par profil HMM est HMMMER [119].

2.2.2.3 Homologues, orthologues, paralogues et xénologues

Il arrive très fréquemment que les homologues d'un gène ne forment pas une seule famille mais plusieurs familles de gènes (famille monogénique et famille multigénique). En effet, l'histoire d'un gène est souvent impactée par de nombreux événements de gènes comme les pertes, duplications et transferts horizontaux de gènes (figure 2.5). La duplication d'un gène d'une famille d'orthologues génère une famille paralogue à la famille initiale. Lorsqu'un transfert d'un gène survient, la famille issue de ce transfert est xénologue par rapport à la famille d'origine. Les xénologues et les paralogues présentent de façon générale des rôles cellulaires différents [247]. Il est donc préférable, pour une analyse fonctionnelle, de constituer des jeux de données de gènes orthologues puisque ceux-ci remplissent de façon plus probable la même fonction au sein des différentes espèces. Néanmoins, l'orthologie n'est pas une condition suffisante pour affirmer catégoriquement que deux gènes jouent le même rôle cellulaire [247].

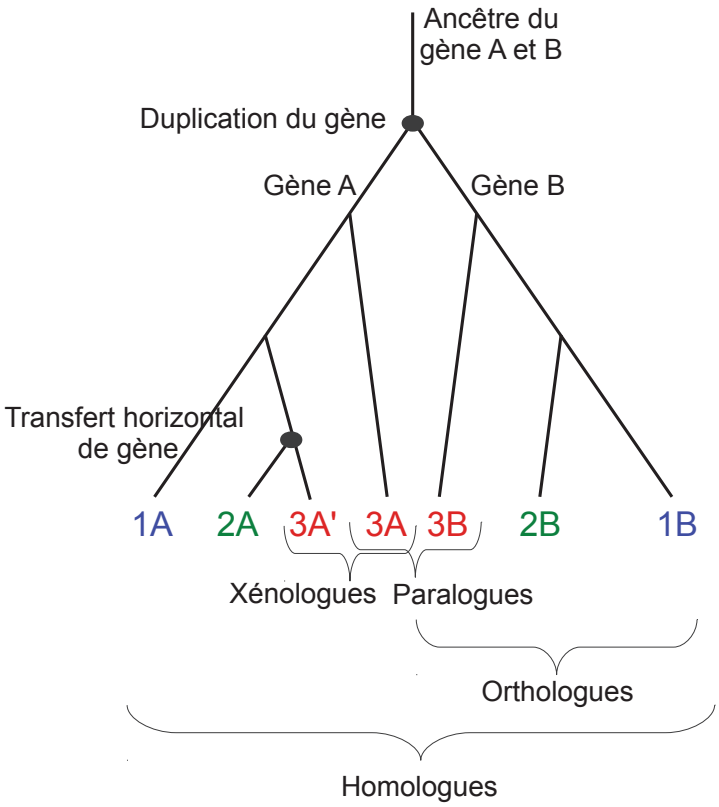


FIGURE 2.5 – Homologie, orthologie, paralogie et xénologie. A l'ancêtre des espèces 1, 2 et 3, un gène s'est dupliqué en A et B. Les deux copies se sont répandues chez les espèces 1, 2 et 3. Une copie A a été transférée de l'espèce 2 à l'espèce 3 donnant le gène A'. 3A et 3B sont paralogues, 3A'et 3A sont xénologues, tous les gènes B sont orthologues. L'ensemble des séquences dérive d'une séquence ancestrale, elles sont toutes homologues.

Pour délimiter des familles de gènes orthologues, il existe plusieurs approches. la plus utilisée est celle des COG (« Clusters of Orthologous Groups ») [458]. Il s'agit d'une technique qui consiste en la comparaison de toutes les séquences d'une base de données entre elles par BLAST et en la clusterisation des familles de gènes. Tout d'abord, un BLAST tout contre tout est effectué pour l'intégralité de la base de données. Entre deux génomes, deux séquences sont regroupées si, pour chacune d'entre elles, le BLAST entre elles correspond à la E-value minimale. Bien que leur efficacité soit incontestable, ces méthodes sont automatisées et peuvent conduire à des erreurs. Notamment, certaines familles paralogues sont regroupées (*e.g.* DivIVA et GpsB chez les *Firmicutes*) entre elles tandis que certaines familles orthologues sont divisées

en plusieurs groupes (*e.g.* DivIC chez les *Firmicutes*). Ces approches sont donc interessantes pour traiter de grandes quantités de données rapidement mais sont inadaptées quand il s'agit de réaliser des analyses précises.
L'approche la plus performante consiste à construire les familles de gènes de façon individuelle par BLAST et/ou profils, puis d'identifier par la phylogénie et en la comparant avec une phylogénie d'espèces les éventuelles sous-familles [247]. C'est cette technique que nous avons utilisé dans le cadre de ma thèse. Bien que cette approche soit coûteuse en temps et ne puisse pas être automatisée facilement, c'est la manière la plus précise de procéder pour identifier les familles d'orthologues.

2.3 Inférence de l'histoire évolutive d'un gène par la phylogénie

une fois l'identification des éventuelles sous-familles auxquelles les homologues identifiés appartiennent, une analyse détaillée de la sous-famille d'interêt doit être réalisée. Il s'agit notamment d'inférer son histoire évolutive et de déterminer les liens de parenté entre les séquences qui la composent. L'approche classique consiste à inférer sa phylogénie. En comparant l'arbre inféré avec une phylogénie des espèces, il est possible d'identifier les événements majeurs ayant impacté la famille d'interêt.

2.3.1 Alignement multiple

En préalable à toute analyse phylogénomique, il est necessaire d'aligner les séquences afin de retrouver les relations d'homologie entre les sites. Initialement chez une espèce ancestrale, une séquence nucléotidique a subit un certain nombre de modifications : les mutations. Ces mutations sont de plusieurs sortes : la substitution qui correspond à la modification ponctuelle d'un résidu, l'insertion qui consiste en l'intégration d'un ou plusieurs résidus et à l'inverse, la délétion. L'accumulation de ces mutations conduit à la dérive de cette séquence qui se propage

au sein des espèces descendantes. On obtient ainsi plusieurs séquences qui se ressemblent mais qui ne sont pas identiques. Concrètement, l'alignement consiste à mettre en face les résidus d'acides nucléiques ou d'acides aminés provenant d'un même résidu ancestral les uns en face des autres.
Plusieurs approches ont été développées afin d'aligner plus de trois séquences comme l'alignement progressif implémenté dans CLUSTAL [65] ou encore l'alignement par transformées de Fourrier rapides implémenté dans MAFFT [238].
Néanmoins, l'alignement multiple peut conduire à des biais lors de l'inférence phylogénétique [62]. En effet, un grand nombre de gaps dans l'alignement ou des positions évoluant trop rapidement peuvent engendrer des difficultés à aligner le jeu de données. Il est donc nécessaire d'éliminer les positions non informatives ou dont l'alignement est ambigu. Cette opération est appelée nettoyage ou « Trimming ». Pour cela, plusieurs logiciels ont été développés comme Gblocks [62] ou BMGE [72].

2.3.2 Inférence phylogénétique

A partir d'un alignement du jeu de données de séquences biologiques, il est possible de construire l'arbre phylogénétique de la famille de gènes. Plusieurs méthodes ont été développées mais nous n'allons présenter que les méthodes les plus utilisées actuellement (Maximum de vraisemblance et Bayésien).
Un arbre phylogénétique est une représentation des liens de parenté entre des gènes homologues, c'est-à-dire ayant tous un ancêtre commun. Il s'agit mathématiquement d'un graphe acyclique composé de nœuds et d'arêtes (branches). Les branches distales sont les feuilles de l'arbre et correspondent aux séquences utilisées pour inférer l'arbre tandis que les branches internes représentent les ancêtre communs de ces séquences. Généralement, un arbre phylogénétique est dit binaire car chaque branche interne possède deux branches filles, bien que le manque de signal et l'utilisation de certains logiciels conduisent à la production d'arbres non binaires (aussi appelés multifurqués). Dans un certain nombre de cas, il est nécessaire d'orienter un arbre, c'est-à-dire de lui donner un sens chronologique. Pour cela, il faut raciner

l'arbre, c'est-à-dire sélectionner une branche qui sera basale à l'arbre et qui représentera l'ancêtre commun de la famille étudiée [370]. La méthode la plus utilisée est l'enracinement par un groupe externe. Cela consiste en l'ajout d'une ou plusieurs séquences homologues provenant d'un groupe taxonomique extérieur.

2.3.2.1 Modèles évolutifs

Un modèle évolutif (ou de substitution) permet de décrire le processus évolutif des séquences. Il existe des modèles de substitution protéiques ou nucléiques. Soit une séquence S, chaque site i de S est dans un état Ei (A, T, G ou C pour une séquence nucléique par exemple). Un modèle de substitution rend compte de la probabilité de substitution (taux d'évolution) entre chaque état Ei et tous les autres états possibles pour décrire comment évolue un site et plus généralement une séquence (figure 2.6). Dans la plupart des cas, un processus de Markov est utilisé pour décrire le processus évolutif. Un processus de Markov correspond à une modélisation d'une succession d'états au cours du temps avec des probabilités associées à chaque changement d'état [412]. Ainsi, on peut distinguer deux cas : les probabilités de substitution (changement de l'état Ei) et les fréquences d'équilibre (maintien à l'état Ei).

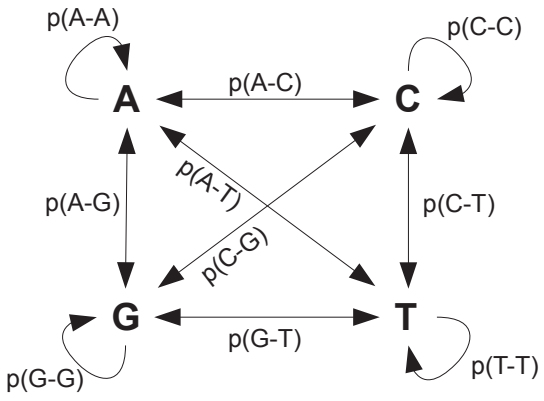


FIGURE 2.6 – Modèle général de substitution des acides nucléiques. Les flèches représentent les transitions.

Dans la plupart des modèles, plusieurs hypothèses sont faites quant aux processus évolutifs.

Les processus sont considérés comme réversibles, c'est-à-dire que les probabilités de passer de A à T et de T à A par exemple, sont égales. De plus, le système est considéré comme étant stationnaire, c'est-à-dire qu'il a atteint un équilibre. Aussi, il est postulé que tous les sites évoluent à la même vitesse, c'est l'hypothèse d'uniformité. Également, le processus est considéré comme homogène, c'est-à-dire qu'il ne varie pas au cours du temps et est le même pour tous les sites. Enfin, les sites sont tous considérés comme indépendants les uns des autres.
La plupart de ces hypothèses sont en réalité des simplifications non réalistes d'un point de vue biologique. Plusieurs méthodes ont donc été développées afin de s'affranchir de ces hypothèses. Par exemple, la plupart des modèles sont corrigés par une estimation globale du taux d'évolution pour s'affranchir de l'hypothèse d'uniformité. Un facteur correctif r est donc ajouté au modèle pour rendre compte de l'hétérogénéité du taux d'évolution entre les sites. Classiquement, une distribution gamma est utilisée [518]. La distribution gamma, exprimée en fonction de r, change de forme en fonction d'un paramètre nommé alpha. Pour un alignement donné, le paramètre alpha est estimé donnant ainsi la forme de la distribution. La distribution est ensuite discretisée (classiquement 4 catégories), chaque site étant classé ainsi dans une catégorie. Il existe aussi des modèles plus complexes qui permettent de s'affranchir de l'hypothèse d'homogénéité comme le modèle CAT [262]. Il s'agit de catégoriser les sites en fonction de leur composition et d'appliquer un processus évolutif différent pour chaque catégorie, notamment en utilisant différentes matrices de substitution.
Il existe une multitude de modèles décrivant le processus évolutif des séquences. Le modèle nucléique le plus utilisé est le modèle GTR [459], modèle qui calcule les taux de substitution à partir du jeu de données. Pour les séquences protéiques, plusieurs modèles ont été proposés comme le modèle WAG [502] ou LG [264] qui se basent sur des jeux de données de référence pour calculer les taux de substitution entre les acides aminés.

2.3.2.2 Inférence phylogénétique en maximum de vraisemblance

Afin d'estimer l'adéquation entre les données (alignement) avec un modèle évolutif et une topologie donnée, il est possible de calculer la vraisemblance. La vraisemblance $L(\theta)$ est définie comme la probabilité d'observer les données S sachant les paramètres du modèle et de l'arbre

θ :

$$L(\theta) = \mathbb{P}(S|\theta) \tag{2.1}$$

Le paramètre θ comprend la topologie de l'arbre τ, le vecteur des longueurs des branches b et les paramètres du modèle ϑ.

Afin d'obtenir l'arbre le plus vraisemblable parmi tous les arbres possibles pour un nombre de feuilles données, il est théoriquement nécessaire d'explorer l'ensemble des topologies. Néanmoins, cette exploration exhaustive n'est techniquement pas possible en raison du très grand nombre de topologies possibles pour un nombre même restreint de feuilles. Les topologies sont donc explorées selon des heuristiques. La construction d'un arbre en maximum de vraisemblance s'effectue donc par itérations successives. A la première itération, un arbre initial est généré. Pour chaque site i de l'alignement donné, la vraisemblance de l'arbre est calculée. Cette vraisemblance correspond à la somme des probabilités conditionnelles de l'ensemble des scénarios possibles expliquant les états observés dans l'alignement à cette position i. Ces probabilités conditionnelles sont calculées à partir des paramètres de l'arbre (τ, b, ϑ). La somme des vraisemblances par site donne la vraisemblance de l'arbre. Classiquement, la topologie n est ensuite modifiée pour donner la topologie $n+1$ dont la vraisemblance est calculée. Les longueurs de branches sont aussi modifiées pour maximiser la vraisemblance. Si la vraisemblance de la topologie $n+1$ est plus élevée, c'est elle qui va à son tour être modifiée pour la prochaine itération. Dans le cas contraire, la topologie n va être modifiée différemment et utilisée pour la prochaine itération. Les topologies sont modifiées principalement par deux algorithmes : le NNI (« Nearest Neighbor Interchange ») ou le SPR (« Subtree Pruning and Regrafting ») que nous ne détaillerons pas ici. La topologie avec le maximum de vraisemblance est alors sélectionnée. Parmi les programmes de construction d'arbres par maximum de vraisemblance, on peut citer PhyML [183], IQ-TREE [348] ou encore RAxML [439].

2.3.2.3 Inférence phylogénétique bayésienne

L'inférence phylogénétique par l'approche Bayésienne est similaire à celle du Maximum de vraisemblance. La différence principale est l'utilisation d'une distribution *a priori* des paramètres de l'arbre, basé sur le théorème de Bayes. L'utilisation d'une distribution *a priori* permet d'établir un *prior*, c'est-à-dire une idée préconçue de la distribution statistique de certains paramètres, comme les longueurs de branches. Dans le cas d'une distribution uniforme, on parle de *prior* non informatif. Cette technique permet d'implémenter une connaissance biologique supplémentaire dans le modèle mais nécessite une extrême précaution dans le choix des *prior*. La limitation du théorème de Bayes est qu'il est nécessaire en théorie de connaître la probabilité des données sachant l'arbre pour l'ensemble des arbres possibles. C'est pourquoi des méthodes heuristiques sont utilisées comme la MCMCMC (« Metropolis Coupling of Markov Chain Monte-Carlo »). La chaîne de Markov avec la technique de Monte-Carlo (MCMC) est semblable à un marcheur dans la montagne (espace des topologies) qui cherche le sommet (maximum de vraisemblance) (figure 2.7). La chaîne avance tant que la vraisemblance croit. Si elle décroit, la chaîne avance avec une probabilité d'autant plus faible que la différence de vraisemblance est élevée. Le couplage de Métropolis correspond au fait qu'il n'y ait pas une mais plusieurs chaînes, qu'elles parcourent le paysage des topologies avec des pas plus ou moins long (chaînes chaudes et froides) et qu'elles intervertissent leur longueur de pas de temps à autres (chaîne froide qui devient chaude, *et vice versa*). Ainsi, les chaînes arriveront théoriquement à trouver le maximum de vraisemblance. On parle alors de convergence des chaînes. A partir de l'ensemble des itérations et en enlevant la phase d'approche (« burn-in »), on peut déduire la topologie consensus qui correspond aux topologies les plus probables. De nombreux programmes permettent d'inférer des phylogénies par des méthodes bayésiennes comme MrBayes [406] ou PhyloBayes [261]. L'inférence bayésienne permet d'obtenir des phylogénies de bonne qualité mais le temps de calcul peut être très conséquent, surtout si les chaînes convergent difficilement.

FIGURE 2.7 – Illustration du principe des MCMC. Les ellipses correspondent à des maxima locaux de vraisemblance, chaque point correspond à l'itération d'une chaîne. Les chaînes froides sont en bleu et les chaînes chaudes sont en rouge. Plus le nombre d'itérations est grand, plus le paysage est exploré. Tiré de [374].

2.3.2.4 Tests de robustesse des bipartitions

La robustesse de chaque branche de l'arbre doit ensuite être testée. Pour la méthode du maximum de vraisemblance, plusieurs métriques ont été développées. Certaines sont considérées plus optimistes que d'autres. La procédure la plus utilisée est celle du bootstrap non paramétrique [138]. Il s'agit tout d'abord de créer un réplicat de bootstrap en échantillonnant de façon aléatoire avec remise des positions de l'alignement initial pour obtenir un nouvel alignement de taille égale à l'initial, ce qui revient à accorder un poids variable à chaque position. Un arbre est ensuite généré de la même manière que pour l'arbre initial sur la base de cet alignement. La procédure est répétée typiquement 100 ou 1000 fois. Ensuite, pour chaque bipartition de l'arbre initial, le nombre de phylogénies issues des réplicats de bootstrap qui contient la même bipartition est compté puis projeté sur la phylogénie initiale. Cette méthode, bien que très utilisée, nécessite des temps de calcul conséquents, surtout si l'inférence des arbres se fait au maximum de vraisemblance. Des méthodes de bootstrap rapides ont ainsi été développées afin d'accélérer la procédure comme le « RAxML rapid bootstrap » implémenté dans RAxML [440] ou l'« ultrafast bootstrap » implémenté dans IQ-TREE [323], [348]. L'interprétation des valeurs de bootstrap peut se révéler parfois difficile puisqu'il n'existe pas de seuil pour déterminer si la branche est supportée ou non. Il existe ainsi d'autres tests de robustesse comme le aLRT (« approximate Likelihood Ratio Test ») ou encore le SH-aLRT(« Shimodaira–Hasegawa

approximate Likelihood Ratio Test ») [9], [428]. Ceux-ci ont pour avantage d'être calculés plus rapidement que les bootstraps classiques et le seuil de significativité de 5% est couramment admis pour déterminer si une branche est supportée. Pour l'inférence bayésienne, le support utilisé est la probabilité postérieure (PP), c'est-à-dire la probabilité de l'existence d'une bipartition en prenant en compte les priors. Il est à noter que le bootstrap est en général considéré comme plus pessimiste que l'ensemble des autres tests. Récemment, une approche alternative au bootstrap a été proposée [267]. Celle-ci consiste en le calcul non pas de la présence ou l'absence de la bipartition de l'arbre de boostrap dans l'arbre initial mais en un calcul d'une distance graduelle entre la bipartition observée dans l'arbre de boostrap et la bipartition la plus similaire de l'arbre initial, comprise entre zéro et un.

2.3.3 Arbre d'espèces et identification d'événements évolutifs

Afin d'identifier les événements évolutifs qui ont affecté une famille de gènes, il est nécessaire de comparer sa phylogénie à un arbre d'espèces : la réconciliation.
L'arbre d'espèces est censé refléter les relations de parentés entre espèces. Pour cela, il est nécessaire d'utiliser des familles servant marqueurs génétiques. Les familles utilisées comme marqueurs doivent remplir plusieurs conditions. Elles doivent être conservées au sein du taxon étudié et présenter un faible nombre d'événements de duplications et de transferts horizontaux. De plus, les familles ne doivent pas présenter un taux d'évolution trop élevé (saturation substitutionnelle). Les marqueurs couramment utilisés sont les protéines/ARN ribosomiques et les protéines impliquées dans le traitement de l'information génétique (transcription et traduction). Historiquement, les ARN ribosomiques, comme l'ARN 16S ont été utilisés [509]. Les protéines ribosomiques sont également considérés comme des bons marqueurs [396]. Une autre technique est d'utiliser toutes les familles de gènes unicopie conservées dans un taxon (« Single gene core »). Chaque famille est alignée. Les alignements résultants sont nettoyés puis concaténés en une supermatrice qui va servir de base à l'inférence d'une phylogénie.

La comparaison entre un arbre de gènes et un arbre d'espèces, également appelée réconci-

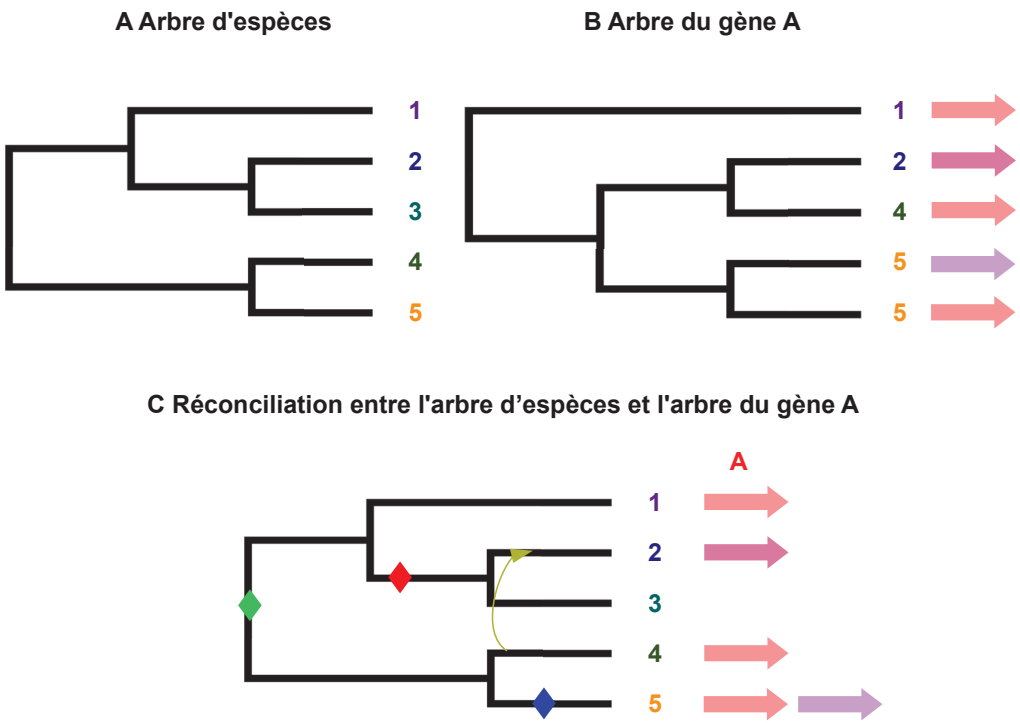


FIGURE 2.8 – Principe de la réconciliation. (A) Arbre des espèces 1, 2, 3, 4 et 5. (B) Arbre du gène A chez les espèces 1, 2, 3, 4 et 5. L'arbre n'est pas concordant avec l'arbre d'espèces. (C) Réconciliation des deux arbres qui explique au mieux l'arbre du gène A. Acquisition de gène : diamant vert, perte de gène : diamant rouge, duplication de gène : diamant bleu, transfert horizontal de gène : flèche jaune.

liation, se fait alors en comparant les topologies des deux arbres. Il est possible de comparer qualitativement les deux arbres en visualisant les topologies. Le principe est d'identifier des événements évolutifs :

- Apparitions de gènes
- Spéciations
- Pertes de gènes
- Duplications de gènes
- Transferts horizontaux de gène (« out » pour les donneurs et « in » pour les receveurs)

Des méthodes quantitatives de réconciliation ont également été développées. Ces techniques pouvant également servir dans la détection de co-évolution sont de ce fait expliquées en section 2.4.2.4.

2.4 Annotation fonctionnelle des génomes

L'annotation des gènes sur un génome consiste à identifier le début et la fin des gènes (annotation syntaxique) mais également à inférer le rôle cellulaire de ces-dits gènes (annotation fonctionnelle). La première étape que nous ne détaillerons pas ici est donc la délimitation des séquences géniques telles que les cadres ouverts de lecture ou les ARN non codants. Plusieurs programmes ont été développés à cet effet comme Prodigal [212] ou Glimmer [90]. Il est ensuite nécessaire d'identifier la fonction potentielle d'un gène. L'approche classique est d'utiliser la similarité de séquence avec des séquences déjà caractérisées. Nous allons également voir qu'il existe d'autres méthodes basées sur l'analyse phylogénomique des séquences.

2.4.1 Annotation par homologie et orthologie

Un des principaux enjeu de la phylogénomique est l'inférence de la fonction des gènes au sein d'un génome afin de les annoter [123]. Le principe de base de l'annotation fonctionnelle utilisé actuellement par la plupart des bases de données de génomes repose sur l'homologie [400], [17], [421]. Deux séquences sont considérées comme homologues si elles présentent un ancêtre commun. La technique principale pour identifier les homologues d'une séquence consiste concrètement à rechercher les séquences présentant une similarité de séquence. Un des algorithmes le plus utilisé est celui du programme BLAST, détaillé dans la section 2.2.2.1 [4] . Ainsi, toutes les séquences ayant été prouvées comme étant impliquées dans un processus cellulaire sont utilisées pour identifier les séquences qui pourraient remplir la même fonction au sein d'autres génomes. Néanmoins, cette méthodologie se heurte à plusieurs difficultés. Tout d'abord, près d'un quart des gènes issus des génomes séquencés n'ont pas d'homologue caractérisé biologiquement et donc pas de fonction assignée selon une étude de 2013 [10]. Ensuite, l'homologie entre deux

séquences ne signifie pas forcément qu'elles présentent la même fonction cellulaire. En effet, au sein d'une famille d'homologues, on peut discerner différentes sous-familles qui peuvent présenter des rôles cellulaires très différents [247]. Les séquences appartenant à la même sous-famille dérivant d'une séquence ancestrale et acquises par transmission verticale (par spéciation) sont définies comme orthologues. A l'inverse, un groupe issu d'une duplication de gène est considéré comme paralogue de la famille du gène ayant été dupliquée. Il est admis que deux orthologues ont plus de chances de remplir le même rôle au sein de la cellule que deux homologues, bien que l'orthologie ne soit pas une condition suffisante pour affirmer catégoriquement que deux séquences ont le même rôle [247].

2.4.2 Données phylogénomiques : comparaison entre familles de gènes

Bien que les méthodes d'annotation basées sur l'homologie soient efficaces, une proportion conséquente des séquences issues du séquençage de génomes n'ont pas de fonction assignée [10]. D'autres méthodes ont donc été développées afin de pallier ce manque de connaissance. Une de ces techniques consiste à comparer l'histoire évolutive et l'organisation génomique des familles de gènes pour déterminer le rôle des gènes dont la fonction est inconnue [369], [303]. En effet, une corrélation entre l'histoire évolutive de deux gènes ou de leur organisation génomique peut indiquer une relation fonctionnelle [123]. Ce lien évolutif peut se caractériser concrètement de plusieurs manières dans la cellule. Il peut s'agir d'une interaction protéine-protéine directe ou indirecte au sein d'un complexe multiprotéique. Une des deux protéines peut aussi agir en amont de l'autre en régulant son expression de façon directe ou indirecte. Dans le cas d'un voisinage conservé entre deux gènes, on peut faire l'hypothèse qu'une synchronisation de la transcription des deux gènes soit importante au sein de la cellule. Les liens fonctionnels peuvent donc être de différentes natures et la phylogénomique n'est pas capable de différencier ces cas. Ainsi, seules les sciences expérimentales peuvent permettre de déterminer plus précisément la nature de ces liens. Malgré tout, ces approches peuvent permettre d'aiguiller l'approche expérimentale [288], [384]. Le plus connu des logiciels fournissant ce type d'informations et

étant le plus utilisé par les biologistes est STRING [449].

2.4.2.1 Analyse du voisinage génomique

Concept d'opéron et de synténie Chez les procaryotes, certains gènes sont organisés en opéron [216]. Il s'agit d'une unité génique fonctionnelle dans laquelle on retrouve plusieurs gènes voisins impliqués dans le même processus cellulaire. L'ensemble des gènes présents dans un opéron sont sous le contrôle d'un même promoteur et sont co-transcrits puis co-traduits. Il s'agit du moyen le plus simple et le moins coûteux pour la cellule de produire en même temps plusieurs constituants d'un même système biologique. Ainsi, une pression de sélection peut s'exercer sur les génomes afin que plusieurs gènes restent vicinaux. On parle alors de voisinage conservé entre les deux gènes. De façon intéressante, il a été montré par Lathe et collègues [263] que les clusters de gènes pouvaient avoir une dynamique complexe entre eux et que le concept d'opéron était plus large que ce que les biologistes entendent. Ainsi, certains clusters fusionnent avec d'autres et se réarrangent pour former d'autres clusters. Les auteurs proposent alors la notion de « uber-operon » (de *über* : au-dessus) pour décrire l'ensemble des clusters qui se réorganisent les uns avec les autres. De façon encore plus large, deux gènes étant voisins, même sans être en opéron, peuvent être reliés fonctionnellement [231]. En effet, le génome étant compacté la plupart du temps et la décompaction nécessaire à la transcription des gènes s'effectuant sur des grandes portions de génome, deux gènes voisins seront probablement transcrits au même moment, même sans être sous le contrôle du même promoteur.
Il est à noter que le terme synténie est également utilisé pour désigner le voisinage entre gènes [385], [84] bien que celui-ci signifiait initialement la présence de deux gènes sur un même chromosome [365].

Principe de l'analyse des synténies De façon générale, le voisinage génomique conservé de deux gènes peut indiquer un lien fonctionnel (figure 2.9). L'analyse du voisinage génomique consiste alors concrètement en l'analyse de la position des gènes les uns par rapport aux autres le long du chromosome, ainsi que de la conservation de cette organisation au sein du vivant. La

distance entre deux gènes peut être mesurée en paires de bases ou en nombre de gènes séparant ces derniers.

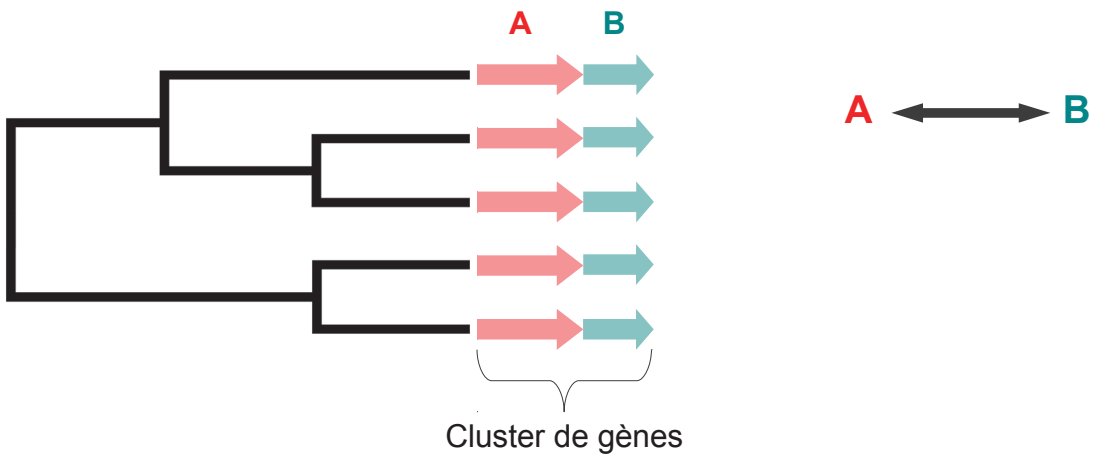


FIGURE 2.9 – Principe de prédiction de lien fonctionnel par voisinage génomique. L'arbre représente les relations de parenté entre espèces. Les gènes A et B sont systématiquement voisins et sont donc potentiellement reliés fonctionnellement.

Méthode qualitative A partir du concept de synténie conservée, un grand nombre d'études ont mis en évidence des relations fonctionnelles entre des gènes. Certaines études utilisent des approches qualitatives afin de caractériser le voisinage entre plusieurs gènes [288], [384], c'est-à-dire en visualisant les contextes génomiques et en identifiant visuellement les voisinages génomiques et leur conservation . De nombreux logiciels/plate-formes permettent de visualiser le contexte génomique d'un gène comme MGcV [361], EasyFig [447] ou encore STRING [449].

Limites des méthodes qualitative Bien que de nombreux logiciels de visualisation de contextes génomiques soient disponibles, ceux-ci présentent des inconvénients techniques majeurs. La plupart de ces outils reposent en effet sur des bases de données en ligne composées d'un faible nombre de génomes ne permettant ainsi qu'une appréciation partielle de la conservation des voisinages génomiques. De plus, le rendu des figures générées ne présente en général que peu de flexibilité. Dans ce contexte, nous avons développé le logiciel GeneSpy [157]. Nous développons ce point dans le chapitre 5.

Méthode quantitative Néanmoins, pour mieux caractériser les synténies, il peut être intéressant de développer des approches quantitatives. Ainsi, un certain nombre de publications ont proposé des méthodes pour quantifier le voisinage génomique d'un ensemble de gènes et sa conservation chez les différents organismes [455], [404], [231], [211], [431]. Elles diffèrent en de nombreux points comme la définition du voisinage génomique, les métriques pour quantifier la conservation du voisinage génomique ou encore la méthode de regroupement des gènes synténiques. Nous allons ici présenter quelques exemples de méthodes.

Les méthodes les plus simples consistent en l'identification des gènes voisins conservés chez quelques génomes. Par exemple, Tamames et collègues ont comparé 2 génomes de bactéries et ont inféré une synténie si deux gènes se situent l'un à coté de l'autre dans les deux génomes [455]. Autre exemple, Rogozin et collègues considèrent qu'il y a synténie entre deux gènes si au moins 3 génomes sur les 23 étudiés sont espacés de 0, 1 ou 2 gènes [404].

D'autres méthodes ont été développées afin d'estimer statistiquement la conservation d'une synténie. Ainsi, Huynen et collègues ont considéré qu'une synténie existe entre deux gènes si la probabilité de les observer l'un à coté de l'autre est bien supérieure à la probabilité de les retrouver voisins par le hasard [211]. Autre exemple, Junier et collègues ont appliqué une méthode qui semble plus performante et sur un grand nombre de génomes [231]. A partir de 1108 génomes, ils ont calculé la distribution de la distance génomique entre deux COG puis l'ont comparée avec une distribution uniforme. Un seuil de p-value a ensuite été fixé à 5% et un réseau a été construit à partir de ces liens.

Limites des méthodes quantitative Ainsi plusieurs méthodes ont été proposées mais aucune d'entre elles n'a été considérée comme méthode standard. De plus, la définition même de synténie n'est pas clairement établie. En effet, la mesure de la vicinité entre deux gènes peut s'exprimer en nombre de gènes qui les séparent sur un génome, ou par une distance en nombre de nucléotides, et aucun seuil standard n'a été proposé. Enfin, il est intéressant de noter que la plupart de ces études n'utilisent pas l'information phylogénétique des espèces pour étudier les synténies. En effet, il est attendu que le nombre de synténies conservées soit fonction du temps de divergence entre les organismes. Ainsi, deux organismes proches présenteront en

moyenne plus d'adjacence de gènes conservées que des organismes plus distants, sans que cela ait de rapport avec des liens fonctionnels entre gènes. En faisant correspondre les voisinages génomiques des gènes avec une phylogénie d'espèces, nous avons dans cette étude reconstruit les synténies ancestrales et utilisé cette information pour inférer les liens fonctionnels.

2.4.2.2 Méthode de la pierre de rosette et fusion de domaines

Principe Une autre méthode pour identifier des liens fonctionnels est la méthode de « Pierre de Rosette » [80]. Cette méthode consiste en l'identification d'un lien fonctionnel entre deux protéines distinctes chez un organisme si, chez un autre organisme, les deux protéines sont fusionnées en un même gène (figure 2.10). l'association des deux gènes permet de relier deux gènes fonctionnellement et ainsi de prédire la fonction d'un gène si l'autre est connu, telle la « Pierre de Rosette » qui contenait le même texte en hiéroglyphes, démotique et en grec. De façon générale, les portions fusionnées dans les protéines issues des gènes « Pierre de Rosette » sont des domaines fonctionnels [367].

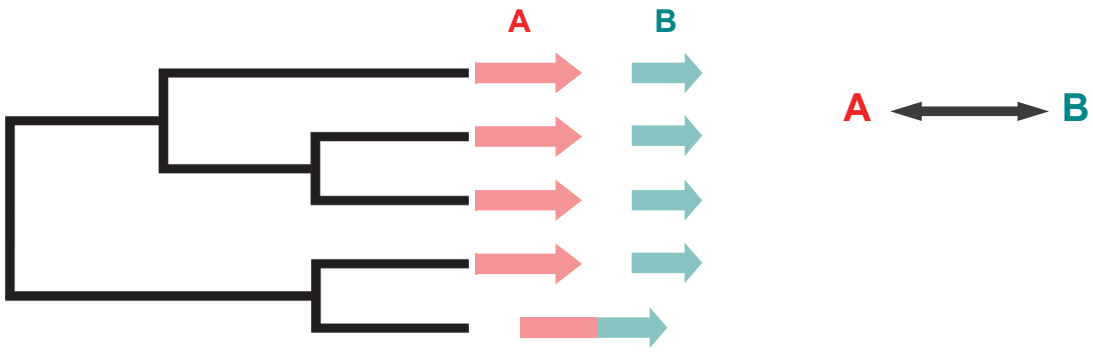


FIGURE 2.10 – Principe de prédiction de lien fonctionnel par la méthode de la pierre de Rosette. L'arbre représente les relations de parenté entre espèces. Les gènes A et B sont fusionnés dans une espèce et sont donc potentiellement reliés fonctionnellement.

Bases de données et méthodes Plusieurs bases de données de domaines fonctionnels de protéines sont disponibles comme Pfam [146] ou Interpro [210]. Les domaines Pfam consistent en des profils HMM issus d'alignements de séquences vérifiées manuellement. A partir d'alignements multiples de séquences entières, les bornes des domaines sont définies en identifiant des régions conservées. Interpro correspond au regroupement de l'information provenant de plusieurs bases de données de domaines dont Pfam [210]. Ces bases de données peuvent être utilisées par HMMER [119] afin d'identifier les domaines contenus dans les séquences protéiques d'un jeu de données et de détecter des potentielles familles de gènes reliées par un domaine commun.
Un exemple de l'application de la méthode de la « Pierre de Rosette » est celle utilisée dans l'étude de Marcotte et collègues [304] dans laquelle de nombreuses interactions protéine-protéine ont été inférées chez *E coli*.

2.4.2.3 Méthode des profils phylogénétiques

Principe Un profil phylogénétique représente la distribution taxonomique d'un gène, c'est-à-dire le profil de présence/absence du gène chez un ensemble d'organismes [369] (figure 2.11). Ce profil correspond au résultat de l'ensemble des événements évolutifs ayant affecté la famille de gènes durant la diversification du taxon incluant les organismes analysés. Un hypothèse couramment admise est que deux gènes ayant co-évolué sont reliés fonctionnellement. Ainsi, si deux gènes présentent des profils phylogénétiques similaires (co-présence ou co-absences), il est probable que ces deux gènes soient reliés fonctionnellement [369]. Concrètement, un profil phylogénétique est un vecteur binaire qui correspond à l'absence (0) ou la présence (1) d'un gène au sein de plusieurs génomes. De façon importante, les génomes nécessitent d'être complets afin d'être certain qu'une absence n'est pas due à une incomplétude d'assemblage ou de séquençage.

Profils binaires et continus De nombreuses méthodes ont été proposées pour construire les profils phylogénétiques [241], [225]. Le plus simple type de profil phylogénétique correspond

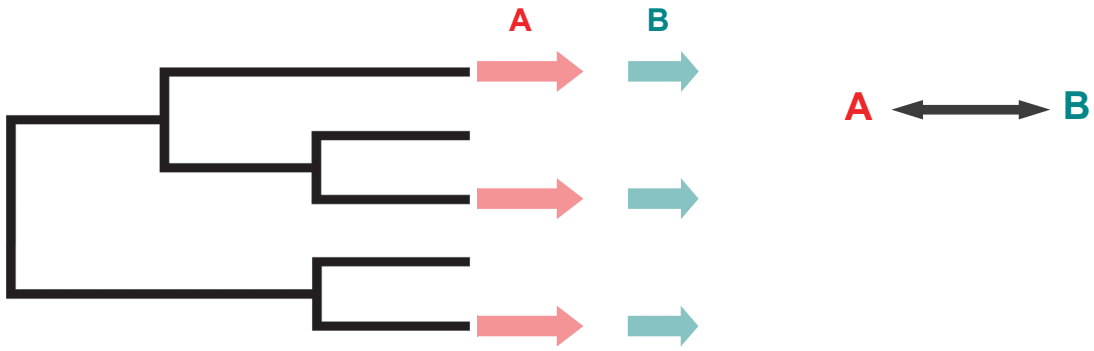


FIGURE 2.11 – Principe de prédiction de lien fonctionnel par profils phylogénétiques. L'arbre représente les relations de parenté entre espèces. Les gènes A et B sont présents ou absents chez les mêmes espèces et sont donc potentiellement reliés fonctionnellement.

au profil binaire basé sur l'homologie. Il s'agit de reconstruire une famille d'homologues et de projeter la présence/absence à chaque espèce étudiée. Cette technique est la plus simple mais la moins précise car elle ne prend en compte les événements évolutifs ayant pu affecter l'histoire évolutive des gènes étudiés, notamment les duplications et les transferts horizontaux.

Pour pallier cette difficulté, les profils phylogénétiques continus ont été utilisés [81], [229]. La présence d'un gène au sein d'un génome est exprimée par une valeur oscillant entre 0 et 1. Les gènes provenant d'un génome de référence sont alignés avec toutes les séquences de chaque génome. La plus petite E-value pour chaque gène du génome de référence et pour un génome testé est sélectionnée puis utilisée pour calculer le score. Ainsi, la divergence des séquences est quantitative et il est possible de faire la différence entre un orthologue et un paralogue. Cette technique est donc plus précise qu'un profil binaire.

Enfin, la dernière technique correspond à la construction d'un profil phylogénétique binaire basé sur l'orthologie [46], [490], [313]. Il s'agit donc de distinguer de façon précise les gènes issus d'une transmission verticale au sein d'une famille d'homologues. Ainsi, le cas des paralogues est écarté.

Métriques de comparaison des profils Après construction des profils phylogénétiques, il est nécessaire de les comparer pour évaluer le degré de coévolution entre les familles de gènes. Pour ceci, plusieurs métriques sont utilisées. La plus usitée est l'information mutuelle [46], [490], [313], [211]. Il s'agit d'une fonction qui estime la quantité d'information que deux séries de données apporte l'une à l'autre. Il existe d'autres alternatives possibles comme le coefficient de Pearson (ou coefficient ϕ pour des caractères binaires) [168], la distance de Hamming [369] ou encore le test exact de Fisher [22].

Méthodes utilisant la phylogénie d'espèces Néanmoins, le problème de la plupart des méthodes utilisées est l'absence de prise en compte du biais phylogénétique (figure 2.12A). En effet, la présence d'un gène chez des espèces proches phylogénétiquement n'a pas le même poids que pour des espèces éloignées. Pour éviter ce biais, plusieurs méthodes qui consistent à prendre en compte la phylogénie d'espèces dans l'estimation de la co-occurrence entre deux familles de gènes ont été développées (figure 2.12) [362], [526], [489].
La méthode implémentée dans STRING est de collapser les branches de l'arbre d'espèces dont les espèces filles possèdent des profils phylogénétiques identiques pour les deux familles de gènes considérées [489]. Ainsi, le profil à la branche collapsée correspond à l'état ancestral (figure 2.12B).

La deuxième méthode pour prendre en compte le biais phylogénétique est de reconstruire les états ancestraux. Le premier type d'approche est celle de la parcimonie (figure 2.12C). Elle nécessite de fixer les coûts associés aux gains et de pertes. Il est possible d'utiliser des coûts égaux suivant l'algorithme de Fitch [278]. Une autre possibilité est de n'autoriser qu'un seul gain suivant l'algorithme de Dollo [21]. Ainsi, les événements de gains et pertes inférés peuvent être considérés comme indépendants, ce qui justifie l'utilisation des métriques telles que le coefficient ϕ ou l'information mutuelle [241]. Néanmoins, la solution la plus parcimonieuse ne prend pas en compte les incertitudes. Dans ce but, un modèle qui prend en compte les 100 solutions suboptimales a été proposé par Zhou et collègues [526]. Une autre solution permettant de prendre en compte les incertitudes dans la reconstruction des états ancestraux est d'utiliser

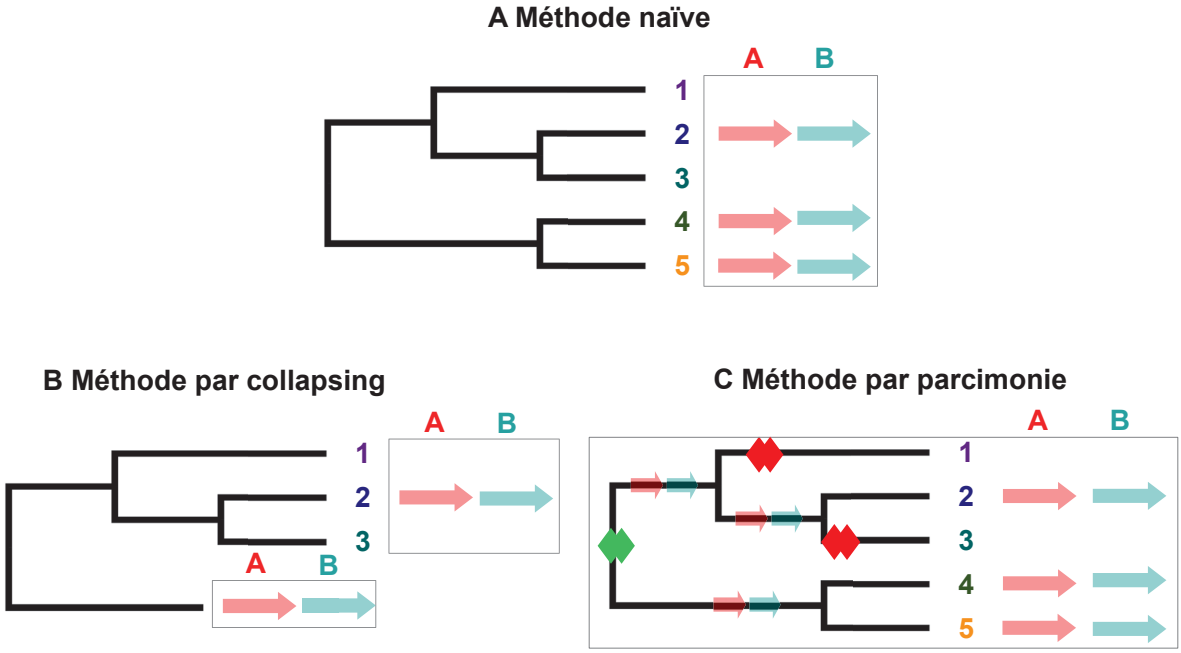


FIGURE 2.12 – Trois méthodes de comparaison de profils phylogénétiques. (A) Méthode naive. les profils sont construits sans prendre en compte la phylogénie. (B) Méthode de collapsing utilisée par STRING [489]. (C) Méthode par parcimonie. Les états de présence/absence au branches de l'arbre d'espèces sont inférés par parcimonie. Les états et/ou les événements sont utilisés pour établir la corrélation des deux familles.

un modèle pour calculer une vraisemblance. Par exemple, Barker et collègues ont développé des méthodes basées sur des modèles de Markov [22], [21]. Les méthodes par vraisemblance, bien que plus précises, demandent des ressources de calcul bien plus importantes.

Limites des profils phylogénétiques Bien que la méthode des profils phylogénétiques se soit imposée comme méthode de prédiction de liens fonctionnels entre familles de gènes, elle présente une limitation majeure. En effet, les profils phylogénétiques ne reflètent qu'une image partielle de l'histoire des gènes car de multiples duplications et transferts horizontaux peuvent avoir eu lieu sans impacter le profil phylogénétique. Les méthodes de reconstruction des états ancestraux sont donc basées sur l'hypothèse que l'ensemble des gènes de la famille considérée

sont tous orthologues et issus de spéciations. Cette hypothèse ne se vérifie que rarement en pratique et particulièrement pour les procaryotes chez qui de nombreux gènes sont transférés d'une espèce à l'autre de manière horizontale.

2.4.2.4 Méthodes de réconciliation

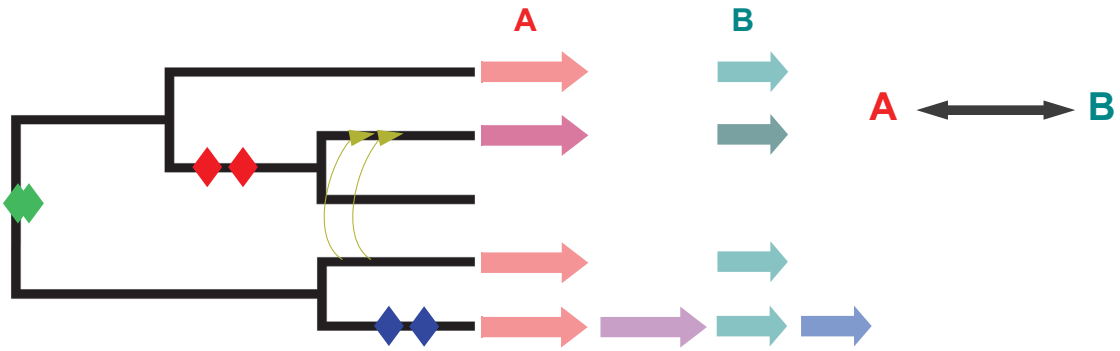


FIGURE 2.13 – Principe de prédiction de lien fonctionnel par co-évolution. L'arbre représente les relations de parenté entre espèces. Les familles de gènes A et B ont été affectées par les mêmes événements évolutifs (co-pertes, co-transferts, co-duplications) et sont donc potentiellement reliés fonctionnellement.

Principe Une méthode possible pour identifier des liens fonctionnels entre des familles de gènes de façon alternative aux profils phylogénétiques est l'analyse de la coévolution. Il s'agit de comparer les histoires des familles de gènes (figure 2.13). Il est possible de comparer directement deux phylogénies de gènes et de calculer une distance ou un score de similitude [172]. Néanmoins, cette méthode n'est pas basée sur un modèle et ne fourni aucune information sur les événements évolutifs. L'alternative est la réconciliation entre les phylogénies de gènes avec un arbre d'espèces [409] (figure 2.8). Plus précisément, la réconciliation correspond à l'association de chaque branche de l'arbre de gènes à une branche de l'arbre d'espèces [409], [100]. Chaque association est accompagnée d'un événement évolutif ayant affecté l'histoire du gène. Ce procédé permet d'identifier précisément les événements évolutifs ayant impacté les familles

de gènes et de les replacer sur la phylogénie des espèces. Il est ensuite possible de compter le nombre d'événements en communs entre deux scénarios de deux familles de gènes.

Méthodes La première approche utilisée pour réconcilier un arbre de gène et d'espèces est la parcimonie. Cela consiste en la sélection de la solution qui minimise le coût du scénario évolutif, sachant la phylogénie des espèces et le coût de chaque événement. La première publication à utiliser la réconciliation décrit l'histoire évolutive des globines chez les vertébrés [174] avec un modèle ne prenant en compte que les duplications et les pertes (« DL model »). Ce modèle a depuis été complexifié en prenant également en compte les transferts horizontaux (« DLT model »). L'implémentation des transferts horizontaux parait absolument nécessaire, particulièrement pour l'analyse des gènes de procaryotes. En effet, le nombre d'événements de transfert horizontaux chez les procaryotes a été estimé comme étant très élevé [253], [27]. Certains programmes ont depuis été implémentés avec ce modèle comme ecceTERA [420]. L'estimation des coûts est néanmoins une tache très complexe. En effet, les coûts optimaux d'événements peuvent varier fortement en fonction des familles de gènes concernées. Une solution pour choisir les coûts a été proposée par Scornavacca et collègues, basée sur la fréquence d'événements inférés à partir de coûts fixes. Des méthodes probabilistes comme celle implémentée dans ALE [450] ont également été développées, basées non pas sur la parcimonie mais sur une estimation d'une vraisemblance à partir d'un modèle.

Méthodes de comparaison de réconciliation La méthode la plus simple consiste en le comptage du nombre de co-événements entre deux familles par branche. Cette méthode n'a, à notre connaissance, jamais été utilisée de façon quantitative. Chan et collègues ont proposé une méthode pour comparer deux distributions de gènes par une approche probabiliste qui n'a néanmoins pas encore été implémentée dans un programme [64]. Il est aussi possible à partir des événements d'établir les états de présence/absence aux branches et feuilles de l'arbre d'espèces et de construire un profil phylogénétique basé sur la réconciliation (Cf 2.4.2.3).

Limites de la réconciliation Malgré l'élégance de la méthode, la réconciliation souffre de quelques limites. Le principal écueil provient de la qualité des phylogénies utilisées. En effet, la topologie de l'arbre de gènes et d'espèces peuvent présenter des incertitudes à certaines branches. Ces inconsistances qui proviennent d'un manque de signal dans les alignements ou de signaux contradictoires, mènent à des erreurs lors de la réconciliation. Pour résoudre en partie ce problème, certains programmes comme ecceTERA offrent la possibilité de faire une amalgamation d'arbres de gènes à partir d'une distribution d'arbres [420]. La distribution d'arbres, souvent obtenue à partir des réplicats de bootstrap, permet de tenir compte des incertitudes et d'augmenter la qualité de la réconciliation. Le programme ProfilNJ [353] permet aussi de rectifier l'arbre de gènes selon l'arbre d'espèces, notamment au niveau des branches présentant un support statistique faible. D'autres approches comme celle utilisée par DeCoSTAR [106] permettent de prendre en compte les adjacences entre les gènes comme les synténies pour compenser le manque de signal phylogénétique.

2.5 Les *Firmicutes*

2.5.1 Généralités

L'arbre du vivant peut être décomposé en 3 domaines : Les Eucaryotes, les Archées et les Bactéries [509]. Le niveau taxonomique inférieur chez les bactéries correspond à l'embranchement ou au *phylum*. Au sein des bactéries, on retrouve un grand nombre de *phyla*. Dans les plus représentés et plus étudiés, on peut citer les *Proteobacteria*, les *Actinobacteria*, les *Cyanobacteria*, les *Bacteroidetes* ou les *Firmicutes*.

Les *Firmicutes* ont été pour la première fois décrit par Gibbons et Murray en 1978 [165] comme englobant l'ensemble des bactéries à Gram positif. Ce groupe regroupait donc les bactéries à Gram positif à faible taux de GC (*Firmicutes* actuels) et à fort taux de GC (*Actinobacteria*). En 2001, Garrity et Holt proposent de regrouper dans les *Firmicutes* les bactéries à Gram positif à faible taux de GC avec les *Mollicutes*, taxon de bactéries sans paroi cellulaire mais également d'exclure les *Actinobacteria* [160]. Cette classification est de nouveau remise en cause par Lud-

wig et Schleifer en 2005 [289]. Dans cette étude basée sur une phylogénie de gènes conservés comme *rpoB*, les auteurs montrent que les *Mollicutes* et les *Firmicutes* typiques présentent une distance phylogénétique conséquente, bien que les *Mollicutes* dérivent des *Firmicutes* [510]. A cela s'ajoute que les *Mollicutes* possèdent également des caractéristiques morphologiques distinctes des *Firmicutes* types comme l'absence de paroi. Les *Mollicutes* sont ainsi classés dans le *phylum* des *Tenericutes*.

Au sein des *Firmicutes* selon la classification actuelle, on retrouve donc une grande majorité de bactéries à Gram positif bien que certaines présentent une double membrane comme les *Negativicutes* [302] ou les *Halanaerobiales* [525]. Il a d'ailleurs été suggéré que la présence de la double membrane serait un caractère ancestral chez les *Firmicutes* [11]. De façon plus générale, les bactéries incluses dans le *phylum* des *Firmicutes* présentent une grande diversité de phénotypes. Ainsi, on retrouve des cellules sphériques, en bâtonnet, filamenteuses, avec ou sans flagelle, sporulantes ou non....

Au sein de cette diversité, de nombreuses bactéries sont d'un grand intérêt médical et industriel. En effet, un certain nombre de bactéries pathogènes appartient aux *Firmicutes* comme *Peptoclostridium/Clostridoides difficile*, *Staphylococcus aureus*, *Listeria monocytogenes* ou encore *Streptococcus pneumoniae*. La compréhension de leur physiologie peut donc permettre de mieux appréhender leurs pathologies associées et de développer des stratégies thérapeutiques. De façon intéressante, les *Firmicutes* représentent une part importante des bactéries présentes au sein du microbiote intestinal [471]. Elles remplissent donc aussi un rôle essentiel dans la physiologie animale et notamment au niveau du métabolisme [349]. Un déséquilibre de la composition du microbiote peut générer des pathologies telles que l'obésité et le diabète [464]. Enfin, certains *Firmicutes* sont utilisés dans l'industrie et particulièrement l'agroalimentaire. Par exemple, certaines espèces du genre *Lactobacillus* font partie du procédé de fabrication d'un grand nombre de produits laitiers. Le grand intérêt pour les *Firmicutes* en font l'un des *phyla* bactérien les plus séquencés après les *Proteobacteria*, comme en témoigne les 42 691 génomes disponibles sur le serveur du NCBI.

Classe	Ordre	Famille	Exemple d'espèce
Clostridia	*Clostridiales*	*Clostridiaceae*	*Clostridium botulinum*
		Eubacteriaceae	*Eubacterium limosum*
		Heliobacteriaceae	*Heliobacterium modesticaldum*
		Lachnospiraceae	*Roseburia hominis*
		Oscillospiraceae	*Oscillibacter valericigenes*
		Peptococcaceae	*Desulfotomaculum gibsonia*
		Peptostreptococcaceae	*Peptoclostridium difficile*
		Ruminococcaceae	*Ruminococcus albus*
		Symbiobacteriaceae	*Symbiobacterium thermophilum*
		Syntrophomonadaceae	*Syntrophothermus lipocalidus*
		Gracilibacteraceae	*Gracilibacter thermotolerans*
	Halanaerobiales	*Halanaerobiaceae*	*Halanaerobium praevalens*
		Halobacteroidaceae	*Halobacteroides halobius*
	Thermoanaerobacterales	*Thermoanaerobacteraceae*	*Thermoanaerobacter italicus*
		Thermodesulfobiaceae	*Thermodesulfobium narugense*
	Natranaerobiales	*Natranaerobiaceae*	*Natranaerobius thermophilus*
Bacilli	*Bacillales*	*Bacillaceae*	*Bacillus subtilis*
		Listeriaceae	*Listeria monocytogenes*
		Paenibacillaceae	*Paenibacillus durus*
		Planococcaceae	*Planococcus kocurii*
		Sporolactobacillaceae	*Sporolactobacillus inulinus*
		Staphylococcaceae	*Staphylococcus aureus*
	Lactobacillales	*Aerococcaceae*	*Aerococcus urinae*
		Carnobacteriaceae	*Carnobacterium maltaromaticum*
		Enterococcaceae	*Enterococcus faecalis*
		Lactobacillaceae	*Lactobacillus casei*
		Leuconostocaceae	*Leuconostoc kimchii*
		Streptococcaceae	*Streptococcus pneumoniae*
Erysipelotrichia	*Erysipelotrichales*	*Erysipelotrichaceae*	*Erysipelothrix rhusiopathiae*
Negativicutes	*Acidaminococcales*	*Acidaminococcaceae*	*Acidaminococcus intestini*
	Veillonellales	*Veillonellaceae*	*Veillonella parvula*
	Selenomonadales	*Selenomonadaceae*	*Selenomonas sputigena*
Tissierellia	*Tissierellales*	*Peptoniphilaceae*	*Finegoldia magna*

TABLE 2.1 – Taxonomie des *Firmicutes* selon le NCBI [137]. Les genres et souches ne figurent pas sur le tableau. La famille en rouge n'est pas représentée dans cette étude. La couleur des Ordres correspond à celle représentée sur la figure 2.14.

2.5.2 Taxonomie et Systématique des *Firmicutes*

Le *phylum* des *Firmicutes* est divisé en différents *taxa*. Nous utiliserons ici la taxonomie fournie par le NCBI [137]. On dénombre 5 Classes au sein des *Firmicutes* : Les *Clostridia*, les *Bacilli*, Les *Erysipelotrichia*, les *Negativicutes* et les *Tissierellia*. La composition des classes en ordres et familles est décrite table 2.1. Une phylogénie des *Firmicutes* basée sur une concaténation de 47 protéines ribosomiques a été proposée par Antunes et collègues [11]. Nous allons nous baser sur cette phylogénie pour décrire les relations de parenté chez les *Firmicutes* (figure 2.14).
Le groupe le plus basal est constitué des *Natranaerobiales* et *Halanaerobiales*, deux ordres des *Clostridia*. On retrouve ensuite deux groupes correspondant aux autres *Clostridia* et aux *Bacilli*. Au sein du groupe des *Clostridia*, deux groupes émergent. Deux classes se branchent dans les *Clostridia* : les *Tissierellia* émergent à partir du premier groupe et les *Negativicutes* à partir du deuxième. Dans la branche correspondant aux *Bacilli*, les *Bacillales* sont en position basale. Les *Erysipelotrichia* émergent au sein des *Bacillales* (non présenté sur la figure 2.14) puis les *Lactobacillales* branchent après l'émergence des *Listeriaceae*.

Il est à noter que certains points au sein de la phylogénie et de la taxonomie des *Firmicutes* restent encore mal définis. Tout d'abord, certaines familles décrites notamment chez les *Clostridia* ne sont pas monophylétiques dans l'arbre d'espèces comme par exemple les *Thermoanaerobacterales* (figure 2.14). D'autre part, certains organismes classés comme faisant partie des *Firmicutes*, comme les *Thermodesulfobiaceae*, semblent finalement en être exclus [519],[523]. Enfin, les relations entre les groupes peuvent varier en fonction des études [11], [523]. Nous débattrons de ces ambiguïtés de topologie dans le chapitre 3.

FIGURE 2.14 – Phylogénie des *Firmicutes* selon Antunes *et al.* [11] inférées à partir de 47 protéines ribosomiques (PhyML, LG+G4, 218 taxa, 5551 positions). Les couleurs correspondent aux ordres : Jaune : groupe externe, vert : *Halanaerobiales*, noir : *Natranaerobiales*, rouge : *Bacillales*, rose : *Lactobacillales*, marron : *Thermoanaerobacterales*, bleu : *Clostridiales* et *Tissierellales*, violet :*Negativicutes* (*Acidaminococcales*, *Veillonellales* et *Selenomonadales*). La barre d'échelle correspond au nombre moyen de substitution par site. Les *Erysipelotrichia* ne sont pas représentés dans l'arbre.

Chapitre 3

Application de l'approche de co-évolutions aux familles de gènes du cycle cellulaire

Sommaire

3.1 Base de données des protéomes de *Firmicutes* 135

3.1.1 Construction des bases de données de séquences et d'annotation . . 135

3.1.2 Composition des bases de données 135

3.2 Inférence de la phylogénie des espèces de *Firmicutes* 138

3.2.1 Construction de l'arbre d'espèces daté des *Firmicutes* 138

3.2.2 Analyse de la topologie de l'arbre d'espèces des *Firmicutes* 143

3.3 Construction des familles de gènes impliqués dans le cycle cellulaire 148

3.3.1 Méthodes de construction des familles 148

3.3.2 Composition des familles 161

3.4 Événements évolutifs majeurs des familles de gènes du cycle cellulaire chez les *Firmicutes* 171

3.4.1 Reconstruction de l'histoire évolutive des familles de gènes 172

3.4.2 Événements évolutifs majeurs du cycle cellulaire chez les *Firmicutes* 179

3.5 Inférence de relations fonctionnelles par approche corrélative . . . 203

3.5.1 Méthodes de corrélation 203

3.5.2 Liens fonctionnels entre familles de gènes du cycle cellulaire 220

3.6 Histoires évolutives détaillées et implications fonctionnelles 235

3.6.1 Description et interprétation fonctionnelle de l'histoire des clans . . 235

3.6.2 Description et interprétation fonctionnelle de liens évolutifs ponctuels 263

3.6.3 Observations générales et conclusion 270

3.7 Identification de nouvelles familles potentiellement impliquées dans le cycle cellulaire 276

3.7.1 Méthode d'identification des familles de gènes corrélant avec les familles du cycle cellulaire 277

3.7.2 Familles corrélées évolutivement aux familles du cycle cellulaire . . . 279

3.8 Conclusion 286

3.1 Base de données des protéomes de *Firmicutes*

3.1.1 Construction des bases de données de séquences et d'annotation

Une base de données locale de protéomes complets de procaryotes a été construite à partir des protéomes disponibles sur le FTP du NCBI (ftp://ftp.ncbi.nlm.nih.gov/). Les fichiers FAA (« Fasta Amino Acid ») ont été récupérés en priorité à partir de RefSeq [187] puis, si les fichiers n'étaient pas disponibles, à partir de GenBank [31]. Les fichiers FAA ont ensuite été concaténés puis formatés avec le programme FORMATDB de la suite BLASTALL 2.2.25 [5]. Au total, 4 466 protéomes ont été concaténés dans la base de données. Une base de données d'annotation de ces mêmes génomes a parallèlement été construite : GFF (« Generic Feature Format »), GBFF (« GenBank Flat-File ») et GFM (« GFF Minimal content »). Ces fichiers contiennent différentes informations relatives aux gènes présents dans un génome comme la position, le numéro d'accession de la protéine associée ou encore la fonction biochimique. Les fichiers GBFF sont très complets en terme d'informations mais difficiles à exploiter tandis que les fichiers GFF et GFM contiennent moins d'informations mais sont sous format tabulé ce qui facilite leur utilisation par des scripts.
Des bases de données de *taxa* ont été construites à partir de la base de données des 4 466 protéomes de procaryotes. L'attribution de chaque génome aux différents *taxa* est basée sur la taxonomie du NCBI (https://www.ncbi.nlm.nih.gov/taxonomy, [137]). Ainsi, une base de données de séquences protéiques de 937 génomes de *Firmicutes* a été construite (annexe .1).

3.1.2 Composition des bases de données

La base de données de protéomes complets de procaryotes téléchargée à partir du FTP du NCBI est composée en majorité de *Proteobacteria*, de *Firmicutes* et d'*Actinobacteria* (figure 3.1).

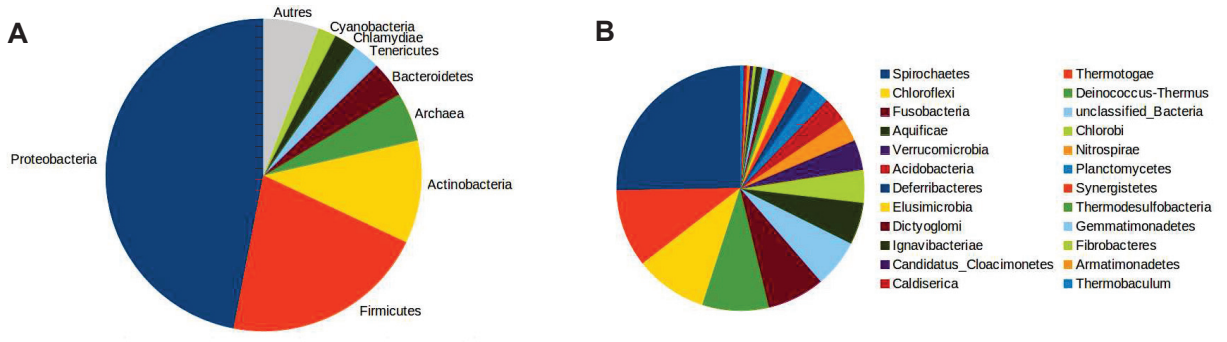


FIGURE 3.1 – Composition en *phyla*/domaine de la base de données de procaryotes. (A) Phyla/domaines majoritaires. (B) Phyla du groupe Autres.

Parmi les 4 466 protéomes de procaryotes, 927 protéomes sont annotés comme provenant de *Firmicutes* et ont été concaténés pour former une base de données de *Firmicutes*. Il est important de noter que les *Mollicutes* n'ont pas été inclus dans la base de données. Initialement, les *Mollicutes* ont été décrits comme étant groupés avec les *Firmicutes* mais ce regroupement a été plus récemment remis en cause [88], [289]. De plus, les caractéristiques morphologiques des *Mollicutes* sont très différentes des *Firmicutes* comme l'absence de paroi.

La majorité des protéomes de *Firmicutes* provient des *Bacilli* (762) et des *Clostridia* (160) (figure 3.2). Le reste des classes est nettement sous-représenté. Aussi, une sur-représentation de certaines familles au sein des *Bacilli* et *Clostridia* est observée. Chez les *Bacillales*, les trois familles sur-représentés sont les *Bacillaceae* (211), les *Staphylococcaceae* (94) et les *Listeriaceae* (69). Chez les *Lactobacillales*, les deux familles majoritaires sont les *Streptococcaceae* (202) et les *Lactobacillaceae* (90). Enfin, chez les *Clostridiales*, les protéomes de *Clostridiaceae* (58) sont les plus représentés.

Cette sur-représentation peut induire des biais, notamment dans les études corrélatives. Elle est due au fait que quelques espèces d'intérêt médical (*e.g. Streptococcaceae*) ou industriel (*e.g. Lactobacillaceae*) sont très étudiés et de nombreuses souches ont été séquencées de nombreuses fois. De plus, considérer un grand nombre de protéomes demande de plus grands moyens de calculs lors des analyses. Pour limiter les biais liés à la redondance taxonomique, nous avons donc décidé d'effectuer un échantillonnage taxonomique représentatif de la diversité des *Firmicutes*. Pour ceci, nous avons sélectionné une souche par espèce ce qui a conduit à

l'obtention de 306 souches. La distribution des taxons représentés dans la base de données est présentée figure 3.2. Il apparaît que la sur-représentation de certaines familles est largement diminuée suggérant que les biais d'échantillonnage ont été minimisés. Il est important de noter que pour les familles de gènes étudiés dans la section 3.3, l'échantillonnage taxonomique a été effectué *a posteriori*, c'est-à-dire que les familles de gènes ont d'abord été construites à partir des 937 protéomes de *Firmicutes* puis dans un second temps, les jeux de données de séquences des familles ont été échantillonnés.

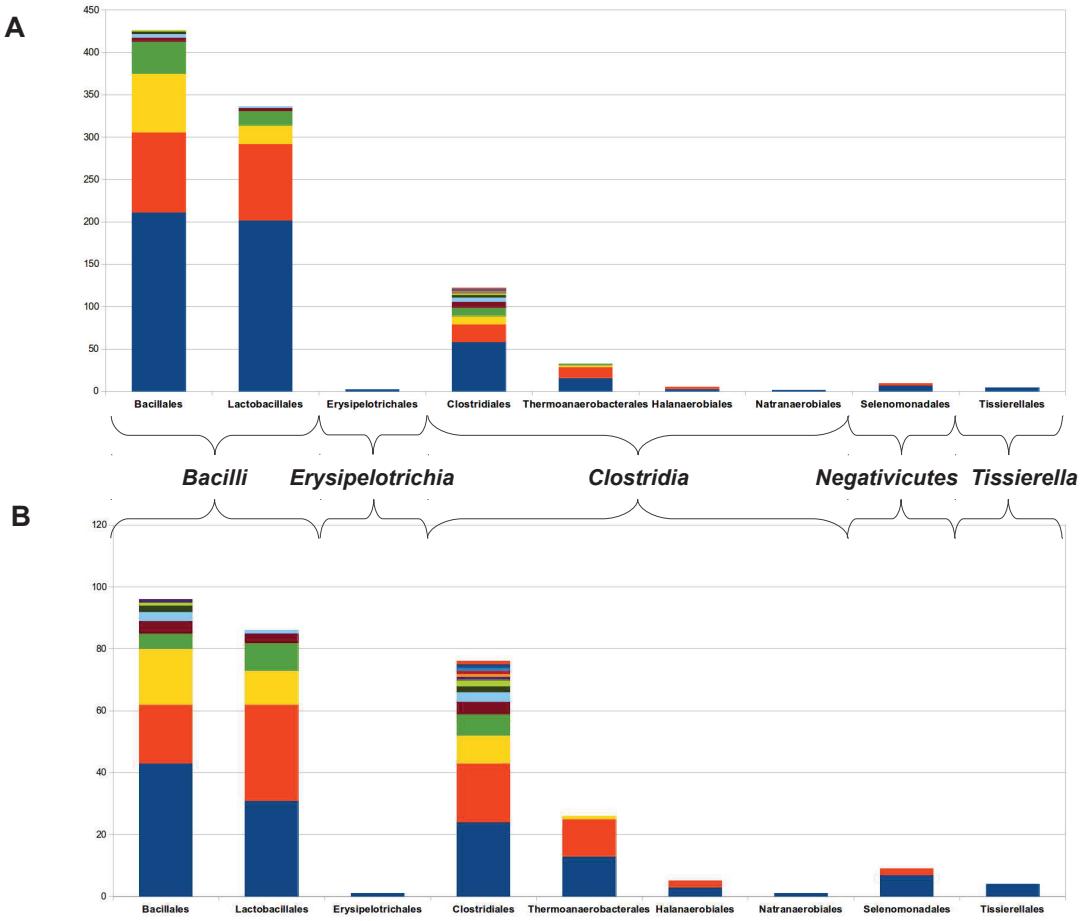


FIGURE 3.2 – Composition en classes/ordres/familles de la base de données des *Firmicutes*. (A) Avant échantillonnage taxonomique (927 protéome). (B) Après échantillonnage taxonomique (304 protéomes). Les différentes couleurs indiquent les familles.

3.2 Inférence de la phylogénie des espèces de *Firmicutes*

Afin de mener à bien l'analyse des protéines du cycle cellulaire chez les *Firmicutes*, nous avons eu besoin d'un arbre d'espèces des *Firmicutes*. L'arbre nécessitait d'être composé rigoureusement des mêmes souches dans la base de données, nous avons donc du le construire. Nous avons décidé d'utilisé les protéines ribosomiques qui sont considérées comme de bons marqueurs chez les procaryotes [396]. L'arbre a ensuite été inféré à l'aide d'une matrice spécifique des protéines ribosomiques de *Firmicutes* et d'un modèle hétérogène. Il a ensuite fallut également daté l'arbre afin d'optimiser les résultats de réconciliation.

3.2.1 Construction de l'arbre d'espèces daté des *Firmicutes*

3.2.1.1 Construction du jeu de données de séquences ribosomiques

Un arbre d'espèces des *Firmicutes* a été inféré à partir des protéines ribosomiques. Tout d'abord, les séquences des protéines ribosomiques des 937 protéomes de *Firmicutes* présents dans la base de données ont été récoltées. Un groupe extérieur de 9 espèces a été ajouté afin de raciner l'arbre (5 *Actinobacteria* et 4 *Cyanobacteria*, annexe .3). Pour extraire les protéines ribosomiques, nous avons utilisé l'outil RiboDB [219]. Au total, 53 protéines ribosomiques ont été extraites. Cependant, 147 protéomes présents dans la base de données des 937 *Firmicutes* étaient absents de la base de données de RiboDB. Un pipeline a donc été créé afin de récupérer les séquences manquantes directement à partir de la base de données locale des *Firmicutes*. Les séquences appâts ont été sélectionnées de façon aléatoire pour chaque protéine ribosomique puis une recherche par BLASTP a été effectuée contre le protéome manquant. La séquence cible présentant la E-value la plus faible a alors été sélectionnée. Cette procédure ne permet néanmoins pas de détecter la présence de paralogues. Le nombre de protéines ribosomiques identifiées pour chaque souche des 937 *Firmicutes* est présenté en annexe .4.
Les séquences ont ensuite été échantillonnées avec une souche par espèce pour obtenir 306 souches de *Firmicutes* (annexe .2).

3.2.1.2 Construction de la supermatrice

Chaque famille de protéine ribosomique a été alignée avec MAFFT (option Linsi), à partir des séquences protéiques. Les alignements ont ensuite été nettoyés avec BMGE (BLOSUM30) [238], [72]. Le nombre de séquences et de positions pour chaque famille est présenté en annexe .5. La supermatrice a été alors construite en assemblant tous les alignements nettoyés en un seul et même alignement. La matrice contient 315 séquences et 5 901 positions d'acides aminés.

3.2.1.3 Inférence phylogénétique

Une phylogénie a ensuite été inférée au maximum de vraisemblance à l'aide du programme IQ-TREE 1.4.1 [348] 3.3. La matrice de substitution a été construite directement à partir d'un jeu de données de protéines ribosomiques de *Firmicutes* (1 687 séquences, 6 200 positions) via l'outil de atgc-montpellier (http://www.atgc-montpellier.fr/ReplacementMatrix/, [76]). Afin de prendre en compte l'hétérogénité des processus évolutifs aux différents sites, nous avons utilisé 20 catégories de sites. Ce nombre de catégories est un bon compromis entre temps de calcul et prise en compte de l'hétérogénité du processus évolutif. L'algorithme PMSF (« Posterior Mean Site Frequency model ») a été utilisé pour déterminer les paramètres des catégories de sites [492]. Afin de s'affranchir de l'hypothèse d'uniformité du processus évolutif (vitesse d'évolution égale pour tous les sites), nous avons utilisé 4 catégories de taux de substitution suivant la loi Gamma. Les supports ont été estimés par trois méthodes : 100 réplicats de « Bayesian-like transformation of aLRT » (abayes), 1 000 réplicats de SH-aLRT et 1 000 réplicats de bootstraps ultra-rapides [323]. Un support mixte a ensuite été calculé sur la base des trois supports (table 3.1).

SH-aLRT > 80				SH-aLRT < 80			
UltrafastBootstrap > 95		UltrafastBootstrap < 95		UltrafastBootstrap > 95		UltrafastBootstrap < 95	
Abayes > 0,8	Abayes < 0,8	Abayes > 0,8	Abayes > 0,8	Abayes > 0,8	Abayes > 0,8	Abayes > 0,8	Abayes > 0,8
100	85	85	70	85	70	70	5

TABLE 3.1 – Algorithme de supports composites.

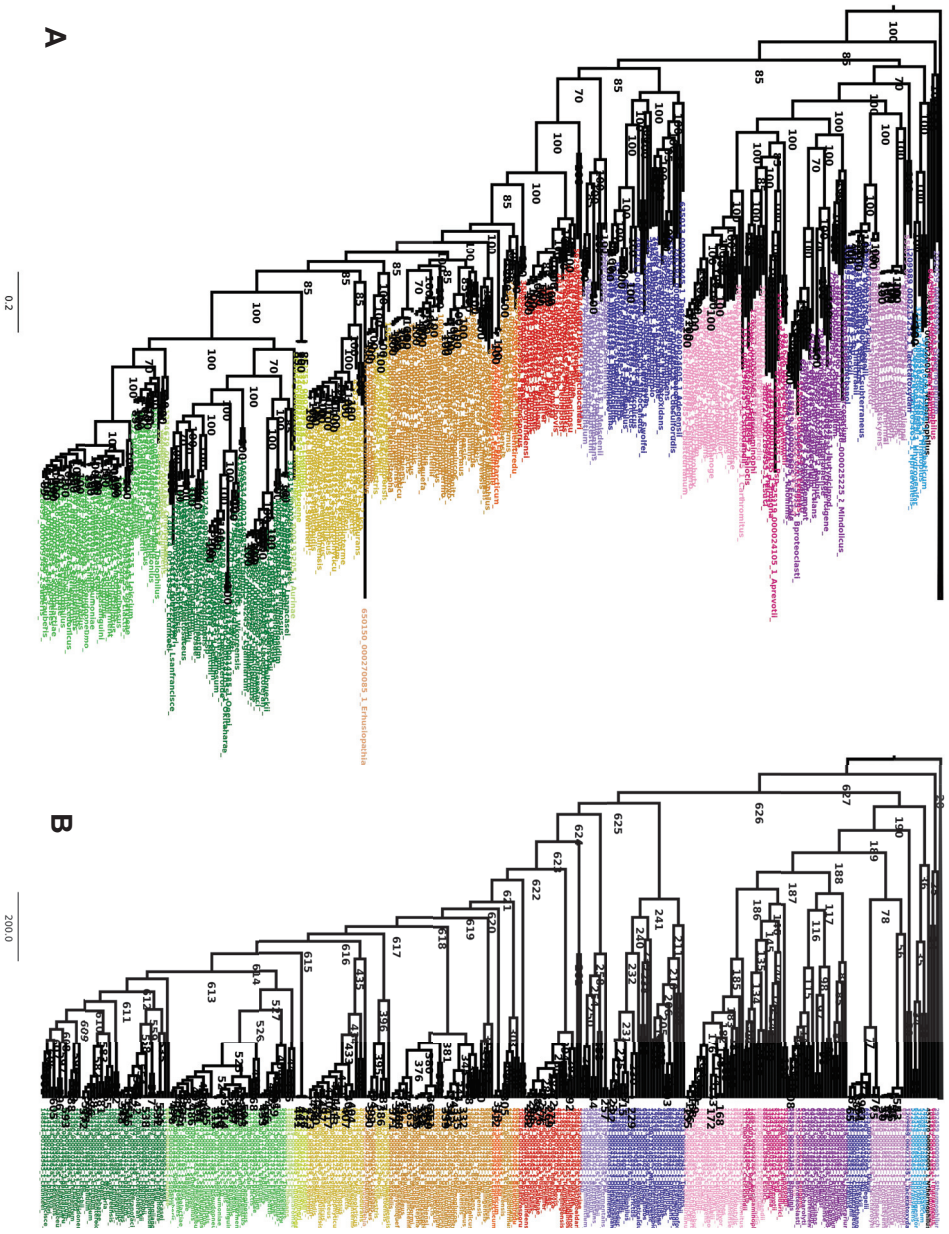


FIGURE 3.3 – Phylogénie des *Firmicutes* inférées à partir de protéines ribosomiques. (A) Phylogénie non datée. 53 protéines ribosomiques ont été utilisées. IQ-TREE, Matrice spécifique des *Firmicutes*+G4+C20+PMSF, 515 séquences, 5901 positions. Les nombres aux nœuds correspondent au support mixte, la barre d'échelle représente le nombre moyen de substitutions par site. (B) Arbre daté par inférence bayésienne. Phylobayes, Matrice spécifique des *Firmicutes*+G4+C20+U-gam+Birth-Death. Les nombres aux nœuds correspondent à une numérotation arbitraire. La barre d'échelle représente le nombre de millions d'années. Pour les deux panels : la couleur correspond aux différentes familles représentées figure 3.6.

3.2.1.4 Datation de l'arbre

L'arbre a ensuite été daté par une approche bayésienne en utilisant la suite PhyloBayes 4.1c [261], figure 3.3. Afin de dater un arbre phylogénétique, il est nécessaire d'utiliser un modèle qui modélise les vitesses d'évolution dans chaque branche de l'arbre. Le modèle le plus simple consiste en l'horloge moléculaire qui postule que la vitesse d'évolution est constante. Un modèle plus réaliste a été ensuite proposé : l'horloge moléculaire relaxée [103]. Il est alors possible d'estimer la vitesse d'évolution par branche. Deux types de modèles d'horloge relaxée ont été proposés : le modèle corrélé ou la vitesse d'évolution d'une branche mère influence celle des branches filles et le modèle non corrélé ou chaque branche est associée à une vitesse d'évolution indépendante. Dans ce contexte, nous avons utilisé le modèle de l'horloge relaxée suivant une loi gamma non corrélée. Lors de la datation, la phylogénie est calibrée au temps géologique, c'est-à-dire que les nœuds sont datés (divergence du temps). Il est possible de donner un *prior* concernant la distribution de l'age des nœuds. Ainsi, nous avons utilisé un *prior* de divergence de temps suivant le processus de « birth death ». Il est également possible de calibrer l'arbre d'espèces, c'est-à-dire de donner des nœuds dont la date absolue est connue. Concernant les procaryotes, peu d'informations géologiques sont disponibles sur l'âge des nœuds. Nous n'avons donc pas calibré l'arbre. Enfin, concernant le modèle évolutif, nous avons choisi un modèle à 20 catégories de sites, 4 catégories de taux de substitution et la matrice d'échangeabilité spécifique des *Firmicutes* pour être en accord avec la méthode d'inférence de la phylogénie.
Deux chaînes ont été lancées en parallèle sur 16 854 et 17 514 itérations. La convergence a été vérifiée en monitorant l'évolution du logarithme de la vraisemblance avec l'outil Tracer [395], figure 3.5. Également, deux indices sont donnés par les programmes bpcomp (mesure des divergences maximales entre les chaînes) et tracecomp (mesure des divergences maximales entre les chaînes et des tailles effectives) de la suite PhyloBayes. La convergence est estimée comme acceptable lorsque la divergence maximale $< 0{,}3$ et la taille effective minimale > 50. Après plusieurs tests avec différents échantillonnages et différentes valeurs de burn'in, nous avons retenu un échantillonnage de 40 (1 arbre sur 40 est échantillonné) et un burn'in de 5 000 itérations. La divergence maximale était de 0 et la taille effective minimale de 85, indiquant ainsi une bonne convergence. Cet échantillonnage a permis d'obtenir 296 et 312 arbres pour la

chaîne 1 et 2 respectivement. La convergence entre les deux chaînes étant satisfaisante, nous avons pu utilisé une des deux chaînes. La première chaîne a été arbitrairement choisie pour générer le chronogramme en utilisant readdiv de la suite PhyloBayes. L'arbre généré n'étant pas exactement paramétrique, une correction a été effectuée. Un algorithme de rectification de longueurs de branches a ainsi été implémenté en utilisant la librairie ete3 [208]. Le principe de cet algorithme est présenté figure 3.4. Enfin, les nœuds internes ont été nommés en utilisant ecceTERA 1.2.4 [420].

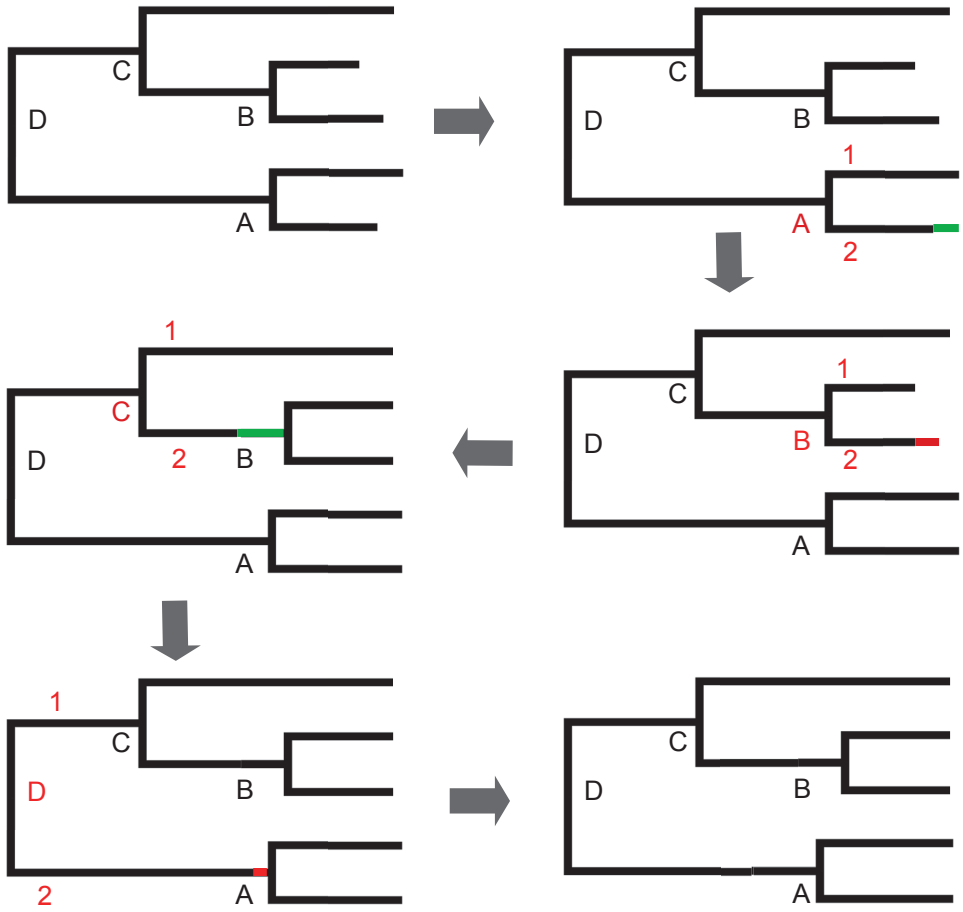


FIGURE 3.4 – Algorithme de correction des arbres pseudo-ultramétriques. L'arbre initial n'est pas exactement ultramétrique. Pour chaque nœud interne (A,B,C,D), la longueur des branches filles est ajustée pour que l'ensemble des feuilles issues de chaque branche fille soit à la même distance du nœud considéré. L'arbre final est parfaitement ultramétrique.

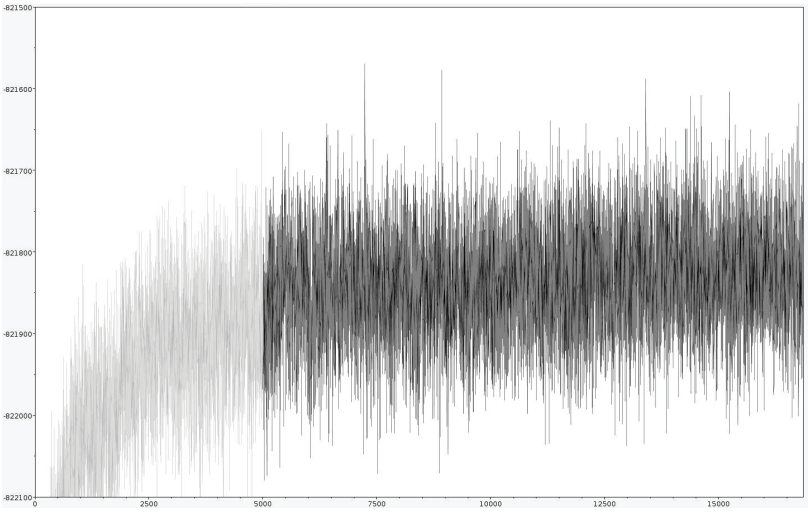


FIGURE 3.5 – Convergence de chaîne lors de la datation de la phylogénie des *Firmicutes*. L'axe des abscisses correspond au nombre d'itérations et l'axe des ordonnées au logarithme de la vraisemblance. La zone éclaircie correspond au burn-in. La chaîne semble converger à partir de l'itération 5000.

3.2.2 Analyse de la topologie de l'arbre d'espèces des *Firmicutes*

Annotation de l'arbre des espèces Afin d'estimer la concordance entre l'arbre d'espèces (systématique) et la taxonomie, les feuilles de l'arbre d'espèces ont été colorées en fonction de la taxonomie fournie par le NCBI et particulièrement par familles. Néanmoins, nous avons tenté de regrouper les familles pour minimiser le nombre de couleurs. Pour définir les groupes de familles, nous avons d'abord inféré une phylogénie d'espèces basée sur les protéines ribosomiques construite avec une souche représentative de chaque famille de *Firmicutes* (figure 3.6). Les familles groupées ensemble sur cette phylogénie ont alors été associées à la même couleur. Les souches de l'arbre d'espèces des 306 *Firmicutes* ont ensuite été associées à la couleur de leur famille d'appartenance 3.3. Cependant, il apparaît que certains groupes semblent paraphylétiques comme par exemple les *Clostridiaceae* représentés en rose. Cela provient principalement du fait que les familles annotées par le NCBI ne sont pas monophylétiques (*e.g. Thermoanaerobacteraceae*, *Clostridiaceae*). Ces incohérences traduisent une incongruence entre la systématique et la taxonomie. De plus, certains regroupements dans la phylogénie des familles

semblent provenir d'artefacts de reconstruction phylogénétique comme par exemple celui des *Negativicutes* avec les *Symbiobacteriaceae*. Il semblerait en effet que le phénomène d'attraction des longues branches soit responsable d'un tel groupement. Bien que ce code couleur ne soit pas optimal, nous avons choisi de le conserver.

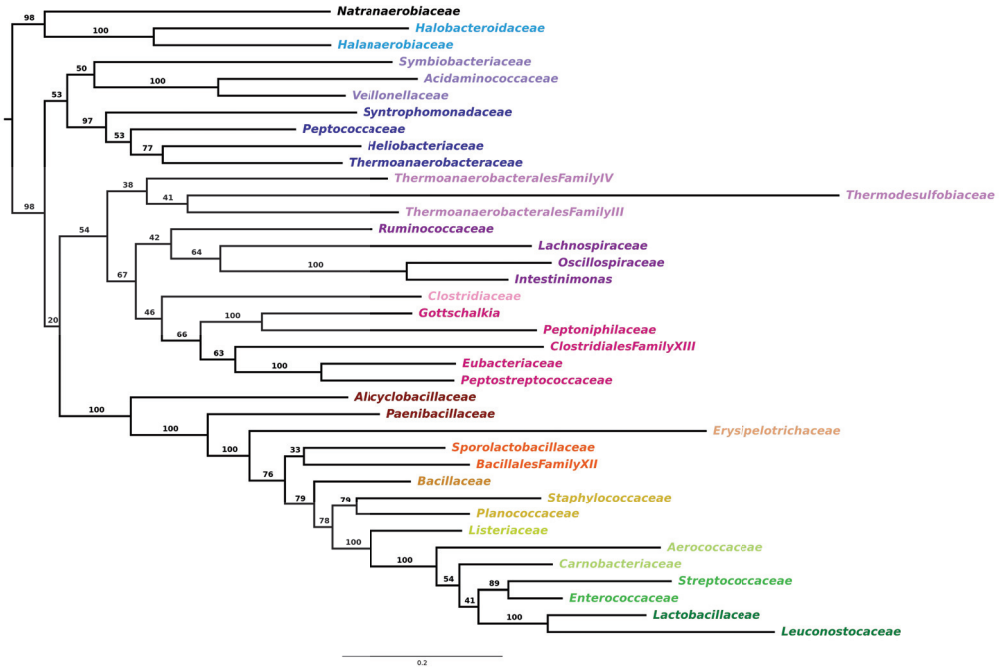


FIGURE 3.6 – Phylogénie des familles des *Firmicutes*. Une souche par familles a été sélectionnée, 53 familles de protéines ribosomiques ont été concaténées. PhyML, LG+I+G4, 38 séquences, 6133 positions. La couleur a été attribuée en fonction des groupements de familles. La barre d'échelle représente le nombre moyen de substitutions par site. Les nombres aux branches correspondent aux valeurs de bootstrap non paramétrique (100).

L'arbre ainsi inféré et annoté en fonction de la taxonomie semble globalement bien supporté (support de 95 en moyenne). La distribution des supports est présentée figure 3.7. Ainsi, 261/313 des branches présentent un support de 100 et seulement 14 branches présentent un support inférieur à 85. Les deux branches présentant un support de 5 correspondent au groupement de *Bacillus cereus* avec *Bacillus thuringiensis* et *Streptococcus gordonii* avc *Streptococcus sanguinis*. La coloration de l'arbre par la taxonomie indique que la plupart des familles de *Firmicutes* sont bien constituées malgré certaines inconsistances énoncées précédemment et que

les classes sont groupées ensemble (*Clostridia*, *Bacilli*, *Negativicutes*, *etc...*). Cette phylogénie parait donc être cohérente.

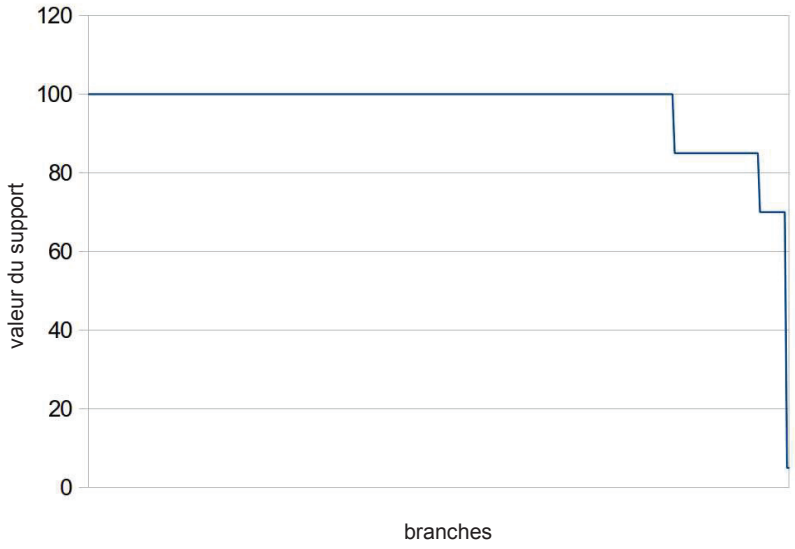


FIGURE 3.7 – Distribution des supports par branche de l'arbre d'espèces des *Firmicutes*.

Pour faciliter la description de l'arbre des espèces, nous avons décidé de définir plusieurs groupes en nous basant uniquement sur la systématique c'est-à-dire sur l'arbre d'espèces. Nous avons ainsi délimité des groupes d'espèces en clusterisant l'arbre à partir de deux seuils présentés en figure 3.8.

Ainsi, les *Clostridia* se divisent en 3 groupes : les groupes 0, 1 et 2. Le groupe 0 est composé des *Symbiobacteriaceae*, *Thermaerobacter* et *Sulfobacillus*. Le groupe 1 se subdivise en 4 groupes A, B, C et D. Le groupe 1A correspond au *Natranaerobiales* et *Halanaerobiales*, le groupe 1B de *Tepidanaerobacter acetatoxydans* et *Thermosediminibacter oceani* et le groupe 1C d'une partie des *Thermoanaerobacteraceae*, des *Clostridia* famille III et IV. Le dernier groupe 1D est subdivisé en deux groupes D1 et D2. Le sous-groupe D1 contient les *Ruminococcaceae*, les *Oscillospiraceae*, les *Lachnospiraceae*, les *Intestimonas*, une partie des *Eubacteriaceae* et les *Gottschalkia*. Le sous-groupe D2 contient les *Tissierellia*, une partie des *Eubacteriaceae*, les *Clostridiaceae* et les *Peptostreptococcaceae*. Le dernier groupe des *Clostridia* 2 contient les *Peptococcaceae*, les *Heliobacteriaceae*, les *Syntrophomonadaceae* et une partie des *Thermoanaerobacteraceae*. Les autres clades sont séparés classiquement en *Negativicutes* et *Bacilli*.

Il est aussi intéressant de noter que les souches de *Coprothermobacter proteolyticus* et *Thermodesulfobium narugense*, pourtant annotées comme faisant partie des *Firmicutes*, se placent dans le groupe extérieur. Ce résultat est corroboré par l'étude de Kunisawa et collègues qui, à partir de l'ARN 16S, ont fait la même observation [254]. Nous avons donc exclu ces deux souches pour les analyses des familles de gènes.

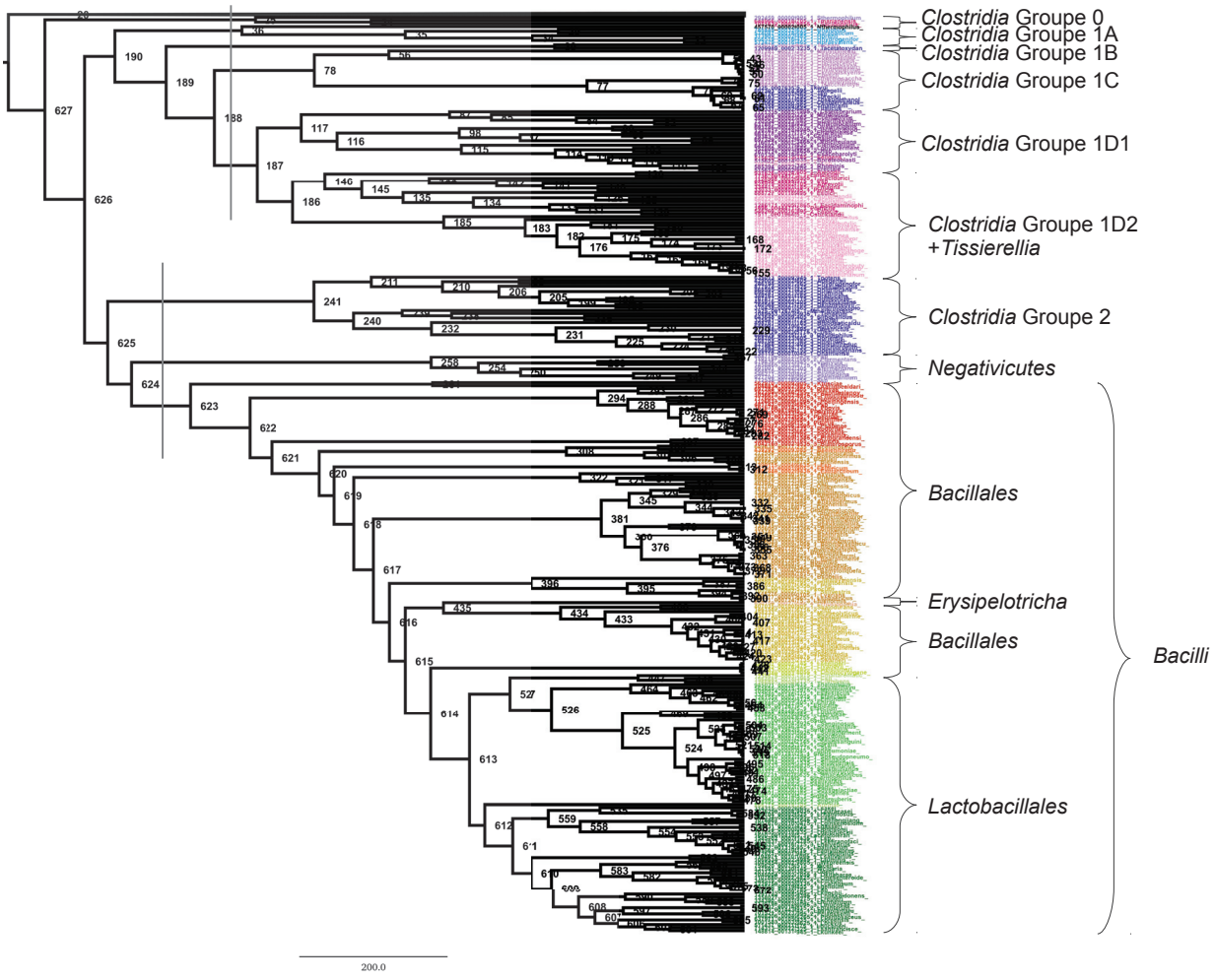


FIGURE 3.8 – Phylogénie des*Firmicutes* de cette étude. Les groupes ont été délimités par les deux seuils représentés en gris.

Comparaison avec la phylogénie de Antunes *et al.*, 2016 Nous avons ensuite comparé la topologie obtenue à partir de notre jeu de données avec celle obtenue par Antunes et collègues [11] (figure 3.9). Cet arbre est, d'après nos connaissances, celui qui semble le plus jus-

tifié méthodologiquement dans la littérature. En effet, les auteurs ont utilisé un grand nombre de souches de *Firmicutes* (205) ainsi qu'un grand nombre de marqueurs (47 protéines ribosomiques). De plus, deux méthodes reconnues comme performantes ont été utilisées pour inférer l'arbre (inférence bayésienne et maximum de vraisemblance).
Dans la topologie de Antunes et collègues, plusieurs bipartitions profondes diffèrent de celles obtenues dans notre étude (figure 3.10). Tout d'abord, l'émergence des *Bacilli* est plus basale dans cette topologie, c'est-à-dire que ce groupe émerge au deuxième nœud en partant de la racine. De plus, le groupe *Clostridia* 1A est à la racine alors qu'il est regroupé avec les 1B, 1C et 1D dans notre arbre d'espèces. Enfin, le groupe 0 n'est pas à la racine comme dans notre topologie mais groupé avec le groupe *Clostridia* 2 et les *Negativicutes*. Il y a donc trois différences majeures entre les deux arbres. Ces différences sont corroborées par les supports faibles aux branches concernés (figure 3.10). Néanmoins, il est à noter que les branches de l'arbre bayésien de Antunes et collègues présentent des supports très élevés (probabilités postérieures égales à 1) pour la quasi-totalité des bipartitions.

Image non disponible

FIGURE 3.9 – Phylogénie des*Firmicutes* de Antunes *et al.* [11]. Les groupes ont été délimités par les deux seuils représentés en gris.

Ces différences de topologies proviennent de plusieurs différences méthodologiques. Premièrement, l'échantillonnage taxonomique des deux arbres est différent. En effet, nous avons 304 souches de *Firmicutes* contre 205 dans l'étude de Antunes. De plus, l'étude de Antunes est focalisée sur les *Negativicutes* et présente donc plus de génomes de ce clade (38 contre 9). Ensuite, bien que les deux phylogénies aient été construites à partir d'une supermatrice de protéines ribosomiques, nous avons plus de positions (5901 contre 5551) puisque nous avons utilisé 6 marqueurs de plus. Enfin, les modèles utilisés ne sont pas les mêmes. Nous avons

utilisé une matrice spécifique des protéines ribosomiques des *Firmicutes* alors que l'arbre en maximum de vraisemblance de Antunes a été inféré à partir du modèle LG pour l'arbre en maximum de vraisemblance et du modèle CAT pour l'arbre inféré par approche bayésienne. Il est difficile d'estimer lequel des deux arbres est le plus proche de la réalité et cela nécessiterait une investigation plus profonde. Il faudrait notamment optimiser l'échantillonnage taxonomique, surtout dans le groupe des *Clostridia* qui souffre d'une sous-représentation. Également, il serait peut-être nécessaire d'utiliser des méthodes permettant de diminuer la saturation du signal phylogénétique comme avec la méthode Slow-Fast [49], [249].

3.3 Construction des familles de gènes impliqués dans le cycle cellulaire

3.3.1 Méthodes de construction des familles

3.3.1.1 Constitution de la liste des protéines appâts

Une veille bibliographique a été effectuée pour identifier les protéines ayant été démontrées comme étant impliquées dans le cycle cellulaire chez *S. aureus*, *E. coli*, *B. subtilis* et *S. pneumoniae* et autres bactéries modèles. Parmi celles-ci, les protéines Fts (« Filamentous temperature sensitive »), Min (« Minicell »), Mre (« MuRein E »), Mra (« MuRein A »), Mur (« MURein »), Xer (« CER-specific recombination »), Rec (« RECombination »), Spo (« SPOrulation »), Rod (« ROD-shape »), Scp (« Segration and Condensation Protein), Wal (« cell WALl »), Zap (« Z-ring Associated Protein »), Zip (« Z-ring Interacting Protein »), Cps (« Capsular Polysaccharide Synthesis »), Dna («DNA »), Dac (« D-Alanine Carboxypeptidase »), Div (« DIVision »), Glm (« GLucosaMine »), Nag (« N-AcetylGlucosamine »), PBP (« Penicillin-Binding Protein ») et Fem (« Factors Essential for Methicillin resistance ») ont été sélectionnées étant donné qu'elles sont nommées ainsi de part leur fonction dans le cycle cellulaire. Pour trois protéines étudiées lors de l'analyse, les publications concernant leur ca-

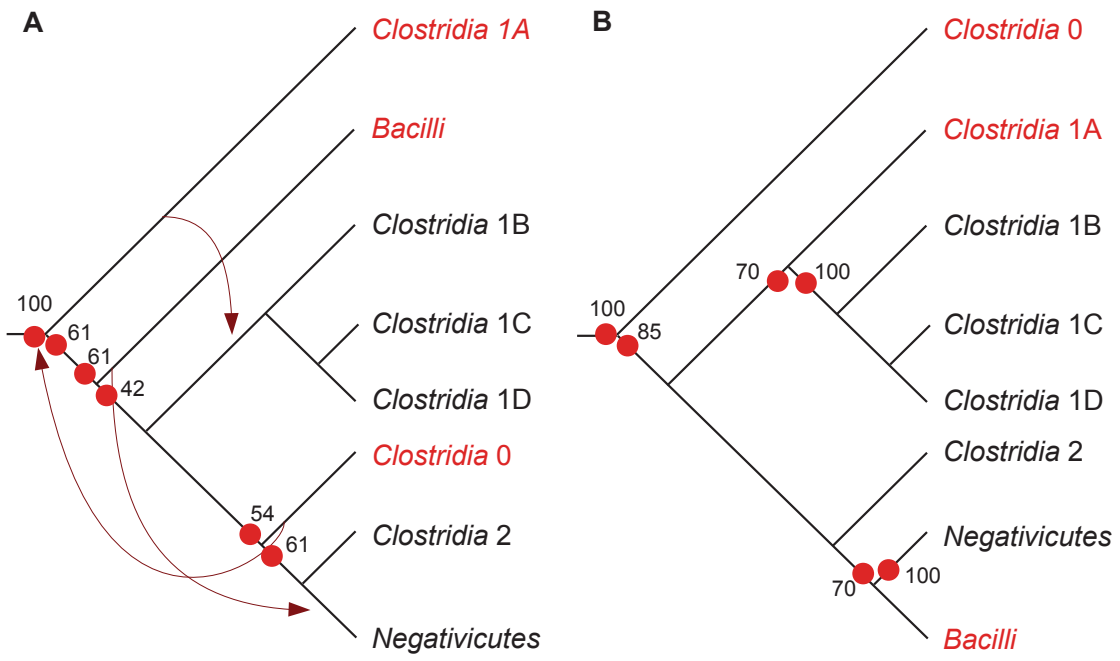


FIGURE 3.10 – Comparaison des deux topologies.(A) Topologie inférée par [11]. (B) Topologie inférée dans cette étude. Les arbres ont été collapsés en sous-groupes précédemment décrits. Les groupes en rouge sont les groupes placés différemment dans les deux topologies. Les branches mères et sœurs des branches concernées sont représentées par des points rouges avec leur support associé. Les flèches rouges correspondent aux trois modifications de topologie à effectuer sur l'arbre de [11] pour obtenir celle de cette étude.

ractérisation n'est pas encore parues. Elles sont dénommée CDP1, CDP3 et CDP4 pour « Cell Division Protein ».
La veille bibliographique effectuée a ainsi permis d'identifier 261 protéines impliquées dans la réplication, la ségrégation chromosomique, la synthèse, le remodelage et le recyclage du peptidoglycane, la division, l'élongation, la sporulation et la synthèse de la capsule (annexes .6, .7, figure 3.11). Certaines de ces familles présentaient des histoires évolutives trop complexes pour identifier des sous-familles (*e.g.* PBPE). La complexité de la construction de ces familles provenait principalement du fait qu'elles étaient impactées par de nombreux transferts horizontaux de gènes. D'autres protéines ne présentaient qu'un intérêt restreint pour l'étude de par leur rôle très éloigné du cycle cellulaire (*e.g.* MreA). Nous les avons donc écarté de l'analyse. Pour la sporulation, la réplication et la synthèse de la capsule, nous avons délibérément choisi quelques marqueurs puisque ces trois processus impliquent un très grand nombre de protéines. Pour un processus donné, nous avons donc choisis les protéines qui étaient essentielles au processus (*e.g.* DnaA), dont les gènes respectifs présentaient des synténies qualitativement conservées avec les autres gènes du cycle cellulaire (*e.g.* PriA) ou étudiées par le laboratoire (*e.g.* CpsD). Nous avons donc sélectionné 13 protéines de la sporulation, 21 de la réplication et 4 de la synthèse de la capsule comme marqueurs. Pour ce qui est des protéines impliquées dans le métabolisme du peptidoglycane, 33 protéines n'ont pas été inclues. Ces dernières correspondent principalement à des hydrolases dont l'histoire évolutive étaient complexes avec de nombreux transferts horizontaux de gènes. Nous avons donc analysé 148 des 261 protéines impliquées dans le cycle cellulaire (figure 3.11).

L'analyse des contextes génomique des protéines de division nous a amené à inclure 15 protéines dont les gènes correspondant sont localisées au voisinage de clusters de gènes du cycle cellulaire chez *B. subtilis* et *S. pneumoniae* et dont la localisation chromosomique semblait conservée (figure 3.12). Il s'agit des protéines NudF, IleS, LeuS, ValS, AlaS, Mtf, Pfs, SunL, TilS, Mfd mais aussi PCDP4/6/7/8/10 (« Putative Cell Division Protein »). Au total, 163 séquences ont été utilisées pour initier la recherche d'homologues.
Les analyses phylogénétiques ont révélé l'existence de 21 familles homologues aux familles

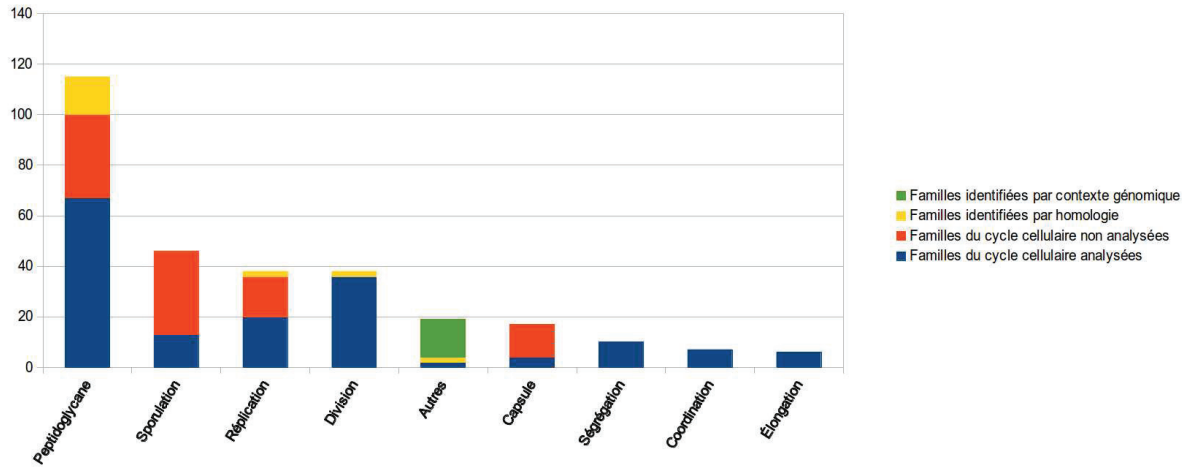


FIGURE 3.11 – Familles de gènes étudiées par processus cellulaire. Les familles non analysées sont principalement impliquées dans l'hydrolyse du peptidoglycane, la sporulation, la réplication et la production de la capsule.

initialement étudiées qui de par leur proximité phylogénétique et/ou leur contexte génomique présentent un intérêt : Alr-like, CozE4, DnaB2, FtsH2, FtsH3, RodA2, GatD2, GatD3, GatD4, IleS2, LysA2, LysA3, MurA2, MurB2, MurG2, YabM, PBP-AX1, PBP-BX1, MGT2, SbcC1, SbcC2 (table .7). Ces familles ont été aussi sélectionnées dans la plupart des cas de par leur cohérence relative avec la taxonomie. En effet, les familles paralogues à la famille d'origine qui présentaient des incohérences majeures avec la taxonomie n'ont pas été analysées (sauf la famille Alr-like).

3.3.1.2 Présentation générale du pipeline de d'analyse

Un pipeline d'analyses a été développé afin de détecter les homologues de chaque protéine présents dans la base de données protéique de *Firmicutes*. Ce pipeline couple deux approches de recherche de similarité : la recherche par BLASTP et la recherche par profil HMM (« Hidden Markov Model »). L'intérêt de coupler ces deux approches de recherche de similarité de séquences a déjà été démontré dans le cas de RiboDB [219]. Le pipeline ainsi développé a été testé initialement sur un jeu de données de 30 protéines de la division afin de tester ses

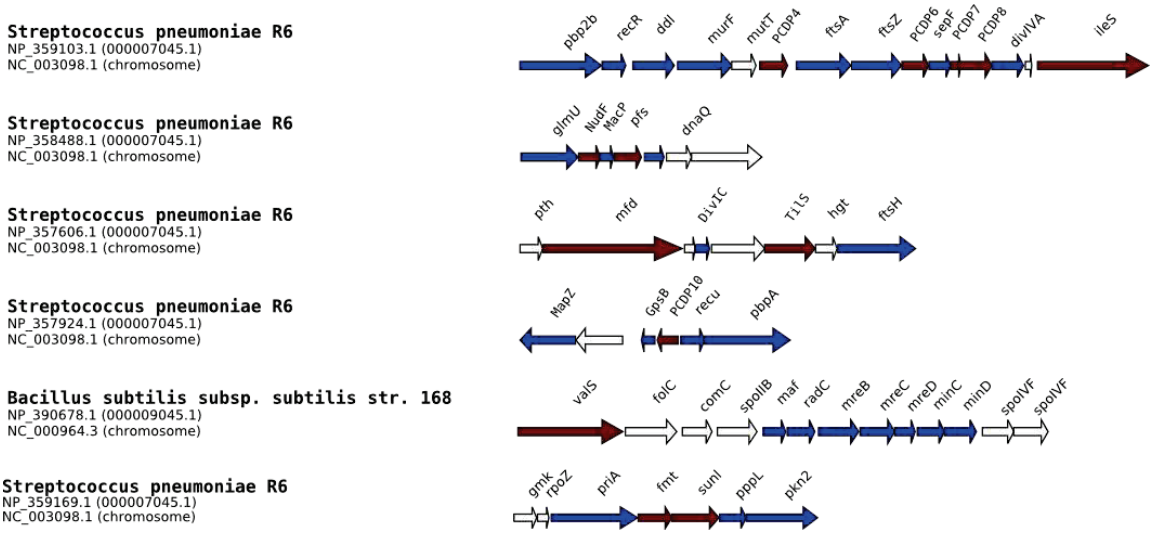


FIGURE 3.12 – Contexte génomique des clusters du cycle cellulaire chez *S. pneumoniae* et *B. subtilis*. Les gènes colorés ont été analysés. Les gènes en bleu sont impliqués dans le cycle cellulaire, les gènes en rouge sont impliqués dans d'autres processus cellulaires.

performances, d'optimiser ses paramètres et d'implémenter des options. L'ensemble des étapes du pipeline est résumé sur la figure 3.13.

3.3.1.3 Séquences initiales utilisées

Les séquences protéiques appâts ont été sélectionnées dans les protéomes suivants par ordre de priorité (annexes .6, étape 1 de la figure 3.13). Le génome de *S. pneumoniae* R6 a été choisi comme étant la référence. Si un gène n'est pas présent ou mal annoté chez *S. pneumoniae* R6, nous avons utilisé des séquences de *S. pneumoniae* D39 comme alternative, notamment pour les gènes de production de la capsule. Si la famille d'intérêt n'est pas représentée dans ces deux *Streptococcus*, nous avons sélectionné les gènes présents chez *B. subtilis* str. 168 ou *E. coli* K-12. Enfin, si aucun représentant de la famille considérée n'est présent dans ces quatre protéomes, nous avons sélectionné les séquences appâts chez d'autres bactéries. Les séquences des protéines codées par chaque gène ont été récupérées en format FASTA à partir de la base

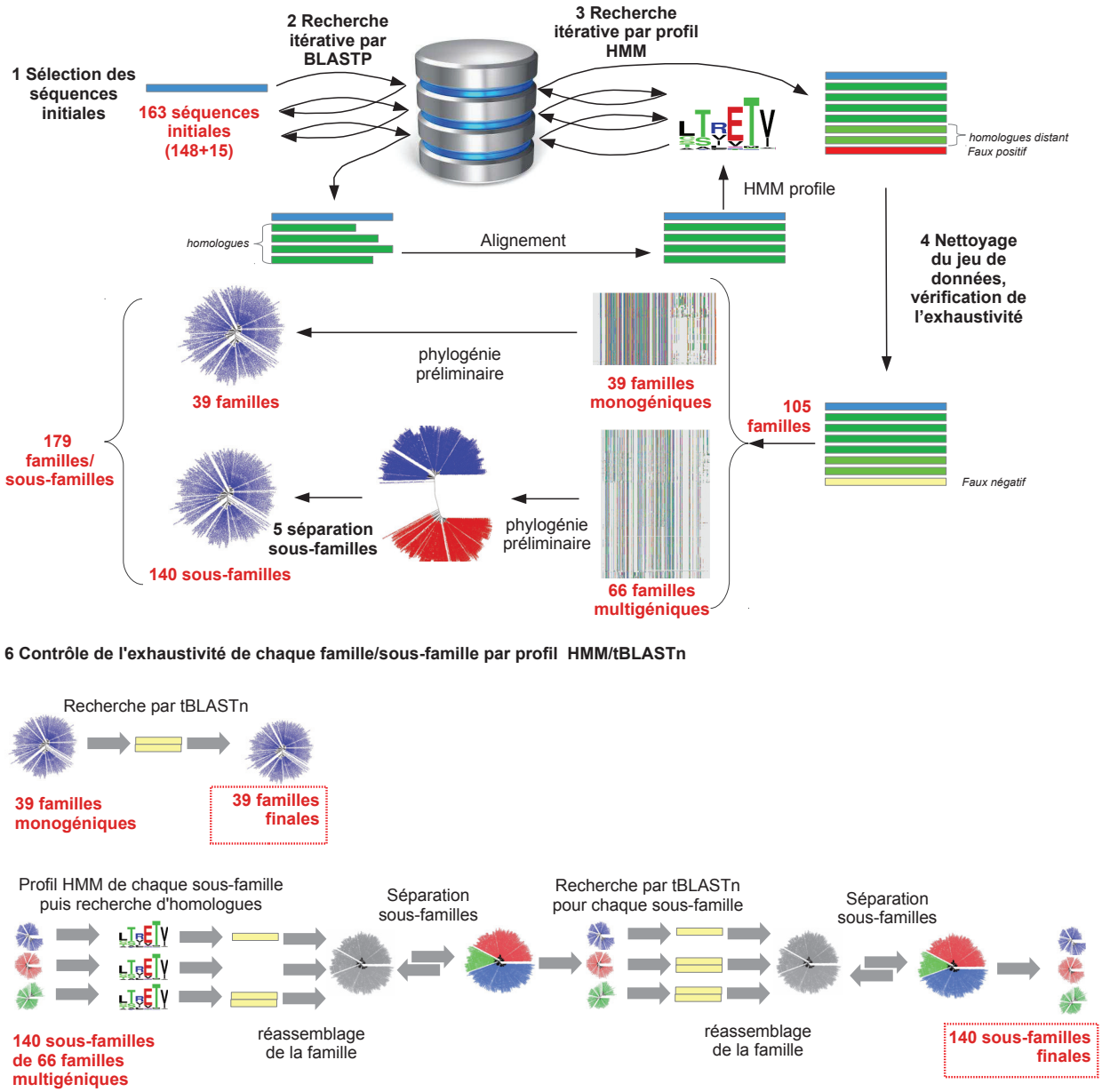


FIGURE 3.13 – Pipeline de construction de familles de gènes. En gras : les étapes principales numérotées, en rouge : jeux de données de séquences, en encadrés rouges : jeux de données finaux de familles de gènes.

de données UniProt [45].

3.3.1.4 Recherche par similarité avec BLASTP

Nous avons d'abord utilisé une recherche de similarité de façon itérative en utilisant BLASTP (étape 2 de la figure 3.13). Le principe de ce type de recherche est présenté figure 3.14.
Tout d'abord, un BLASTP est effectué à partir de la séquence protéique contre la base de données protéique des *Firmicutes* en utilisant la suite BLASTALL 2.2.25 [5]. Les recherches ont été réalisées avec les paramètres par défaut hormis les paramètres de sortie (-v 20 000 et -b 20 000) permettant d'afficher dans la grande majorité des cas la totalité des séquences cibles. À partir de la sortie de BLASTP, un seuil d'E-value est établi à partir des longueurs des séquences cibles. Les séquences cibles sont parcourues par ordre croissant de E-value. Si quatre séquences à la suite présentent une longueur supérieure à 7/5 ou inférieure à 3/5 de la longueur de la séquence appât ou de l'alignement, le seuil de E-value correspond à la E-value de la séquence située avant les quatre séquences précédentes. Toutes les séquences dont la E-value est inférieure ou égale à ce seuil sont ensuite sélectionnées. Si la condition précédente n'est jamais satisfaite, le seuil de E-value est fixé à 10^{-4} pour les séquences appâts de longueur supérieure à 150 acides aminés et de 1 pour les séquences de longueur inférieure à 150 acides aminés. Une nouvelle séquence appât est alors sélectionnée parmi les séquences dont la E-value est inférieure au seuil selon l'équation 3.1. Cette opération est répétée jusqu'à ce que le nombre total de séquences récupérées soit constant. Les séquences sont ensuite alignées à l'aide du programme MAFFT 7.123b (option auto) [238].

Soit n le nombre de séquences présentant une E-value inférieure au seuil, l la longueur de la séquence appât et p tel que $p = 50$ si $l \leq 150$ et $p = 75$ si $l > 150$, l'indice I de la séquence appât sélectionnée dans la liste ordonnée des séquences présentant une E-value inférieure au seuil s'exprime :

$$I = \left[\frac{np}{100}\right] \tag{3.1}$$

Des options ont été implémentées afin d'améliorer la détection des homologues lointains :

- -f : la séquence appât pour la prochaine itération est la dernière de la liste des séquences sélectionnées lors du BLASTP ($p = 100$)
- -b : Matrice BLOSUM45 utilisée au lieu de BLOSUM62
- -w : la fenêtre d'initiation de l'alignement local lors du BLASTP est fixée à 2 acides aminés au lieu de 3

Les cas ou ces options ont été utilisées sont présentés en annexe .6.

3.3.1.5 Recherche par profil HMM

Les séquences présentes dans le jeu de données issu de la recherche de similarité par BLASTP sont ensuite utilisées pour effectuer une recherche par profil HMM (étape 3 de la figure 3.13). Les séquences sont d'abord échantillonnées (une souche par espèce) puis alignées. Si le nombre de séquences est inférieur à 400, l'alignement est fait par MAFFT avec l'option linsi, sinon par avec l'option auto [238]. Les potentiels régions spécifiques de quelques séquences (comme des domaines additionnels ou des fusions de protéines) situés aux extrémités N-terminale et C-terminale sont ensuite éliminés par l'utilisation de Gblocks. Gblocks permet de détecter les blocs conservés puis les régions amont du premier bloc et aval du dernier bloc conservé sont ensuite éliminés de l'alignement [62].
Un profil HMM est construit à partir de l'alignement avec le programme HMMbuild de la suite HMMER 3.1b2 [119]. Le profil est ensuite utilisé pour requêter la base de données de séquences protéiques des *Firmicutes* par le programme HMMsearch. Les recherches ont été réalisées avec les paramètres par défaut. À partir de la sortie, toutes les séquences présentant une E-value inférieure à 0,01 et absentes du jeu de données issu de la recherche par BLASTP sont sélectionnées.

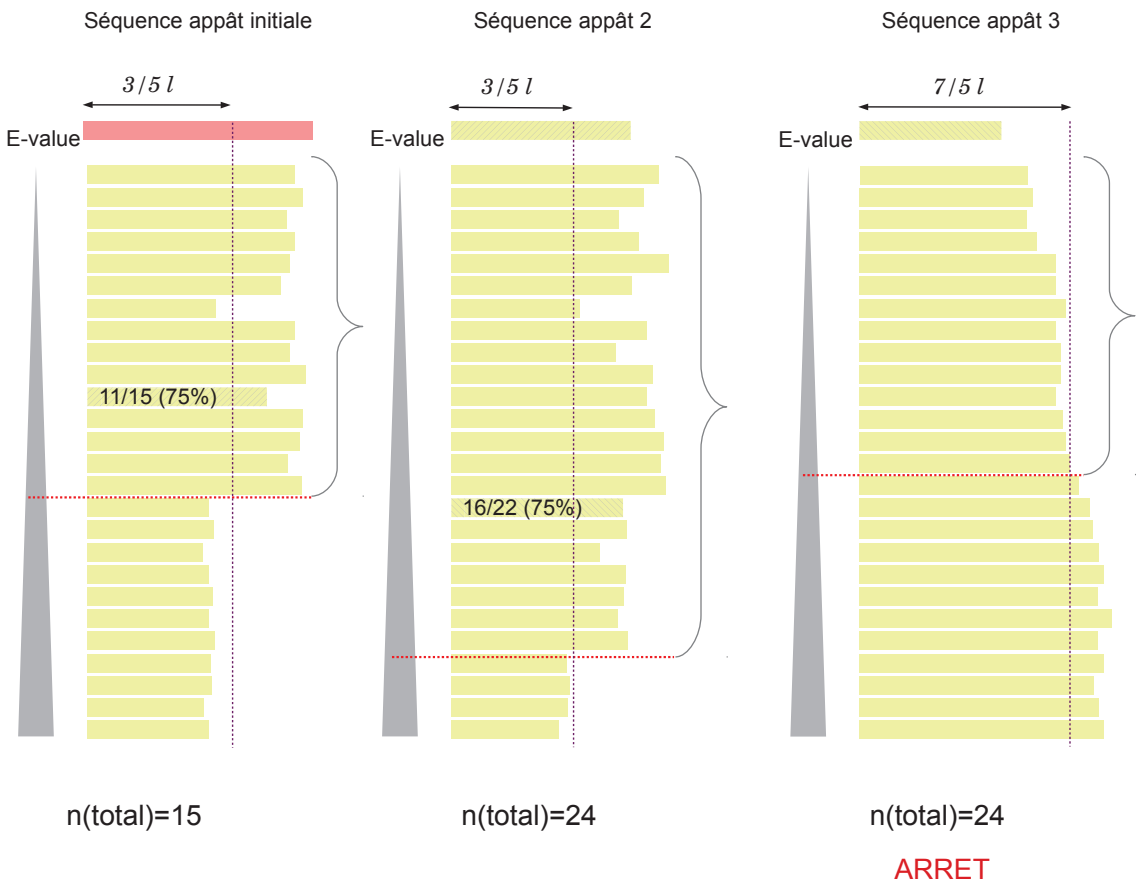


FIGURE 3.14 – Principe de la recherche itérative par BLASTP. La séquence appât initiale est présentée en rouge, les séquences issues de la base de données en jaune, le gradient de E-value en triangle gris, le seuil de E-value en trait rouge pointillé et le seuil de longueur des séquences en trait violet pointillé, les séquences sélectionnées dans le jeu de données d'homologues par une accolade. Lors du premier BLASTP, à partir de la 16ème séquence, les séquences sont plus courte, le seuil est donc choisi comme étant la E-value correspondant à la séquence précédente. La 11ème séquence est utilisée comme nouvelle séquence appât. À la troisième itération, le nombre d'homologues sélectionnés n'a pas augmenté, l'itération est stoppée.

L'homologie de chacune de ces séquences avec les séquences issues de la recherche par BLASTP est vérifiée à l'aide de BLASTP. Une base de données de 10 protéomes représentatifs de la diversité des *Firmicutes* et de quelques autre bactéries a été préalablement constituée (annexe .8). Les séquences initialement présentes dans le jeu de données issu de la recherche par BLASTP provenant de ces 10 protéomes sont considérées comme les séquences de référence. Une recherche par BLASTP est effectuée à partir de chacune des séquences ayant été détectées avec HMMsearch contre la base de données des 10 protéomes. Si au moins une séquence de référence est située dans les premières séquences cibles (deux fois le nombre de séquences références) avec une E-value inférieure à 10^{-4}, la séquence testée est considérée comme un homologue et est ajoutée au jeu de données. Le pipeline offre aussi la possibilité d'être moins stringent en sélectionnant les séquences tests si au moins une séquence de référence a été détectée par le BLASTP avec une E-value inférieure ou égale à 10^{-4}.
Si la recherche par profil a conduit à l'ajout de séquences par rapport au jeu de données issu de la recherche par BLASTP, une seconde recherche par profil est effectuée de la même manière. L'ensemble des séquences ayant été détectées comme étant des homologues sont ensuite alignées à l'aide du programme MAFFT (option auto) [238].

3.3.1.6 Vérification des jeux de données

Tous les alignements sont systématiquement vérifiés visuellement en utilisant Aliview 1.18.1 et Seaview 4.6.1 [260], [175] (étape 3 de la figure 3.13). Si une séquence ne parait pas alignée de façon satisfaisante avec les autres, un BLASTP réciproque est effectué à partir de cette séquence. Plus précisément, une recherche par BLASTP est réalisée à partir de cette séquence contre la base de données des 937 protéomes de *Firmicutes*. La séquence est conservée uniquement si les premières séquences cibles appartiennent à la famille de gènes d'origine. La visualisation de contexte génomique du gène associé à la séquence peut permettre de valider ou non l'homologie de cette séquence, notamment si le gène présente le même contexte génomique qu'un grand nombre de gènes de la famille considérée.
Les séquences détectées par la recherche par profil HMM mais considérées comme non-homologue après recherche par BLASTP contre la base de données des 10 protéomes sont systématique-

ment vérifiées. Dans certains cas, il est ainsi nécessaire d'intégrer des séquences considérées comme non-homologues par le pipeline mais dont l' homologie est vérifiée manuellement. Cette vérification est effectuée comme expliqué précédemment par BLASTP réciproque et par visualisation des contextes génomiques.

3.3.1.7 Identification des sous-familles

Les alignements de séquences homologues sont ensuite utilisées afin d'inférer des phylogénies préliminaires à l'aide du programme Fasttree 2.1.8 [389] (modèle WAG+G4). Certains alignements ont été préalablement nettoyés à l'aide de BMGE 1.1 (BLOSUM30) [72] mais d'autres n'ont pas été nettoyés. En effet, certains alignement ne possèdent pas suffisamment de positions conservées et informatives pour être nettoyés. Les phylogénies préliminaires sont ainsi inférées directement à partir de l'alignement original.
À partir des phylogénies préliminaires, nous avons observé deux cas de figures. Dans le premier cas, certaines familles ne présentent en moyenne qu'une seule copie et sont *a priori* orthologues bien que des duplications et des transferts horizontaux aient pu avoir eu lieu ponctuellement et récemment dans certains *taxa*. Ces familles sont qualifiées de monogéniques. Dans le deuxième cas, les familles présentent plus de une copie par génome en moyenne. Les phylogénies indiquent l'existence de plusieurs sous familles d'orthologues, paralogues entre-elles. Il est nécessaire de découper les familles d'homologues en sous-familles monogéniques (étape 5 de la figure 3.13). Cette opération en apparence simple s'est révélée plus complexe que prévue. En effet, le manque de signal phylogénétique ou l'existence de transferts horizontaux et/ou des pertes multiples peut rendre complexe l'analyse de l'histoire évolutive des familles. L'identification des sous-familles s'appuie sur plusieurs types d'informations :

- La topologie de l'arbre
- La distribution taxonomique des homologues
- Le contexte génomique des homologues
- La composition en domaine des homologues
- La longueur des séquences des homologues

– L'alignement des séquences des homologues
– Les liens de parenté des homologues *Firmicutes* avec les homologues des autres bactéries

La convergence de ces informations permet dans la plupart des cas de proposer un découpage fiable. Le découpage en sous-familles des familles mutligéniques est présenté en annexe .11. Néanmoins, le découpage d'une familles multigénique en familles monogéniques est beaucoup plus délicat que dans le cas des familles monogéniques. La difficulté provient de plusieurs facteurs, notamment la phylogénie qui n'est pas toujours cohérente avec l'arbre d'espèces, un contexte génomique non discriminant ou encore une faible distance phylogénétique avec les autres sous-familles. Afin de classer les familles en fonction de la confiance que l'on peu accorder à leur construction, nous avons défini un indice de fiabilité. Cet indice qualitatif est basé sur la distance phylogénétique entre la sous-famille considérée et les autres, sur la présence ou non d'un contexte génomique discriminant, sur la cohérence entre la phylogénie de la famille et la phylogénie des espèces et sur la possibilité de discriminer la sous-famille au sein de l'alignement multiple (régions conservées spécifiques des différentes sous-familles). L'indice est compris entre 0 et 2 (0 : peu fiable, 2 : fiable). Les différents cas rencontrés sont présentés en section 3.3.2.

3.3.1.8 Contextes génomiques et composition en domaines

La visualisation des contextes génomiques permet de discriminer les différentes sous-familles dans le cas où une sous-famille possède un contexte génomique spécifique et conservé. Néanmoins, les outils disponibles au début de ma thèse ne permettaient pas la visualisation des contextes génomiques de l'ensemble des gènes de mes familles puisque les bases de données sur lesquels ces outils reposent ne sont pas concordantes avec la base de données de protéomes de *Firmicutes* que nous avons construite. Il a donc été nécessaire de développer un outil adapté à n'importe quelle base de données locale. Nous avons donc développé GeneSpy, un visualiseur de contexte génomique [157]. Le chapitre 5 est dédié à l'implémentation de cet outil.
La composition en domaines des séquences permet aussi dans certains cas le découpage en sous-familles des familles multigéniques bien que la plupart du temps, cet information ne soit pas discriminante. La composition en domaine fonctionnels des séquences est étudiée en uti-

lisant les profils HMM de la base de données Pfam (http://pfam.xfam.org/). Les domaines Pfam au sein des séquences d'intérêt sont détectés par l'utilisation du programme HMMscan de la suite HMMER 3.1b2 [119].

3.3.1.9 Contrôle de la complétude des familles de gènes

La complétude des familles de gènes considérées comme orthologues est systématiquement vérifiée (étape 6 de la figure 3.13).
Pour vérifier la complétude des sous-familles des familles multigéniques, un profil spécifique de chaque sous-famille d'intérêt est tout d'abord construit puis utilisé pour requêter la base de données de protéomes de *Firmicutes*. Dans quelques cas, des séquences additionnelles ont été détectées et ajoutées aux jeux de données. Le découpage en sous-familles a ensuite été réalisé comme expliqué précédemment.
Pour toutes les familles (monogéniques et multigéniques), un tBLASTn supplémentaire est systématiquement réalisé afin d'identifier les séquences mal annotées ou annotées en tant que pseudogènes. La recherche est uniquement effectuée sur les protéomes pour lesquels aucun homologue de la famille considérée n'a été détecté. Plusieurs séquences sélectionnées aléatoirement dans le jeu de données initial sont utilisées comme séquences requête. Toutes les séquences présentant une E-value inférieure à 10^{-4} et une longueur comprise entre 1/2 et 3/2 de la longueur de la séquence requête sont sélectionnées. Toutes les séquences qui se chevauchent sont considérées comme se référant au même gène et celle qui présente parmi celles-ci la E-value la plus faible est sélectionnée. Les pseudogènes sont annotés par concaténation de quatre éléments : « PSDG », le placement de la séquence dans les résultats de tBASTn, la position génomique moyenne du gène et la E-value correspondante. Pour les familles multigéniques, la recherche par tBLASTn est effectuée pour chaque sous-famille. Si des séquences sont ajoutées, le découpage de la famille en sous-familles est de nouveau effectué.
Une étape supplémentaire de vérification est effectuée à ce stade au niveau des alignements. Toutes les séquences dont l'alignement apparaît ambigu sont de nouveau testées par BLASTP réciproque et retirées des jeux de données si considérées comme non-homologues.

3.3.2 Composition des familles

3.3.2.1 Nomenclature des familles et généralités

Pour chaque famille d'intérêt, les homologues présents dans la base de données des 937 *Firmicutes* ont été identifiés (annexe .6). Il est important de noter que nous parlons ici de familles de gènes et non de protéines. En effet, bien que nous ayons utilisé les séquences protéiques, nous considérerons ici qu'il s'agit d'une famille de gènes. De plus, nous parlerons de famille/sous-familles de gènes et non d'orthologues. En effet, même si une grande majorité des séquences présentes dans les jeux de données construits sont orthologues, il existe un certain nombre de séquences paralogues issues de duplications et xénologues issues de transferts horizontaux.

Concernant la nomenclature des familles de gènes, nous avons décidé de leur attribuer le nom de la protéine qui a été caractérisée au sein de la famille. Les familles additionnelles ont été numérotées de 2 à plus (exemple : RodA2). Les PBP ont été nommées d'après la classification de Sauvage et collègues [414] avec un code à une lettre et un chiffre (exemple : A3). Les PBPs supplémentaires ont été nommées AX1 et BX1. Quatre familles présentent une ambiguité de nom : LytB chez *S. pneumoniae* avec LytB de *B. subtilis* et DnaB chez *E. coli* avec DnaB chez *B. subtilis*. LytB et DnaB de *B. subtilis* ont été renommées LytBBS et DnaBBS. Les familles n'ayant aucune fonction assignée sont dénommées par PCDP pour « Putative Cell Division Protein ». Il est à noter que 16 protéines ne possèdent pas d'homologue chez *Firmicutes* et que 8 protéines possèdent des homologues mais aucun orthologue chez les *Firmicutes*. Pour les autres cas, nous avons reconstruit les familles/sous-familles de gènes chez les *Firmicutes* ce qui correspond à 179 familles/sous-familles. La distribution de la taille de familles et leur composition respective en familles/sous-familles monogéniques est présentée figure 3.15. Nous avons constaté une très grande diversité des cas. Comme expliqué précédemment, les familles d'homologues peuvent sur la base de la taxonomie être classées en familles monogéniques ($\approx$ une seule famille d'orthologues) et multigéniques ($\approx$ plusieurs familles d'orthologues). Nous allons dans la section suivante décrire les différents cas de figure.

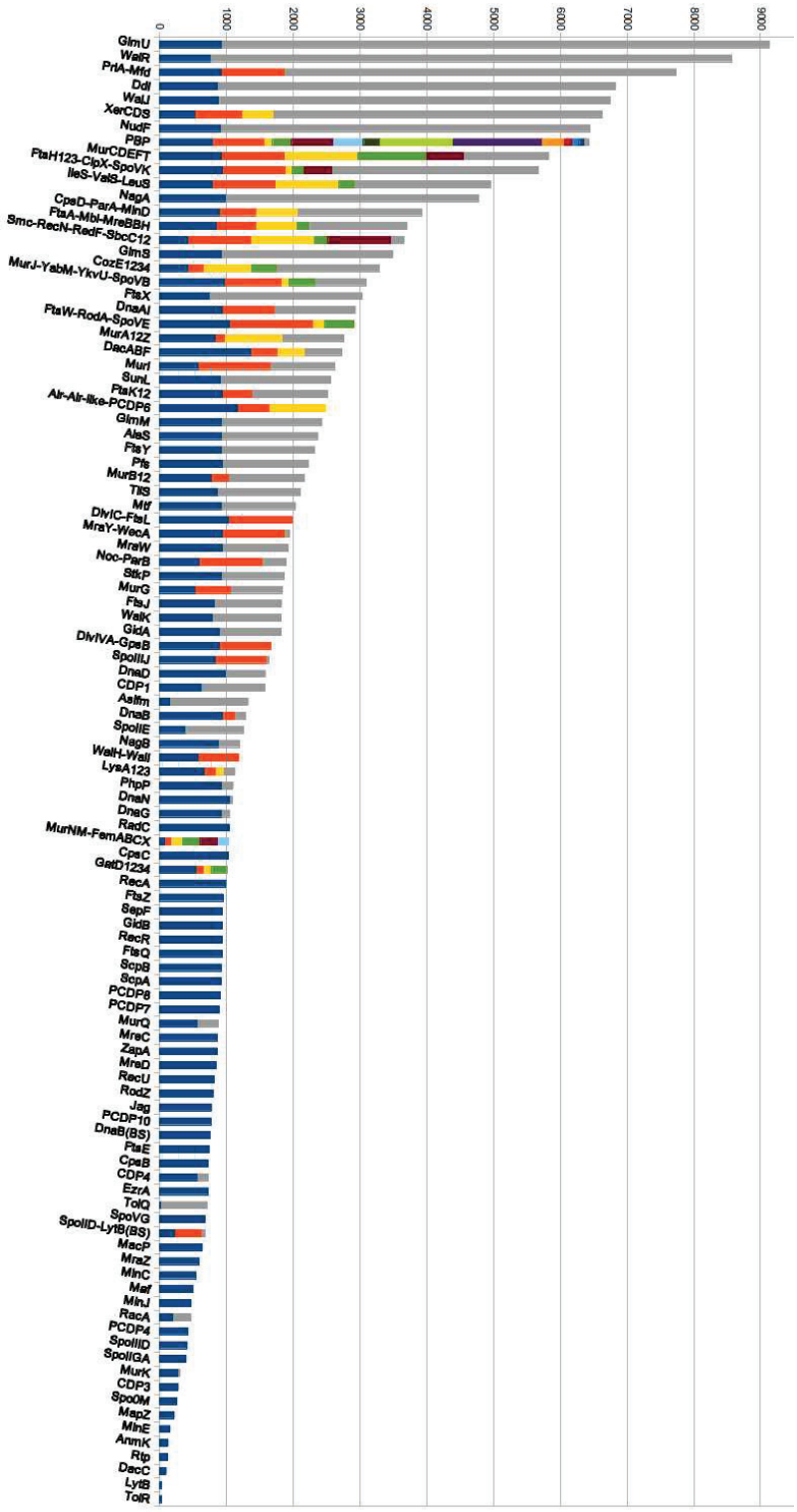


FIGURE 3.15 – Distribution de la taille des familles de gènes (en nombre de séquences). Les barres colorées correspondent à des familles/sous-familles analysées. les barres grises correspondent à des sous-familles non analysées.

3.3.2.2 Les familles monogéniques

On compte ainsi 39 familles monogéniques. Il s'agit du cas le plus simple ou la majorité des protéomes contiennent une seule copie de la famille d'homologues. Un des cas des plus représentatifs est la famille du gène *ftsZ*. Au moins une copie dans les 927 protéomes et seulement 17 copies surnuméraires issues de duplications et de transferts sont retrouvées. En comparant la phylogénie de la famille FtsZ avec la phylogénie des espèces de *Firmicutes*, la majorité des séquences semblent issues de transmission verticale à partir de l'ancêtre des *Firmicutes* bien que certains transferts horizontaux/duplications entre *Firmicutes* semblent avoir eu lieu (figure 3.16). Nous avons donc considéré que cette famille était monogénique. Les copies surnuméraires sont rares, présentent des grandes branches et sont branchées principalement avec les séquences de *Negativicutes*.

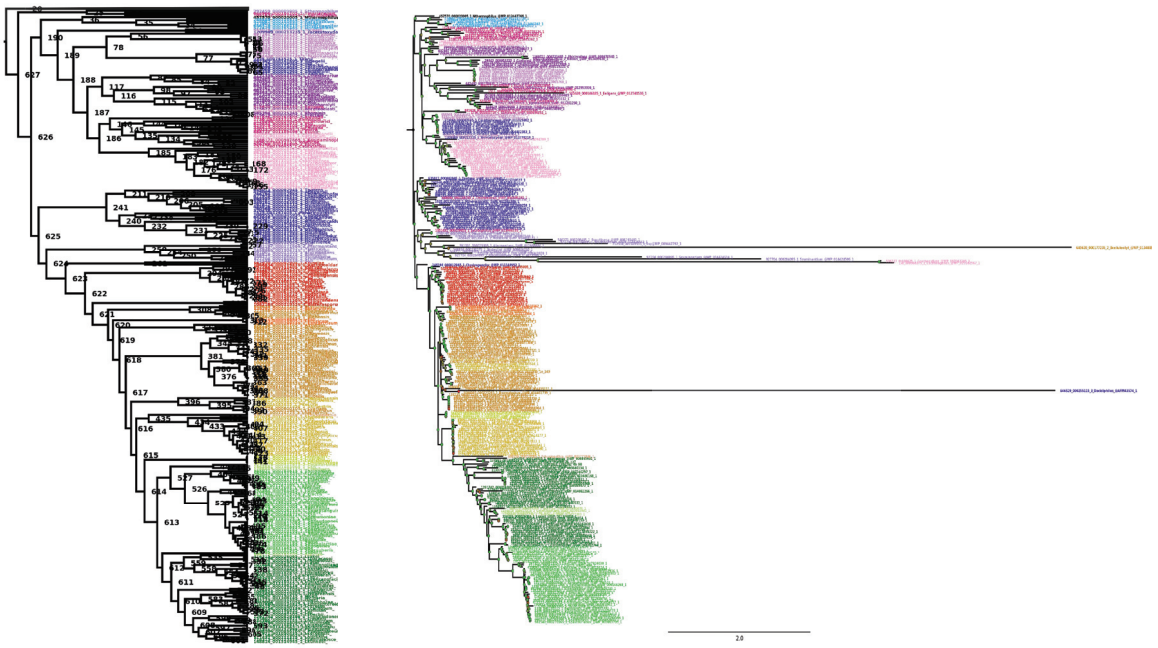


FIGURE 3.16 – Familles monogéniques : cas de la famille FtsZ. À gauche, la phylogénie des *Firmicutes*, à droite, la phylogénie de la famille FtsZ. Les couleurs correspondent aux groupes taxonomiques. Les topologies sont très similaires excepté quelques groupes.

3.3.2.3 Les familles multigéniques

66 familles multigéniques ont été identifiées, regroupant dans un certain nombre de cas plusieurs sous-familles impliquées dans le cycle cellulaire. Nous allons présenter ici différentes familles afin d'illustrer la diversité de complexité lors du découpage en sous-familles.

Les sous-familles très fiables Pour 90 sous-familles, la délimitation a été aisée principalement grâce à une distance phylogénétique élevée avec les autres sous-familles et une taxonomie cohérente (indice de fiabilité 2). La famille GidA illustre particulièrement bien ce cas. Cette sous-famille est distante du reste des autres sous-familles (0,43) et sa topologie est cohérente avec la arbre d'espèces (figure 3.17AB). De plus, le contexte génomique est très conservé et les séquences de cette sous-famille sont clairement identifiables dans l'alignement multiple (figure 3.17CD).

Les sous-familles modérément fiables 40 familles ont été plus difficile à délimiter de par des distances phylogénétiques faibles avec les autres sous-familles et une cohérence modérée avec la taxonomie mais une convergence des différentes informations (régions conservées spécifiques de la sous-famille au sein de l'alignement, contextes génomiques discriminants, taxonomie) ont permis d'être relativement confiant dans les délimitations proposées (indice de fiabilité 1). Deux difficultés majeures ont été rencontrées : un manque de signal phylogénétique ou une incohérence entre l'arbre d'espèces et l'arbre de la sous-famille provenant d'une histoire complexe.
La famille FtsL-DivIC illustre parfaitement le premier cas. La phylogénie est peu soutenue en raison du fait que peu de positions sont conservées au sein de l'alignement multiple (1985 séquences, 12 positions avec nettoyage et 659 sans nettoyage) (figure 3.18AB). Le manque de signal phylogénétique induit des mauvaise positions au sein de l'arbre. Néanmoins, le contexte génomique permet de discriminer les deux sous-familles (figure 3.18C).

Dans certains cas, le fait de construire des sous-arbres permet d'augmenter le signal phylogénétique. Par exemple, les familles Mbl, MreB et MreBH ne sont pas résolue dans l'arbre de

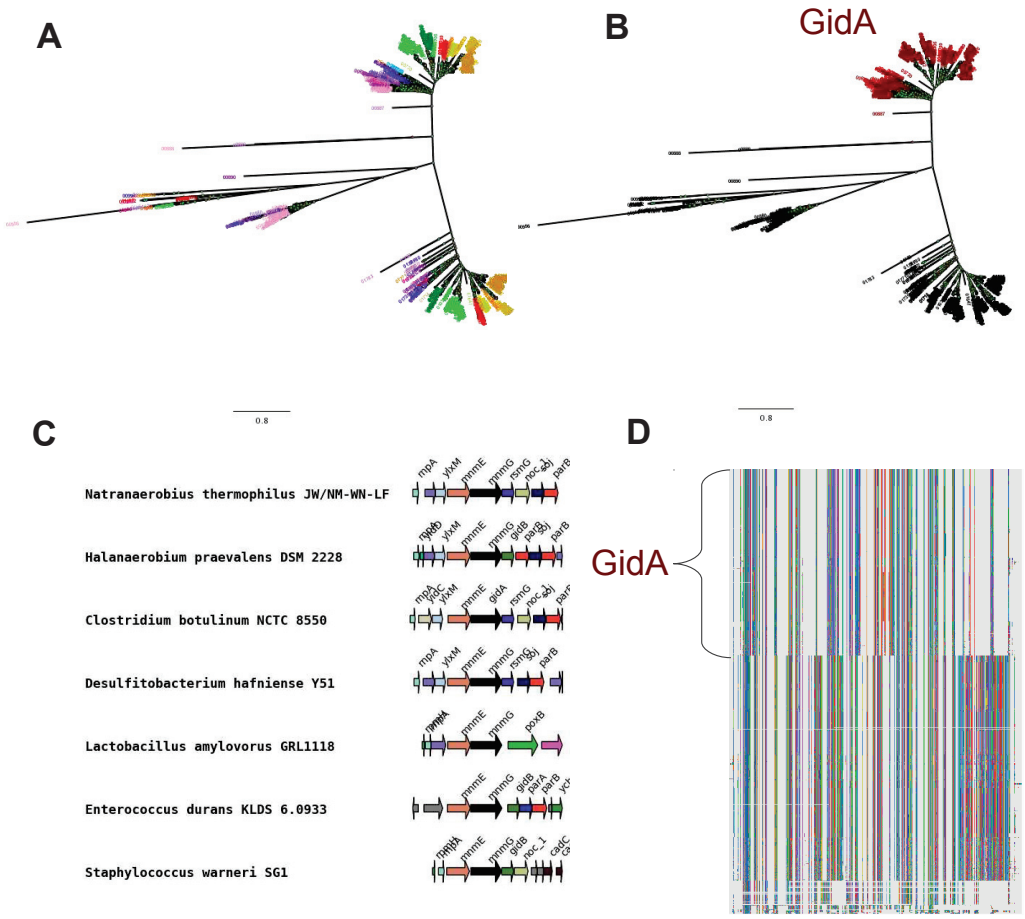


FIGURE 3.17 – Famille multigénique : le cas simple de la famille multigénique GidA. (A) Phylogénie de la famille GidA. Les couleurs représentent la taxonomie. (B) Phylogénie de la famille GidA. La famille GidA est montrée en rouge. (C) Contexte génomique très conservé de GidA (voisin de mnmE). (D) Alignement de la famille multigénique de GidA. La sous-famille de GidA est très distincte du reste de la famille.

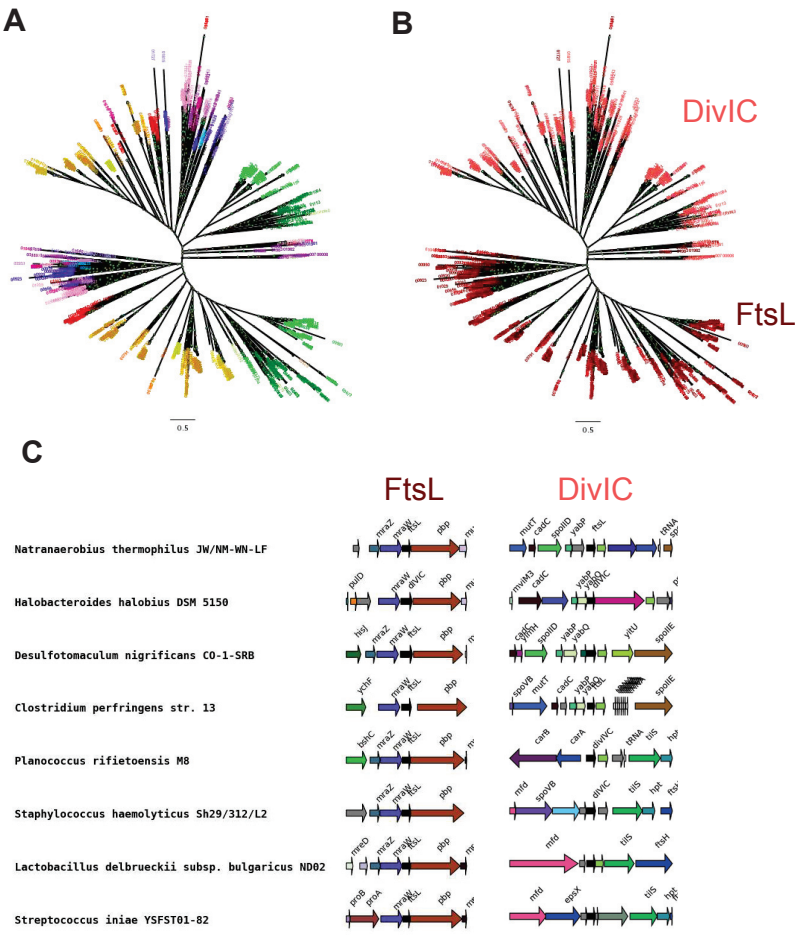


FIGURE 3.18 – Famille multigénique : le cas modérément complexe de la famille FtsL/DivIC. (A) Phylogénie de la famille FtsL/DivIC. Les couleurs représentent la taxonomie (B) Phylogénie de la famille FtsL/DivIC. Les couleurs montrent les deux sous-familles FtsL et DivIC. Quelques séquences sont mélangées au niveau des *Clostridia* (bleu/violet). (C) Contexte génomique conservé de FtsL et DivIC.

la famille FtsA-Mbl-MreB-MreBH-DnaK mais le fait d'exclure FtsA et DnaK permet d'inférer un sous-arbre dans lequel les trois familles sont clairement distinctes (figure 3.19).

Le deuxième cas correspond à des sous-familles dont la phylogénie ne correspond peu ou pas à celle des espèces. Dans certains cas, un groupe basal qui présente une incohérence avec la taxonomie est observé. Un des cas les plus explicites est celui de la famille GlmU ou un groupe

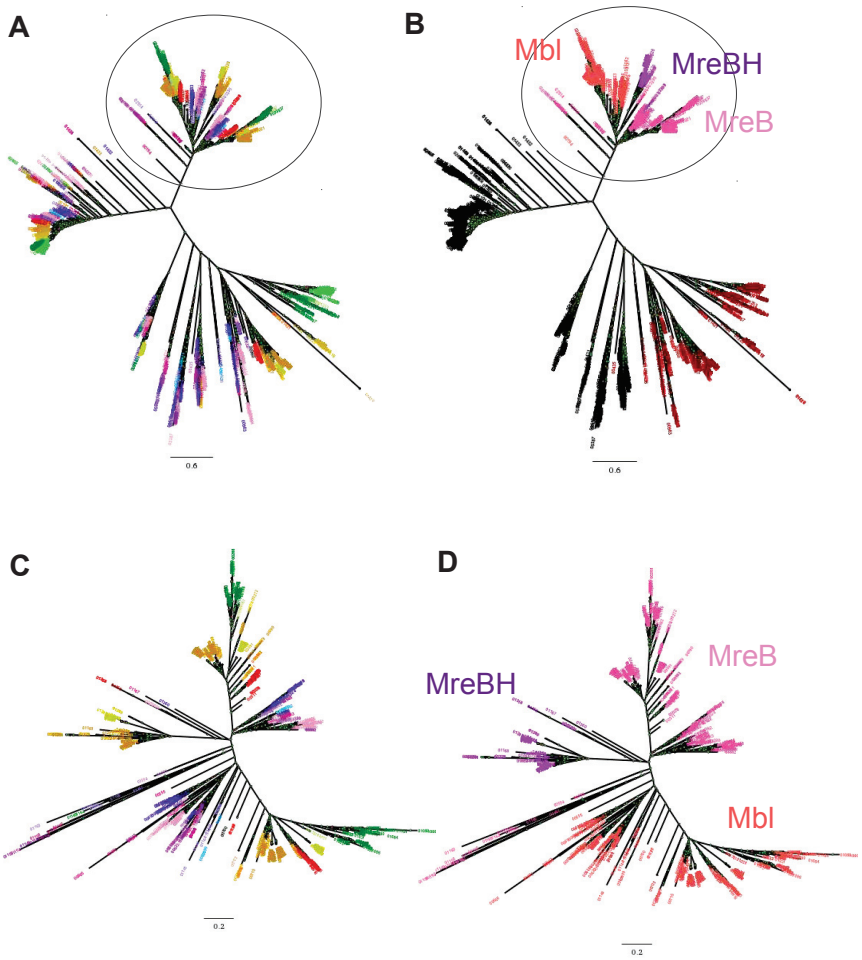


FIGURE 3.19 – Famille multigénique : le cas modérément complexe de la famille FtsA/MreB/MreBH/Mbl. (A) Phylogénie de la famille FtsA/MreB/MreBH/Mbl coloré par la taxonomie. (B) Phylogénie de la famille FtsA/MreB/MreBH/Mbl coloré par sous-familles. (C) Sous-arbre de la partie entourée de l'abre A et B correspondant à MreB/MreBH/Mbl, coloré par la taxonomie. (D) Sous-arbre de la partie entourée de l'abre A et B correspondant à MreB/MreBH/Mbl, coloré par les sous-groupes. Le sous-arbre améliore la résolution entre les sous-familles.

basal constitué de *Bacilli* et de *Clostridia* sans cohérence taxonomique et redondant avec le reste de la famille est retrouvé (figure 3.20). Dans ce cas, la convergence des informations telles que le contexte génomique ou le positionnement au sein de la phylogénie au niveau des bactéries ont permis de guider le découpage de la sous-famille.

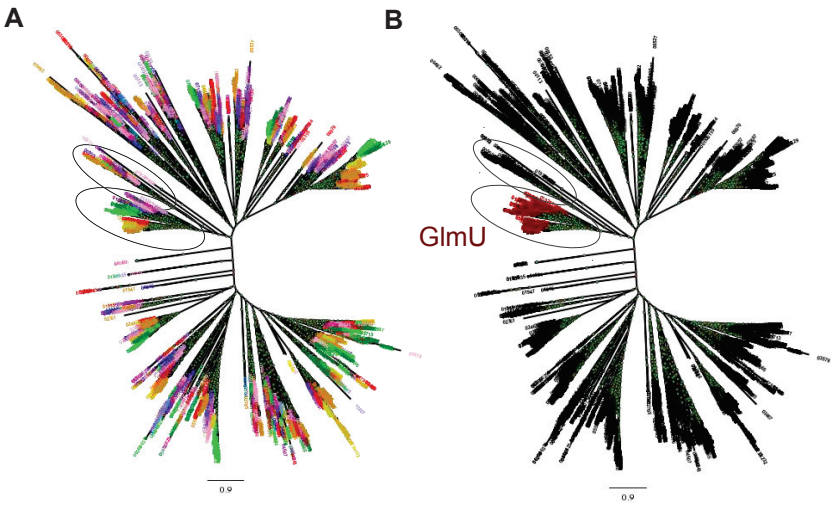


FIGURE 3.20 – Famille multigénique : le cas très complexe de la famille GlmU. (A) Phylogénie de la famille GlmU, coloré par la taxonomie. (B) Phylogénie de la famille GlmU, la sous-famille GlmU est colorée en rouge et entourée. Le deuxième groupe entouré correspond au groupe basal dont la taxonomie est redondante avec le groupe rouge et peu cohérente avec l'arbre d'espèces.

Les sous-familles peu fiables 9 sous-familles correspondent à des situations plus complexes liées à la non conservation des contextes génomiques, à des faibles distances phylogénétiques avec les autres sous-familles, des incohérences avec l'arbre d'espèces et des phylogénies peu soutenues (indice de fiabilité 0). La famille des PBPs de type A illustre ce cas. La phylogénie dans cette famille est très peu résolue et les sous-familles se recoupent (figure 3.21). Également, le contexte génétique n'est pas du tout conservé pour la plupart des groupes. Il a donc été nécessaire de faire un choix en prenant en compte le peu d'informations disponibles afin de construire les sous-familles.

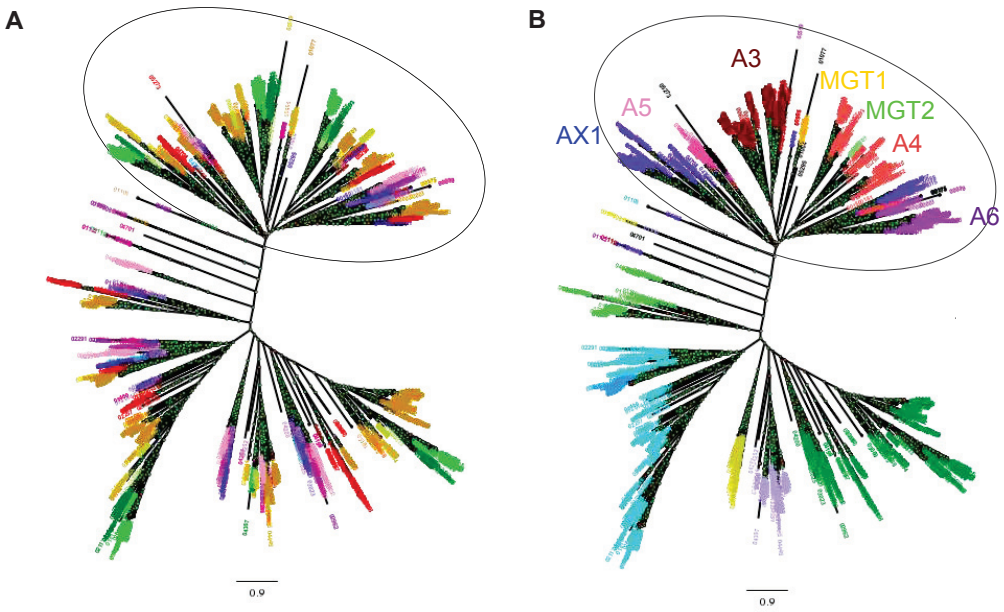


FIGURE 3.21 – Famille multigénique : le cas très complexe de la famille des PBPs type A. (A) Phylogénie des PBPs, coloré par la taxonomie. Le groupe de type A est entouré. (B) Phylogénie des PBPs, coloré par les sous-familles. Celles-ci sont mélangées car la phylogénie n'est que peu résolue. L'indice de confiance pour ces familles est très bas.

3.3.2.4 Conclusion sur la méthode

A travers ces quelques exemples, il est assez aisé de se rendre compte de la diversité et la complexité des cas rencontrés pour chaque famille/sous-famille. Les cas les plus complexes ont été ceux des familles multigéniques. Pour découper ces familles en sous-familles, nous nous sommes appuyés sur diverses informations phylogénétiques et génomiques. Nous avons également défini un indice de fiabilité. Néanmoins, nous avons décelé plusieurs limites à cette méthodologie.

Tout d'abord, dans certains cas, aucune cohérence entre les informations analysées n'a pu être mise en évidence (indice 0, table .6). Ensuite, certaines informations ne sont pas discriminantes pour la plupart des cas. C'est le cas par exemple de la composition en domaines. Enfin, les indices de fiabilité que nous avons défini ne sont que qualitatifs. Il aurait été préférable de définir des métriques associées à chaque critère utilisé.

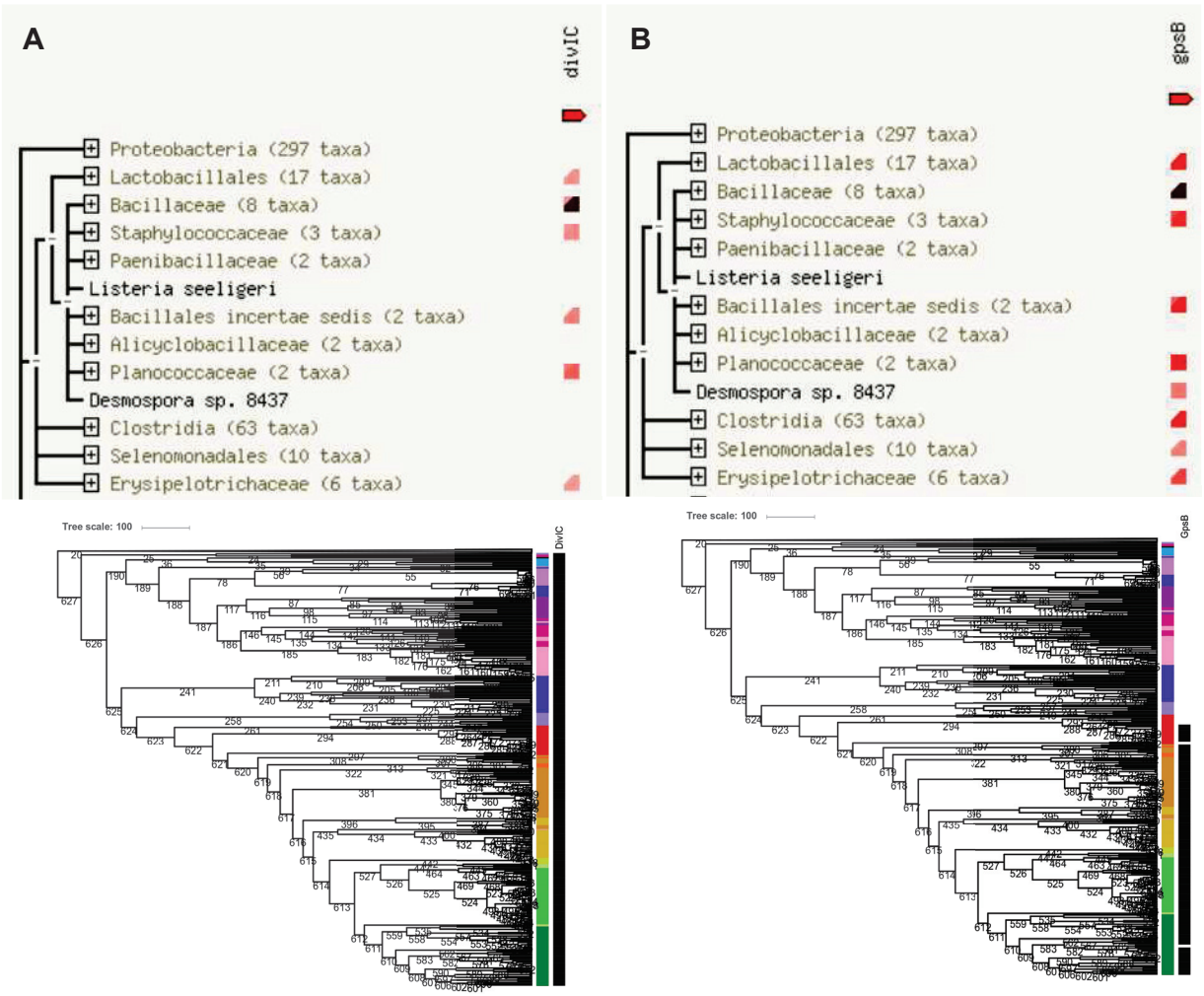


FIGURE 3.22 – Comparaison de deux familles construites par les COGs et dans cette étude. (A) Distribution taxonomique de la famille DivIC par la méthode des COGs (STRING) et par notre étude. La distribution taxonomique de la famille COG DivIC est très éparse chez les *Firmicutes* alors que nous retrouvons une copie de DivIC chez tous les *Firmicutes*. (B) Distribution taxonomique de la famille GpsB par la méthode des COGs (STRING, https://string-db.org/) et par notre étude. La majorité des *Firmicutes* possèdent la famille COG GpsB alors que nous ne retrouvons GpsB que chez les *Bacilli*.

Malgré cela, il est à noter que les résultats présentés par la suite et notamment les analyses corrélatives ne sont que peu impactés par les incertitudes existant quant au découpage de certaines sous-familles. En effet, le nombre de cas où le découpage est ambigu est relativement faible. De plus, les incertitudes ne concernent en général qu'un nombre restreint de séquences.

Il est important de souligner que la qualité des jeux de données semble bien meilleure que celle des COGs. En effet, dans de nombreux cas, les familles inférées par les COG semblent ne pas être correctement assemblées. Par exemple, la famille DivIC est séparées en plusieurs familles chez les *Firmicutes* et la sous-famille DivIVA n'est pas séparée de la sous-famille GpsB (figure 3.22).
Il est important de noter qu'aucun outil automatique ne permet à ce jour de délimiter précisément les familles en raison de la grande complexité et diversité des cas rencontrés. Ainsi, seule l'approche manuelle semble adaptée pour la délimitation des familles bien que cette opération soit fastidieuse et chronophage. Du fait que les solutions soient définie manuellement, certaines solutions proposées peuvent être subjectives et ne correspondent pas forcément à la réalité. Néanmoins, le découpage des familles a été discuté au sein de l'équipe afin de trouver les solutions optimales.

3.4 Événements évolutifs majeurs des familles de gènes du cycle cellulaire chez les *Firmicutes*

À partir des jeux de données de familles des gènes, il est possible de reconstruire l'histoire évolutive des gènes et d'identifier les événements évolutifs majeurs ayant affecté ces familles de gènes. L'histoire évolutive peut se décliner en deux types : l'histoire individuelle de chaque gène et l'histoire organisationnelle des gènes au sein des génomes. Deux approches ont été utilisées afin de reconstruire l'histoire individuelle des familles de gènes : La reconstruction des états ancestraux par les profils phylogénétiques et la réconciliation. Concernant l'histoire organisationnelle des familles de gènes, nous avons utilisé la technique de reconstruction des synténies ancestrales. Nous avons à partir de ces trois méthodes identifié plusieurs points chauds de l'histoire évolutive des gènes du cycle cellulaire chez les *Firmicutes*. Nous avons également tenté de comparer les données obtenues à partir de ces trois méthodes.

3.4.1 Reconstruction de l'histoire évolutive des familles de gènes

3.4.1.1 Concordance de souches entre arbre d'espèces et familles de gènes

Les jeux de données initiaux de séquences de familles de gènes ont été construits à partir de la base de données des 937 *Firmicutes*. Afin de reconstruire l'histoire évolutive des familles de gènes chez les *Firmicutes*, nous avons utilisé l'arbre d'espèces présenté section 3.2. Néanmoins, cet arbre d'espèces a été échantillonné avec une souche par espèces. Nous avons donc échantillonné les jeux de données de familles de gènes de la même façon que l'arbre d'espèces puis réaligné les séquences correspondantes avec MAFFT (option linsi) [238] (annexe .2). Il est important de noter que les séquences provenant de *Coprothermobacter proteolyticus* et *Thermodesulfobium narugense* n'ont pas été inclues dans les jeux de données car bien qu'étant annotés comme appartenant aux *Firmicutes*, ces deux espèces ne se groupent pas avec le reste des *Firmicutes* au sein de l'arbre d'espèces.

3.4.1.2 Inférence des événements de gènes par réconciliation

Inférence phylogénétique Les alignements ont été nettoyés à l'aide de BMGE (BLOSUM30). Pour 17 familles, le nombre de positions en acides aminés était inférieur à 100 (annexe .9). Pour l'inférence des phylogénies correspondant à chaque famille, le modèle d'évolution a été sélectionné parmi les modèles WAG, LG, VT et JTT d'après le BIC (« Bayesian Information Criterion ») en utilisant ModelFinder [234] inclu dans IQ-TREE [348]. Nous avons en effet restreint le nombre de modèles testés (18 par défaut) afin de minimiser le temps de calcul et car les quatre modèles sélectionnés sont ceux qui empiriquement sont le plus souvent sélectionnés. Nous avons tout d'abord inféré les phylogénies en utilisant IQ-TREE. Néanmoins, le caractère non-déterministe de l'exploration des topologies implémenté dans IQ-TREE génère des topologies différentes à chaque inférence pour un même jeu de données si l'alignement ne possède que peu de sites informatifs. La génération de plusieurs topologies pour un même arbre peut se révéler problématique dans l'exploitation de ces arbres.
Pour savoir si la quantité d'information phylogénétique au sein des alignements des familles

de gènes étudiées était suffisante pour que l'utilisation d'IQ-TREE génère une seule topologie pour différents lancements, nous avons inféré trois fois de façon indépendante les phylogénies de 145 familles de gènes. Pour chaque famille, nous avons calculé la distance de Robinson-Foulds entre les trois topologies générées. Seulement 11 des 145 familles de gènes ne présentaient aucune différence pour les trois topologies inférées. Les familles de gènes que nous avons étudié semblent ne pas posséder suffisamment de positions informatives puisque les topologies inférées par IQ-TREE ont fortement divergé. De façon intéressante, PhyML et RAxML semblent générer des topologies plus similaires entre elles que entre les topologies générées par IQ-TREE (exemple présenté en figure 3.23).

Ces résultats nous ont conduit a utiliser un programme d'inférence phylogénétique basé sur un algorithme déterministe afin d'obtenir une seule et même topologie, les inconsistances se traduisant non pas par des topologies différentes mais par des supports de branches faibles. Nous avons donc utilisé RaxML 0.5.0b qui utilise un algorithme d'exploration des topologies pseudo-déterministe [439]. RAxML possède l'avantage de fournir des résultats robustes comparables à ceux générés par PhyML et semble plus rapide que PhyML [441].
Les phylogénies ont été inférées avec les modèles sélectionné par IQ-TREE comme précédemment expliqué. L'ensemble des modèles choisis sont présentés en annexes .7, .12. 100 réplicats de bootstraps non paramétriques ont également été générés. Les supports de bootstraps ont ensuite été corrigés par l'approche des boostraps de transfert [267] particulièrement adaptés à des alignements avec un grand nombre de séquences (plusieurs centaines) et un signal phylogénétique modéré. Les arbres ont été visualisés grâce à Figtree 1.4.3 et iTOL (`http://tree.bio.ed.ac.uk/software/figtree/`, `https://itol.embl.de/`, [271]).

Réconciliation Les réconciliations ont été réalisées à l'aide de ecceTERA 1.2.4 [420]. Une approche d'amalgamation a été utilisée afin de tenir compte au mieux des bipartitions non soutenues et des alternatives de scénarios probables. Ainsi, les arbres issus des réplicats de bootstraps ont été utilisés en tant que distribution d'arbres de gènes. L'arbre d'espèces daté et corrigé des *Firmicutes* a été utilisé. Les coûts d'événements ont été fixés initialement à 1

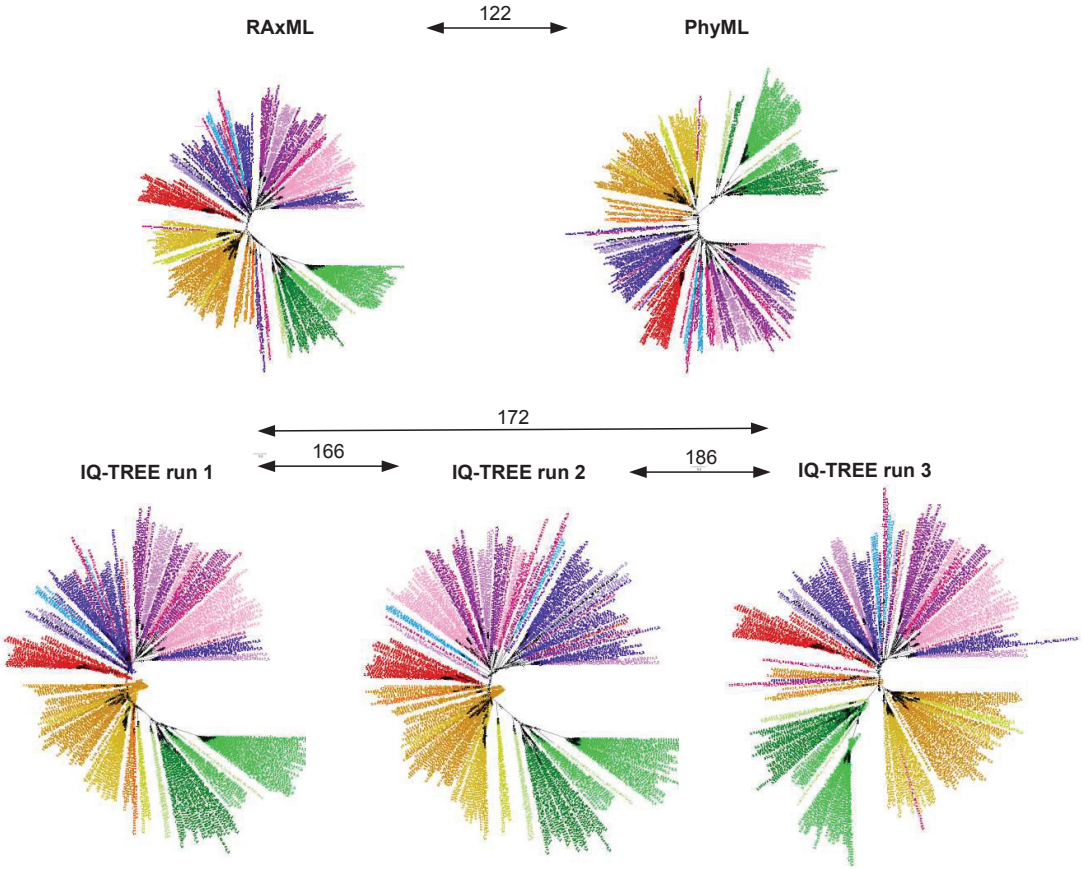


FIGURE 3.23 – Différences de topologies en fonction des programmes d'inférence, le cas de la famille ZapA. Les deux topologies du haut inférées par RAxML et PhyML présentent une distance de Robinson-Foulds de 122 tandis que les trois topologies inférées par IQ-TREE par le même jeu de données présentent entre elles des distances plus élevées (166, 172, 186). 275 séquences, 57 positions d'acides aminés, LG+G4 pour RAxML, PhyML, IQ-TREE, WAG+G4 pour Fasttree.

pour les pertes, 2 pour les duplications et 3 pour les transferts horizontaux (coûts par défaut). Le poids d'amalgamation a d'abord été fixé à zéro (par défaut). Une itération de l'algorithme d'estimation des coûts implémenté dans ecceTERA a été effectuée pour estimer les coûts spécifiques de chaque événement et le poids d'amalgamation pour chaque famille. Les transferts horizontaux depuis les lignées éteintes/non échantillonnées ont été autorisés. Les fichiers de sortie de réconciliation ont ensuite été convertis en RecPhyloXML, un format de réconciliation

en XML particulièrement facile a exploiter par des scripts [107].

3.4.1.3 Reconstruction des états ancestraux par les profils phylogénétiques

Les profils phylogénétiques ont été construits à partir des identifiants de séquences de chaque famille protéique. Si au moins une copie de la famille est présente dans une souche, la valeur attribuée à la souche est 1, sinon 0. Les états ancestraux issus des profils phylogénétiques ont été inférés par parcimonie à l'aide d'un programme développé pendant ma thèse utilisant la librairie python ete3 [208]. Plus précisément, la présence ou l'absence a été projeté sur les feuilles de l'arbre d'espèces des *Firmicutes* (figure 3.24). Nous aurions pu utiliser des programmes tels que Mesquite [296] mais nous avons voulu maîtriser au mieux les paramètres et les sorties.
L'arbre est d'abord parcouru des feuilles jusqu'à la racine afin de déterminer les coûts associés à chaque état (présence/absence). Pour le cas des feuilles, l'état est déjà déterminé. Pour les branches internes, il est nécessaire d'estimer le coût des états en fonction des branches/feuilles filles. Soit 0 ou 1 l'état de présence ou d'absence à chaque branche b et 1-1, 0-1, 1-0, 0-0 les états de présence ou d'absence dans l'ordre des deux branches filles $f1$ et $f2$. Pour chaque branche b, les coûts des 8 scénarios associés sont alors calculés (figure 3.24). Le coût de chaque scénario correspond à la somme de deux types de coûts : le premier correspond au coût de l'état supposé aux branches filles sachant l'état réel ou inféré préalablement et le deuxième au coût de transition entre l'état de la branche b à la branche fille f (c'est-à-dire le coût d'une perte ou d'un gain). Pour chaque état 1 et 0 de la branche b, le scénario dont le coût est le moins élevé est sélectionné comme étant la solution la plus parcimonieuse (figure 3.24). Dans le cas de la racine, le coût d'un gain supplémentaire est ajouté à tous les coûts pour l'état 1. Chaque nœud présente ainsi deux solutions possibles (1 ou 0), chacun associées à un scénario pour les deux branches filles. L'arbre est ensuite parcouru de la racine jusqu'aux feuilles. Pour la racine, la solution la moins coûteuse est sélectionnées. Le scénario associé à cet état défini ensuite l'état aux deux deux branches filles. Ce processus itératif est répété jusqu'aux feuilles. En cas d'égalité entre les deux états, la présence est privilégiée de façon arbitraire.
Concernant les coûts d'événements (gain et perte), la littérature décrit principalement deux méthodes : L'algorithme de Fitch qui établi un coût égal pour les gains et les pertes [278]

ou l'algorithme de Dollo qui n'autorise qu'un seul gain [21]. Nous avons décidé utiliser une solution intermédiaire, c'est-à-dire que nous avons maximisé le coût des gains mais malgré tout autorisé plusieurs gains. Pour choisir la valeurs des coûts nous avons testé plusieurs possibilités sur quelques familles et nous avons sélectionné les coûts qui présentaient les résultats les plus cohérents de façon qualitative. Les coûts ont été fixés à 2 pour une perte de gène et 6 pour une apparition. Les coûts aux feuilles ont été attribués en fonction de l'état binaire de la synténie : Pour une présence : coût d'une présence : 0, coût d'une absence : 1 000 000 000. Pour une absence : coût d'une absence : 0, coût d'une présence : 1 000 000 000.

3.4.1.4 Reconstruction des synténies ancestrales

Les synténies ancestrales ont été inférées par une approche de parcimonie, tout comme les états de présence/absence à partir des profils phylogénétiques. Néanmoins, la génération de la matrice binaire de synténies a nécessité des traitements supplémentaires. L'ensemble des étapes de l'inférence des synténies ancestrales est présenté figure 3.25.
Tout d'abord, une matrice binaire de voisinage entre chaque famille protéique pour chaque espèce des 304 *Firmicutes* a été construite. Seuls les couples de gènes dont la distance entre les extrémités les plus proximales des deux gènes était inférieure ou égale à 2 000 paires de bases à partir ont été considérés comme voisins (les distances sont issues des fichiers GFM). La matrice a ensuite été corrigée par une approche de regroupement par lien unique (« Single Linkage Clustering », SLC) en utilisant la librairie NetworkX [418]. Ainsi, Les voisins par gènes intermédiaires ont donc été considérés comme vicinaux. La matrice binaire de synténie a ensuite été utilisée comme pour l'inférence des états de présence/absence expliquée précédemment.
Pour choisir la valeurs des coûts (gain et perte) nous avons testé plusieurs possibilités sur quelques familles et nous avons sélectionné les coûts qui présentaient les résultats les plus cohérents de façon qualitative. Les mêmes coûts que pour la reconstruction des états de présence/absence ont été choisis, c'est-à-dire 2 pour une perte de synténie et 6 pour un gain. Les coûts aux feuilles ont été attribués en fonction de l'état binaire de la synténie : Pour une présence : coût d'une présence : 0, coût d'une absence : 1 000 000 000. Pour une absence : coût d'une absence : 0, coût d'une présence : 1 000 000 000.

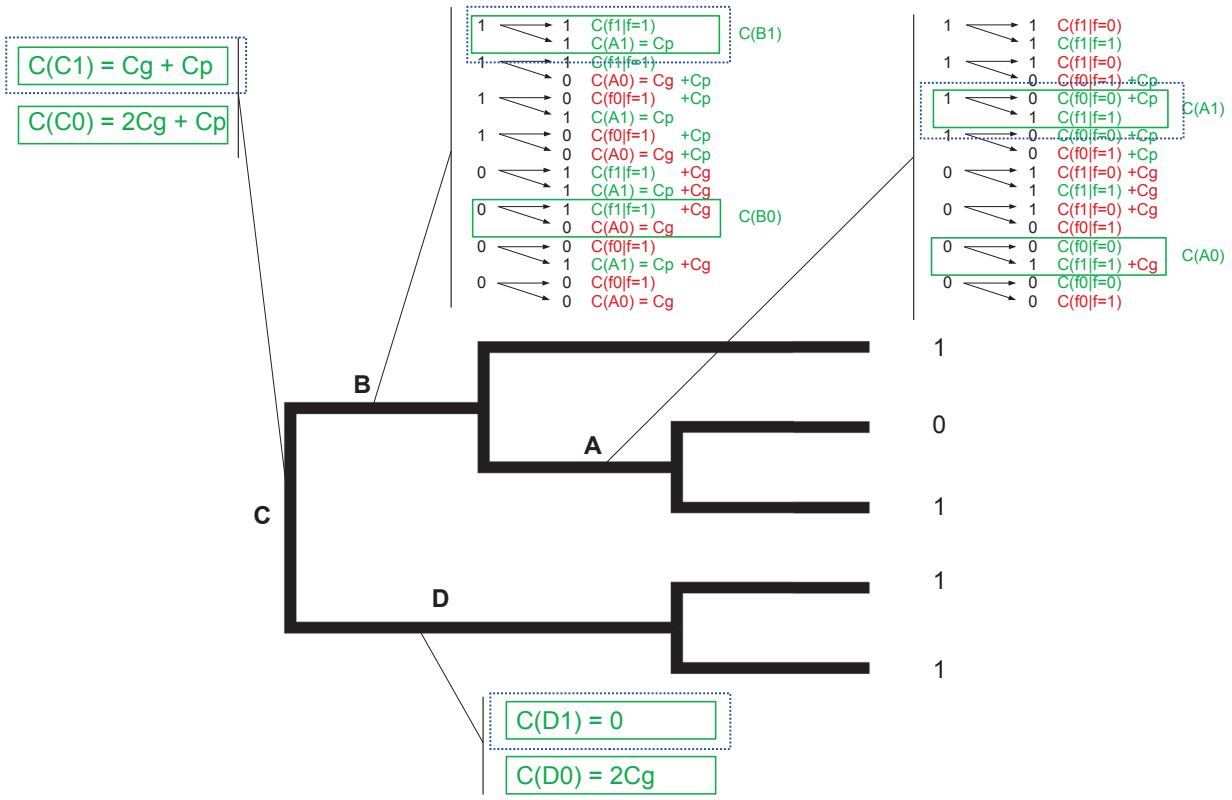


FIGURE 3.24 – Algorithme de reconstruction d'états ancestraux par parcimonie. Les états initiaux sont binaires (0 ou 1). À la branche A, tous les scénarios sont envisagés et les coûts relatifs sont calculés. Deux types de coûts existent : le premier correspond au coût de l'état supposé aux branches filles ($C(f|f)$, $C(A1)$, *etc...*) et le deuxième au coût de transition entre l'état de la branche A à la branche fille (Cg ou Cp). Les coûts indiqués en vert correspondent au moins élevés et ceux indiqués en rouge aux plus élevés. Le coût minimal pour chaque solution à cette branche (0 et 1) est choisi (rectangle vert). Pour les autres branches, la même procédure est appliquée. Un coût supplémentaire Cg est ajouté pour l'état 1 de la racine. À partir de l'état le moins coûteux à la racine et du scénario associé, les états sont inférés à toutes les branches (rectangles bleus). Cg : coût d'un gain, Cp : coût d'une perte, $C(fe|fE)$: coût d'un état e d'une feuille f sachant l'état E de la feuille f, $C(Xe)$: coût d'un état e pour la branche X, inféré préalablement.

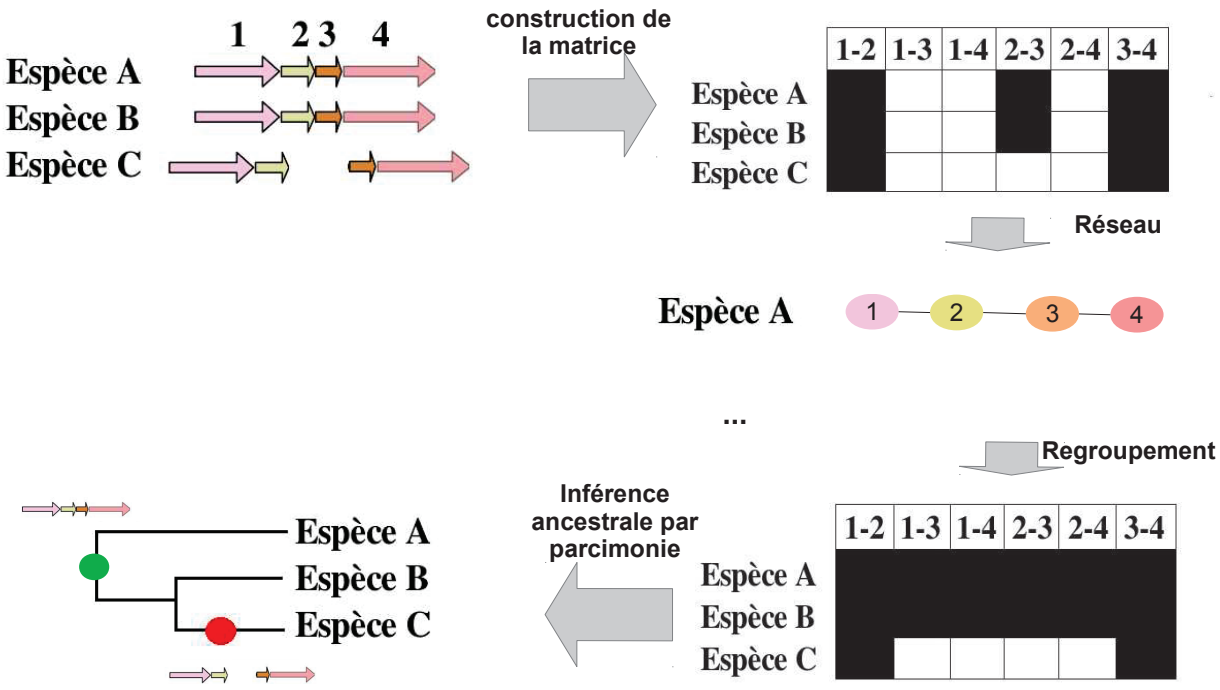


FIGURE 3.25 – Principe de la reconstruction des synténies ancestrales. Les gènes 1, 2, 3 et 4 sont présents chez les espèces A, B et C. 1, 2, 3 et 4 forment un cluster chez A et B. le cluster est dissocié chez C. Tout d'abord, la matrice de voisinage est construite pour tous les couples de gènes. La matrice est ensuite corrigée par SLC. Enfin, la matrice permet d'inférer les synténies ancestrales aux branches internes de l'arbre d'espèces. Les événements de clusters peuvent ensuite être inférés (rond vert : gain, rond rouge : perte).

Les synténies ancestrales pour chaque couple ont ensuite été utilisées pour construire les clusters ancestraux. Pour chaque branche, les gènes ont été regroupés par SLC et chaque groupement constitue un cluster ancestral de gènes. Les événements de gains, modifications ou pertes de clusters de gènes ont ensuite été inférés à partir de ces résultats. Nous avons défini le gain de cluster par l'apparition d'un cluster à une branche alors que celui-ci était absent à la branche mère. La modification d'un cluster correspond à l'événement d'éloignement ou de rapprochement d'un ou plusieurs gènes au cluster à la condition qu'il reste au moins deux gènes dans le cluster après la modification. Enfin, la perte correspond à la dislocation totale du cluster de gènes, c'est-à-dire qu'aucun des gènes n'est voisin après l'événement.

3.4.2 Événements évolutifs majeurs du cycle cellulaire chez les *Firmicutes*

3.4.2.1 Inférence des apparitions de gènes

L'apparition des familles de gènes est une information capitale en phylogénomique [97]. Cela constitue la première étape dans la reconstruction de l'histoire évolutive des familles de gènes. Cette information permet aussi d'identifier des événements de multiples d'apparitions de gènes (points chauds). L'apparition d'un gène peut provenir d'une duplication d'un gène préexistant, d'un transfert horizontal de gène d'un autre clade, d'un réarrangement de domaines fonctionnels ou de façon très rare, *de novo* [8]. Ainsi, nous avons inféré l'apparition ancestrale de chaque famille de gènes du cycle cellulaire chez les *Firmicutes*. Pour cela, nous avons utilisé deux approches, les profils phylogénétiques et la réconciliation puis nous avons comparé les deux méthodes.

Inférence des apparitions de gènes par profil phylogénétiques Pour chaque famille, nous avons tout d'abord inféré la présence/absence des représentants de la famille à chaque nœud de la phylogénie des *Firmicutes* à partir des profils de présence/absence aux feuilles de l'arbre d'espèces. Les événements de gains ont ensuite été projetés sur l'arbre d'espèces. Néanmoins, cette technique peut conduire à l'inférence de plusieurs apparitions pour une même famille. L'arbre d'espèces étant daté, nous pouvons dater relativement les branches. Ainsi, pour chaque famille de gènes, la branche comportant un événement d'apparition de gènes la moins distantes de la racine a été considérée comme étant la branche d'apparition. Les branches présentant le plus d'apparitions sont représentés aux nœuds associés en figure 3.26.
La distribution du nombre d'apparitions par branche montre que la majorité des événements sont concentrés en quelques branches (figure 3.26B). Ainsi, trois « points chauds » sont retrouvés tandis que la majorité des branches ne sont associées à aucune apparition (figure 3.26C). Le premier point chaud correspond à la racine des *Firmicutes* et comporte 117 apparitions. Cela suggère que la majorité des gènes du cycle cellulaire étudiés étaient présents chez l'ancêtre des

A

Nœud		Nombre d'acquisitions
Racine *Firmicutes*	1	117
Émergence *Clostridia sauf Clostridia groupe 0*	2	13
Émergence *Paenibacillaceae*/Autres *Bacilli*	3	11
Émergence *Bacillaceae*/Autres *Bacilli*	4	4
Émergence *Brevibacillus*/Autres *Bacilli*	5	4
Émergence *Staphylococcaceae*	6	3
Émergence *Bacilli*	7	3
Pelosinus fermentans JBW45		2
Émergence *Clostridia Groupe 1D*		2
Émergence *Clostridia Groupe 1C/1D*		2
Émergence *Negativicutes*		2

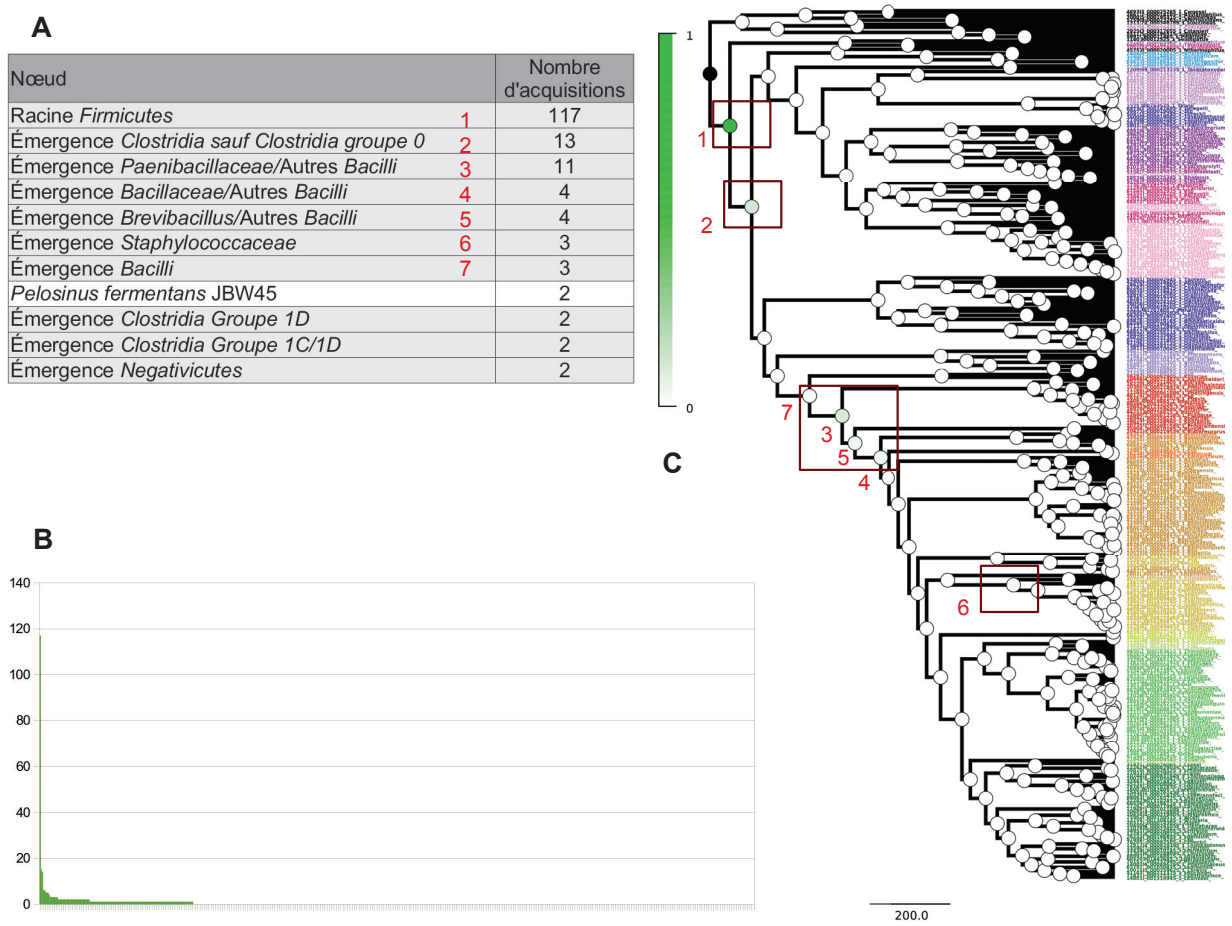


FIGURE 3.26 – Apparitions inférées par parcimonie. (A) Branches (ou nœuds) présentant deux apparitions de gènes ou plus. Les lignes blanches correspondent à des feuilles (souches), les lignes en gris à des branches (ou nœuds) internes. Les branches/feuilles présentant le plus d'événements sont numérotées et indiquées sur le panneau (C). (B) Distribution du nombre d'apparitions par branche. (C) Projection du nombre d'apparitions sur la phylogénie des *Firmicutes*. Les nombres en rouge ainsi que les encadrés correspondent à la table (A).

Firmicutes. Le deuxième point chaud comporte 13 apparitions et correspond à la deuxième branche fille de la racine, soit l'émergence des *Clostridia* excepté les *Clostridia* groupe 0. Ces apparitions sont donc très anciennes. Il est à noter que le groupe 0 des *Clostridia* est composé de trois espèces seulement. Une absence dans les trois espèces de ce groupe exclu systématiquement la possibilité d'une apparition à la racine des *Firmicutes*. Ce groupe est peut-être mal placé puisque la phylogénie des *Firmicutes* inférée par Antunes et collègues propose une émergence plus tardive de ce groupe (figure 3.10). Si cette hypothèse est exacte, alors il est tentant de faire l'hypothèse que les gènes inférés comme acquis chez les *Firmicutes* à cette branche ont été en fait acquis à la racine des *Firmicutes*.
Le troisième point chaud correspond à l'émergence des *Paenibacillaceae* et des autres *Bacilli*. Plus précisément, 11 gènes sont inférés comme étant apparus à cette branche. De plus, les branches voisines de cette branche correspondant à l'émergence des *Bacilli*, des *Bacillaceae*/autres *Bacilli* et des *Brevibacillus*/autres *Bacilli* comportent respectivement 3, 4 et 4 apparitions. L'émergence et les étapes précoces de la diversification des *Bacilli* semblent donc accompagnées de nombreuses apparitions de gènes impliqués dans le cycle cellulaire. Il est tentant de faire également l'hypothèse que ces gènes ont été acquis au niveau de la même branche mais que des pertes dans les groupes basaux des *Bacilli* soient responsable de la dispersion apparente des événements d'apparition sur l'ensemble des branches basales de ce taxon.

Inférence des apparitions de gènes par réconciliation À partir des réconciliations calculées par ecceTERA, nous avons également projeté les événements d'apparitions sur l'arbre d'espèces des *Firmicutes* (figure 3.27). Les trois mêmes points chauds identifiés par la technique des profils phylogénétiques sont observés (figure 3.27C). Également, la distribution du nombre d'apparitions par branche est similaire à celle générée précédemment (figure 3.27BC).

En effet, la racine des *Firmicutes* et la seconde branche fille de la racine présentent le plus grand nombre d'apparitions. Cependant, la branche qui correspond à l'émergence des *Clostridia* excepté les *Clostridia* groupe 0 se situe en tête avec 62 apparitions de gènes contre 53 à la racine des *Firmicutes*. Cela s'explique par le fait que la présence de nombreux gènes chez les espèces des *Clostridia* groupe 0 sont interprétées comme le fruit de transferts horizontaux

A

Nœud		Nombre d'acquisitions ancestrales
Émergence *Clostridia* sauf *Clostridia groupe 0*	1	62
Racine *Firmicutes*	2	52
Émergence *Paenibacillaceae*/Autres *Bacilli*	3	12
Émergence *Bacillaceae*/Autres *Bacilli*	4	6
Émergence *Brevibacillus*/Autres *Bacilli*	5	5
Émergence *Negativicutes*	6	4
Émergence *Staphylococcaceae*	7	3
Émergence *Bacilli*	8	3
Émergence *Exiguobacterium*/Autres *Bacilli*		3
Émergence *Clostridia* groupe 1		3
Émergence *Bacillus/Geobacillus*		2
Émergence *Lactobacilllales*		2
Émergence *Clostridia* groupe 1D		2
Émergence *Amphibacillus/Anoxybacillus*/Autres *Bacilli*		2

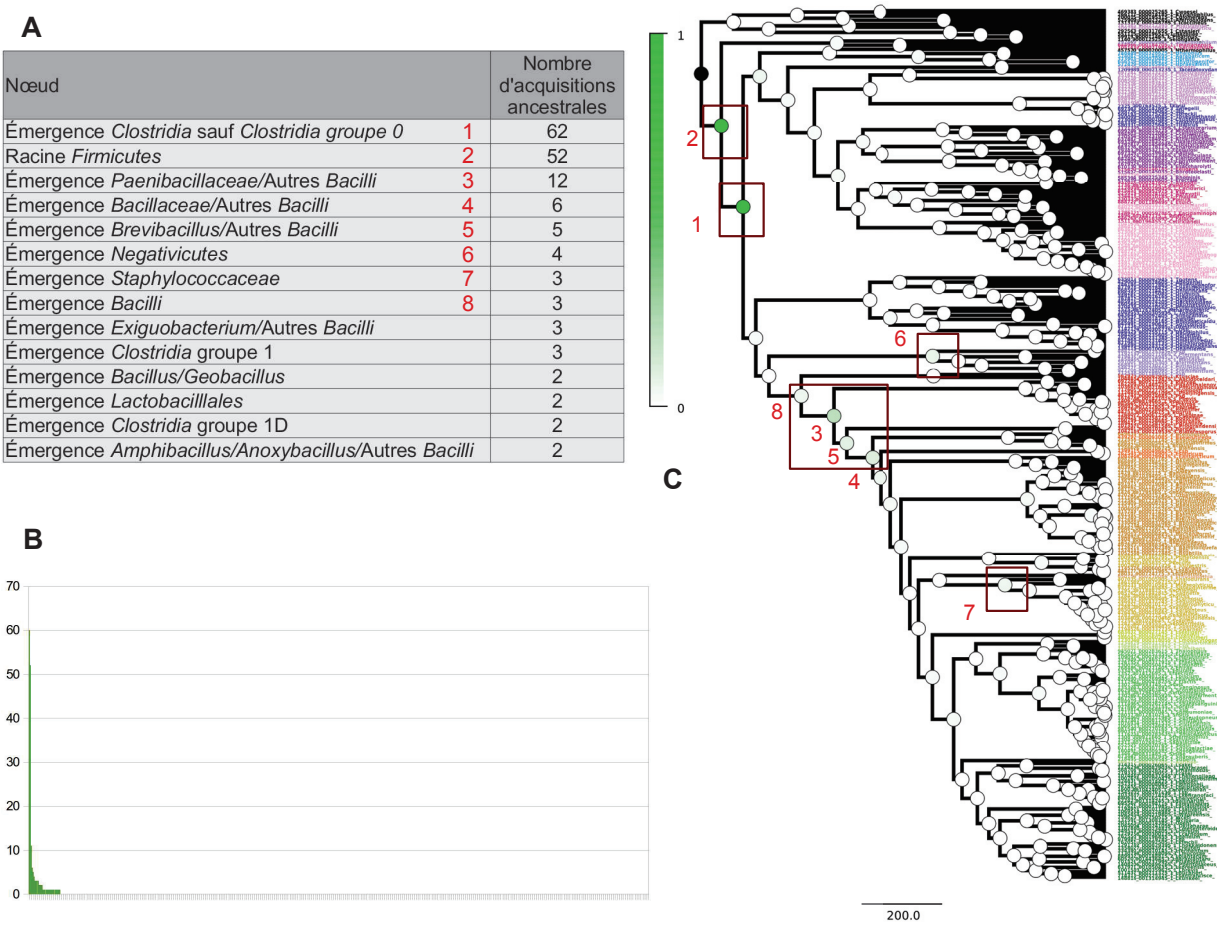


FIGURE 3.27 – Apparitions inférées par réconciliation. (A) Branches (ou nœuds) présentant deux apparitions de gènes ou plus. Les lignes blanches correspondent à des feuilles (souches), les lignes en gris à des branches (ou nœuds) internes. Les branches/feuilles présentant le plus d'événements sont numérotées et indiquées sur le panneau (C). (B) Distribution du nombre d'apparitions par branche. (C) Projection du nombre d'apparitions sur la phylogénie des *Firmicutes*. Les nombres en rouge ainsi que les encadrés correspondent à la table (A).

par l'approche de réconciliation (Cf 3.4.2.2). Trois explications sont envisageable. Tout d'abord, il est possible de faire l'hypothèse que l'arbre d'espèces correspond à la réalité et que les arbres de gènes sont de bonne qualité ce qui signifierait que les apparitions de gènes sont bien placées. La deuxième hypothèse est que l'arbre d'espèces est vrai mais que les espèces de *Clostridia* groupe 0 ont réellement acquis par transfert horizontal de nombreuses familles de gènes. Dans ce cas, les inférences réalisées par l'approche de réconciliation seraient les plus réalistes. La troisième hypothèse consiste à considérer que le groupe des *Clostridia* groupe 0 est mal placé dans l'arbre d'espèces ce qui induit pour un grand nombre de familles de gènes l'inférence de transferts horizontaux vers ce taxon (Cf 3.4.2.2). Les deux dernières hypothèses sont supportées par le fait que les séquences des *Clostridia* groupe 0 présentent des longues branches et par la grande divergence au sein des arbres de gènes et d'espèces. De plus la branche sœur de celle de l'émergence du groupe 0 des *Clostridia* est modérément supportée (85). Cela suggère alors que la majorité de ces familles de gènes auraient été présentes à la racine et que la prépondérance de cette branche est un artefact, tout comme pour l'inférence par profils phylogénétiques.
Concernant le troisième point chaud, il correspond aux mêmes branches que par l'inférence par profils phylogénétiques (figure 3.26). Les quatre branches basales des *Bacilli* comportent dans l'ordre 3, 12, 5 et 6 apparitions de gènes du cycle cellulaire. Les mêmes hypothèses que pour les profils phylogénétiques peuvent être faites pour ce point chaud.

Comparaison des deux méthodes Afin de mesurer la congruence des deux méthodes, nous avons calculé la distance en nombre de branches pour chaque famille de gènes entre la position de l'apparition inférée par profil phylogénétique et celle inférée par réconciliation au sein de l'arbre d'espèces. Les résultats sont présentés figure 3.28. Ainsi, 84 familles de gènes sur 179 présentent la même position d'apparition et 63 une différence d'une seule branche. Des résultats identiques ou très similaires avec les deux approches sont donc constatés pour une grande majorité de familles de gènes. Cependant, les deux méthodes donnent des positions différentes de 2 branches ou plus pour 32 familles de gènes. Ces différences s'expliquent par l'implémentation des transferts horizontaux de gènes lors de la réconciliation. Le cas le plus parlant est celui de la famille LysA3. Il s'agit d'une famille dont la distribution taxonomique est très partielle et qui de par la phylogénie semble avoir subi de nombreux transferts horizontaux

de gènes (figure 3.28).

A

Famille	Distance (en nœuds)	Famille	Distance (en nœuds)
LysA3	20	SbcCD1	4
DnaB2	18	CozE3	4
Aslfm	14	Pfs	4
B1	9	FemX	4
DacC	7	Spo0M	3
GatD1	7	AX1	3
RecU	6	LysA2	3
CozE2	6	GpsB	3
Rtp	6	Noc	2
XerC	6	NagA	2
PCDP2	6	GatD4	2
TolR	5	FtsW	2
TolQ	5	MreBH	2
SbcCD2	5	Asr	2
WalH	5	PCDP10	2
RodA2	5	DnaD	2

B

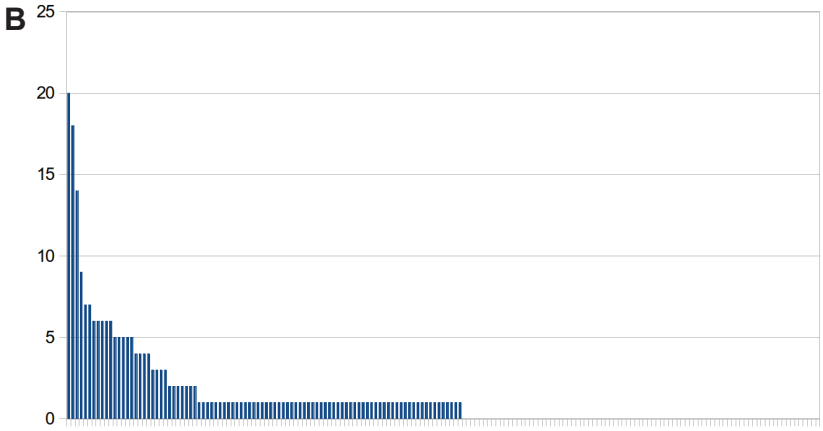


FIGURE 3.28 – Distances en nœuds des apparitions inférées par les deux méthodes. (A) Familles de gènes dont l'inférence de l'apparition présente une distance de deux nœuds ou plus entre les deux méthodes. (B) Distribution de la distance en nœuds entre l'apparition inférées par les deux méthodes par famille de gènes.

L'approche par profil phylogénétique a inféré l'apparition de cette famille à l'émergence des *Erysipelotrichia* et des *Staphylococcaceae* ainsi que plusieurs apparitions secondaires chez les autres espèces. De façon contradictoire, la réconciliation a inféré que l'apparition était située à l'émergence de *Clostridium cellulolyticum* H10 qui a été suivie d'une succession de transferts horizontaux de gènes vers d'autres espèces. La comparaison entre les deux méthodes met ainsi

en exergue les cas limites dont l'histoire évolutive est particulièrement difficile à reconstruire. Néanmoins, pour ce cas, il semblerait que la réconciliation ne donne pas la solution optimale puisqu'il n'est que peu probable que la famille soit apparue à l'émergence d'une souche puis ait été transférée à un très grand nombre d'espèces.

Identification de points chauds d'apparitions de gènes De façon générale, les deux approches mettent en évidence deux points chauds : la racine des *Firmicutes* et l'émergence des *Bacilli*. Le premier point chaud d'apparitions de gènes peut être expliqué par deux hypothèses. Tout d'abord, les gènes ont pu apparaître à l'émergence des *Firmicutes*. L'autre hypothèse est que les gènes étaient présents de façon ancienne chez les bactéries. Il serait donc logique de retrouver l'apparition de ces gènes au niveau de la racine des *Firmicutes*. La deuxième hypothèse est soutenue par le fait qu'un grand nombre des protéines impliquées dans le cycle cellulaire sont conservées chez les bactéries et notamment retrouvées chez *E. coli* [379].
Le point chaud à l'émergence des *Bacilli* est plus surprenant. De nombreux gènes semblent être acquis à cette branche mais peu de relations fonctionnelles entre ces gènes n'ont être décrites. Nous discuterons plus en détail de ce point en section 3.6. L'origine de ces familles semble variées. Certaines familles semblent issues de duplications (*e.g.* GpsB à partir de DivIVA), de transferts horizontaux (*e.g.* XerS) ou *de novo*/de transfert horizontaux de lignées éteintes (*e.g.* EzrA).
L'ensemble des apparitions de gènes pour les trois points chauds et la concordance entre les deux méthodes utilisées sont présentés en annexe .13.

3.4.2.2 Inférence des pertes, transferts horizontaux et duplications de gènes

La reconstruction de l'histoire des familles de gènes dans une lignée ne consiste pas uniquement à déterminer leur point d'apparition. Cela consiste à appréhender la manière dont ils ont été transmis au sein d'un clade (transmission verticale, horizontale ou duplication) mais également s'il ont été perdus au cours de l'évolution. L'approche par profils phylogénétique

permet uniquement d'inférer des pertes de gènes tandis que la réconciliation offre la possibilité d'inférer l'ensemble des événements possibles. Nous avons donc utilisé ces deux approches décrites précédemment pour identifier les événements majeurs ayant affecté le pool des gènes du cycle cellulaire chez les *Firmicutes*.

Inférence des pertes massives de gènes par profils phylogénétiques L'inférence des états ancestraux des familles de gènes au sein de l'arbre d'espèces des *Firmicutes* par les profils phylogénétiques a été réalisée. La distribution des pertes de gènes par branche semble suivre le même type de distribution que celle des apparitions de gènes mais la concentration en certaines branches est moins évidente. Néanmoins le graphe a permis de mettre en lumière quelques pertes massives de gènes présentés en figure 3.29. De façon générale, il apparaît que les multiples pertes aient eu lieu sur des branches plus récentes que pour les apparitions de gènes (figure 3.29C) Par exemple, 48 gènes ont été perdus chez la souche *Mageeibacillus indolicus* UPII9-5 dont les gènes de la synthèse de la capsule, les gènes de sporulation, les gènes Min ou encore les gènes Mre. De façon plus surprenante, des gènes essentiels pour la synthèse du peptidoglycane et pour la division sont perdus à cette même branche comme GlmS, GlmU, MurA mais aussi FtsA, composant pourtant essentiel à l'assemblage de l'anneau Z chez une grande majorité de bactéries. L'émergence des souches *Eubacterium sulci* ATCC 35585 et de *Erysipelothrix rhusiopathiae* str. Fujisawa sont aussi accompagnées de nombreuses pertes (37 pour chaque souche). Concernant les branches plus profondes, l'émergence des *Exiguobacterium* et celle de *Acetobacterium woodii* et *Eubacterium limosum* présentent 36 et 32 pertes respectivement (figure 3.29).

Inférence des pertes massives de gènes par réconciliation Les pertes de gènes ont également été inférées par réconciliation. Néanmoins, quelques précisions sont nécessaires sur le calcul du nombre de pertes. Pour des raisons pratiques, les événements de remplacement par transfert horizontal ont été scindés en un événement de perte et un événement d'acquisition par transfert. Pour compter le nombre de pertes de gènes, il a donc fallut soustraire pour chaque branche et pour chaque famille le nombre de transferts au nombre de perte. Lorsque

A

Nœud		Nombre de pertes
Mageeibacillus indolicus UPII9-5	1	48
Eubacterium sulci ATCC 35585	2	37
Erysipelothrix rhusiopathiae str. Fujisawa	3	37
Émergence *Eubacterium limosum/Acetobacterium woodii*	4	36
Émergence *Exiguobacterium*	5	32
Candidatus Arthromitus sp. SFB-rat-Yit	6	25
Émergence *Tissierellia/Peptoniphilus*	7	22
Aerococcus urinae ACS-120-V-Col10a	8	22

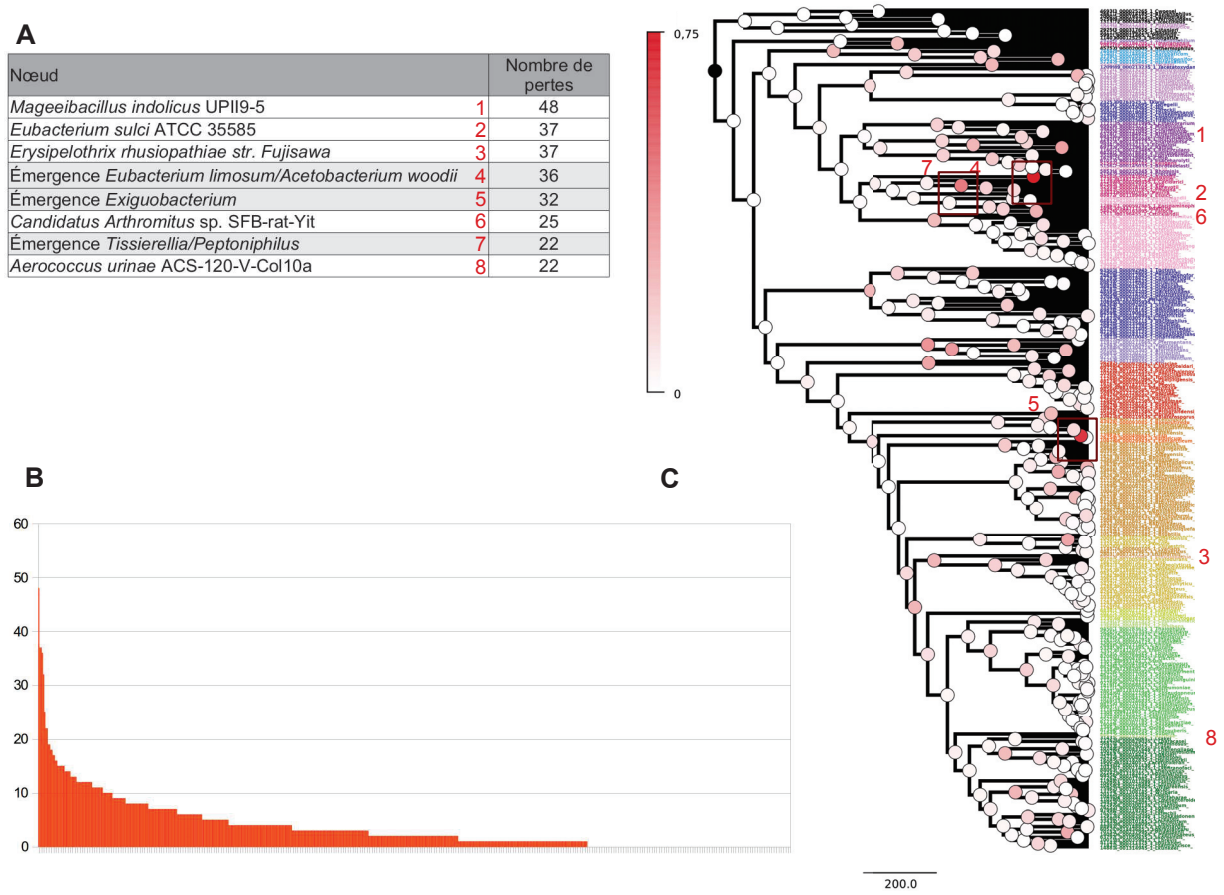


FIGURE 3.29 – Pertes inférées par parcimonie. (A) Branches (ou nœuds) présentant 22 pertes de gènes ou plus. Les lignes blanches correspondent à des feuilles (souches), les lignes en gris à des branches (ou nœuds) internes. Les branches/feuilles présentant le plus d'événements sont numérotées et indiquées sur le panneau (C). (B) Distribution du nombre de pertes par branche. (C) Projection du nombre de pertes sur la phylogénie des *Firmicutes*. Les nombres en rouge ainsi que les encadrés correspondent à la table (A).

des transferts et des pertes du même gène ont lieu à la même branche, nous avons considéré ces événements comme des remplacements homologues.

La distribution des pertes de gènes par branche *via* la réconciliation est similaire à celle calculée par profils phylogénétiques. L'émergence de *Mageeibacillus indolicus* UPII9-5 est accompagnée par de nombreuses pertes (47), tout comme inféré par les profils phylogénétiques (48) (figure 3.30A). L'émergence des souches *Eubacterium sulci* ATCC 35585 et de *Erysipelothrix rhusiopathiae* str. Fujisawa présentent aussi de nombreuses pertes (23 pertes pour chaque branche) mais moins nombreuses que par la méthode des profils phylogénétiques. Cela provient du fait que les gènes manquant ont été perdus à des branches ancestrales. Concernant les branches profondes, l'émergence des *Exiguobacterium* et celle de *Acetobacterium woodii* et *Eubacterium limosum* sont accompagnées par 25 et 28 pertes de gènes respectivement, en accord avec l'approche par profils phylogénétiques (figure 3.30A).

Comparaison entre les méthodes Certaines branches présentent de nombreuses pertes par la réconciliation qui ne sont pas retrouvés par l'approche des profils comme l'émergence des *Negativicutes* et celle des *Staphylococcaceae/Erisypelotrichia* qui comportent 37 et 30 pertes de gènes respectivement. Pour expliquer ces nombreuses pertes, il est nécessaire d'analyser les événements de transferts horizontaux (figure 3.32). Pour l'émergence des *Staphylococcaceae/Erysipelotrichia*, La branche correspondant à l'émergence des *Staphylococcaceae* présente 28 transferts horizontaux sans remplacement tandis que pour la branche correspondant aux *Erysipelotrichia*, 37 pertes sont inférées par approche de profil phylogénétique. Ces résultats suggèrent que les pertes à cette branche sont probablement dues à un biais méthodologique résultant de deux causes (figure 3.31). En effet, les séquences des *Staphylococcaceae* présentant souvent des longues branches se placent mal au sein des arbres de gènes et le génome d'*Erysipelotrichia* présente de nombreuses absences de gènes (annexe .10). Le cumul de ces deux conditions place ainsi une partie des pertes chez *Erysipelothrix rhusiopathiae* à la branche mère, c'est-à-dire à l'émergence des *Staphylococcaceae/Erysipelotrichia*. Le même effet est observé pour l'émergence des *Negativicutes* (17 pertes pour l'émergence des *Selenomonas/Veillonella/Megasphaera* par profils phylogénétiques et 39 acquisitions par transferts

A

Nœud		Nombre de pertes
Mageeibacillus indolicus UPII9-5	1	47
Émergence *Negativicutes*	2	37
Émergence *Staphylococcaceae/Erysipelotrichia*	3	30
Émergence *Acetobacterium woodii/Eubacterium limosum*	4	28
Émergence *Staphyococcaceae*/Autres *Bacilli*	5	28
Émergence *Exiguobacterium*	6	25
Émergence *Bacillus groupe cereus*	7	24
Émergence *Tissierellia/Peptoniphilus*	8	24
Erysipelothrix rhusiopathiae str. Fujisawa	9	23
Eubacterium sulci ATCC 35585	10	23

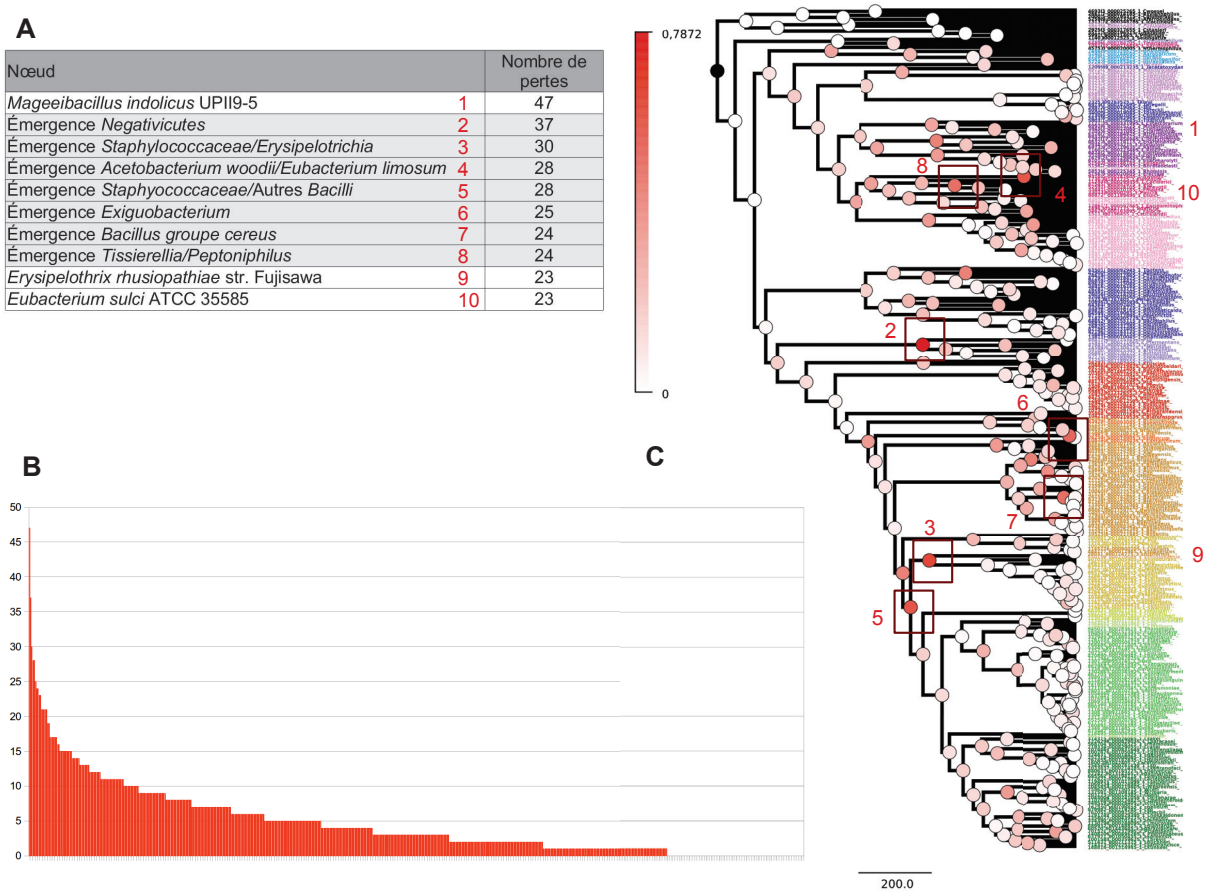


FIGURE 3.30 – Pertes inférées par réconciliation (A) Branches (ou nœuds) présentant 23 pertes de gènes ou plus. Les lignes blanches correspondent à des feuilles (souches), les lignes en gris à des branches (ou nœuds) internes. Les branches/feuilles présentant le plus d'événements sont numérotées et indiquées sur le panneau (C). (B) Distribution du nombre de pertes par branche. (C) Projection du nombre de pertes sur la phylogénie des *Firmicutes*. Les nombres en rouge ainsi que les encadrés correspondent à la table (A).

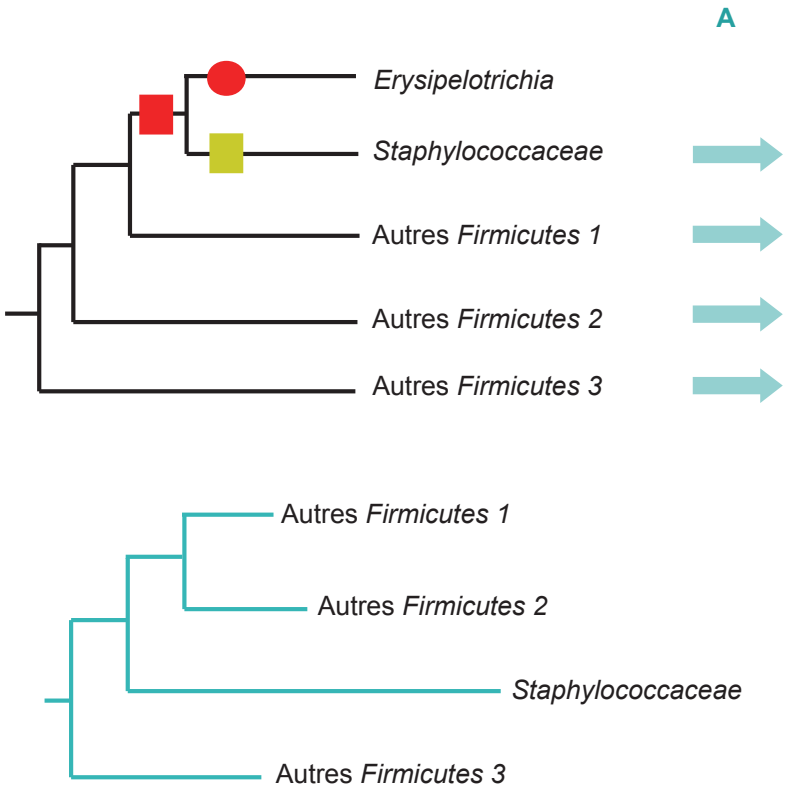


FIGURE 3.31 – Différences d'inférence des pertes entre profils phylogénétiques et réconciliation lié aux grandes branches. L'arbre du haut représente la phylogénie des espèces associé à la distribution taxonomique du gène A. Le gène A est uniquement absent chez *Erysipelotrichia*. L'approche par profil phylogénétique infère une perte à l'émergence de ce groupe (rond rouge). La phylogénie du gène A place les séquences *Staphylococcaceae* avec le groupe Autre *Firmicutes* 3. ce groupe est potentiellement mal placé à cause de sa longue branche et non pas à cause d'un réel transfert horizontal. La réconciliation infère une perte ainsi à l'émergence des *Erysipelotrichia*/*Staphylococcaceae* puis une réacquisition par transfert chez les *Staphylococcaceae*.

pour l'émergence des *Pelosinus*). De façon générale, le positionnement de ces pertes reste donc peu résolu car il est difficile de savoir si le positionnement des groupes concernés dans les phylogénies de gènes est du à un manque de signal ou à un réel transfert horizontal de gène. Néanmoins, les grandes branches associées aux séquences de ces deux groupes (*Staphylococcaceae* et *Pelosinus*) pourrait indiquer qu'il s'agisse d'un biais de reconstruction. Cela suggérerais que les pertes massives ont eu lieu à l'émergence des *Selenomonas/Veillonella/Megasphaera* et des *Erysipelotrichia*.

Identification de points chauds de pertes de gènes Bien que les deux méthodes n'identifient pas exactement les mêmes branches comme étant des points chauds évolutifs pour les pertes, il existe une certaine convergence. L'ensemble de ces résultats suggèrent que de nombreuses pertes ont eu lieu chez *Mageeibacillus indolicus* UPII9-5, *Eubacterium sulci* ATCC 35585, *Erysipelothrix rhusiopathiae* str. Fujisawa mais également à l'émergence des *Exiguobacterium*, des *Selenomonas/Veillonella/Megasphaera* (*Negativicutes*, des *Tissierellia/Peptoniphilus* et de *Acetobacterium woodii/Eubacterium limosum*. Ces résultats sont corroborés par l'absence de nombreux gènes du cycle cellulaire chez les souches concernées (annexe .10). Ainsi, il est tentant de faire l'hypothèse que ces taxons présentent des moyens d'opérer et de réguler leur cycle cellulaire très différents de ceux connus chez les bactéries modèles. Aussi, ces résultats sont renforcés par le fait que les *Tissierellia*, les *Negativicutes* et les *Erysipelotrichia* constituent des classes à part entière [289]. L'équipement génétique lié au cycle cellulaire étant essentiel à la morphologie des cellules, un changement de cet équipement peut être à l'origine d'une transformation radicale conduisant ainsi à la formation d'une nouvelle lignée. Pour le cas des *Negativicutes*, les *Pelosinus* (le premier groupe à émerger chez les *Negativicutes*) semblent avoir conservé plus de gènes du cycle cellulaire analysés ce qui semble indiquer qu'ils pourraient présenter des grandes différences phénotypiques par rapport au reste des *Negativicutes*. Concernant *Mageeibacillus indolicus* UPII9-5 appartenant à la famille des *Ruminococcaceae*, les pertes massives associées à son émergence nous conduisent à faire l'hypothèse que cette souche devrait présenter des différences notoire en terme de mécanisme du cycle cellulaire par rapport aux autres *Ruminococcaceae*. Les mêmes hypothèses peuvent être faites pour *Eubacterium sulci* ATCC 35585 et *Acetobacterium woodii/Eubacterium limosum*.

De façon générale, les gènes les plus perdus sont les gènes Spo (sporulation), Mre (elongation), Min (localisation anneau Z), Cps (capsule) Wal (régulation du peptidoglycane) et Anm/Nag/Mur (recyclage du peptidoglycane). Concernant les gènes Spo, Cps, Anm/Nag/Mur, ces résultats ne sont pas surprenants car la sporulation, la production de la capsule et le recyclage du peptidoglycane ne sont pas des processus essentiels et sont retrouvés de façon partielle chez les bactéries. La perte des gènes Mre et Min est plus surprenante et indique probablement l'existence de mécanismes alternatifs chez les *phyla* concernés.

Il est enfin important de préciser que ces pertes ne semblent pas résulter d'une mauvaise qualité de séquençage/assemblage des génomes. Pour vérifié cela, une technique possible est de construire des familles de gènes conservées et d'identifier de potentiels génomes avec un grand nombre de gènes manquant. L'absence de nombreux marqueurs indique ainsi que le génome a été mal séquencé/assemblé. Nous avons donc vérifié le nombre de protéines ribosomiques (considérées comme conservées chez les bactéries) identifiées dans chaque génome par RiboDB [219], figure .4. Il apparaît ainsi que seul un génome contient moins de 48 protéines ribosomiques sur les 53 initiales (*Thermoanaeobacter mathranii* subsp.*mathranii* str. A3). De plus, l'émergence de la souche correspondant à ce génome n'a pas été inféré comme accompagnée de multiples pertes. Le biais de séquençage et d'assemblage peut donc être écarté.

Autres événements

A

Nœud		Nombre de transferts
Symbiobacterium thermophilum IAM 14863	1	43
Émergence *Pelosinus*	2	39
Erysipelothrix rhusiopathiae str. Fujisawa	3	36
Émergence *Clostridia* groupe 0	4	35
[Clostridium] acidurici 9a	5	34
Émergence *Bacillus* groupe *cereus*	6	33
Eubacterium sulci ATCC 35585	7	31
Thermaerobacter marianensis DSM 12885	8	30
Carboxydothermus hydrogenoformans Z-2901	9	30
Jeotgalibacillus malaysiensis	10	29
Émergence *Staphylococcaceae*	11	28

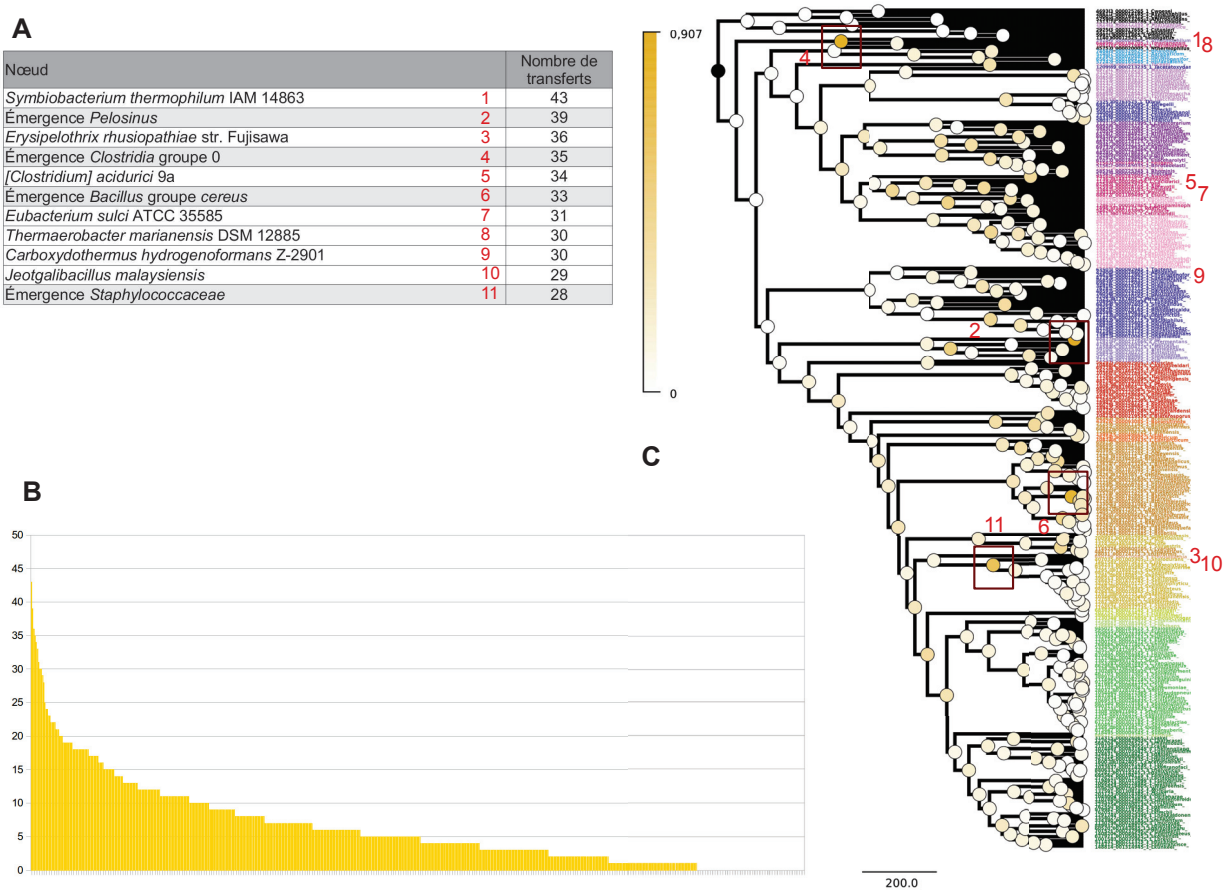


FIGURE 3.32 – Transferts horizontaux inférés par réconciliation (A) Branches (ou nœuds) présentant 28 transferts horizontaux de gènes ou plus. Les lignes blanches correspondent à des feuilles (souches), les lignes en gris à des branches (ou nœuds) internes. Les branches/feuilles présentant le plus d'événements sont numérotées et indiquées sur le panneau (C). (B) Distribution du nombre de transferts horizontaux par branche. (C) Projection du nombre de transferts horizontaux sur la phylogénie des *Firmicutes*. Les nombres en rouge ainsi que les encadrés correspondent à la table (A).

Concernant les transferts horizontaux de gènes sans remplacement, la distribution du nombre de transferts par branche semble présenter les mêmes caractéristiques que pour les pertes (figure 3.32). L'émergence de *Symbiobacterium thermophilum* et des *Clostridia* groupe 0 présentent 43 et 35 transferts respectivement. Il s'agit la encore probablement d'un biais, soit d'un mauvais positionnement de ces *taxa* dans l'arbre d'espèces, soit des séquences dans les arbres de gènes qui a d'ailleurs généré l'incertitude du positionnement des présences ancestrales de nombreux gènes (Cf section 3.4.2.1). L'émergence des *Pelosinus* est aussi associée à un grand nombre de transferts mais il s'agit encore une fois probablement d'un biais dont nous avons discuté dans la section 3.4.2.2, de même pour *Erysipelothrix rhusiopathiae* str. Fujisawa. Nous avons également projeté les événements de remplacements homologues par transfert horizontal de gènes sur la phylogénie des espèces (figure 3.33). La distribution du nombre de remplacements homologues par branche semble similaire à celles des pertes et des transferts horizontaux. La branche présentant le plus de remplacements homologues correspond à l'émergence de *Caldanaerobacter subterraneus* subsp. *tengcongensis* MB4. Ce résultat est confirmé par une étude sur l'espèce *Caldanaerobacter subterraneus* où il a été notamment montré que 5% des gènes ont été potentiellement acquis par transferts horizontaux [411]. Les émergences des *Caldicellulosiruptor/Mahella*, de *Mageeibacillus indolicus* UPII9-5, de *Erysipelothrix rhusiopathiae* str. Fujisawa et de *Streptococcus thermophilus/salivarius* semblent aussi avoir été accompagnées par un grand nombre de remplacements homologues par transfert horizontal. Il est intéressant de noter qu'une étude a suggéré que de nombreux îlots génomiques provenant de transferts horizontaux composeraient le génome de *Streptococcus thermophilus*, confirmant ainsi le nombre élevé de transferts trouvés à l'ancêtre de cette espèce et de *Streptococcus salivarius* [203]. Les remplacements homologues concernent une plus grande variété de processus concernés que pour les pertes de gènes. Il est néanmoins intéressant de noter que certains clusters de gènes entiers semblent être remplacés comme par exemple le cluster Mre/Min à l'émergence des *Caldicellulosiruptor*.

Enfin, les événements de duplication ont été analysés (figure 3.34). Il semble que peu de duplications aient eu lieu durant la diversification des *Firmicutes* pour les gènes du cycle cellulaire. La majorité des duplications observées concernent les branches distales, c'est-à-dire qu'elles semblent récentes. Ces résultats ne sont pas surprenant au vue de la méthodologie appliquée

A

Nœud		Nombre de pertes avec remplacement
Caldanaerobacter subterraneus subsp. *tengcongensis* MB4	1	41
Émergence *Caldicellulosiruptor/Mahella*	2	33
Mageeibacillus indolicus UPII9-5	3	30
Erysipelothrix rhusiopathiae str. Fujisawa	4	28
Émergence *Streptococcus thermophilus/salivarius*	5	28
Émergence *Bacillus weihenstephanensis/mycoides*	6	26
Streptococcus gordonii str. Challis substr. CH1	7	25
Émergence *Pediococcus*	8	25

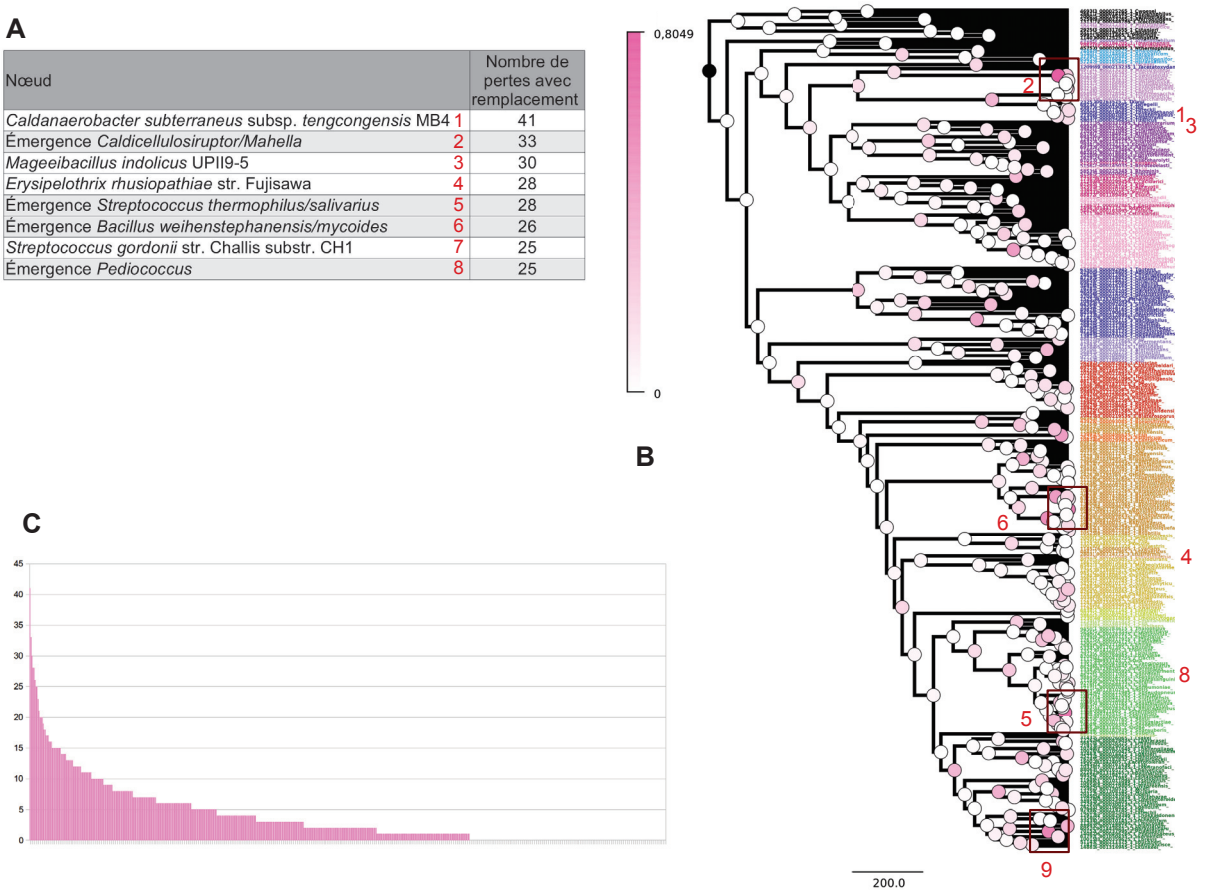


FIGURE 3.33 – Remplacements homologues inférées par réconciliation (A) Branches (ou nœuds) présentant 25 remplacements homologues de gènes ou plus. Les lignes blanches correspondent à des feuilles (souches), les lignes en gris à des branches (ou nœuds) internes. Les branches/feuilles présentant le plus d'événements sont numérotées et indiquées sur le panneau (C). (B) Distribution du nombre de remplacements homologues par branche. (C) Projection du nombre de remplacements homologues sur la phylogénie des *Firmicutes*. Les nombres en rouge ainsi que les encadrés correspondent à la table (A).

pour construire les sous-familles. En effet, des duplications anciennes n'apparaissent que peu puisque nous avons effectué un découpage en sous-familles des familles multigéniques. Les duplications identifiées ici correspondent donc principalement à des duplications tardives difficilement détectables par notre approche de découpage des familles.

3.4.2.3 Histoire évolutive des clusters de gènes du cycle cellulaire

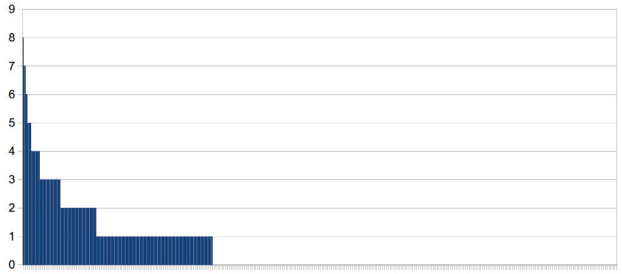
À partir des matrices de synténies aux feuilles de l'arbre d'espèces, nous avons reconstruit les synténies ancestrales à chaque branche interne pour tous les gènes du cycle cellulaire. Les clusters de gènes ont ensuite été reconstruits pour chaque branche en regroupant les gènes par SLC. Nous avons ensuite inféré les événements de gains, de pertes et de modifications des clusters.

Reconstruction des clusters de gènes chez l'ancêtre des *Firmicutes* La reconstruction des clusters ancestraux a donc permis d'inférer des clusters présents chez l'ancêtre des *Firmicutes* (table 3.2). Néanmoins, le premier groupe à émerger dans l'arbre d'espèces, le groupe 0 des *Clostridia* a induit des incertitudes dans la projection des apparitions de gènes car sa position au sein de l'arbre ne semble pas incontestable (Cf section 3.4.2.1). Nous avons donc considéré les clusters inférés à cette branche et à la racine comme clusters ancestraux aux *Firmicutes* (table 3.2). Il apparaît qu'une vingtaine de clusters soient présents chez l'ancêtre des *Firmicutes*. Les clusters du cycle cellulaire identifiés précédemment chez *S. pneumoniae* et *B. subtilis* semblent conservés (figure 3.12). La plupart des gènes contenus dans ces clusters sont reliés fonctionnellement comme le cluster ScpAB (ségrégation du chromosome), FtsEX (système FtsEX de dégradation des protéines), NagAB (synthèse des précurseurs du peptidoglycane) ou encore Mur/Mre (élongation et localisation du septum).
Le cluster de division et paroi cellulaire (« Division and Cell-Wall cluster ») contenant notamment FtsZ et FtsA semble également conservé. Cependant, le cluster est divisé en deux parties. Cette observation est du au fait que de nombreux génomes possèdent effectivement un cluster scindé et notamment chez les *Clostridia* (Cf section 3.6). Le même constat est fait

C

Nœud		Nombre de duplications
Clostridium acetobutylicum EA 2018	1	8
Alicyclobacillus acidocaldarius subsp. acidocaldarius Tc-4-1	2	7
Clostridium bornimense M2/40	3	7
Clostridium cellulovorans 743B	4	6
Weissella cibaria CH2	5	6
Émergence *Pelosinus*	6	5
Sulfobacillus acidophilus TPY	7	5
Butyrivibrio proteoclasticus B316	8	5
Clostridium lentocellum DSM 5427	9	5
Émergence *Bacillus* groupe *cereus*		4
Symbiobacterium thermophilum IAM 14863		4
Paenibacillus sp. Y412MC10		4
Bacillus megaterium WSH-002		4
Desulfotomaculum acetoxidans DSM 771		4
Selenomonas ruminantium subsp. *lactilytica* TAM6421		4
Eubacterium limosum SA11		4
Clostridium perfringens str. 13		4
Veillonella parvula DSM 2008		4

B



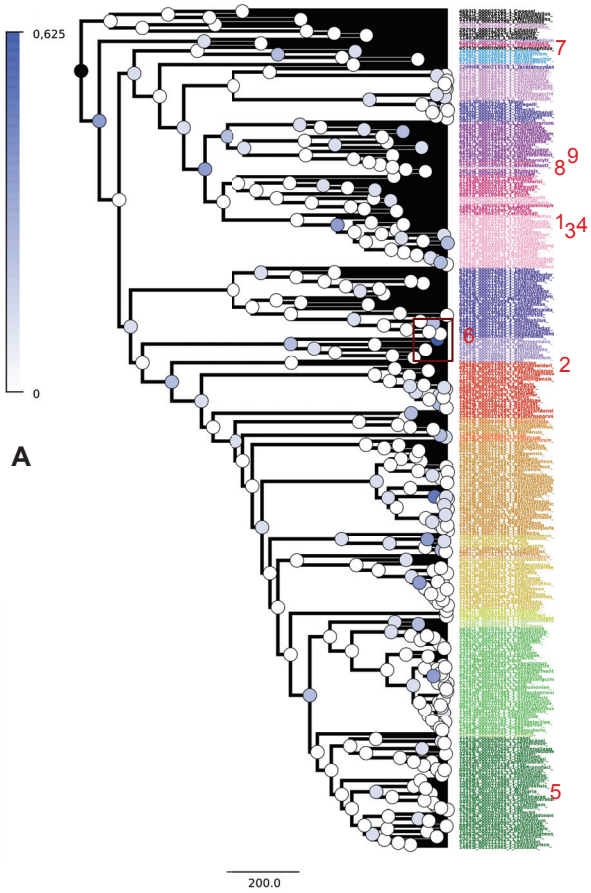


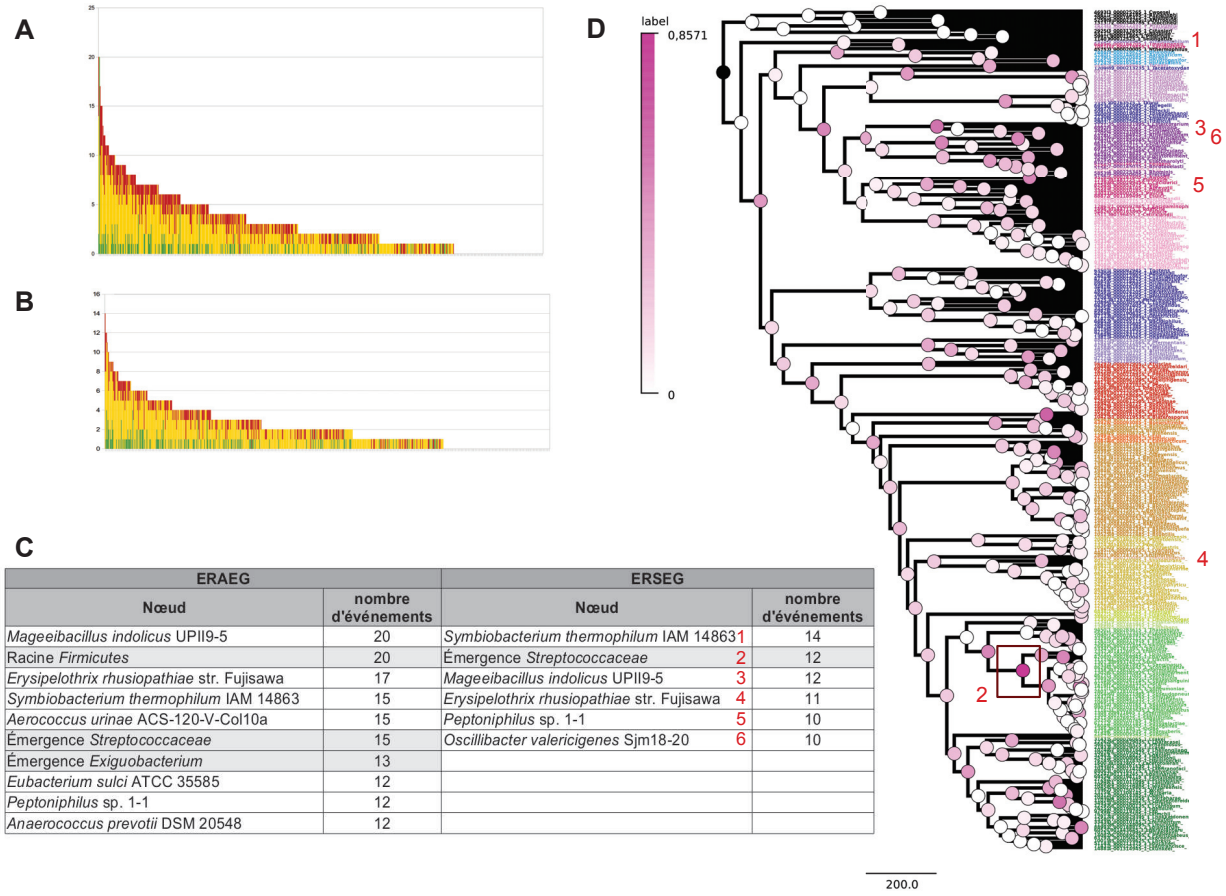
FIGURE 3.34 – Duplications inférées par réconciliation (A) Branches (ou nœuds) présentant quatre duplications par transferts horizontaux de gènes ou plus. Les lignes blanches correspondent à des feuilles (souches), les lignes en gris à des branches (ou nœuds) internes. Les branches/feuilles présentant le plus d'événements sont numérotées et indiquées sur le panneau (C). (B) Distribution du nombre de duplications par branche. (C) Projection du nombre de duplications sur la phylogénie des *Firmicutes*. Les nombres en rouge ainsi que les encadrés correspondent à la table (A).

Cluster	Gènes contenus dans le cluster (Racine des Firmicutes)	Gènes contenus dans le cluster (Émergence des Firmicutes hors Clostrida groupe 0)
ScpAB	LysA1, ScpA, ScpB	DacB, LysA1, ScpA, ScpB
MurKQ	MurK, MurQ	MurK, MurQ
Wal	WalH, WalI, WalK, WalR	WalH, WalI, WalJ, WalK, WalR
FtsEX	FtsX, FtsE	MinJ, FtsE, FtsX
NagAB	NagA, NagB	NagA, NagB
Min/Mre	B5, Maf, MinC, MinD, MinE, MreB, MreC, MreD, RadC, RodA	B5, Maf, MinC, MinD, MinE, MreB, MreC, MreD, RadC, RodA
GlmMS	GlmM, GlmS	GlmM, GlmS
StkP	Mtf, PhpP, PriA, StkP, SunL	Mtf, PhpP, PriA, StkP, SunL
DCW1	B4, FtsA, FtsL, FtsQ, FtsZ, MraW, MraY, MraZ, MurA2, MurB2, MurE, MurF, MurG1, SpoIIGA, SpoVE	B4, FtsA, FtsL, FtsQ, FtsZ, MraW, MraY, MraZ, MurA2, MurB2, MurC, MurD, MurE, MurF, MurG1, SpoIIGA, SpoVE
DCW2	DivIVA, IleS1, PCDP6, PCDP7, SepF	DivIVA, IleS1, PCDP6, PCDP7, PCDP8, SepF
Ori1	DnaA, DnaN, RecF	DnaA, DnaN, RecF
Ori2	GidA, GidB, Jag, Noc, ParA, ParB, SpoIIIJ1	GidA, GidB, Jag, Noc, ParA, ParB, SpoIIIJ1
	FtsJ, RecN	FtsJ, RecN
	AlaS, CozE4	AlaS, CozE4
	FtsY, Smc	FtsY, Smc
	Alr_like, FemX	Alr_like, FemX
	Mbl, MurA1, SpoIID, SpoIIID	Mbl, MurA1, SpoIID, SpoIIID
	DacF, XerD	DacF, XerD, NudF
	FtsH1, SpoIIE, TilS	DivIC, FtsH1, SpoIIE, TilS
	GlmU, Mfd, SpoVG, YabM	GlmU, Mfd, SpoVG, YabM
Cps		CpsB, CpsC, CpsD
		FtsH3, FtsK1, RecA, RodZ
		BX1, FtsW

TABLE 3.2 – Clusters de gènes inférés à l'ancêtre des *Firmicutes*. Les familles en vert correspondent aux familles ajoutées aux clusters à l'émergence des *Firmicutes* hors *Clostridia*.

pour le cluster d'origine de réplication (Ori) mais il semble que cela provienne d'un artefact de reconstruction. En effet, les génomes sont représentés de façon linéaire dans les fichiers avec un début et une fin. le cluster Ori1 est situé majoritairement aux des génomes tandis que le cluster Ori2 est situé à la fin. L'inférence des synténies ne prenant pas en compte la circularisatition du génome génère donc deux clusters.
Huit Clusters sont également identifiés comme étant présents chez l'ancêtre des *Firmicutes* mais dont les gènes les constituant n'ont pas été décrits comme étant reliés fonctionnellement. Trois clusters sont inférés à l'émergence des *Firmicutes* hors *Clostridia* groupe 0 mais pas à la racine des *Firmicutes*. Cela provient potentiellement du biais lié au positionnement du groupe 0 des *Clostridia*.

Réarrangements de clusters indépendants des événements de gènes Les événements liés aux clusters de gènes peuvent être des gains, des pertes ou des modifications. Nous avons défini le gain de cluster par l'apparition d'un cluster à une branche alors que celui-ci était absent à la branche ancestrale. La modification d'un cluster correspond à l'événement d'éloignement ou de rapprochement d'un ou plusieurs gènes au cluster à la condition qu'il reste au moins deux gènes dans le cluster après la modification. Enfin, la perte correspond à la dislocation totale du cluster de gènes, c'est-à-dire qu'aucun des gènes n'est voisin après l'événement. Néanmoins, une perte ou une apparition d'un gène au sein d'un cluster induit la modification de ce dernier. Afin de ne pas présenter une information redondante avec les sections 3.4.2.1 et 3.4.2.2, nous avons également inféré les événements de réarrangement de clusters en supprimant ceux qui était concomitant avec des événements de perte/apparition de gènes inférés par les profils phylogénétiques. Nous avons choisi les inférences d'absence/présence par profils phylogénétiques plutôt que par réconciliation car c'est le même modèle de parcimonie qui a été utilisé pour les profils phylogénétiques et pour les synténies ancestrales. Nous parlerons donc d'Événements de Réarrangements Sans Événement de Gènes (ERSEG) et d'Événements de Réarrangements Avec Événement de Gènes (ERAEG).



ERAEG		ERSEG		
Nœud	nombre d'événements	Nœud		nombre d'événements
Mageeibacillus indolicus UPII9-5	20	*Symbiobacterium thermophilum* IAM 14863	1	14
Racine *Firmicutes*	20	Émergence *Streptococcaceae*	2	12
Erysipelothrix rhusiopathiae str. Fujisawa	17	*Mageeibacillus indolicus* UPII9-5	3	12
Symbiobacterium thermophilum IAM 14863	15	*Erysipelothrix rhusiopathiae* str. Fujisawa	4	11
Aerococcus urinae ACS-120-V-Col10a	15	*Peptoniphilus* sp. 1-1	5	10
Émergence *Streptococcaceae*	15	*Oscillibacter valericigenes* Sjm18-20	6	10
Émergence *Exiguobacterium*	13			
Eubacterium sulci ATCC 35585	12			
Peptoniphilus sp. 1-1	12			
Anaerococcus prevotii DSM 20548	12			

FIGURE 3.35 – Événements de clusters de gènes inférées par parcimonie. (A) Distribution du nombre d'événements de réarrangements de clusters par branche (ERAEG). (B) Distribution du nombre d'événements de réarrangements de clusters par branche (ERSEG). Rouge : pertes du cluster, jaune : modification du cluster, vert :gain du cluster. (C) Branches (ou nœuds) présentant 12 ou 10 événements réarrangements de clusters de gènes ou plus. Les lignes blanches correspondent à des feuilles (souches), les lignes en gris à des branches (ou nœuds) internes. Les branches/feuilles présentant le plus d'événements sont numérotées et indiquées sur le panneau (D). (D) Projection du nombre d'événements de clusters de gènes (ERSEG) sur la phylogénie des *Firmicutes*. Les nombres en rouge ainsi que les encadrés correspondent à la table (C).

Inférence des événements de réorganisation génomique De façon générale, la distribution du nombre d'événements par branche montre qu'un petit nombre de branches présente la majorité des événements. Aussi, la suppression des événements affectant les clusters de gènes concomitants avec les événements de pertes/apparitions de gènes a des effet drastiques sur les résultats (figure 3.35A,B,C). Notamment, une diminution importante du nombre global d'événements est constatée (figure 3.35A,B). De plus, certaines branches présentant un grand nombre d'ERAEG ne sont pas associées à un grand nombre de ERSEG (figure 3.35C). Cela indique ainsi que certaines branches étaient associées à un grand nombre d'événements uniquement parce que des gènes apparaissaient directement organisés en clusters, soit parce que les gènes perdus au sein des clusters étaient réellement perdus à ces branches. C'est le cas par exemple de la racine des *Firmicutes*. 20 événements de réarrangements sont inférés par ERAEG alors que 0 événements sont inférés par ERSEG. Ce résultat est cohérent puisque toutes les apparitions de clusters à cette branche sont concomitantes avec les apparitions de gènes inclus dans ces clusters. Nous n'allons ici décrire que les événements inférés par ERSEG car non redondants avec la section 3.4.2.1 et 3.4.2.2.
Les branches associées avec un grand nombre de pertes correspondent principalement à l'émergence de souches : *Symbiobacterium thermophilum* IAM 14863, *Erysipelothrix rhusiopathiae* str. Fujisawa et *Mageeibacillus indolicus* UPII9-5, *Peptoniphilus* sp. 1-1 et *Oscillibacter valericigenes* Sjm18-20. Il est intéressant de noter que pour la plupart de ces souches, de multiples pertes ont été inférées dans la section 3.4.2.2. Il semblerait donc que les pertes de nombreux gènes aient été accompagnés des réarrangements d'autres clusters de gènes. Concernant les branches plus profondes, seule l'émergence des *Streptococcaceae* semble être accompagnée de nombreux réarrangements de clusters de gènes du cycle cellulaire (12).

3.4.2.4 Événements majeurs des gènes du cycle cellulaire chez les *Firmicutes*

Il semble donc que plusieurs événements majeurs aient affecté l'équipement génétique du cycle cellulaire pendant la diversification des *Firmicutes*. Il est important de noter que ces branches constituent les points chauds les plus importants mais que d'autres branches ont été affectés par de nombreux événements évolutifs. Nous ne présentons ici que les points chauds

majoritaires de façon non-exhaustive.

Tout d'abord, il est vraisemblable qu'un grand nombre des gènes du cycle cellulaire étudiées soient anciens chez les *Firmicutes*. De 62 à 130 sur les 179 étudiés sont inférés à la racine ou juste après la divergence du groupe *Clostridia* 0, selon la méthode utilisée (profil phylogénétique, réconciliation, ...). Il est important de noter que 81 à 98 de ces gènes ont été inférés comme étant en cluster de gènes (20 à la racine et 23 à la divergence du groupe *Clostridia* 0). Au sein des *Clostridia*, quelques branches semblent avoir été affectés par de nombreux événements évolutifs.

L'émergence des *Selenomonas/ Veillonella/ Megasphaera* (ou des *Negativicutes*), d'*Eubacterium sulci* ATCC 35585 d'*Acetobacterium woodii/ Eubacterium limosum*, des *Tissierellia* (incluant ou non *Peptoniphilus* sp. 1-1), ont été principalement accompagnées de pertes de gènes du cycle cellulaire (de 22 à 37 pertes). L'émergence des *Caldicellulosiruptor* et de *Caldanaerobacter subterraneus* subsp. tengcongensis MB4 présentent un grand nombre de remplacements homologues homologue par transfert horizontal de gène. L'émergence de la souches *Symbiobacterium thermophilum* IAM 14863, de *Peptoniphilus* sp. 1-1 et de *Oscillibacter valericigenes* Sjm18-20 ont été accompagnées de multiples réarrangements de clusters de gènes du cycle cellulaire. Enfin, l'émergence de la souche *Mageeibacillus indolicus* UPII9-5 a conduit à de nombreuses pertes de gènes, remplacements homologues de gènes et réarrangements de clusters.

L'émergence des *Bacilli* a été accompagnée de l'apparition de 11 à 26 gènes en fonction de la méthode d'inférence et de la prise en compte ou non des branches voisines. L'émergence des *Exiguobacterium* a été accompagnée d'un grand nombre de pertes. Ensuite, l'apparition de la souche *Erysipelothrix rhusiopathiae* str. Fujisawa est accompagnée de nombreuses pertes, remplacements homologues et réarrangements de clusters. L'émergence des *Streptococcaceae* s'est accompagnée d'un grand nombre de réarrangements de clusters de gènes. Enfin, l'apparition des *Streptococcus thermophilus/ salivarius* s'est accompagnée de nombreuses remplacements homologues.

3.5 Inférence de relations fonctionnelles par approche corrélative

3.5.1 Méthodes de corrélation

3.5.1.1 Déclinaisons des types d'informations évolutives

Nous avons comparé les histoires évolutives des 179 familles de gènes afin de déterminer si certaines présentaient des similitudes et étaient corrélées. La comparaison par paire de ces 179 familles de gènes nous a permis de construire une matrice de 15 931 couples (matrice triangulaire). Trois méthodes ont été utilisées pour reconstruire l'histoire des familles de gènes.

- Les profils phylogénétiques : co-occurrences et co-événements (gènes)
- Les phylogénies et leur réconciliation respective avec la phylogénie d'espèces : co-événements et co-occurrences (gènes)
- Les contextes génomiques : synténies (organisation génomique)

Nous avons ainsi quantifier la corrélation/similarité pour chaque couple de familles de gènes à partir des résultats générés par ces trois méthodes. Deux approches sont possible pour le calcul des coefficients de corrélation/similarité. En effet, il est possible d'étudier les données évolutives soit uniquement au feuilles de l'arbre (c'est-à-dire chez les espèces séquencées uniquement) soit en prenant aussi en compte les informations générées par la reconstruction de l'histoire évolutive aux branches internes de l'arbre d'espèces (c'est-à-dire chez les souches séquencées et leurs ancêtres). Les données générées par la reconstruction de l'histoire évolutive des familles peuvent être déclinées en deux catégories : les événements ayant eu lieu sur chaque branche ou l'état de présence/absence induit par ces événements (par exemple, une perte de gène à une branche induit que le gène est présent chez le nœud père et que le gène est absent dans le nœud fils).
Nous avons donc calculé différentes métriques pour calculer les coefficients de corrélation/similarité entre les familles de gènes résumées en table 3.3. Le « p » minuscule correspond à l'utilisation

des profils de présence/absence, et le « e » minuscule correspond aux événements. Le « P » majuscule correspond aux profils phylogénétiques, « R » à la réconciliation et « S » aux synténies. Enfin, « F » correspond à l'utilisation des feuilles, c'est-à-dire des génomes et « FN » à l'utilisation des feuilles et des nœuds (ou branches) ancestraux.

Espace d'analyse	Type d'information	Profil phylogénétique	Réconciliation	Synténie
Feuilles	Présence/ absence	pP(F) : MI+Phi		pS(F) : J
Nœuds et feuilles		pP(FN) : MI+Phi	pR(FN) : MI+Phi	pS(FN) : J
	Événements	eP(FN) : J	eR(FN) : J	

TABLE 3.3 – Méthodes utilisées pour comparer les histoires de familles de gènes. p :profil présence/absence, e : événements, P : Profil phylogénétique, R : Réconciliation, S : Synténie, F : Feuilles, FN : feuilles et Nœuds, MI : Information Mutuelle, J : Jaccard.

3.5.1.2 Métriques utilisées

Les coefficients utilisés sont l'information mutuelle et le coefficient ϕ (ou coefficient de Pearson pour les caractères binaires) pour les profils de présence/absence, inférés directement par les profils phylogénétiques ou par la réconciliation. En effet, ces deux métriques permettent la comparaison de deux vecteurs binaires et ont été utilisées pour des analyses de même nature dans la littérature [46], [490], [313], [211], [168]. L'information mutuelle correspond à la mesure de dépendance statistique entre deux variables tandis que le coefficient ϕ permet de mesurer l'intensité de liaison entre deux variables binaires. Nous avons utilisé la moyenne des deux valeurs obtenues par les deux métriques car le coefficient ϕ a tendance a être optimiste tandis que l'information mutuelle est plutôt pessimiste (figure 3.36). Il est important de noter que, bien que l'information mutuelle mesure la dépendance statistique entre deux variables et le coefficient ϕ mesure une corrélation, nous parlerons de corrélation entre les familles lorsque nous parlerons de ces métriques.

La synténie entre deux familles au sein de l'ensemble des branches concernées correspond à

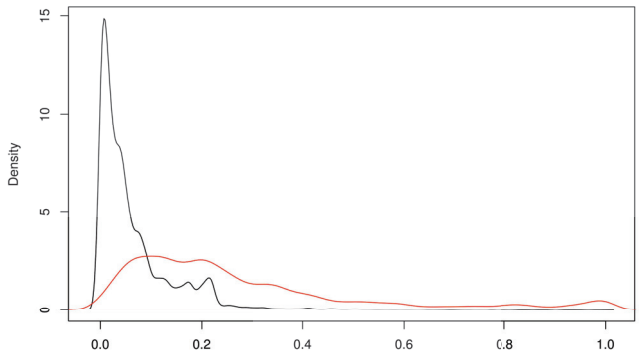


FIGURE 3.36 – Comparaison de la distribution l'information mutuelle et de la valeur absolue du coefficient ϕ. La matrice des 15 931 liens pour les 179 familles de gènes a été utilisée à partir des données pP(FN) (corrélation par les états présence/absence à partir des profils phylogénétiques sur l'ensemble des feuilles et nœuds). Le coefficient ϕ est plus optimiste que l'information mutuelle.

un seul vecteur binaire. Le coefficient de similarité de Jaccard qui permet de mesurer la similarité entre deux ensembles parait être une mesure adaptée à ce type de données. Nous avons donc utilisé le coefficient de Jaccard qui mesure l'étendue de la synténie au sein des branches qui possèdent au moins un des deux gènes.
Enfin, pour les événements de gènes, nous avons aussi utilisé le coefficient de Jaccard qui estime la similarité de l'ensemble des événements de deux familles de gènes à chaque branche de l'arbre d'espèces.

Afin d'estimer la chance d'avoir obtenu de tels résultats par hasard, nous avons décidé de calculer un Z-score pour toutes les métriques. Pour un couple de famille de gènes A et B, nous avons calculé la moyenne des deux Z-scores des deux familles A et B. Chaque Z-score prend en compte la moyenne et la variance des valeurs obtenues pour chaque famille de gènes avec l'ensemble des autres familles de gènes. Les Z-scores ont ensuite été normalisés par la valeur maximale (différentes pour les valeurs positives et négatives) pour obtenir un score compris

entre -1 et 1. Les couples présentant un score de 1 sont très corrélés tandis que les résultats proches de 0 ou négatif ne sont pas corrélés. L'utilisation du Z-score pour estimer la chance d'avoir obtenu un résultat par hasard pour un couple de familles de gènes A et B nécessite de faire hypothèse que la moyenne des résultats obtenus entre A et toutes les autres familles et B et toutes les autres familles relève du hasard.

3.5.1.3 Corrélation/Similarité inférée par les profils phylogénétiques

Corrélation entre états : Coefficient ϕ La première approche pour identifier des corrélations entre les familles de gènes consiste en la comparaison des profils phylogénétiques aux feuilles de l'arbre d'espèces.

Soit deux familles de gène A et B et $S(G)$ la liste ordonnée des états de présences aux feuilles de l'arbre d'espèces pour la famille de gène G.

Soit $\alpha, \beta, \gamma, \delta$ tels que :

$$\alpha = \sum_{i=1}^{n} S_i(A)\ S_i(B) \tag{3.2}$$

$$\beta = \sum_{i=1}^{n} \overline{S_i(A)}\ \overline{S_i(B)} \tag{3.3}$$

$$\gamma = \sum_{i=1}^{n} S_i(A)\ \overline{S_i(B)} \tag{3.4}$$

$$\delta = \sum_{i=1}^{n} \overline{S_i(A)}\ S_i(B) \tag{3.5}$$

Le coefficient de corrélation ϕ entre les distributions de A et B est défini par :

$$\phi_{\mathrm{pP(F)}}(A, B) = \frac{(\alpha\beta - \gamma\delta)}{\sqrt{(\alpha+\gamma)(\alpha+\delta)(\beta+\gamma)(\beta+\delta)}} \tag{3.6}$$

$$\text{Avec } -1 \leq \phi_{\text{pP(F)}}(A, B) \leq 1$$

Les états ancestraux ayant été inférés à partir des profils phylogénétiques, nous avons également calculé cette métrique pour l'ensemble des nœuds et des feuilles de l'arbre d'espèces $\phi_{\text{pP(FN)}}(A, B)$.

Corrélation entre états : Information mutuelle Soit deux familles de gène A et B et $S(G)$ la liste ordonnée des états de présences aux feuilles de l'arbre d'espèces pour la famille de gène G.

La fréquence de l'état S du gène A s'exprime :

$$f_{S(A)} = \frac{1}{n} \sum_{i=1}^{n} S_i(A) \tag{3.7}$$

La fréquence jointe des états $S(A)$ et $S(B)$ s'exprime :

$$f_{S(A),S(B)} = \frac{1}{n} \sum_{i=1}^{n} S_i(A) S_i(B) \tag{3.8}$$

L'information mutuelle entre les profils phylogénétiques de A et B est définie par :

$$MI_{\text{pP(F)}}(A, B) = \sum_{X=\{S(A),\overline{S(A)}\}} \sum_{Y=\{S(B),\overline{S(B)}\}} f_{X,Y} \log \frac{f_{X,Y}}{f_X f_Y} \tag{3.9}$$

$$\text{Avec } 0 \leq MI_{\text{pP(F)}}(A, B) \leq 1$$

Les états ancestraux ayant été inférés à partir des profils phylogénétiques, nous avons également calculé cette métrique pour l'ensemble des nœuds et des feuilles de l'arbre d'espèces $MI_{\text{pP(FN)}}(A, B)$.

Similarité entre événements : Coefficient de Jaccard À partir des états de présence/absence aux nœuds et feuilles de l'arbre, nous avons également implémenté les événements de pertes et de gains. La perte a été inférée si une branche était associée à une absence alors que la branche ancestrale était associée à une présence. Le gain de gène a été défini comme l'inverse de cette condition. Cela nous a permis de calculer la corrélation des événements entre deux familles de gènes.

Soit deux familles de gènes A et B, T l'arbre d'espèces de taille n, i un nœud de T et $E(i, G)$ l'ensemble des événements de la famille de gène G au nœud i.

Le coefficient de Jaccard entre les événements des familles de gène A et B est défini par :

$$J_{\mathrm{eP(FN)}}(A, B) = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{E(i, A) \cap E(i, B)}{E(i, A) \cup E(i, B)} \right) \tag{3.10}$$

$$\text{Avec } E(i, A) + E(i, B) \neq 0$$

$$\text{Et } 0 \leq J_{\mathrm{eP(FN)}}(A, B) \leq 1$$

3.5.1.4 Corrélation/Similarité inférées par les réconciliations

Les événements issus des réconciliations (perte, gain, transfert, duplication) ont été comparés entre chaque couple de famille. De façon importante, les remplacements de gène par perte -transfert ont été divisés en deux événements indépendants de perte et de transfert. Aussi, les transferts « in » (vers la branche) ont été pris en compte mais pas les transferts « out » (à partir de la branche). L'espèce donneuse lors d'un transfert n'a ainsi pas été prise en compte.

Similarité entre événements : Coefficient de Jaccard Soit deux familles de gènes A et B, T l'arbre d'espèces de taille n, i un nœud de T et $E(i, G)$ l'ensemble des événements de la famille de gène G au nœud i. Le coefficient de Jaccard $J_{\mathrm{eR(FN)}}(A, B)$ entre les événements des familles de gène A et B est défini comme dans l'équation 3.10.

Corrélation entre états : Coefficient ϕ et Information mutuelle À partir des données de réconciliation, il est également possible d'inférer les états ancestraux de présence/absence au nœuds profonds de l'arbre d'espèces. Toute spéciation présente dans le fichier RecPhyloXML à un nœud est comptabilisée.

Soit deux familles de gènes A et B et $S(G)$ la liste ordonnée des états de présences aux feuilles et nœuds de l'arbre d'espèces pour la famille de gène G, le coefficient de corrélation $\phi_{\mathrm{pR(FN)}}(A, B)$ entre les distributions de A et B issues des réconciliations est défini comme dans l'équation 3.6. De même, l'information mutuelle $MI_{\mathrm{pR(FN)}}(A, B)$ entre les distributions de A et B issues des réconciliations est défini comme dans l'équation 3.9.

3.5.1.5 Similarité inférée par les synténies de gènes

Enfin nous avons utilisé les informations liées à l'organisation génomique des gènes du cycle cellulaire.

Similarité entre états aux feuilles : Coefficient de Jaccard Tout d'abord, nous avons utilisé l'information aux feuilles de l'arbre d'espèces, c'est-à-dire sans prendre en compte l'information évolutive.

Soit deux familles de gènes A et B, $P(G)$ l'ensemble des états de présence du gène G au sein des feuilles de l'arbre d'espèces T, et $N(A, B)$ l'ensemble des états de synténie entre les deux familles A et B au sein des feuilles de l'arbre d'espèces T.

Le coefficient de Jaccard exprimant la synténie entre A et B est défini par :

$$J_{\mathrm{pS(F)}}(A, B) = \frac{P(A) \cap P(B) \cap N(A, B)}{P(A) \cup P(B) \cup N(A, B)} \tag{3.11}$$

$$\text{Avec } 0 \leq J(S(A, B)) \leq 1$$

Similarité entre états aux feuilles et nœuds : Coefficient de Jaccard Nous avons également pris en compte les états ancestraux inférés par parcimonie. Pour cela, nous avons inféré l'état de présence/absence de synténie entre chaque couple pour chaque nœud de l'arbre d'espèces. Néanmoins, pour appliquer le même calcul que pour le $J_{\mathrm{pS(F)}}(A, B)$ précédemment décrit, il est nécessaire d'avoir connaissance des états de présence/absence des gènes aux nœuds internes. Pour cela, nous avons utilisé les états ancestraux inférés par les profils phylogénétiques plutôt que ceux issus des réconciliations car basés sur le même modèle et les mêmes coûts. Les états de vicinité entre les gènes ont ainsi été comparés.

Soit deux familles de gènes A et B, $P(G)$ l'ensemble des états de présence du gène G (avérés ou inférés par parcimonie) au sein des branches et feuilles de l'arbre d'espèces T, et $N(A, B)$ l'ensemble des états de synténie entre les deux familles A et B au sein de l'arbre d'espèces T, le coefficient de Jaccard exprimant la synténie entre A et B est défini comme dans l'équation 3.11.

3.5.1.6 Normalisation des coefficients de corrélation/similarité

L'ensemble de ces coefficients de corrélation/similarité ne prennent pas en compte la probabilité d'être obtenu par le hasard. Il est donc nécessaire de comparer ces valeurs avec une valeur correspondant à une valeur attendue sous l'hypothèse nulle. Étant donné le grand nombre de familles comparée, on peut considérer la valeur moyenne du coefficient de corrélation/similarité d'une famille de gène avec les autres familles de gène comme étant une estimation acceptable de la valeur attendue sous l'hypothèse nulle. Dans ce contexte, une métrique intéressante pour estimer la significativité d'une valeur de coefficient de corrélation/similarité est le Z-score. Plus précisément, nous avons calculé la moyenne entre les Z-scores de chaque membre du couple de famille de gènes.

Pour les coefficient de Jaccard :

Soit la moyenne arithmétique μ et l'écart type σ des coefficients de Jaccard pour la famille

de gènes G et l'ensemble des familles de gènes F :

$$\mu_{\mathrm{J}}(G,F) = \frac{1}{|F \setminus G|} \sum_{f \in \{F \setminus G\}} J(G,f) \tag{3.12}$$

$$\sigma_{\mathrm{J}}(G,F) = \sqrt{\frac{1}{|F \setminus G|} \sum_{f \in \{F \setminus G\}} J(G,f)^2 - \left(\frac{1}{|F \setminus G|} \sum_{f \in \{F \setminus G\}} J(G,f) \right)^2} \tag{3.13}$$

Le Z-score moyen s'exprime :

$$Z_{\mathrm{J}}(A,B,F) = \frac{1}{2} \left(\frac{J(A,B) - \mu_J(A,F)}{\sigma_J(A,F)} + \frac{J(A,B) - \mu_J(B,F)}{\sigma_J(B,F)} \right) \tag{3.14}$$

Quatre Z-scores de coefficients de Jaccard ont été calculés :

- $Z_{\mathrm{eR(FN)}}(A,B,F)$
- $Z_{\mathrm{eP(FN)}}(A,B,F)$
- $Z_{\mathrm{pS(F)}}(A,B,F)$
- $Z_{\mathrm{pS(FN)}}(A,B,F)$

Pour les informations mutuelles et coefficients ϕ :

Soit le coefficient mixte de co-occurrence tel que :

$$C(A,B) = \frac{1}{2} |\phi(A,B)| + MI(A,B) \tag{3.15}$$

Soit la moyenne arithmétique μ et l'écart type σ des coefficients mixtes de co-occurrence pour la famille de gènes G et l'ensemble des familles de gènes F :

$$\mu_C(G,F) = \frac{1}{|F \setminus G|} \sum_{f \in \{F \setminus G\}} C(G,f) \tag{3.16}$$

$$\sigma_C(G,F) = \sqrt{\frac{1}{|F \setminus G|} \sum_{f \in \{F \setminus G\}} C(G,f)^2 - \left(\frac{1}{|F \setminus G|} \sum_{f \in \{F \setminus G\}} C(G,f) \right)^2} \tag{3.17}$$

Le Z-score moyen s'exprime :

$$Z_{\mathrm{C}}(A,B,F) = \frac{1}{2} \left(\frac{C(A,B) - \mu_C(A,F)}{\sigma_C(A,F)} + \frac{C(A,B) - \mu_C(B,F)}{\sigma_C(B,F)} \right) \tag{3.18}$$

Trois Z-scores de coefficients mixtes ont été calculés :

- $Z_{\mathrm{pP(F)}}(A,B,F)$
- $Z_{\mathrm{pP(FN)}}(A,B,F)$
- $Z_{\mathrm{pR(FN)}}(A,B,F)$

Les Z-scores ont ensuite été normalisés à partir de la valeur maximale négative pour les valeurs négatives et positive pour les valeurs postives.

$$\text{D'où } -1 \leq Z_{\mathrm{C}}(A,B,F) \leq 1$$

3.5.1.7 Distribution et comparaison des scores de corrélation/similarité

La distribution des Z-scores de corrélation/similarité sous forme de graphe de densité est présentée par type d'information, à savoir les profils absence/présence (figure 3.37), les événements (figure 3.38) et les synténies (figure 3.39). Elle correspondent aux courbes rouges. Afin d'estimer la possible redondance des métriques, nous avons aussi calculé la différence des

valeurs de scores de corrélation/similarité pour chaque paire de familles de gènes entre les différentes méthodes. Les distributions des différences entre les métriques sont représentées par les courbes jaunes. La distribution des scores basés sur l'état d'absence/présence par les trois méthodes (pP(FN), pR(FN) et pP(F)) montre que la grande majorité des couples présente une valeur de score de corrélation/similarité proche de zéro ou négatif (figure 3.37). Une infime fraction présente des scores supérieurs à 0,5. Une allure de courbe très similaire pour le pP(FN) et le pR(FN) avec un épaulement droit qui correspond à un score de zéro est constatée. La distribution des différences entre ces deux dernières méthodes montre que les méthodes donnent des résultats très proches avec des différences maximale de 0,2 et une valeur moyenne de 0,015. La distribution de pP(F) semble différente des deux autres méthodes. La comparaison entre pP(F) et pP(FN) ou pR(FN) montre de plus grandes variations dans les résultats générés par cette approche. En effet, la moyenne de la différence entre pP(F) et pP(FN) et pP(F) et pR(FN) est de 0,07. Il apparaît donc qu'utiliser les états aux branches internes de l'arbre d'espèces modifie de façon conséquente les résultats. Les paires de familles de gènes présentant une grande différence dans les deux approches semblent être des familles très conservées dont le nombre de génomes ne présentant pas une copie est très rare mais également les familles très peu conservées. Le fait de prendre en compte les états ancestraux semble diminuer le signal de corrélation porté par ces absences. Parmi ces paires de familles de gènes, certains couples semblent pourtant faire sens biologiquement (*e.g.* StkP avec PhpP absents tous deux uniquement chez *Lactobacillus sanfranciscensis* TMW 1.1304) tandis que de nombreux autres couples ne semblent *a priori* faire aucun sens biologique (*e.g.* SunL avec GlmM absents tous deux uniquement chez les *Oenococcus*). La prise en compte des états ancestraux dans la mesure de la corrélation des états de présence/absence semble donc diminuer la valeur des coefficients de corrélation des familles dont le rapport absence sur présence avoisine zéro ou un.

La distribution des Z-scores de similarité basés sur les événements (eP(FN) et eR(FN)) est centrée sur zéro avec une faible quantité de couples présentant des scores élevés (figure 3.38). Il est intéressant de noter qu'une grande partie des scores sont négatifs, indiquant que les familles comprises dans ces couples présentent des valeurs du coefficient de Jaccard plus faibles que la moyenne des coefficients des couples incluant les deux familles concernées. Le cas

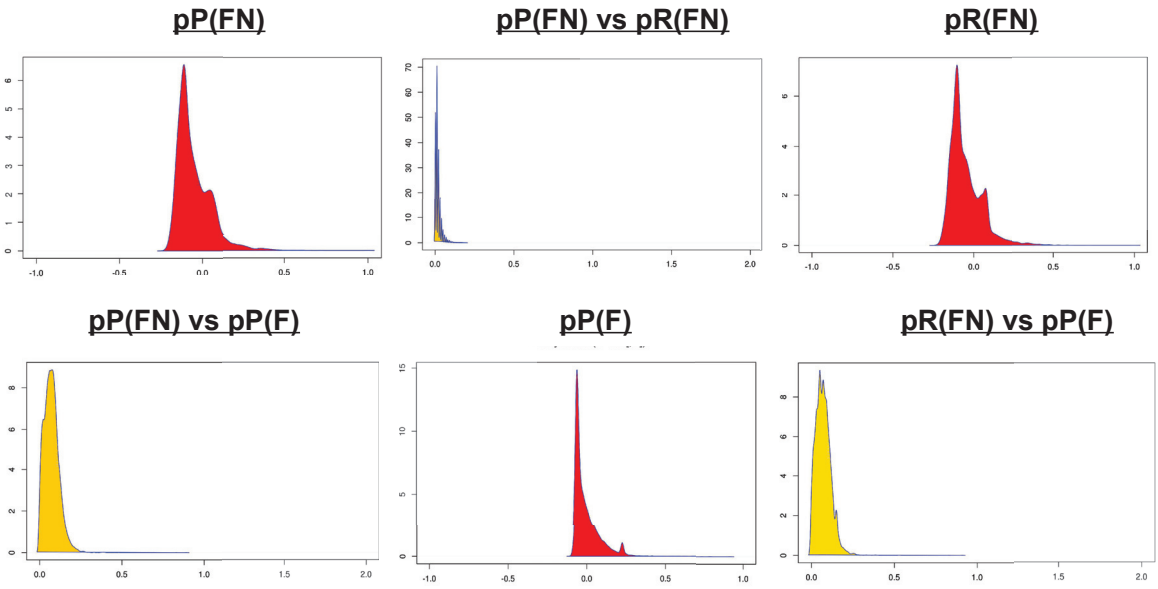


FIGURE 3.37 – Distribution des Z-score de corrélation par les profils de présence/absence. Densité en fonction du Z-score. Les courbes rouges correspondent aux distributions des Z-score, les courbes jaunes aux distributions des différences de Z-score entre deux méthodes par couple de familles de gènes.

typique est celui d'un couple avec un coefficient de similarité faible d'une famille qui corrèle avec de nombreuses autres familles et d'une famille qui ne corrèle avec pratiquement aucune famille. En effet, en reprenant l'équation du Z-score 3.14, le Z-score de la famille qui ne corrèle pratiquement pas avec d'autres familles est proche de zéro tandis que celui de l'autre famille est négatif (la valeur du coefficient étant plus faible que celle du coefficient moyen). Les familles ne corrélant que peu avec d'autres familles sont typiquement celles qui présentent une distribution taxonomique très restreinte comme MGT1, LysA2, LytB ou encore MurM.
La comparaison entre les deux méthodes eP(FN) et eR(FN) montre des différences (distance maximale de 1,74 et moyenne de 0,068, figure 3.38). Ces différences s'expliquent par le fait que les deux méthodes ne se basent pas sur les mêmes données (arbre phylogénétique et profil phylogénétique) ni sur les mêmes modèles (événements de gains et pertes pour les profils phylogénétiques et événements de gains, pertes, duplications et transferts horizontaux pour la réconciliation).

Pour les événements et les états d'absence/présence provenant de la même information (eP(FN) et pP(FN), eR(FN) et pR(FN)), il est intéressant de comparer les résultats entre ces deux approches. Des différences sont observées (eP(FN) et pP(FN) : maximum = 1,87, moyenne = 0,088 ; eR(FN) et pR(FN) : maximum = 0,82, moyenne = 0.067) montrant ainsi qu'observer les états de présence/absence ou les événements de gènes ne donnent pas les mêmes résultats.

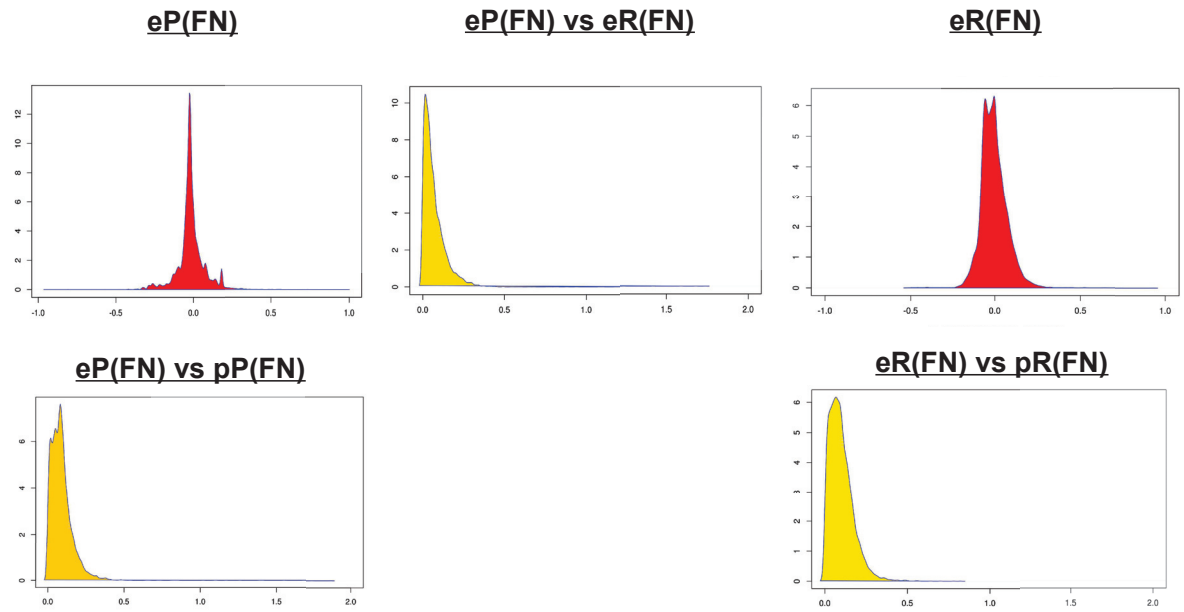


FIGURE 3.38 – Distribution des Z-score de similarité par les événements. Densité en fonction du Z-score. Les courbes rouges correspondent aux distributions des Z-score, les courbes jaunes aux distributions des différences de Z-score entre deux méthodes par couple de familles de gènes.

Enfin, l'analyse des synténies par les méthodes pS(F) et pS(FN) montrent une distribution centrée sur zéro avec une fraction infime de scores supérieurs à 0,1 (figure 3.39). En effet, une densité très élevée correspondant au score nul est constatée. Il semble donc que la synténie soit plus discriminante dans les corrélations/similarités entre familles de gènes que les deux autres méthodes.

La comparaison entre les deux méthodes montre que la majorité des couples présentent une différence de score de similarité égale à zéro. Ainsi, la prise en compte de l'état de synténie aux branches internes de l'arbre d'espèces ne semble pas induire des différences majeures dans

le calcul de similarité entre les familles de gènes basé sur la synténie.

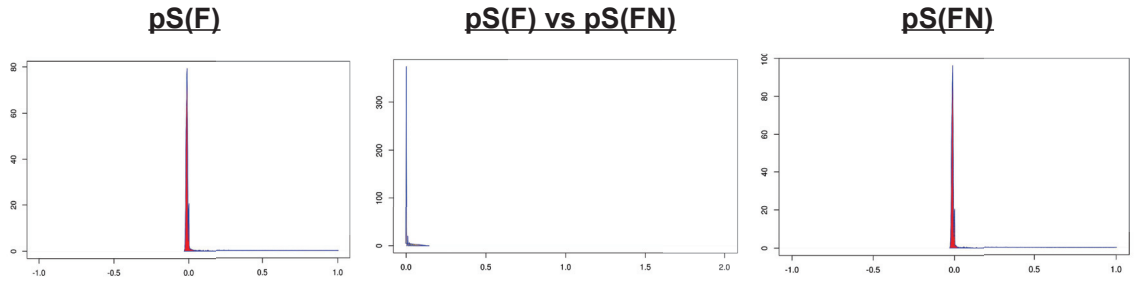


FIGURE 3.39 – Distribution des Z-score de similarité par les synténies. Densité en fonction du Z-score. Les courbes rouges correspondent aux distributions des Z-score, la courbe jaune à la distribution des différences de Z-score entre les deux méthodes par couple de familles de gènes.

3.5.1.8 Score intégré de corrélation

Afin de faciliter l'exploitation des résultats de corrélation/similarité entre les familles de gènes, nous avons décidé de calculer un score global prenant en compte les événements de gènes, les profils absence/présence de gènes et les synténies. Il est à noter que pour des raisons de clarté, nous parlerons de score de corrélation et non de score de corrélation/similarité.
Pour chaque type d'information, nous avons utilisé différentes méthodes :

- Profils absence/présence
  - pP(F)
  - pP(FN)
  - pR(FN)
- Événements
  - eP(FN)
  - eR(FN)
- Synténie
  - pS(F)
  - pS(FN)

Nous avons donc du choisir la méthode à utiliser. Concernant les événements, nous avons préféré utiliser les données de réconciliation (eR(FN)) plutôt que ceux des profils car étant plus complètes en terme de diversité d'événements et étant plus justifier d'un point de vue évolutif. Concernant les états de présence/absence de gènes, nous avons décidé d'utiliser le pP(FN) car cette approche prend en compte la phylogénie d'espèces. Enfin, les données de synténies présentant sensiblement les mêmes résultats, nous avons décidé de prendre la pS(FN) qui a l'avantage de prendre en compte les états aux branches internes. Ce choix aurait pu être optimisé notamment en analysant le comportement de l'ensemble des combinaisons mais nous n'avons pas pu par manque de temps.
Nous avons donc calculé un score moyen qui intègre les scores de corrélation issus des trois types de données événements, états de présence/absence et synténies, tel que :

$$Z_{\text{score}_{\text{g}}\text{lobal}}(A,B,F) = \frac{1}{3}\left(Z_{\text{pP(FN)}}(A,B,F) + Z_{\text{eR(FN)}}(A,B,F) + Z_{\text{pS(FN)}}(A,B,F)\right) \quad (3.19)$$

Les valeurs négatives ont été rapportées à zéro tel que :

$$0 \leq Z_{\text{score}_{\text{g}}\text{lobal}}(A,B,F) \leq 1$$

La distribution des scores moyens est présentée figure 3.40A. Les scores moyens ont également été convertis en distances afin de créer une matrice de distances entre les familles de gènes. Les distances ont été normalisées par la transformation arc sinus [73], tel que :

$$d(A,B) = Arcsin(1 - Z_{\text{score}_{\text{g}}\text{lobal}}(A,B,F)) \quad (3.20)$$

Les arbres de distances ont été inférés par Fastme 2.1.5 en utilisant la matrice de distance générée précédemment (méthode de « Neighbor Joining », NJ) [266].

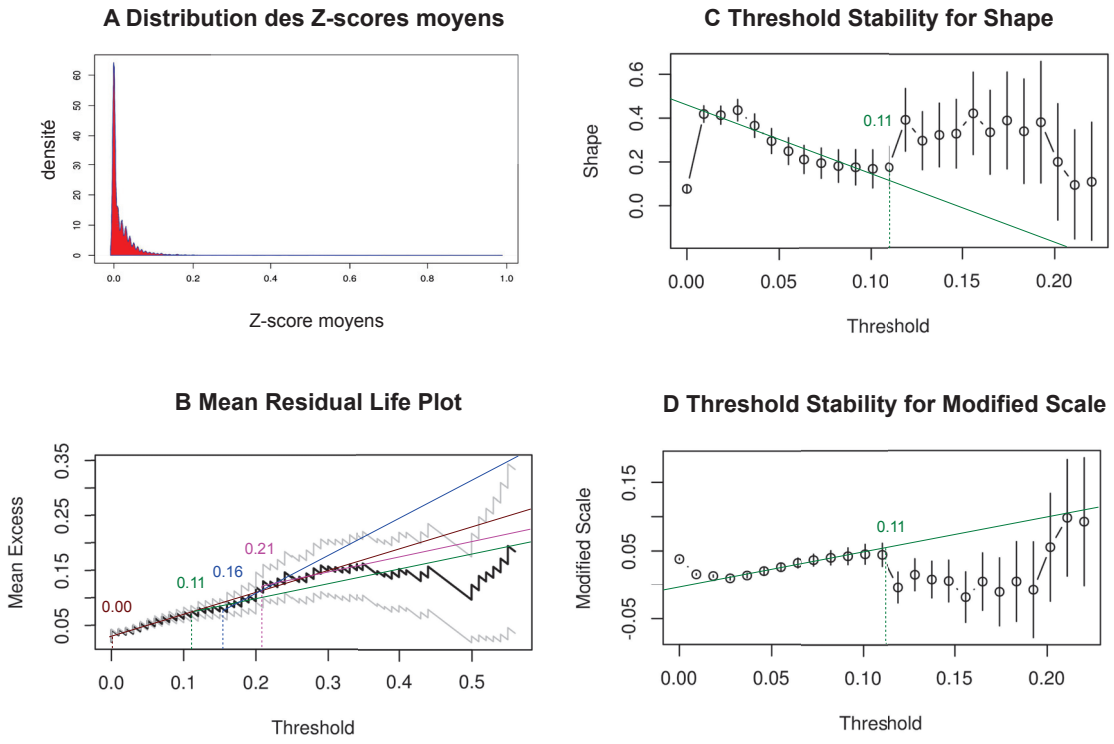


FIGURE 3.40 – Sélection du seuil de score moyen. (A) La distribution des scores moyens suit une loi de puissance. (B) Graphe MRL. Les droites colorées correspondent à des portions linéaires, les nombres correspondent au seuil associé à la partie basse de la portion linéaire. (C) Graphe TSS. La linéarité semble perdue vers 0,11. (D) Graphe TSMS. La linéarité semble également perdue vers 0,11.

3.5.1.9 Sélection du seuil de score moyen

La représentation la plus simple pour décrire les liens inférés entre les familles de gènes est celle d'un graphe dont les nœuds représentent les familles de gènes et les arêtes les liens évolutifs ou fonctionnels. Néanmoins, cette représentation nécessite l'utilisation d'un seuil pour ne retenir que les arêtes porteuses d'une information et éliminer celles qui reflètent le bruit de fond. Pour estimer le seuil de score moyen le plus adéquat, nous nous somme tout d'abord

demandé quelle était la nature de la distribution des scores moyens de corrélation. Il semble par l'appréciation qualitative de la distribution que celle-ci suit une loi de puissance (figure 3.40A). Cette distribution peut être qualifiée de distribution de Pareto [345]. La distribution de Pareto généralisée (GPD) est une distribution classique observée dans de nombreux champs scientifiques (sociologie, biologie, météorologie, *etc.*.) qui peut être décomposée en deux parties : la tête correspondant à la majorité des observations centrées vers zéro et la queue qui correspond aux valeurs extrêmes. Cette distribution a déjà été décrite pour des systèmes biologiques telles que les réseaux métaboliques [222] ou les réseaux d'interaction protéine-protéine [221]. Le test de Kolmogorov-Smirnov sur le jeu de données des scores moyens indique une p-value de 0,76, indiquant ainsi que l'adéquation du jeu de données avec le modèle de la GPD n'est pas rejeté [309].

Plusieurs techniques ont été développées afin d'estimer le seuil pour lequel les valeurs sont considérées comme extrêmes à partir d'une distribution de Pareto GPD. C'est justement ces valeurs qui présentent un intérêt pour la construction du graphe puisque elles correspondent à des couples plus corrélés que la majorité des couples. La technique proposée par Coles est d'interpréter les graphes MRL (« Mean Residual Life »), TSS (« Threshold Stability for Shape ») et TSMS (« Threshold stability for Modified Scale ») [69]. Nous n'allons pas décrire les fondements de cette techniques mais simplement son application. Le MRL correspond grossièrement à l'adéquation des données extrêmes avec le modèle de Pareto généralisé en fonction du seuil de valeurs extrêmes. Nous avons donc tracé ces graphes à partir des scores moyens de corrélation (figure 3.40). Pour trouver le seuil optimal sur ce graphe, il faut identifier les portions linéaires. La borne inférieure de cette portion linéaire correspond au seuil optimal. Néanmoins, ce graphe est très difficile à interpréter puisque plusieurs portions semblent linéaires et que par conséquent, plusieurs seuils peuvent être choisis (figure 3.40B).

Pour affiner le choix du seuil, il est possible d'utiliser les TSS et TSMS qui correspondent à la stabilité des paramètres du modèle à partir des données extrêmes sélectionnées par le seuil en fonction du seuil (figure 3.40CD). Sur ces graphes, on recherche une valeur de seuil à partir de laquelle les paramètres ne sont plus stables (perte de linéarité). Les TSS et TSMS des coefficients moyens de corrélation présentés en figure nous indiquent une perte de linéarité pour un seuil de 0,11. Il semble donc justifiable d'utiliser un seuil de 0,11. Pour tester l'adéquation

entre les valeurs extrêmes sélectionnées par le seuil de 0,11 et la distribution de Pareto, nous avons tracé les graphes PP (« Probability Plot »), QP (« Quantile Plot »), RLP (« Return Level Plot ») et DP (« Density Plot ») (figure 3.41). Ceux-ci indiquent que le choix parait raisonnable, puisque les modèles (lignes bleues) semble être en adéquation avec les distributions observées. L'utilisation du seuil de 0,11 permet d'obtenir un jeu de données de 539 scores de corrélation moyens.
Le choix du seuils a été effectuée par l'utilisation des librairies R. L'adéquation de la la distribution avec la loi de puissance (ou loi de Pareto généralisée) a été testée par le test de Kolmogorov-Smirnov *via* la fonction fit_power_law de la librairie igraph. La sélection du seuil a été effectuée d'après la théorie des extrêmes en traçant les graphes MRL (« Mean Residual Life »), TSS (« Threshold Stability for Shape ») et TSMS (Threshold Stability for Modified Scale ») par les fonctions mrlplot et tcplot de la librairie POT. L'adéquation de la distribution des valeurs extrêmes pour le seuil avec la distribution de Pareto généralisée a été testée par les fonctions gpd.fit et gpd.diag de la librairie ismev. Le graphe de corrélation a été générés à l'aide de Cytoscape 3.5.1 [427].

3.5.2 Liens fonctionnels entre familles de gènes du cycle cellulaire

Afin de représenter les corrélations évolutives entre les familles de gènes du cycle cellulaire, nous avons utilisé plusieurs types de graphes. Cette analyse nous a permis de mettre en lumière des clans de familles de gènes qui présenteraient potentiellement des liens fonctionnels. Cette partie décrit de façon générale les différents graphes et les liens forts entre familles de gènes. L'histoire évolutive détaillée de l'ensemble des clans et les implications fonctionnelles sont décrites section 3.6.

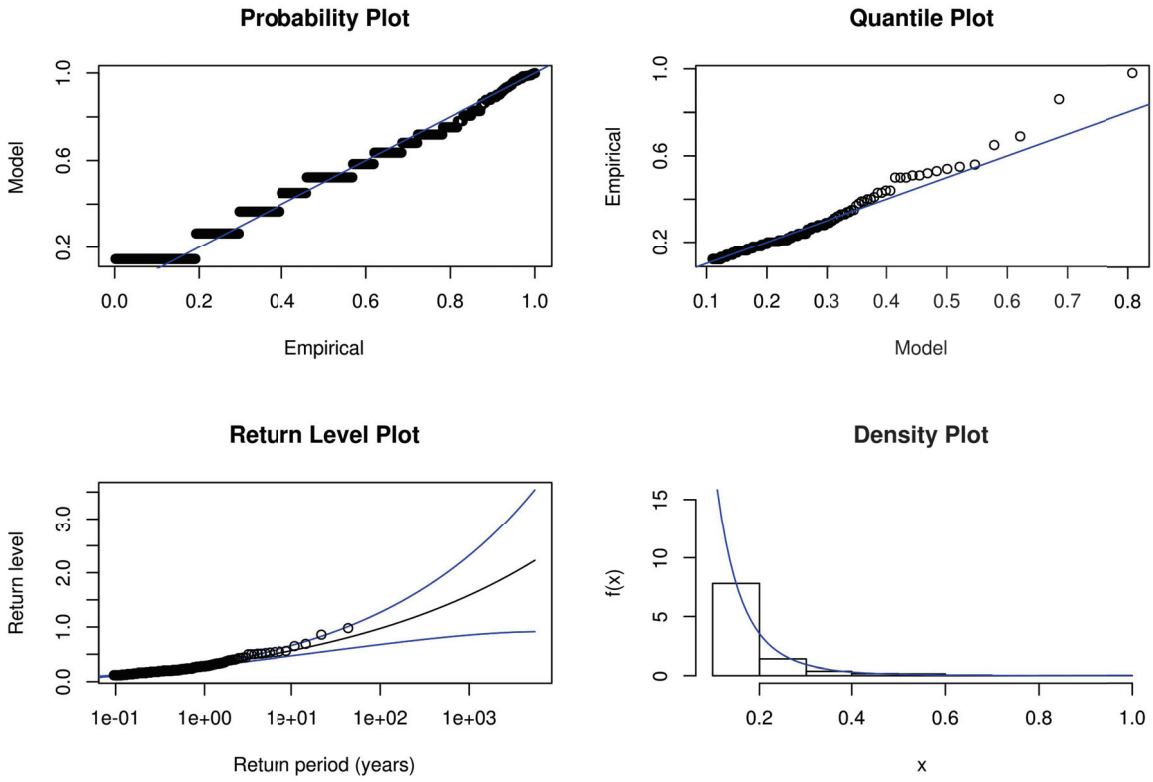


FIGURE 3.41 – Adéquation des valeurs extrêmes avec la GPD. Les quatre graphes indiquent que la distribution des valeurs extrêmes avec le seuil de 0,11 est en adéquation avec la GPD. Les courbes bleues représentent le modèle GPD, les points noirs les valeurs extrêmes.

3.5.2.1 Représentation des liens évolutifs en réseau

Le graphe obtenu est présenté figure 3.42. Tout d'abord, la majorité des familles de gènes sont liées par au moins une arête (177/179). Seules les familles LytB et MGT2 ne sont corrélées à aucune autre famille de gènes. Ceci s'explique par le fait que ces deux familles possèdent des distribution taxonomiques très restreintes et leurs gènes sont isolés dans les génomes, c'est-à-dire qu'ils n'appartiennent à aucun des clusters décrits.

Plusieurs clans de familles de gènes hyperconnectées peuvent être observés sur le graphe

Clan	Familles de gènes (réseaux)	Familles de gènes (arbre)
Mre/Min	MreC, MreD, MreB, MinC, MinD, MinE, MinJ, Maf, ValS, RodA, RadC, Mbl, B5	MreC, MreD, MreB, MinC, MinD, MinE, Maf, ValS, RodA, RadC, B5
Ori1	DnaA, RecF, DnaN	DnaA, RecF, DnaN
Ori2	FtsJ, RecN, SpoIIIJ1, GidA, GidB, ParA, ParB, Noc, Jag	FtsJ, RecN, SpoIIIJ1, GidA, GidB, ParA, ParB, Noc, Jag
StkP	StkP, PriA, PhpP, Mtf, SunL	StkP, PriA, PhpP, Mtf, SunL
Fem	FemA, FemB, fmh, MGT1	FemA, FemB, fmh, MGT1
Cps	CpsB, CpsC, CpsD	CpsB, CpsC, CpsD
FtsH	FtsH, Mfd, YabM, TilS, DivIC, SpoIIE	FtsH, Mfd, YabM, TilS, DivIC
Scp/Xer/Smc	ScpA, ScpB, XerC, XerD,Smc,FtsY	ScpA, ScpB, XerC, XerD
DCW1	FtsZ, FtsA, MurF, MurE, FtsQ, FtsL, MurD, MurG1, MraZ, MraY, MraW, B4	FtsZ, FtsA, FtsQ, FtsL, MurD, MurG1, MraZ, MraY, MraW, B4
DWC2	DivIVA, IleS1, SepF, PCDP6, PCDP7, PCDP8	DivIVA, IleS1, SepF, PCDP6, PCDP7, PCDP8
Spo	SpoIIE, SpoVB, SpoIIGA, SpoVE, SpoIID, SpoIIID, DacB, DacF, B6,MurA2, CozE4	SpoIIE, SpoVB, SpoIIGA, SpoVE, SpoIID, SpoIIID, DacB, DacF, B6,MurA2, CozE4, FtsH3, Mbl, MurA1
Wal	WalH, WalI, WalJ, WalK, WalR, MurZ	WalH, WalI, WalJ, WalK, WalR, MurZ
Bacilli	RecU, GpsB, A3, PCDP10, CDP1, CDP4, DacA, EzrA, DnaI, DnaBBS, MurJ, MacP, SpoIIIJ2, BX1, MinE	RecU, GpsB, A3, PCDP10, CDP1, CDP4, DacA, EzrA, DnaI, DnaBBS, MurJ, MacP, Pfs
Nag/Mur	NagA, NagB, MurQ, MurK , AnmK	NagA, NagB, MurQ, MurK , AnmK

TABLE 3.4 – Composition en familles de gènes des clans par le réseau et l'arbre. Les familles qui diffèrent au sein d'un même clan entre les deux méthodes sont indiqués en vert.

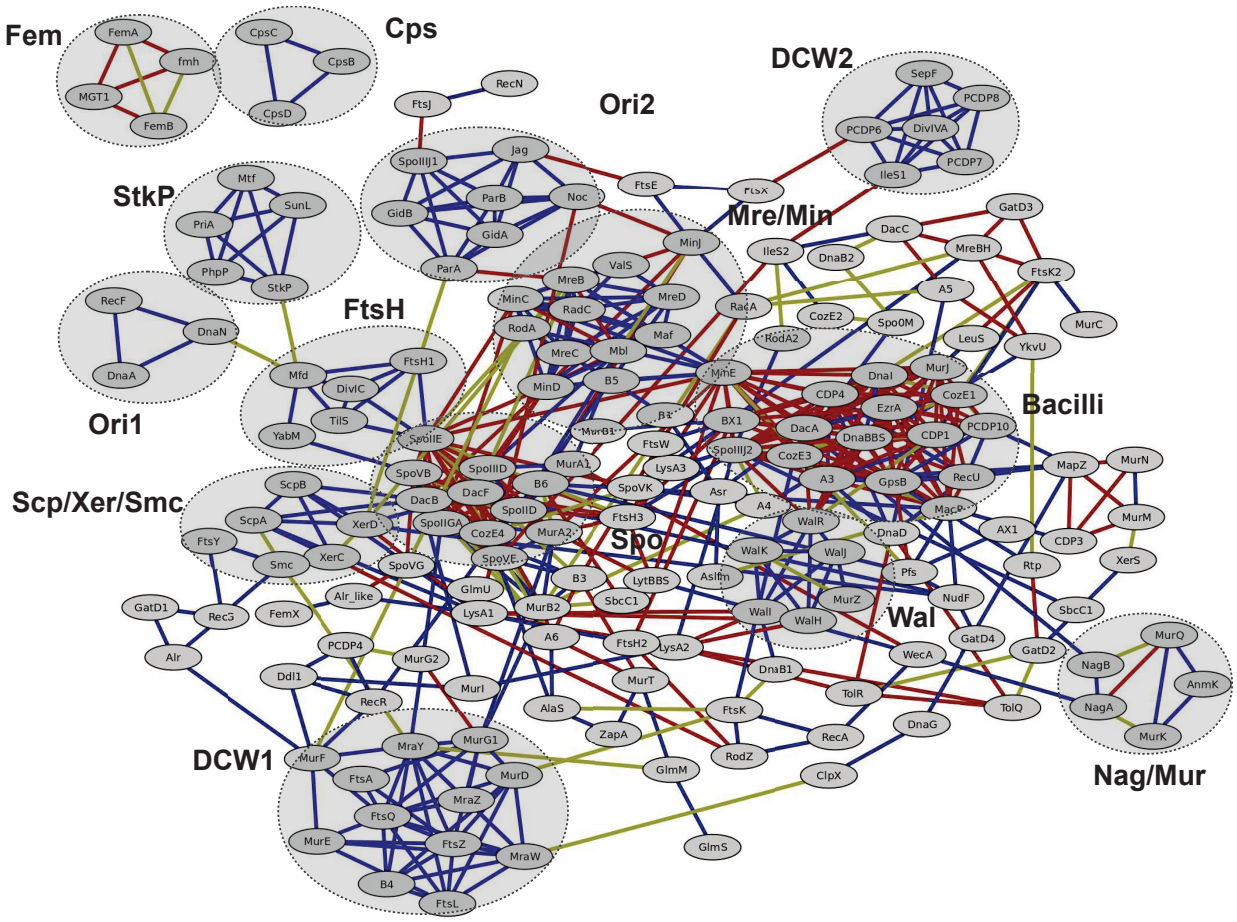


FIGURE 3.42 – Réseau de relations évolutives entre familles du cycle cellulaire. Les ellipses représentent les familles de gènes, les ellipses grisées représentent les clans, les liens évolutifs correspondent aux arêtes. Arêtes bleues : synténie majoritaire, arêtes rouges : co-occurrence majoritaire, arêtes jaunes : co-événements majoritaires.

(figure 3.42, table 3.4). Les clans ont été nommés en fonction de la littérature (*e.g.* DCW, Division and Cell Wall), des familles de gènes présents dans le clan (*e.g.* Fem), de la famille majeure du cycle cellulaire présent dans le clan (*e.g.* StkP) ou de l'émergence des familles de gènes contenus dans le clan (*e.g. Bacilli*). De façon générale, trois types de clans en fonction du type de lien majoritaire entre les composants peuvent être distingués sur la figure 3.42. Le premier type correspond aux clans dont les familles de gènes sont reliés par la distribution taxonomiques : les clans Spo et *Bacilli*. Le deuxième type correspond à des clans principalement liés par le contexte génomique : Les clans, Ori 1 et 2, DCW 1et 2, StkP, FtsH, Mre/Min, Wal. En-

fin, certains clans sont constitués de liens mixtes issus du contexte génomique, des événements évolutifs et de la distribution taxonomique : les clans Scp/Xer/Smc, Fem, Cps et Nag/Mur. Le même réseau de corrélation est présenté en figure 3.43, mais dont la couleur des arêtes représentent la force de corrélation. Certaines arêtes sont très marquées du fait que l'ensemble des informations évolutives suggèrent un lien évolutif entre les familles de gènes concernées. Ainsi, les trois clusters mixtes Fem, Cps et Nag/Mur décrits précédemment semblent étroitement associés, ce qui suggère des liens fonctionnels. D'autres liens ponctuels très marqués ne formant pas de clan à proprement parler sont retrouvés. Ainsi, FtsE est très corrélé à FtsX (score=0,56), de même pour TolR et TolQ (score=0,98), Asr et Alsfm (score=0,50), MurM et MurN (score=0,65) et FtsW et BX1 (score=0,43).

3.5.2.2 Représentation des liens évolutifs en arbre de distances

À partir de la matrice de distances entre les familles de gènes issus des scores moyens de corrélation, un arbre de distance a été généré (figure 3.44). Cette méthode a l'avantage de ne pas nécessiter l'utilisation d'un seuil et d'intégrer l'ensemble des scores moyens mais ne rend pas compte du détail des liens évolutifs. Les clans identifiés dans le réseau de corrélation sont très similaires à ceux retrouvés sur l'arbre de distances. Néanmoins, quelques différences sont retrouvées (table 3.4). Ainsi, MinJ et Mbl ne se situent plus dans le clan Mre/Min. MinJ est groupé avec FtsE et FtsX tandis que Mbl se situe dans le clan Spo. Les familles SpoIIE et SpoVB ne sont pas non plus retrouvées dans le clan FtsH mais au sein du clan Spo. En effet, ces deux familles appartiennent au clan FtsH par le contexte et au clan Spo par la distribution taxonomique dans le réseau de corrélation. Ainsi, il semble que le signal fourni par la distribution taxonomique soit plus fort que pour le contexte génomique. En ce qui concerne le clan DCW1, les familles MurE et MurF sont situées plus basalement dans l'arbre et sont donc exclues du clan. Enfin, le clan *Bacilli* n'est plus composé des familles SpoIIIJ2, BX1 et MinE. SpoIIIJ2 se place avec LysA3, BX1 avec FtsW et MinE au sein du clan Mre/Min. Ces groupements sont induits par les hauts coefficients de corrélation issus des synténies (SpoIIIJ2-LysA3 : 0,534 ; BX1-FtsW : 0,916 ; MinE-MinC : 0,344).

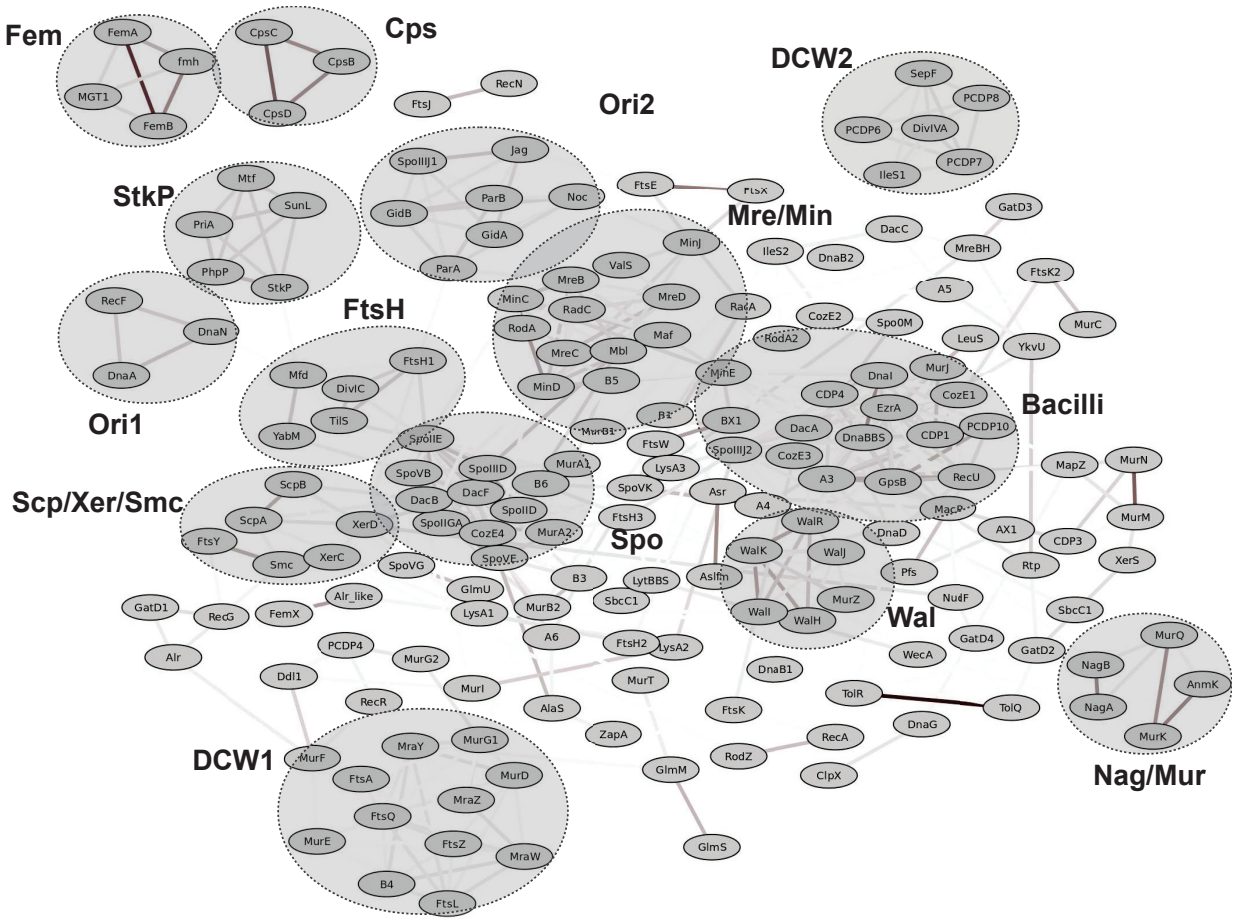


FIGURE 3.43 – Réseau de relations évolutives entre familles du cycle cellulaire. Les ellipses représentent les familles de gènes, les ellipses grisées représentent les clans, les liens évolutifs correspondent aux arêtes. La couleur des arêtes représente la quantité totale de signal de co-évolution.

3.5.2.3 Validation qualitative de l'inférence des liens fonctionnels

Choix de la méthode de validation L'approche précédemment décrite a permis de quantifier les corrélation évolutives entre les familles de gènes du cycle cellulaire. Ces corrélations évolutives peuvent théoriquement être le reflet de liens fonctionnels au sein des cellules bactériennes. Néanmoins, cette méthode n'ayant jamais été utilisée de la sorte dans la littérature (utilisation simultanée des différents types d'information évolutive, calcul des Z-scores, choix du seuil, *etc* ...), il est nécessaire de la valider. Il est possible de valider quantitativement la

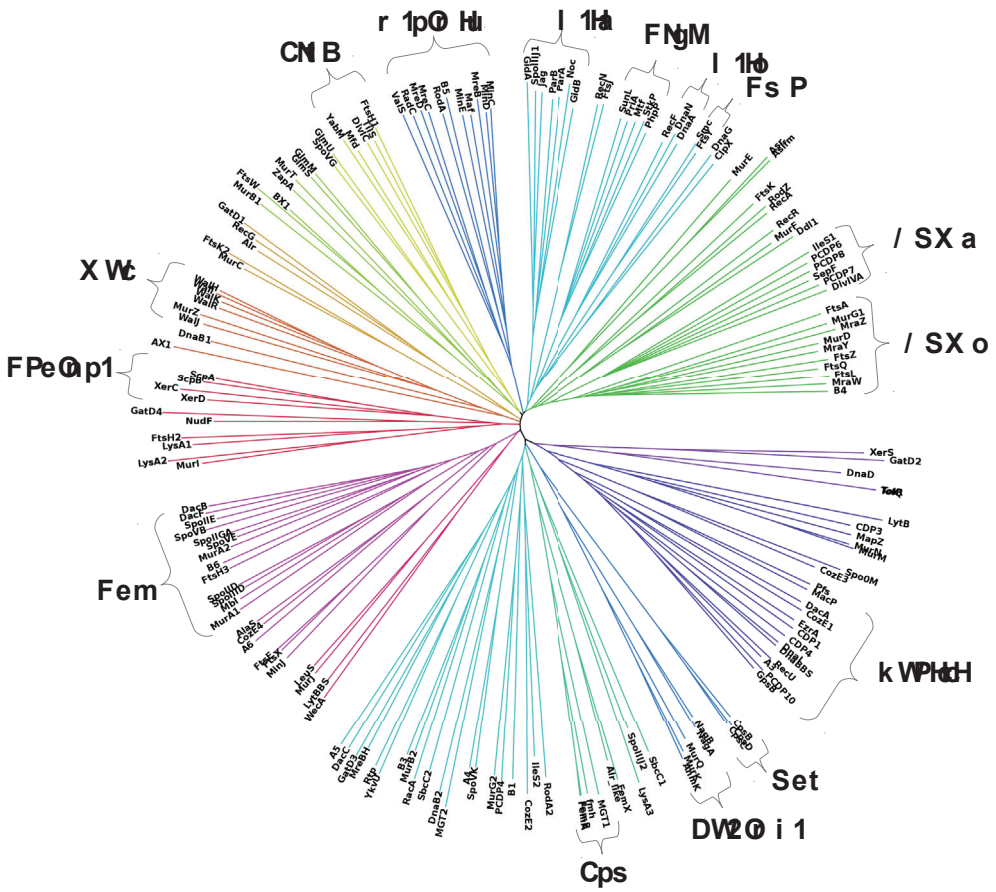


FIGURE 3.44 – Arbre de distances de co-évolution des familles du cycle cellulaire. La couleur des branches correspond aux groupes inférés par la topologie. Les clans sont représentés par des accolades.

qualité d'une méthode en estimant sa sensibilité et sa spécificité en calculant le nombre de faux positifs et de faux négatifs. Cependant, cette validation dans notre cas nécessite l'obtention d'un jeu de données de référence qui inclurait la présence et l'absence de liens fonctionnels démontrés. Il est très difficile de créer un tel jeu de données pour plusieurs raisons.

Tout d'abord, le choix de la nature des liens fonctionnels entre deux familles de gènes serait délicat. En effet, il pourrait s'agir d'interaction protéine-protéine, de la participation à un même complexe ou de la même localisation cellulaire, ... Les interactions protéine-protéine sont les liens fonctionnels les plus utilisés car faciles à obtenir grâce à de nombreuses méthodes de détection mais ne couvrent pas la totalité de l'information fonctionnelle qui existe potentielle-

ment entre deux protéines au sein d'une cellule.

De plus, il est communément admis que les interactions protéine-protéine ne sont pas conservées au sein des espèces [299]. À cela s'ajoute le manque de données sur les non-interactions protéine-protéine bien que quelques bases de données aient été développées à cet effet [40]. Face à ces difficultés, nous avons décidé de recourir à une approche qualitative. Nous avons donc mis en lumière un certain nombre de liens fonctionnels entre des familles de gènes obtenus par notre approche qui sont décrits dans la littérature, estimant ainsi la pertinence de notre approche. Plusieurs des systèmes moléculaires du cycle cellulaire décrits dans la littérature et dans le chapitre 1 ont été retrouvés par notre approche corrélative, soulignant ainsi l'intérêt de notre approche. Nous allons décrire ici ces liens par système cellulaire.

Réplication et ségrégation des chromosomes Plusieurs familles de gènes impliquées dans la réplication semblent reliées, notamment au sein du clan Ori1 (table 3.4). Le lien fonctionnel entre DnaA et DnaN a été prouvé puisque DnaN est impliqué directement dans la régulation de DnaA chez *B. subtilis* et *E. coli* [191],[57]. Le potentiel lien entre DnaA, DnaN et RecF a aussi été décrit puisqu'il semblerait que les trois gènes soient co-transcrits chez *E. coli* [373]. Concernant la ségrégation des chromosomes, trois clans sont retrouvés dont certains liens fonctionnels internes ont déjà été décrits. Au sein du clan Ori2 (table 3.4), les deux familles ParA et ParB semblent reliés fonctionnellement. Ces deux familles forment un système bien décrit impliqué dans la ségrégation de l'origine de réplication chez de nombreuses espèces [151] [23] [214] [341]. Au sein du clan Scp/Xer/Smc, ScpA et ScpB semblent reliés évolutivement par le contexte génomique et ScpB et Smc par des co-événements évolutifs. Ces trois gènes forment un système de ségrégation de l'hétérochromatine décrit chez *B. subtilis* [435],[351]. Enfin, FtsY et Smc ont déjà été décrits comme étant co-transcrits chez *B. subtilis* [233].

Division cellulaire Concernant la division cellulaire, les clans DCW 1 et 2 ont été décrits à maintes reprises dans la littérature [310], [134], [456]. Le cluster DCW est très conservé chez les bactéries. Il est composé de nombreux gènes impliqués dans la division cellulaire et la synthèse

du peptidoglycane comme ceux codants pour les protéines FtsZ, FtsA, FtsQ, FtsL, DivIVA, SepF, MraY ou encore MurG. L'ensemble de ces gènes sont intimement liés fonctionnellement puisqu'il contribuent au phénomène de septation et constituent le divisome chez *E. coli* et de nombreuses autres espèces [472],[343]. Il est aussi intéressant de noter que la PBPB4 qui a été montrée comme étant recrutée au niveau du site de division chez *B. subtilis* est inférée comme faisant partie du cluster DCW [416]. Néanmoins, l'ensemble du divisome n'est pas représenté dans le cluster DCW et certaines de ces familles sont retrouvés ailleurs dans le réseau et dans l'arbre de distances. Néanmoins, des liens fonctionnels démontrés sont retrouvés entre certaines d'entre elles. Par exemple, FtsE et FtsX qui forment un transporteur ABC nécessaire à la division cellulaire chez *E. coli* ont été retrouvés comme étant fortement liés évolutivement [87]. Les familles TolQ et TolR impliquées dans l'invagination de la membrane externe lors de la division cellulaire chez *E. coli* présentent également un fort lien évolutif [163].

Élongation cellulaire L'élongasome est représenté principalement par le clan Mre/Min au sein du réseau de corrélations évolutives (table 3.4). La structure en opéron des gènes codants pour les protéines RodA, MreB, MreC, MreD, MinC, MinD et MinE a déjà été décrite [274]. L'ensemble des protéines Mre sont les constituants principaux de l'élongasome [472]. Également la famille PBPBB5 semble reliée à l'élongasome. La contribution de cette famille de PBP à l'élongation a été démontrée chez *B. subtilis* [499]. Il est intéressant de noter que la famille RodZ, un composant pourtant important dans l'élongation n'est pas retrouvé comme étant corrélé avec le reste de l'élongasome (principalement lié à RecA).

Synthèse du peptidoglycane De nombreux liens entre les familles de gènes impliqués dans la synthèse du peptidoglycane ont été inférés par l'analyse et démontrés dans la littérature. Tout d'abord, les familles GlmM et GlmS impliquées dans l'initiation de la voie de synthèse des précurseurs du peptidoglycane sont retrouvées comme étant fortement liés [227], [18]. Ensuite, le clan Nag/Mur est composé des familles AnmK, MurQ, MurK, NagA et NagB qui ont été prouvées comme étant responsables du recyclage du peptidoglycane chez *E. coli* [226]. Enfin, les systèmes spécifiques de la synthèse des ponts interpeptidiques chez *S. pneumoniae* (MurN-

MurM) [144], *Staphylococcus aureus* (FemA-FemB-Fmh) [486],[248] et *E. faecium* (Asr-Aslfm) [29] sont retrouvés au sein du réseau de corrélation.

Sporulation Concernant les familles de gènes impliquées dans la sporulation, la plupart est associée au clan Spo (table 3.4). Les familles SpoIIF, SpoVB, SpoIIID, SpoIID, SpoIIGA et SpoVE ont toutes été décrites comme étant impliquées dans le processus de sporulation chez *B. subtilis* [378]. Les deux PBP DacB et DacF impliquées spécifiquement dans la sporulation font également partie du clan [382]. La PBPB6 est également retrouvée dans le clan Spo, en accord avec des travaux la proposant comme étant impliquée dans la sporulation, bien que son rôle exact ne soit pas encore connu [500]. Il est à noter que les familles de gènes impliquées dans la sporulation comme SpoVK, SpoVG, SpoIIIJ et SpoIIIJ ne sont pas retrouvées dans ce clan.

Synthèse de la capsule Seuls 4 gènes impliqués dans la synthèse de la capsule ont été analysés. Néanmoins, les trois familles CpsB, CpsC et CpsD impliqués dans la régulation de la synthèse de la capsule, notamment chez *S. pneumoniae* [521] sont fortement liées d'après le réseau de corrélations, par la synténie, la co-occurrence et les co-évenements. Ces trois familles forment un clan séparé du reste des familles.

Régulation du cycle cellulaire et autres systèmes La kinase StkP et sa phosphatase associée PhpP [359] sont retrouvées comme reliés au sein du clan StkP (table 3.4). Également, le clan Wal est composé des familles WalI, WalJ, WalH, WalK et WalR. L'ensemble de ces familles sont impliquées dans la régulation de la synthèse du peptidoglycane chez *S. aureus* [105],[384]. Enfin, les deux familles GidA et GidB inférées comme fortement reliés évolutivement sont toutes deux impliquées dans la modification des ARN [429].

Conclusion sur la qualité de l'approche l'analyse des clans mis en évidence dans le réseau et dans l'arbre de distances révèle que ces clans recoupent les systèmes majeurs décrits dans la littérature. Ceci indique que l'approche que nous avons développée est efficace pour

mettre en évidence des liens fonctionnels, au moins d'un point de vue qualitatif. Néanmoins, il est difficile d'estimer le nombre de liens détectés alors qui n'auront pas de sens au niveau biologique. Les méthodes mises en œuvre pour prendre en compte le maximum d'information et les coefficients ainsi que leurs scores sont espérés être suffisants pour ne pas introduire un grand nombre de faux positifs. Aussi, le nombre de faux négatifs est difficile à estimer mais un certain nombre de liens démontrés expérimentalement ne sont pas inférés par notre approche (*e.g.* RodA et RodZ, FtsZ et ZapA, ...). Les faux négatifs peuvent provenir du fait que nous n'avons étudié les familles de gènes qu'au niveau des *Firmicutes* mais pas au niveau des sous-groupes des *Firmicutes*, que nous n'avons pas été exhaustifs dans les familles de gènes étudiées ou encore à cause de biais méthodologiques.

3.5.2.4 Nouveaux liens fonctionnels inférés

Un certain nombre de liens fonctionnels inférés par l'étude corrélative n'a jamais été décrit à notre connaissance dans la littérature. Les familles de gènes concernées retrouvées dans le réseau de corrélation et dans l'arbre de distances sont présentés table 3.5. Afin de faciliter la compréhension, nous avons séparé les cas en trois groupes : Les clans majoritairement constitués de familles dont les liens fonctionnels n'ont jamais été décrits, les clans composés de quelques liens non décrits puis les liens ponctuels entre quelques familles n'appartenant pas à proprement parler à des clans.

Clans majoritairement non décrits dans la littérature Le clan *Bacilli* qui correspond aux gènes principalement retrouvés ou exclus chez les *Bacilli* est composé de familles de gènes qui contribuent à des machineries moléculaires très variées. On retrouve des familles impliquées dans la réplication (RecU, DnaI, DnaB(BS)), dans la régulation de la septation/élongation (GpsB, EzrA), dans la synthèse du peptidoglycane (PBPA3, PBPBX1 DacA), dans la régulation des PBPs (CDP4, MacP, CozE3), dans la régulation du positionnement de l'anneau Z (MinE), dans la sporulation (SpoIIIJ2) et dont la fonction n'est pas connue (PCDP10). L'ensemble de ces familles de gènes présentent entre elles un Z-score moyen de corrélation par les

Clan	Liens fonctionnels inférés non décrits dans la littérature						Type de corrélation prépondérant	Z-score
Bacilli	Clan entier						Co-occurrence	0,3
Bacilli	RecU	PCDP10	GpsB	A3			Synténie, co-occurrence	0,53/0,3
FtsH	Clan entier						Synténie	0,3
Spo	SpoIIGA/IID/IIE/IIID/VB/DacB/F/B6	CozE4	MurA2	MurA1	FtsH3	Mbl	Co-occurrence, co-événements	0,17/0,12
DCW2	DivIVA/SepF	PCDP6	PCDP7	PCDP8	IleS1		Synténie	0,41
Ori2	ParA/B	GidA	GidAB	Noc	Jag	SpoIIIJ1	Synténie	0,35
StkP	StkP/PhpP	PriA	SunL	Mtf			Synténie	0,49
Mre/Min	MreB/C/D/MinC/D/E/J/RodA/B5	ValS	RadC	Maf			Synténie	0,22
Wal	WalH/I/J/K/R	MurZ					Synténie	0,22
Scp/Xer	ScpA/B	XerC	XerD				Synténie	0,33
Smc	Smc	RecG					Synténie	0,27
	MreBH	GatD3	DacC	A5			Co-occurrence, co-événements	0,25/0,14
	FtsW	BX1					Synténie, co-événements	0,92/0,36
	B3	MurB2					Synténie, co-événements	0,42/0,32
	MurG2	PCDP4	B1				Co-événements	0,26
	Ddl1	RecR	PCDP4				Synténie	0,39
	AlaS	CozE4	A6				Synténie	0,52
	RodZ	RecA	FtsK				Synténie	0,53
	GatD1	RecG	Alr				Synténie	0,56
	GlmU	SpoVG					Synténie	0,93
	FtsH2	LysA1					Synténie	0,57
	A4	SpoVK					Synténie	0,98
	MurJ	LeuS					Synténie	0,94
	DnaG	ClpX					Synténie	0,53
	Alr_like	FemX					Synténie	0,95
	FtsJ	RecN					Synténie	1
	MurC	FtsK2					Synténie	0,86
	FtsE/X	MinJ					Synténie	0,46

TABLE 3.5 – Liens fonctionnels inférés entre les familles de gènes. Les Z-scores indiqués correspondent à la moyenne des Z-scores entre toutes les familles de la ligne. Pour les lignes avec une case grisée, les Z-scores correspondent à la moyenne des Z-scores entre toutes les familles de la ligne en blanc avec les familles de la case grisée.

profils phylogénétiques élevé (Z-score=0,30). Les familles MinE et PBPBX1 sont anticorrélées aux autres familles du clan indiquant ainsi que celles-ci remplissent peut-être des rôles complémentaires de celles acquises à l'émergence des *Bacilli*. Certaines familles au sein du clan des *Bacilli* présentent également une synténie. Ainsi, PCDP10, GpsB, RecU et PBPA3 présentent un Z-score moyen de corrélation par les synténies élevé (Z-score=0,52).
La composition en familles de gènes du clan FtsH est également très surprenante car elle regroupe les familles FtsH1 (protéase du divisome), DivIC (également composant du divisome), TilS (tRNA(Ile)-lysidine synthase, modification des ARN), Mfd (facteur de couplage transcription-réparation), SpoIIE et SpoVB (protéines de la sporulation) et YabM (homologue de MurJ dont la fonction n'est pas connue) sont retrouvées. Ces familles sont corrélées principalement par la synténie (Z-score=0,30) et forment un seul et même cluster.

Clans avec certains liens fonctionnels non décrits Certains liens évolutifs n'ayant jamais été décrits dans la littérature sont également observés au niveau d'autres clans.
Tout d'abord, le clan Spo présente certaines familles non reliées fonctionnellement de façon évidente avec la sporulation. Les liens de ces familles avec les autres familles de la sporulation sont de différentes natures. La famille CozE4 semble très corrélée par les profils phylogénétiques aux autres familles de la sporulation et notamment avec DacB (Z-score=0,341). Les familles de gènes MurA1 et MurA2 semblent plus corrélées avec les familles de la sporulation par la synténie que par les profils phylogénétiques ou les événements évolutifs. MurA1 est corrélé par le contexte génomique avec SpoIID et SpoIIID (Z-score=0,61/0,42) tandis que MurA2 est corrélée avec SpoIIGA (Z-score=0,31). Les corrélation au niveau des profils phylogénétiques sont légèrement plus faibles pour MurA2 (Z-score=0,21) et quasi nulle pour MurA1. FtsH3 semble corrélé surtout par les profils phylogénétiques avec les familles de la sporulation avec un Z-score moyen de 0,20. Enfin, concernant Mbl, il semble que cette famille soit surtout reliées à SpoIID par le contexte génomique (Z-score=0,49) et présente un Z-score moyen issu des profils phylogénétiques assez faible (Z-score=0,12).
Le clan DCW2 qui contient des familles de gènes impliquées dans la division cellulaire (DivIVA et SepF) contient également des familles de gènes de fonction inconnue (PCDP6/7/8) mais aussi une famille de gènes impliquées dans la synthèse des ARNt (IleS1, Isoleucyl-ARNt

synthétase).

Au sein du clan Ori2, des familles de gènes impliquées dans de divers mécanismes sont retrouvés. Certaines sont impliquées dans la ségrégation chromosomique (ParA/B), dans le positionnement de l'anneau Z (Noc), dans la sporulation (SpoIIIJ1, Jag) ou dans la modification des ARN (GidA/B). L'ensemble de ces familles de gènes présentent un Z-score moyen de corrélation par synténie élevé (Z-score=0,39).

Au sein du clan StkP, en plus de la kinase StkP et de sa phosphatase associée PhpP, d'autres familles sans lien fonctionnel apparent sont retrouvées. Ainsi, les familles PriA (Primase, réplication du chromosome), SunL (ARN méthyltransferase) et Mtf (Methionyl-ARNt-formyl synthétase) sont liés StkP et PhpP, notamment par le contexte génomique (Z-score=0,48/0,50/0,50 respectivement).

Le clan Mre/Min (élongasome et système Min) présente trois familles de gènes dont les fonctions ne sont pas reliées de manière apparente avec les autres familles de gènes qui composent le clan. Il s'agit des familles ValS (Valyl-ARNt synthétase), Maf (nucléotide pyrophosphatases qui présente des analogies structurales avec les ARNr synthétases [322], [460]) et RadC (impliquée dans la réparation de l'ADN). Les liens de ces trois protéines avec les autres familles du clan est due à un Z-score de corrélation élevé à partir des synténies. VaS et RadC corrèlent uniquement avec MreB, MreC, MreD, MinC et MinD (Z-score=0,19/0,35) mais pas avec RodA, B5 et MinJ. Maf corrèle avec tous les Mre/Min, RodA et B5 (Z-score=0,31).

Le clan Wal composé des familles WalH, WalI, WalJ, WalK et WalR contient également la famille MurZ en son sein. Cette dernière est liées principalement par son contexte génomique puisqu'elle présente un Z-score moyen de 0,22 à partir des synténies.

Au sein du clans Scp/Xer/Smc, un lien évolutif entre ScpA/B et XerC/D est observé. Ces deux groupes de familles de gènes sont impliqués respectivement dans la ségrégation chromosomique et dans la résolution des dimères de chromosome. Le lien qui les unit provient principalement du contexte génomique (Z-score moyen entre les deux groupes de 0,33). Néanmoins, le lien entre ScpA/B et XerC semble artefactuel puisque la synténie entre ces groupes observés n'est retrouvée que dans 3 génomes. Le Z-score surestime la synténie car XerC n'est que très peu voisin avec d'autres gènes étudiés. Un lien évolutif entre Smc et RecG est observé mais n'a jamais été décrit. Il s'agit là encore d'un lien issu du contexte génomique avec un Z-score entre

les deux familles de 0,27.

Liens ponctuels non décrits dans la littérature De très nombreux liens ponctuels entre des familles de gènes jamais décrits dans la littérature ont été détectés par la méthode de corrélation évolutive. Nous n'allons pas tous les décrire ici mais énumérer les plus importants/intéressants, la plupart sont représentées dans la table 3.5. Ainsi, MreBH, GatD3, DacC et PBPA5 présentent des profils phylogénétiques très similaires (Z-score=0,25). Ces quatre familles de gènes sont en effet présentes de façon très éparse chez les *Bacillales* et chez quelques *Clostridia*. Autre exemple, les familles FtsW et PBPBX1 semblent extrêmement corrélées de par leur contexte génomique mais également par leurs événements évolutifs (Z-score=0,920,36). Une situation similaire est observée pour les familles MurB2 et B3 qui sont aussi corrélées par la contexte génomique et les co-événements (Z-score=0,42/0,32). Dans ce cas, les distributions taxonomiques sont corrélées chez les *Bacillales* mais pas chez les autres clades. Les familles MurG2, PCDP4 et B1 semblent corrélées par les données issues des événements évolutifs (Z-score=0.26). Il est intéressant de noter que leur distribution taxonomique n'est que très peu corrélée. Le reste des liens est issu principalement des contexte génomiques comme entre FtsEX et MinJ, SpoVK et PBPA4 ou encore Alr-like FemX.

Conclusion Il existe donc au sein du réseau de corrélation de nombreux liens inférés n'ayant pas été démontrés expérimentalement. Il n'est pas possible de savoir en l'état si ces liens correspondent à une réalité biologique ou à des artefacts méthodologiques. Pour vérifier cela, il serait nécessaire de tester ces liens. Il serait par exemple judicieux de tester des interactions protéines-protéines chez différents *Firmicutes* par différentes méthodes comme la résonance plasmonique de surface ou le double hybride.

3.6 Histoires évolutives détaillées et implications fonctionnelles

De nombreux liens décrits ou non dans la littérature entre les familles du cycle cellulaire ont donc été inférés par notre approche. Nous avons disséqué précisément l'histoire évolutive des clans et de quelques familles présentant un intérêt particulier, soit au niveau de l'organisation génomique, soit au niveau des événements de gènes.
Pour les événements de gènes il est possible d'utiliser les données de réconciliation ou issues des profils phylogénétiques. Nous avons uniquement présenté les données issues des profils phylogénétiques principalement par manque de temps. Néanmoins, nous décrirons quelques résultats issus de la réconciliation puisque dans certains cas, les résultats sont bien meilleurs, notamment pour les familles dont la transmission n'est pas globalement verticale chez les *Firmicutes*. Également, nous nous sommes principalement focalisés sur les événements ayant eu lieu dans les branches profondes de l'arbre d'espèces dans un soucis de clarté. Les branches dans cette section seront dénommées par des nombres indiqués sur les figures.

3.6.1 Description et interprétation fonctionnelle de l'histoire des clans

3.6.1.1 Clan *Bacilli*

En recoupant les deux représentations des liens évolutifs (réseau et arbre de distances), le clan *Bacilli* est composé des familles GpsB, PBPA3, RecU, PCDP10, DnaB(BS), DnaI, CDP4, CDP1, DacA, MacP, CozE1. Les familles MinE, BX1, CozE3, SpoIIIJ2 et MurJ sont inclues uniquement dans le clan dans le réseau de corrélation tandis que Pfs y est regroupé uniquement dans l'arbre de distances. Les événements majeurs ayant affecté ces familles sont présentés figure 3.45.

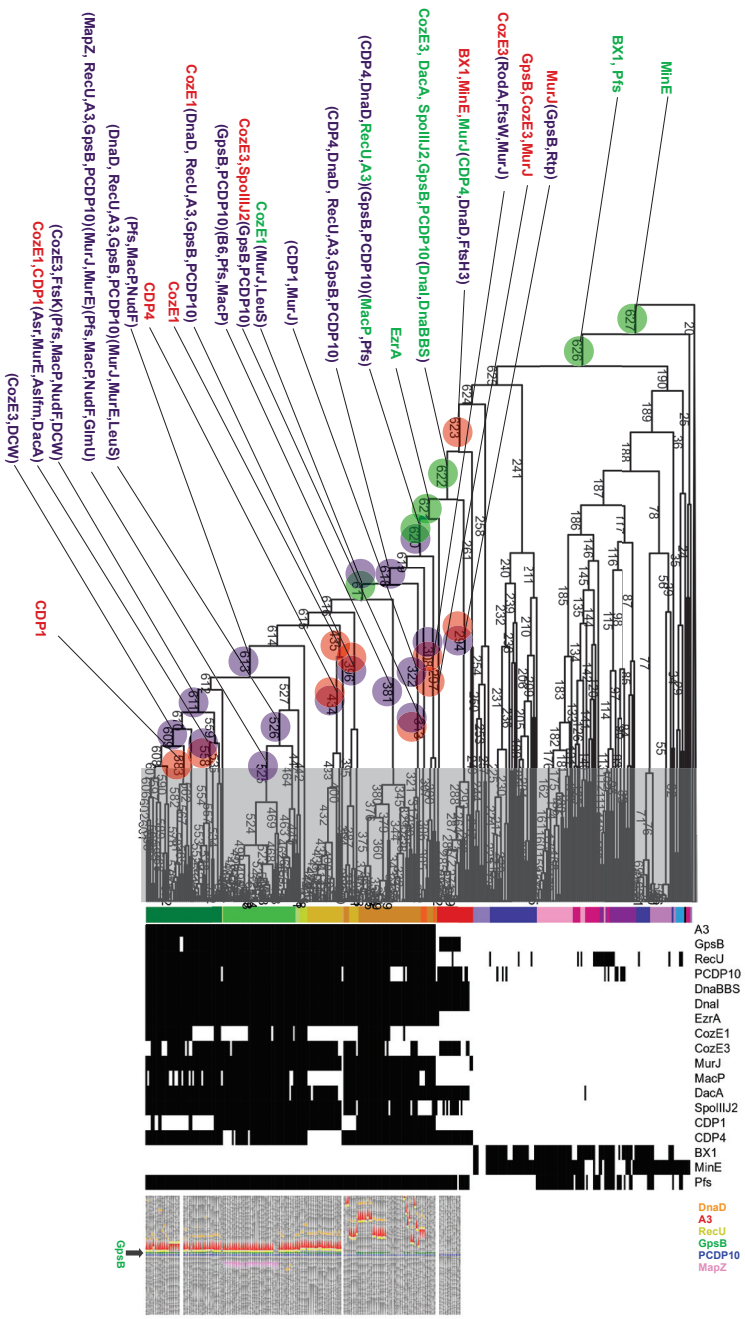


FIGURE 3.45 – Histoire évolutive du clan *Bacilli*. Les événements évolutifs sont représentés par des ronds (vert : apparition, rouge : perte, violet : modification de clusters de gènes). La zone de l'arbre grisée correspond aux branches ou les événements évolutifs n'ont pas été représentés. La bande colorée correspond à la taxonomie. La distribution des familles de gènes est représentée par des rectangles noirs (présence) et blancs (absence). Le contexte génomique de GpsB est représenté.

Description Il semble que MinE et BX1 soient anciens chez les *Firmicutes* (branche 627,626) et aient été perdus à l'émergence des *Bacilli* (branche 623). Ces pertes sont concomitantes avec l'apparition de MurJ et CDP4. À la branche suivante (622), CozE3, DacA, SpoIIIJ2, GpsB, PCDP10, DnaI et DnaBBS apparaissent. DnaI et DnaBBS présentent une synténie qui sera conservée chez tous les *Bacilli*. À la branche 621, EzrA apparaît puis a la branche 620, trois gènes apparaissent : RecU, A3 et MacP. MacP est voisin de Pfs et RecU et A3 sont en contexte avec CDP4 et DnaD. GpsB et PCDP10 sont également inférés comme étant voisins génomiques à cette branche. Les copies de la famille RecU chez les *Clostridia* semblent provenir de transferts horizontaux au vue de la réconciliation et de la phylogénie (figure 3.45). GspB semble également être acquis par transmission horizontale chez les *Paenibacillaceae*, indiquant que GpsB aurait pu en fait apparaître à la branche 619 et non 621 (annexe .11). Les deux clusters GpsB/PCDP10 et RecU/A3/DnaD/CDP4 s'associent à la branche 618. La synténie entre ces deux clusters n'est pas conservée chez les *Bacillales* mais semble plus ténue à partir de la branche 616. CDP4 est dissocié du cluster sur plusieurs branches (396, 434, 526). À l'émergence des *Streptococcaceae* (525), DnaD est également dissocié du cluster et MapZ y est associé. Concernant le cluster Pfs/MacP, NudF s'y inclut à la branches 613 de façon concomitante avec une modification majeure de MacP qui perd son domaine C-terminal et acquiert un domaine supplémentaire en N-terminal. À l'émergence des *Streptococcaceae* (525), GlmU est également associé au cluster ainsi que CDP3 qui apparaît à cette branche. À l'émergence des *Lactobacillaceae/Leuconostocaceae*, Le cluster NudF/Pfs/MacP fusionne avec le cluster DCW.

Interprétation fonctionnelle Tout d'abord la perte de MinE et BX1 quasiment concomitante avec le reste des familles du clan *Bacilli* indique une potentielle redondance fonctionnelle avec le reste des familles du clan. En effet, il est tentant de faire l'hypothèse que les familles apparues à l'émergence des *Bacilli* constituent un système de remplacement de MinE et BX1. Néanmoins, cette hypothèse est à nuancer car il est possible que nous ayons sous-échantillonné les familles impliquées dans le cycle cellulaire et que l'émergence des *Bacilli* soit en fait accompagnée de multiples pertes d'autres familles. MurE et BX1 ne seraient ainsi que des représentants d'un plus grand ensemble de familles de gènes. De plus, MinE étant impli-

qué dans le positionnement de FtsZ [188] et BX1 dans la réticulation du peptidoglycane, la redondance fonctionnelle avec les familles apparaissant à l'émergence des *Bacilli* et impliquées dans d'autres fonctions cellulaires parait peu évidente.

Les familles GpsB, PCDP10, RecU, A3, DnaD et CDP4 présentent potentiellement des liens fonctionnels au vue de leur histoire évolutive. GpsB et PCDP10 co-apparaissent et sont quasi-systématiquement synténiques indiquant un lien fonctionnel très probable, de même pour RecU et A3. DnaD semble aussi être un voisin conservé de RecU et A3 mais cette synténie est plus labile, de même pour CDP4. Les deux clusters GpsB/PCDP10 et CDP4/DnaD/RecU/A3 présentent également entre eux une synténie, labile chez les *Bacillaceae* et conservée chez les autres clades. Ceci indique également une relation fonctionnelle probable entre ces deux clusters. Il est intéressant de noter que les deux clusters présentent des directions opposées, indiquant potentiellement la présence d'un promoteur bidirectionnel contrôlant l'expression des gènes des deux clusters. Le rôle de PCDP10 n'est pas décrit dans la littérature à ce jour et GpsB est impliqué dans la coordination élongation-division [68], [148]. Néanmoins, des résultats préliminaires effectués durant mon stage de Master suggèrent que la délétion de PCDP10 chez *Streptococcus* pneumoniae a des effets sur la morphologie des cellules et que PCDP10 interagirait avec GpsB et DivIVA. Il est tentant de faire l'hypothèse que PCDP10 serait aussi impliqué dans la coordination élongation-division. GpsB a aussi été montré comme étant impliqué dans la localisation de A3 chez *B. subtilis* ce qui corrobore le lien fonctionnel entre les deux clusters [68]. Les trois autres familles sont impliquées dans la régulation des PBP (CDP4), l'initiation de la réplication (DnaD, [218] et la résolution des jonctions Holliday (RecU, [371]. Elles sont donc toutes trois impliquées dans les processus liés à l'ADN. Le cluster GpsB/PCDP10/RecU/A3/DnaD/CDP4 formerait donc probablement une unité fonctionnelle de coordination entre l'élongation, la division, la réplication et la ségrégation du chromosome.

MacP et Pfs sont synténiques dès l'apparition de MacP, puis NudF s'associe au cluster. La encore, le lien fonctionnel entre MacP, Pfs et NudF n'est pas évident. MacP est un régulateur de PBPB5 chez *Streptococcus* pneumoniae [141], Pfs est une 5'-méthylthioadenosine/S-adenosylhomocysteine nucléosidase impliquée dans la méthylation des ARN, la synthèse d'acides aminés et quorum-sensing [20] et NudF est prédite comme étant une ADP-ribose pyrophosphatase impliquée dans de nombreux processus cellulaires [363]. Il est possible que MacP joue

d'autres rôles au sein de la cellule, notamment dans le métabolisme. Aussi, la modification de la structure de MacP concomitante avec l'inclusion de NudF dans le cluster indique potentiellement un changement fonctionnel de MacP lié à NudF. Le seul domaine conservé avant et après la modification structurale de MacP est le domaine prédit comme étant transmembranaire. MacP étant toujours voisin de Pfs, il est tentant de faire l'hypothèse que Pfs serait relié fonctionnellement avec le domaine transmembranaire de MacP.
Concernant DnaI et DnaBBS, il apparaît clairement que ce couple est relié fonctionnellement (co-apparition et synténie), ce qui d'ailleurs a déjà été prouvé expérimentalement bien que le rôle de DnaBBS ne soit pas encore élucidé [218].
MurJ présente une synténie conservée avec LeuS à partir de la branche 617 puis avec MurE dès l'émergence des *Streptococcaceae*/Enteroccaceae (526). Le lien entre LeuS et MurJ est difficile à expliquer puisque MurJ est la flippase du Lipide II dans la synthèse du peptidoglycane [425] et LeuS est une Leucyl-ARNt synthétase.Le lien fonctionnel entre MurJ et MurE est plus facilement explicable puisque MurE est aussi impliquée dans la synthèse du peptidoglycane.
SpoIIIJ2, EzrA, DacA, et CozE3 ne semblent pas présenter de contexte génomique conservé avec des familles de gènes du cycle cellulaire analysées. Leur lien fonctionnel avec le reste des autres familles est donc plus nuancé.
Enfin, CozE1 ne semble pas présenter une co-occurrence avec le reste des familles du clan *Bacilli* et semble appartenir a ce clan de par son lien évolutif avec DacA. En effet, DacA est ponctuellement voisin génomique de CozE1. Étant donné que les deux familles ne sont pratiquement jamais en contexte avec d'autres familles analysées, la corrélation a été sur-évaluée par le Z-score. CozE1 n'est donc probablement présent dans ce cluster que par artefact méthodologique.

3.6.1.2 Clans DCW1 et DCW2

Les clans DCW1 et DCW2 sont constitués des familles FtsZ, FtsA, FtsL, FtsQ, MraW, MraY, MraZ, MurD, MurG1, B4, SepF, PCDP6, PCDP7, PCDP8, DivIVA et IleS1. La majorité de ces familles sont impliquées dans la division cellulaire et la synthèse du peptidoglycane et sont regroupées dans le cluster DCW. Les événements majeurs ayant affectés les clans DCW1

et DCW2 sont présentés figure 3.46. Dans un souci de clarté, les modifications des clusters de gènes sont indiquées uniquement si au moins deux gènes sont perdus/acquis au sein du cluster.

Description L'ensemble des gènes présents dans les clan DCW1 et DCW2 sont inférés comme étant présents chez l'ancêtre des *Firmicutes*. Les deux clans sont séparés l'un de l'autre en terme de contexte génomique chez l'ancêtre. Le cluster de gènes correspondant au clan DCW1 contient les gènes MraW, MraY, MraZ, FtsA, FtsQ, FtsL, FtsZ, MurF, MurG1 et B4 mais également des gènes qui ne sont pas inclus dans les clan DCW1 comme MurA2, MurB2, MurE, SpoIIGA et SpoVE. Le clan DCW2 forme un cluster de gènes distinct de DCW1 et regroupe SepF, PCDP6, PCDP7, PCDP8, DivIVA et IleS1. Néanmoins, l'analyse qualitative du profil des synténies (figure 3.46) suggère que l'absence de voisinage entre les deux clusters chez les *Clostridia* serait plus du à un éloignement génomique des deux clans chez les *Clostridia* et que les clusters DCW1 et 2 chez l'ancêtre *Firmicutes* présenteraient une vicinité.
Durant la diversification des *Clostridia*, plusieurs événements majeurs ont eu lieu. Chez les groupe 0 et le groupe 1A, les clans DCW1 et 2 semblent être relativement proches mais néanmoins espacés par plusieurs gènes impliqués dans la sporulation. Chez le reste du groupe 1 (B, C, D), les deux clusters sont éloignés l'un de l'autre. Durant l'émergence du groupe 1, DivIVA est perdu à plusieurs branches (113, 120, 142) ainsi que FtsA (146, 116, 55), IleS1 (186, 115), MurG1 (185, 115), PCDP7 (144, 115, 90) et MraZ (133, 29). Les contextes génomiques de FtsZ et DivIVA chez les *Clostridia* (figure 3.46) du groupe 1 indiquent que les deux clusters sont éloignés au sein de la majorité des génomes. Cependant, à l'émergence des *Clostridium* (branche 185), FtsA et FtsZ ont été séparés du cluster DCW1 et le reste du DCW1 se place à coté du DCW2. Chez les *Clostridia* du groupe 2, les deux clusters sont rapprochés mais séparés par des gènes de sporulation (SpoIIGA, SigE, SigG, ...). Les *Negativicutes* présentent aussi deux clusters distincts DCW1 et 2.
A l'émergence des *Bacilli* (branche 623), les deux clusters sont inférés comme étant voisins mais néanmoins toujours séparés par un certain nombre de gènes de sporulation. À partir de l'émergence des *Staphylococcaceae* et autres *Bacilli* (branche 616), les deux clusters sont proches voisins. Cette branche correspond à la perte massive des gènes de la sporulation (figure 3.46), probablement responsable du rapprochement des deux clusters. À l'émergence des

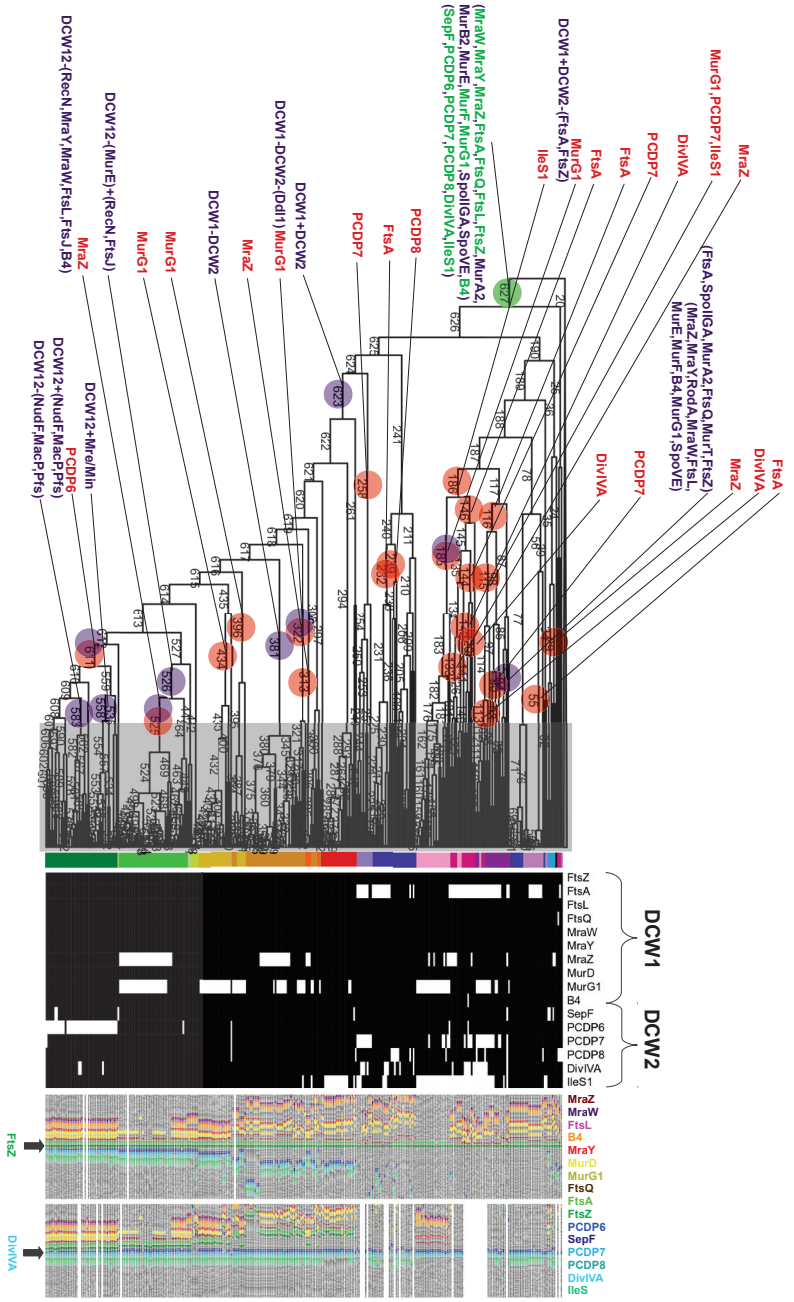


FIGURE 3.46 – Histoire évolutive du clan DCW. Les événements évolutifs sont représentés par des ronds (vert : apparition, rouge : perte, violet : modification de clusters de gènes). La zone de l'arbre grisée correspond aux branches ou les événements évolutifs n'ont pas été représentés. La bande colorée correspond à la taxonomie. La distribution des familles de gènes est représentée par des rectangles noirs (présence) et blancs (absence). Les contextes génomiques de FtsZ et DivIVA sont représentés.

Streptococcaceae (525), le cluster DCW est divisé en deux : RecN, MraY, MraW, FtsL, FtsJ, B4 d'un coté et FtsA, PCDP6, PCDP7, PCDP8, MurD, IleS1, SepF, MurG1, FtsQ, DivIVA, FtsZ de l'autre. À l'émergence des *Lactobacillaceae*/*Leuconostocaceae* (611), le petit cluster NudF/MacP/Pfs est inclus dans le cluster DCW puis de nouveau éloigné à l'émergence des *Leuconostocaceae* (branche 583). À cette même branche, MraZ semble avoir subi une remplacement homologue (annexe .11). De façon intéressante, le cluster Mre/Min est inclus dans le DCW à la branche 558.

Interprétation fonctionnelle Les deux clans DCW1 et DCW2 semblent, de par le contexte génomique, constitué de deux unités fonctionnelles distinctes chez les *Firmicutes*. Ces deux unités sont séparées chez la plupart des *Clostridia*, distants de quelques gènes de la sporulation chez la plupart des *Bacillales* et à coté chez la plupart des *Lactobacillales*.
Il est intéressant d'observer que certains taxons présentent des clusters DCW atypiques. Chez les Clostridium, FtsZ et FtsA sont séparés du cluster DCW suggérant une régulation transcriptionnelle atypique des gènes de division cellulaire. Il est possible d'émettre l'hypothèse que la régulation temporelle et spatiale du divisome peut être différente bien qu'aucune donnée expérimentale ne soit disponible pour ces espèces. Chez les *Streptococcaceae*, le cluster DCW est aussi très différent de ceux observés chez les *Bacilli*. En effet, grand nombre de gènes ne sont plus localisés dans le cluster. Cette observation a déjà été faite par Massidda et collègues [310]. De cette observation, il est possible de faire l'hypothèse que ce changement d'organisation soit lié la manière atypique qu'a notamment *S. pneumoniae* de se diviser. En effet, *S. pneumoniae* a la particularité d'effectuer l'élongation cellulaire et la division au milieu de la cellule [200]. Il est intéressant de noter que Tamames et collègues ont généré un arbre de distances de quelques espèces de bactéries à partir de la composition et de l'ordre des gènes au sein du cluster DCW [456]. Ils ont mis en évidence une corrélation entre la morphologie des cellules et l'organisation du cluster DCW. Ils ont aussi observé de façon concordante avec notre étude que *Streptococcus pneumoniae* présentait un cluster DCW différent de celui des Bacillus.Enfin, la plupart des *Lactobacillaceae* possèdent les clusters Mre/Min et NudF/MacP/Pfs au sein du cluster DCW. Il est tentant de faire l'hypothèse que la fixation dans ces lignées de l'association de ces clusters n'est pas du au hasard et que les *Lactobacillaceae* régulent de façon coordonnée les gènes des

trois clusters.

Certains gènes pourtant prouvés comme étant essentiels au bon déroulement de la division cellulaire sont absents chez certains *taxa*. Ainsi, FtsA est perdu chez un grand nombre de *Clostridia*. FtsA est la protéine majeure d'ancrage à la membrane de FtsZ permettant ainsi le positionnement et la formation de l'anneau Z [376]. Cette absence indique donc qu'il existe probablement un système alternatif pour l'ancrage à la membrane de l'anneau Z chez ces espèces. De même, DivIVA est perdu chez certains *Clostridia* bien que cela soit plus rare.

Plusieurs gènes au sein du cluster DCW n'ont pas une fonction reliée à la division cellulaire. Leur association étroite au sein du cluster DCW est donc surprenante. La famille IleS est impliquée dans la traduction (Isoleucyl-ARNt synthétase) et la famille MraW dans la méthylation des ARN et particulièrement l'ARN 16S [60]. La nature de la relation fonctionnelle entre ces familles et le reste des familles du cluster DCW reste donc à être déterminée. De manière intéressante, Eraso et collègues proposent que MraW jouerait le rôle d'antagoniste à MraZ. Ce dernier a été montré comme étant un régulateur transcriptionnel agissant comme represseur sur des promoteurs spécifiques P*mra* situés notamment en amont du cluster DCW chez *E. coli* [125]. Enfin, quelques gènes n'ont pas de fonction identifiée à ce jour : PCDP6, 7 et 8. La délétion de ces gènes chez *Streptococcus* impacte la morphologie des cellules mais de façon relativement faible [134]. Les défauts morphologiques chez les bactéries peuvent résulter de problèmes au niveau de la division ou plus généralement du cycle cellulaire ou d'effets pléiotropiques. La conservation de ces gènes dans de cluster DCW et ces résultats indiquent qu'ils pourraient jouer un rôle non négligeable dans la division cellulaire.

3.6.1.3 Clan Spo

Le clan Spo (de Sporulation) est composé des familles SpoIID, SpoIIE, SpoIIGA, SpoIIID, SpoVB, SpoVE, B6, DacB, DacF, FtsH3, CozE4, MurA2, MurA1 et Mbl. Leur histoire évolutive est présentée figure 3.47.

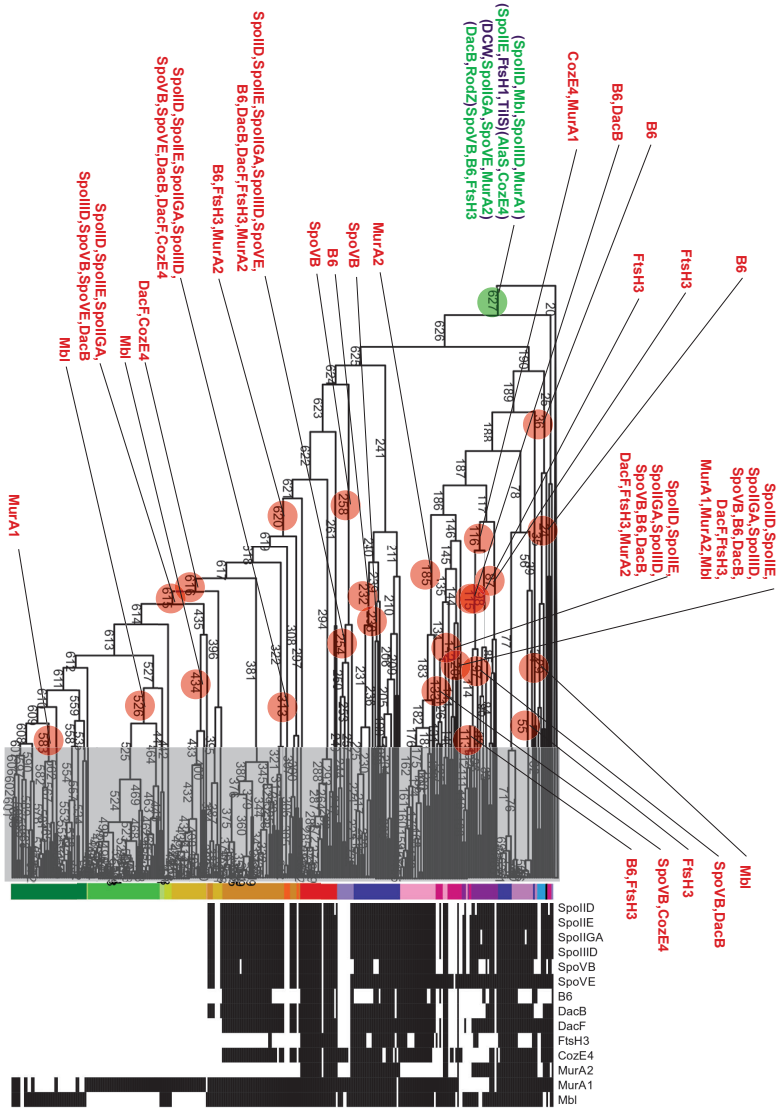


FIGURE 3.47 – Histoire évolutive du clan Spo. Les événements évolutifs sont représentés par des ronds (vert : apparition, rouge : perte). La zone de l'arbre grisée correspond aux branches ou les événements évolutifs n'ont pas été représentés. La bande colorée correspond à la taxonomie. La distribution des familles de gènes est représentée par des rectangles noirs (présence) et blancs (absence).

Description L'ensemble de ces gènes est inféré comme étant présent chez l'ancêtre des *Firmicutes*. Ils sont répartis dans cinq clusters de gènes distincts. Il semble que cinq pertes majeures des gènes de la sporulation aient eu lieu durant la diversification des *Firmicutes* : branche 142 (*Tissierellia*), 120 (*Acetobacterium woodii*/*Eubacterium limosum*), 254 (*Selenomonas*/*Veillonella*/*Megasphaera*), 313 (*Exiguobacterium*) et 615 (*Staphylococcaceae*/Autres *Bacilli*). Ces pertes massives conduisent à des profils phylogénétiques très similaires entre la plupart des familles du clan Spo.

Interprétation fonctionnelle L'ensemble des gènes étudiés impliqués dans la sporulation semble perdu massivement sur les mêmes branches même si ces gènes ne sont pas situés dans les mêmes clusters. Ces résultats suggèrent que l'équipement génétique impliqué dans la sporulation est majoritairement perdu en bloc. Les espèces ne présentant plus les gènes de la sporulation peuvent donc être inféré comme non sporulantes. Ainsi, les *Tissierellia*, les *Selenomonas*, les *Veillonella*, les *Megasphaera*, les *Exiguobacterium* , *Acetobacterium woodii*, *Eubacterium limosum*, les *Staphylococcaceae*, les *Listeriaceae*, les *Erysipelotrichia* et les *Lactobacillales* sont suggérés comme étant non-sporulants.

B6, DacB et DacF sont trois PBP qui ont été prouvées comme étant impliquées dans la sporulation. Les co-pertes de ces familles avec les familles de la sporulation confirment l'existence de liens fonctionnels.

De façon plus surprenante, les familles FtsH3, CozE4 et MurA2 présentent une distribution taxonomique similaire au reste des familles du clan Spo. FtsH3 est un paralogue de FtsH1, montré comme étant une protéase du divisome chez *B. subtilis* [498]. Ces résultats suggèrent que cette familles correspond à une protéase spécifique de la sporulation.

CozE4 est une famille paralogue à CozE1 qui a été caractérisé chez *Staphylococcus aureus* et *Streptococcus pneumoniae* comme étant impliqué dans le contrôle de l'élongation et la régulation de PBP [442] [142]. Il est donc tentant de faire l'hypothèse que CozE4 remplisse le même rôle que CozE1 mais spécifiquement pour les PBP impliquées dans la sporulation, et particulièrement DacB dont la distribution taxonomique corrèle particulièrement avec celle de CozE4. Il est intéressant de noter que les familles CozE4 et FtsH3 n'ont jamais été décrites dans la littérature et que c'est notre analyse qui a révélé leur existence.

La famille MurA2, une famille paralogue à MurA1 qui a été décrite comme étant impliquée dans la première étape de synthèse des précurseurs du peptidoglycane [52] présente une distribution taxonomique similaire aux protéines de la sporulation. La famille MurA1 est aussi reliée évolutivement à la sporulation par le synténie puisqu'elle forme un cluster de gènes avec Mbl, SpoIID et SpoIIID. Les deux familles MurA1 et MurA2 sont donc reliées évolutivement à la sporulation mais pas aux autres processus qui nécessitent également la synthèse du peptidoglycane tel que l'élongation et la division.
La synténie entre Mbl et SpoIID et SpoIIID est aussi surprenante. La famille Mbl est une famille paralogue à celle de MreB et a été démontrée initialement comme impliquée dans l'élongation [239]. Elle forme comme MreB des patchs nécessaires à la synthèse du peptidoglycane périphérique lors de l'élongation chez *B. subtilis* [239]. Néanmoins, il a été également prouvé chez *Streptomyces coelicolor* que Mbl contribuait à la synthèse du peptidoglycane durant la sporulation [193]. Il est donc tentant de faire l'hypothèse que Mbl serait également impliqué dans la sporulation chez les *Firmicutes* comme *B. subtilis*.

3.6.1.4 Clan Mre/Min

Le clan Mre/Min est composé des familles MreB, MreC, MreD, MinC, MinD, MinE, MinJ, B5, RodA, Maf, ValS, et RadC. L'histoire évolutive de ce clan est présentée figure 3.48.

Description L'ensemble des familles de ce clan est inféré comme ancestral aux *Firmicutes*. Également, elles sont toutes inférées comme étant localisées dans le même cluster de gènes (cluster principal) excepté ValS et MinJ. Le cluster de gènes est conservé chez tous les *Firmicutes* bien que de nombreuses familles aient été perdues chez plusieurs *taxa*. Les familles MreC, MreD, RodA, ValS et RadC sont très conservées et peu affectées par des pertes de gènes. La famille MinJ semble avoir été perdue à de nombreuses branches et souvent co-perdue avec les familles MinC, MinD ou MinE (branches 120, 98, 135, 435, 526). La famille MreB est co-perdue avec MinC, MinD et MinJ chez l'ancêtre des Staphylococcaceae/*Erysipelotrichia* (branche 435) et celui des *Streptococcaceae*/*Enterococcaceae* (branche 526). La famille MinE

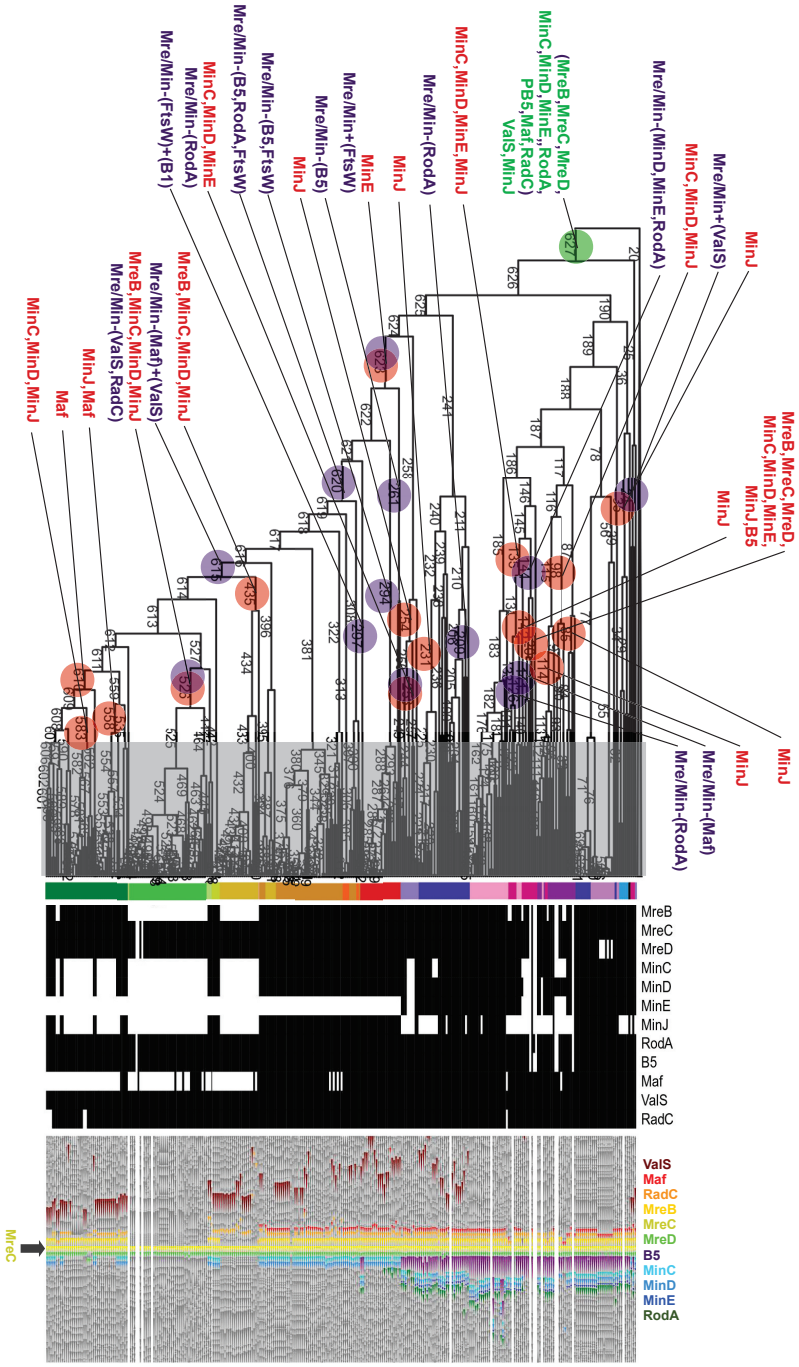


FIGURE 3.48 – Histoire évolutive du clan Mre/Min. Les événements évolutifs sont représentés par des ronds (vert : apparition, rouge : perte, violet : modification de clusters de gènes). La zone de l'arbre grisée correspond aux branches ou les événements évolutifs n'ont pas été représentés. La bande colorée correspond à la taxonomie. La distribution des familles de gènes est représentée par des rectangles noirs (présence) et blancs (absence). Le contexte génomique de MreC est représenté.

est perdue à l'émergence des *Bacilli* (branche 623). Enfin, la famille Maf est perdue chez la plupart des *Staphylococcaceae* et *Lactobacillales* mais certaines espèces semblent avoir gardé une copie, probablement transmise de façon verticale au vue de la phylogénie (annexe .12). Les copies des *Streptococcaceae* semblent néanmoins avoir été acquises par transmission horizontale ce qui indiquerait une perte avec ré-acquisition par transfert horizontal à l'émergence des *Streptococcaceae* (branche 525).

En terme d'organisation génomique, plusieurs modifications sont observées. Tout d'abord, le cluster principal semble exclure RodA et PBPB5 chez la plupart des *Bacilli* sauf chez les *Brevibacillus*. L'approche par parcimonie indique que la synténie du cluster principal avec B5 a eu lieu dans les branches 294 (émergence des *Paenibacillaceae*) et 620. La synténie entre le cluster principal et RodA a été perdue durant la diversification des *Paenibacillaceae* et également à la branche 620. À partir de la branche 615, Maf ne fait plus partie du cluster principal. À l'émergence des *Streptococcaceae/Enterococcaceae* (526), la synténie de RadC avec le cluster principal est perdue de façon concomitante avec la perte de MreB, MinC, MinD et MinJ. Il semble que la famille ValS soit voisine du cluster principal chez les *Clostridia* groupe 0 et 2 et chez les *Bacilli* bien que de nombreux gènes soient intercalés entre le cluster principal et les gènes ValS. À l'émergence des *Streptococcaceae/Enterococcaceae*, ValS n'est plus à la proximité du cluster principal.

Interprétation fonctionnelle De façon générale, l'ensemble de ces familles de gènes excepté MinJ forme un cluster conservé suggérant un lien fonctionnel étroit entre ces familles. Ce cluster de gènes a déjà été décrit dans la littérature [274]. Les familles MreB, MreC, MreD, RodA et B5 sont impliquées dans la synthèse périphérique de peptidoglycane durant la phase d'élongation tandis que MinC, MinD, MinE et MinJ forment un système de positionnement de l'anneau Z [472]. Il existerait donc un lien fonctionnel étroit entre ces deux systèmes. La conservation des familles MreC, MreD, RodA, ValS et RadC chez les *Firmicutes* indique des fonctions essentielles. Les familles MreB, MinC, MinD, MinE et MinJ ont été perdues à de nombreuses reprises et souvent co-perdues. Cette observation associée aux synténies conservées indique un lien évolutif d'autant fort entre ces familles.

Concernant les familles RodA et B5, la perte de synténie avec le cluster principal chez les

Bacilli parait surprenante. En effet, B5 et RodA ont été montrées comme impliquées dans l'élongation en association avec MreB/MreC/MreD chez *B. subtilis* [499], [472] mais ne sont retrouvées en cluster avec ces dernières familles que chez les *Clostridia*.
Les familles Maf et RadC font aussi partie du cluster principal Mre/Min mais n'ont jamais été prouvées comme étant reliées fonctionnellement avec le reste des familles du cluster. Maf a été montrée comme ayant un rôle dans la formation du septum de division chez *B. subtilis* [54]. Il possède des similarité structurales avec certaines ARNt synthétases [322] et a une activité nucléotide pyrophosphatase [460]. La famille RadC est quant à elle impliquée dans la réparation des cassure de l'ADN [415]. Il existe donc probablement un lien fonctionnel encore méconnu entre ces familles et celles impliquées dans l'élongation et le positionnement de l'anneau Z.
La famille ValS (Valyl-ARNt synthétase) semble également présenter une synténie avec le cluster principal mais de façon plus distante.

3.6.1.5 Clans Ori1 et Ori2

Les clans Ori1 et Ori2 sont composés respectivement des familles DnaA, DnaN, RecF et SpoIIIJ1, GidA, GidB, Jag, Noc, ParA, ParB. L'ensemble de ces gènes est à proximité de DnaA, marqueur de l'origine de réplication [295]. Leur histoire évolutive est présentée figure 3.49.

Description L'ensemble de ces familles est inféré comme étant ancestral aux *Firmicutes*. Étant donné que les deux clusters Ori1 et Ori2 sont situés majoritairement au début et à la fin des génomes et que l'inférence des synténies ancestrales ne prend pas en compte le fait que l'ADN est circulaire, deux clusters distincts ont été inférés. Nous avons donc rectifié ceci en indiquant que les deux clusters sont synténiques chez l'ancêtre des *Firmicutes*. La majorité des familles des deux clans sont très conservées mais Jag, Noc et ParA sont perdues à de nombreuses branches. Le cluster principal composé de l'ensemble des familles de gènes du clan est très conservé mais semble se dissocier à la branche 614 (ancêtre des *Listeriaceae*/*Lactobacillales*). En effet, GidB, Noc, ParA et ParB forment un cluster indépendant éloigné de l'origine de répli-

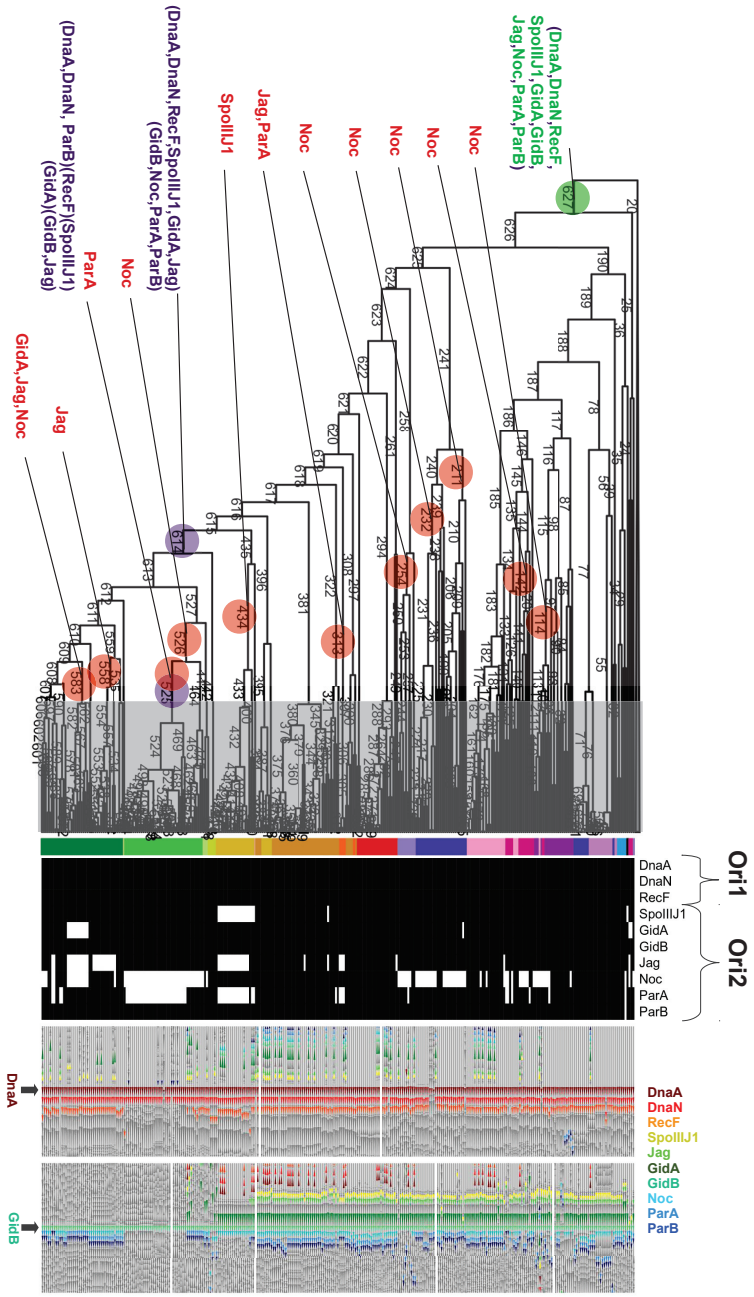


FIGURE 3.49 – Histoire évolutive des clans Ori1 et Ori2. Les événements évolutifs sont représentés par des ronds (vert : apparition, rouge : perte, violet : modification de clusters de gènes). La zone de l'arbre grisée correspond aux branches ou les événements évolutifs n'ont pas été représentés. La bande colorée correspond à la taxonomie. La distribution des familles de gènes est représentée par des rectangles noirs (présence) et blancs (absence). Les contextes génomiques de DnaA et GidB sont représentés.

cation. À l'émergence des *Streptococcaceae*, une dissociation du cluster est également observée ce qui a conduit à 5 locus chez les *Streptococcaceae* : DnaA/DnaN/ParB, GidB/Jag, GidA, RecF et SpoIIIJ1.

Interprétation fonctionnelle L'ensemble de ces familles semble ainsi relié fonctionnellement de par leur proximité au sein des génomes de *Firmicutes*. La nature de ces liens n'est pas évidente à inférer mais il semble que la majorité des familles en question est impliquée dans les mécanismes cellulaires informationnels liés à l'ADN et à l'ARN. En effet, DnaA et DnaN sont impliqués dans l'initiation de la réplication [191], RecF dans la recombinaison homologue [184], ParA et ParB dans la ségrégation du chromosome [214] et GidA et GidB dans la modification des ARNt [155], [405]. Il semblerait donc que ces processus cellulaires soient liés. Ce n'est pas surprenant puisque la réplication, la ségrégation et la réparation doivent être cordonnées afin de mener à bien la séparation du matériel génétique lors du cycle cellulaire.
Néanmoins, le rôle de GidA et GidB parait plus éloigné fonctionnellement. En effet, la modification des ARNt n'a jamais été prouvée comme étant couplée au reste des processus décrits précédemment. Cependant, la délétion de GidA semble altérer la division cellulaire et l'expression de YmgF (composant du divisome absent chez les *Firmicutes*) chez *E. coli* et [279]. GidA et GidB agiraient donc peut-être en régulant l'expression des gènes du cycle cellulaire par modification des ARNt.
Enfin, Jag et SpoIIIJ1 ont été montrés comme étant impliqués de concert dans la sporulation chez *B. subtilis* [130]. Jag présente un domaine de liaison à l'ARN et son activité est régulée par phosphorylation par la kinase StkP chez *S. pneumoniae* [477]. Jag a également été montrée comme étant impliquée dans la régulation de l'élongation chez *S. pneumoniae* [443]. SpoIIIJ1 est impliquée directement dans la sporulation en régulant l'activité du facteur sigma F [130].
Les clans Ori1 et Ori2 forment donc une unité régulant et coordonnant plusieurs processus cellulaires liés à l'ADN et aux ARN durant le cycle cellulaire. Il est intéressant de noter que chez les *Streptococcaceae*, le clan est dissocié ce qui suggère que la coordination entre la réplication, la ségrégation, la réparation et la modification des ARNt pourrait être différente chez ce taxon.

3.6.1.6 Clan Scp/Xer et Smc

Le clan Scp/Xer/Smc est composé des familles ScpA, ScpB, XerC, XerD, Smc et FtsY. La dénomination de clan est discutable pour ce cas car les deux groupes Scp/Xer et Smc/FtsY ne sont reliés que par une arête dans le réseau et distants dans l'arbre de distance. De meme, le lien entre ScpA/B et XerC/D semble relativement ténu. En effet, XerC ne présente de réelle synténie conservée avec aucune des familles analysées et seulement un lien issu des profils phylogénétiques avec XerD. Pourtant, ScpA et ScpB ont des Z-scores de corrélation issu des synténies élevés avec XerC (Z-score=0,331). Ce score élevé s'explique par le fait que la famille XerC n'est pratiquement jamais voisine avec les familles analysées. Les seules familles voisines détectées sont ScpA, ScpB et SpoIIIJ1 pour seulement trois génomes. La synténie est donc largement surévaluée par le Z-score. Néanmoins, le contexte de XerC semble extrêmement conservé (DprA, topoisomerase I, TrmF0, ...) et le fait que ScpA et ScpB soient intégrés dans ce cluster de gènes chez les *Acidaminococcus* ne relève pas forcément du hasard.
Malgré cela, nous les avons traité ensemble afin de mettre en lumière leur histoire commune figure 3.50.

Description Toutes les familles du clan sont inférées comme étant présentes chez l'ancêtre des *Firmicutes*. ScpA et ScpB sont en contexte génomique avec LysA1, XerD avec DacF et Smc avec FtsY. ScpA et ScpB sont systématiquement voisins alors que la synténie avec LysA1 est perdu à de nombreuses branches. XerC et XerD sont perdus à plusieurs branches, notamment chez les *Clostridia* mais aussi à l'émergence des *Streptococcaceae* (525). D'après la phylogénie et le scénario de réconciliation de la famille XerD, il semble que la présence d'une copie au sein des génomes de *Paenibacillaceae* provienne d'un remplacement homologue à la branche 294 (annexe .12), induisant ainsi une co-perte de XerC et XerD à cette branche. XerD est proche de ScpA et ScpB au sein des génomes de certains *Firmicutes* mais surtout à partir de la branche 615 (émergence des *Staphylococcaceae*/*Erysipelotrichia*/autres *Bacilli*). Cette synténie est perdue à l'émergence des *Streptococcaceae*. L'histoire de la famille Smc et de celle ScpA/ScpB semblent très similaires puisque les profils phylogénétiques sont strictement identiques. Néanmoins, Smc n'est pas relié avec ScpA/B par le contexte génomique mais avec FtsY. Le voisinage entre ces deux familles est très conservé. À partir de l'émergence des *Streptococcaceae*, un certain nombre de gènes s'insèrent entre les deux.

Interprétation fonctionnelle Le lien fonctionnel entre ScpA, ScpB et Smc est déjà caractérisé puisque ces trois familles forment un système impliqué dans la ségrégation des chromosome. Le lien fonctionnel entre ScpA/B et XerD, une famille impliquée dans la résolution des dimères de chromosome, n'a jamais été démontré. Néanmoins, il ne parait pas surprenant que la ségrégation et la résolution des dimères de chromosome soient couplés puisqu'il existe déjà un lien entre les deux processus cellulaires. En effet, XerC et XerD interagissent directement avec FtsK qui est impliqué dans la ségrégation de la région *ter* du chromosome [71],[177]. Le lien entre Smc et FtsY est plus compliqué à expliquer. En effet, FtsY est une protéine impliquée dans la translocation de protéines et la biogenèse des protéines membranaires, qui ferait probablement partie du divisome [290], [422], [291]. Néanmoins, il a été démontré que FtsY et Smc étaient co-transcrits [233] . Il est possible que le lien fonctionnel entre FtsY et Smc constitue un moyen supplémentaire de coordination entre la division et la ségrégation chromosomique. Enfin, il est intéressant de noter que la perte de XerD à l'émergence des *Streptococcaceae* est accompagnée d'une perte progressive du voisinage entre Smc et FtsY. Il serait tentant de faire l'hypothèse

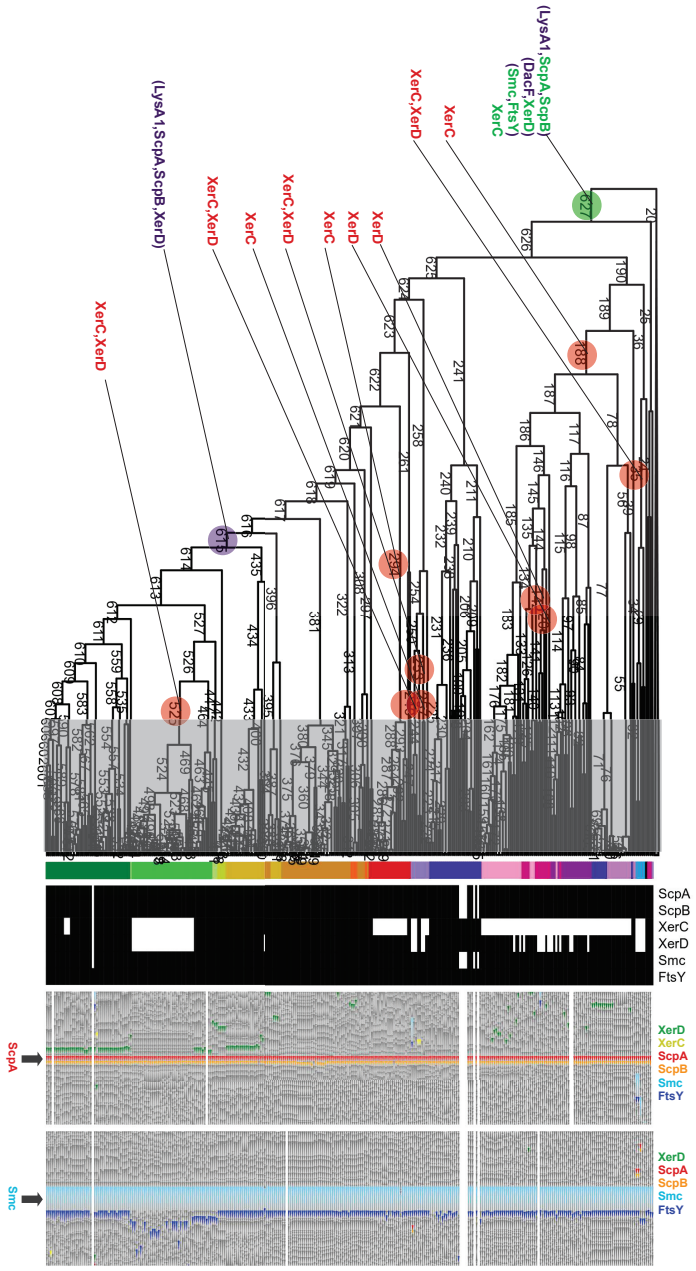


FIGURE 3.50 – Histoire évolutive du clan Scp/Xer/Smc. Les événements évolutifs sont représentés par des ronds (vert : apparition, rouge : perte, violet : modification de clusters de gènes). La zone de l'arbre grisée correspond aux branches ou les événements évolutifs n'ont pas été représentés. La bande colorée correspond à la taxonomie. La distribution des familles de gènes est représentée par des rectangles noirs (présence) et blancs (absence). Les contextes génomiques de ScpA et Smc sont représentés.

que cette perte ait eu un impact sur le lien fonctionnel entre Smc et FtsY.

3.6.1.7 Autres clans

Clan Cps Le clan Cps est composé des familles CpsB, CpsC et CpsD, toutes impliquées dans la régulation de la synthèse de la capsule [352]. L'histoire évolutive de ce clan est présentée figure 3.51. Les trois familles sont inférées comme étant anciennes chez les *Firmicutes*. CpsC et CpsD sont inférés comme ayant apparu à la branche 626 (émergence des *Clostridia* hors group 0) mais il est plus probable qu'elles aient été présentes chez l'ancêtre des *Firmicutes* avec une perte chez les *Clostridia* groupe 0. De façon intéressante, de nombreux génomes possèdent plusieurs copies de ces trois familles. Bien que l'ensemble des familles de gènes sont globalement en accord avec la phylogénie d'espèces, de nombreux transferts horizontaux tardifs sont observés (annexe .12).
Les profils phylogénétiques des trois familles sont très similaires et de nombreuses co-pertes sont observées durant la diversification des *Firmicutes*. Il est intéressant de noter que CpsC semble plus conservée que CpsB et CpsD indiquant peut-être une fonction plus essentielle que les autres, notamment chez les *Clostridia*. Les trois familles forment un cluster chez la plupart des *Firmicutes*. Cependant, quelques *taxa* ne présentent pas de cluster avec les trois gènes comme les *Paenibacillaceae*. L'analyse de la phylogénie des trois familles indique une remplacement homologue par transfert horizontal à l'émergence des *Streptococcaceae*. Cela suggère un co-transfert de ces trois gènes chez l'ancêtre des *Streptococcaceae*. Il est aussi intéressant de noter que l'ordre des gènes dans le cluster chez les *Streptococcaceae* est différent de la plupart des *Firmicutes* (CpsB, CpsC, CpsD). De façon plus générale, le clan Cps semble complètement isolé du reste des clans. Or, la production de la capsule a été montrée comme liée fonctionnellement avec le cycle cellulaire chez *S. pneumoniae*. Ces résultats suggèrent que ce lien n'est probablement pas conservé chez les *Firmicutes*.

Clan Fem Le clan Fem est composé des familles FemA, FemB, Fmh et MGT1. L'histoire évolutive de ces familles est présentée figure 3.52. FemA, FemB et Fmh co-apparaissent à

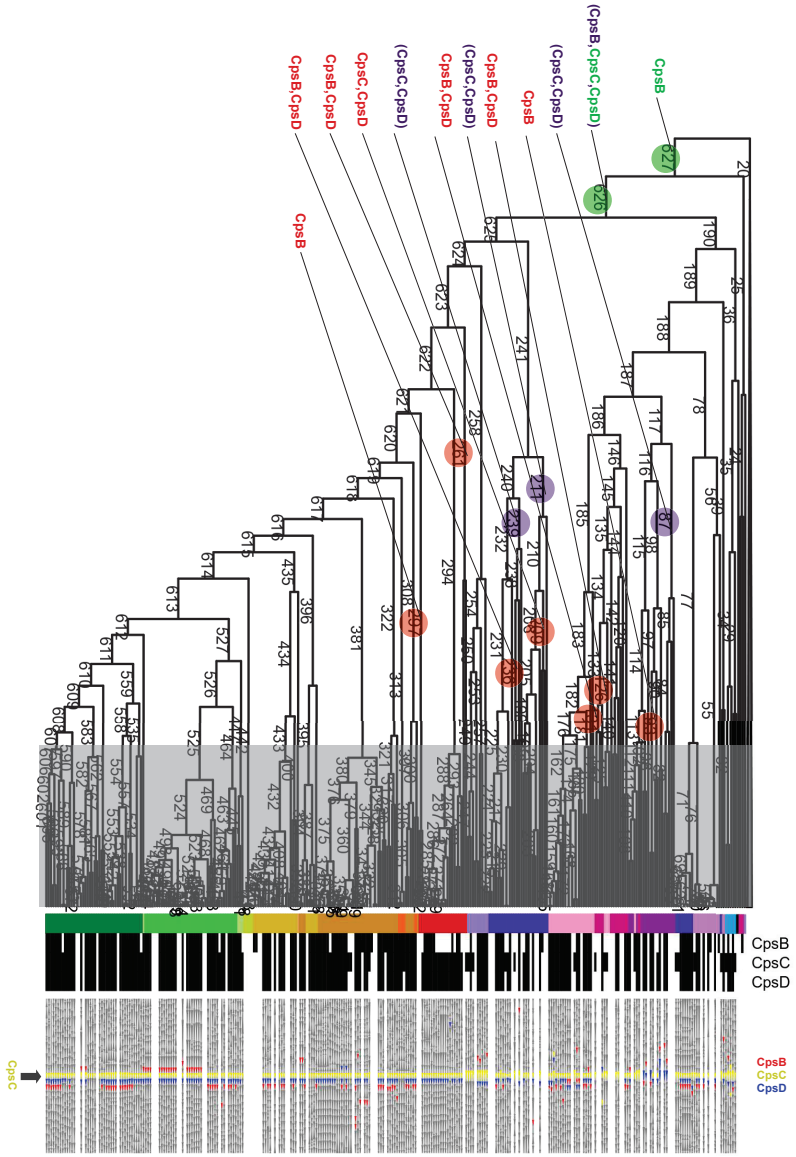


FIGURE 3.51 – Histoire évolutive du clan Cps. Les événements évolutifs sont représentés par des ronds (vert : apparition, rouge : perte, violet : modification de clusters de gènes). La zone de l'arbre grisée correspond aux branches ou les événements évolutifs n'ont pas été représentés. La bande colorée correspond à la taxonomie. La distribution des familles de gènes est représentée par des rectangles noirs (présence) et blancs (absence). Le contexte génomique de CpsC est représenté.

l'émergence des *Staphylococcaceae*. Ces trois familles font partie de la même famille multigénique. Leur apparition est probablement due à une succession de duplications puisque ces trois familles sont très proches au sein de la phylogénie et leurs gènes sont voisins chez *Salinococcus halodurans* et *Jeotgalibacillus*. FemA et FemB forment un cluster de gènes conservé chez tous les *Staphylococcaceae*. MGT1 apparaît à l'émergence des *Staphylococcus*. Le regroupement de MGT1 dans le clan Fem est du à sa distribution taxonomique qui est très similaire à celles de FemA, FemB et Fmh. MGT1 est une PBP impliquée dans la réticulation du peptidoglycane et FemA/FemB/Fmh sont impliquées dans la synthèse du pont interpeptidique du peptidoglycane. Un lien fonctionnel entre MGT1 et FemA/FemB/Fmh est donc peu surprenant. Néanmoins, étant donné la distribution réduite de ces familles, le lien fonctionnel entre MGT1 et FemA/FemB/Fmh par la distribution taxonomique pourrait être artefactuel.

Clan StkP Le clan StkP est composé des familles StkP, PhpP, SunL, PriA et Mtf. Toutes les familles de ce clan sont inférées comme étant présentes à l'ancêtre des *Firmicutes* et formant un cluster de gène (figure 3.53). Les familles sont extrêmement conservées ainsi que le cluster de gène. Il est donc naturel de penser qu'un lien fonctionnel étroit existe entre ces familles de gènes. Le lien entre StkP et PhpP est évident puisqu'il s'agit d'une Sérine/Thréonine kinase et de sa phosphatase associée [359]. StkP est impliqué dans de nombreux processus cellulaires tels que la division cellulaire, la synthèse du peptidoglycane ou le métabolisme [300]. Néanmoins, le reste des familles n'est apparemment pas relié ni à StkP, ni à PhpP. En effet, PriA a été montré comme impliqué dans la ré-initiation de la réplication [195], SunL est une ARN méthyltransferase et Mtf est une Methionyl-ARNt synthétase [297]. Le cluster du clan StkP constituerait donc une unité de régulation de processus variés tels que la division cellulaire, la réplication et la régulation de la traduction.

Clan FtsH Le clan FtsH est composé de cinq familles : FtsH1, DivIC, Mfd, TilS et YabM. Ce sont des familles très conservées qui sont liées essentiellement par le contexte génomique (figure 3.54). Les cinq familles sont présentes chez l'ancêtre des *Firmicutes* et organisées en deux clusters de gènes en association avec des familles impliquées dans la sporulation :

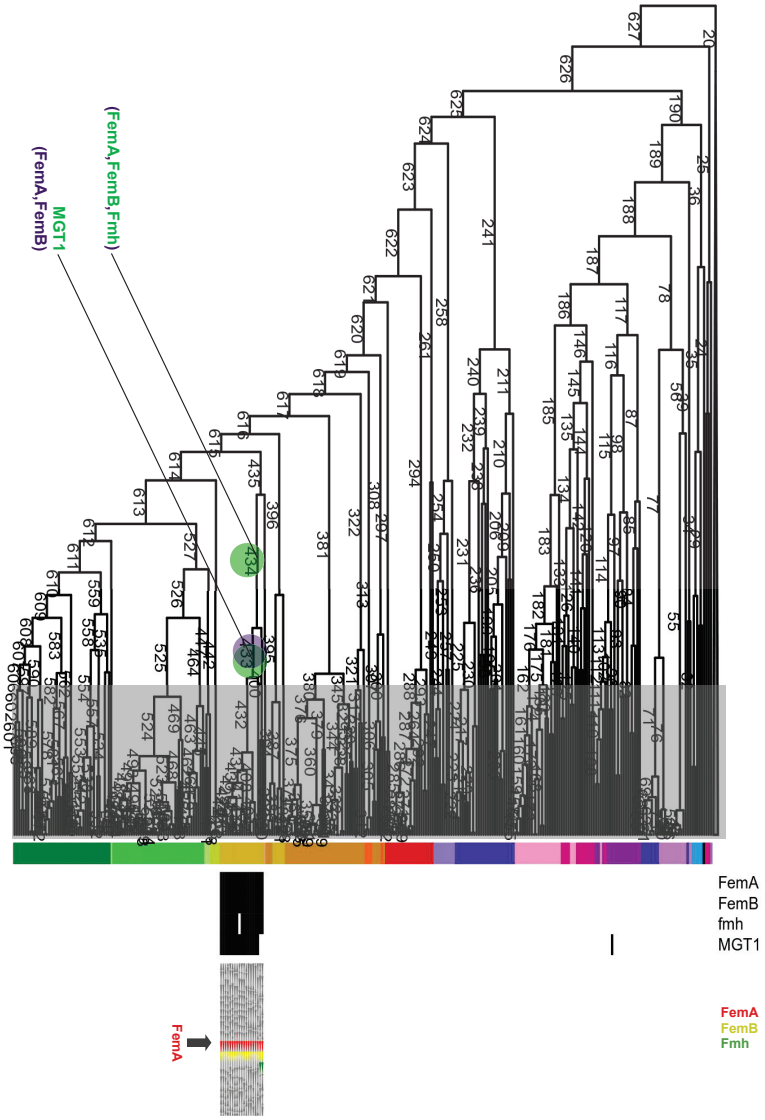


FIGURE 3.52 – Histoire évolutive du clan Fem. Les événements évolutifs sont représentés par des ronds (vert : apparition, rouge : perte, violet : modification de clusters de gènes). La zone de l'arbre grisée correspond aux branches ou les événements évolutifs n'ont pas été représentés. La bande colorée correspond à la taxonomie. La distribution des familles de gènes est représentée par des rectangles noirs (présence) et blancs (absence). Le contexte génomique de FemA est représenté.

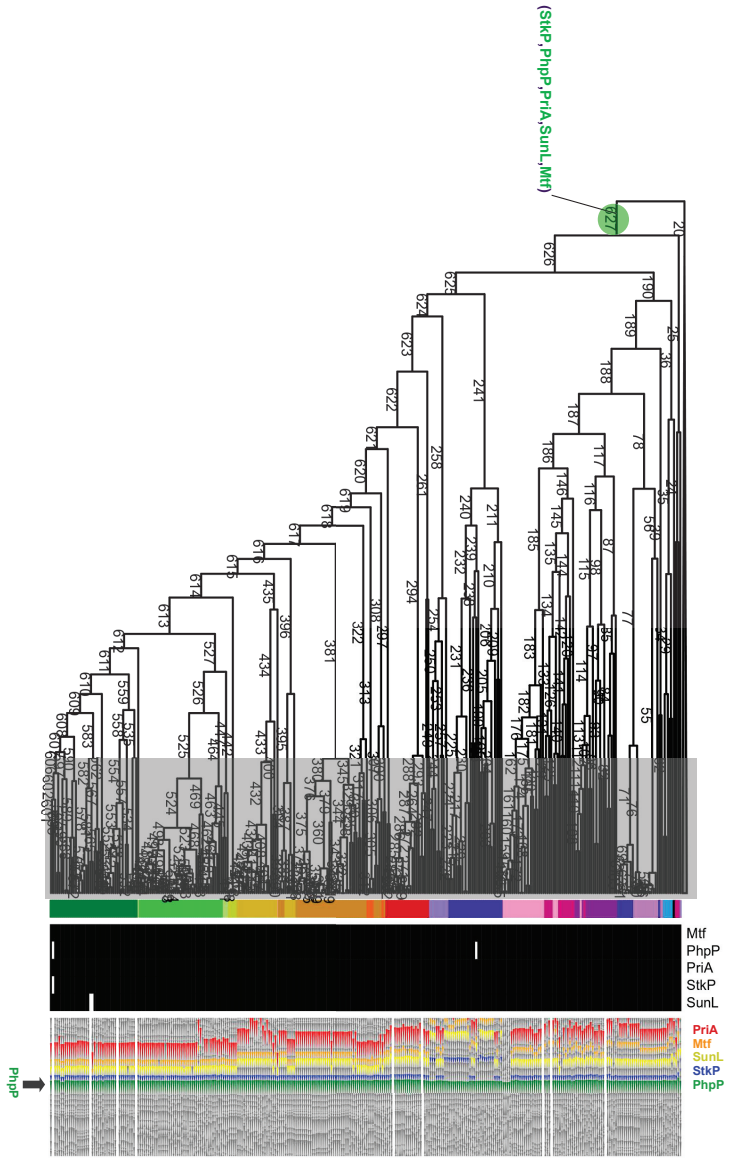


FIGURE 3.53 – Histoire évolutive du clan StkP. Les événements évolutifs sont représentés par des ronds (vert : apparition). La zone de l'arbre grisée correspond aux branches ou les événements évolutifs n'ont pas été représentés. La bande colorée correspond à la taxonomie. La distribution des familles de gènes est représentée par des rectangles noirs (présence) et blancs (absence). Le contexte génomique de PhpP est représenté.

FtsH1/SpoIIE/TilS et Mfd/SpoVG/GlmU/YabM. Ces deux clusters sont proches aux sein des génomes chez la plupart des *Firmicutes* mais séparés par un certain nombre de gènes eux aussi impliqués dans la sporulation (YabP, YabQ, ...).
Il apparaît donc que l'ensemble de ces familles soient étroitement reliées au processus de sporulation. Il est difficile de faire des hypothèses sur la nature des liens fonctionnels entre l'ensemble de ces familles puisque ces gènes sont impliqués dans des processus cellulaires très variés : FtsH1 et DivIC sont des composants du divisome [498], [491], Mfd est impliqué dans le couplage transcription-réparation de l'ADN [330], TilS est impliquée dans la modification des ARN [448] et YabM est un homologue de MurJ dont la fonction n'a pas encore été décrite [482]. À cela s'ajoutent les familles de la sporulation qui présentent une synténie conservée avec ces dernières comme SpoIIE et SpoVG. Néanmoins, quelques liens fonctionnels au sein de ce clan ont déjà été démontrés. Par exemple, il a été prouvé que TilS favorise la transcription de FtsH1 chez *B. subtilis* [282]. Ce clan constituerait donc potentiellement une unité de régulation de nombreux processus cellulaires tels que la division cellulaire, la sporulation ou la réparation.

Clan Wal Le clan Wal est composé des familles WalH, WalI, WalJ, WalK, WalR et MurZ. Les gènes codant pour WalH, WalI, WalK et WalR sont inférés comme étant présents et en cluster chez l'ancêtre des *Firmicutes* (figure 3.55). WalJ et MurZ sont inférées comme ayant apparu à la branche 626 (émergence des *Clostridia* hors groupe 0). Néanmoins, encore une fois, le placement de l'apparition de ces familles à cette branche n'est peut-être qu'artefactuel et qu'elle est en fait placée à la racine. WalH, WalI, WalK et WalR semblent co-perdus à de nombreuses branches, notamment chez les *Clostridia* et les *Negativicutes*. Ils semblent être aussi systématiquement organisés en clusters au sein des génomes. WalJ est également majoritairement en contexte génomique avec WalH, WalI, WalK et WalR mais semble avoir une histoire évolutive indépendante ce ces derniers. En effet, sa distribution taxonomique est très différente de celle des autres gènes Wal. MurZ semble avoir une distribution taxonomique relativement similaire à WalJ et est en contexte avec le cluster Wal chez quelques *Clostridia*.

WalK et WalR forment un système à deux composant impliqué notamment dans la régulation de la synthèse du peptidoglycane et le métabolisme montré chez plusieurs organismes

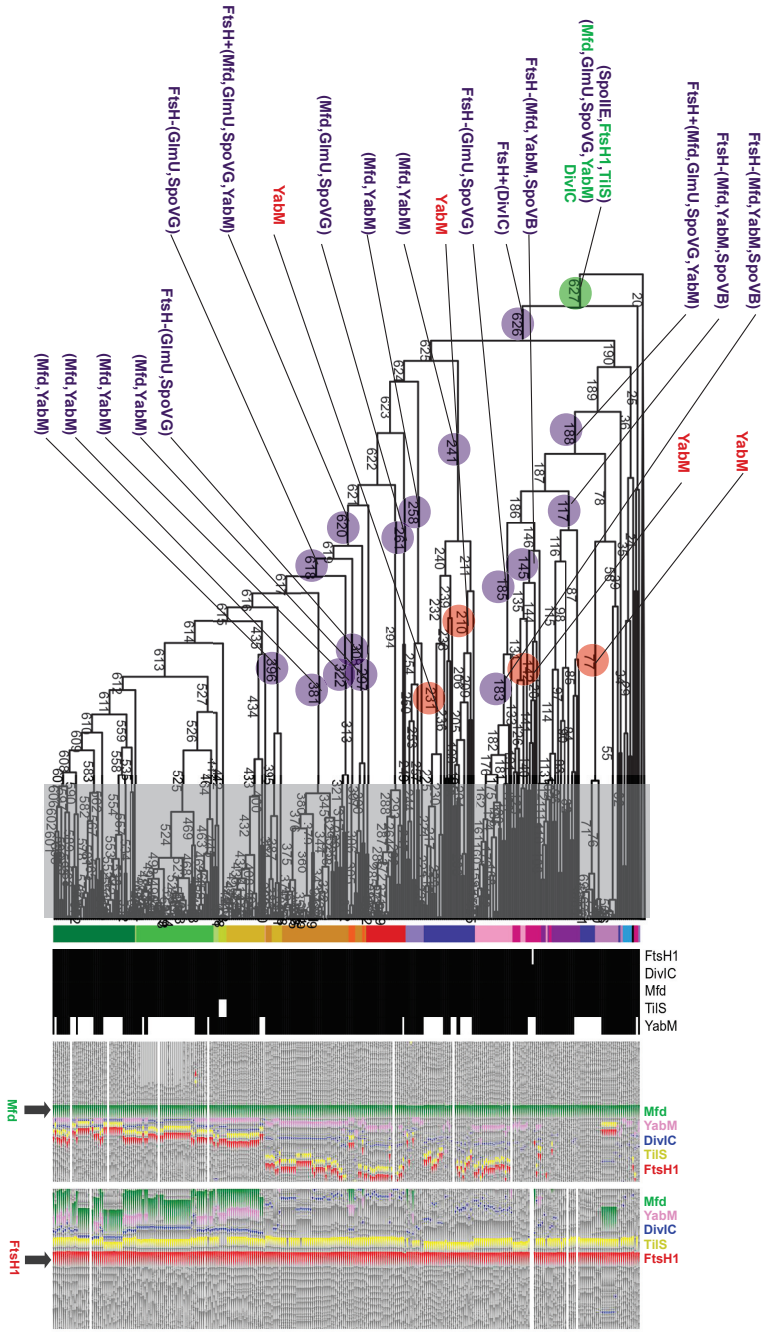


FIGURE 3.54 – Histoire évolutive du clan FtsH. Les événements évolutifs sont représentés par des ronds (vert : apparition, rouge : perte, violet : modification de clusters de gènes). La zone de l'arbre grisée correspond aux branches ou les événements évolutifs n'ont pas été représentés. La bande colorée correspond à la taxonomie. La distribution des familles de gènes est représentée par des rectangles noirs (présence) et blancs (absence). Les contextes génomiques de FtsH1 et Mfd sont représentés. FtsH : cluster contenant FtsH1.

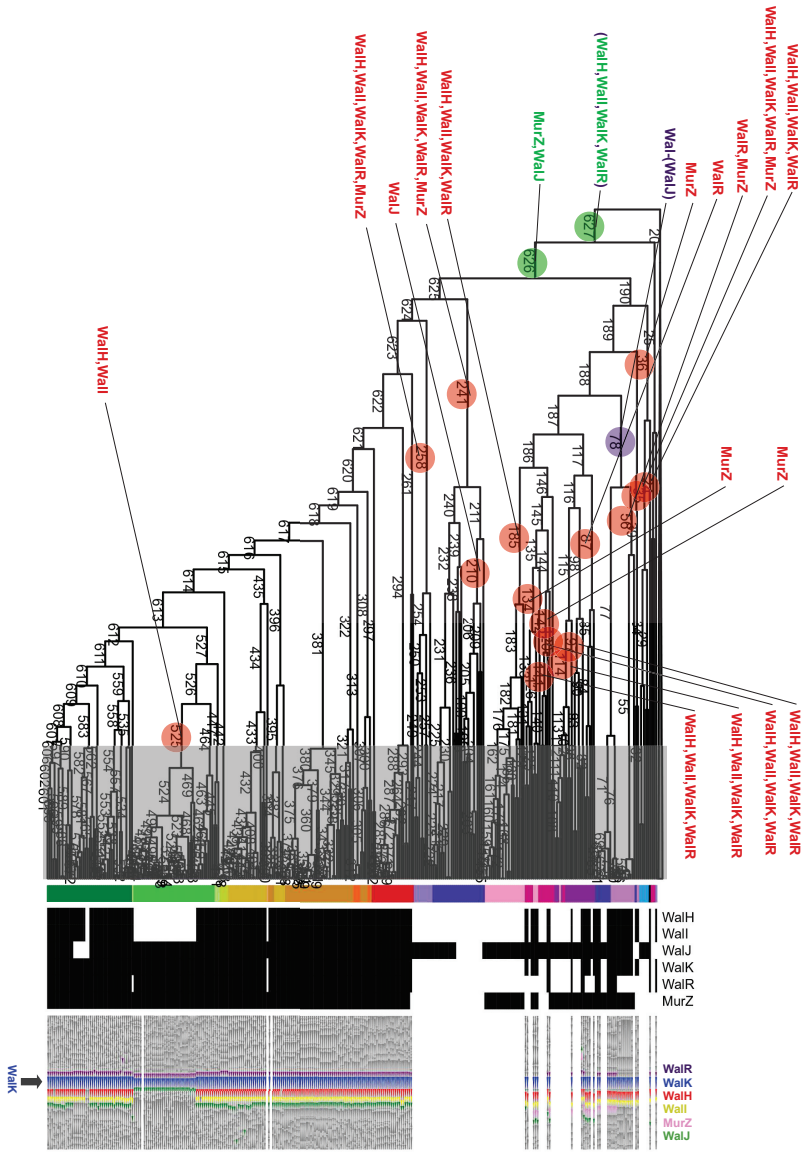


FIGURE 3.55 – Histoire évolutive du clan Wal. Les événements évolutifs sont représentés par des ronds (vert : apparition, rouge : perte, violet : modification de clusters de gènes). La zone de l'arbre grisée correspond aux branches ou les événements évolutifs n'ont pas été représentés. La bande colorée correspond à la taxonomie. La distribution des familles de gènes est représentée par des rectangles noirs (présence) et blancs (absence). Le contexte génomique de WalK est représenté.

tels que *S. pneumoniae*, *B. subtilis* et *S. aureus* [89]. Notamment, WalR et WalK sembleraient réguler l'activité et l'expression d'hydrolases chez *S. pneumoniae*, *B. subtilis* et S aureus. WalH et WalI ont été décrites comme interagissant avec WalK et inhibant son activité kinase chez *B. subtilis* [451]. La famille WalJ semble avoir un rôle plus éloigné de ce système bien qu'étant relié avec celui-ci, ce qui expliquerait son histoire évolutive différente des autres familles Wal. En effet, il a été suggéré que WalJ serait impliquée dans la coordination des processus de ségrégation et de division cellulaire [451]. Enfin, MurZ semble relié fonctionnellement au familles Wal et plus particulièrement avec WalJ. Sachant que le système Wal est impliqué dans la régulation de la synthèse du peptidoglycane et que MurZ est impliqué dans l'étape initiale de la synthèse des précurseurs du peptidoglycane [52], il est possible de faire l'hypothèse que le système Wal pourrait réguler directement la synthèse du peptidoglycane en régulant l'activité de MurZ.

Clan Nag/Mur Ce clan est composé des familles AnmK, MurK, MurQ, NagA et NagB. Ces familles sont impliquées dans le recyclage du peptidoglycane [226]. Ces familles sont reliées principalement par leur distribution taxonomique (figure 3.56) et la réconciliation. Néanmoins, leurs histoires évolutives sont très complexes car impactées par de multiples transferts horizontaux et/ou paralogies cachées. Leur contexte génomique varie énormément d'une espèce à l'autre mais il semble qu'elles présentent un certain nombre de synténies.

3.6.2 Description et interprétation fonctionnelle de liens évolutifs ponctuels

Certains liens évolutifs ponctuels sont surprenants au vue de la littérature. Nous allons décrire succinctement ces liens de façon non-exhaustive.

3.6.2.1 Lien fonctionnel BX1-FtsW

Les familles BX1 et FtsW semble très corrélées de par leur histoire évolutive (figure 3.57). La famille BX1 correspond à une familles de PBP qui n'a, à notre connaissance, pas encore

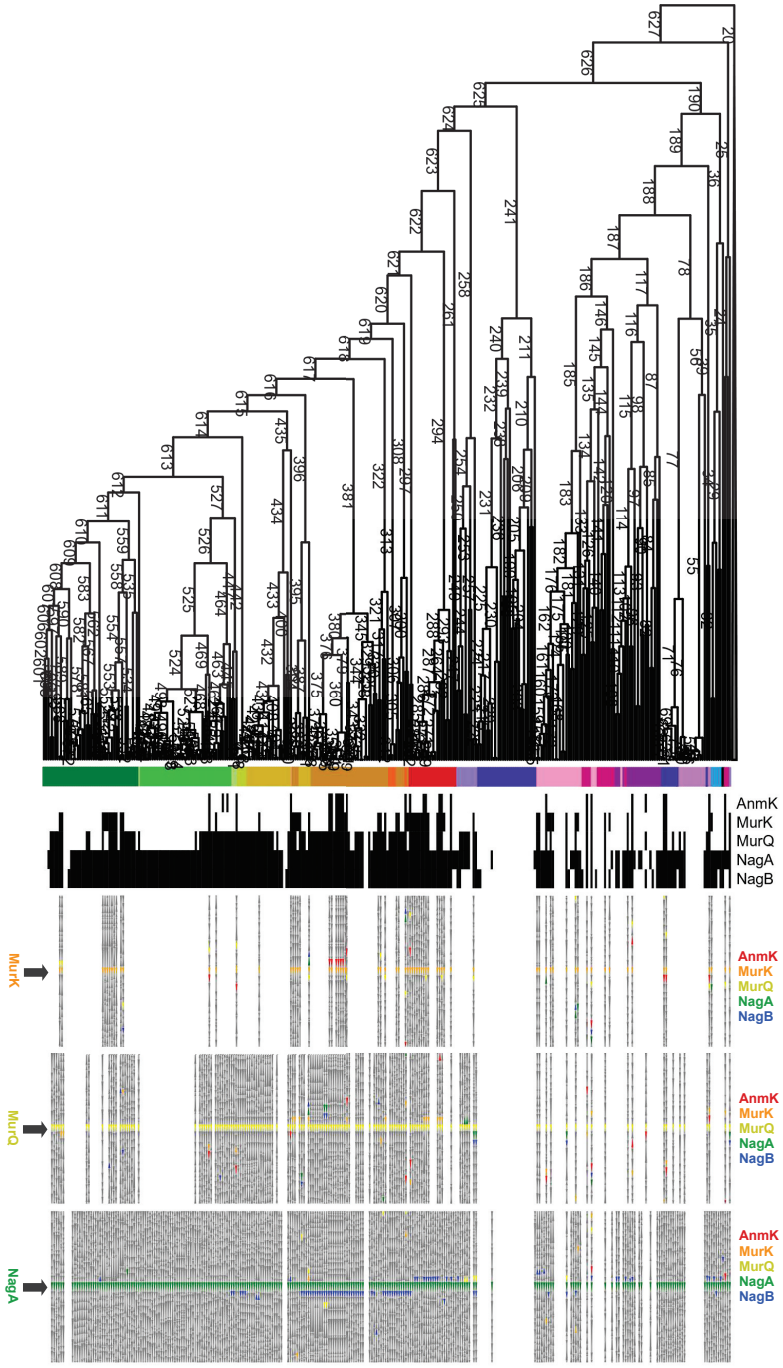


FIGURE 3.56 – Distribution taxonomique et contexte génomique du clan Nag/Mur. La bande colorée correspond à la taxonomie. La distribution des familles de gènes est représentée par des rectangles noirs (présence) et blancs (absence). Les contextes génomiques de MurK, MurQ et NagA sont représentés.

été caractérisée expérimentalement et FtsW est impliquée dans la synthèse du peptidoglycane, soit en tant que flippase du Lipide II, soit en tant que peptidoglycane polymérase [327], [312]. Les distributions taxonomiques de ces dernières sont quasiment identiques chez les *Clostridia* notamment de par de nombreuses co-pertes. De plus, elles présentent une synténie très conservées chez les *Clostridia* et une forme fusionnée des deux gènes est même retrouvée chez *Lachnoclostridium phytofermentans*. À l'émergence des *Bacilli*, BX1 est perdu mais pas FtsW. Cela suggère que FtsW présente potentiellement un mécanisme différent chez les *Clostridia* et les *Bacilli*. Il a déjà été montré que FtsW interagissait avec une PBP, PBPB3 chez Mycobacterium tuberculosis [82]. Également, FtsW nécessite le recrutement de PBPB4 afin de localiser au septum chez *Bacillus subtilis* [154]. Néanmoins, aucune corrélation n'est observée entre FtsW et ces PBP.

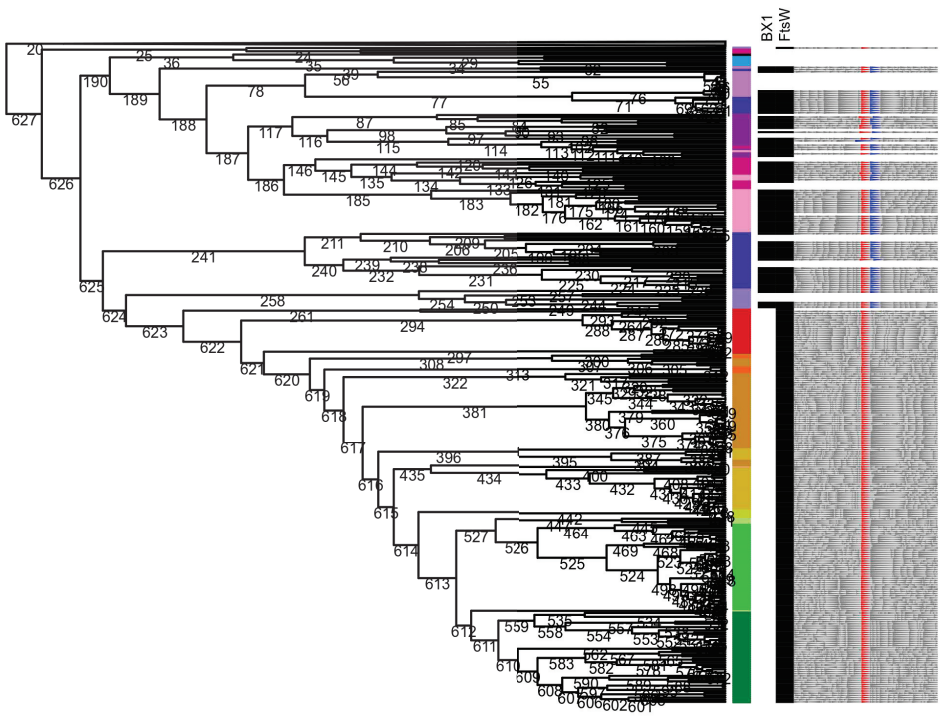


FIGURE 3.57 – Histoire évolutive de BX1 et FtsW. La bande colorée correspond à la taxonomie. La distribution des familles de gènes est représentée par des rectangles noirs (présence) et blancs (absence). Le contexte génomique de FtsW est représenté.

3.6.2.2 Lien fonctionnel RecN-FtsJ

Une synténie conservée est retrouvée entre les familles RecN et FtsJ (figure 3.58). La synténie est observée chez tous les *Firmicutes* excepté les *Caldicellulosiruptor* et les *Staphylococcaceae*. Un gène codant pour un régulateur transcriptionnel est retrouvé entre les deux gènes chez la plupart des espèces. La protéine FtsJ chez *E. coli* a été décrite comme méthylant l'ARN 23S [55] tandis que RecN est impliqué dans la réparation des cassures double brin chez *B. subtilis* [242]. Ces résultats suggèrent qu'il pourrait exister un couplage fonctionnel entre ces gènes et donc par extension entre la réparation de l'ADN et la régulation de la traduction.

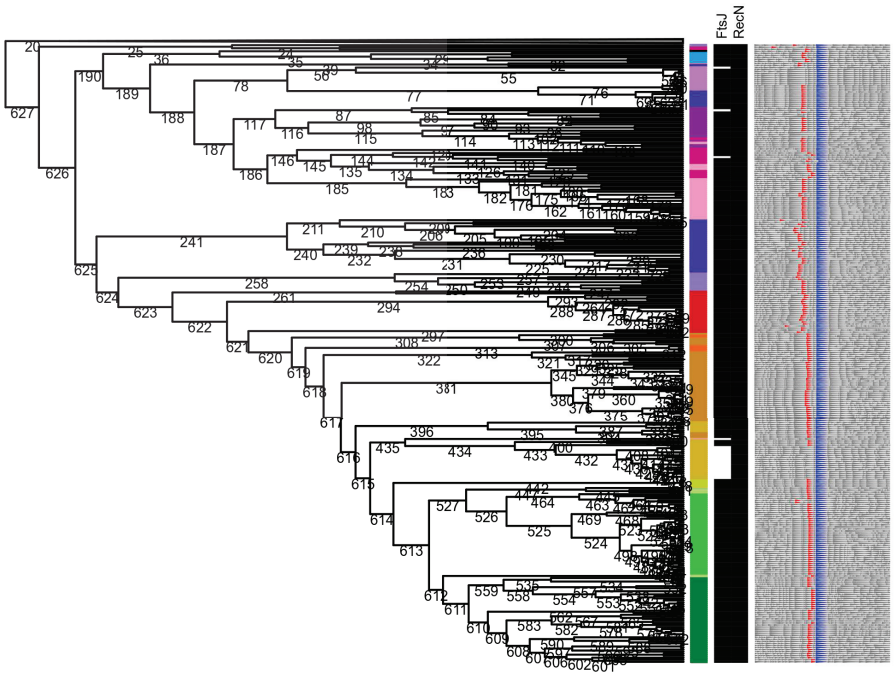


FIGURE 3.58 – Histoire évolutive de FtsJ et RecN. La bande colorée correspond à la taxonomie. La distribution des familles de gènes est représentée par des rectangles noirs (présence) et blancs (absence). Le contexte génomique de RecN est représenté.

3.6.2.3 Lien fonctionnel FtsE/X-MinJ

Les familles FtsE/X et MinJ présentent également une synténie et dans une moindre mesure des profils phylogénétiques similaires (figure 3.59). Les gènes codant pour MinJ sont cependant

distants de plusieurs gènes de ceux codant pour FtsE/X. Cela suggère un lien fonctionnel déjà démontré entre le positionnement de l'anneau Z et le divisome.

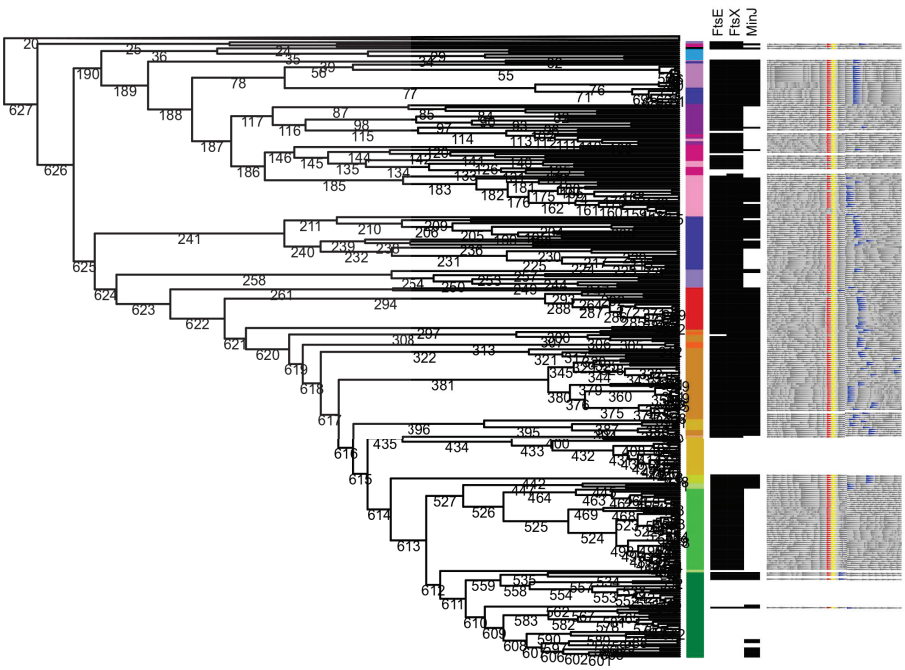


FIGURE 3.59 – Histoire évolutive de FtsE, FtsX et MinJ. La bande colorée correspond à la taxonomie. La distribution des familles de gènes est représentée par des rectangles noirs (présence) et blancs (absence). Le contexte génomique de FtsX est représenté.

3.6.2.4 Lien fonctionnel MapZ-MurN/M

Une corrélation des profils phylogénétiques des familles MapZ, MurN, MurM et CDP3 est observée (figure 3.60). D'après la distribution taxonomique et le scénario de réconciliation, MapZ, MurN et MurM émergent à la branche 526 (ancêtre des *Streptococcaceae*/*Enterococcaceae*). Ces résultats suggèrent un lien fonctionnel entre MapZ, MurN et MurM. Pour le cas de CDP3, la situation est plus complexe. En effet, la phylogénie semble indiquer que CDP3 est apparu à la branche 525 (ancêtre des *Streptococcaceae*) mais présente un grand nombre de séquences d'autres *Lactobacillales* groupées ensemble mais très hétérogène en terme de taxonomie (annexe .11). Ces séquences sont notamment mélangées avec des copies plasmidiques indiquant

potentiellement des transferts par échanges de plasmides. La corrélation avec CDP3 parait donc artéfactuelle et sur-évaluée non seulement parce que CDP3 apparaît à une branche différente des trois autres familles mais aussi parce que les quatre familles de gènes ne sont que peu étendue au sein des *Firmicutes* .

MapZ a été montrée comme étant impliquée dans le positionnement de l'anneau Z chez *S. pneumoniae* [149] et MurN/MurM synthétisent le pont interpeptidique lors de la synthèse des précurseurs du peptidoglycane [284]. La synthèse du peptidoglycane s'effectuant au septum chez *Streptococcus pneumoniae*, il est tentant de faire l'hypothèse qu'il existerait un lien direct entre MapZ et la synthèse du peptidoglycane via MurN/MurM.

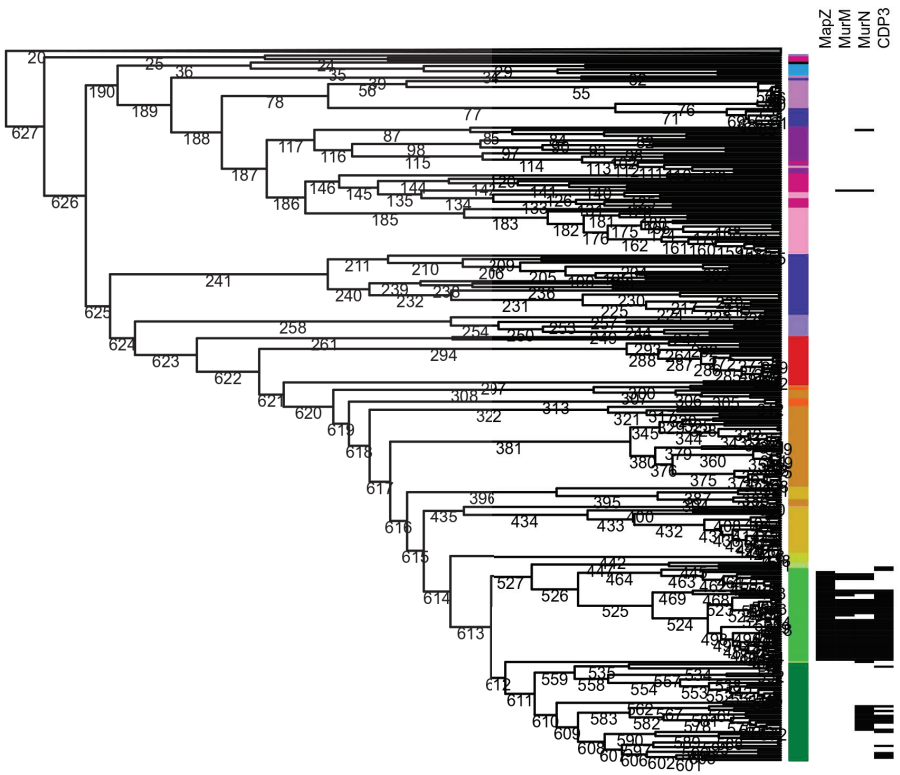


FIGURE 3.60 – Distribution taxonomique de MapZ/MurN/M et CDP3. La bande colorée correspond à la taxonomie. La distribution des familles de gènes est représentée par des rectangles noirs (présence) et blancs (absence).

3.6.2.5 Lien fonctionnel MurG1-MurG2

Enfin, les familles MurG1 et MurG2, impliquées dans la production du Lipide II durant la synthèse du peptidoglycane [315], présentent des distributions taxonomiques opposées (figure 3.61). Ces deux familles sont des paralogues très proches comme observé sur la phylogénie de la famille MurG1/MurG2 (figure 3.61). Il semblerait donc que deux copies de MurG aient été présentes chez l'ancêtre des *Firmicutes* et que des pertes différentielles aient eu lieu lors de la diversification des *Firmicutes*. En terme de contexte génomique, MurG1 est situé systématiquement dans le cluster DCW tandis que MurG2 ne présente pas un contexte conservé. Néanmoins, chez les *Streptococcus*, MurG2 est localisé dans le cluster DCW.

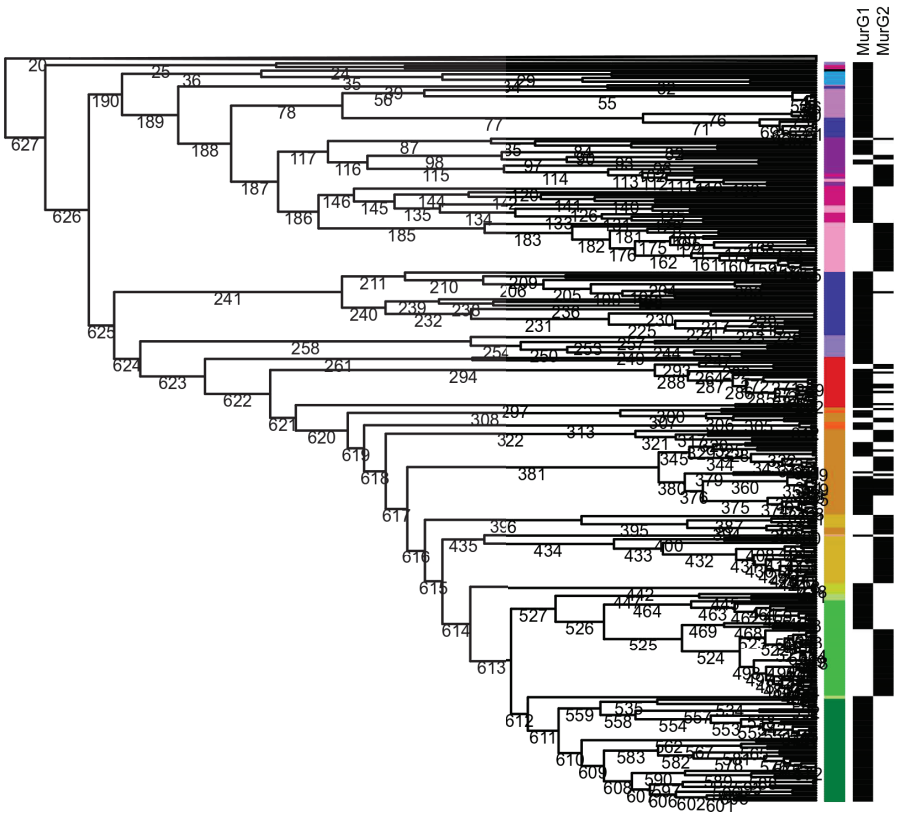


FIGURE 3.61 – Distribution taxonomique de MurG1 et MurG2. La bande colorée correspond à la taxonomie. La distribution des familles de gènes est représentée par des rectangles noirs (présence) et blancs (absence).

3.6.3 Observations générales et conclusion

3.6.3.1 Le cycle cellulaire est relié fonctionnellement à la traduction

Au cours de cette analyse, nous avons étudié les gènes associés ou situés à proximité des clusters de gènes impliqués dans le cycle cellulaire chez *B. subtilis* et *S. pneumoniae*. Nos données montrent qu'une majorité de ces gènes sont impliqués dans la modification des ARN, la transcription et la traduction. La fonction cellulaire de chacune de ces familles est présentée table 3.6.

Famille	Clan/Famille lié(e) évolutivement	Activité	Rôle cellulaire
Mtf	Clan SktP	Méthionyl-ARNt-formyl synthétase	Traduction
SunL	Clan SktP	ARN méthyltransferase	Méthylation
TilS	Clan FtsH	ARNt(Ile)-lysidine synthétase	Transcription de FtsH chez *B. subtilis*
IleS	Clan DCW1	Isoleucyl-tRNA synthétase	Traduction
ValS	Clan Mre/Min	Valyl-ARNt synthétase	Traduction
LeuS	MurJ	Leucyl-ARNt synthétase	Traduction
FtsJ	RecN	ARN 23S méthyltransferase	Mutant : défaut de division chez *E. coli*
Pfs	MacP	5'-méthylthioadenosine/S-adenosylhomocysteine nucléosidase	Méthylation, synthèse des polyamines, des vitamines, quorum-sensing
MraW	Clan DCW1	ARN 16S méthyltransferase	Transcription des gènes du DCW chez *E. coli*
Maf	Clan Mre/Min	nucléotide pyrrophosphatases	Amplification : arrêt de la division chez *B. subtilis*
GidA	Clan Ori2	Méthylènetetrahydrofolate--ARNt-(uracile-5-)-méthyltransferase	Mutant : défaut de division chez *S. suis*, *S. mutans* mais pas chez *S. pyogenes*
GidB	Clan Ori2	ARN méthyltransferase	Mutant : défaut de division chez *Salmonella* si conditions de stress
Jag	Clan Ori2	Fixation aux ARN	Traduction de FtsA, et de protéines de la synthèse de la paroi chez *S. pneumoniae*

TABLE 3.6 – Familles impliquées dans la traduction ou interagissant avec les ARN.

Ainsi, les clans majeurs mis en évidence dans notre étude contiennent au moins une famille impliquée dans des processus liés aux ARN. Ces familles sont impliquées dans la traduction/modification des ARN à différents niveaux, comme indiqué figure 3.62. Maf impacte le taux global d'ARN, GidA et TilS modifient les anti-codons des ARNt, Pfs est impliqué dans le cycle Homocystéine/Méthionine, Mtf modifie spécifiquement les Methiony-ARNt, SunL, GidB, MraW et FtsJ méthylent les ARN et LeuS, ValS, et AlaS synthétisent les ARNt. Sous l'hypothèse que ces gènes sont co-transcrits avec les gènes du cycle cellulaire au vue de leur contexte génomique, il existerait donc un couplage entre le cycle cellulaire et la régulation de la traduction chez les bactéries, particulièrement au niveau transcriptionnel. De façon intéressante, les

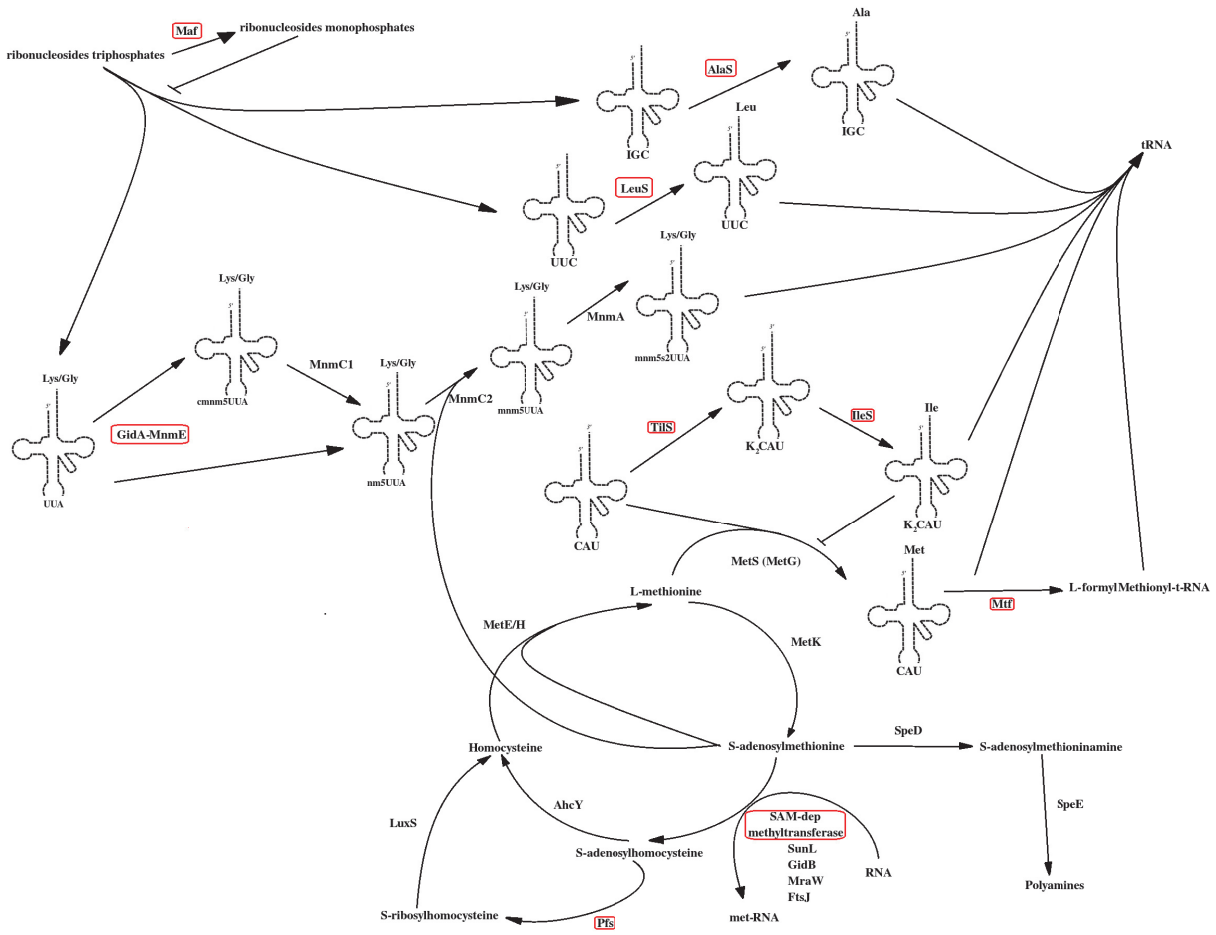


FIGURE 3.62 – Rôles des familles de gènes impliquées dans la traduction. Les familles étudiées dans cette étude sont entourées en rouge.

famille Maf [54], MraW [125], FtsJ [466], Jag [524], GidA [155] et GidB [429] ont été décrites comme également impliquées dans la division cellulaire. Il a déjà été montré que MraZ module la transcription des gènes du cluster DCW [125] et que TilS régule la transcription de FtsH [282]. De plus, Jag a été décrite comme régulant la production de nombreuses protéines de synthèse du peptidoglycane et de la division cellulaire [524]. Les autre familles (Mtf, SunL, TilS, IleS, ValS, LeuS, Pfs) n'ont pas été montrées comme étant impliquées dans le cycle cellulaire ou la division cellulaire.

Il est tentant de faire l'hypothèse que l'ensemble des familles nommées précédemment régulent la transcription et/ou la traduction des gènes et/ou ARN impliqués dans le cycle cellulaire.

Ces familles de gènes seraient peut-être la clé de la compréhension de la temporalité des événements moléculaires pendant le cycle cellulaire. Pour aller plus loin, il est nécessaire d'estimer la chance d'avoir un gène impliqué dans la traduction au sein d'un cluster de gènes par pur hasard. En effet, si les gènes impliqués dans la traduction et la modification des ARN sont très nombreux, la chance d'obtenir au hasard un tel gène à un locus donné sera très élevée.

3.6.3.2 Implications fonctionnelles liées à deux points chauds évolutifs

Émergence des *Streptococcaceae* L'émergence des *Streptococcaceae* est accompagnée de multiples événements évolutifs. Tout d'abord, plusieurs familles de gènes sont perdues : WalI, WalH, ParA, XerC, XerD et MraZ. La famille CDP3 est acquise et CpsB/CpsC/CpsD ont subi un remplacement homologue. De façon concomitante, plusieurs clusters de gènes ont été réorganisés. Le cluster DCW a été scindé en deux : RecN, MraY, MraW, FtsL, FtsJ, B4 d'une part et FtsA, PCDP6, PCDP7, PCDP8, MurD, IleS1, SepF, MurG1, FtsQ, DivIVA, FtsZ d'autre part. Le cluster Mre/Min perd également deux gènes qui se localisent ailleurs dans le génome : RadC et ValS. La distance entre les gènes codant pour Smc et FtsY devient plus grande et MapZ est associé au cluster GpsB/PCDP10/RecU/PBPA3 tandis que DnaD est séparé de ce dernier. Enfin, le cluster d'origine de réplication est fragmenté en 5 sous-clusters.
Plusieurs hypothèses peuvent être formulées à propos de ce point chaud. Tout d'abord, ParA qui a été montré comme étant le partenaire de ParB est perdu [151] [23] [214] [341] alors que CpsD, un homologue de ParA est acquis par transfert horizontal de même que CpsB et CpsC. Il semblerait donc qu'un remplacement non orthologue de ParA par CpsD ait eu lieu. CpsD a été montré comme interagissant avec ParB chez *Streptococcus pneumoniae* [352]. L'autophosphorylation de CpsD semble médier l'interaction CpsD-ParB. Également, CDP3 qui est acquis de façon concomitante a été récemment décrit comme interagissant avec CpsD et ParB et coordonnant la ségrégation des chromosomes, la division cellulaire et la production de la capsule [318]. La perte de ParA, l'acquisition par transfert horizontal de CpsB/C/D et le gain de CDP3 sont donc potentiellement à l'origine d'une innovation en terme de couplage entre la production de la capsule, la division cellulaire et la ségrégation chromosomique. Au vue de l'ensemble de ces observations et du fait que le clan Cps semble complètement isolé du reste

des clans, il est envisageable de faire l'hypothèse que le lien entre la production de la capsule et la division cellulaire est spécifique des *Streptococcaceae*.
A cela s'ajoute le fait que le cluster d'origine de réplication dans lequel se trouve ParB a été fragmenté en cinq sous-clusters. Les gènes codant pour ParB/DnaA/DnaN restent à l'origine de réplication tandis que ceux codant pour RecF, SpoIIIJ1, GidA, GidB et Jag sont localisés ailleurs dans le chromosome. Il est possible que ce réarrangement chromosomique soit lié aux événements décrits ci-avant. Les deux clusters impliqués dans la ségrégation chromosomique sont également modifiés puisque XerD initialement situé à proximité de ScpA et ScpB est perdu et que la synténie entre Smc et FtsY ne semble plus conservée. L'ensemble de ces résultats suggèrent que l'émergence des *Streptococcaceae* soit accompagnée de la mise en place de mécanismes de régulation du cycle cellulaire très différents des autres *Firmicutes* et notamment au niveau de la ségrégation chromosomique.

Émergence de *Acetobacterium woodii/Eubacterium limosum* *Acetobacterium woodii* et *Eubacterium limosum* sont deux *Clostridia* qui émergent au sein du groupe 1D2. L'émergence de ces deux espèces est accompagnée d'un grand nombre de pertes, comme indiqué dans la section 3.4.2.2. La quasi-totalité des clans Wal, Mre/Min, Spo, Nag/Mur ainsi que les familles XerD, BX1, FtsW et DivIVA sont perdus à la branche correspondante. Tout d'abord, ces résultats indiquent que *Acetobacterium woodii* et *Eubacterium limosum* ne sont pas capable de sporuler ce qui d'ailleurs a été montré [88]. L'absence des familles du clan Mre/Min indiquent que la synthèse de peptidoglycane lors de l'élongation doit être très différente chez ces espèces par rapport aux bactéries modèles. L'élongation chez ces bactéries pourrait donc être réalisée par un système différent restant à découvrir. De plus, RodA et SpoVE sont présents mais FtsW est perdu ainsi que RodA2. RodA et SpoVE sont donc les seuls membres de la famille RodA/FtsW/SpoVE. L'absence des familles du clan Nag/Mur indique que ces bactéries ne recyclent pas le peptidoglycane ou par un autre moyen restant à découvrir. Enfin, ces deux espèces ne présentent aucune des recombinases responsables de la résolution des dimères de chromosome XerC, XerD et XerS. *Acetobacterium woodii* et *Eubacterium limosum* possèdent donc probablement un système alternatif afin d'éviter les dimères de chromosomes et les dégâts que cela génère.

3.6.3.3 De nouveaux systèmes sont à décrire pour certains processus cellulaires

Les systèmes de positionnement de l'anneau Z Trois systèmes régulant le positionnement de l'anneau Z et présents chez les *Firmicutes* ont été caractérisés. Il s'agit du système Min impliquant MinC, MinD, MinE, DivIVA et MinJ, le système NO impliquant Noc et le système MapZ. Les familles de gènes SsgA/B, PomZ, MipZ et SlmA décrits respectivement chez *Myxococcus xanthus*, *Streptomyces coelicolor*, *Caulobacter crescentus* et *Escherichia coli* sont absentes chez les *Firmicutes*. La distribution taxonomique des systèmes NO, Min et MapZ a déjà été étudiée chez les *lactobacillales* dans une revue que nous avons publié en 2016 [158]. La distribution de ces systèmes de régulation de l'anneau Z chez l'ensemble des *Firmicutes* est présentée figure 3.63. Certains taxa ne présentent aucun de ces systèmes comme les *Leuconostocaceae*, les *Acidaminococcus*, *Aerococcus urinae*, certains *Lactobacillaceae*, *Mageeibacillus indolicus*, *Acetobacterium woodii*, *Finegoldia magna*, *Megasphaera elsdenii* et *Veillonella parvula*. Ces résultats suggèrent que des systèmes alternatifs existent chez ces *taxa* dans le positionnement de l'anneau Z.

Les systèmes de la résolution des dimères de chromosomes Lors de la réplication, la réparation des cassures génère des dimères de chromosomes. Les recombinases XerC, XerD et XerS sont responsable de la résolution des dimères de chromosomes [71], [177], [419], [265]. La distribution taxonomique de ces trois familles représentée en figure 3.64 indique que certains *taxa* ne possèdent aucune de ces recombinases : les *Halanaerobiales*, les *Ruminococcus*, les *Selenomonas*, *Acetohalobium arabaticum*, *Acetobacterium woodii*, *Clostridium cellulovorans*, *Eubacterium limosum*, *Mageeibacillus indolicus*, *Parvimonas micra*, *Peptoniphilus*, *Veillonella parvula* et *Weissella koreensis*. La résolution des dimères de chromosomes s'effectue donc par un système alternatif chez ces espèces. Néanmoins, il est à noter que la famille XerC/D/S est une familles multigénique très large qui comprend d'autres sous-familles non étudiées (annexe .11). Il est donc possible que d'autres membres de la famille XerC/D/S aient pris le relais chez ces espèces pour la résolution des dimères. Une étude spécifique de la distribution taxonomique de chaque sous-famille de cette famille serait nécessaire pour aller plus loin.

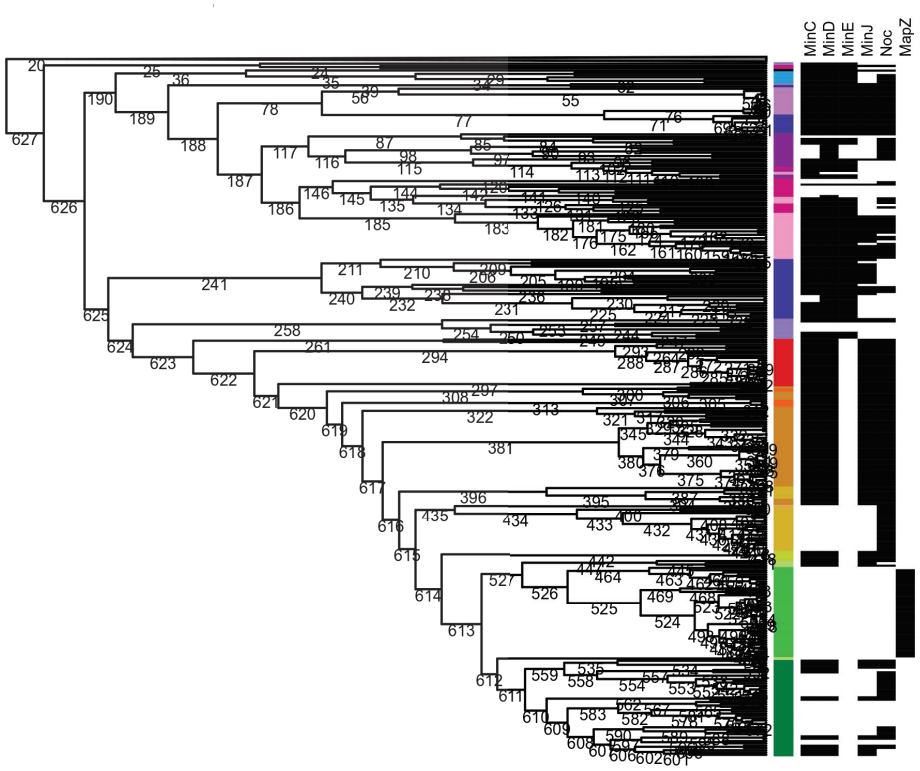


FIGURE 3.63 – Distribution taxonomique des systèmes de positionnement de l'anneau Z. La bande colorée correspond à la taxonomie. La distribution des familles de gènes est représentée par des rectangles noirs (présence) et blancs (absence).

3.6.3.4 Conclusion

L'approche que nous avons utilisé pour établir des liens entre les familles de gènes s'appuie sur des informations évolutives et génomiques et non sur les fonction de gènes ou les liens fonctionnels déjà connus. Il s'agit donc d'une méthode sans *a priori*. Nous avons ainsi mis en exergue de nombreux potentiels liens fonctionnels entre les familles de gènes du cycle cellulaire. Certains d'entre eux ont déjà été démontrés expérimentalement, ce qui valide la pertinence de l'approche, tandis que d'autres restent à prouver. Il apparaît que tous les processus du cycle cellulaire sont reliés évolutivement, au niveau de l'organisation génomique (contexte génomique) et de l'équipement génétique (profils phylogéniques et réconciliation). La traduction semble aussi être intimement liée avec le cycle cellulaire. De nombreux gènes impliqués dans le cycle cellulaire restent probablement à découvrir.

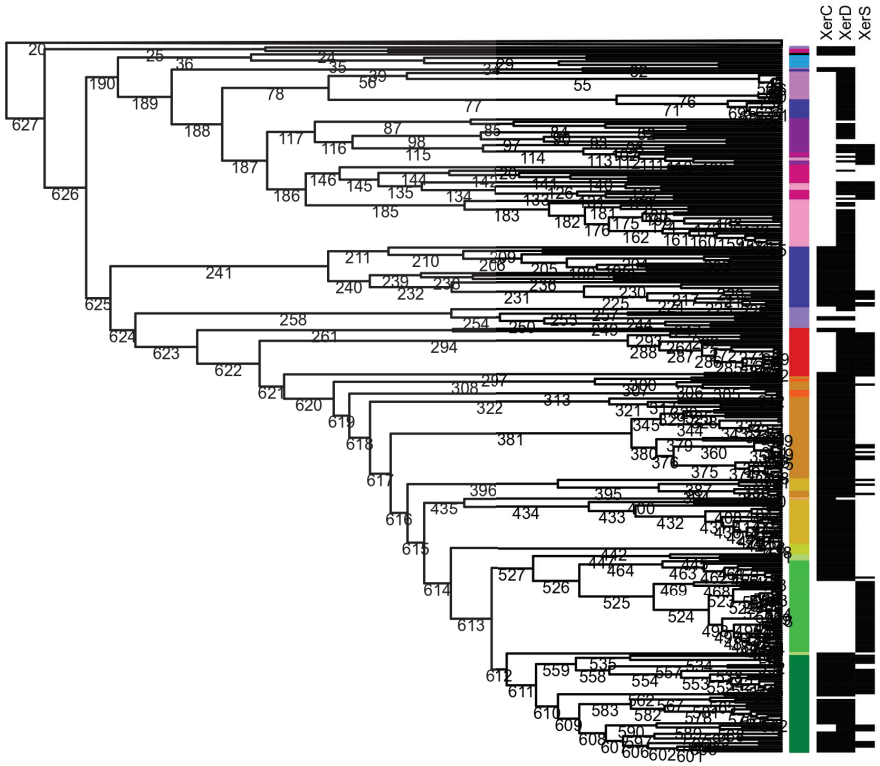


FIGURE 3.64 – Distribution taxonomique des systèmes de résolution de dimères de chromosomes. La bande colorée correspond à la taxonomie. La distribution des familles de gènes est représentée par des rectangles noirs (présence) et blancs (absence).

3.7 Identification de nouvelles familles potentiellement impliquées dans le cycle cellulaire

Nous avons appliqué l'approche développée afin d'essayer d'identifier les familles de gènes possiblement reliées au cycle cellulaire, mais jamais caractérisées. Nous avons ainsi recherché dans les familles de gènes des *Firmicutes* celles présentant des histoires évolutives similaires à celles impliquées dans le cycle cellulaire. Bien que les résultats présentés dans cette section soient préliminaires, il apparaissait intéressant de les présenter pour illustrer une autre application possible de notre approche.

3.7.1 Méthode d'identification des familles de gènes corrélant avec les familles du cycle cellulaire

Pour construire les familles de gènes des *Firmicutes*, nous avons utilisé une méthode de clusterisation qui a pour l'avantage d'être rapide mais plus grossière que les méthodes développées dans la section 3.3. Nous avons d'abord construit une base de données de protéomes de 304 *Firmicutes* correspondant à une souche par espèce (annexe .2).Une comparaison des séquences protéiques des 304 protéomes de base de données avec elle-même a été réalisée (BLASTP all vs all, option de sortie tabulaire -m 8).
Les familles de gènes ont ensuite été construites à partir de ce fichier à l'aide de SiLiX 1.2.9 [320]. Le fonctionnement de SiLiX repose sur le SLC. Deux séquences de la base de données sont regroupées si elles présentent un pourcentage d'identité supérieur à un seuil donné et que l'alignement par pair est suffisamment long. Cette approche génère ainsi des clusters de séquences qui constituent des familles de gènes. Les seuils d'identité ont été fixés de 0,3 à 0,7 et le taux de recouvrement à 80% (défaut). Afin d'estimer quel seuil de pourcentage d'identité utiliser, nous avons vérifié pour 10 familles du cycle cellulaire si les bases de données issues de SiLiX contenaient ces familles. Ces familles (DnaI, EzrA, FemA, IleS1, MinE, MurG1, ParA, SbcC1, SpoIIE et WalK) ont été choisies pour la confiance accordée à leur construction, la variété de leur distribution taxonomique et le fait qu'elles appartiennent à des familles monogéniques et multigéniques. Nous avons donc calculé les coefficients de corrélation à partir de l'équation 3.11 entre la distribution taxonomique de chacune des 10 familles et l'ensemble des familles de chaque base de données SiLiX. Nous avons ensuite testé si chaque famille avait été détectée dans les familles construites par SiLiX (figure 3.65). Ainsi, le seuil maximisant le nombre de familles retrouvées est 0,4. Nous avons donc utilisé cette base de données de familles de gènes. L'ensemble des familles de cette base de données ont été nommées FT pour Familles de la base de données Totale.

Le profil phylogénétique des familles générées a ensuite été comparé avec celui des 179 familles de gène du cycle cellulaire (nommés FCC pour Familles du Cycle Cellulaire) en utilisant

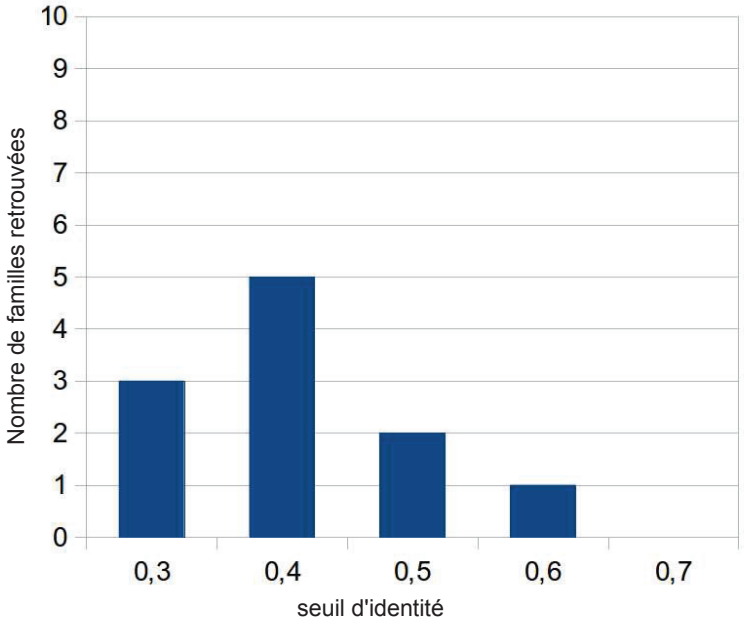


FIGURE 3.65 – Nombre de familles retrouvées dans les bases de données SiLiX par seuil d'identité.

la moyenne du coefficient de corrélation ϕ et de l'information mutuelle comme décrit dans l'équation 3.15. Les couples possédant un score global supérieur à 0,7 en ont été sélectionnés. Les FT associées sont qualifiées de FT proies. Il est aussi important de rappeler que pour certains processus du cycle cellulaire, nous n'avons pas analysé tous les gènes (*e.g.* sporulation). Cette omission peut générer des faux négatifs.

Afin d'identifier les FT proies qui présenteraient un lien particulièrement étroit avec le cycle cellulaire, nous avons voulu savoir quelles FT proies présentaient également une synténie conservée avec des gènes du cycle cellulaire. Nous avons ainsi calculé les coefficients de similarité à partir des synténies entre les familles identifiées et celles précédemment construites, d'après l'équation 3.11.

3.7.2 Familles corrélées évolutivement aux familles du cycle cellulaire

3.7.2.1 Description et annotation des FT proies

Description générale des FT proies En utilisant la méthode précédemment décrite, 62 FCC corrèlent avec au moins une FT (annexe .14). Au total, 185 FT proies ont été identifiées. De façon attendue, certaines FT proies correspondaient à des FCC. Pour connaître le nombre de ces familles, nous avons compté le nombre FT proies dont plus de 50% des séquences appartenaient à une FCC. Ainsi, 51/185 FT proies remplissaient cette condition. Nous avons voulu savoir combien parmi elles avaient été détéctées par la FCC qui leur était associée. Ainsi, 40 de ces 51 FT ont été identifiées par leur FCC associée. Ces FCC sont principalement celles qui ont été construite avec facilité (familles monogéniques ou sous-familles d'une famille multigénique avec une grande distance phylogénétique avec les autres sous-familles). Les 11 restantes correspondaient à des FCC identifiées par d'autres FCC.
En ramenant au nombre de FCC total, 40/179 FCC ont donc corrélé avec une FT correspondant à cette même FCC. Pour les 139 autres cas, il existe une discordance entre les familles générées avec SiLiX et les familles reconstruites par le pipeline. Ainsi, les deux méthodes semblent fournir des familles différentes.
Les FCC des clans *Bacilli* et Spo corrèlent avec un grand nombre de FT. Ces deux clans sont principalement constitués de liens issus des profils phylogénétiques. Il semblerait donc que de nombreux autres gènes appartiennent à ces clans. Concernant le clan *Bacilli*, il existe peut-être un biais qui expliquerait le grand nombre de FT proies. En effet, il existe une grande distance phylogénétique entre les *Bacilli* et le reste des *Firmicutes* (*Clostridia*, *Negativicutes* et *Tissierellia*). Il est possible que l'utilisation de SiLiX ait séparé des familles pourtant orthologues de par cette distance phylogénétique. Pour vérifier cela, il serait nécessaire de faire l'analyse phylogénétique de toutes les FT.

Annotations fonctionnelles des FT proies Nous avons voulu savoir dans quels processus cellulaires les FT proies étaient impliquées. Pour cela, nous avons utilisé les annotations

Code	Fonction cellulaire	Fonction cellulaire générale
J	Translation, ribosomal structure and biogenesis	Information support system
A	RNA processing and modification	Information support system
K	Transcription	Information support system
L	Replication, recombination and repair	Cell cycle
B	Chromatin structure and dynamics	Information support system
D	Cell cycle control, cell division, chromosome partitioning	Cell cycle
Y	Nuclear structure	Other
V	Defense mechanisms	Other
T	Signal transduction mechanisms	Other
M	Cell wall/membrane/envelope biogenesis	Cell cycle
N	Cell motility	Other
Z	Cytoskeleton	Cell cycle
W	Extracellular structures	Other
U	Intracellular trafficking, secretion, and vesicular transport	Other
O	Posttranslational modification, protein turnover, chaperones	Information support system
X	Mobilome: prophages, transposons	Other
C	Energy production and conversion	Metabolism
G	Carbohydrate transport and metabolism	Metabolism
E	Amino acid transport and metabolism	Metabolism
F	Nucleotide transport and metabolism	Metabolism
H	Coenzyme transport and metabolism	Metabolism
I	Lipid transport and metabolism	Metabolism
P	Inorganic ion transport and metabolism	Metabolism
Q	Secondary metabolites biosynthesis, transport and catabolism	Metabolism
R	General function prediction only	Other
S	Function unknown	Unknown function

TABLE 3.7 – Correspondance entre la classification COG et les fonction cellulaires générales.

fonctionnelles des séquences des FT dans la base de données des protéomes et nous les avons comparé avec les annotations des COG disponibles sur ftp://ftp.ncbi.nih.gov/pub/COG/COG2014/static/lists/listCOGs.html.

Tout d'abord, nous avons récupéré la liste des annotations COG qui associe un code de fonction cellulaire générale avec des annotations fonctionnelles. Nous avons ensuite identifié l'annotation majoritaire de chaque FT proie. Néanmoins, les annotations COG ne contiennent pas l'ensemble des annotations fonctionnelles de la base de données des protéomes utilisés pour construire les FT. Nous avons donc utilisé un algorithme basé sur les mots clés.

L'annotation fonctionnelle de chaque FT a d'abord été découpée en mots puis les mots non spécifiques (mots de liaisons comme « and », « or » et les mots génériques comme « protein », « family ») ont été supprimés. Chaque annotation COG a subi le même traitement. L'annotation COG présentant le plus de mots identiques aux mots de l'annotation fonctionnelle de la FT a été sélectionnée. Le code COG associé à l'annotation la plus proche de l'annotation de la FT est ensuite attribuée. Les codes COG étant nombreux, nous les avons clusterisé en

fonctions cellulaires générales (table 3.7). Cette technique bien que imprécise permet d'avoir un ordre d'idée de la fonction cellulaire des FT.
L'annotation majoritaire est le cycle cellulaire (203 occurrences). 131 FT proies ont été annotées comme n'ayant pas de fonction connues. Nous avons ensuite observé la proportion de FT annotées par fonction cellulaire générales, identifiées par chaque FCC (figure 3.66). Les FCC du clan Spo corrèlent avec un grand nombre de FT également impliqués dans le cycle cellulaire. Les FCC du clan *Bacilli* semblent corréler avec des FT de fonctions plus variées.

Analyse des synténies des FT proies avec les FCC Afin de sélectionner les FT proies qui seraient les plus susceptibles d'être impliquées dans le cycle cellulaire, nous avons analysé les synténies entre chaque FT proie et chaque FCC appât. Pour cela, nous avons calculé le coefficient de Jaccard aux feuilles comme explicité dans la section 3.5.1.3 entre les FCC et les FT proies (non redondantes avec des FCC). Les FT proies possédant un coefficient de Jaccard supérieur ou égal à 0,12 avec une ou plusieurs FCC sont présentées annexe .15. La plupart des FT avec des coefficient de Jaccard élevés sont annotées comme impliqués dans la sporulation (35/65). En effet, la majorité des FCC corrélant avec les FT qui présentent une synténie conservée avec des FCC font partie du clan Spo (34/65). 14 FT proies corrèlent avec les FCC du clan *Bacilli*. 12 FT sont annotées comme n'ayant pas de fonction connue. Le reste des FT sont annotées comme étant impliquées dans de divers processus.

3.7.2.2 Focus sur quelques familles intéressantes

Nouvelles familles impliquées dans la sporulation Sur l'ensemble des familles de gènes du clan Spo, huit familles ont corrélé avec au moins une FT :DacB, DacF, SpoIID, SpoIIID, SpoIIE, SpoIIGA, SpoVE, SpoVB. La grande majorité des FT proies corrélant avec des FCC du clan Spo sont des familles de gènes impliquées dans la sporulation (27/36, annexe .16) et seulement SpoIID et SpoIIID ont corrélé avec leur équivalent FT. Ces résultats indiquent que l'approche utilisée semble efficace dans l'identification des familles de la sporulation.
Traag et collègues ont déjà effectué une étude similaire sur les familles impliquées dans la

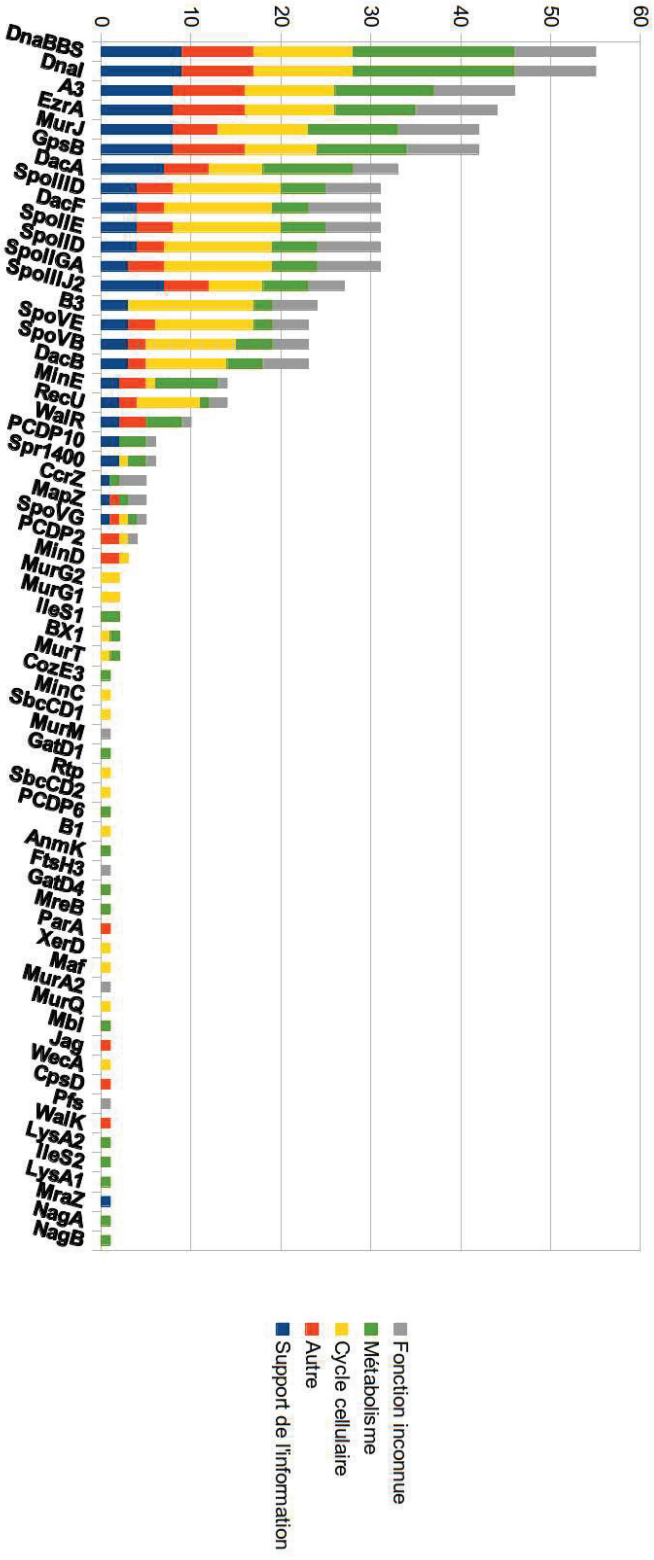


FIGURE 3.66 – Distribution des fonction cellulaires générales des FT par FCC.

sporulation[468]. Dans cette étude, les auteurs ont d'abord construit les profils phylogénétiques des gènes de *B. subtilis* pour 626 génomes de procaryotes. Ils ont ensuite identifié l'ensemble des familles qui présentaient des profils phylogénétiques à ceux des familles de la sporulation. Neuf d'entre elles n'ont jamais été décrites comme étant impliquées dans la sporulation (annexe .16). Trois d'entre elles ont été retrouvées par notre analyse.
Nous avons également identifié 6 autres familles potentiellement impliques dans la sporulation. Parmi celles-ci, trois présentent une synténie avec une FCC. Ainsi, l'endopeptidase La présente un voisinage conservé avec ClpX. Il pourrait s'agir d'une endopeptidase spécifique de la sporulation. Également, le gène codant pour une protéine membranaire est synténique de celui codant pour SpoVB. Enfin, une glucide kinase est voisine génomique de MurJ. Les trois autres familles pourraient aussi être impliquées dans la sporulation bien qu'elles ne présentent pas de voisinage avec des FCC.

Nouvelles familles du clan *Bacilli* La quasi-totalité des familles de gènes du clan *Bacilli* ont corrélé avec des FT : A3, CDP1, CDP4, DacA, DnaBBS, DnaI, EzrA, GpsB, MinE, MurJ, PCDP10, RecU, MacP, SpoIIIJ2. Ces FT sont annotées comme étant impliquées dans de nombreux processus cellulaires comme la compétence, le métabolisme ou la traduction (annexe .17). 13 d'entre elles n'ont pas de fonction assignée.
Néanmoins, étant donné que les *Bacilli* et les *Clostridia* présentent une distance phylogénétique importante, il est possible que nombre de ces FT aient été mal reconstruites et soient la résultante de la séparation d'une vraie famille en plusieurs FT. Pour vérifier cela, nous avons fait un BLASTP en prenant une séquence au hasard de chaque FT proie puis avons récupéré toutes les séquences avec une E-value inférieure à 0,001. Les séquences ont ensuite été alignées pour généré une phylogénie. Chaque arbre a été inspecté visuellement afin de déterminer si les familles étaient bien spécifiques des *Bacilli* ou du reste des *Firmicutes*. Il s'avère que 22/79 FT n'ont été que la résultante d'une mauvaise clusterisation tandis que 52/79 FT ont été correctement construites (annexe .17). 15 FT correspondent à des FCC dont 9 qui ont été correctement construites. En effet, les FT correspondant à PCDP8, MreC, DnaG, B4 et B5 corrèlent avec les FCC du clan des *Bacilli* alors qu'elles ne corrèlent pas dans l'analyse section 3.5. 27 FT semblent être spécifiques des *Bacilli*, tandis que 16 FT sont présentes uniquement

chez les autres *Firmicutes* (dont 9 en incluant également les *Paenibacillaceae*). Concernant les FT spécifiques des *Bacilli*, les processus cellulaires dans lesquels elles sont impliquées sont principalement le métabolisme, la compétence et la traduction. 10 FT d'entre elles n'ont pas de rôle connu. Celles-ci pourraient être potentiellement impliquées dans le cycle cellulaire. Par exemple, deux d'entre elles présentent un voisinage conservé avec des FCC. Ainsi la famille FAM007382 est synténique avec les familles MurA1 et Mbl et la famille FAM007232 présente un voisinage conservé avec CozE3. Ces dernières représentent des candidats particulièrement intéressants. Les familles FAM004126 et FAM004208 présentent un contexte génomique respectivement avec MurI (synthèse du peptidoglycane) chez les *Streptococcaceae* et YabP (sporulation) chez les *Bacillales* indiquant qu'il pourrait aussi s'agir de potentiels bon candidats. Les FT spécifiques des autres *Firmicutes* semblent avoir des rôles cellulaires très variés. Ainsi, il semblerait que l'émergence des *Bacilli* soit accompagnée d'un grand nombre de pertes et d'apparitions de gènes impliqués dans diverse fonctions. Étant donné que ces gènes ne sont pas spécifiquement impliqués dans le cycle cellulaire, il est difficile de prédire que les gènes de fonction inconnue seraient aussi impliqués dans le cycle cellulaire. Néanmoins, ceux-ci serait de bons candidats et nécessiteraient d'être testés expérimentalement.

Autres familles détectées

Familles corrélant avec B3 La famille B3 corrèle avec 24 FT dont la moitié sont impliquées dans la sporulation (annexe .18). Ces données suggèrent que PBPB3 est impliqué dans la sporulation bien que cette famille ne fasse pas partie du clan Spo. La famille B3 est en effet impliquée dans la sporulation, son nom usuel est d'ailleurs SpoVD. Chez *B. subtilis*, B3 est impliqué dans la synthèse du peptidoglycane durant la sporulation [77]. Le reste des FT corrélant avec la famille B3 sont aussi potentiellement impliquées dans la sporulation (dont 5 sans fonction prédite).

Familles corrélant avec WalR Dix FT corrèlent avec la famille WalR (annexe .18). Les FT correspondant à PCDP10 et PCDP8 semblent corréler avec la FCC WalR. Cette corrélation

n'est pas retrouvée dans l'analyse corrélative de la section 3.5, indiquant que ces deux FT sont probablement ma reconstruites. Le reste des familles détectées sont principalement impliquées dans le métabolisme.

Familles corrélant avec MapZ La famille MapZ semble être corrélée à cinq FT (annexe .18). Ainsi, une transposase, une protéine membranaire, une protéine hypothétique (parfois annotée comme antigène-A du pneumocoque), une perméase et une superoxide dismutase. Néanmoins, la famille MapZ émergeant à l'ancêtre des *Streptococcaceae* et *Enterococcaceae*, il est possible tout comme pour le clan *Bacilli* que la distance phylogénétique conséquente entre ce groupe et le reste des *Firmicutes* ait conduit à la formation de familles artefactuelles. Pour vérifier cela, nous avons appliqué la même procédure que la section 3.7.2.2. Ainsi, aucune des familles ne semble spécifique des *Streptococcaceae* et *Enterococcaceae* suggérant que les FT détectées sont donc mal construites et non corrélées à la famille MapZ. Le fait qu'aucune FT ne corresponde à l'émergence des *Streptococcaceae* et *Enterococcaceae* peut paraître surprenant. Cela peut être du à la construction des FT par SiLiX mais également par un seuil trop stringent de détection des corrélations.

3.7.2.3 Conclusion

L'approche de génomique comparative par la méthode de clusterisation a permis d'identifier quelques familles potentiellement impliquées dans le cycle cellulaire. La plupart d'entre elles corrèlent avec les familles des clan Spo et *Bacilli*. Nous avons potentiellement identifié 6 nouvelles familles impliquées dans la sporulation dont 3 semblent être des bons candidats au vue de leur contexte génomique. 43 familles ont corrélé avec le clan *Bacilli* et 12 d'entre elles n'ont pas de fonction connue. Parmi ces dernières, 2 semblent particulièrement être de bons candidats de par leur contexte génomique. Néanmoins, la méthode par clusterisation semble générer un certain nombre de familles dont la qualité n'est pas optimale. En effet, de nombreuses familles semblent tronquées, probablement par un seuil d'agrégation trop élevé. De même, certaines familles peuvent être agrégées générant ainsi des familles trop larges. Il aurait

éventuellement fallu utilisé plusieurs seuils. La méthode par clusterisation, bien que simple et rapide ne peut remplacer l'analyse phylogénétique dans la reconstruction des familles de gènes. Ces observations soulignent l'impact important de la construction des familles dans ce type d'analyses.

3.8 Conclusion

Nous avons donc à partir des séquences de protéines impliquées dans le cycle cellulaire reconstruit les familles de gènes associées. Nous avons ensuite reconstruit l'histoire évolutive de chaque famille tant au niveau individuel par la réconciliation et les profils phylogénétiques que au niveau organisationnel par l'analyse du contexte génomique. L'ensemble de ces données ont été utilisées afin d'établir des corrélations évolutives entre les différentes familles. Ces corrélations ont permis d'inférer des potentiels liens fonctionnels dont certains ont déjà été démontré et d'autres qui restent à démontrer. Nous avons aussi identifié quelques potentielles nouvelles familles de gènes impliquées dans le cycle cellulaire en comparant les familles de gènes du cycle cellulaire avec des familles de gènes inférées par méthode de clusterisation. L'ensemble des connaissance générées à partir de cette analyse pourra servir à de futures études phylogéniques mais également expérimentales. Il serait notamment nécessaire de prouver ces liens fonctionnels inférés par des méthodes expérimentales.

Chapitre 4

Étude de l'histoire évolutive des domaines PASTA chez les *Lactobacillales*

Sommaire

4.1 Le rôle des domaines PASTA 288

4.2 Matériel et méthodes 290

4.2.1 Assemblage des jeux de données 290

4.2.2 Groupement des domaines PASTA 291

4.2.3 Inférence des phylogénies 292

4.2.4 Construction des arbres d'espèces 292

4.3 Résultats/Discussion 294

4.3.1 Analyse phylogénétique des domaines PASTA chez les *Lactobacillales* 294

4.3.2 Les différents rôles des domaines PASTA de StkP chez *S. pneumoniae* 306

4.4 Conclusion 311

Nous avons précédemment décrits l'histoire évolutive des gènes du cycle cellulaire. Pour cela, nous nous sommes focalisés principalement sur les événements affectant les gènes entiers et leurs contextes génomiques. Dans cette partie nous allons décrire l'histoire évolutive des domaines de la protéine StkP au sein des *Lactobacillales* et nous allons également discuter des implications fonctionnelles. Ce chapitre renvoie directement à la publication :

Zucchini L, Mercy C, Garcia PS, et al. PASTA repeats of the protein kinase StkP interconnect cell constriction and separation of *Streptococcus* pneumoniae. Nat Microbiol. 2018 ;3(2) :197-209.

Dans cette publication, les membre du laboratoire MMSB ont caractérisé fonctionnellement le rôle des différents domaines de la protéine StkP chez *S. pneumoniae*. J'ai quant à moi effectué l'analyse phylogénétique des domaines de StkP chez les *Streptococcaceae*. Nous décrirons plus largement l'histoire des domaines StkP au sein des *Lactobacillales* dans ce chapitre. Nous décririons aussi brièvement le contenu de la publication et les principaux résultats associés.

4.1 Le rôle des domaines PASTA

Comme expliqué dans le chapitre 1, la protéine StkP appartient à la famille des Sérine/Thréonine kinases de type Hanks [359], [444]. Il s'agit d'une famille retrouvée chez la plupart des organismes vivants capables de phosphoryler des protéines sur des résidus Sérine ou Thréonine [444]. Cette famille est impliquée dans de nombreux processus cellulaires et notamment dans la division cellulaire et la morphogenèse, la synthèse du peptidoglycane, la virulence, ou encore le métabolisme chez les bactéries [300], [223], [377], [444], [321], [372]. Chez *S. pneumoniae*, StkP est principalement impliquée dans le contrôle de la division cellulaire [150], [148], [28].
Les protéines de cette famille possèdent une organisation modulaire .Elles sont en effet constituées de plusieurs domaines fonctionnels plus ou moins conservés. Chez les bactéries, elles sont

principalement composées d'un domaine kinase intracellulaire, d'un domaine transmembranaire et de plusieurs domaines PASTA extracellulaires (« PBP And Serine-Threonine kinase Associated ») [520]. Le domaine kinase est conservé au sein de la famille StkP mais les domaines PASTA sont présents en nombre variable chez les différentes bactéries [407]. Cette diversité est probablement due à une histoire évolutive complexe des domaine PASTA.

Les domaines PASTA sont peu conservés en séquence mais partagent une structure globulaire similaire [520]. Ils sont retrouvés principalement au sein de kinases de type eucaryote (type Hanks) et de PBPs. Ils ont été montrés comme interagissant avec certains fragments de peptidoglycane (muropeptide) et avec les β-lactames [325], [298], [438]. De façon intéressante, l'interaction entre le muropeptide et les domaines PASTA semble être espèce-spécifique chez les bactéries [325], [423]. En effet, les domaines PASTA d'une espèce n'interagissent qu'avec les muropeptide de la même espèce ou avec celui d'espèces dont la composition du peptidoglycane est similaire. Le rôle des domaines PASTA n'est pas encore clairement élucidé mais il a été proposé que les domaines PASTA joueraient le rôle de senseurs qui moduleraient l'activité kinase de StkP, tout comme les kinases des systèmes à deux composants [372]. Cependant, cette hypothèse n'explique pas les variations observées au niveau de la répétition des domaines PASTA ni même pourquoi les domaines PASTA sont répétés. De plus, il a été montré que certains domaines PASTA ne fixent pas les muropeptides [56] [332]. Il est donc tentant de faire l'hypothèse que les domaines PASTA pourraient avoir d'autres fonctions que celle de simples senseurs et modulateurs de l'activité kinase de StkP.

Dans cette étude, nous avons caractérisé le rôle des domaines PASTA de StkP chez *S. pneumoniae*. Chez cette bactérie, StkP contient quatre domaines PASTA. Le rôle de chaque domaine PASTA a été étudié. Les membres du laboratoire MMSB ont montré que les domaines PASTA 1, 2 et 3 possèdent des rôles similaires et que leur présence conditionne l'activité kinase de StkP. Le domaine PASTA 4, qui est le plus distal, semble posséder un rôle supplémentaire. En effet, il interagit avec l'hydrolase du peptidoglycane LytB afin de réguler l'épaisseur du peptidoglycane. L'histoire évolutive de ces domaines chez les *Lactobacillales* et plus spécifiquement chez les *Streptococcaceae* a été reconstruite. Nous avons mis en évidence que les domaines PASTA peuvent être séparés en plusieurs sous-familles et qu'ils possèdent une histoire évolutive complexe, particulièrement chez les *Enterococcaceae*. Nous avons mis en évidence que la

sous-famille C, correspondant au domaine PASTA 4 de *S. pneumoniae*, semble évoluer plus rapidement que les autres sous-familles de domaines PASTA et que l'interaction avec LytB semble restreinte à quelques *Streptococcus*.

4.2 Matériel et méthodes

4.2.1 Assemblage des jeux de données

Deux bases de données de protéomes complets ont été construites à partir des données disponibles sur le FTP du NCBI (ftp://ftp.ncbi.nlm.nih.gov/). La base de *Lactobacillales* regroupe 223 protéomes (Février 2015). Une base de données spécifique des *Streptococcaceae* contenant 200 protéomes a aussi été construite (Janvier 2016).
Le profil HMM des domaines PASTA (PF03793) issu de la base de données Pfam (http://pfam.xfam.org/) a été utilisé pour requêter les deux bases de données. Une recherche de toutes les séquences protéiques comportant au moins un domaine PASTA a été effectuée pour chacune des deux bases de données à l'aide du programme HMMsearch de la suite HMMER v3.1b1 [119]. Les séquences présentant une E-value inférieure à 0,01 ont ensuite été collectées dans un fichier FASTA. Les séquences potentiellement mal-annotées ont été récupérées par tBLASTn en utilisant la séquence de StkP de *S. pneumoniae* comme graine (NP_359169.1). La composition en domaines de ces séquences a alors été analysées grâce au programme HMMscan de la suite HMMER. Les séquences présentant un domaine Pkinase (PF00069) en plus d'un ou plusieurs domaines PASTA ont été considérées comme étant des StkP, tandis que celles combinant un ou plusieurs domaines PASTA associés à un domaine Transpeptidase (PF00905) ont été classées comme homologues de PBPs. Les séquences ne présentant ni domaine Pkinase, ni domaine Transpeptidase ont été regroupés dans la classe Autres. La position de chaque domaine a été déterminée à partir des résultats de l'HMMSCAN. Les domaines présentant une i-value inférieure à 0,5 ont été considérés. Les domaines ont été numérotés en fonction de leur ordre dans la séquence, de N-terminal à C-terminal.
Une stratégie similaire a été utilisée pour identifier les séquences apparentées à LytB. Nous

avons pour cela utilisé les domaines Glucosaminidase (PF01832) et CW_binding_1 (PF01473). La phylogénie des séquences des domaines PASTA a été inférée avec Fasttree [389] pour chacune des familles de *Lactobacillales* (*Aerococcaceae*, *Leuconostocaceae*, *Streptococcaceae*, *Lactobacillaceae*, *Carnobacteraceae*). Ces arbres nous ont servi à identifier les sous-familles de domaines PASTA. Pour chaque sous-famille, les séquences des domaines PASTA ont été alignées et utilisées pour construire de nouveaux profils HMM, spécifique de chaque sous-famille. Ces profils ont été utilisés pour rechercher de potentielles nouvelles séquences présentant une domaine PASTA non détectés par le profil générique PF03793 par l'utilisation de HMMsearch. Aucune nouvelle séquence présentant un domaine PASTA n'a été ajoutée. Néanmoins, nous avons détecté des domaines additionnels au sein des séquences présentant au moins un domaine PASTA par l'utilisation de ces profils à l'aide du programme HMMscan. Ce fut particulièrement le cas des domaines PASTA en position 1 des PBPs.

4.2.2 Groupement des domaines PASTA

Pour les analyses, un échantillonnage des séquences des domaines PASTA a été réalisé en conservant une souche par espèce afin de limiter la redondance taxonomique (1233 séquences à 370 séquences) (figure 4.1, étape 1). Les séquences de domaines PASTA spécifiques aux kinases StkP ont ensuite été sélectionnés (217 séquences) (figure 4.1, étape 2). Les séquences de domaines PASTA étant très divergentes chez les *Lactobacillales*, les séquences ont été séparés en 6 ensembles correspondant chacun à une famille de *Lactobacillales* (figure 4.1, étape 3). Les domaines PASTA ont ensuite été classés en groupes selon leur positionnement dans la séquence de StkP et la phylogénie (figure 4.1, étape 4). Les séquences de domaines PASTA de chaque famille ont été analysées par paire de familles (figure 4.1, étape 5). Pour chaque paire de familles, les distances patristiques moyennes entre chaque groupe de position ont été mesurées à l'aide de ete3 [208]. Plus précisément, pour chaque paire de groupes de position, nous avons calculé la distance patristique de chaque séquence d'un groupe avec chaque séquence de l'autre groupe. Ces distances ont permis de construire une matrice de distances patristiques (figure 4.1, étape 5). Un arbre de distances a été inféré à partir de cette matrice à l'aide de Fastme

2.1.5 (méthode NJ) [266] (figure 4.1, étape 6).

4.2.3 Inférence des phylogénies

Les alignements multiples ont été construits en utilisant MAFFT v7.123b [238] avec l'option L-ins-i puis nettoyés à l'aide de BMGE 1.1 [72] avec l'option BLOSUM30. Les alignements de l'ensemble des domaines PASTA des *Lactobacillales* n'ont pas été nettoyés étant donné que très peu de positions étaient conservées (figure 4.1, étape 1, 2).
Les modèles évolutifs ont été sélectionnés sur la base du critère BIC (« Bayesian Information Criterion ») en utilisant l'outil ModelFinder [234] inclus dans IQ-TREE [348]. Seuls les modèles disponibles dans PhyML ont été testés. Le modèle WAG+G4 a été sélectionné pour une très grande majorité des jeux de données.
Les phylogénies au maximum de vraisemblance ont été inférées à l'aide de PhyML 3.1 [183]. La robustesse des branches a été évaluée par le test SH-like implémenté dans PhyML.
La phylogénie des domaines PASTA spécifiques de StkP chez les *Streptococcaceae* (figure 4.12) a été inférée par une approche bayésienne en utilisant MrBayes [406] avec des modèles mixtes et une distribution gamma avec quatre catégories de sites. Pour chaque analyse, deux runs de quatre chaînes ont été lancés en parallèle. La convergence des chaînes a été vérifiée en utilisant Tracer [395] et les premiers 25% des topologies ont été écartés (« burn-in »).

4.2.4 Construction des arbres d'espèces

La phylogénie des *Lactobacillales* est celle de l'arbre d'espèces des *Firmicutes* contenant le sous-arbre des *lactobacillales* inféré au chapitre 3. Néanmoins, l'échantillonnage taxonomique entre cet arbre d'espèces et la base de données des domaines PASTA n'est pas rigoureusement identique. En effet, 15 espèces ne sont pas représentées dans le jeu de données des domaines PASTA (1 *Carnobacteraceae*, 1 *Enterococcaceae*, 9 *Lactobacillaceae*, 1 *Streptococcaceae* et 3 *Leuconostocaceae*. Nous avons malgré cela utilisé cet arbre pour inféré l'histoire des domaines PASTA chez les *lactobacillales*, principalement par manque de temps.

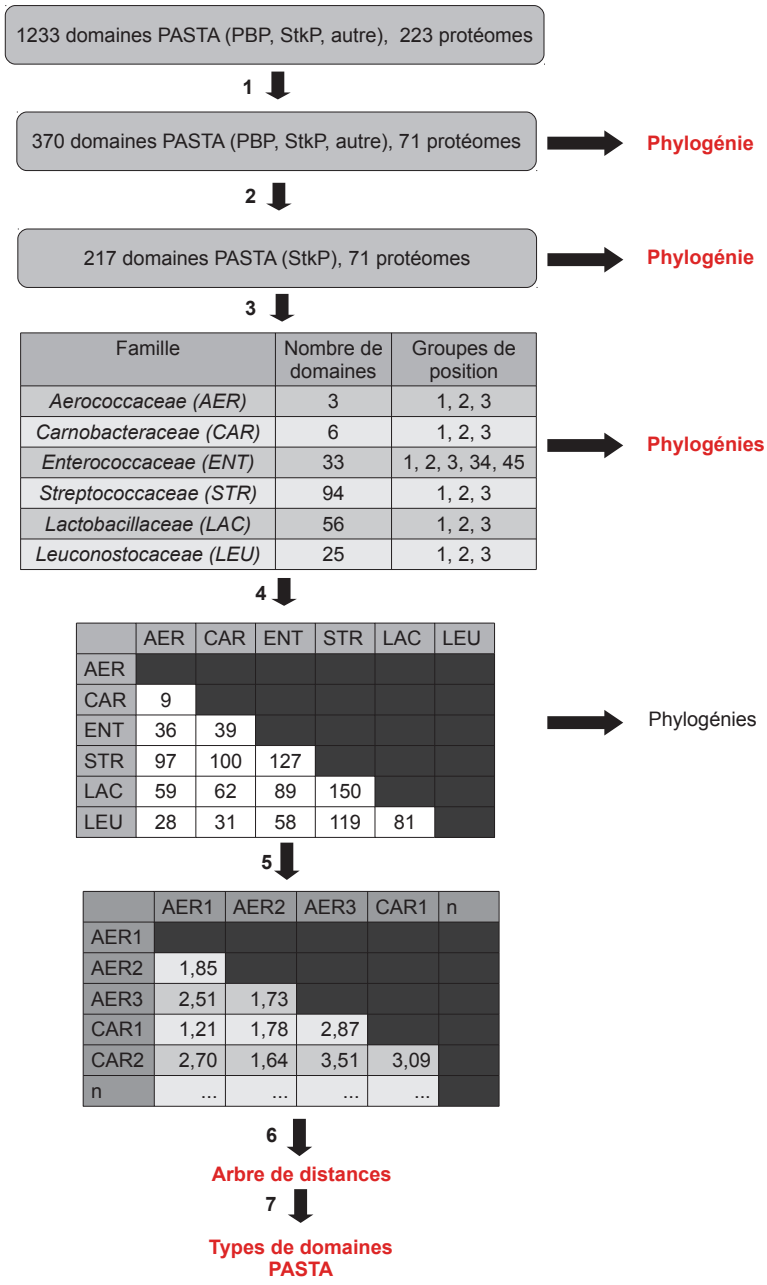


FIGURE 4.1 – Approche de groupement des domaines PASTA. 1 : échantillonnage taxonomique, 2 : sélection des séquences de StkP, 3 : découpage en familles et classification en groupes de position, 4 : groupement des jeux de données entre familles par paire (les nombres correspondent au nombre de séquences), 5 : distances patristiques moyennes entre les groupes de position, matrice de distances patristiques, 6 : arbre de distances patristiques, 7 : séparation des domaines PASTA en types. Les éléments indiqués en rouge sont présentés en figures dans le manuscrit.

L'arbre de référence des *Streptococcaceae* a été inféré en utilisant les protéines ribosomiques extraites à partir de RiboDB [219]. Les souches *Lactobacillus casei LOCK919*, *Leuconostoc mesenteroides* subsp. *mesenteroides* ATCC 8293, *Carnobacterium maltaromaticum* LMA28 et *Enterococcus faecalis* D32 ont été utilisées comme groupe externe. Les séquences ont été alignées à l'aide de MAFFT v7.123b (option L-ins-i) [238]. Les alignements ont été nettoyés en utilisant BMGE 1.1 (BLOSUM30) [72] puis concaténés en une supermatrice. Le modèle LG+I+G4 a été sélectionnés sur la base du critère BIC (« Bayesian Information Criterion ») en utilisant ModelFinder [234] inclus dans IQ-TREE [348] puis utilisé pour inférer la phylogénie à l'aide de PhyML 3.1 [183]. La robustesse des branches a été évaluée par 100 réplicats de bootstrap.
La reconstruction de l'histoire évolutive de l'organisation des domaines PASTA au sein des séquences de StkP des *Lactobacillales* et les *Streptococcaceae* (c'est-à-dire l'inférence des états ancestraux, des événements de duplication, perte et transfert de domaines) a été effectuée de façon qualitative par une approche de parcimonie (minimisation du nombre d'événements).

4.3 Résultats/Discussion

4.3.1 Analyse phylogénétique des domaines PASTA chez les *Lactobacillales*

4.3.1.1 Classification des domaines PASTA

Nous avons tout d'abord voulu savoir si les domaines PASTA pouvaient être catégorisés en groupes selon leurs relations de parenté (phylogénie) et leur position relative dans les séquences de StkP/PBP2X.
A partir de la base de données des *Lactobacillales*, nous avons identifié toutes les séquences qui contiennent au moins un domaine PASTA. Il est apparu que toutes les souches possédaient une copie de StkP et une copie de PBP2X. La phylogénie de l'ensemble des domaines PASTA

chez les *Lactobacillales* sans nettoyage de l'alignement multiple est présentée figure 4.2. Les séquences de domaines PASTA sont colorées en fonction de la protéine dans laquelle ils sont retrouvés. Ainsi, les domaines PASTA des PBPs sont annotés en vert, ceux des Sérine/thréonine kinases en bleu et les autres séquences en gris. Les domaines PASTA issus des séquences de PBPs semblent former deux groupes monophylétiques et sont groupés en fonction de leur position. Nous avons donc défini deux groupes, α et β. Les domaines PASTA des PBPs en position 1 correspondent au groupe α et ceux en position 2 au groupe β. Ces résultats suggèrent que les domaines PASTA des PBPs n'ont pas changé de position au cours de la diversification des *Lactobacillales*. Les séquences des domaines PASTA retrouvées chez 13 protéines autres que les kinases et les PBPs se groupent avec les séquences issues des kinases.

Les séquences des domaines PASTA provenant uniquement des protéines StkP ont été alignées pour inférer la phylogénie des domaines PASTA des kinases (figure 4.3). Le positionnement de chaque domaine PASTA au sein de la séquence originale a été projeté sur la phylogénie. Les domaines PASTA en position 1 semblent se regrouper mais les autres domaines forment un groupe hétérogène et les supports sont globalement très faibles. Ce mélange semble provenir du fait que les séquences des domaines PASTA sont très courtes (64 acides aminés en moyenne) et que les séquences sont très divergentes. Nous avons donc analysé les phylogénies des domaines PASTA de StkP au niveau des familles de *Lactobacillales* (*Aerococcaceae*, *Carnobacteraceae*, *Leuconostocaceae*, *Lactobacillaceae*, *Streptococcaceae*). Sur les phylogénies, des clusters de séquences correspondant aux positions des domaines de StkP sont identifiables (figure 4.4 à 4.8). Chaque groupe semblait corréler avec la position au sein des séquences suggérant une transmission verticale à partir d'un ancêtre pour tous les domaines PASTA à une position données pour chaque famille. Chaque groupe a donc été nommé par la position majoritaire des séquences incluses dans ce groupe. La majorité des groupes sont monophylétiques, mais pour les *Leuconostocaceae* et les *Streptococcaceae*, le groupe des domaines PASTA en position 3 est paraphylétique (figures 4.5,4.7). De plus, pour les *Enterococcaceae*, les groupes correspondant à la position 3 et 2 sont paraphylétiques et deux groupes présentent un mélange de deux positions (3-4 et 4-5) (figure 4.8).

Nous avons ensuite voulu connaître les liens de parenté entre les groupes correspondant

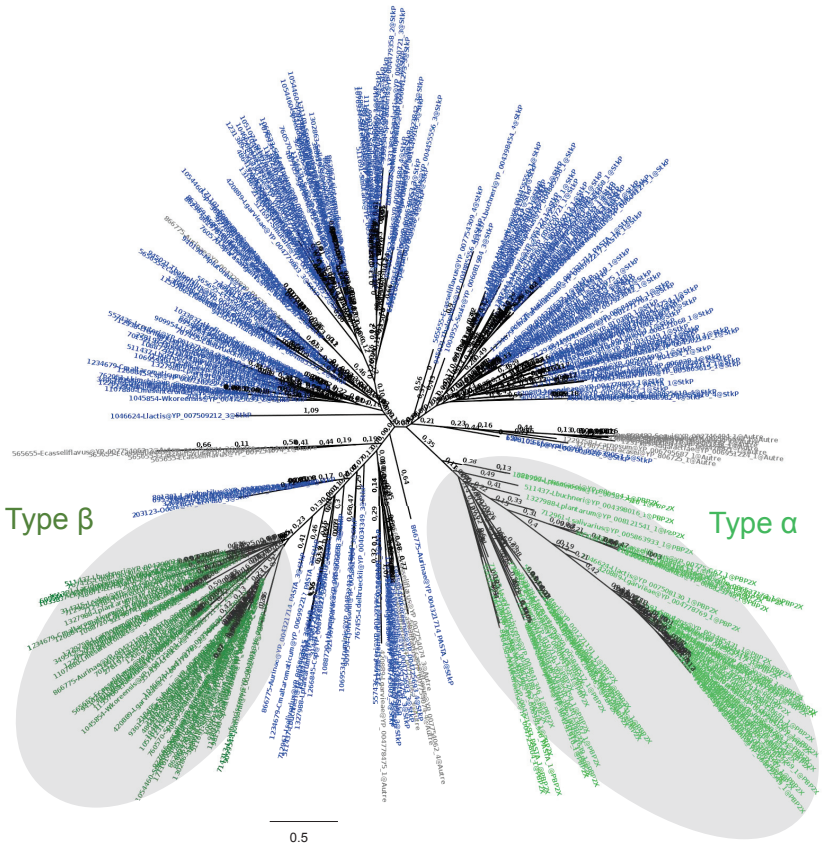


FIGURE 4.2 – Phylogénie des domaines PASTA chez les *Lactobacillales*. PhyML, WAG+G4, 307 séquences, 101 positions d'acides aminés. Bleu : domaines de StkP, vert : domaines de PBP, gris : domaines d'autres protéines. La barre d'échelle indique le taux moyen de substitution par site. Les nombre aux branches correspondent à des valeurs de SH-like.

aux positions des domaines PASTA dans les séquences chez les différentes familles de *Lactobacillales*. L'approche classique consisterait à reconstruire la phylogénie de l'ensemble des domaines PASTA de StkP chez les *Lactobacillales*. Cependant, comme expliqué précédemment, les séquences des domaines PASTA de StkP chez les *Lactobacillales* sont très divergentes conduisant ainsi à une phylogénie insuffisament résolue. Ainsi, pour contourner ce problème, nous avons conduit cette analyse en considérant les familles de *lactobacillales* par paire et non pas l'ensemble des *Lactobacillales* à la fois (figure 4.1, étape 4). Par exemple, une phylogénie des séquences des domaines PASTA des *Carnobacteraceae* et des *Streptococcaceae* a été reconstruite. La distance patristique moyenne entre chaque groupe de position précédemment établi

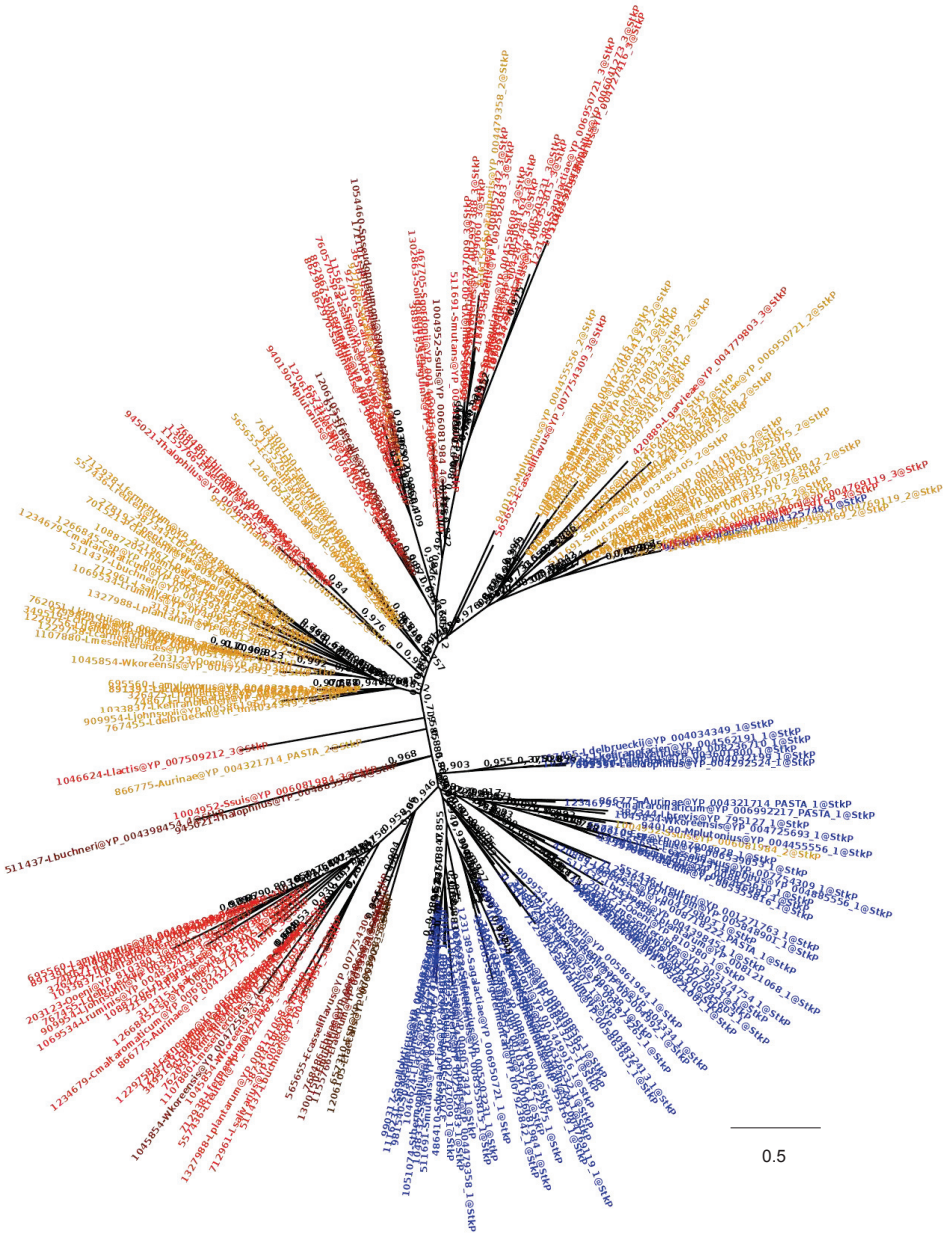


FIGURE 4.3 – Phylogénie des domaines PASTA de StkP chez les *Lactobacillales*. PhyML, WAG+G4, 217 séquences, 85 positions d'acides aminés. La couleur des feuilles représente la position relative des domaines PASTA dans la séquence. Bleu : position 1, jaune : position 2, rouge : position 3, rouge foncé : position 4, marron : position 5. La barre d'échelle indique le taux moyen de substitution par site. Les nombre aux branches correspondent à des valeurs de SH-like.

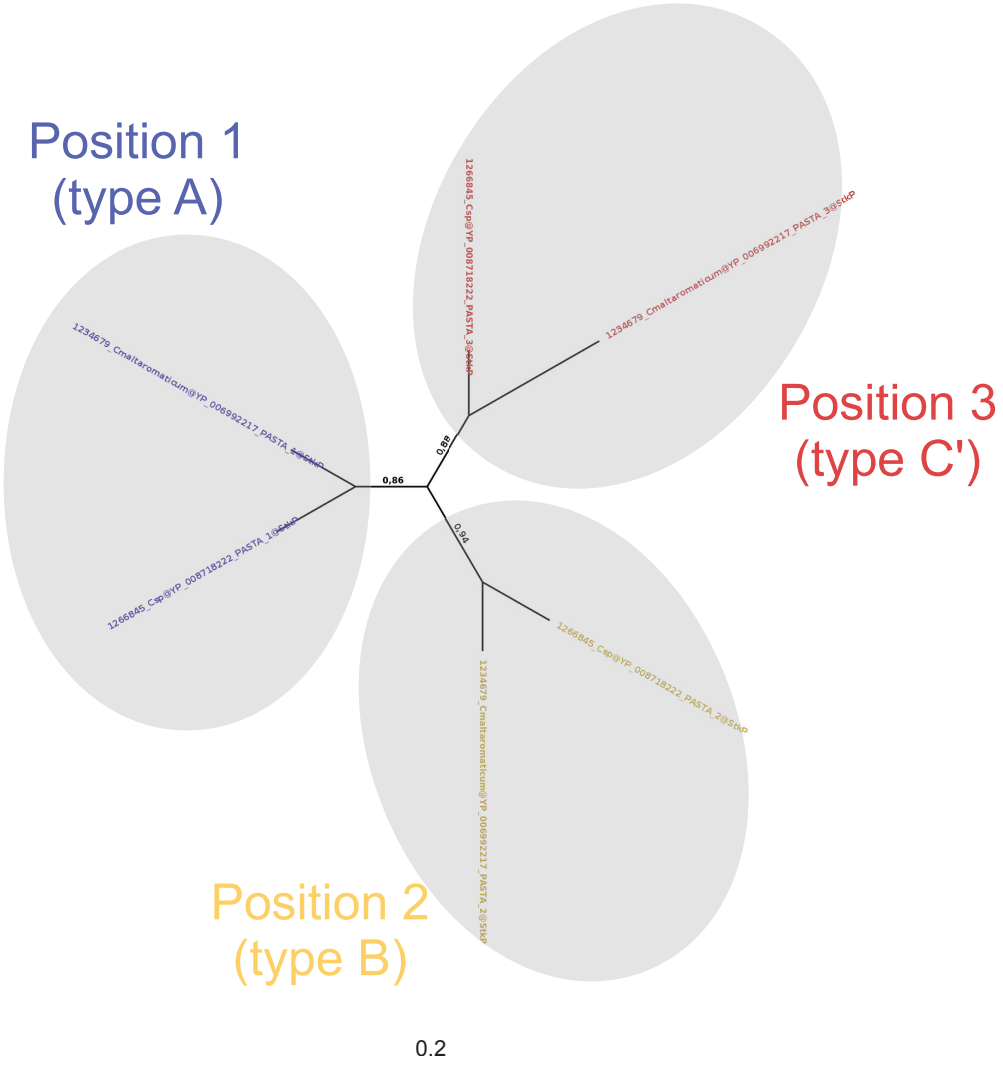


FIGURE 4.4 – Phylogénie des domaines PASTA de StkP chez les *Carnobacteraceae*. PhyML, WAG+G4, 6 séquences, 58 positions d'acides aminés. La couleur des feuilles représente la position relative des domaines PASTA dans la séquence. Bleu : position 1, jaune : position 2, rouge : position 3, rouge foncé : position 4, marron : position 5. Les groupes de position sont représentés par des ellipses grises. La barre d'échelle indique le taux moyen de substitution par site. Les nombre aux branches correspondent à des valeurs de SH-like.

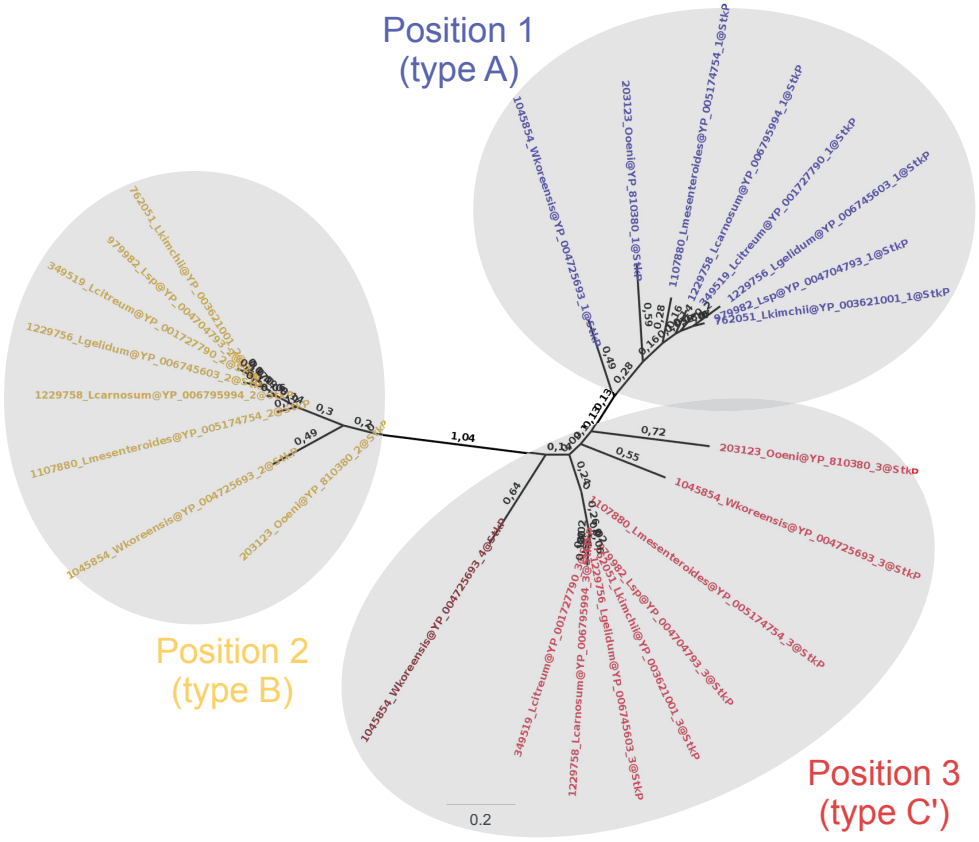


FIGURE 4.5 – Phylogénie des domaines PASTA de StkP chez les *Leuconostocaceae*. PhyML, WAG+G4, 25 séquences, 61 positions d'acides aminés. La couleur des feuilles représente la position relative des domaines PASTA dans la séquence. Bleu : position 1, jaune : position 2, rouge : position 3, rouge foncé : position 4, marron : position 5. Les groupes de position sont représentés par des ellipses grises. La barre d'échelle indique le taux moyen de substitution par site. Les nombre aux branches correspondent à des valeurs de SH-like.

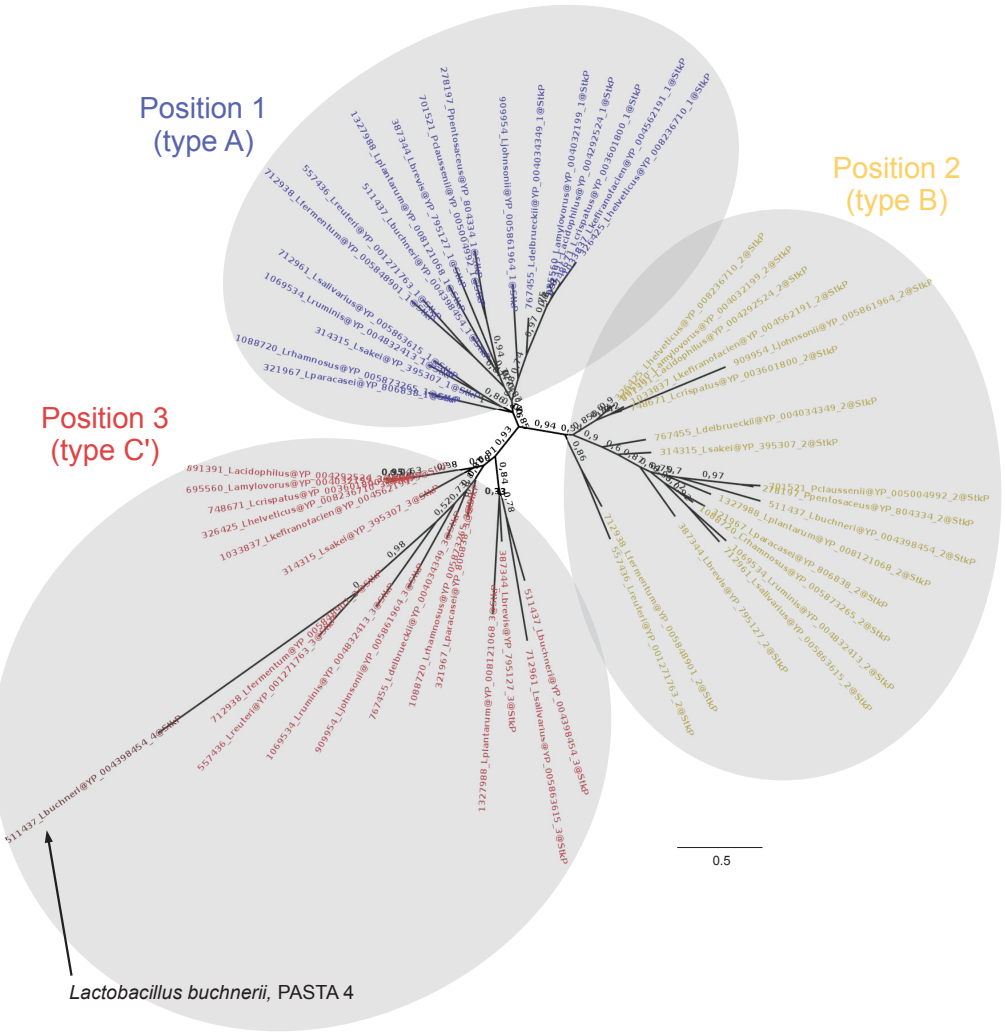


FIGURE 4.6 – Phylogénie des domaines PASTA de StkP chez les *Lactobacillaceae*. PhyML, WAG+G4, 56 séquences, 45 positions d'acides aminés. La couleur des feuilles représente la position relative des domaines PASTA dans la séquence. Bleu : position 1, jaune : position 2, rouge : position 3, rouge foncé : position 4, marron : position 5. Les groupes de position sont représentés par des ellipses grises. La barre d'échelle indique le taux moyen de substitution par site. Les nombre aux branches correspondent à des valeurs de SH-like.

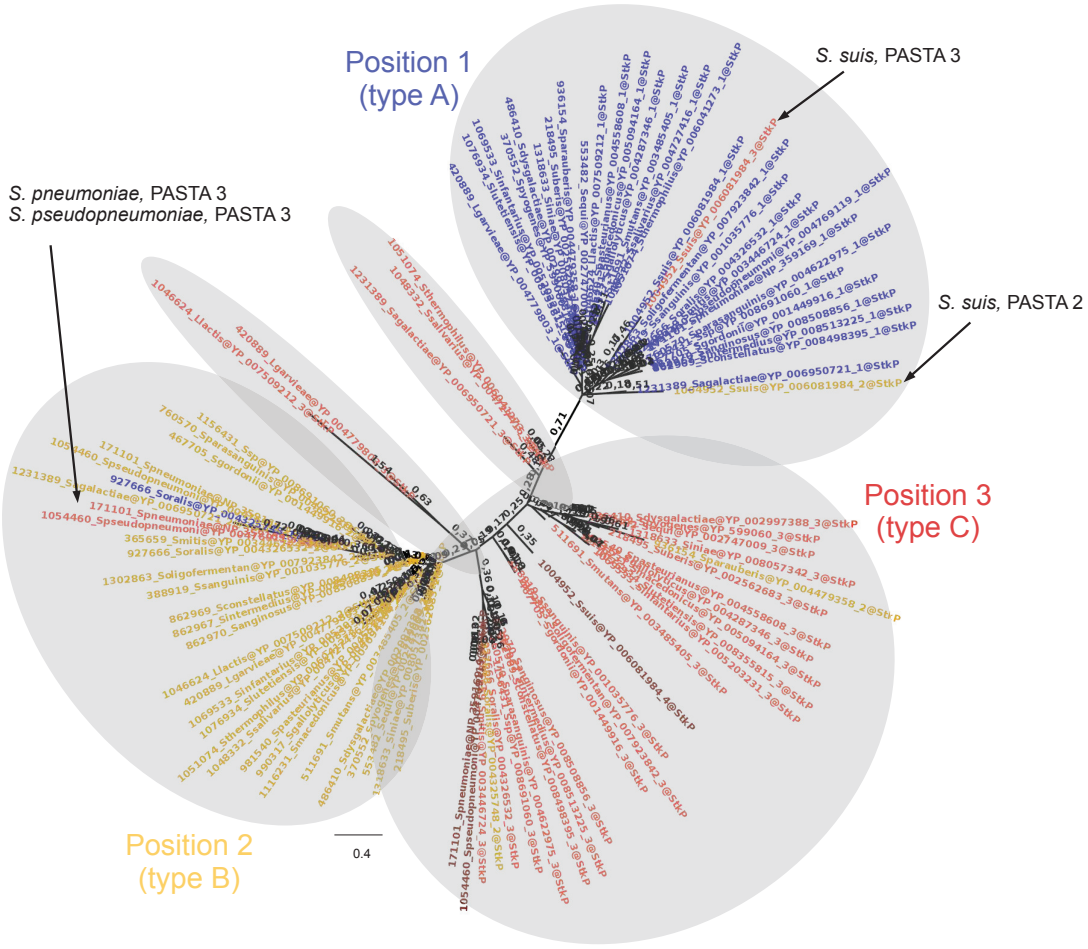


FIGURE 4.7 – Phylogénie des domaines PASTA de StkP chez les *Streptococcaceae*. PhyML, WAG+G4, 94 séquences, 50 positions d'acides aminés. La couleur des feuilles représente la position relative des domaines PASTA dans la séquence. Bleu : position 1, jaune : position 2, rouge : position 3, rouge foncé : position 4, marron : position 5. Les groupes de position sont représentés par des ellipses grises. La barre d'échelle indique le taux moyen de substitution par site. Les nombre aux branches correspondent à des valeurs de SH-like.

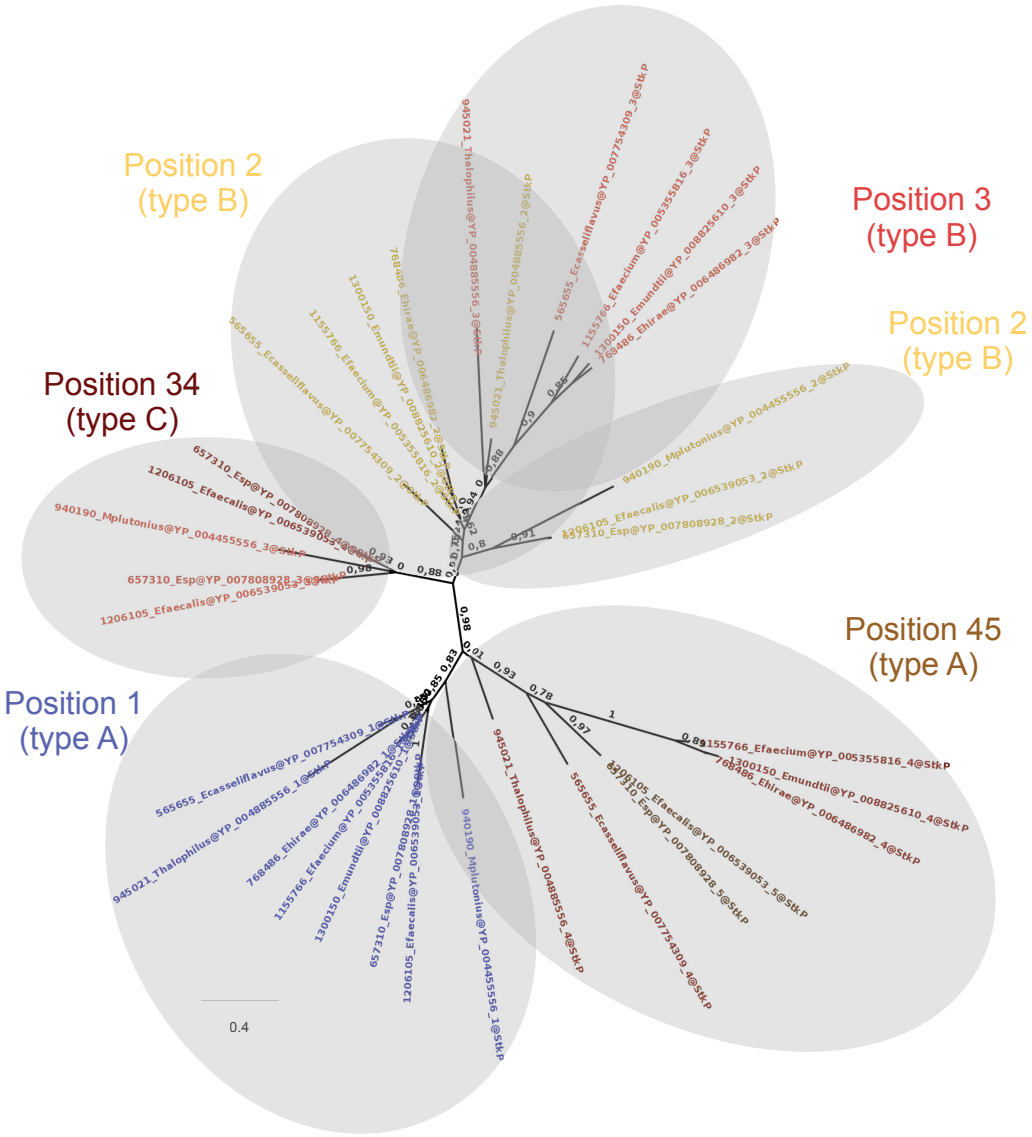


FIGURE 4.8 – Phylogénie des domaines PASTA de StkP chez les *Enterococcaceae*. PhyML, WAG+G4, 33 séquences, 57 positions d'acides aminés. La couleur des feuilles représente la position relative des domaines PASTA dans la séquence. Bleu : position 1, jaune : position 2, rouge : position 3, rouge foncé : position 4, marron : position 5. Les groupes de position sont représentés par des ellipses grises. La barre d'échelle indique le taux moyen de substitution par site. Les nombre aux branches correspondent à des valeurs de SH-like.

a été mesurée (figure 4.1, étape 5). Nous avons ainsi pu calculer une distance moyenne entre chaque groupe de chaque familles des *Lactobacillales* sous forme de matrice de distance. Cette matrice a permis de construire un arbre de distance entre tous les groupes (figure 4.9).
Les groupes correspondant à la même position des domaines au sein des séquences de StkP semblent se regrouper. Les clusters ainsi formés nous ont permis de définir des types de domaines PASTA, les groupes de position 1 correspondant au type A, les groupes de position 2 au type B et les groupes de position 3 au type C. Néanmoins, quelques exceptions sont observées. En effet, le groupe de la position 3 chez les *Enterococcaceae* se place avec les groupes de position 2, le groupe de position 2 des *Aerococcaceae* est proche des groupes de position 3 et le groupe de position 4-5 des *Enterococcaceae* se place avec les groupes de position 1. Il apparaît aussi que les groupes de position 3 forment deux clusters, les *Enterococcaceae* et *Streptococcaceae* d"une part et les autres *Lactobacillales* d'autre part. Nous avons donc délimité deux types : C et C'. Il est intéressant de noter l'étroite relation entre le positionnement des domaines PASTA dans les séquences de PBPs et StkP et leurs relation phylogénétique. Cette observation a déjà été faite chez les *Actinobacteria* [355].

A partir de cette classification, nous avons pu représenter la diversité de l'organisation modulaire des domaines PASTA de StkP chez les *Lactobacillales* (figure 4.10). Chez les *Carnobacteraceae*, *Leuconostocaceae* et *Lactobacillaceae*, l'organisation retrouvée chez quasiment toutes les espèces correspond à un motif ABC'. Une exception est retrouvée chez *Lactobacillus buchneri* ou un domaine de type C' additionnel est situé en position terminale et chez *Aerococcus urinae* ou le domaine de type B est remplacé par un type C'. Chez les *Enterococcaceae* et *Streptococcaceae*, l'organisation des domaines PASTA de StkP est plus variable (figure 4.10). Chez les *Streptococcaceae*, l'organisation majoritaire correspond à un motif ABC. Néanmoins, chez *S. pneumoniae* et S. *pseudopneumoniae*, un domaine PASTA additionnel de type B est retrouvé en position médiane (motif ABBC). Chez *S. suis*, le domaine de type B a été remplacé par deux domaines de type A (motif AAAC). Enfin, chez *S. parauberis*, le domaine de type B est absent (motif AC). Concernant les *Enterococcaceae*, l'organisation majoritaire correspond à un motif ABBA. Chez *Enterococcus sp.* et *Enterococcus faecalis*, les domaines PASTA sont organisés en motif ABCCA et chez *Melissococcus plutonius*, les domaines PASTA suivent l'ordre

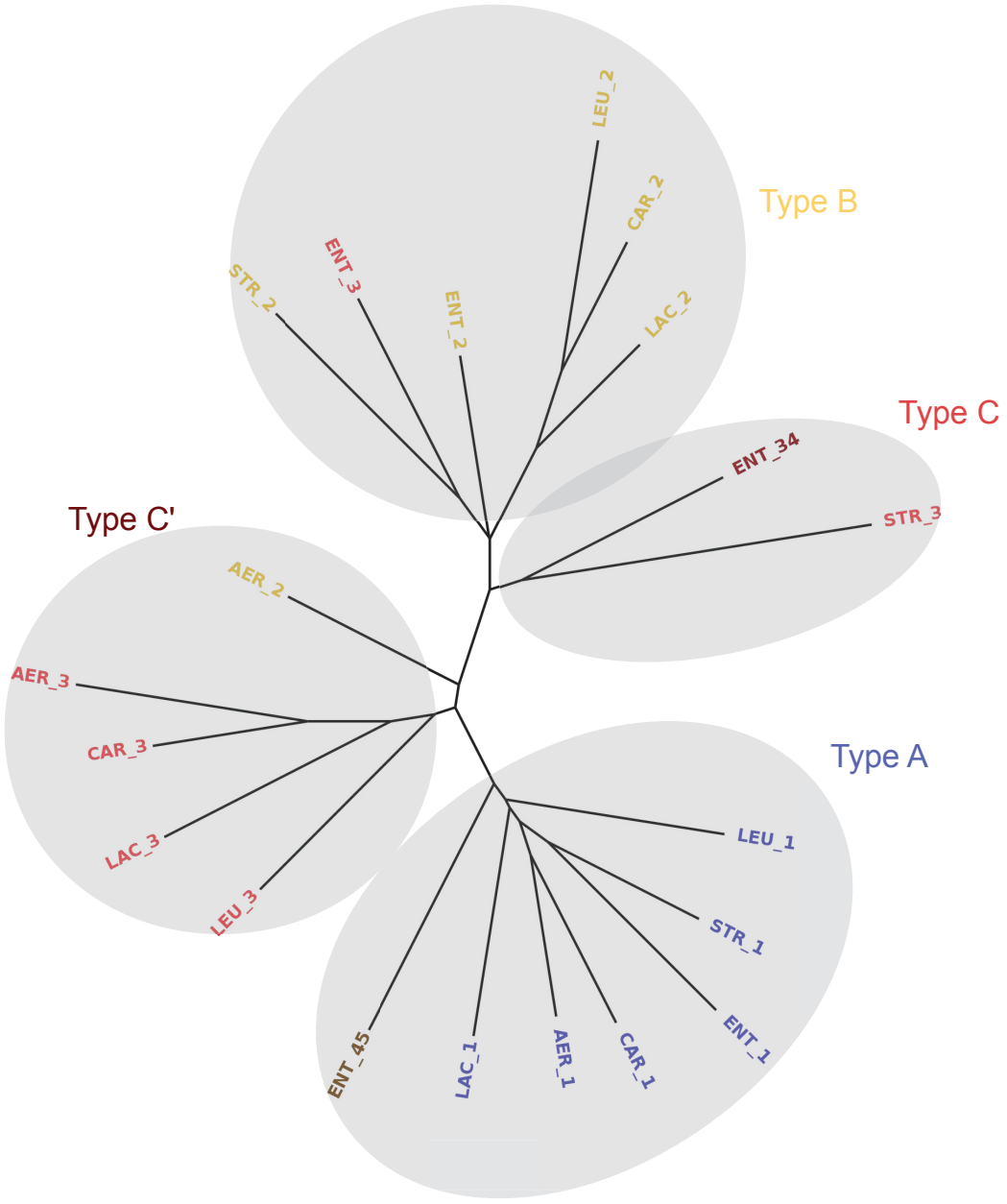


FIGURE 4.9 – Arbre de distances construit sur la base des distances patristiques moyennes entre groupes de position. Les feuilles correspondent aux groupes de position définis précédemment. La longueur des branches correspond aux distances patristiques moyennes entre les séquences des différents groupes. La couleur des feuilles représente la position relative majoritaire des domaines PASTA dans le groupe de position. Bleu : position 1, jaune : position 2, rouge : position 3, rouge foncé : position 34, marron : position 45. Les types de domaines PASTA sont représentés par des ellipses grises.

ABC.

FIGURE 4.10 – Diversité de la composition en domaines de StkP chez *Lactobacillales*. le domaine Pkinase est représenté en rectangle gris, les domaines PASTA en rectangles colorés. Les astérisques correspondent à l'organisation majoritaire au sein de la famille.

4.3.1.2 Histoire évolutive des domaines PASTA

A partir de ces résultats et de la phylogénie de référence des *Firmicutes* présentée au chapitre 3, nous avons pu établir un scénario évolutif pour retracer l'histoire évolutive des domaines PASTA de StkP chez les *lactobacillales* (figure 4.11). La protéine StkP chez l'ancêtre des *Lactobacillales* est inférée comme présentant trois domaines PASTA selon un schéma ABC'. À l'émergence des *Aerococcaceae*, le domaine de type B a été remplacé par un domaine de type C'. Étant donné la position de ce domaine dans l'arbre des groupes (figure 4.9), il semblerait que l'apparition du domaine PASTA de type C' soit la conséquence d'un transfert horizontal et non une duplication du domaine PASTA de type C' présent initialement dans la séquence de StkP.

À l'émergence de *Lactobaillus buchneri*, un domaine PASTA de type C' a été acquis en position terminale mais son origine est difficile à inférer puisque cette séquence présente une longue branche dans la phylogénie des domaines PASTA et que son positionnement n'est pas supporté (SH-like=0) (figure 4.6).

À l'émergence des *Streptococcaceae*, le domaine PASTA de type C' a été remplacé par un type C. Il est néanmoins difficile d'affirmer s'il s'agit d'un événement de recombinaison avec d'autres domaines PASTA ou d'une accumulation de mutations du domaine PASTA de type C' qui serait à l'origine de l'apparition du domaine PASTA de type C. À l'émergence de *S. suis*, le domaine PASTA de type B est remplacé par deux domaines PASTA de type A. Le premier semble issu d'un transfert horizontal à partir de *Streptococcus agalactiae* tandis que le second

provient d'une duplication du domaine PASTA de type A de *S. suis* (figure 4.7. L'émergence de *S. pneumoniae* et *S. pseudopneumoniae* est accompagnée d'une duplication du domaine PASTA de type B. Ce scénario est supporté par la phylogénie (figure 4.7).
À l'émergence des *Enterococcaceae*, le motif général est ABBA. Le domaine PASTA de type B semble avoir subi une duplication d'après la phylogénie (figure 4.8). Le domaine PASTA de type C' a été perdu et remplacé par un domaine PASTA de type A. Ce dernier ne semble pas provenir d'une duplication du domaine PASTA de type A de l'ancêtre des *Enterococcaceae* mais d'un transfert d'après l'arbre de distances (figure 4.9). À l'émergence de *M. plutonius* un domaine PASTA de type C remplace les deux domaines PASTA terminaux de types B et A. Similairement, à l'émergence de *E. faecalis*, *E. sp.*, un domaine PASTA de type C remplace le deuxième domaine PASTA de type B puis est dupliqué. L'origine des domaines PASTA de type C est difficile à inférer. En effet, ceux ci sont groupés ensemble dans la phylogénie des domaines PASTA des *Enterococcaceae* (figure 4.8) et se groupent avec les domaines PASTA de type C des *Streptococcaceae* sur l'arbre de distance (figure 4.9). Deux scénarios sont ainsi plausibles. Le premier scénario consiste à expliquer l'apparition des domaines PASTA de type C à l'aide de deux transferts horizontaux à partir des *Streptococcaceae*. La seconde possibilité est que le domaine PASTA de type C était présent à l'ancêtre des *Enterococcaceae* mais qu'il a été perdu à de nombreuses branches.

4.3.2 Les différents rôles des domaines PASTA de StkP chez *S. pneumoniae*

Dans l'étude que nous avons mené, les membres du laboratoire MMSB ont caractérisé le rôle fonctionnel de chacun des quatre domaines PASTA de StkP chez *S. pneumoniae*. J'ai quant à moi étudié plus spécifiquement l'histoire évolutive des domaines PASTA au sein des *Streptococcaceae*. Pour cela, j'ai utilisé la base de données spécifique des *Streptococcaceae*.

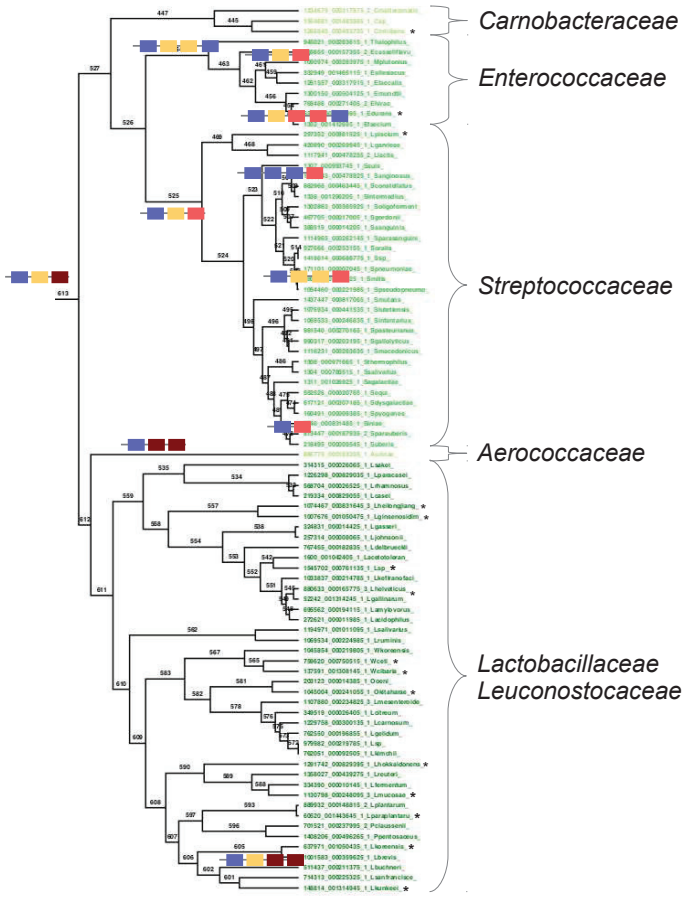


FIGURE 4.11 – Histoire évolutive des domaines PASTA de StkP chez les *Lactobacillales*. L'état ancestral des *Lactobacillales* est représenté à la racine. Les motifs sur les autres branches représentent des états de changement par rapport aux états des branches ancestrales. La phylogénie des *Lactobacillales* est issue de la phylogénie des *Firmicutes* présentée au chapitre 3. Les astérisques correspondent aux espèces non représentées dans l'analyse.

4.3.2.1 Le rôle fonctionnel des différents domaines PASTA de StkP chez *S. pneumoniae*

Dans cette étude, de nombreux mutants de *S. pneumoniae* ont été construits afin de caractériser le rôle respectif des différents domaines PASTA de StkP. Nous n'allons pas détailler l'ensemble des résultats obtenus mais dégager les idées principales.

Chez *S. pneumoniae*, StkP est impliquée principalement dans la division et la morphogenèse [148], [28]. Chaque domaine PASTA de StkP chez *S. pneumoniae* a ainsi été délété afin d'observer les potentiels défauts de division cellulaire. Il a été observé que seule la délétion du domaine PASTA 4 induit un défaut de division cellulaire, les cellules sont plus allongées. De plus, la délétion du domaine PASTA 4 entraîne la formation de chaînes, c'est-à-dire que les cellules ne se séparent pas normalement en fin de division. L'influence du nombre de domaines PASTA sur l'activité kinase de StkP a également été étudiée. D'après les résultats, plus le nombre de domaines PASTA diminue, plus l'activité kinase de StkP diminue. Il semble donc que le nombre de domaines PASTA soit essentiel à l'activité kinase de StkP. La localisation de StkP en fonction des domaines PASTA délétés a ensuite été étudiée. Les résultats suggèrent que le domaine PASTA 4 est particulièrement important pour la localisation de StkP au site de division. D'après l'ensemble de ces résultats, le domaine PASTA 4 semble remplir un rôle différent des trois autres. Les résultats suggèrent ainsi que l'ensemble des domaines PASTA sont important pour l'activité kinase tandis que le domaine PASTA 4 conditionne la localisation de StkP et la morphologie des cellules.

Comme expliqué précédemment, les mutants exprimant StkP sans domaine PASTA 4 semblent former des chaînes de cellules. Ce phénotype a déjà été décrit dans la littérature et est caractéristique de celui d'un mutant LytB, une hydrolase du peptidoglycane qui permet la séparation entre les cellules filles après la division [156]. La protéine StkP et particulièrement le domaine PASTA 4 serait potentiellement impliqué dans la localisation ou l'activité de LytB. Afin de prouver cette hypothèse, la localisation de LytB a été observée en présence et en l'absence du domaine PASTA 4. Ainsi, la délétion du domaine PASTA 4 semble affecter la localisation de LytB. Afin de savoir quels résidus du domaine PASTA 4 seraient impliquées dans la localisation de LytB, la structure calculée du domaine PASTA 4 a été réalisée. Celui-ci présente un motif

de « main crochue » non retrouvé dans les structures modélisées des trois autres domaines PASTA et impliquant les résidus R633, E636, K642, R644 et K646. Des mutations ponctuelles ont ensuite été effectuées sur ces résidus (en Alanines). Le mutant correspondant semble également présenter une délocalisation de la protéine LytB. Cela suggère que ces 5 résidus sont impliqués dans l'interaction avec LytB.
Une question reste néanmoins non résolue. En effet, pourquoi les mutants *S. pneumoniae* délétés des domaines PASTA 1, 2 et/ou 3 conservent-t-ils un phénotype similaire à celui du sauvage alors que l'activité kinase de StkP, pourtant essentielle à la morphogenèse [148], [150], est d'autant plus faible que le nombre de domaines PASTA est réduit ? L'hypothèse qui a été émise est que les domaines PASTA 1, 2 et 3 serviraient de « règle » pour positionner le domaine PASTA 4 à la bonne distance de la membrane. LytB serait ainsi d'autant plus proche de la membrane que le nombre de domaines PASTA proximaux serait faible. Elle pourrait donc hydrolyser le peptidoglycane proche de la membrane et ainsi réduire l'épaisseur de la paroi cellulaire. L'activité kinase de StkP, nécessaire à la septation [150], dans ce contexte n'a donc pas besoin d'être très efficace puisque la constriction de l'anneau Z nécessiterait moins de « puissance » pour séparer les deux cellules filles. Le faible nombre de domaines PASTA et ainsi la diminution de l'activité kinase de StkP serait donc compensés par la localisation plus proche de la membrane de LytB par le domaine PASTA 4 (sous réserve que le domaine PASTA 4 est encore présent). Pour vérifier cette hypothèse, l'épaisseur du peptidoglycane chez les mutants délétés du domaine PASTA 3, PASTA 2 et 3 puis PASTA 1, 2 et 3 a été observée. Les résultats indiquent que l'épaisseur du septum diminue d'autant plus que le nombre de domaines PASTA délétés est grand, suggérant ainsi que l'hypothèse formulée précédemment est vérifiée.

4.3.2.2 Les caractéristiques spécifiques du PASTA terminal

L'analyse fonctionnelle des domaines PASTA chez *S. pneumoniae* a révélé l'importance du domaine PASTA 4 par rapport aux autres domaines PASTA, notamment du fait qu'il interagit avec la protéine LytB. Les analyses phylogéniques indiquent que ce domaine appartient au type C. Nous avons calculé la distance patristique par paire entre les séquences de chaque type de domaine PASTA mais également entre les séquences du domaine kinase chez les *Streptococ-*

caceae (figure 4.12, 4.13). Il semble que les domaines PASTA présentent une vitesse évolutive significativement supérieure à celle du domaine kinase. De plus, le domaine de type C présente des distances patristiques par paire plus élevées que pour les deux autres types. Ces résultats indiquent que les domaines de type C évoluent plus vite que les deux autres.

Le domaine PASTA 4 de StkP chez *S. pneumoniae* étant particulier par rapport aux autres domaines principalement par le fait qu'il interagit avec LytB, nous avons voulu savoir si l'interaction de LytB avec le domaine PASTA 4 mise en évidence chez *S. pneumoniae* était généralisable à l'ensemble des *Streptococcaceae*. Pour cela, nous avons donc déterminé quelles espèces contenaient une copie de LytB. Cette dernière étant composée d'un domaine Glucosaminidase et d'une répétition de domaines CW_binding_1 (qui forme un domaine de liaison aux cholines), nous avons d'abord identifié tous les gènes présentant un domaine glucosaminidase chez les *Streptococcaceae* (figure 4.14) puis nous avons restreint la distribution taxonomique de LytB aux séquences qui possédaient aussi une répétition de domaines CW_binding_1, soit un domaine de liaison à la choline (figure 4.14). Nous avons ensuite projeté la présence ou l'absence des résidus du domaine PASTA de type C montrés comme étant impliqués dans l'interaction avec LytB chez *S. pneumoniae* (figure 4.15). Les résultats montrent que seules les espèces *S. pneumoniae*, *S. pseudopneumoniae*, *S. mitis* et *S. oralis* possèdent LytB et la majorité des résidus du domaines PASTA type C impliqués dans l'interaction avec LytB. L'interaction du domaine PASTA de type C avec LytB serait donc spécifique de ces espèces.

Il apparaît donc que le domaine PASTA de type C présente un taux d'évolution plus élevé que pour les deux autres types A et B mais que l'interaction entre LytB et le domaine PASTA de type C soit uniquement conservée chez *S. pneumoniae*, *S. pseudopneumoniae*, *S. mitis* et *S. oralis*. Il est donc tentant de faire l'hypothèse que les domaines de type C chez les autres *Streptococcaceae* posséderaient donc des rôles ou tout du moins des partenaires différents. Cependant, étant donné la diversité de l'organisation modulaire des hydrolases possédant un domaine glucosaminidase, il est possible que la divergence des domaines PASTA de type C chez les *Streptococcaceae* soit la conséquence d'une adaptation avec la composition en domaines des hydrolases. Le domaine PASTA de type C et les hydrolases fonctionneraient donc comme un véritable système « clé-serrure » ou chaque espèce/groupe d'espèces posséderait un système

spécifique non transposable à d'autres *taxa*. Cette hypothèse est renforcée par le fait que les domaines PASTA de StkP ont été prouvés comme interagissant avec du muropeptide provenant uniquement de la même espèce ou d'une espèce produisant une paroi de composition similaire [325], [423]. Il est aussi possible de faire l'hypothèse que les domaines PASTA de type C interagissent avec d'autres protéines dont la nature n'est pas encore connue, éventuellement non impliquées dans la division cellulaire. Cela serait consistant avec le fait que la protéines StkP semble impliquée dans des processus cellulaires variés en fonction de l'espèce considérée [300], [223], [377], [444], [321], [372].

4.4 Conclusion

Nous avons donc à travers cette étude reconstruit l'histoire évolutive des domaines PASTA de StkP chez les *Lactobacillales*. Cette analyse s'est révélée particulièrement difficile du fait de la petite taille des domaines PASTA et de leurs fortes divergences limitant ainsi la robustesse des arbres reconstruits. nous a permis de classer les domaines PASTA en fonction de leur relation phylogénétique. Nous avons montré que l'ordre des différents types de domaines PASTA au sein des séquences de PBPs et de StkP sont conservés chez la plupart des espèces de *Lactobacillales*. StkP chez les *Streptococcaceae* et les *Enterococcaceae* présente néanmoins de grandes variabilités dans la composition en domaines PASTA et une histoire évolutive plus complexe que les autres familles de *Lactobacillales*. La cause de cette complexité n'est pas connue. Le couplage entre l'approche expérimentale et phylogénétique a permis d'identifier le caractère spécial du domaine PASTA de type C chez les *Streptococcaceae*. Cette étude a donc permis de mettre en évidence l'histoire complexe de StkP chez les *Lactobacillales* et les implications fonctionnelles que cela pouvait avoir sur la physiologie des bactéries. De manière plus générale, ce type d'analyse réalisée à l'échelle des domaines fonctionnels d'une protéine est complémentaire de celles développées au chapitre 3.

FIGURE 4.12 – Phylogénie des domaines PASTA chez les *Streptococcaceae* inférée par approche bayésienne. MrBayes, Modèle mixte+G4, 110 séquences, 45 positions. La racine est composée des séquences de domaines PASTA de PBPs. Les différents types de domaines PASTA de StkP sont représentés par les couleurs : bleu : type A, jaune : type B, rouge, type C. Le diamètre des cercles aux branches est proportionnel à la valeur de probabilité postérieures (PP). Les branches présentant une PP inférieure à 0,5 ont été collapsée. La barre d'échelle indique le taux moyen de substitution par site. Tiré de [527].

FIGURE 4.13 – Distribution des distances patristiques par paire issues des arbres bayésiens de chaque type de domaines. Pkinase : gris, domaine PASTA type A : bleu, domaine PASTA type B : jaune, domaine PASTA type C : rouge. La borne inférieure est supérieure de la boite représente les percentiles de 25% et 75%. La barre dans la boite représente la médiane. Les points représentent les valeurs aberrantes. Un teste bilatéral de Wilcoxon ($* \ p \leq 0,01$) indique que les distributions sont toutes différentes.Tiré de [527].

FIGURE 4.14 – Distribution taxonomique des protéines à domaine Glucosaminidase chez les *Streptococcaceae*. L'arbre d'espèces des *Streptococcaceae* a été inféré à partir des séquences de protéines ribosomiques (PhyML, LG+I+G4, 34 séquences, 6.255 positions). La barre d'échelle indique le taux moyen de substitution par site. Le diamètre des cercles aux branches est proportionnel à la valeur de bootstrap associée supérieure à 50%. Tiré de [527].

FIGURE 4.15 – Distribution taxonomique de StkP, de LytB et des résidus impliqués dans l'interaction entre le domaine PASTA C et LytB chez les *Streptococcaceae*. La conservation des résidus du domaine PASTA C impliqués dans l'interaction avec LytB sont représentés par des carrés cyan. La conservation de LytB est représentée par des carrés violets. L'arbre d'espèces des *Streptococcaceae* a été inféré à partir des séquences de protéines ribosomiques (PhyML, LG+I+G4, 34 séquences, 6.255 positions). La barre d'échelle indique le taux moyen de substitution par site. Le diamètre des cercles aux branches est proportionnel à la valeur de bootstrap associée supérieure à 50%. Tiré de [527].

Chapitre 5

Développement d'un logiciel de visualisation de contextes génomiques, GeneSpy

Sommaire

5.1 Contexte 316

5.1.1 La visualisation du contexte génomique : principes et enjeux 316

5.1.2 Les logiciels disponibles 318

5.2 GeneSpy, un outil flexible et facile d'utilisation pour l'exploration des contextes génomiques 320

5.2.1 Librairies, environnement et implémentation 320

5.2.2 Construction et gestion de la base de données 320

5.2.3 Importation des identifiants de gènes 324

5.2.4 Options de visualisation des figures 326

5.2.5 Exportation des données 327

5.3 Conclusion 328

Au cours de l'analyse de l'histoire évolutive des familles de gènes du cycle cellulaire, la visualisation du contexte génomique a été nécessaire afin non seulement de guider l'identification des sous-familles de gènes au sein de familles multigéniques mais également de mettre en évidence de potentiels liens fonctionnels entre familles de gènes. Au début de ma thèse, aucun outil performant ne permettait l'analyse des bases de données locales n'était disponible. J'ai donc entrepris de développer mon propre logiciel qui puisse répondre à mes attentes. Ce logiciel nommé GeneSpy a fait l'objet d'une publication :

Garcia PS, Jauffrit F, Grangeasse C, Brochier-armanet C. GeneSpy, a user-friendly and flexible genomic context visualizer. Bioinformatics. 2018.

Nous allons ici décrire plus en détail l'implémentation, le fonctionnement de GeneSpy et les options disponibles qui ne figurent pas dans la publication.

5.1 Contexte

5.1.1 La visualisation du contexte génomique : principes et enjeux

L'exploration de l'organisation spatiale des gènes au sein des génomes est une composante essentielle en phylogénomique. En effet, la position relative des gènes les uns par rapport aux autres peut être d'une importance capitale pour un organisme [53], [231], [454].
Chez les procaryotes, une grande majorité de gènes est organisée en opérons qui consistent en de véritables unités de transcription [216]. Les gènes au sein d'un opéron sont co-transcrits c'est-à-dire que la machinerie de transcription produit un ARNm polycistronique contenant l'ensemble des gènes. Par la suite, l'ARNm est traduit gène par gène ce qui génère les protéines associées aux gènes de façon concomitante. De manière générale, il est admis que les protéines produites à partir des gènes d'un même opéron sont reliées fonctionnellement. De façon plus large, la vicinité de plusieurs gènes au sein d'un chromosome peut indiquer une relation fonc-

tionnelle même si ceux-ci ne sont pas en opéron [231]. En effet, l'ADN dans une cellule est compacté et seulement certaines régions sont décompactées à un instant t. La décompaction de l'ADN étant nécessaire à la transcription, lorsque une région est décompactée, l'ensemble des gènes présents sur cette portion est susceptible d'être transcrit. La colocalisation des gènes codant pour des protéines reliées fonctionnellement permet donc d'assurer que tous les partenaires sont exprimés en même temps.

Il existe aussi une régulation spatiale de l'expression des gènes au sein de la cellule. Chez les bactéries, le chromosome est extrêmement structuré et certaines régions du chromosome sont localisées précisément à des endroits de la cellule. Il a été suggéré que les ARNm étaient produits puis traduits au même endroit ce qui suggère qu'une colocalisation au sein du génome peu impliquer une colocalisation des ARNm [344]. Cette colocalisation des ARNm pourrait donc faciliter la production des protéines impliquées dans un même processus au même endroit et ainsi faciliter leur interaction. Il a même été montré par Sobetzko et collègues que le positionnement relatif à l'origine de réplication des gènes impliqués dans le cycle cellulaire était corrélé à leur profil d'expression temporel [434]. La position d'un gène au sein du chromosome est donc d'une importance certaine pour mener à bien les divers processus cellulaires.

Le génome étant une structure très plastique, le positionnement absolu (par rapport à l'origine de réplication) et relatif (les gènes les uns par rapport aux autres) varie en fonction des espèces. Cette plasticité est due aux nombreux réarrangements chromosomiques qui conduisent aux déplacements, pertes, recombinaisons, troncatures et duplications des gènes. Néanmoins, durant l'évolution et la diversification du vivant, certaines organisations géniques sont conservées comme par exemple le cluster de division cellulaire (« Division and Cell Wall Cluster ») où l'on retrouve un certain nombre de gènes impliqués dans la division cellulaire regroupés et organisés dans un ordre précis [263], [484], [310]. Cela est du au fait que la disruption de ce type de clusters mène à une dérégulation des processus impliquant les gènes qu'ils contiennent, conduisant ainsi à la diminution de la « fitness » et la viabilité de la cellule. Néanmoins, la disruption d'un cluster de gènes peut, dans certains cas, être une innovation bénéfique pour la cellule ce qui peut conduire à sa fixation au sein du génome.

De façon pratique, l'étude de la vicinité entre gènes permet d'identifier des relations fonctionnelles entre gènes [231]. Il est ainsi possible de faire des hypothèses sur la fonction des gènes

dont la fonction n'a pas encore été caractérisée [288], [384]. L'analyse des contextes génomiques des gènes présente aussi un intérêt dans l'identification d'orthologues [93]. Il existe en effet de nombreux cas de familles multigéniques constituées de plusieurs sous-familles (ou orthologues). Dans ces cas, il est parfois nécessaire de différencier les sous-familles. Ce fut notamment le cas au cours de l'analyse du cycle cellulaire effectuée durant ma thèse. Les sous-familles présentent parfois des contextes génomiques conservés ce qui peut permettre d'identifier formellement et distinctement les groupes d'orthologues.

5.1.2 Les logiciels disponibles

Il est donc d'un intérêt certain d'observer l'organisation des gènes au sein des génomes notamment pour le type d'analyses effectuées durant ma thèse. Il existe de nombreux outils permettant d'observer le contexte génomiques de familles de gènes au sein de plusieurs génomes [361], [308], [354], [447]. Malheureusement, ces outils comportent des désavantages majeurs lors des analyses (table .19).

Tout d'abord, la plupart des outils informatiques permettant de visualiser les contextes génomiques sont basés sur des bases de données en ligne sur lesquels l'utilisateur n'a pas la main [361], [308], [354]. Il n'est donc pas possible de les utiliser pour l'analyse de données de l'utilisateur. Il n'est aussi en général pas possible d'observer les contextes génomiques d'eucaryotes et de procaryotes avec un même outil [361], [308], [354]. Aussi, la plupart des outils en ligne n'utilisent que des génomes complets excluant ainsi l'analyse des génomes à l'état de « Draft » ou d'« Assembly » [361], [308], [354]. Enfin, l'utilisateur n'a pas le controle sur les mises à jour des bases de données qui peuvent dans certains cas ne pas être suffisamment régulières.

Certains outils ne permettent pas ou difficilement l'utilisation de données de l'utilisateur. Le programme Easyfig par exemple ne prend en entrée que des fiches GenBank de régions spécifiques de génomes qui sont difficilement générables de façon automatisée [447]. Autre exemple, SyntTax ne prend en entrée que des mots clés ou une séquence ce qui exclu l'utilisation d'identifiants des bases de données les plus utilisées [354].

La sortie visuelle des contextes dans la plupart des outils n'est pas optimale (figure 5.1). Tout

d'abord, l'aspect purement visuel est parfois de mauvaise qualité et manque de lisibilité. Également, les informations affichées sur les figures telles que les noms de gènes ou d'organismes ne sont pas systématiquement indiquées ce qui rend difficile l'analyse consécutive [447]. En général, il est difficile de mettre en forme facilement la figure (échelle, taille de la fenêtre génomique, informations affichées, largeur des flèches) ou de choisir simplement les couleurs des familles de gènes [308], [354].

Enfin, les fonctionnalités proposées par les différents outils sont très limitées. Il n'est pas exemple pas possible de faire des sous-sélections de gènes, d'explorer de façon dynamique et interactive les génomes ou encore d'obtenir facilement les informations relatives aux gènes affichés.

J'ai donc décidé de développer un outil de visualisation de contexte génomique qui puisse s'adapter à mes besoin dans le cadre de mon analyse mais qui puisse également offrir des fonctionnalités utiles à l'ensemble de la communauté scientifique, GeneSpy [157].

FIGURE 5.1 – Exemples de figures de contextes génomiques générées par différents programmes. (A) GeConT. (B) JContextExplorer. (C) MGcV. (D) EasyFig. (E) SynTax. (F) GeneSpy.

5.2 GeneSpy, un outil flexible et facile d'utilisation pour l'exploration des contextes génomiques

5.2.1 Librairies, environnement et implémentation

GeneSpy est un logiciel développé en Python 2.7 utilisant les librairies Matplotlib, Tkinter et SQLite. GeneSpy est disponible gratuitement sur https://lbbe.univ-lyon1.fr/GeneSpy/ sous la licence CeCILL 1. La documentation est disponible sur le site et sur un document PDF déposé dans l'archive du programme. Afin de rendre disponible l'outil à un maximum de chercheurs, celui-ci a été développé sur les trois systèmes d'exploitation MacOS, Windows et Linux bien que son utilisation soit optimisée sur Linux. Le programme est composé d'un script python GeneSpy.py, de plusieurs fichiers de dépendance tels que les fichiers de paramètres, de chemins et de couleurs. Le programme comporte une interface graphique interactive qui permet simplement de générer des contextes génomiques de n'importe quel gène. L'interface est composée de plusieurs boites et d'un menu (figure 5.2).

Trois classes principales ont été implémentées pour constituer le noyau du programme (figure 5.3). Ces trois classes gèrent la requête du gène, la production des figures SVG produites par Matplotlib et l'inclusion de ces figures dans une « Frame » générée par Tkinter. Le format universel d'entrée correspond à un format tabulaire constitué de deux éléments : le nom du fichier GFM (ou numéro d'assemblage) et le numéro d'accession du gène : $< Assemblage >< Accession >$.

5.2.2 Construction et gestion de la base de données

Les bases de données GeneSpy sont constituées de trois éléments : Les fichiers GFM (« GFF Minimal content »), une liste des souches contenues dans la base de données et un fichier SQL qui permet d'accélérer les requêtes (figure 5.4).

Les fichiers GFM sont générés à partir des fichiers GFF (« Generic Feature File ») (Annexe

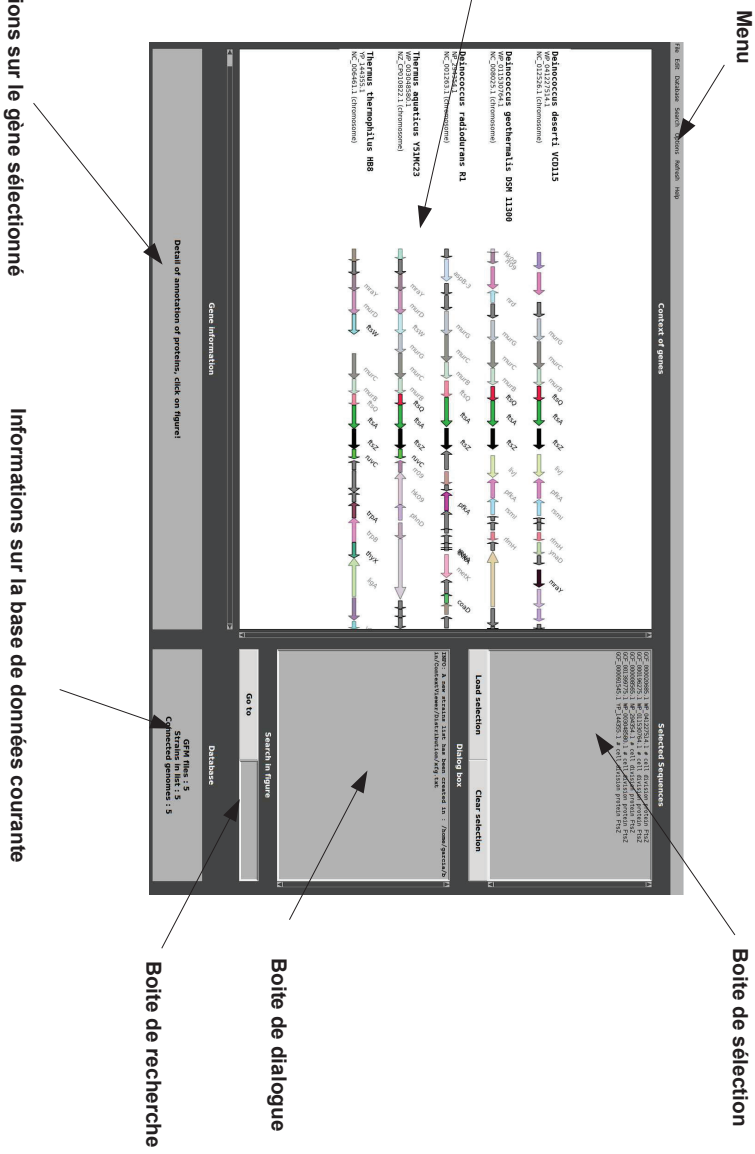


FIGURE 5.2 – Interface utilisateur de GeneSpy.

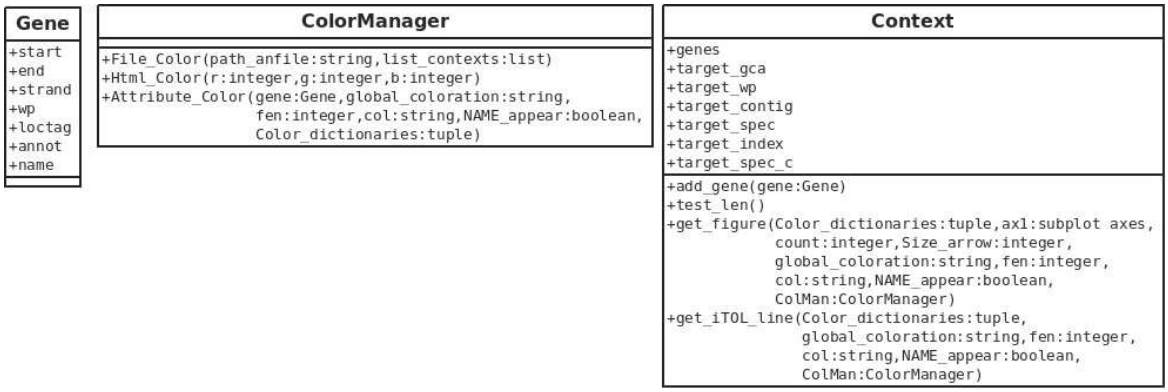


FIGURE 5.3 – Diagramme de classes de GeneSpy. il existe trois classes : Gene, Context et ColorManager. Les attributs sont représentés par des + et les méthodes par de +().

.20). Ces derniers contiennent l'ensemble des informations d'annotation des génomes et correspondent à un format standard généré par la grande majorité des programmes d'annotation comme RAST ou PROKKA [17], [421] (Annexe .20). Lors de la construction de la base de données, les informations nécessaires au fonctionnement de GeneSpy sont indexées dans les fichiers GFM, un format qui a pour avantage d'être très léger et très simple à manipuler : à chaque gène est associé un numéro de contig, sa position, son sens, son locus tag, son numéro d'accession, sa fonction biochimique et son nom. GeneSpy offre la possibilité de télécharger les fichiers GFFs à partir du FTP du NCBI (`ftp://ftp.ncbi.nlm.nih.gov/`) directement *via* l'interface. Pour cela, il est nécessaire de fournir une liste de liens FTP facilement accessible sur le site du NCBI (`https://www.ncbi.nlm.nih.gov/genome/browse/`). Les fichiers GFFs provenant de RAST et PROKKA [17], [421] sont également compatibles avec GeneSpy ce qui permet d'inclure dans des analyses des données non publiées.

La liste de souches est également générée lors de la construction de la base de données. Il s'agit d'un simple fichier texte qui établi la correspondance entre le nom du fichier GFM, du nom de souche affiché sur l'interface et du nom de la souche exporté en format iTOL [271] (Annexe .20). Dans le cas de fichiers provenant du NCBI, les noms des souches sont obtenus par les fichiers de rapport d'assemblage téléchargés en même temps que les GFF. Si les fichiers proviennent d'une autre base de données ou que les rapports d'assemblage ne sont pas présents dans le répertoire des fichiers GFF, le nom est directement extrait du nom de fichier GFF. Il

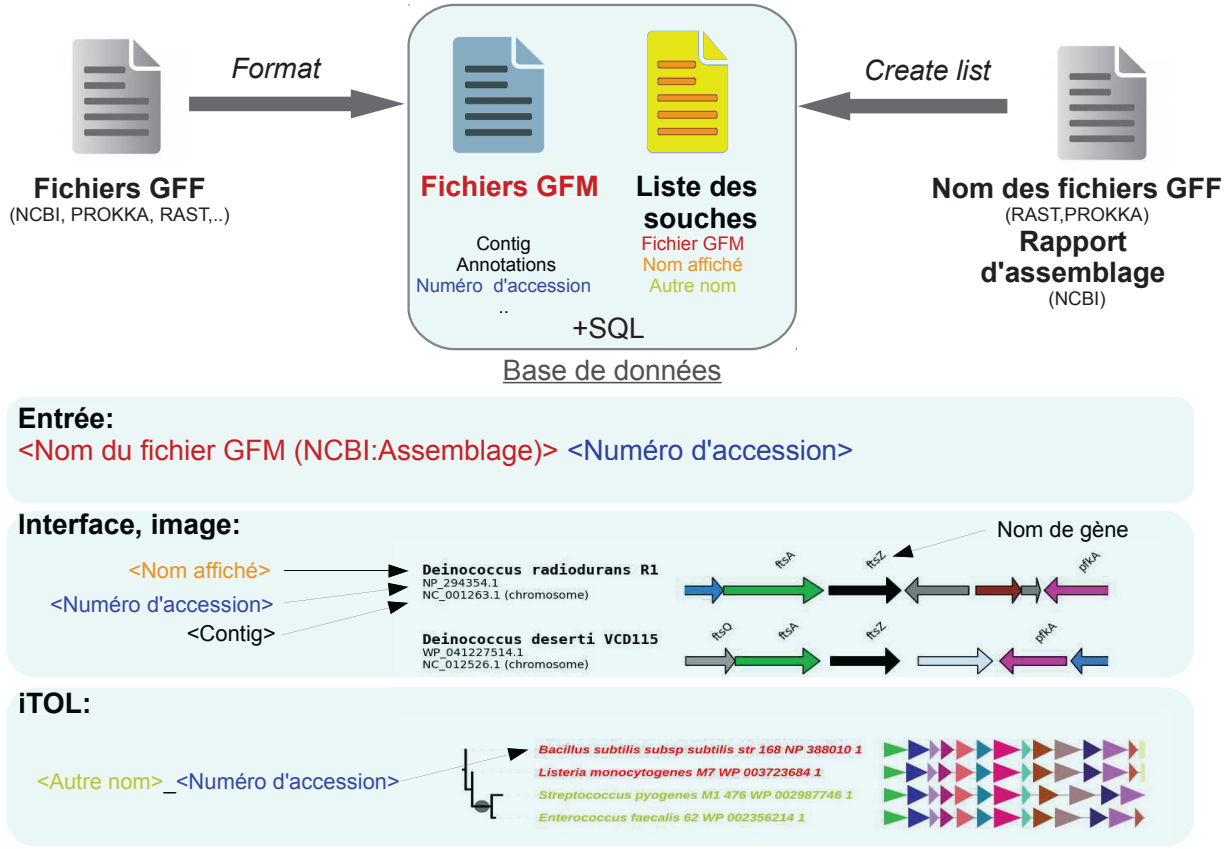


FIGURE 5.4 – Fonctionnement interne de GeneSpy. Les fichiers GFF et les rapports d'assemblage permettent de construire la base de données constituée de fichiers GFM et d'une liste des souches. Le format d'entrée permet d'afficher les contextes génomiques directement sur l'inteface mais aussi de projeter les contextes génomiques sur une phylogénie à l'aide d'iTOL [271].

est possible par l'interface d'éditer la liste de souches mais également de créer des nouvelles listes en faisant une sous-sélection. Il est même possible d'échantillonner automatiquement une souche par espèce de façon aléatoire pour limiter la redondance taxonomique et ne pas surcharger les analyses. La nouvelle liste ainsi créée permet de requêter uniquement les génomes qui y sont contenus tout en gardant l'ensemble du dossier contenant les fichiers GFM.

Enfin, le fichier SQL permet de faire la correspondance entre les numéros d'accession et les numéros d'assemblage donnés en entrée. La requête SQL permet à partir d'un numéro d'ac-

cession de retrouver les numéros d'assemblage associés ce qui permet de reformer l'identifiant de base $< Assemblage >< Accession >$.
Le chemin de la base de données (dossier contenant les fichiers GFM et la liste de souches) doit être choisis via un menu dédié à cet effet. La correspondance entre la liste de souches et les fichiers GFM est vérifiée et affichée en temps réel sur l'interface.

5.2.3 Importation des identifiants de gènes

Le chargement d'identifiants peut être effectué par de nombreux moyens et tolère de nombreux formats. Dans tous les cas, il est nécessaire d'obtenir un identifiant au format GeneSpy : $< Assemblage >< Accession >$ (figure 5.5, Annexe .20).

Tout d'abord, il est possible de requêter la base de données en effectuant une recherche par mot clé. Il suffit de fournir un nom d'espèce/souche ou même une partie du nom et un mot clé (annotation fonctionnelle, numéro d'accession, locus tag , ...). Le résultat de la recherche est alors converti en format GeneSpy.
Ensuite, une multitude de formats contenant des numéros d'accession sans le numéro d'assemblage sont acceptés (Annexe .20). Pour ces fichiers, il est nécessaire de faire la conversion $< Accession >$ en $< Assemblage >< Accession >$ via la base de données SQL.

- Un fichier de sortie de BLASTP effectué contre la base de données du NCBI si la base de données de GeneSpy provient du NCBI ou contre une base de données protéique locale si elle correspond exactement à celle de GeneSpy.
- Une collection de fichiers GenBank/GenPept de gènes/protéines.
- Un fichier contenant une liste de numéros d'accession dans un simple fichier texte.

Dans tous les cas, les identifiants sont stockés dans la boite de sélection et l'utilisateur doit ensuite les charger pour obtenir le contexte génomique des gènes cibles.
Les identifiants peuvent être sauvegardés sous un format texte de type $< Assemblage >< Accession >$. Ces fichiers peuvent ensuite être directement chargés via l'interface. Il est aussi possible de sauvegarder une session c'est-à-dire que l'ensemble des paramètres et des identi-

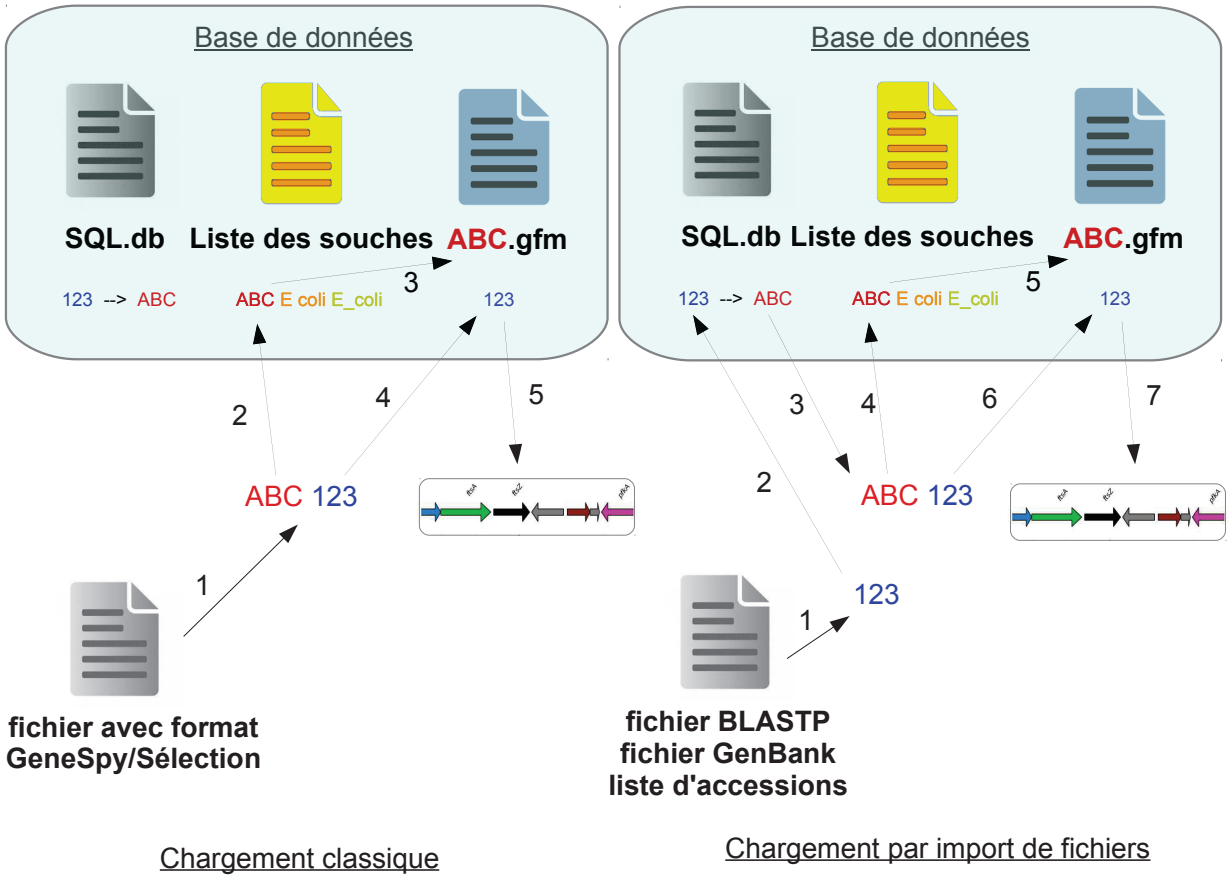


FIGURE 5.5 – Gestion des identifiants GeneSpy. (gauche) : Le fichier contient directement les identifiants $< Assemblage >< Accession >$ en format tabulaire. 1 : vérification du format, 2 : recherche de l'assemblage dans la liste, 3 : vérification que le fichier GFM associé existe, 4 : recherche de l' accession et des gènes voisins dans le GFM, 5 : génération de la figure. (droite) : Le fichier contient uniquement les numéros d'accession. 1 : vérification du format, 2 : recherche de tous les numéro d'assemblages associés au numéro d'accession, 3 : formation de l'identifiant $< Assemblage >< Accession >$ en format tabulaire. 4, 5, 6, 7 identique au panneau gauche.

fiants est sauvegardé.

Enfin, il est possible grâce à la fonction Convert_IDs du script GeneSpy_tools.py de charger des identifiants dans n'importe quel format et de les transformer directement en $< Assemblage >< Accession >$. La condition nécessaire est que $< Assemblage >$ et $< Accession >$ soient initialement présents dans l'identifiant. Cette fonction est particulièrement intéressante puisqu'elle

permet de charger des identifiants provenant d'autres programmes tels que Figtree (http://tree.bio.ed.ac.uk/software/figtree/) ou Aliview [260] par un simple copier/coller à partir de ces programmes directement dans la boite de sélection de GeneSpy.

5.2.4 Options de visualisation des figures

GeneSpy laisse la possibilité à l'utilisateur d'adapter à son goût les figures grâce à de nombreuses options. Il est en effet possible de trier les contextes par ordre alphabétique des noms des souches, d'afficher/cacher les noms des gènes, les numéros d'accession, les numéros de contig, les locus tag, de régler la taille et l'échelle de la région génomique observée et aussi la largeur des flèches représentant les gènes.
GeneSpy fourni aussi une multitude de fonctions afin de colorer les familles de gène (géré par la classe ColorManager). La fonction initiale colorie les gènes en fonction de l'annotation fonctionnelle et du nom du gène. Une couleur aléatoire est attribuée à chaque annotation fonctionnelle et à chaque nom. La correspondance entre les couleurs et les annotations est stockée dans le fichier AnnotCol.txt tandis que celle entre les couleurs les noms est sauvegardée dans le fichier NameCol.txt. Une optimisation de l'annotation des familles géniques est faite par une approche de regroupement par lien unique (« Single linkage clustering », SLC). Si dans un GFM, une annotation fonctionnelle et un nom de gène sont attribués à un même gène, une correspondance entre l'annotation fonctionnelle et le nom est crée (stockée dans AnnotName.txt) et tout gène étant annoté par l'annotation fonctionnelle reliée au nom prendra le nom et la couleur du nom. Également, les noms de gènes affichés sont légèrement transformés à l'affichage sur l'interface. Les noms de gènes perdent en effet leur majuscule et un maximum de 6 caractères est autorisé. La correspondance entre le nom réel du gène et le nom affiché est sauvegardée dans le fichier NameName.txt. Le stockage des couleurs et des annotations dans des fichiers permet ainsi de garder les mêmes couleurs d'une session à l'autre. Il est à noter que la banque de données de couleurs s'enrichit à chaque nouvelle portion de génome explorée. Il est possible d'éditer n'importe quel liens via un menu dédié (changer la couleur d'une annotation fonctionnelle/d'un nom ou la correspondance entre une annotation fonctionnelle et un nom).

La couleur peut aussi être adaptée à son goût de façon très précise en utilisant plusieurs fonctions de coloration spécifiques. Tout d'abord, il est possible de donner une ou plusieurs annotations de gène à GeneSpy pour que ces gènes soient colorés en bleu (fonction Monochrome). GeneSpy permet aussi d'importer un fichier contenant la correspondance entre des couleurs et des annotations ou des identifiants GeneSpy (Annexe .20). Cette flexibilité permet ainsi de construire très rapidement et facilement des figures de contexte génomique, quelques soient les besoin de l'utilisateur.

5.2.5 Exportation des données

GeneSpy permet d'exporter les figures générées en de nombreux formats. Tout d'abord, les figures peuvent être enregistrées aux formats PDF, SVG, PNG, EPS, TIF, JPG (figure 5.6A).

Également, les contextes peuvent être exportés dans un format compatible avec iTOL [271]. Cette fonctionnalité permet de projeter les contextes sur une phylogénie ce qui peut être extrêmement utile lors de l'analyse des événements évolutifs ayant affecté une région génomique (figure 5.6B). La correspondance entre les feuilles et la sortie de GeneSpy est faite par le nom de la feuille. il est donc nécessaire d'avoir exactement le même nom aux feuilles et dans la sortie de GeneSpy. Pour être le plus flexible possible, GeneSpy fourni quatre possibilités de noms de feuilles :

- $< Assemblage >$
- $< Assemblage >_< Accession >$
- $< Nomsouche >$
- $< Nomsouche >_< Accession >$

Les noms de souches correspondent à la troisième colonne de la liste des souches et peuvent être éditer soit de façon automatique via un script soit de façon manuelle sur l'interface dans le menu d'édition de liste.

5.3 Conclusion

GeneSpy a été développé dans un but de flexibilité et de facilité d'utilisation afin de pallier à ce manque retrouvé dans la plupart des autres logiciels. La portabilité a été maximisée afin que GeneSpy soit opérationnel sur le plus grand nombre de machines. Grâce à l'utilisation d'une base de donnés locale, GeneSpy permet de visualiser le contexte génomique de n'importe quel gène sur n'importe quel génome quelque soit sont état de complétude ou le domaine du vivant auquel l'organisme appartient. Les données peuvent être importées de très nombreuses façon et les figures de contextes sont très facilement adaptables. Enfin les figures peuvent être exportées en de nombreux formats et notamment en format compatible avec iTOL permettant ainsi de projeter le contexte génomique sur une phylogénie [271]. GeneSpy est donc particulièrement adapté pour les analyses phylogénomiques mais aussi pour générer des figures prêtes à être publiées. Néanmoins, l'utilisation des deux librairies Tkinter et Matplotlib induit certains écueils comme une fuite de mémoire et une lenteur dans le rafraîchissement des figures sur certains systèmes d'exploitation. Il serait ainsi envisageable de développer une nouvelle version basée sur d'autres librairies comme PyQt ou wxPython pour l'interface et ggplot-python pour les figures. Il serait aussi intéressant de compiler totalement la base de données en SQL.

FIGURE 5.6 – Exemples de figures générées par GeneSpy (Tiré de [157]). (A) Contexte génomique de *rpsN* chez 10 espèces de *Deinococcus-Thermus*. Les gènes codant pour les protéines ribosomiques sont colorés. (B) Projection des contextes génomiques de *rpsN* sur une phylogénie des espèces des *Deinococcus-Thermus* basé sur des séquences d'ARN 16S. Les gènes codant pour les protéines ribosomiques sont colorés en gris. L'émergence de *Deinococcus ficus* d'une part et de *Deinococcus reticulitermitis, wulumuqiensis, radiodurans* d'autre part sont accompagnées de réarrangements du cluster de gènes codant pour les protéines ribosomiques. La figure a été générée par iTOL [271].

Chapitre 6

Discussion générale et perspectives

6.1 Limites et apports de la méthode

6.1.1 Méthodes de détections des homologues et choix des seuils

Pour reconstruire les familles de gènes du cycle cellulaire, nous avons d'abord lancé des BLASTP itératifs à partir des séquences initiales. Lors d'un BLASTP, la sélection du seuil de E-value affecte énormément la qualité et la complétude d'un jeu de données de séquences présumées homologues. Nous avons choisi de nous baser sur le critère de longueur des séquences et des alignements par paire tandis que la plupart des études se basent sur des E-values fixes. Ce choix m'a semblé être plus justifié que celui d'un seuil arbitraire appliqué massivement à toutes les familles car toutes les familles de gènes présentent de grandes variabilités de composition et d'histoires évolutives. Ainsi, d'une famille à l'autre, un même seuil peut être trop stringent ou dans d'autres cas trop permissif. Néanmoins, le fait de sélectionner les séquences sur la base de leur longueur présente aussi un biais. En effet, le postulat de base est que les séquences homologues présentent des tailles similaires ce qui est loin d'être systématiquement le cas. De même, cela suppose aussi qu'une séquence présentant une taille similaire à celle de la séquence initiale serait homologue à cette dernière ce qui n'est pas non plus une règle absolue.

Le fait d'effectuer des BLASTP de façon itérative permet d'identifier la majorité des homologues et donc de réduire le nombre de faux négatifs. Cependant, dans quelques cas, l'approche itérative peut introduire un grand nombre de faux positifs et notamment à cause de la sélection de nouvelles graines non homologues.
Nous effectuons également une recherche par profil HMM pour réduire au maximum le nombre de faux négatifs. L'approche classique consiste à considérer comme homologues toutes les séquences dont la E-value est inférieure à 0,01. Là encore, l'utilisation d'un seuil arbitraire est très risquée. Nous avons donc décidé d'introduire une étape de vérification basée sur des BLASTP réciproques pour s'assurer que chaque nouvelle séquence analysée appartienne à la famille de gène initiale. Néanmoins, les résultats sont très dépendants de la base de données utilisée pour les BLASTP réciproques. Nous avons utilisé une base de données contenant une dizaine de génomes représentatifs de la diversité des *Firmicutes* afin d'optimiser la rapidité de la procédure, mais nous n'avons pas testé l'influence de l'échantillonnage taxonomique. La sélection des souches de cette base de données n'est donc peut-être pas optimale. Malgré les potentielles faiblesses de la détection des homologues, nous avons en cumulant les méthodes diminué le nombre de faux négatifs et en vérifiant systématiquement les alignements, évincé les faux positifs. De plus, les résultats semblent satisfaisants et similaires à ce qu'aurait donné une analyse manuelle.

6.1.2 Le cas complexe des familles multigéniques

A travers les nombreux cas rencontrés durant l'analyse des familles de gènes du cycle cellulaire, nous avons identifié principalement deux types de familles : les familles monogéniques et les familles multigéniques. Néanmoins, la limite entre les deux est parfois très mince. En effet, certaines familles comportaient uniquement un groupe de quelques séquences divergentes et redondantes avec la taxonomie du groupe principal. Il a été difficile de les classer en familles monogéniques ou multigéniques.
Pour le cas des familles multigéniques, nous avons tenté de séparer les différentes sous-familles. Ces dernières consistent en des groupes qui ont été majoritairement transmis de façon verticale.

Il s'agit donc théoriquement de familles constituées majoritairement d'orthologues. Néanmoins, nous avons systématiquement observé que certains groupes taxonomiques sont mal placés par rapport à la phylogénie des espèces au sein des phylogénies des sous-groupes. Ces inconsistances entre les phylogénies des gènes et des espèces peuvent être dues à de multiples transferts horizontaux/paralogies cachées mais aussi à des biais de reconstruction, notamment dans le cas de *taxa* dont les séquences évoluent systématiquement plus vite que les séquences des autres *taxa*. Les phylogénies de gènes présentent de nombreuses branches non soutenues et pour quelques familles de gènes, ont été construites à partir d'alignement multiples avec peu de positions conservées. Il est donc difficile de déterminer si les inconsistances entre phylogénies de gènes et d'espèces reflètent une réalité biologique ou une insuffisance de signal phylogénétique. Il est difficile de lever toutes les incertitudes mais il est important d'en avoir conscience lors de l'analyse des résultats.

6.1.3 Méthode de reconstruction des états ancestraux et coûts d'événements

Afin de reconstruire l'histoire évolutive des familles de gènes du cycle cellulaire, nous avons utilisé trois techniques basées sur la parcimonie et qui nécessitent de pondérer le coût de chaque type d'événement.

La première méthode consiste en la reconstruction de l'histoire des gènes à partir des profils phylogénétiques. Cette méthode bien que rapide et relativement facile à implémenter ne prend pas en compte les duplications et les transferts. L'hypothèse sous-jacente est donc que tous les gènes de la famille considérée sont orthologues. Pour choisir la valeurs des coûts (gains et pertes) nous avons utilisé une solution intermédiaire entre l'algorithme de Fitch (même coût pour les deux événements) et Dollo (un seul gain possible) en maximisant le coût d'un gain mais en laissant la possibilité d'en inférer plusieurs. Nous avons testé plusieurs possibilités sur quelques familles et nous avons sélectionné les coûts qui présentaient les résultats les plus cohérents de façon qualitative.

La deuxième technique consiste en la réconciliation des phylogénies de gènes avec l'arbre

d'espèces. Contrairement à l'approche par profils phylogénétiques, cette méthode permet de prendre en compte une plus grande variété d'événements comme les pertes, les transferts horizontaux ou les duplications. Néanmoins, comme nous l'avons observé précédemment, il existe de nombreux biais de reconstruction pouvant affecter les phylogénies. Les résultats inférés par réconciliation sont donc très sensibles à ces biais. Enfin, le choix des coûts est très délicat. Nous avons opté pour la solution implémentée dans ecceTERA proposée par [420] mais l'estimation des coûts reste un problème épineux.

L'approche basée sur l'étude de l'évolution des synténies de gènes nécessite de fixer un seuil de proximité en nombre de nucléotides (2kb) permettant de décider si deux gènes sont voisins ou non. Ce seuil a été également testé sur quelques familles mais nécessiterait d'avantage d'analyses pour être optimisé. Aussi, il aurait été possible d'utiliser un seuil de nombre de gènes situés entre les deux gènes considérés plutôt qu'une distance en nombre de nucléotides. Enfin, l'approche nécessite de pondérer les coûts d'événements (gain et perte de synténie) et est donc soumis aux mêmes questionnements que les approches décrites précédemment.

6.1.4 Méthodes corrélatives basées sur la reconstruction des états ancestraux par parcimonie

A partir de l'inférence des histoires évolutives des familles de gènes du cycle cellulaire, nous avons inféré des liens fonctionnels entre ces familles. Pour les trois types d'approches, nous avons utilisé des méthodes de parcimonie pour inférer les évenements ancestraux.

Concernant les profils phylogénique, la technique que nous avons utilisée, basée sur la reconstruction des états ancestraux, a été montrée comme plus spécifique que les techniques dites « naïves », c'est-à-dire qui ne prennent pas en compte l'arbre d'espèces [241]. Pour ce qui est de la réconciliation, nous avons implémenté une méthode plus simple basée sur le calcul du coefficient de Jaccard. La méthode utilisée pour inférer les liens évolutifs à partir de la reconstruction des états ancestraux de synténie est une technique qui n'a, à ma connaissance , pas encore été décrite. Néanmoins, le programme DeCoSTAR permet, en réconciliant une phylogénie de gènes et un arbre d'espèces de prendre en compte les synténies et reconstruire

les synténies ancestrales [106].

Enfin, nous avons calculé un score global intégrant les résultats des trois méthodes pour définir la force de corrélation entre les familles. Ce score est basé sur la moyenne des Z-scores obtenus par les trois méthodes. Il serait necessaire d'optimiser ce score notamment en étudiant son comportement en fonction des métriques utilisées pour son calcul.

6.1.5 Méthodes de détection de potentielles nouvelles familles du cycle cellulaire

L'analyse des familles de gènes présentes chez les 304 *Firmicutes* nous a permis d'identifier plusieurs familles présentant des caractéristiques proches en termes de synténie et d'histoires évolutives de celles du cycle cellulaire. Pour quantifier la corrélation entre les familles, nous avons utilisé les profils phylogénétiques. Néanmoins, cette méthode présente plusieurs biais. Tout d'abord, nous avons vu que les familles générées par la technique de clusterisation ne sont pas toujours construites correctement. Ceci est du probablement à l'utilisation des mêmes critères de clusterisation pour toutes les familles. De plus, nous avons utilisé des profils phylogénétiques « naïfs » (sans tenir compte de l'arbre d'espèces) qui ont été décrits comme moins performants dans la détection des liens fonctionnels que les profils prenant en compte la phylogénie des espèces [241]. Cette méthode nécessiterait donc d'être optimisée.

6.1.6 Développement d'un outil de visualisation de contextes génomiques

Durant ma thèse j'ai également développé le logiciel GeneSpy, un visualiseur de contextes génomiques. Il a en effet été nécessaire pour mener à bien mes analyses de développer cet outil puisqu'aucun de ceux déjà existant ne répondait à mes attentes. GeneSpy possède ainsi de nombreuses fonctionnalités et fonctionne avec une base de données locale ce qui a pour avantage pour l'utilisateur de pouvoir observer le contexte génomique de gènes présents dans une base de données qu'il aura lui même conçu. GeneSpy permet aussi de projeter très sim-

plement le contexte génomique sur une phylogénie ce qui n'a jamais été proposé de façon aussi simple par les autres logiciels. GeneSpy est disponible en libre accès sur le site https://lbbe.univ-lyon1.fr/GeneSpy/ qui a été visité plus de 300 fois dans plus de 40 pays depuis juin 2018.

6.2 Inférences de liens fonctionnels et identification de nouveaux composants du cycle cellulaire

6.2.1 Reconstruction des familles de gènes du cycle cellulaire et de leur histoire

Durant ma thèse, j'ai identifié précisément les familles de gènes impliquées dans le cycle cellulaire chez les *Firmicutes*. J'ai ainsi délimité les familles pouvant être considérée comme quasiment orthologues. Il est tentant de faire l'hypothèse que les séquences au sein de chaque famille possèdent la même fonction. Leur histoire évolutive respective a été inférée, au niveau individuel et organisationnel. Ces résultats nous ont montré que l'émergence des *Bacilli* était accompagnée d'une multitude d'acquisitions de gènes du cycle cellulaire et que l'émergence des *Streptococcaceae* était concomitante avec de nombreux réarrangements organisationnels. Concernant les *Streptococcaceae*, cela parait cohérent avec les données expérimentales puisque notamment *S. pneumoniae*, l'organisme modèle de ce taxon, présente des mécanismes de division très différents des autres *Firmicutes* comme *B. subtilis*. Aussi, nous avons pu montré que certaines espèces ne présentent aucun des systèmes décrits pour un processus cellulaire donné comme la localisation de l'anneau Z.

6.2.2 Identification des familles potentiellement liées fonctionnellement

La méthode que j'ai développée durant ma thèse nous a conduit à identifier un grand nombre de liens fonctionnels déjà décrits dans la littérature, ce qui a permis de montrer l'interêt de l'approche, mais également de proposer des liens jamais décrits. Cela suggère que la méthode utilisée malgré les nombreux biais cités précédemment est efficace dans la prédiction des liens fonctionnels. A travers les différents exemples de clans observés au sein du réseau de corrélation, nous avons constaté que deux clans de familles de gènes sont unis sur la base de leurs profils phylogénétiques tandis que les autres clans sont principalement reliés par les contextes génomiques. Tous les processus du cycle cellulaire semblent ainsi reliés évolutivement entre eux. Également, de nombreuses familles impliquées dans la traduction et la modification des ARN semblent liés aux cycle cellulaire. Enfin, un grand nombre de potentiels liens fonctionnels jamais décrits ont été inférés entre les familles du cycle cellulaire comme par exemple entre CozE4 et les familles de la sporulation ou encore entre PBPBX1 et FtsW.

6.2.3 Identification de nouvelles familles du cycle cellulaire

Nous avons identifié un certain nombre de familles de gènes présentant des similarités avec les familles du cycle cellulaire mais dont la fonction n'est soit pas directement reliée avec ce processus, soit inconnue. Ainsi, 43 familles de gènes corrélant avec le clan *Bacilli*, 9 avec le clan Spo, 12 avec la famille B3, 3 avec la famille SpoVG, 1 avec MurA2 et 8 avec la famille WalR. Onze d'entre elles présentent une synténie conservée avec certaines familles du cycle cellulaire ce qui en ferait de bons candidats pour une validation expérimentale.

6.3 Perspectives

6.3.1 Amélioration de la méthode

Concernant la méthode, plusieurs améliorations pourraient être faites.
Il serait tout d'abord intéressant d'analyser l'ensemble des familles de gènes du cycle cellulaire puisque nous avons dans cette étude écarté un certain nombre de protéines pourtant impliquées dans le cycle cellulaire mais qui présentaient des histoire évolutives trop complexes. Il existe aussi une réelle nécessité d'automatiser certaines étapes de la procédure de découpage en sous-familles car cette étape est extrêmement chronophage et laborieuse. Il serait par exemple intéressant d'identifier sur un arbre d'espèces les bipartitions correspondant le plus probablement à l'émergence d'une sous-famille dans le cas des familles multigéniques. De plus, la méthode manuelle peut introduire des erreurs notamment à cause de l'appréciation subjective des arbres qui peut être biaisée par un grand nombre de paramètres.
Concernant l'arbre d'espèces des *Firmicutes*, il serait nécessaire de vérifier la topologie et notamment le positionnement du groupe 0 des *Clostridia* qui semble mal placé d'après nos analyses. Il serait aussi intéressant d'inclure les *Mollicutes* dans l'analyse puisque, bien qu'étant classés dans un *phylum* différent, ceux-ci émergent au sein des *Firmicutes* [510].
Ensuite, les états ancestraux (de gènes et de clusters de gènes) ont été inférés par parcimonie. Nous pourrions utiliser des méthodes de vraisemblance, plus justifiées au niveau biologique et statistique, afin d'estimer au mieux les états ancestraux. Des méthodes de réconciliation par vraisemblance ont déjà été implémentées [450]. De plus, la méthode DeCoSTAR a été développée afin de manipuler de façon conjointe l'arbre d'espèces, l'arbre de gènes et les synténies pour reconstruire l'histoire des familles de gènes [106]. Le principe de DeCoSTAR est de corriger de façon itérative les réconciliations par l'information de synténie afin d'optimiser les résultats de réconciliation. L'utilisation d'une telle méthode pourrait permettre d'obtenir des réconciliations et des inférences de synténies ancestrales plus robustes.
La corrélation entre les histoires évolutives des familles de gènes a été mesurée par des coefficients de corrélation simples. Une méthode statistique plus précise a été décrite afin d'estimer la co-évolution de deux familles de gènes à partir de distributions de réconciliation [64] mais

aucun outil n'a été rendu disponible à ce jour à la communauté scientifique. Il serait intéressant d'utiliser cette approche sur nos jeux de données. Pour le calcul du score global, il serait nécessaire d'optimiser le choix des métriques prises en compte.
Il serait également intéressant de refaire l'analyse non pas pour tous les *Firmicutes* mais au niveau de certains sous-groupes, notamment pour observer d'éventuelle corrélations spécifiques de certains *taxa* comme les *Bacilli*

6.3.2 Perspectives expérimentales

Durant ma thèse, nous avons, à partir de ces méthodes, reconstruit les familles de gènes impliquées dans le cycle cellulaire chez les *Firmicutes* et leur histoire évolutive associée. Cette connaissance pourra servir au biologistes expérimentaux afin d'orienter leur recherche sur la caractérisation de protéines mais également de découvrir de nouveaux systèmes de régulation du cycle cellulaire chez certains organismes qui ne présentent aucun système décrit pour un processus cellulaire donné. Néanmoins, de nombreuses familles impliquées dans le cycle cellulaire restent encore à analyser et particulièrement impliquées dans la sporulation, la maturation du peptidoglycane, la production de la capsule et la réplication.
Nous avons aussi inféré de nombreux liens évolutifs entre les familles du cycle cellulaire. Il serait maintenant nécessaire de tester *in vivo* les liens fonctionnels inférés. Il s'agirait notamment de tester d'éventuelles interactions entre produits de gènes chez plusieurs bactéries (et notamment *S. pneumoniae* qui est le modèle de mon laboratoire) par des techniques comme le double hybride, la co-immunoprécipitation ou encore la résonance plasmonique de surface. L'ensemble des gènes détectés comme étant potentiellement impliqués dans le cycle cellulaire nécessiteraient aussi d'être testés expérimentalement, notamment en observant le phénotype cellulaire des mutants et la localisation cellulaire chez quelques bactéries. J'ai déjà initié ce travail pendant mon stage de master en caractérisant la protéine de la famille PCDP10 chez *S. pneumoniae* mais ces expériences restent encore préliminaires.
Cette connaissance sur le cycle cellulaire pourrait ensuite servir à de projets plus appliqués et

notamment au développement de molécules antibiotiques ciblant spécifiquement des protéines du cycle cellulaire. Aussi, les pipelines développés durant ma thèse pourraient être utilisés pour d'autres problématiques biologiques et l'analyse d'autres systèmes.

Bibliographie

[1] D. W. Adams, L. J. Wu, and J. Errington. Cell cycle regulation by the bacterial nucleoid. *Current Opinion in Microbiology*, 22 :94–101, dec 2014.

[2] S. G. Addinall, C. Cao, and J. Lutkenhaus. FtsN, a late recruit to the septum in *Escherichia coli*. *Molecular Microbiology*, 25(2) :303–309, 1997.

[3] M. A. Al-Hinai, S. W. Jones, and E. T. Papoutsakis. The Clostridium sporulation programs : diversity and preservation of endospore differentiation. *Microbiology and molecular biology reviews : MMBR*, 79(1) :19–37, mar 2015.

[4] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. Basic local alignment search tool. *Journal of Molecular Biology*, 215(3) :403–410, oct 1990.

[5] S. F. Altschul, T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. Gapped BLAST and PSI-BLAST : a new generation of protein database search programs. *Nucleic acids research*, 25(17) :3389–402, sep 1997.

[6] L. I. Alvarez-Anorve, M. L. Calcagno, and J. Plumbridge. Why Does *Escherichia coli* Grow More Slowly on Glucosamine than on N-Acetylglucosamine ? Effects of Enzyme Levels and Allosteric Activation of GlcN6P Deaminase (NagB) on Growth Rates. *Journal of Bacteriology*, 187(9) :2974–2982, may 2005.

[7] D. E. Anderson, F. J. Gueiros-Filho, and H. P. Erickson. Assembly dynamics of FtsZ rings in *Bacillus subtilis* and *Escherichia coli* and effects of FtsZ-regulating proteins. *Journal of bacteriology*, 186(17) :5775–81, sep 2004.

[8] D. I. Andersson, J. Jerlström-Hultqvist, and J. Näsvall. Evolution of New Functions De Novo and from Preexisting Genes. *Cold Spring Harbor Perspectives in Biology*, 7(6) :a017996, jun 2015.

[9] M. Anisimova and O. Gascuel. Approximate Likelihood-Ratio Test for Branches : A Fast, Accurate, and Powerful Alternative. *Systematic Biology*, 55(4) :539–552, aug 2006.

[10] B. P. Anton, Y.-C. Chang, P. Brown, H.-P. Choi, L. L. Faller, J. Guleria, Z. Hu, N. Klitgord, A. Levy-Moonshine, A. Maksad, V. Mazumdar, M. McGettrick, L. Osmani, R. Pokrzywa, J. Rachlin, R. Swaminathan, B. Allen, G. Housman, C. Monahan, K. Rochussen, K. Tao, A. S. Bhagwat, S. E. Brenner, L. Columbus, V. de Crécy-Lagard, D. Ferguson, A. Fomenkov, G. Gadda, R. D. Morgan, A. L. Osterman, D. A. Rodionov, I. A. Rodionova, K. E. Rudd, D. Söll, J. Spain, S.-y. Xu, A. Bateman, R. M. Blumenthal, J. M. Bollinger, W.-S. Chang, M. Ferrer, I. Friedberg, M. Y. Galperin, J. Gobeill, D. Haft, J. Hunt, P. Karp, W. Klimke, C. Krebs, D. Macelis, R. Madupu, M. J. Martin, J. H. Miller, C. O'Donovan, B. Palsson, P. Ruch, A. Setterdahl, G. Sutton, J. Tate, A. Yakunin, D. Tchigvintsev, G. Plata, J. Hu, R. Greiner, D. Horn, K. Sjölander, S. L. Salzberg, D. Vitkup, S. Letovsky, D. Segrè, C. DeLisi, R. J. Roberts, M. Steffen, and S. Kasif. The COMBREX Project : Design, Methodology, and Initial Results. *PLoS Biology*, 11(8) :e1001638, aug 2013.

[11] L. C. Antunes, D. Poppleton, A. Klingl, A. Criscuolo, B. Dupuy, C. Brochier-Armanet, C. Beloin, and S. Gribaldo. Phylogenomic analysis supports the ancestral presence of LPS-outer membranes in the *Firmicutes*. *eLife*, 5, aug 2016.

[12] S. Ardissone and P. H. Viollier. Interplay between flagellation and cell cycle control in *Caulobacter*. *Current Opinion in Microbiology*, 28 :83–92, dec 2015.

[13] A. Atrih, G. Bacher, G. Allmaier, M. P. Williamson, and S. J. Foster. Analysis of peptidoglycan structure from vegetative cells of *Bacillus subtilis* 168 and role of PBP 5 in peptidoglycan maturation. *Journal of bacteriology*, 181(13) :3956–66, jul 1999.

[14] L. Attaiech, A. Minnen, M. Kjos, S. Gruber, and J.-W. Veening. The ParB- *parS* Chromosome Segregation System Modulates Competence Development in *Streptococcus pneumoniae*. *mBio*, 6(4) :e00662–15, sep 2015.

[15] S. Aung, J. Shum, A. Abanes-De Mello, D. H. Broder, J. Fredlund-Gutierrez, S. Chiba, and K. Pogliano. Dual localization pathways for the engulfment proteins during *Bacillus subtilis* sporulation. *Molecular microbiology*, 65(6) :1534–46, sep 2007.

[16] S. J. Austin, R. J. Mural, D. K. Chattoraj, and A. L. Abeles. Trans- and cis-acting elements for the replication of P1 miniplasmids. *Journal of molecular biology*, 183(2) :195–202, may 1985.

[17] R. K. Aziz, D. Bartels, A. A. Best, M. DeJongh, T. Disz, R. A. Edwards, K. Formsma, S. Gerdes, E. M. Glass, M. Kubal, F. Meyer, G. J. Olsen, R. Olson, A. L. Osterman, R. A. Overbeek, L. K. McNeil, D. Paarmann, T. Paczian, B. Parrello, G. D. Pusch, C. Reich, R. Stevens, O. Vassieva, V. Vonstein, A. Wilke, and O. Zagnitko. The RAST Server : rapid annotations using subsystems technology. *BMC genomics*, 9 :75, feb 2008.

[18] B. Badet, P. Vermoote, P. Y. Haumont, F. Lederer, and F. LeGoffic. Glucosamine synthetase from *Escherichia coli* : purification, properties, and glutamine-utilizing site location. *Biochemistry*, 26(7) :1940–8, apr 1987.

[19] A. Badrinarayanan, T. B. K. Le, and M. T. Laub. Bacterial chromosome organization and segregation. *Annual review of cell and developmental biology*, 31 :171–99, 2015.

[20] Y. Bao, X. Zhang, Q. Jiang, T. Xue, and B. Sun. Pfs promotes autolysis-dependent release of eDNA and biofilm formation in *Staphylococcus aureus*. *Medical Microbiology and Immunology*, 204(2) :215–226, apr 2015.

[21] D. Barker, A. Meade, and M. Pagel. Constrained models of evolution lead to improved prediction of functional linkage from correlated gain and loss of genes. *Bioinformatics*, 23(1) :14–20, jan 2007.

[22] D. Barker and M. Pagel. Predicting Functional Gene Links from Phylogenetic-Statistical Analyses of Whole Genomes. *PLoS Computational Biology*, 1(1) :e3, 2005.

[23] G. Baronian, K. Ginda, L. Berry, M. Cohen-Gonsaud, J. Zakrzewska-Czerwińska, D. Jakimowicz, and V. Molle. Phosphorylation of *Mycobacterium tuberculosis* ParB participates in regulating the ParABS chromosome segregation system. *PloS one*, 10(3) :e0119907, 2015.

[24] H. Barreteau, A. Kovač, A. Boniface, M. Sova, S. Gobec, and D. Blanot. Cytoplasmic steps of peptidoglycan biosynthesis. *FEMS Microbiology Reviews*, 32(2) :168–207, mar 2008.

[25] N. H. Barton, D. E. Briggs, J. A. Eisen, D. B. Goldstein, and N. H. Patel. *Evolution*. Cold Spring Harbor Laboratory Press, 2007.

[26] B. Beall and J. Lutkenhaus. Impaired cell division and sporulation of a *Bacillus subtilis* strain with the ftsA gene deleted. *Journal of bacteriology*, 174(7) :2398–403, apr 1992.

[27] R. G. Beiko, T. J. Harlow, and M. A. Ragan. Highways of gene sharing in prokaryotes. *Proceedings of the National Academy of Sciences*, 102(40) :14332–14337, oct 2005.

[28] K. Beilharz, L. Nováková, D. Fadda, P. Branny, O. Massidda, and J.-W. Veening. Control of cell division in *Streptococcus pneumoniae* by the conserved Ser/Thr protein kinase StkP. *Proceedings of the National Academy of Sciences of the United States of America*, 109(15) :E905–13, apr 2012.

[29] S. Bellais, M. Arthur, L. Dubost, J.-E. Hugonnet, L. Gutmann, J. van Heijenoort, R. Legrand, J.-P. Brouard, L. Rice, and J.-L. Mainardi. Aslfm, the D-Aspartate Ligase Responsible for the Addition of D-Aspartic Acid onto the Peptidoglycan Precursor of *Enterococcus faecium*. *Journal of Biological Chemistry*, 281(17) :11586–11594, apr 2006.

[30] F. O. Bendezú, C. A. Hale, T. G. Bernhardt, and P. A. J. de Boer. RodZ (YfgA) is required for proper assembly of the MreB actin cytoskeleton and cell shape in *E. coli*. *The EMBO Journal*, 28(3) :193–204, feb 2009.

[31] D. A. Benson, M. Cavanaugh, K. Clark, I. Karsch-Mizrachi, J. Ostell, K. D. Pruitt, and E. W. Sayers. GenBank. *Nucleic Acids Research*, 46(D1) :D41–D47, jan 2018.

[32] T. E. Benson, J. L. Marquardt, A. C. Marquardt, F. A. Etzkorn, and C. T. Walsh. Overexpression, purification, and mechanistic study of UDP-N-acetylenolpyruvylglucosamine reductase. *Biochemistry*, 32(8) :2024–30, mar 1993.

[33] T. G. Bernhardt and P. A. J. de Boer. SlmA, a nucleoid-associated, FtsZ binding protein required for blocking septal ring assembly over Chromosomes in *E. coli*. *Molecular cell*, 18(5) :555–64, may 2005.

[34] E. Bi and J. Lutkenhaus. Cell division inhibitors SulA and MinCD prevent formation of the FtsZ ring. *Journal of Bacteriology*, 175(4) :1118–1125, 1993.

[35] S. Bigot, O. A. Saleh, C. Lesterlin, C. Pages, M. El Karoui, C. Dennis, M. Grigoriev, J.-F. Allemand, F.-X. Barre, and F. Cornet. KOPS : DNA motifs that control *E. coli* chromosome segregation by orienting the FtsK translocase. *The EMBO journal*, 24(21) :3770–80, nov 2005.

[36] S. Bigot, V. Sivanathan, C. Possoz, F.-X. Barre, and F. Cornet. FtsK, a literate chromosome segregation machine. *Molecular microbiology*, 64(6) :1434–41, jun 2007.

[37] S. J. Biller and W. F. Burkholder. The *Bacillus subtilis* SftA (YtpS) and SpoIIIE DNA translocases play distinct roles in growing cells to ensure faithful chromosome partitioning. *Molecular microbiology*, 74(4) :790–809, nov 2009.

[38] P. Bisicchia, D. Noone, E. Lioliou, A. Howell, S. Quigley, T. Jensen, H. Jarmer, and K. M. Devine. The essential YycFG two-component system controls cell wall metabolism in *Bacillus subtilis*. *Molecular Microbiology*, 65(1) :180–200, jul 2007.

[39] A. W. Bisson-Filho, Y.-P. Hsu, G. R. Squyres, E. Kuru, F. Wu, C. Jukes, Y. Sun, C. Dekker, S. Holden, M. S. VanNieuwenhze, Y. V. Brun, and E. C. Garner. Treadmilling by FtsZ filaments drives peptidoglycan synthesis and bacterial cell division. *Science (New York, N.Y.)*, 355(6326) :739–743, 2017.

[40] P. Blohm, G. Frishman, P. Smialowski, F. Goebels, B. Wachinger, A. Ruepp, and D. Frishman. Negatome 2.0 : a database of non-interacting proteins derived by literature mining, manual annotation and protein structure analysis. *Nucleic Acids Research*, 42(D1) :D396–D400, jan 2014.

[41] P. M. Blumberg and J. L. Strominger. Covalent affinity chromatography of penicillin-binding components from bacterial membranes. *Methods in enzymology*, 34 :401–5, 1974.

[42] D. Boothby, L. Daneo-Moore, M. L. Higgins, J. Coyette, and G. D. Shockman. Turnover of bacterial cell wall peptidoglycans. *The Journal of biological chemistry*, 248(6) :2161–9, mar 1973.

[43] M. Borisova, J. Gisin, and C. Mayer. The N-Acetylmuramic Acid 6-Phosphate Phosphatase MupP Completes the *Pseudomonas* Peptidoglycan Recycling Pathway Leading to Intrinsic Fosfomycin Resistance. *Mbio*, 8(2) :e00092–17, 2017.

[44] B. Boussau and V. Daubin. Genomes as documents of evolutionary history. *Trends in Ecology & Evolution*, 25(4) :224–232, apr 2010.

[45] E. Boutet, D. Lieberherr, M. Tognolli, M. Schneider, and A. Bairoch. UniProtKB/Swiss-Prot. *Methods in molecular biology (Clifton, N.J.)*, 406 :89–112, 2007.

[46] P. M. Bowers, S. J. Cokus, D. Eisenberg, and T. O. Yeates. Use of Logic Relationships to Decipher Protein Network Organization. *Science*, 306(5705) :2246–2249, dec 2004.

[47] D. Bramhill and C. M. Thompson. GTP-dependent polymerization of *Escherichia coli* FtsZ protein to form tubules. *Proceedings of the National Academy of Sciences of the United States of America*, 91(13) :5813–7, jun 1994.

[48] M. Bramkamp, R. Emmins, L. Weston, C. Donovan, R. A. Daniel, and J. Errington. A novel component of the division-site selection system of *Bacillus subtilis* and a new mode of action for the division inhibitor MinCD. *Molecular Microbiology*, 70(6) :1556–1569, dec 2008.

[49] H. Brinkmann and H. Philippe. Archaea sister group of Bacteria ? Indications from tree reconstruction artifacts in ancient phylogenies. *Molecular Biology and Evolution*, 16(6) :817–825, jun 1999.

[50] R. A. Britton, D. C. Lin, and A. D. Grossman. Characterization of a prokaryotic SMC protein involved in chromosome partitioning. *Genes & development*, 12(9) :1254–9, may 1998.

[51] D. H. Broder and K. Pogliano. Forespore engulfment mediated by a ratchet-like mechanism. *Cell*, 126(5) :917–28, sep 2006.

[52] E. D. Brown, E. I. Vivas, C. T. Walsh, and R. Kolter. MurA (MurZ), the enzyme that catalyzes the first committed step in peptidoglycan biosynthesis, is essential in *Escherichia coli*. *Journal of bacteriology*, 177(14) :4194–7, jul 1995.

[53] J. A. Bryant, L. E. Sellars, S. J. W. Busby, and D. J. Lee. Chromosome position effects on gene expression in *Escherichia coli* K-12. *Nucleic Acids Research*, 42(18) :11383–11392, oct 2014.

[54] Y. X. Butler, Y. Abhayawardhane, and G. C. Stewart. Amplification of the *Bacillus subtilis* maf gene results in arrested septum formation. *Journal of Bacteriology*, 175(10) :3139–3145, 1993.

[55] T. Caldas, E. Binet, P. Bouloc, A. Costa, J. Desgres, and G. Richarme. The FtsJ/RrmJ Heat Shock Protein of *Escherichia coli* is a 23 S Ribosomal RNA Methyltransferase. *Journal of Biological Chemistry*, 275(22) :16414–16419, jun 2000.

[56] L. Calvanese, L. Falcigno, C. Maglione, D. Marasco, A. Ruggiero, F. Squeglia, R. Berisio, and G. D'Auria. Structural and binding properties of the PASTA domain of PonA2, a key penicillin binding protein from *Mycobacterium tuberculosis*. *Biopolymers*, 101(7) :712–9, jul 2014.

[57] J. E. Camara, K. Skarstad, and E. Crooke. Controlled initiation of chromosomal replication in *Escherichia coli* requires functional Hda protein. *Journal of bacteriology*, 185(10) :3244–8, may 2003.

[58] J. L. Camberg, J. R. Hoskins, and S. Wickner. ClpXP protease degrades the cytoskeletal protein, FtsZ, and modulates FtsZ polymer dynamics. *Proceedings of the National Academy of Sciences of the United States of America*, 106(26) :10614–9, jun 2009.

[59] R. Carballido-López, A. Formstone, Y. Li, S. D. Ehrlich, P. Noirot, and J. Errington. Actin Homolog MreBH Governs Cell Morphogenesis by Localization of the Cell Wall Hydrolase LytE. *Developmental Cell*, 11(3) :399–409, 2006.

[60] M. Carrión, M. J. Gómez, R. Merchante-Schubert, S. Dongarrá, and J. A. Ayala. mraW, an essential gene at the dcw cluster of *Escherichia coli* codes for a cytoplasmic protein with methyltransferase activity. *Biochimie*, 81(8-9) :879–888, 1999.

[61] F. Castillo, A. Benmohamed, and G. Szatmari. Xer Site Specific Recombination : Double and Single Recombinase Systems. *Frontiers in microbiology*, 8 :453, 2017.

[62] J. Castresana. Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. *Molecular Biology and Evolution*, 17(4) :540–552, apr 2000.

[63] J. H. Cha and G. C. Stewart. The divIVA minicell locus of *Bacillus subtilis*. *Journal of bacteriology*, 179(5) :1671–83, mar 1997.

[64] Y.-b. Chan, V. Ranwez, and C. Scornavacca. Reconciliation-based detection of co-evolving gene families. *BMC bioinformatics*, 14 :332, nov 2013.

[65] R. Chenna, H. Sugawara, T. Koike, R. Lopez, T. J. Gibson, D. G. Higgins, and J. D. Thompson. Multiple sequence alignment with the Clustal series of programs. *Nucleic acids research*, 31(13) :3497–500, jul 2003.

[66] E. Cho, N. Ogasawara, and S. Ishikawa. The functional analysis of YabA, which interacts with DnaA and regulates initiation of chromosome replication in Bacillus subtils. *Genes & genetic systems*, 83(2) :111–25, apr 2008.

[67] K.-M. Chung, H.-H. Hsu, H.-Y. Yeh, and B.-Y. Chang. Mechanism of regulation of prokaryotic tubulin-like GTPase FtsZ by membrane protein EzrA. *The Journal of biological chemistry*, 282(20) :14891–7, may 2007.

[68] D. Claessen, R. Emmins, L. W. Hamoen, R. A. Daniel, J. Errington, and D. H. Edwards. Control of the cell elongation–division cycle by shuttling of PBP1 protein in *Bacillus subtilis*. *Molecular Microbiology*, 68(4) :1029–1046, may 2008.

[69] S. Coles. *An introduction to statistical modeling of extreme values*. 2001.

[70] B. D. Corbin, Y. Wang, T. K. Beuria, and W. Margolin. Interaction between Cell Division Proteins FtsE and FtsZ. *Journal of Bacteriology*, 189(8) :3026–3035, apr 2007.

[71] F. Cornet, B. Hallet, and D. J. Sherratt. Xer recombination in *Escherichia coli*. Site-specific DNA topoisomerase activity of the XerC and XerD recombinases. *The Journal of biological chemistry*, 272(35) :21927–31, aug 1997.

[72] A. Criscuolo and S. Gribaldo. BMGE (Block Mapping and Gathering with Entropy) : a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evolutionary Biology*, 10(1) :210, jul 2010.

[73] P. Dagnelie. *Théorie et méthodes statistiques : applications agronomiques*. Les Presses agronomiques de Gembloux, 1970.

[74] K. Dai, Y. Xu, and J. Lutkenhaus. Cloning and characterization of ftsN, an essential cell division gene in *Escherichia coli* isolated as a multicopy suppressor of ftsA12(Ts). *Journal of bacteriology*, 175(12) :3790–7, jun 1993.

[75] A. Dajkovic, G. Lan, S. X. Sun, D. Wirtz, and J. Lutkenhaus. MinC Spatially Controls Bacterial Cytokinesis by Antagonizing the Scaffolding Function of FtsZ. *Current Biology*, 18(4) :235–244, feb 2008.

[76] C. C. Dang, V. Lefort, V. S. Le, Q. S. Le, and O. Gascuel. ReplacementMatrix : a web server for maximum-likelihood estimation of amino acid replacement rate matrices. *Bioinformatics*, 27(19) :2758–2760, oct 2011.

[77] R. A. Daniel, S. Drake, C. E. Buchanan, R. Scholle, and J. Errington. The *Bacillus subtilis* spoVD gene encodes a mother-cell-specific penicillin-binding protein required for spore morphogenesis. *Journal of molecular biology*, 235(1) :209–20, jan 1994.

[78] O. Danilova, R. Reyes-Lamothe, M. Pinskaya, D. Sherratt, and C. Possoz. MukB colocalizes with the oriC region and is required for organization of the two *Escherichia coli* chromosome arms into separate cell halves. *Molecular microbiology*, 65(6) :1485–92, sep 2007.

[79] C. Darwin. *The Origin of Species by means of Natural Selection*. 1859.

[80] S. V. Date. The rosetta stone method. *Methods in molecular biology (Clifton, N.J.).*, 2008.

[81] S. V. Date and E. M. Marcotte. Protein function prediction using the Protein Link EXplorer (PLEX). *Bioinformatics*, 21(10) :2558–2559, may 2005.

[82] P. Datta, A. Dasgupta, A. K. Singh, P. Mukherjee, M. Kundu, and J. Basu. Interaction between FtsW and penicillin-binding protein 3 (PBP3) directs PBP3 to mid-cell, controls cell septation and mediates the formation of a trimeric complex involving FtsZ, FtsW and PBP3 in mycobacteria. *Molecular microbiology*, 62(6) :1655–73, dec 2006.

[83] C. Davies, S. W. White, and R. A. Nicholas. Crystal structure of a deacylation-defective mutant of penicillin-binding protein 5 at 2.3-A resolution. *The Journal of biological chemistry*, 276(1) :616–23, jan 2001.

[84] D. A. Dawson, M. Akesson, T. Burke, J. M. Pemberton, J. Slate, and B. Hansson. Gene Order and Recombination Rate in Homologous Chromosome Regions of the Chicken and a Passerine Bird. *Molecular Biology and Evolution*, 24(7) :1537–1552, mar 2007.

[85] P. A. J. de Boer. Classic Spotlight : Discovery of ftsZ. *Journal of bacteriology*, 198(8) :1184, apr 2016.

[86] M. J. L. de Hoon, P. Eichenberger, and D. Vitkup. Hierarchical evolution of the bacterial sporulation network. *Current biology : CB*, 20(17) :R735–45, sep 2010.

[87] E. de Leeuw, B. Graham, G. J. Phillips, C. M. ten Hagen-Jongman, B. Oudega, and J. Luirink. Molecular characterization of *Escherichia coli* FtsE and FtsX. *Molecular microbiology*, 31(3) :983–93, feb 1999.

[88] P. De Vos, G. Garrity, D. Jones, N. R. Krieg, W. Ludwig, F. A. Rainey, K.-H. Schleifer, and W. B. Whitman. *Bergey's manual of systematic bacteriology : Volume 3*. Williams & Wilkins, 2009.

[89] A. Delauné, S. Dubrac, C. Blanchet, O. Poupel, U. Mäder, A. Hiron, A. Leduc, C. Fitting, P. Nicolas, J.-M. Cavaillon, M. Adib-Conquy, and T. Msadek. The WalKR System Controls Major Staphylococcal Virulence Genes and Is Involved in Triggering the Host Inflammatory Response. *Infection and Immunity*, 80(10) :3438–3453, oct 2012.

[90] A. L. Delcher, D. Harmon, S. Kasif, O. White, and S. L. Salzberg. Improved microbial gene identification with GLIMMER. *Nucleic acids research*, 27(23) :4636–41, dec 1999.

[91] A. Derouaux, B. Wolf, C. Fraipont, E. Breukink, M. Nguyen-Disteche, and M. Terrak. The Monofunctional Glycosyltransferase of *Escherichia coli* Localizes to the Cell Division Site and Interacts with Penicillin-Binding Protein 3, FtsW, and FtsN. *Journal of Bacteriology*, 190(5) :1831–1834, mar 2008.

[92] R. Derouiche, H. Bénédetti, J. C. Lazzaroni, C. Lazdunski, and R. Lloubès. Protein complex within *Escherichia coli* inner membrane. TolA N-terminal domain interacts with TolQ and TolR proteins. *The Journal of biological chemistry*, 270(19) :11078–84, may 1995.

[93] C. N. Dewey. Positional orthology : putting genomic evolutionary relationships into context. *Briefings in bioinformatics*, 12(5) :401–12, sep 2011.

[94] V. Diez, G. E. Schujman, F. J. Gueiros-Filho, and D. de Mendoza. Vectorial signalling mechanism required for cell-cell communication during sporulation in *Bacillus subtilis*. *Molecular microbiology*, 83(2) :261–74, jan 2012.

[95] S. C. Dillon and C. J. Dorman. Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. *Nature reviews. Microbiology*, 8(3) :185–95, mar 2010.

[96] A. V. Divakaruni, C. Baida, C. L. White, and J. W. Gober. The cell shape proteins MreB and MreC control cell morphogenesis by positioning cell wall synthetic complexes. *Molecular Microbiology*, 66(1) :174–188, 2007.

[97] T. Domazet-Lošo, J. Brajković, and D. Tautz. A phylostratigraphy approach to uncover the genomic history of major adaptations in metazoan lineages. *Trends in Genetics*, 23(11) :533–539, nov 2007.

[98] J. Dominguez-Escobar, A. Chastanet, A. H. Crevenna, V. Fromion, R. Wedlich-Soldner, and R. Carballido-Lopez. Processive Movement of MreB-Associated Cell Wall Biosynthetic Complexes in Bacteria. *Science*, 333(6039) :225–228, jul 2011.

[99] P. Doublet, J. van Heijenoort, J. P. Bohin, and D. Mengin-Lecreulx. The murI gene of *Escherichia coli* is an essential gene that encodes a glutamate racemase activity. *Journal of bacteriology*, 175(10) :2970–9, may 1993.

[100] J.-P. Doyon, V. Ranwez, V. Daubin, and V. Berry. Models, algorithms and programs for phylogeny reconciliation. *Briefings in Bioinformatics*, 12(5) :392–400, sep 2011.

[101] J. Drummelsmith and C. Whitfield. Gene products required for surface expression of the capsular form of the group 1 K antigen in *Escherichia coli* (O9a :K30). *Molecular microbiology*, 31(5) :1321–32, mar 1999.

[102] J. Drummelsmith and C. Whitfield. Translocation of group 1 capsular polysaccharide to the surface of *Escherichia coli* requires a multimeric complex in the outer membrane. *The EMBO Journal*, 19(1) :57–66, jan 2000.

[103] A. J. Drummond, S. Y. W. Ho, M. J. Phillips, and A. Rambaut. Relaxed Phylogenetics and Dating with Confidence. *PLoS Biology*, 4(5) :e88, mar 2006.

[104] W. Du, J. R. Brown, D. R. Sylvester, J. Huang, A. F. Chalker, C. Y. So, D. J. Holmes, D. J. Payne, and N. G. Wallis. Two active forms of UDP-N-acetylglucosamine enolpyruvyl transferase in gram-positive bacteria. *Journal of bacteriology*, 182(15) :4146–52, aug 2000.

[105] S. Dubrac, I. G. Boneca, O. Poupel, and T. Msadek. New insights into the WalK/WalR (YycG/YycF) essential signal transduction pathway reveal a major role in controlling cell wall metabolism and biofilm formation in *Staphylococcus aureus*. *Journal of bacteriology*, 189(22) :8257–69, nov 2007.

[106] W. Duchemin, Y. Anselmetti, M. Patterson, Y. Ponty, S. Bï¿½rard, C. Chauve, C. Scornavacca, V. Daubin, and E. Tannier. DeCoSTAR : Reconstructing the Ancestral Organization of Genes or Genomes Using Reconciled Phylogenies. *Genome Biology and Evolution*, 9(5) :1312–1319, may 2017.

[107] W. Duchemin, G. Gence, A.-M. Arigon Chifolleau, L. Arvestad, M. S. Bansal, V. Berry, B. Boussau, F. Chevenet, N. Comte, A. A. Davín, C. Dessimoz, D. Dylus, D. Hasic, D. Mallo, R. Planel, D. Posada, C. Scornavacca, G. Szöllősi, L. Zhang, É. Tannier, and V. Daubin. RecPhyloXML : a format for reconciled gene trees. *Bioinformatics*, may 2018.

[108] R. Duman, S. Ishikawa, I. Celik, H. Strahl, N. Ogasawara, P. Troc, J. Lowe, and L. W. Hamoen. Structural and genetic analyses reveal the protein SepF as a new membrane anchor for the Z ring. *Proceedings of the National Academy of Sciences*, 110(48) :E4601–E4610, nov 2013.

[109] K. Duncan, J. van Heijenoort, and C. T. Walsh. Purification and characterization of the D-alanyl-D-alanine-adding enzyme from *Escherichia coli*. *Biochemistry*, 29(9) :2379–86, mar 1990.

[110] L. Duncan, S. Alper, F. Arigoni, R. Losick, and P. Stragier. Activation of cell-specific transcription by a serine phosphatase at the site of asymmetric division. *Science (New York, N.Y.)*, 270(5236) :641–4, oct 1995.

[111] J. Durand-Heredia, E. Rivkin, G. Fan, J. Morales, and A. Janakiraman. Identification of ZapD as a cell division factor that promotes the assembly of FtsZ in *Escherichia coli*. *Journal of bacteriology*, 194(12) :3189–98, jun 2012.

[112] J. M. Durand-Heredia, H. H. Yu, S. De Carlo, C. F. Lesser, and A. Janakiraman. Identification and Characterization of ZapC, a Stabilizer of the FtsZ Ring in *Escherichia coli*. *Journal of Bacteriology*, 193(6) :1405–1413, mar 2011.

[113] J. Dworkin. Protein Targeting during *Bacillus subtilis* Sporulation. In *The Bacterial Spore : from Molecules to Systems*, volume 2, pages 145–156. American Society of Microbiology, feb 2014.

[114] N. A. Dye, Z. Pincus, J. A. Theriot, L. Shapiro, and Z. Gitai. Two independent spiral structures control cell shape in *Caulobacter*. *Proceedings of the National Academy of Sciences of the United States of America*, 102(51) :18608–13, dec 2005.

[115] A. Eberhardt, C. N. Hoyland, D. Vollmer, S. Bisle, R. M. Cleverley, O. Johnsborg, L. S. Håvarstein, R. J. Lewis, and W. Vollmer. Attachment of Capsular Polysaccharide to the Cell Wall in *Streptococcus pneumoniae*. *Microbial Drug Resistance*, 18(3) :240–255, jun 2012.

[116] G. Ebersbach, E. Galli, J. Møller-Jensen, J. Löwe, and K. Gerdes. Novel coiled-coil cell division factor ZapB stimulates Z ring assembly and cell division. *Molecular Microbiology*, 68(3) :720–735, may 2008.

[117] S. R. Eddy. Profile hidden Markov models. *Bioinformatics (Oxford, England)*, 14(9) :755–63, 1998.

[118] S. R. Eddy. Where did the BLOSUM62 alignment score matrix come from ? *Nature biotechnology*, 22(8) :1035–6, aug 2004.

[119] S. R. Eddy. A new generation of homology search tools based on probabilstic inference. In *Genome Informatics 2009*, pages 205–211, oct 2009.

[120] D. H. Edwards and J. Errington. The *Bacillus subtilis* DivIVA protein targets to the division septum and controls the site specificity of cell division. *Molecular microbiology*, 24(5) :905–15, jun 1997.

[121] A. J. F. Egan, J. Biboy, I. van't Veer, E. Breukink, and W. Vollmer. Activities and regulation of peptidoglycan synthases. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 370(1679), oct 2015.

[122] P. Eichenberger, P. Fawcett, and R. Losick. A three-protein inhibitor of polar septation during sporulation in *Bacillus subtilis*. *Molecular microbiology*, 42(5) :1147–62, dec 2001.

[123] J. A. Eisen and C. M. Fraser. Phylogenomics : intersection of evolution and genomics. *Science (New York, N.Y.).*, Jun 2003.

[124] K. Emami, A. Guyet, Y. Kawai, J. Devi, L. J. Wu, N. Allenby, R. A. Daniel, and J. Errington. RodA as the missing glycosyltransferase in *Bacillus subtilis* and antibiotic discovery for the peptidoglycan polymerase pathway. *Nature microbiology*, 2 :16253, jan 2017.

[125] J. M. Eraso, L. M. Markillie, H. D. Mitchell, R. C. Taylor, G. Orr, and W. Margolin. The highly conserved MraZ protein is a transcriptional regulator in *Escherichia coli*. *Journal of Bacteriology*, 196(11) :2053–2066, 2014.

[126] H. P. Erickson. The FtsZ protofilament and attachment of ZipA–structural constraints on the FtsZ power stroke. *Current opinion in cell biology*, 13(1) :55–60, feb 2001.

[127] H. P. Erickson, D. E. Anderson, and M. Osawa. FtsZ in Bacterial Cytokinesis : Cytoskeleton and Force Generator All in One. *Microbiology and Molecular Biology Reviews*, 74(4) :504–528, dec 2010.

[128] H. P. Erickson and D. Stoffler. Protofilaments and rings, two conformations of the tubulin family conserved from bacterial FtsZ to alpha/beta and gamma tubulin. *The Journal of cell biology*, 135(1) :5–8, oct 1996.

[129] J. Errington. Bacterial morphogenesis and the enigmatic MreB helix. *Nature Reviews Microbiology*, 13(4) :241–248, apr 2015.

[130] J. Errington, L. Appleby, R. A. Daniel, H. Goodfellow, S. R. Partridge, and M. D. Yudkin. Structure and function of the spoIIIJ gene of *Bacillus subtilis* : a vegetatively expressed gene that is essential for sigma G activity at an intermediate stage of sporulation. *Journal of general microbiology*, 138(12) :2609–18, dec 1992.

[131] J. Errington, H. Murray, and L. J. Wu. Diversity and redundancy in bacterial chromosome segregation mechanisms. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 360(1455) :497–505, mar 2005.

[132] O. Espéli, R. Borne, P. Dupaigne, A. Thiel, E. Gigant, R. Mercier, and F. Boccard. A MatP-divisome interaction coordinates chromosome segregation with cell division in *E. coli*. *The EMBO journal*, 31(14) :3198–211, may 2012.

[133] O. Espeli, C. Levine, H. Hassing, and K. J. Marians. Temporal regulation of topoisomerase IV activity in *E. coli*. *Molecular cell*, 11(1) :189–201, jan 2003.

[134] D. Fadda, C. Pischedda, F. Caldara, M. B. Whalen, D. Anderluzzi, E. Domenici, and O. Massidda. Characterization of divIVA and other genes located in the chromosomal region downstream of the dcw cluster in *Streptococcus pneumoniae*. *Journal of bacteriology*, 185(20) :6209–14, oct 2003.

[135] A. Fay and J. Dworkin. Bacillus subtilis homologs of MviN (MurJ), the putative *Escherichia coli* lipid II flippase, are not essential for growth. *Journal of bacteriology*, 191(19) :6020–8, oct 2009.

[136] A. Fay, P. Meyer, and J. Dworkin. Interactions Between Late-Acting Proteins Required for Peptidoglycan Synthesis during Sporulation. *Journal of Molecular Biology*, 399(4) :547–561, jun 2010.

[137] S. Federhen. The NCBI Taxonomy database. *Nucleic acids research*, 40(Database issue) :D136–43, jan 2012.

[138] J. Felsenstein. CONFIDENCE LIMITS ON PHYLOGENIES : AN APPROACH USING THE BOOTSTRAP. *Evolution*, 39(4) :783–791, jul 1985.

[139] L. Feng, K. Sheppard, D. Tumbula-Hansen, and D. Söll. Gln-tRNAGln Formation from Glu-tRNAGln Requires Cooperation of an Asparaginase and a Glu-tRNAGln Kinase. *Journal of Biological Chemistry*, 280(9) :8150–8155, mar 2005.

[140] A. K. Fenton and K. Gerdes. Direct interaction of FtsZ and MreB is required for septum synthesis and cell division in *Escherichia coli*. *The EMBO Journal*, 32(13) :1953–1965, jun 2013.

[141] A. K. Fenton, S. Manuse, J. Flores-Kim, P. S. Garcia, C. Mercy, C. Grangeasse, T. G. Bernhardt, and D. Z. Rudner. Phosphorylation-dependent activation of the cell wall synthase PBP2a in *Streptococcus pneumoniae* by MacP. *Proceedings of the National Academy of Sciences*, 115(11) :2812–2817, mar 2018.

[142] A. K. Fenton, L. E. Mortaji, D. T. C. Lau, D. Z. Rudner, and T. G. Bernhardt. CozE is a member of the MreCD complex that directs cell elongation in *Streptococcus pneumoniae*. *Nature Microbiology*, 2 :16237, dec 2016.

[143] R. M. Figge, A. V. Divakaruni, and J. W. Gober. MreB, the cell shape-determining bacterial actin homologue, co-ordinates cell wall morphogenesis in *Caulobacter crescentus*. *Molecular microbiology*, 51(5) :1321–32, mar 2004.

[144] S. R. Filipe, M. G. Pinho, and A. Tomasz. Characterization of the murMN Operon Involved in the Synthesis of Branched Peptidoglycan Peptides in *Streptococcus pneumoniae*. *Journal of Biological Chemistry*, 275(36) :27768–74, jun 2000.

[145] S. R. Filipe, E. Severina, and A. Tomasz. Distribution of the mosaic structured murM genes among natural populations of *Streptococcus pneumoniae*. *Journal of bacteriology*, 182(23) :6798–805, dec 2000.

[146] R. D. Finn, A. Bateman, J. Clements, P. Coggill, R. Y. Eberhardt, S. R. Eddy, A. Heger, K. Hetherington, L. Holm, J. Mistry, E. L. L. Sonnhammer, J. Tate, and M. Punta. Pfam : the protein families database. *Nucleic Acids Research*, 42(D1) :D222–D230, jan 2014.

[147] R. D. Fleischmann, M. D. Adams, O. White, R. A. Clayton, E. F. Kirkness, A. R. Kerlavage, C. J. Bult, J. F. Tomb, B. A. Dougherty, J. M. Merrick, and et al. Whole-genome random sequencing and assembly of haemophilus influenzae rd. *Science (New York, N.Y.).*, Jul 1995.

[148] A. Fleurie, C. Cluzel, S. Guiral, C. Freton, F. Galisson, I. Zanella-Cleon, A.-M. Di Guilmi, and C. Grangeasse. Mutational dissection of the S/T-kinase StkP reveals crucial roles in cell division of *Streptococcus pneumoniae*. *Molecular Microbiology*, 83(4) :746–758, feb 2012.

[149] A. Fleurie, C. Lesterlin, S. Manuse, C. Zhao, C. Cluzel, J. P. Lavergne, M. Franz-Wachtel, B. MacEk, C. Combet, E. Kuru, M. S. VanNieuwenhze, Y. V. Brun, D. Sherratt, and C. Grangeasse. MapZ marks the division sites and positions FtsZ rings in *Streptococcus pneumoniae*. *Nature*, 516(7530) :259–262, 2014.

[150] A. Fleurie, S. Manuse, C. Zhao, N. Campo, C. Cluzel, J.-P. Lavergne, C. Freton, C. Combet, S. Guiral, B. Soufi, B. Macek, E. Kuru, M. S. VanNieuwenhze, Y. V. Brun, A.-M. Di Guilmi, J.-P. Claverys, A. Galinier, and C. Grangeasse. Interplay of the serine/threonine-kinase StkP and the paralogs DivIVA and GpsB in pneumococcal cell elongation and division. *PLoS genetics*, 10(4) :e1004275, apr 2014.

[151] M. A. Fogel and M. K. Waldor. A dynamic, mitotic-like mechanism for bacterial chromosome segregation. *Genes & development*, 20(23) :3269–82, dec 2006.

[152] E. F. E. F. Gale. *The Molecular basis of antibiotic action*. J. Wiley, 1981.

[153] E. Galli and K. Gerdes. FtsZ-ZapA-ZapB interactome of *Escherichia coli*. *Journal of Bacteriology*, 194(2) :292–302, 2012.

[154] P. Gamba, L. W. Hamoen, and R. A. Daniel. Cooperative Recruitment of FtsW to the Division Site of *Bacillus subtilis*. *Frontiers in Microbiology*, 7 :1808, nov 2016.

[155] T. Gao, M. Tan, W. Liu, C. Zhang, T. Zhang, L. Zheng, J. Zhu, L. Li, and R. Zhou. GidA, a tRNA Modification Enzyme, Contributes to the Growth, and Virulence of *Streptococcus* suis Serotype 2. *Frontiers in Cellular and Infection Microbiology*, 6(44), 2016.

[156] P. Garcia, M. P. Gonzalez, E. Garcia, R. Lopez, and J. L. Garcia. Lytb, a novel pneumococcal murein hydrolase essential for cell separation. *Molecular microbiology.*, Feb 1999.

[157] P. S. Garcia, F. Jauffrit, C. Grangeasse, and C. Brochier-Armanet. GeneSpy, a user-friendly and flexible genomic context visualizer. *Bioinformatics*, jun 2018.

[158] P. S. Garcia, J. P. Simorre, C. Brochier-Armanet, and C. Grangeasse. Cell division of *Streptococcus pneumoniae* : think positive ! *Current Opinion in Microbiology*, 34 :18–23, 2016.

[159] E. C. Garner, R. Bernard, W. Wang, X. Zhuang, D. Z. Rudner, and T. Mitchison. Coupled, Circumferential Motions of the Cell Wall Synthesis Machinery and MreB Filaments in *B. subtilis*. *Science*, 333(6039) :222–225, jul 2011.

[160] G. M. Garrity and J. G. Holt. The Road Map to the Manual. In *Bergey's Manual® of Systematic Bacteriology*, pages 119–166. Springer New York, New York, NY, 2001.

[161] B. Geissler and W. Margolin. Evidence for functional overlap among multiple bacterial cell division proteins : compensating for the loss of FtsK. *Molecular microbiology*, 58(2) :596–612, oct 2005.

[162] K. A. Geno, G. L. Gilbert, J. Y. Song, I. C. Skovsted, K. P. Klugman, C. Jones, H. B. Konradsen, and M. H. Nahm. Pneumococcal Capsules and Their Types : Past, Present, and Future. *Clinical microbiology reviews*, 28(3) :871–99, jul 2015.

[163] M. A. Gerding, Y. Ogata, N. D. Pecora, H. Niki, and P. A. J. de Boer. The trans-envelope Tol-Pal complex is part of the cell division machinery and required for proper outer-membrane invagination during cell constriction in *E. coli*. *Molecular microbiology*, 63(4) :1008–25, feb 2007.

[164] J. M. Ghigo, D. S. Weiss, J. C. Chen, J. C. Yarrow, and J. Beckwith. Localization of FtsL to the *Escherichia coli* septal ring. *Molecular microbiology*, 31(2) :725–37, jan 1999.

[165] N. E. Gibbons and R. G. E. Murray. Proposals Concerning the Higher Taxa of Bacteria. *International Journal of Systematic Bacteriology*, 28(1) :1–6, jan 1978.

[166] C. Giefing, K. E. Jelencsics, D. Gelbmann, B. M. Senn, and E. Nagy. The pneumococcal eukaryotic-type serine/threonine protein kinase StkP co-localizes with the cell division apparatus and interacts with FtsZ in vitro. *Microbiology*, 156(6) :1697–1707, jun 2010.

[167] J. Gisin, A. Schneider, B. Nägele, M. Borisova, and C. Mayer. A cell wall recycling shortcut that bypasses peptidoglycan de novo biosynthesis. *Nature Chemical Biology*, 9(8) :491–493, aug 2013.

[168] G. V. Glazko and A. R. Mushegian. Detection of evolutionarily stable fragments of cellular pathways by hierarchical clustering of phyletic patterns. *Genome Biology 2004 5 :5*, 5(5) :R32, apr 2004.

[169] N. W. Goehring, M. D. Gonzalez, and J. Beckwith. Premature targeting of cell division proteins to midcell reveals hierarchies of protein interactions involved in divisome assembly. *Molecular Microbiology*, 61(1) :33–45, jul 2006.

[170] N. W. Goehring, C. Robichon, and J. Beckwith. Role for the nonessential N terminus of FtsN in divisome assembly. *Journal of bacteriology*, 189(2) :646–9, jan 2007.

[171] A. Goffeau, B. G. Barrell, H. Bussey, R. W. Davis, B. Dujon, H. Feldmann, F. Galibert, J. D. Hoheisel, C. Jacq, M. Johnston, E. J. Louis, H. W. Mewes, Y. Murakami, P. Philippsen, H. Tettelin, and S. G. Oliver. Life with 6000 genes. *Science (New York, N.Y.)*, 274(5287) :546, 563–7, oct 1996.

[172] C.-S. Goh, A. A. Bogan, M. Joachimiak, D. Walther, and F. E. Cohen. Co-evolution of proteins with their interaction partners 1 1Edited by B. Honig. *Journal of Molecular Biology*, 299(2) :283–293, jun 2000.

[173] M. Gomez, S. Cutting, and P. Stragier. Transcription of spoIVB is the only role of sigma G that is essential for pro-sigma K processing during spore formation in *Bacillus subtilis*. *Journal of bacteriology*, 177(16) :4825–7, aug 1995.

[174] M. Goodman, J. Czelusniak, G. W. Moore, A. E. Romero-Herrera, and G. Matsuda. Fitting the Gene Lineage into its Species Lineage, a Parsimony Strategy Illustrated by Cladograms Constructed from Globin Sequences. *Systematic Biology*, 28(2) :132–163, jun 1979.

[175] M. Gouy, S. Guindon, and O. Gascuel. SeaView Version 4 : A Multiplatform Graphical User Interface for Sequence Alignment and Phylogenetic Tree Building. *Molecular Biology and Evolution*, 27(2) :221–224, feb 2010.

[176] T. G. W. Graham, X. Wang, D. Song, C. M. Etson, A. M. van Oijen, D. Z. Rudner, and J. J. Loparo. ParB spreading requires DNA bridging. *Genes & development*, 28(11) :1228–38, jun 2014.

[177] I. Grainge, C. Lesterlin, and D. J. Sherratt. Activation of XerCD-dif recombination by the FtsK DNA translocase. *Nucleic acids research*, 39(12) :5140–8, jul 2011.

[178] D. C. Grainger. Structure and function of bacterial H-NS protein. *Biochemical Society Transactions*, 44(6) :1561–1569, dec 2016.

[179] J. A. Gregory, E. C. Becker, and K. Pogliano. Bacillus subtilis MinC destabilizes FtsZ-rings at new cell poles and contributes to the timing of cell division. *Genes & Development*, 22(24) :3475–3488, dec 2008.

[180] S. Gruber and J. Errington. Recruitment of condensin to replication origin regions by ParB/SpoOJ promotes chromosome segregation in *B. subtilis*. *Cell*, 137(4) :685–96, may 2009.

[181] S. Gruber, J.-W. Veening, J. Bach, M. Blettinger, M. Bramkamp, and J. Errington. Interlinked sister chromosomes arise in the absence of condensin during fast replication in *B. subtilis*. *Current biology : CB*, 24(3) :293–8, feb 2014.

[182] F. J. Gueiros-Filho and R. Losick. A widely conserved bacterial cell division protein that promotes assembly of the tubulin-like protein FtsZ. *Genes & development*, 16(19) :2544–56, oct 2002.

[183] S. Guindon, J.-F. Dufayard, V. Lefort, M. Anisimova, W. Hordijk, and O. Gascuel. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies : Assessing the Performance of PhyML 3.0. *Systematic Biology*, 59(3) :307–321, mar 2010.

[184] S. Gupta, S. K. Banerjee, A. Chatterjee, A. K. Sharma, M. Kundu, and J. Basu. Essential protein SepF of mycobacteria interacts with FtsZ and MurG to regulate cell growth and division. *Microbiology*, 161(8) :1627–1638, aug 2015.

[185] T. Hadi, U. Dahl, C. Mayer, and M. E. Tanner. Mechanistic Studies on *N* -Acetylmuramic Acid 6-Phosphate Hydrolase (MurQ) : An Etherase Involved in Peptidoglycan Recycling [†]. *Biochemistry*, 47(44) :11547–11558, nov 2008.

[186] D. P. Haeusser, R. L. Schwartz, A. M. Smith, M. E. Oates, and P. A. Levin. EzrA prevents aberrant cell division by modulating assembly of the cytoskeletal protein FtsZ. *Molecular Microbiology*, 52(3) :801–814, apr 2004.

[187] D. H. Haft, M. DiCuccio, A. Badretdin, V. Brover, V. Chetvernin, K. O'Neill, W. Li, F. Chitsaz, M. K. Derbyshire, N. R. Gonzales, M. Gwadz, F. Lu, G. H. Marchler, J. S. Song, N. Thanki, R. A. Yamashita, C. Zheng, F. Thibaud-Nissen, L. Y. Geer, A. Marchler-Bauer, and K. D. Pruitt. RefSeq : an update on prokaryotic genome annotation and curation. *Nucleic Acids Research*, 46(D1) :D851–D860, jan 2018.

[188] C. A. Hale, H. Meinhardt, and P. A. de Boer. Dynamic localization cycle of the cell division regulator MinE in *Escherichia coli*. *The EMBO journal*, 20(7) :1563–72, apr 2001.

[189] R. Hallez. Métabolisme et cycle cellulaire, deux processus interconnectés chez les bactéries. *médecine/sciences*, 32(10) :843–848, oct 2016.

[190] L. W. Hamoen, J.-C. Meile, W. de Jong, P. Noirot, and J. Errington. SepF, a novel FtsZ-interacting protein required for a late step in cell division. *Molecular Microbiology*, 59(3) :989–999, feb 2006.

[191] M. Hayashi, Y. Ogura, E. J. Harry, N. Ogasawara, and S. Moriya. Bacillus subtilis YabA is involved in determining the timing and synchrony of replication initiation. *FEMS Microbiology Letters*, 247(1) :73–79, jun 2005.

[192] P. J. Hedge and B. G. Spratt. Resistance to beta-lactam antibiotics by re-modelling the active site of an *E. coli* penicillin-binding protein. *Nature*, 318(6045) :478–80, 1985.

[193] A. Heichlinger, M. Ammelburg, E. M. Kleinschnitz, A. Latus, I. Maldener, K. Fl ? ?rdh, W. Wohlleben, and G. Muth. The MreB-like protein Mbl of Streptomyces coelicolor A3(2) depends on MreB for proper localization and contributes to spore wall synthesis. *Journal of Bacteriology*, 193(7) :1533–1542, 2011.

[194] C. Heidrich, M. F. Templin, A. Ursinus, M. Merdanovic, J. Berger, H. Schwarz, M. A. de Pedro, and J. V. Höltje. Involvement of N-acetylmuramyl-L-alanine amidases in cell separation and antibiotic-induced autolysis of *Escherichia coli*. *Molecular microbiology*, 41(1) :167–78, jul 2001.

[195] R. C. Heller and K. J. Marians. Replisome assembly and the direct restart of stalled replication forks. *Nature Reviews Molecular Cell Biology*, 7(12) :932–943, dec 2006.

[196] W. Hennig. *Grundzüge einer Theorie der phylogenetischen Systematik*. 1950.

[197] M. X. Henriques, T. Rodrigues, M. Carido, L. Ferreira, and S. R. Filipe. Synthesis of capsular polysaccharide at the division septum of *Streptococcus pneumoniae* is dependent on a bacterial tyrosine kinase. *Molecular Microbiology*, 82(2) :515–534, oct 2011.

[198] C. M. Hester and J. Lutkenhaus. Soj (ParA) DNA binding is mediated by conserved arginines and is essential for plasmid segregation. *Proceedings of the National Academy of Sciences of the United States of America*, 104(51) :20326–20331, 2007.

[199] D. Higgins and J. Dworkin. Recent progress in Bacillus subtilus sporulation. *FEMS Microbial Reviews*, 36(1) :131–148, 2013.

[200] M. L. Higgins and G. D. Shockman. Study of cycle of cell wall assembly in *Streptococcus* faecalis by three-dimensional reconstructions of thin sections of cells. *Journal of bacteriology*, 127(3) :1346–58, sep 1976.

[201] T. M. Hill, J. M. Henson, and P. L. Kuempel. The terminus region of the *Escherichia coli* chromosome contains two separate loci that exhibit polar inhibition of replication. *Proceedings of the National Academy of Sciences of the United States of America*, 84(7) :1754–8, apr 1987.

[202] R. Holliday. A mechanism for gene conversion in fungi. *Genetical Research*, 5(02) :282, jul 1964.

[203] P. Hols, F. Hancy, L. Fontaine, B. Grossiord, D. Prozzi, N. Leblond-Bourget, B. Decaris, A. Bolotin, C. Delorme, S. Dusko Ehrlich, E. Guédon, V. Monnet, P. Renault, and M. Kleerebezem. New insights in the molecular biology and physiology of *Streptococcus thermophilus* revealed by comparative genomics. *FEMS Microbiology Reviews*, 29(3) :435–463, aug 2005.

[204] J.-V. Höltje. "Three for one" — a Simple Growth Mechanism that Guarantees a Precise Copy of the Thin, Rod-Shaped Murein Sacculus of *Escherichia coli*. In *Bacterial Growth and Lysis*, pages 419–426. Springer US, Boston, MA, 1993.

[205] I. Hörger, E. Velasco, J. Mingorance, G. Rivas, P. Tarazona, and M. Vélez. Langevin computer simulations of bacterial protein filaments and the force-generating mechanism during cell division. *Physical Review E*, 77(1) :011902, jan 2008.

[206] I. Hörger, E. Velasco, G. Rivas, M. Vélez, and P. Tarazona. FtsZ Bacterial Cytoskeletal Polymers on Curved Surfaces : The Importance of Lateral Interactions. *Biophysical Journal*, 94(11) :L81–L83, jun 2008.

[207] Z. Hu, C. Saez, and J. Lutkenhaus. Recruitment of MinC, an inhibitor of Z-ring formation, to the membrane in *Escherichia coli* : role of MinD and MinE. *Journal of bacteriology*, 185(1) :196–203, jan 2003.

[208] J. Huerta-Cepas, F. Serra, and P. Bork. ETE 3 : Reconstruction, Analysis, and Visualization of Phylogenomic Data. *Molecular biology and evolution*, 33(6) :1635–8, 2016.

[209] O. Huisman and R. D'Ari. An inducible DNA replication-cell division coupling mechanism in *E. coli*. *Nature*, 290(5809) :797–9, apr 1981.

[210] S. Hunter, R. Apweiler, T. K. Attwood, A. Bairoch, A. Bateman, D. Binns, P. Bork, U. Das, L. Daugherty, L. Duquenne, R. D. Finn, J. Gough, D. Haft, N. Hulo, D. Kahn, E. Kelly, A. Laugraud, I. Letunic, D. Lonsdale, R. Lopez, M. Madera, J. Maslen, C. McAnulla, J. McDowall, J. Mistry, A. Mitchell, N. Mulder, D. Natale, C. Orengo, A. F. Quinn, J. D. Selengut, C. J. A. Sigrist, M. Thimma, P. D. Thomas, F. Valentin, D. Wilson, C. H. Wu, and C. Yeats. InterPro : the integrative protein signature database. *Nucleic acids research*, 37(Database issue) :D211–5, jan 2009.

[211] M. Huynen, B. Snel, W. Lathe, and P. Bork. Predicting protein function by genomic context : quantitative evaluation and qualitative inferences. *Genome research*, 10(8) :1204–10, aug 2000.

[212] D. Hyatt, G.-L. Chen, P. F. LoCascio, M. L. Land, F. W. Larimer, and L. J. Hauser. Prodigal : prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*, 11(1) :119, mar 2010.

[213] M. Ikeda, M. Wachi, H. K. Jung, F. Ishino, and M. Matsuhashi. The *Escherichia coli* mraY gene encoding UDP-N-acetylmuramoyl-pentapeptide : undecaprenylphosphate phospho-N-acetylmuramoyl-pentapeptide transferase. *Journal of bacteriology*, 173(3) :1021–6, feb 1991.

[214] A. A. Iniesta. ParABS System in Chromosome Partitioning in the Bacterium Myxococcus xanthus. *PLoS ONE*, 9(1) :e86897, jan 2014.

[215] F. Jacob, S. Brenner, and F. Cuzin. On the Regulation of DNA Replication in Bacteria. *Cold Spring Harbor Symposia on Quantitative Biology*, 28(0) :329–348, jan 1963.

[216] F. JACOB, D. PERRIN, C. SANCHEZ, and J. MONOD. Operon : a group of genes with the expression coordinated by an operator. *Comptes rendus hebdomadaires des seances de l'Academie des sciences*, 250 :1727–9, feb 1960.

[217] D. B. A. James, K. Gupta, J. R. Hauser, and J. Yother. Biochemical activities of *Streptococcus pneumoniae* serotype 2 capsular glycosyltransferases and significance of suppres-

sor mutations affecting the initiating glycosyltransferase Cps2E. *Journal of bacteriology*, 195(24) :5469–78, dec 2013.

[218] K. H. Jameson and A. J. Wilkinson. Control of Initiation of DNA Replication in *Bacillus subtilis* and *Escherichia coli*. *Genes*, 8(1), jan 2017.

[219] F. Jauffrit, S. Penel, S. Delmotte, C. Rey, D. M. de Vienne, M. Gouy, J.-P. Charrier, J.-P. Flandrois, and C. Brochier-Armanet. RiboDB Database : A Comprehensive Resource for Prokaryotic Systematics. *Molecular biology and evolution*, 33(8) :2170–2, aug 2016.

[220] P. C. Jennings, G. C. Cox, L. G. Monahan, and E. J. Harry. Super-resolution imaging of the bacterial cytokinetic protein FtsZ. *Micron*, 42(4) :336–341, jun 2011.

[221] H. Jeong, S. P. Mason, A.-L. Barabási, and Z. N. Oltvai. Lethality and centrality in protein networks. *Nature*, 411(6833) :41–42, may 2001.

[222] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A.-L. Barabási. The large-scale organization of metabolic networks. *Nature*, 407(6804) :651–654, oct 2000.

[223] C. Jers, B. Soufi, C. Grangeasse, J. Deutscher, and I. Mijakovic. Phosphoproteomics in bacteria : towards a systemic understanding of bacterial phosphorylation networks. *Expert Review of Proteomics*, 5(4) :619–627, aug 2008.

[224] C. Jiang, P. D. Caccamo, and Y. V. Brun. Mechanisms of bacterial morphogenesis : evolutionary cell biology approaches provide new insights. *BioEssays : news and reviews in molecular, cellular and developmental biology*, 37(4) :413–25, apr 2015.

[225] Z. Jiang. Protein Function Predictions Based on the Phylogenetic Profile Method. *Critical Reviews in Biotechnology*, 28(4) :233–238, jan 2008.

[226] J. W. Johnson, J. F. Fisher, and S. Mobashery. Bacterial cell wall recycling. *Annals of the new york academy*, 1277(1) :54–75, 2013.

[227] L. Jolly, P. Ferrari, D. Blanot, J. Van Heijenoort, F. Fassy, and D. Mengin-Lecreulx. Reaction mechanism of phosphoglucosamine mutase from *Escherichia coli*. *European journal of biochemistry*, 262(1) :202–10, may 1999.

[228] L. J. Jones, R. Carballido-López, and J. Errington. Control of cell shape in bacteria : helical, actin-like filaments in *Bacillus subtilis*. *Cell*, 104(6) :913–22, mar 2001.

[229] R. Jothi, T. M. Przytycka, and L. Aravind. Discovering functional linkages and uncharacterized cellular pathways using phylogenetic profile comparisons : a comprehensive assessment. *BMC Bioinformatics*, 8(1) :173, may 2007.

[230] L. Juan Wu and J. Errington. Nucleoid occlusion and bacterial cell division. *Nature reviews*, 10 :8–12, 2011.

[231] I. Junier and O. Rivoire. Synteny in Bacterial Genomes : Inference, Organization and Evolution. jul 2013.

[232] J. M. Kaguni. Replication initiation at the *Escherichia coli* chromosomal origin. *Current opinion in chemical biology*, 15(5) :606–13, oct 2011.

[233] H. Kakeshita, K. Yamane, A. Oguro, K. Nakamura, and R. Amikura. Expression of the ftsY gene, encoding a homologue of the α subunit of mammalian signal recognition particle receptor, is controlled by different promoters in vegetative and sporulating cells of *Bacillus subtilis*. *Microbiology*, 146(10) :2595–2603, oct 2000.

[234] S. Kalyaanamoorthy, B. Q. Minh, T. K. F. Wong, A. von Haeseler, and L. S. Jermiin. ModelFinder : fast model selection for accurate phylogenetic estimates. *Nature methods*, 14(6) :587–589, jun 2017.

[235] S. Karlin and S. F. Altschul. Methods for assessing the statistical significance of molecular sequence features by using general scoring schemes. *Proceedings of the National Academy of Sciences of the United States of America*, 87(6) :2264–8, mar 1990.

[236] M. L. Karow, P. Glaser, and P. J. Piggot. Identification of a gene, spoIIR, that links the activation of sigma E to the transcriptional activity of sigma F during sporulation in *Bacillus subtilis*. *Proceedings of the National Academy of Sciences of the United States of America*, 92(6) :2012–6, mar 1995.

[237] J. Kato and T. Katayama. Hda, a novel DnaA-related protein, regulates the replication cycle in *Escherichia coli*. *The EMBO journal*, 20(15) :4253–62, aug 2001.

[238] K. Katoh and D. M. Standley. MAFFT Multiple Sequence Alignment Software Version 7 : Improvements in Performance and Usability. *Molecular Biology and Evolution*, 30(4) :772–780, apr 2013.

[239] Y. Kawai, K. Asai, and J. Errington. Partial functional redundancy of MreB isoforms, MreB, Mbl and MreBHp in cell morphogenesis of *Bacillus subtilis*. *Molecular Microbiology*, 73(4) :719–731, 2009.

[240] Y. Kawai, R. A. Daniel, and J. Errington. Regulation of cell wall morphogenesis in *Bacillus subtilis* by recruitment of PBP1 to the MreB helix. *Molecular microbiology*, 71(5) :1131–44, mar 2009.

[241] P. R. Kensche, V. van Noort, B. E. Dutilh, and M. A. Huynen. Practical and theoretical advances in predicting the function of a protein by its phylogenetic distribution. *Journal of the Royal Society, Interface*, 5(19) :151–70, feb 2008.

[242] K. Keyamura, C. Sakaguchi, Y. Kubota, H. Niki, and T. Hishida. RecA Protein Recruits Structural Maintenance of Chromosomes (SMC)-like RecN Protein to DNA Double-strand Breaks. *Journal of Biological Chemistry*, 288(41) :29229–29237, oct 2013.

[243] D. Kiekebusch, K. Michie, L.-O. Essen, J. Löwe, and M. Thanbichler. Localized Dimerization and Nucleoid Binding Drive Gradient Formation by the Bacterial Cell Division Inhibitor MipZ. *Molecular Cell*, 46(3) :245–259, may 2012.

[244] H. Kishida, S. Unzai, D. I. Roper, A. Lloyd, S.-Y. Park, and J. R. H. Tame. Crystal structure of penicillin binding protein 4 (dacB) from *Escherichia coli*, both in the native form and covalently linked to various antibiotics. *Biochemistry*, 45(3) :783–92, jan 2006.

[245] N. Kleckner, J. K. Fisher, M. Stouf, M. A. White, D. Bates, and G. Witz. The bacterial nucleoid : nature, dynamics and sister segregation. *Current opinion in microbiology*, 22 :127–37, dec 2014.

[246] A. L. Koch and R. J. Doyle. Inside-to-outside growth and turnover of the wall of gram-positive rods. *Journal of Theoretical Biology*, 117(1) :137–157, nov 1985.

[247] E. V. Koonin. Orthologs, Paralogs, and Evolutionary Genomics. *Annual Review of Genetics*, 39(1) :309–338, dec 2005.

[248] U. Kopp, M. Roos, J. Wecke, and H. Labischinski1. Staphylococcal Peptidoglycan Interpeptide Bridge Biosynthesis : A Novel Antistaphylococcal Target ? *Microbial Drug Resistance*, 2(1), 1996.

[249] M. Kostka, M. Uzlikova, I. Cepicka, and J. Flegr. SlowFaster, a user-friendly program for slow-fast analysis and its application on phylogeny of Blastocystis. *BMC Bioinformatics*, 9(1) :341, aug 2008.

[250] S. C. Kowalczykowski, D. A. Dixon, A. K. Eggleston, S. D. Lauder, and W. M. Rehrauer. Biochemistry of homologous recombination in *Escherichia coli*. *Microbiological reviews*, 58(3) :401–65, sep 1994.

[251] D. Kraus, H. Kalbacher, J. Buschmann, B. Berger-Bächi, F. Götz, and A. Peschel. Muropeptide modification-amidation of peptidoglycan D-glutamate does not affect the proinflammatory activity of *Staphylococcus aureus*. *Infection and immunity*, 75(4) :2084–7, apr 2007.

[252] T. Kruse, J. Bork-Jensen, and K. Gerdes. The morphogenetic MreBCD proteins of *Escherichia coli* form an essential membrane-bound complex. *Molecular Microbiology*, 55(1) :78–89, oct 2004.

[253] V. Kunin and C. A. Ouzounis. The balance of driving forces during genome evolution in prokaryotes. *Genome research*, 13(7) :1589–94, jul 2003.

[254] T. Kunisawa. Evolutionary relationships of completely sequenced *Clostridia* species and close relatives. *International Journal of Systematic and Evolutionary Microbiology*, 65(11) :4276–4283, nov 2015.

[255] L. L. Lackner, D. M. Raskin, and P. A. J. de Boer. ATP-dependent interactions between *Escherichia coli* Min proteins and the phospholipid membrane in vitro. *Journal of bacteriology*, 185(3) :735–49, feb 2003.

[256] H. Laitinen and A. Tomasz. Changes in composition of peptidoglycan during maturation of the cell wall in pneumococci. *Journal of bacteriology*, 172(10) :5961–7, oct 1990.

[257] J. B. P. A. d. M. Lamarck. *Philosophie zoologique*. 1830.

[258] A. D. Land and M. E. Winkler. The Requirement for Pneumococcal MreC and MreD Is Relieved by Inactivation of the Gene Encoding PBP1a. *Journal of Bacteriology*, 193(16) :4166–4179, aug 2011.

[259] J. M. Lang, A. E. Darling, and J. A. Eisen. Phylogeny of Bacterial and Archaeal Genomes Using Conserved Genes : Supertrees and Supermatrices. *PLoS ONE*, 8(4) :e62510, apr 2013.

[260] A. Larsson. AliView : a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics*, 30(22) :3276–3278, nov 2014.

[261] N. Lartillot, T. Lepage, and S. Blanquart. PhyloBayes 3 : a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics*, 25(17) :2286–2288, sep 2009.

[262] N. Lartillot and H. Philippe. A Bayesian Mixture Model for Across-Site Heterogeneities in the Amino-Acid Replacement Process. *Molecular Biology and Evolution*, 21(6) :1095–1109, jun 2004.

[263] W. C. Lathe, B. Snel, and P. Bork. Gene context conservation of a higher order than operons. *Trends in Biochemical Sciences*, 25(10) :474–479, oct 2000.

[264] S. Q. Le and O. Gascuel. An Improved General Amino Acid Replacement Matrix. *Molecular Biology and Evolution*, 25(7) :1307–1320, apr 2008.

[265] P. Le Bourgeois, M. Bugarel, N. Campo, M.-L. Daveran-Mingot, J. Labonté, D. Lanfranchi, T. Lautier, C. Pagès, and P. Ritzenthaler. The unconventional Xer recombination machinery of Streptococci/Lactococci. *PLoS genetics*, 3(7) :e117, jul 2007.

[266] V. Lefort, R. Desper, and O. Gascuel. FastME 2.0 : A Comprehensive, Accurate, and Fast Distance-Based Phylogeny Inference Program : Table 1. *Molecular Biology and Evolution*, 32(10) :2798–2800, oct 2015.

[267] F. Lemoine, J.-B. Domelevo Entfellner, E. Wilkinson, D. Correia, M. Dávila Felipe, T. De Oliveira, and O. Gascuel. Renewing Felsenstein's phylogenetic bootstrap in the era of big data. *Nature*, 556(7702) :452–456, apr 2018.

[268] R. Lenarcic, S. Halbedel, L. Visser, M. Shaw, L. J. Wu, J. Errington, D. Marenduzzo, and L. W. Hamoen. Localisation of DivIVA by targeting to negatively curved membranes. *The EMBO Journal*, 28(15) :2272–2282, aug 2009.

[269] C. Lesterlin, F.-X. Barre, and F. Cornet. Genetic recombination and the cell cycle : what we have learned from chromosome dimers. *Molecular microbiology*, 54(5) :1151–60, dec 2004.

[270] C. Lesterlin, C. Pages, N. Dubarry, S. Dasgupta, and F. Cornet. Asymmetry of Chromosome Replichores Renders the DNA Translocase Activity of FtsK Essential for Cell Division and Cell Shape Maintenance in *Escherichia coli*. *PLoS Genetics*, 4(12) :e1000288, dec 2008.

[271] I. Letunic and P. Bork. Interactive Tree Of Life (iTOL) : an online tool for phylogenetic tree display and annotation. *Bioinformatics*, 23(1) :127–128, jan 2007.

[272] P. A. Levin, I. G. Kurtser, and A. D. Grossman. Identification and characterization of a negative regulator of FtsZ ring formation in *Bacillus subtilis*. *Proceedings of the National Academy of Sciences of the United States of America*, 96(17) :9642–7, aug 1999.

[273] P. A. Levin and R. Losick. Transcription factor Spo0A switches the localization of the cell division protein FtsZ from a medial to a bipolar pattern in *Bacillus subtilis*. *Genes & development*, 10(4) :478–88, feb 1996.

[274] P. A. Levin, P. S. Margolis, P. Setlow, R. Losick, and D. Sun. Identification of *Bacillus subtilis* genes for septum placement and shape determination. *Journal of bacteriology*, 174(21) :6717–28, nov 1992.

[275] Y. Li, N. K. Stewart, A. J. Berger, S. Vos, A. J. Schoeffler, J. M. Berger, B. T. Chait, and M. G. Oakley. Escherichia coli condensin MukB stimulates topoisomerase IV activity by a direct physical interaction. *Proceedings of the National Academy of Sciences of the United States of America*, 107(44) :18832–7, nov 2010.

[276] Z. Li, M. J. Trimble, Y. V. Brun, and G. J. Jensen. The structure of FtsZ filaments *in vivo* suggests a force-generating role in cell division. *The EMBO Journal*, 26(22) :4694–4708, nov 2007.

[277] E. A. Libby, L. A. Goss, and J. Dworkin. The Eukaryotic-Like Ser/Thr Kinase PrkC Regulates the Essential WalRK Two-Component System in *Bacillus subtilis*. *PLOS Genetics*, 11(6) :e1005275, jun 2015.

[278] D. Liberles, A. Thoren, G. Heijne, and A. Elofsson. The Use of Phylogenetic Profiles for Gene Predictions. *Current Genomics*, 3(3) :131–137, jun 2002.

[279] M. Lies, B. J. Visser, M. C. Joshi, D. Magnan, and D. Bates. MioC and GidA proteins promote cell division in *E. coli*. *Frontiers in Microbiology*, 6 :516, 2015.

[280] D. Liger, A. Masson, D. Blanot, J. van Heijenoort, and C. Parquet. Over-production, purification and properties of the uridine-diphosphate-N-acetylmuramate :L-alanine ligase from *Escherichia coli*. *European journal of biochemistry*, 230(1) :80–7, may 1995.

[281] H. C. Lim, I. V. Surovtsev, B. G. Beltran, F. Huang, J. Bewersdorf, and C. Jacobs-Wagner. Evidence for a DNA-relay mechanism in ParABS-mediated chromosome segregation. *eLife*, 3 :e02758, may 2014.

[282] T.-H. Lin, Y.-N. Hu, and G.-C. Shaw. Two enzymes, TilS and HprT, can form a complex to function as a transcriptional activator for the cell division protease gene ftsH in *Bacillus subtilis*. *The Journal of Biochemistry*, 155(1) :5–16, jan 2014.

[283] Z. Liu, A. Mukherjee, and J. Lutkenhaus. Recruitment of ZipA to the division site by interaction with FtsZ. *Molecular microbiology*, 31(6) :1853–61, mar 1999.

[284] A. J. Lloyd, A. M. Gilbey, A. M. Blewett, G. De Pascale, A. El Zoeiby, R. C. Levesque, A. C. Catherwood, A. Tomasz, T. D. H. Bugg, D. I. Roper, and C. G. Dowson. Characterization of tRNA-dependent Peptide Bond Formation by MurM in the Synthesis of *Streptococcus pneumoniae* Peptidoglycan. *Journal of Biological Chemistry*, 283(10) :6402–6417, mar 2008.

[285] R. L. Lock and E. J. Harry. Cell-division inhibitors : new insights for future antibiotics. *Nature reviews. Drug discovery*, 7(4) :324–338, 2008.

[286] H. H. Low, M. C. Moncrieffe, and J. Löwe. The crystal structure of ZapA and its modulation of FtsZ polymerisation. *Journal of molecular biology*, 341(3) :839–52, aug 2004.

[287] S. Lu, S. Cutting, and L. Kroos. Sporulation protein SpoIVFB from *Bacillus subtilis* enhances processing of the sigma factor precursor Pro-sigma K in the absence of other sporulation gene products. *Journal of bacteriology*, 177(4) :1082–5, feb 1995.

[288] J. Luciano, R. Agrebi, A. V. Le Gall, M. Wartel, F. Fiegna, A. Ducret, C. Brochier-Armanet, and T. Mignot. Emergence and modular evolution of a novel motility machinery in bacteria. *PLoS genetics*, 7(9) :e1002268, sep 2011.

[289] W. Ludwig and K.-H. Schleifer. *Molecular phylogeny of bacteria based on comparative sequence analysis of conserved genes*. Microbial phylogeny and evolution : concepts and controversies.Oxford University Press, 2005.

[290] J. Luirink, C. ten Hagen-Jongman, C. van der Weijden, B. Oudega, S. High, B. Dobberstein, and R. Kusters. An alternative protein targeting pathway in *Escherichia coli* : studies on the role of FtsY. *The EMBO Journal*, 13(10) :2289–2296, may 1994.

[291] J. Lutkenhaus. Regulation of cell division in *E. coli*. *Trends in Genetics*, 6 :22–25, jan 1990.

[292] J. Lutkenhaus. The ParA/MinD family puts things in their place. *Trends in microbiology*, 20(9) :411–8, sep 2012.

[293] J. Lutkenhaus, S. Pichoff, and S. Du. Bacterial cytokinesis : From Z ring to divisome. *Cytoskeleton (Hoboken, N.J.)*, 69(10) :778–90, oct 2012.

[294] X. Ma and W. Margolin. Genetic and functional analyses of the conserved C-terminal core domain of *Escherichia coli* FtsZ. *Journal of bacteriology*, 181(24) :7531–44, dec 1999.

[295] P. Mackiewicz, J. Zakrzewska-Czerwinska, A. Zawilak, M. R. Dudek, and S. Cebrat. Where does bacterial replication start ? Rules for predicting the oriC region. *Nucleic acids research*, 32(13) :3781–91, jul 2004.

[296] W. Maddison and D. Maddison. Mesquite : a modular system for evolutionary analysis, 2004.

[297] D. Mader, M. Liebeke, V. Winstel, K. Methling, M. Leibig, F. Götz, M. Lalk, and A. Peschel. Role of N-terminal protein formylation in central metabolic processes in *Staphylococcus aureus*. *BMC Microbiology*, 13(1) :7, jan 2013.

[298] B. Maestro, L. Novaková, D. Hesek, M. Lee, E. Leyva, S. Mobashery, J. M. Sanz, and P. Branny. Recognition of peptidoglycan and β-lactam antibiotics by the extracellular

domain of the Ser/Thr protein kinase StkP from *Streptococcus pneumoniae*. *FEBS letters*, 585(2) :357–63, jan 2011.

[299] S. Maggi, O. Massidda, G. Luzi, D. Fadda, L. Paolozzi, and P. Ghelardini. Division protein interaction web : Identification of a phylogenetically conserved common interactome between *Streptococcus pneumoniae* and *Escherichia coli*. *Microbiology*, 154(10) :3042–3052, 2008.

[300] S. Manuse, A. Fleurie, L. Zucchini, C. Lesterlin, and C. Grangeasse. Role of eukaryotic-like serine/threonine kinases in bacterial cell division and morphogenesis. *FEMS Microbiology Reviews*, 40(1) :41–56, jan 2016.

[301] S. Manuse, N. L. Jean, M. Guinot, J. P. Lavergne, C. Laguri, C. M. Bougault, M. S. Vannieuwenhze, C. Grangeasse, and J. P. Simorre. Structure-function analysis of the extracellular domain of the pneumococcal cell division site positioning protein MapZ. *Nature Communications*, 7 :12071, 2016.

[302] H. Marchandin, C. Teyssier, J. Campos, H. Jean-Pierre, F. Roger, B. Gay, J.-P. Carlier, and E. Jumas-Bilak. Negativicoccus succinicivorans gen. nov., sp. nov., isolated from human clinical samples, emended description of the family *Veillonella*ceae and description of *Negativicutes* classis nov., Selenomonadales ord. nov. and Acidaminococcaceae fam. nov. in the bacterial *phylum Firmicutes*. *International Journal of Systematic and Evolutionary Microbiology*, 60(6) :1271–1279, jun 2010.

[303] E. M. Marcotte. Computational genetics : finding protein function by nonhomology methods. *Current Opinion in Structural Biology*, 10(3) :359–365, jun 2000.

[304] E. M. Marcotte, M. Pellegrini, H. L. Ng, D. W. Rice, T. O. Yeates, and D. Eisenberg. Detecting protein function and protein-protein interactions from genome sequences. *Science (New York, N.Y.)*, 285(5428) :751–3, jul 1999.

[305] W. Margolin. FtsZ and the division of prokaryotic cells and organelles. *Nature Reviews Molecular Cell Biology*, 6(11) :862–871, 2005.

[306] A. L. Marston and J. Errington. Selection of the midcell division site in *Bacillus subtilis* through MinD-dependent polar localization and activation of MinC. *Molecular microbiology*, 33(1) :84–96, jul 1999.

[307] B. S. Marteyn, G. Karimova, A. K. Fenton, A. D. Gazi, N. West, L. Touqui, M.-C. Prevost, J.-M. Betton, O. Poyraz, D. Ladant, K. Gerdes, P. J. Sansonetti, and C. M. Tang. ZapE Is a Novel Cell Division Protein Interacting with FtsZ and Modulating the Z-Ring Dynamics. *mBio*, 5(2) :e00022–14–e00022–14, mar 2014.

[308] C. E. Martinez-Guerrero, R. Ciria, C. Abreu-Goodger, G. Moreno-Hagelsieb, and E. Merino. GeConT 2 : gene context analysis for orthologous proteins, conserved domains and metabolic pathways. *Nucleic Acids Research*, 36(suppl_2) :W176–W180, jul 2008.

[309] F. J. Massey. The Kolmogorov-Smirnov Test for Goodness of Fit. *Journal of the American Statistical Association*, 46(253) :68–78, mar 1951.

[310] O. Massidda, D. Anderluzzi, L. Friedli, and G. Feger. Unconventional organization of the division and cell wall gene cluster of *Streptococcus pneumoniae*. *Microbiology*, 144(11) :3069–3078, nov 1998.

[311] O. Massidda, L. Nováková, and W. Vollmer. From models to pathogens : How much have we learned about *Streptococcus pneumoniae* cell division ? *Environmental Microbiology*, 15(12) :3133–3157, 2013.

[312] A. J. Meeske, E. P. Riley, W. P. Robins, T. Uehara, J. J. Mekalanos, D. Kahne, S. Walker, A. C. Kruse, T. G. Bernhardt, and D. Z. Rudner. SEDS proteins are a widespread family of bacterial cell wall polymerases. *Nature*, 537(7622) :634–638, 2016.

[313] J. C. Mellor, I. Yanai, K. H. Clodfelter, J. Mintseris, and C. DeLisi. Predictome : a database of putative functional links between proteins. *Nucleic acids research*, 30(1) :306–9, jan 2002.

[314] D. Mengin-Lecreulx, T. Falla, D. Blanot, J. van Heijenoort, D. J. Adams, and I. Chopra. Expression of the *Staphylococcus aureus* UDP-N-acetylmuramoyl- L-alanyl-D-glutamate :L-lysine ligase in *Escherichia coli* and effects on peptidoglycan biosynthesis and cell growth. *Journal of bacteriology*, 181(19) :5909–14, oct 1999.

[315] D. Mengin-Lecreulx, L. Texier, M. Rousseau, and J. van Heijenoort. The murG gene of *Escherichia coli* codes for the UDP-N-acetylglucosamine : N-acetylmuramyl-(pentapeptide) pyrophosphoryl-undecaprenol N-acetylglucosamine transferase involved

in the membrane steps of peptidoglycan synthesis. *Journal of bacteriology*, 173(15) :4625–36, aug 1991.

[316] D. Mengin-Lecreulx and J. van Heijenoort. Identification of the glmU gene encoding N-acetylglucosamine-1-phosphate uridyltransferase in *Escherichia coli*. *Journal of bacteriology*, 175(19) :6150–7, oct 1993.

[317] R. Mercier, M.-A. Petit, S. Schbath, S. Robin, M. El Karoui, F. Boccard, and O. Espéli. The MatP/matS site-specific system organizes the terminus region of the *E. coli* chromosome into a macrodomain. *Cell*, 135(3) :475–85, oct 2008.

[318] C. Mercy, J.-P. Lavergne, J. Slager, A. Ducret, P. S. Garcia, M.-F. Noirot-Gros, N. Dubarry, J. Nourikyan, J.-W. Veening, and C. Grangeasse. RocS drives chromosome segregation and nucleoid occlusion in *Streptococcus pneumoniae*. *bioRxiv*, page 359943, jul 2018.

[319] A. Mhammedi-Alaoui, M. Pato, M. J. Gama, and A. Toussaint. A new component of bacteriophage Mu replicative transposition machinery : the *Escherichia coli* ClpX protein. *Molecular microbiology*, 11(6) :1109–16, mar 1994.

[320] V. Miele, S. Penel, and L. Duret. Ultra-fast sequence clustering from similarity networks with SiLiX. *BMC Bioinformatics*, 12(1) :116, apr 2011.

[321] I. Mijakovic, C. Grangeasse, and K. Turgay. Exploring the diversity of protein modifications : special bacterial phosphorylation systems. *FEMS microbiology reviews*, 40(3) :398–417, may 2016.

[322] G. Minasov, M. Teplova, G. C. Stewart, E. V. Koonin, W. F. Anderson, and M. Egli. Functional implications from crystal structures of the conserved *Bacillus subtilis* protein Maf with and without dUTP. *Proceedings of the National Academy of Sciences of the United States of America*, 97(12) :6328–33, jun 2000.

[323] B. Q. Minh, M. A. T. Nguyen, and A. von Haeseler. Ultrafast approximation for phylogenetic bootstrap. *Molecular biology and evolution*, 30(5) :1188–95, may 2013.

[324] A. Minnen, L. Attaiech, M. Thon, S. Gruber, and J.-W. Veening. SMC is recruited to oriC by ParB and promotes chromosome segregation in *Streptococcus pneumoniae*. *Molecular Microbiology*, 81(3) :676–688, aug 2011.

[325] M. Mir, J. Asong, X. Li, J. Cardot, G.-J. Boons, and R. N. Husson. The Extracytoplasmic Domain of the *Mycobacterium tuberculosis* Ser/Thr Kinase PknB Binds Specific Muropeptides and Is Required for PknB Localization. *PLoS Pathogens*, 7(7) :e1002182, jul 2011.

[326] T. Mohammadi, A. Karczmarek, M. Crouvoisier, A. Bouhss, D. Mengin-Lecreulx, and T. den Blaauwen. The essential peptidoglycan glycosyltransferase MurG forms a complex with proteins involved in lateral envelope growth as well as with proteins involved in cell division in *Escherichia coli*. *Molecular Microbiology*, 65(4) :1106–1121, aug 2007.

[327] T. Mohammadi, V. van Dam, R. Sijbrandi, T. Vernet, A. Zapun, A. Bouhss, M. Diepeveen-de Bruin, M. Nguyen-Distèche, B. de Kruijff, and E. Breukink. Identification of FtsW as a transporter of lipid-linked cell wall precursors across the membrane. *The EMBO Journal*, 30(8) :1425–1432, apr 2011.

[328] L. G. Monahan and E. J. Harry. Identifying how bacterial cells find their middle : A new perspective. *Molecular Microbiology*, 87(2) :231–234, 2013.

[329] L. G. Monahan, A. T. Liew, A. L. Bottomley, and E. J. Harry. Division site positioning in bacteria : One size does not fit all. *Frontiers in Microbiology*, 5 :19, 2014.

[330] J. Monnet, W. Grange, T. R. Strick, and N. Joly. Mfd as a central partner of transcription coupled repair. *Transcription*, 4(3) :109–13, 2013.

[331] J. M. Monteiro, A. R. Pereira, N. T. Reichmann, B. M. Saraiva, P. B. Fernandes, H. Veiga, A. C. Tavares, M. Santos, M. T. Ferreira, V. Macário, M. S. VanNieuwenhze, S. R. Filipe, and M. G. Pinho. Peptidoglycan synthesis drives an FtsZ-treadmilling-independent step of cytokinesis. *Nature*, 554(7693) :528–532, feb 2018.

[332] C. Morlot, L. Bayle, M. Jacq, A. Fleurie, G. Tourcier, F. Galisson, T. Vernet, C. Grangeasse, and A. M. Di Guilmi. Interaction of Penicillin-Binding Protein 2x and Ser/Thr protein kinase StkP, two key players in *Streptococcus pneumoniae* R6 morphogenesis. *Molecular microbiology*, 90(1) :88–102, oct 2013.

[333] C. Morlot, L. Pernot, A. Le Gouellec, A. M. Di Guilmi, T. Vernet, O. Dideberg, and A. Dessen. Crystal structure of a peptidoglycan synthesis regulatory factor (PBP3) from

Streptococcus pneumoniae. *The Journal of biological chemistry*, 280(16) :15984–91, apr 2005.

[334] J. K. Morona, R. Morona, D. C. Miller, and J. C. Paton. Streptococcus pneumoniae capsule biosynthesis protein CpsB is a novel manganese-dependent phosphotyrosine-protein phosphatase. *Journal of bacteriology*, 184(2) :577–83, jan 2002.

[335] J. K. Morona, J. C. Paton, D. C. Miller, and R. Morona. Tyrosine phosphorylation of CpsD negatively regulates capsular polysaccharide biosynthesis in streptococcus pneumoniae. *Molecular microbiology*, 35(6) :1431–42, mar 2000.

[336] A. Mukherjee and J. Lutkenhaus. Guanine nucleotide-dependent assembly of FtsZ into filaments. *Journal of bacteriology*, 176(9) :2754–8, may 1994.

[337] A. Mukherjee and J. Lutkenhaus. Dynamic assembly of FtsZ regulated by GTP hydrolysis. *The EMBO Journal*, 17(2) :462–469, jan 1998.

[338] E. Mulder and C. L. Woldringh. Actively replicating nucleoids influence positioning of division sites in *Escherichia coli* filaments forming cells lacking DNA. *Journal of bacteriology*, 171(8) :4303–14, aug 1989.

[339] D. Münch, T. Roemer, S. H. Lee, M. Engeser, H. G. Sahl, and T. Schneider. Identification and in vitro Analysis of the GatD/MurT Enzyme-Complex Catalyzing Lipid II Amidation in *Staphylococcus aureus*. *PLoS Pathogens*, 8(1) :e1002509, jan 2012.

[340] J. M. Munita and C. A. Arias. Mechanisms of Antibiotic Resistance. *Microbiology spectrum*, 4(2), 2016.

[341] H. Murray and J. Errington. Dynamic Control of the DNA Replication Initiation Protein DnaA by Soj/ParA. *Cell*, 135(1) :74–84, oct 2008.

[342] N. Nanninga. Morphogenesis of *Escherichia coli*. *Microbiology and molecular biology reviews : MMBR*, 62(1) :110–29, mar 1998.

[343] P. Natale, M. Pazos, and M. Vicente. The *Escherichia coli* divisome : Born to divide, dec 2013.

[344] K. Nevo-Dinur, S. Govindarajan, and O. Amster-Choder. Subcellular localization of RNA and proteins in prokaryotes. *Trends in Genetics*, 28(7) :314–322, jul 2012.

[345] M. E. J. Newman. Power laws, Pareto distributions and Zipf's law. Technical report, 2006.

[346] C. Neylon, A. V. Kralicek, T. M. Hill, and N. E. Dixon. Replication termination in *Escherichia coli* : structure and antihelicase activity of the Tus-Ter complex. *Microbiology and molecular biology reviews : MMBR*, 69(3) :501–26, sep 2005.

[347] W.-L. Ng, K. M. Kazmierczak, and M. E. Winkler. Defective cell wall synthesis in *Streptococcus pneumoniae* R6 depleted for the essential PcsB putative murein hydrolase or the VicR (YycF) response regulator. *Molecular Microbiology*, 53(4) :1161–1175, jul 2004.

[348] L.-T. Nguyen, H. A. Schmidt, A. von Haeseler, and B. Q. Minh. IQ-TREE : A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution*, 32(1) :268–274, jan 2015.

[349] M. Nieuwdorp, P. W. Gilijamse, N. Pai, and L. M. Kaplan. Role of the Microbiome in Energy Regulation and Metabolism. *Gastroenterology*, 146(6) :1525–1533, may 2014.

[350] E. Nogales, K. H. Downing, L. A. Amos, and J. Löwe. Tubulin and FtsZ form a distinct family of GTPases. *Nature structural biology*, 5(6) :451–8, jun 1998.

[351] S. Nolivos and D. Sherratt. The bacterial chromosome : architecture and action of bacterial SMC and SMC-like complexes. *FEMS microbiology reviews*, 38(3) :380–92, may 2014.

[352] J. Nourikyan, M. Kjos, C. Mercy, C. Cluzel, C. Morlot, M. F. Noirot-Gros, S. Guiral, J. P. Lavergne, J. W. Veening, and C. Grangeasse. Autophosphorylation of the Bacterial Tyrosine-Kinase CpsD Connects Capsule Synthesis with the Cell Cycle in *Streptococcus pneumoniae*. *PLoS Genetics*, 11(9) :e1005518, 2015.

[353] E. Noutahi, M. Semeria, M. Lafond, J. Seguin, B. Boussau, L. Guéguen, N. El-Mabrouk, and E. Tannier. Efficient Gene Tree Correction Guided by Genome Evolution. *PLOS ONE*, 11(8) :e0159559, aug 2016.

[354] J. Oberto. SyntTax : a web server linking synteny to prokaryotic taxonomy. *BMC Bioinformatics*, 14(1) :4, jan 2013.

[355] H. Ogawara. Distribution of PASTA domains in penicillin-binding proteins and serine/threonine kinases of Actinobacteria. *The Journal of Antibiotics*, 69(9) :660–685, sep 2016.

[356] R. Okazaki, T. Okazaki, K. Sakabe, K. Sugimoto, and A. Sugino. Mechanism of DNA chain growth. I. Possible discontinuity and unusual secondary structure of newly synthesized chains. *Proceedings of the National Academy of Sciences of the United States of America*, 59(2) :598–605, feb 1968.

[357] M. A. Oliva, S. C. Cordell, and J. Löwe. Structural insights into FtsZ protofilament formation. *Nature Structural & Molecular Biology*, 11(12) :1243–1250, dec 2004.

[358] J. O'Neill. The Review on Antimicrobial Resistance Chaired by Jim O'Neill December 2014. Technical Report December, 2014.

[359] M. Osaki, T. Arcondeguy, A. Bastide, C. Touriol, H. Prats, and M.-C. Trombe. The StkP/PhpP Signaling Couple in *Streptococcus pneumoniae* : Cellular Organization and Physiological Characterization. *Journal of Bacteriology*, 191(15) :4943–4950, aug 2009.

[360] M. Osawa, D. E. Anderson, and H. P. Erickson. Curved FtsZ protofilaments generate bending forces on liposome membranes. *The EMBO Journal*, 28(22) :3476–3484, nov 2009.

[361] L. Overmars, R. Kerkhoven, R. J. Siezen, and C. Francke. MGcV : the microbial genomic context viewer for comparative genome analysis. *BMC Genomics*, 14(1) :209, apr 2013.

[362] M. Pagel. Detecting Correlated Evolution on Phylogenies : A General Method for the Comparative Analysis of Discrete Characters. *Proceedings of the Royal Society B : Biological Sciences*, 255(1342) :37–45, jan 1994.

[363] L. Palazzo, B. Thomas, A.-S. Jemth, T. Colby, O. Leidecker, K. Feijs, R. Zaja, O. Loseva, J. Puigvert, I. Matic, T. Helleday, and I. Ahel. Processing of protein ADP-ribosylation by Nudix hydrolases. *Biochemical Journal*, 468(2) :293–301, jun 2015.

[364] C. J. Paredes, K. V. Alsaker, and E. T. Papoutsakis. A comparative genomic view of clostridial sporulation and physiology. *Nature Reviews Microbiology*, 3(12) :969–978, dec 2005.

[365] E. Passarge, B. Horsthemke, and R. A. Farber. Incorrect use of the term synteny. *Nature Genetics*, 23(4) :387–387, dec 1999.

[366] J. E. Patrick and D. B. Kearns. MinJ (YvjD) is a topological determinant of cell division in *Bacillus subtilis*. *Molecular Microbiology*, 70(5) :1166–1179, dec 2008.

[367] L. Patthy. Modular exchange principles in proteins. *Current Opinion in Structural Biology*, 1(3) :351–361, jun 1991.

[368] J. Pei and N. V. Grishin. COG3926 and COG5526 : a tale of two new lysozyme-like protein families. *Protein science : a publication of the Protein Society*, 14(10) :2574–81, oct 2005.

[369] M. Pellegrini, E. M. Marcotte, M. J. Thompson, D. Eisenberg, and T. O. Yeates. Assigning protein functions by comparative genome analysis : protein phylogenetic profiles. *Proceedings of the National Academy of Sciences of the United States of America*, 96(8) :4285–8, apr 1999.

[370] D. Penny. Criteria for optimising phylogenetic trees and the problem of determining the root of a tree. *Journal of molecular evolution*, 8(2) :95–116, aug 1976.

[371] A. R. Pereira, P. Reed, H. Veiga, and M. G. Pinho. The Holliday junction resolvase RecU is required for chromosome segregation and DNA damage repair in *Staphylococcus aureus*. *BMC Microbiology*, 13(1) :18, dec 2013.

[372] S. F. F. Pereira, L. Goss, and J. Dworkin. Eukaryote-Like Serine/Threonine Kinases and Phosphatases in Bacteria. *Microbiology and Molecular Biology Reviews*, 75(1) :192–212, mar 2011.

[373] I. Pérez-Roger, M. García-Sogo, J. P. Navarro-Aviñó, C. López-Acedo, F. Macián, and M. E. Armengod. Positive and negative regulatory elements in the dnaA-dnaN-recF operon of *Escherichia coli*. *Biochimie*, 73(2-3) :329–34, 1991.

[374] G. Perrière and C. Brochier-Armanet. *Concepts et méthodes en phylogénie moléculaire*. Springer, 2010.

[375] S. Pichoff and J. Lutkenhaus. Unique and overlapping roles for ZipA and FtsA in septal ring assembly in *Escherichia coli*. *The EMBO journal*, 21(4) :685–93, feb 2002.

[376] S. Pichoff and J. Lutkenhaus. Tethering the Z ring to the membrane through a conserved membrane targeting sequence in FtsA. *Molecular microbiology*, 55(6) :1722–34, mar 2005.

[377] N. Pietack, D. Becher, S. R. Schmidl, M. H. Saier, M. Hecker, F. M. Commichau, and J. Stülke. In vitro Phosphorylation of Key Metabolic Enzymes from *Bacillus subtilis :* PrkC Phosphorylates Enzymes from Different Branches of Basic Metabolism. *Journal of Molecular Microbiology and Biotechnology*, 18(3) :129–140, 2010.

[378] P. J. Piggot and D. W. Hilbert. Sporulation of *Bacillus subtilis*. *Current Opinion in Microbiology*, 7(6) :579–586, dec 2004.

[379] M. G. Pinho, M. Kjos, and J. W. Veening. How to get (a)round : Mechanisms controlling growth and division of coccoid bacteria. *Nature Reviews Microbiology*, 11(9) :601–614, 2013.

[380] M. Plomp, A. M. Carroll, P. Setlow, and A. J. Malkin. Architecture and Assembly of the *Bacillus subtilis* Spore Coat. *PLoS ONE*, 9(9) :e108560, sep 2014.

[381] F. Pompeo, E. Foulquier, B. Serrano, C. Grangeasse, and A. Galinier. Phosphorylation of the cell division protein GpsB regulates PrkC kinase activity through a negative feedback loop in *B acillus subtilis*. *Molecular Microbiology*, 97(1) :139–150, jul 2015.

[382] D. L. Popham, M. E. Gilmore, and P. Setlow. Roles of low-molecular-weight penicillin-binding proteins in *Bacillus subtilis* spore peptidoglycan synthesis and spore properties. *Journal of bacteriology*, 181(1) :126–32, jan 1999.

[383] L. Postow, N. J. Crisona, B. J. Peter, C. D. Hardy, and N. R. Cozzarelli. Topological challenges to DNA replication : conformations at the fork. *Proceedings of the National Academy of Sciences of the United States of America*, 98(15) :8219–26, jul 2001.

[384] O. Poupel, M. Moyat, J. Groizeleau, L. C. S. Antunes, S. Gribaldo, T. Msadek, and S. Dubrac. Transcriptional analysis and subcellular protein localization reveal specific features of the essential walkr system in staphylococcus aureus. *PLoS ONE*, 11(3) :1–32, 2016.

[385] J. F. Poyatos and L. D. Hurst. The determinants of gene order conservation in yeasts. *Genome biology*, 8(11) :R233, 2007.

[386] F. Pratviel-Sosa, D. Mengin-Lecreulx, and J. van Heijenoort. Over-production, purification and properties of the uridine diphosphate N-acetylmuramoyl-L-alanine :D-glutamate ligase from *Escherichia coli*. *European journal of biochemistry*, 202(3) :1169–76, dec 1991.

[387] D. M. Prescott and P. L. Kuempel. Bidirectional replication of the chromosome in *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America*, 69(10) :2842–5, oct 1972.

[388] K. D. Price and R. Losick. A four-dimensional view of assembly of a morphogenetic protein during sporulation in *Bacillus subtilis*. *Journal of bacteriology*, 181(3) :781–90, feb 1999.

[389] M. N. Price, P. S. Dehal, and A. P. Arkin. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. *PLoS ONE*, 5(3) :e9490, mar 2010.

[390] N. P. Price and F. A. Momany. Modeling bacterial UDP-HexNAc : polyprenol-P HexNAc-1-P transferases. *Glycobiology*, 15(9) :29R–42R, sep 2005.

[391] R. Priyadarshini, D. L. Popham, and K. D. Young. Daughter Cell Separation by Penicillin-Binding Proteins and Peptidoglycan Amidases in *Escherichia coli*. *Journal of Bacteriology*, 188(15) :5345–5355, aug 2006.

[392] J. C. Quintela, M. Caparrós, and M. A. de Pedro. Variability of peptidoglycan structural parameters in gram-negative bacteria. *FEMS microbiology letters*, 125(1) :95–100, jan 1995.

[393] C. R. H. Raetz and C. Whitfield. Lipopolysaccharide endotoxins. *Annual review of biochemistry*, 71(1) :635–700, jun 2002.

[394] K. S. Ramamurthi and R. Losick. ATP-driven self-assembly of a morphogenetic protein in *Bacillus subtilis*. *Molecular cell*, 31(3) :406–14, aug 2008.

[395] A. Rambaut, A. J. Drummond, D. Xie, G. Baele, and M. A. Suchard. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Systematic Biology*, apr 2018.

[396] H. G. Ramulu, M. Groussin, E. Talla, R. Planel, V. Daubin, and C. Brochier-Armanet. Ribosomal proteins : Toward a next generation standard for prokaryotic systematics ? *Molecular Phylogenetics and Evolution*, 75 :103–117, jun 2014.

[397] D. M. Raskin and P. a. J. D. Boer. MinDE-Dependent Pole-to-Pole Oscillation of Division Inhibitor MinC in *Escherichia coli* MinDE-Dependent Pole-to-Pole Oscillation of Division Inhibitor MinC in *Escherichia coli*. *Journal Of Bacteriology*, 181(20) :6419–6424, 1999.

[398] G. Real, S. Autret, E. J. Harry, J. Errington, and A. O. Henriques. Cell division protein DivIB influences the Spo0J/Soj system of chromosome segregation in *Bacillus subtilis*. *Molecular Microbiology*, 55(2) :349–367, nov 2004.

[399] R. Reyes-Lamothe, E. Nicolas, and D. J. Sherratt. Chromosome Replication and Segregation in Bacteria. *Annu. Rev. Genet*, 46(1) :121–43, 2012.

[400] E. J. Richardson and M. Watson. The automatic annotation of bacterial genomes. *Briefings in bioinformatics*, 14(1) :1–12, jan 2013.

[401] A. I. Rico, M. García-Ovalle, J. Mingorance, and M. Vicente. Role of two essential domains of *Escherichia coli* FtsA in localization and progression of the division ring. *Molecular microbiology*, 53(5) :1359–71, sep 2004.

[402] P. Rico-Lastres, R. Díez-Martínez, M. Iglesias-Bexiga, N. Bustamante, C. Aldridge, D. Hesek, M. Lee, S. Mobashery, J. Gray, W. Vollmer, P. García, and M. Menéndez. Substrate recognition and catalysis by LytB, a pneumococcal peptidoglycan hydrolase involved in virulence. *Scientific Reports*, 5(1) :16198, dec 2015.

[403] C. Robinow and E. Kellenberger. The bacterial nucleoid revisited. *Microbiological reviews*, 58(2) :211–32, jun 1994.

[404] I. B. Rogozin, K. S. Makarova, J. Murvai, E. Czabarka, Y. I. Wolf, R. L. Tatusov, L. A. Szekely, and E. V. Koonin. Connected gene neighborhoods in prokaryotic genomes. *Nucleic acids research*, 30(10) :2212–23, may 2002.

[405] M. J. Romanowski, J. B. Bonanno, and S. K. Burley. Crystal structure of the *Escherichia coli* glucose-inhibited division protein B (GidB) reveals a methyltransferase fold. *Proteins : Structure, Function and Genetics*, 47(4) :563–567, 2002.

[406] F. Ronquist, M. Teslenko, P. van der Mark, D. L. Ayres, A. Darling, S. Höhna, B. Larget, L. Liu, M. A. Suchard, and J. P. Huelsenbeck. MrBayes 3.2 : efficient Bayesian

phylogenetic inference and model choice across a large model space. *Systematic biology*, 61(3) :539–42, may 2012.

[407] A. Ruggiero, P. De Simone, G. Smaldone, F. Squeglia, and R. Berisio. Bacterial cell division regulation by Ser/Thr kinases : a structural perspective. *Current protein & peptide science*, 13(8) :756–66, dec 2012.

[408] N. Ruiz. Lipid Flippases for Bacterial Peptidoglycan Biosynthesis. *Lipid Insights*, 8s1 :LPI.S31783, jan 2015.

[409] L. Y. Rusin, E. V. Lyubetskaya, K. Y. Gorbunov, and V. A. Lyubetsky. Reconciliation of gene and species trees. *BioMed research international*, 2014 :642089, mar 2014.

[410] M. Sánchez, A. Valencia, M. J. Ferrándiz, C. Sander, and M. Vicente. Correlation between the structure and biochemical activities of FtsA, an essential cell division protein of the actin family. *The EMBO journal*, 13(20) :4919–25, oct 1994.

[411] F. Sant'Anna, A. Lebedinsky, T. Sokolova, F. Robb, and J. Gonzalez. Analysis of three genomes within the thermophilic bacterial species *Caldanaerobacter subterraneus* with a focus on carbon monoxide dehydrogenase evolution and hydrolase diversity. *BMC Genomics*, 16(1) :757, dec 2015.

[412] M. Sarich, J.-H. Prinz, and C. Schütte. Markov Model Theory. In *Advances in experimental medicine and biology*, volume 797, pages 23–44. 2014.

[413] J. Sassine, M. Xu, K. R. Sidiq, R. Emmins, J. Errington, and R. A. Daniel. Functional redundancy of division specific penicillin-binding proteins in *Bacillus subtilis*. *Molecular microbiology*, 106(2) :304–318, oct 2017.

[414] E. Sauvage, F. Kerff, M. Terrak, J. A. Ayala, and P. Charlier. The penicillin-binding proteins : Structure and role in peptidoglycan biosynthesis. *FEMS Microbiology Reviews*, 32(2) :234–258, 2008.

[415] C. J. Saveson and S. T. Lovett. Tandem repeat recombination induced by replication fork defects in *Escherichia coli* requires a novel factor, RadC. *Genetics*, 152(1) :5–13, may 1999.

[416] D.-J. Scheffers, L. J. F. Jones, and J. Errington. Several distinct localization patterns for penicillin-binding proteins in *Bacillus subtilis*. *Molecular microbiology*, 51(3) :749–64, feb 2004.

[417] F. Schneider, R. Krämer, and A. Burkovski. Identification and characterization of the main ?-alanine uptake system in *Escherichia coli*. *Applied Microbiology and Biotechnology*, 65(5) :576–82, oct 2004.

[418] D. A. Schult. Exploring network structure, dynamics, and function using NetworkX. In *Proceedings of the 7th Python in science conference*, 2008.

[419] S. A. Sciochetti, P. J. Piggot, D. J. Sherratt, and G. Blakely. The ripX locus of *Bacillus subtilis* encodes a site-specific recombinase involved in proper chromosome partitioning. *Journal of bacteriology*, 181(19) :6053–62, oct 1999.

[420] C. Scornavacca, E. Jacox, and G. J. Szöllősi. Joint amalgamation of most parsimonious reconciled gene trees. *Bioinformatics (Oxford, England)*, 31(6) :841–8, mar 2015.

[421] T. Seemann. Prokka : rapid prokaryotic genome annotation. *Bioinformatics*, 30(14) :2068–2069, jul 2014.

[422] A. Seluanov and E. Bibi. FtsY, the prokaryotic signal recognition particle receptor homologue, is essential for biogenesis of membrane proteins. *The Journal of biological chemistry*, 272(4) :2053–5, jan 1997.

[423] I. M. Shah, M.-H. Laaberki, D. L. Popham, and J. Dworkin. A Eukaryotic-like Ser/Thr Kinase Signals Bacteria to Exit Dormancy in Response to Peptidoglycan Fragments. *Cell*, 135(3) :486–496, oct 2008.

[424] L.-T. Sham, S. M. Barendt, K. E. Kopecky, and M. E. Winkler. Essential PcsB putative peptidoglycan hydrolase interacts with the essential FtsXSpn cell division protein in *Streptococcus pneumoniae* D39. *Proceedings of the National Academy of Sciences*, 108(45) :E1061–E1069, nov 2011.

[425] L.-T. Sham, E. K. Butler, M. D. Lebar, D. Kahne, T. G. Bernhardt, and N. Ruiz. MurJ is the flippase of lipid-linked precursors for peptidoglycan biogenesis. *Science*, 345(6193) :220–222, jul 2014.

[426] L.-T. Sham, H.-C. T. Tsui, A. D. Land, S. M. Barendt, and M. E. Winkler. Recent advances in pneumococcal peptidoglycan biosynthesis suggest new vaccine and antimicrobial targets. *Current Opinion in Microbiology*, 15(2) :194–203, apr 2012.

[427] P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, and T. Ideker. Cytoscape : a software environment for integrated models of biomolecular interaction networks. *Genome research*, 13(11) :2498–504, nov 2003.

[428] H. Shimodaira and M. Hasegawa. Multiple Comparisons of Log-Likelihoods with Applications to Phylogenetic Inference. *Molecular Biology and Evolution*, 16(8) :1114–1116, aug 1999.

[429] D. C. Shippy and A. A. Fadl. RNA modification enzymes encoded by the gid operon : Implications in biology and virulence of bacteria. *Microbial Pathogenesis*, 89 :100–107, dec 2015.

[430] J. K. Singh, R. D. Makde, V. Kumar, and D. Panda. SepF Increases the Assembly and Bundling of FtsZ Polymers and Stabilizes FtsZ Protofilaments by Binding along Its Length. *Journal of Biological Chemistry*, 283(45) :31116–31124, nov 2008.

[431] A. U. Sinha and J. Meller. Cinteny : flexible analysis and visualization of synteny and genome rearrangements in multiple organisms. *BMC Bioinformatics*, 8(1) :82, 2007.

[432] E. Small, R. Marrington, A. Rodger, D. J. Scott, K. Sloan, D. Roper, T. R. Dafforn, and S. G. Addinall. FtsZ polymer-bundling by the *Escherichia coli* ZapA orthologue, YgfE, involves a conformational change in bound GTP. *Journal of molecular biology*, 369(1) :210–21, may 2007.

[433] M. Smith, C. de Vries, D. Langley, G. King, and R. Wake. The *Bacillus subtilis* DNA Replication Terminator. *Journal of Molecular Biology*, 260(1) :54–69, jul 1996.

[434] P. Sobetzko, A. Travers, and G. Muskhelishvili. Gene order and chromosome dynamics coordinate spatiotemporal gene expression during the bacterial growth cycle. *Proceedings of the National Academy of Sciences of the United States of America*, 109(2) :E42–50, jan 2012.

[435] J. Soppa, K. Kobayashi, M.-F. Noirot-Gros, D. Oesterhelt, S. D. Ehrlich, E. Dervyn, N. Ogasawara, and S. Moriya. Discovery of two novel families of proteins that are

proposed to interact with prokaryotic SMC proteins, and characterization of the *Bacillus subtilis* family members ScpA and ScpB. *Molecular microbiology*, 45(1) :59–71, jul 2002.

[436] B. G. Spratt. Escherichia coli resistance to beta-lactam antibiotics through a decrease in the affinity of a target for lethality. *Nature*, 274(5672) :713–5, aug 1978.

[437] B. G. Spratt and A. B. Pardee. Penicillin-binding proteins and cell shape in *E. coli*. *Nature*, 254(5500) :516–7, apr 1975.

[438] F. Squeglia, R. Marchetti, A. Ruggiero, R. Lanzetta, D. Marasco, J. Dworkin, M. Petoukhov, A. Molinaro, R. Berisio, and A. Silipo. Chemical Basis of Peptidoglycan Discrimination by PrkC, a Key Kinase Involved in Bacterial Resuscitation from Dormancy. *Journal of the American Chemical Society*, 133(51) :20676–20679, dec 2011.

[439] A. Stamatakis. Using RAxML to Infer Phylogenies. In *Current Protocols in Bioinformatics*, volume 51, pages 6.14.1–6.14.14. John Wiley & Sons, Inc., Hoboken, NJ, USA, sep 2015.

[440] A. Stamatakis, P. Hoover, and J. Rougemont. A Rapid Bootstrap Algorithm for the RAxML Web Servers. *Systematic Biology*, 57(5) :758–771, oct 2008.

[441] A. Stamatakis, T. Ludwig, and H. Meier. RAxML-III : a fast program for maximum likelihood-based inference of large phylogenetic trees. *Bioinformatics*, 21(4) :456–463, feb 2005.

[442] G. A. Stamsås, I. S. Myrbråten, D. Straume, Z. Salehian, J.-W. Veening, L. S. Håvarstein, and M. Kjos. CozEa and CozEb play overlapping and essential roles in controlling cell division in *Staphylococcus aureus*. *Molecular Microbiology*, jun 2018.

[443] G. A. Stamsås, D. Straume, A. Ruud Winther, M. Kjos, C. A. Frantzen, and L. S. Håvarstein. Identification of EloR (Spr1851) as a regulator of cell elongation in *Streptococcus pneumoniae*. *Molecular Microbiology*, 105(6) :954–967, sep 2017.

[444] I. A. Stancik, M. S. Šestak, B. Ji, M. Axelson-Fisk, D. Franjevic, C. Jers, T. Domazet-Lošo, and I. Mijakovic. Serine/Threonine Protein Kinases from Bacteria, Archaea and Eukarya Share a Common Evolutionary Origin Deeply Rooted in the Tree of Life. *Journal of Molecular Biology*, 430(1) :27–32, 2018.

[445] M. Stouf, J.-C. Meile, and F. Cornet. FtsK actively segregates sister chromosomes in *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America*, 110(27) :11157–62, jul 2013.

[446] M. P. Strauss, A. T. F. Liew, L. Turnbull, C. B. Whitchurch, L. G. Monahan, and E. J. Harry. 3D-SIM super resolution microscopy reveals a bead-like arrangement for FtsZ and the division machinery : implications for triggering cytokinesis. *PLoS biology*, 10(9) :e1001389, 2012.

[447] M. J. Sullivan, N. K. Petty, and S. A. Beatson. Easyfig : a genome comparison visualizer. *Bioinformatics*, 27(7) :1009–1010, apr 2011.

[448] T. Suzuki and K. Miyauchi. Discovery and characterization of tRNA Ile lysidine synthetase (TilS). *FEBS Letters*, 584(2) :272–277, jan 2010.

[449] D. Szklarczyk, J. H. Morris, H. Cook, M. Kuhn, S. Wyder, M. Simonovic, A. Santos, N. T. Doncheva, A. Roth, P. Bork, L. J. Jensen, and C. von Mering. The STRING database in 2017 : quality-controlled protein-protein association networks, made broadly accessible. *Nucleic acids research*, 45(D1) :D362–D368, jan 2017.

[450] G. J. Szöllősi, W. Rosikiewicz, B. Boussau, E. Tannier, and V. Daubin. Efficient Exploration of the Space of Reconciled Gene Trees. jun 2013.

[451] H. Szurmant, M. A. Mohan, P. M. Imus, and J. A. Hoch. YycH and YycI Interact To Regulate the Essential YycFG Two-Component System in *Bacillus subtilis*. *Journal of Bacteriology*, 189(8) :3280–3289, apr 2007.

[452] A. Taghbalout, A. Landoulsi, R. Kern, M. Yamazoe, S. Hiraga, B. Holland, M. Kohiyama, and A. Malki. Competition between the replication initiator DnaA and the sequestration factor SeqA for binding to the hemimethylated chromosomal origin of *E. coli* in vitro. *Genes to cells : devoted to molecular & cellular mechanisms*, 5(11) :873–884, nov 2000.

[453] H. Takada and H. Yoshikawa. Essentiality and function of WalK/WalR two-component system : the past, present, and future of research. *Bioscience, Biotechnology, and Biochemistry*, 82(5) :741–751, may 2018.

[454] T. Takizawa, K. J. Meaburn, and T. Misteli. The meaning of gene positioning. *Cell*, 135(1) :9–13, oct 2008.

[455] J. Tamames, G. Casari, C. Ouzounis, and A. Valencia. Conserved Clusters of Functionally Related Genes in Two Bacterial Genomes. *Journal of Molecular Evolution*, 44(1) :66–73, jan 1997.

[456] J. Tamames, M. González-Moreno, J. Mingorance, A. Valencia, and M. Vicente. Bringing gene order into bacterial shape. *Trends in Genetics*, 17(3) :124–126, mar 2001.

[457] I. S. Tan and K. S. Ramamurthi. Spore formation in *Bacillus subtilis*. *Environmental microbiology reports*, 6(3) :212–25, jun 2014.

[458] R. L. Tatusov, M. Y. Galperin, D. A. Natale, and E. V. Koonin. The COG database : a tool for genome-scale analysis of protein functions and evolution. *Nucleic acids research*, 28(1) :33–6, jan 2000.

[459] S. Tavaré. Some mathematical questions in biology : DNA sequence analysis. *American Mathematical Society*, 17 :57, 1986.

[460] A. Tchigvintsev, D. Tchigvintsev, R. Flick, A. Popovic, A. Dong, X. Xu, G. Brown, W. Lu, H. Wu, H. Cui, L. Dombrowski, J. Joo, N. Beloglazova, J. Min, A. Savchenko, A. Caudy, J. Rabinowitz, A. Murzin, and A. Yakunin. Biochemical and Structural Studies of Conserved Maf Proteins Revealed Nucleotide Pyrophosphatases with a Preference for Modified Nucleotides. *Chemistry & Biology*, 20(11) :1386–1398, nov 2013.

[461] M. Thanbichler. Synchronization of Chromosome Dynamics and Cell Division in Bacteria. *Cold Spring Harb Perspect Biol*, 2(1) :a000331, 2009.

[462] M. Thanbichler and L. Shapiro. MipZ, a Spatial Regulator Coordinating Chromosome Segregation with Cell Division in *Caulobacter*. *Cell*, 126(1) :147–162, jul 2006.

[463] H. B. Thomaides, M. Freeman, M. El Karoui, and J. Errington. Division site selection protein DivIVA of *Bacillus subtilis* has a second distinct function in chromosome segregation during sporulation. *Genes & development*, 15(13) :1662–73, jul 2001.

[464] H. Tilg and A. Kaser. Gut microbiome, obesity, and metabolic dysfunction. *The Journal of clinical investigation*, 121(6) :2126–32, jun 2011.

[465] D. J. Tipper, J. L. Strominger, and J. C. Ensign. Structure of the cell wall of *Staphylococcus aureus*, strain Copenhagen. VII. Mode of action of the bacteriolytic peptidase

from Myxobacter and the isolation of intact cell wall polysaccharides. *Biochemistry*, 6(3) :906–20, mar 1967.

[466] T. Tomoyasu, T. Yuki, S. Morimura, H. Mori, K. Yamanaka, H. Niki, S. Hiraga, and T. Ogura. The *Escherichia coli* FtsH protein is a prokaryotic member of a protein family of putative ATPases involved in membrane functions, cell cycle control, and gene expression. *Journal of Bacteriology*, Mar 1993.

[467] E. Toro, S.-H. Hong, H. H. Mcadams, and L. Shapiro. Caulobacter requires a dedicated mechanism to initiate chromosome segregation.

[468] B. A. Traag, A. Pugliese, J. A. Eisen, and R. Losick. Gene Conservation among Endospore-Forming Bacteria Reveals Additional Sporulation Genes in *Bacillus subtilis*. *Journal of Bacteriology*, 195(2) :253–260, jan 2013.

[469] B. A. Traag and G. P. van Wezel. The SsgA-like proteins in actinomycetes : small proteins up to a big task. *Antonie van Leeuwenhoek*, 94(1) :85–97, jun 2008.

[470] A. Treuner-Lange, K. Aguiluz, C. van der Does, N. Gómez-Santos, A. Harms, D. Schumacher, P. Lenz, M. Hoppert, J. Kahnt, J. Muñoz-Dorado, and L. Søgaard-Andersen. PomZ, a ParA-like protein, regulates Z-ring formation and cell division in *Myxococcus xanthus*. *Molecular Microbiology*, 87(2) :235–253, jan 2013.

[471] P. J. Turnbaugh, R. E. Ley, M. Hamady, C. M. Fraser-Liggett, R. Knight, and J. I. Gordon. The Human Microbiome Project. *Nature*, 449(7164) :804–810, oct 2007.

[472] A. Typas, M. Banzhaf, C. A. Gross, and W. Vollmer. From the regulation of peptidoglycan synthesis to bacterial growth and morphology. *Nature reviews. Microbiology*, 10(2) :123–36, dec 2012.

[473] A. Typas, M. Banzhaf, B. van den Berg van Saparoea, J. Verheul, J. Biboy, R. J. Nichols, M. Zietek, K. Beilharz, K. Kannenberg, M. von Rechenberg, E. Breukink, T. den Blaauwen, C. A. Gross, and W. Vollmer. Regulation of peptidoglycan synthesis by outer-membrane proteins. *Cell*, 143(7) :1097–109, dec 2010.

[474] T. Uehara and J. T. Park. Growth of *Escherichia coli* : significance of peptidoglycan degradation during elongation and septation. *Journal of bacteriology*, 190(11) :3914–22, jun 2008.

[475] T. Uehara, K. R. Parzych, T. Dinh, and T. G. Bernhardt. Daughter cell separation is controlled by cytokinetic ring-activated cell wall hydrolysis. *The EMBO Journal*, 29(8) :1412–1422, apr 2010.

[476] T. Uehara, K. Suefuji, N. Valbuena, B. Meehan, M. Donegan, and J. T. Park. Recycling of the Anhydro- N -Acetylmuramic Acid Derived from Cell Wall Murein Involves a Two-Step Conversion to N Recycling of the Anhydro- N -Acetylmuramic Acid Derived from Cell Wall Murein Involves a Two-Step Conversion to N -Acetylglucosamine-Phosphat. *Journal of Bacteriology*, 187(11) :3643–3649, 2005.

[477] A. Ulrych, N. Holečková, J. Goldová, L. Doubravová, O. Benada, O. Kofroňová, P. Halada, and P. Branny. Characterization of pneumococcal Ser/Thr protein phosphatase phpP mutant and identification of a novel PhpP substrate, putative RNA binding protein Jag. *BMC Microbiology*, 16(1) :1–19, 2016.

[478] F. van den Ent, L. A. Amos, and J. Löwe. Prokaryotic origin of the actin cytoskeleton. *Nature*, 413(6851) :39–44, sep 2001.

[479] F. Van Den Ent, C. M. Johnson, L. Persons, P. De Boer, and J. Löwe. Bacterial actin MreB assembles in complex with cell shape protein RodZ. *EMBO Journal*, 29(6) :1081–1090, 2010.

[480] S. van Teeffelen, S. Wang, L. Furchtgott, K. C. Huang, N. S. Wingreen, J. W. Shaevitz, and Z. Gitai. The bacterial actin MreB rotates, and rotation depends on cell-wall assembly. *Proceedings of the National Academy of Sciences*, 108(38) :15822–15827, sep 2011.

[481] A. Varma and K. D. Young. In *Escherichia coli*, MreB and FtsZ direct the synthesis of lateral cell wall via independent pathways that require PBP 2. *Journal of Bacteriology*, 191(11) :3526–3533, 2009.

[482] P. Vasudevan, J. McElligott, C. Attkisson, M. Betteken, and D. L. Popham. Homologues of the *Bacillus subtilis* SpoVB Protein Are Involved in Cell Wall Metabolism. *Journal of Bacteriology*, 191(19) :6012–6019, oct 2009.

[483] P. Vasudevan, A. Weaver, E. D. Reichert, S. D. Linnstaedt, and D. L. Popham. Spore cortex formation in *Bacillus subtilis* is regulated by accumulation of peptidoglycan precursors under the control of sigma K. *Molecular microbiology*, 65(6) :1582–94, sep 2007.

[484] M. Vicente and J. Errington. Structure, function and controls in microbial division. *Molecular microbiology*, 20(1) :1–7, apr 1996.

[485] P. H. Viollier and L. Shapiro. Spatial complexity of mechanisms controlling a bacterial cell cycle. *Current Opinion in Microbiology*, 7(6) :572–578, dec 2004.

[486] W. Vollmer, D. Blanot, and M. A. De Pedro. Peptidoglycan structure and architecture. 2008.

[487] W. Vollmer, B. Joris, P. Charlier, and S. Foster. Bacterial peptidoglycan (murein) hydrolases. *FEMS Microbiology Reviews*, 32(2) :259–286, 2008.

[488] C. von Linne. *Systema naturae per regna tria naturae secundum classes, ordines, genera ... - Carolus Linnaeus - Google Livres*. 1735.

[489] C. von Mering, M. Huynen, D. Jaeggi, S. Schmidt, P. Bork, and B. Snel. STRING : a database of predicted functional associations between proteins. *Nucleic acids research*, 31(1) :258–61, jan 2003.

[490] C. von Mering, L. J. Jensen, B. Snel, S. D. Hooper, M. Krupp, M. Foglierini, N. Jouffre, M. A. Huynen, and P. Bork. STRING : known and predicted protein-protein associations, integrated and transferred across organisms. *Nucleic Acids Research*, 33(Database issue) :D433–D437, dec 2004.

[491] I. Wadenpohl and M. Bramkamp. DivIC stabilizes FtsL against RasP cleavage. *Journal of Bacteriology*, 192(19) :5260–5263, 2010.

[492] H.-C. Wang, B. Q. Minh, E. Susko, and A. J. Roger. Modeling Site Heterogeneity with Posterior Mean Site Frequency Profiles Accelerates Accurate Phylogenomic Estimation. *Systematic Biology*, 67(2) :216–235, mar 2018.

[493] J. D. Wang and P. A. Levin. Metabolism, cell growth and the bacterial cell cycle. *Nature reviews. Microbiology*, 7(11) :822–7, 2009.

[494] Q. M. Wang, R. B. Peery, R. B. Johnson, W. E. Alborn, W.-K. Yeh, and P. L. Skatrud. Identification and Characterization of a Monofunctional Glycosyltransferase from *Staphylococcus aureus*. *Journal of Bacteriology*, 183(16) :4779–4785, aug 2001.

[495] X. Wang and J. Lutkenhaus. The FtsZ protein of *Bacillus subtilis* is localized at the division site and has GTPase activity that is dependent upon FtsZ concentration. *Molecular microbiology*, 9(3) :435–42, aug 1993.

[496] X. Wang and D. J. Sherratt. Independent segregation of the two arms of the *Escherichia coli* ori region requires neither RNA synthesis nor MreB dynamics. *Journal of bacteriology*, 192(23) :6143–53, dec 2010.

[497] A. D. Warth and J. L. Strominger. Structure of the peptidoglycan from vegetative cell walls of *Bacillus subtilis*. *Biochemistry*, 10(24) :4349–58, nov 1971.

[498] W. Wehrl, M. Niederweis, and W. Schumann. The FtsH protein accumulates at the septum of *Bacillus subtilis* during cell division and sporulation. *Journal of Bacteriology*, 182(13) :3870–3873, 2000.

[499] Y. Wei, T. Havasy, D. C. McPherson, and D. L. Popham. Rod shape determination by the *Bacillus subtilis* class B penicillin-binding proteins encoded by pbpA and pbpH. *Journal of bacteriology*, 185(16) :4717–26, aug 2003.

[500] Y. Wei, D. C. McPherson, and D. L. Popham. A mother cell-specific class B penicillin-binding protein, PBP4b, in *Bacillus subtilis*. *Journal of bacteriology*, 186(1) :258–61, jan 2004.

[501] T. Weitao, S. Dasgupta, and K. Nordström. Role of the mukB gene in chromosome and plasmid partition in *Escherichia coli*. *Molecular microbiology*, 38(2) :392–400, oct 2000.

[502] S. Whelan and N. Goldman. A General Empirical Model of Protein Evolution Derived from Multiple Protein Families Using a Maximum-Likelihood Approach. *Molecular Biology and Evolution*, 18(5) :691–699, may 2001.

[503] C. L. White and J. W. Gober. MreB : pilot or passenger of cell wall synthesis ? *Trends in Microbiology*, 20(2) :74–79, feb 2012.

[504] C. L. White, A. Kitich, and J. W. Gober. Positioning cell wall synthetic complexes by the bacterial morphogenetic proteins MreB and MreD. *Molecular Microbiology*, 76(3) :616–633, mar 2010.

[505] C. Whitfield. Biosynthesis and Assembly of Capsular Polysaccharides in *Escherichia coli*. 2006.

[506] J. Wild, J. Hennig, M. Lobocka, W. Walczak, and T. Kłopotowski. Identification of the dadX gene coding for the predominant isozyme of alanine racemase in *Escherichia coli* K12. *Molecular & general genetics : MGG*, 198(2) :315–22, 1985.

[507] J. Willemse, J. W. Borst, E. de Waal, T. Bisseling, and G. P. van Wezel. Positive control of cell division : FtsZ is recruited by SsgB during sporulation of Streptomyces. *Genes & Development*, 25(1) :89–99, jan 2011.

[508] C. R. Woese and G. E. Fox. Phylogenetic structure of the prokaryotic domain : the primary kingdoms. *Proceedings of the National Academy of Sciences of the United States of America.*, Nov 1977.

[509] C. R. Woese, O. Kandler, and M. L. Wheelis. Towards a natural system of organisms : proposal for the domains Archaea, Bacteria, and Eucarya. *Proceedings of the National Academy of Sciences of the United States of America*, 87(12) :4576–9, jun 1990.

[510] M. Wolf, T. Müller, T. Dandekar, and J. D. Pollack. Phylogeny of firmicutes with special reference to mycoplasma (mollicutes) as inferred from phosphoglycerate kinase amino acid sequence data. *International journal of systematic and evolutionary microbiology.*, May 2004.

[511] L. J. Wu and J. Errington. Bacillus subtilis SpoIIIE protein required for DNA segregation during asymmetric cell division. *Science (New York, N.Y.)*, 264(5158) :572–5, apr 1994.

[512] L. J. Wu and J. Errington. RacA and the Soj-Spo0J system combine to effect polar chromosome segregation in sporulating *Bacillus subtilis*. *Molecular microbiology*, 49(6) :1463–75, sep 2003.

[513] L. J. Wu and J. Errington. Coordination of cell division and chromosome segregation by a nucleoid occlusion protein in *Bacillus subtilis*. *Cell*, 117(7) :915–25, jun 2004.

[514] B. Xayarath and J. Yother. Mutations Blocking Side Chain Assembly, Polymerization, or Transport of a Wzy-Dependent *Streptococcus pneumoniae* Capsule Are Lethal in the Absence of Suppressor Mutations and Can Affect Polymer Transfer to the Cell Wall. *Journal of Bacteriology*, 189(9) :3369–3381, may 2007.

[515] D. C. Yang, K. M. Blair, and N. R. Salama. Staying in Shape : the Impact of Cell Shape on Bacterial Survival in Diverse Environments. *Microbiology and molecular biology reviews : MMBR*, 80(1) :187–203, mar 2016.

[516] D. C. Yang, N. T. Peters, K. R. Parzych, T. Uehara, M. Markovski, and T. G. Bernhardt. An ATP-binding cassette transporter-like complex governs cell-wall hydrolysis at the bacterial cytokinetic ring. *Proceedings of the National Academy of Sciences of the United States of America*, 108(45) :E1052–60, nov 2011.

[517] X. Yang, Z. Lyu, A. Miguel, R. McQuillen, K. C. Huang, and J. Xiao. GTPase activity–coupled treadmilling of the bacterial tubulin FtsZ organizes septal cell wall synthesis. *Science*, 355(6326) :744–747, feb 2017.

[518] Z. Yang. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites : approximate methods. *Journal of molecular evolution*, 39(3) :306–14, sep 1994.

[519] P. Yarza and R. Munoz. The All-Species Living Tree Project. *Methods in Microbiology*, 41 :45–59, jan 2014.

[520] C. Yeats, R. D. Finn, and A. Bateman. The PASTA domain : a beta-lactam-binding domain. *Trends in biochemical sciences*, 27(9) :438, sep 2002.

[521] J. Yother. Capsules of *Streptococcus pneumoniae* and Other Bacteria : Paradigms for Polysaccharide Biosynthesis and Regulation. *Annual Review of Microbiology*, 65(1) :563–581, oct 2011.

[522] L. E. Zawadzke, T. D. Bugg, and C. T. Walsh. Existence of two D-alanine :D-alanine ligases in *Escherichia coli* : cloning and sequencing of the ddlA gene and purification and characterization of the DdlA and DdlB enzymes. *Biochemistry*, 30(6) :1673–82, feb 1991.

[523] W. Zhang and Z. Lu. Phylogenomic evaluation of members above the species level within the *phylum Firmicutes* based on conserved proteins. *Environmental Microbiology Reports*, 7(2) :273–281, apr 2015.

[524] J. J. Zheng, A. J. Perez, H.-C. T. Tsui, O. Massidda, and M. E. Winkler. Absence of the KhpA and KhpB (JAG/EloR) RNA-binding proteins suppresses the requirement for PBP2b by overproduction of FtsA in *Streptococcus pneumoniae* D39. *Molecular Microbiology*, 106(5) :793–814, dec 2017.

[525] T. Zhilina, G. Zavarzin, E. Bulygina, V. Kevbrin, G. Osipov, and K. Chumakov. Ecology, Physiology and Taxonomy Studies on a New Taxon of Haloanaerobiaceae, Haloincola saccharolytica gen. nov., sp. nov. *Systematic and Applied Microbiology*, 15(2) :275–284, may 1992.

[526] Y. Zhou, R. Wang, L. Li, X. Xia, and Z. Sun. Inferring Functional Linkages between Proteins from Evolutionary Scenarios. *Journal of Molecular Biology*, 359(4) :1150–1159, jun 2006.

[527] L. Zucchini, C. Mercy, P. S. Garcia, C. Cluzel, V. Gueguen-Chaignon, F. Galisson, C. Freton, S. Guiral, C. Brochier-Armanet, P. Gouet, and C. Grangeasse. PASTA repeats of the protein kinase StkP interconnect cell constriction and separation of *Streptococcus pneumoniae*. *Nature Microbiology*, 3(2) :197–209, 2018.

[528] E. Zuckerkandl and L. Pauling. Molecules as documents of evolutionary history. *Journal of theoretical biology.*, Mar 1965.

Annexes

.1 Génomes et protéomes complets de *Firmicutes* contenus dans la base de données locale.

Identifiant taxonomique	Numéro d'assemblage	Nom de la souche
931626	GCA_000247605.1	Acetobacterium woodii DSM 1030
574087	GCA_000144695.1	Acetohalobium arabaticum DSM 5501
591001	GCA_000025305.1	Acidaminococcus fermentans DSM 20731
568816	GCA_000230275.1	Acidaminococcus intestini RyC-MR95
866775	GCA_000193205.1	Aerococcus urinae ACS-120-V-Col10a
521098	GCA_000024285.1	Alicyclobacillus acidocaldarius subsp. acidocaldarius DSM 446
1048834	GCA_000219875.1	Alicyclobacillus acidocaldarius subsp. acidocaldarius Tc-4-1
293826	GCA_000016985.1	Alkaliphilus metalliredigens QYMF
350688	GCA_000018325.1	Alkaliphilus oremlandii OhILAs
1313	GCA_001255215.1	Alloactinosynnema sp. L-07
429009	GCA_000024605.1	Ammonifex degensii KC4
698758	GCA_000307165.1	Amphibacillus xylanus NBRC 15112
525919	GCA_000024105.1	Anaerococcus prevotii DSM 20548
491915	GCA_000019045.1	Anoxybacillus flavithermus WK1
198467	GCA_001187595.1	Anoxybacillus gonensis G2
1412898	GCA_000494835.1	Bacillus amyloliquefaciens CC178
692420	GCA_000196735.1	Bacillus amyloliquefaciens DSM 7
1390	GCA_001023595.1	Bacillus amyloliquefaciens G341
1091041	GCA_000242855.2	Bacillus amyloliquefaciens IT-45
1292358	GCA_000835145.1	Bacillus amyloliquefaciens KHG19
1415165	GCA_000508265.1	Bacillus amyloliquefaciens LFB112
1390	GCA_000833005.1	Bacillus amyloliquefaciens L-H15
1001582	GCA_000204275.1	Bacillus amyloliquefaciens LL3
1390	GCA_000973485.1	Bacillus amyloliquefaciens L-S60
1390	GCA_001483885.1	Bacillus amyloliquefaciens MBE1283
999891	GCA_000195515.1	Bacillus amyloliquefaciens TA208
1034836	GCA_000221645.1	Bacillus amyloliquefaciens XH7
1126211	GCA_000262385.1	Bacillus amyloliquefaciens Y2
1392	GCA_000742655.1	Bacillus anthracis 2000031021
1392	GCA_000832965.1	Bacillus anthracis 2002013094
1392	GCA_000875715.1	Bacillus anthracis A1144
1392	GCA_000830095.1	Bacillus anthracis Ames A0462
1392	GCA_000833065.1	Bacillus anthracis Ames_BA1004
1392	GCA_000832665.1	Bacillus anthracis BA1015
1392	GCA_000832725.1	Bacillus anthracis BA1035
1392	GCA_000833125.1	Bacillus anthracis Canadian Bison
1392	GCA_000747335.1	Bacillus anthracis Cvac02
1392	GCA_000725325.1	Bacillus anthracis HYU01
1392	GCA_000747375.1	Bacillus anthracis Han
1392	GCA_000832465.1	Bacillus anthracis K3
1392	GCA_000832505.1	Bacillus anthracis Ohio ACB
1392	GCA_000832425.1	Bacillus anthracis PAK-1
1392	GCA_000832585.1	Bacillus anthracis Pasteur
1392	GCA_000832745.1	Bacillus anthracis RA3
1392	GCA_000832565.1	Bacillus anthracis SK-102
1392	GCA_000832445.1	Bacillus anthracis Vollum 1B
592021	GCA_000022865.1	Bacillus anthracis str. A0248
743835	GCA_000512835.1	Bacillus anthracis str. A16
673518	GCA_000512775.1	Bacillus anthracis str. A16R
198094	GCA_000007845.1	Bacillus anthracis str. Ames
261594	GCA_000008445.1	Bacillus anthracis str. 'Ames Ancestor'
568206	GCA_000021445.1	Bacillus anthracis str. CDC 684
768494	GCA_000258885.1	Bacillus anthracis str. H9401
1392837	GCA_000583105.1	Bacillus anthracis str. SVA11
260799	GCA_000008165.1	Bacillus anthracis str. Sterne
1452727	GCA_000833275.1	Bacillus anthracis str. Turkey32
1449979	GCA_000832785.1	Bacillus anthracis str. V770-NP-1R
261591	GCA_000742895.1	Bacillus anthracis str. Vollum
720555	GCA_000165925.1	Bacillus atrophaeus 1942
1452	GCA_000830075.1	Bacillus atrophaeus NRS 1221A
1330043	GCA_000831065.1	Bacillus bombysepticus str. Wang
649639	GCA_000177235.2	Bacillus cellulosilyticus DSM 2522
572264	GCA_000022505.1	Bacillus cereus 03BB102
451709	GCA_000832865.1	Bacillus cereus 03BB108
1396	GCA_000789315.1	Bacillus cereus 03BB87
1396	GCA_000832765.1	Bacillus cereus 3a
405534	GCA_000021225.1	Bacillus cereus AH187
405535	GCA_000021785.1	Bacillus cereus AH820
222523	GCA_000008005.1	Bacillus cereus ATCC 10987
226900	GCA_000007825.1	Bacillus cereus ATCC 14579
526977	GCA_000832845.1	Bacillus cereus ATCC 4342
405532	GCA_000021205.1	Bacillus cereus B4264
1454382	GCA_000832385.1	Bacillus cereus D17
288681	GCA_000011625.1	Bacillus cereus E33L
347495	GCA_000239195.1	Bacillus cereus F837/76
1396	GCA_000832525.1	Bacillus cereus FM1
1396	GCA_000978375.1	Bacillus cereus FORC_005
1217984	GCA_000292415.1	Bacillus cereus FRI-35
1396	GCA_000724585.1	Bacillus cereus FT9
269801	GCA_000832805.1	Bacillus cereus G9241
405531	GCA_000021305.1	Bacillus cereus G9842
334406	GCA_000283675.1	Bacillus cereus NC7401
1396	GCA_001277915.1	Bacillus cereus NJ-W
361100	GCA_000013065.1	Bacillus cereus Q1
1396	GCA_000835185.1	Bacillus cereus S2-8
637380	GCA_000143605.1	Bacillus cereus biovar anthracis str. CI
79880	GCA_000737305.2	Bacillus clausii ENTPro
66692	GCA_000009825.1	Bacillus clausii KSM-K16
941639	GCA_000217835.1	Bacillus coagulans 2-6
345219	GCA_000169195.2	Bacillus coagulans 36D1
1121088	GCA_000832905.1	Bacillus coagulans DSM 1 = ATCC 7050
1398	GCA_000876545.1	Bacillus coagulans HM-08

1398	GCA_001039495.1	Bacillus coagulans S-lac
315749	GCA_000017425.1	Bacillus cytotoxicus NVH 391-98
135735	GCA_000972245.2	Bacillus endophyticus Hbe603
272558	GCA_000011145.1	Bacillus halodurans C-125
1367477	GCA_000473245.1	Bacillus infantis NRRL B-14911
1246626	GCA_000706725.1	Bacillus lehensis G1
279010	GCA_000008425.1	Bacillus licheniformis DSM 13 = ATCC 14580
592022	GCA_000025805.1	Bacillus megaterium DSM 319
1348623	GCA_000832985.1	Bacillus megaterium NBRC 15308 = ATCC 14581
1452722	GCA_001050455.1	Bacillus megaterium Q3
545693	GCA_000025825.1	Bacillus megaterium QM B1551
1006007	GCA_000225265.1	Bacillus megaterium WSH-002
796606	GCA_000724485.1	Bacillus methanolicus MGA3
1405	GCA_000742855.1	Bacillus mycoides 219298
1405	GCA_000832605.1	Bacillus mycoides ATCC 6462
766760	GCA_000408885.1	Bacillus paralicheniformis ATCC 9945a
1648923	GCA_000876525.1	Bacillus paralicheniformis BL-09
398511	GCA_000005825.2	Bacillus pseudofirmus OF4
1408	GCA_000590455.1	Bacillus pumilus B6033
1408	GCA_001191605.1	Bacillus pumilus GR-8
1408	GCA_001431145.1	Bacillus pumilus NJ-M2
1408	GCA_001431785.1	Bacillus pumilus NJ-V2
315750	GCA_000017885.2	Bacillus pumilus SAFR-032
1408	GCA_000972685.1	Bacillus pumilus W3
439292	GCA_000093085.1	[Bacillus] selenitireducens MLS10
1479	GCA_001050115.1	Bacillus smithii DSM 4216
666686	GCA_000242895.3	Bacillus sp. 1NLA3E
1570330	GCA_000827045.1	Bacillus sp. BH072
1127744	GCA_000259365.1	Bacillus sp. JS
1628753	GCA_000978495.1	Bacillus sp. LM 4-2
98228	GCA_000829195.1	Bacillus sp. OxB-1
1446792	GCA_000815145.1	Bacillus sp. Pc3
756828	GCA_000800825.1	Bacillus sp. WP8
1565991	GCA_000747345.1	Bacillus sp. X1(2014)
1574141	GCA_000877815.1	Bacillus sp. YP1
1423	GCA_000772125.1	Bacillus subtilis ATCC 13952
1423	GCA_000772165.1	Bacillus subtilis ATCC 19217
1204342	GCA_000523045.1	Bacillus subtilis BEST7003
1204343	GCA_000328745.1	Bacillus subtilis BEST7613
1409	GCA_000952895.1	Bacillus subtilis BS34A
1423	GCA_000953615.1	Bacillus subtilis BS49
936156	GCA_000186745.1	Bacillus subtilis BSn5
1423	GCA_000772205.1	Bacillus subtilis Bs-916
1453990	GCA_000973605.1	Bacillus subtilis HJ5
1136873	GCA_000971925.1	Bacillus subtilis KCTC 1028
1415167	GCA_000497485.1	Bacillus subtilis PY79
1220533	GCA_000293765.1	Bacillus subtilis QB928
1423	GCA_000782835.1	Bacillus subtilis SG6
1423	GCA_000959025.1	Bacillus subtilis T30
1423	GCA_001037985.1	Bacillus subtilis TO-A JPC
1423	GCA_001015095.1	Bacillus subtilis UD1022
1233100	GCA_000338735.1	Bacillus subtilis XF-1
645657	GCA_000209795.2	Bacillus subtilis subsp. natto BEST195
96241	GCA_000816805.1	Bacillus subtilis subsp. spizizenii
1052585	GCA_000227465.1	Bacillus subtilis subsp. spizizenii TU-B-10
655816	GCA_000146565.1	Bacillus subtilis subsp. spizizenii str. W23
135461	GCA_000827065.1	Bacillus subtilis subsp. subtilis
1147161	GCA_000344745.1	Bacillus subtilis subsp. subtilis 6051-HGW
224308	GCA_000009045.1	Bacillus subtilis subsp. subtilis str. 168
1221328	GCA_000699525.1	Bacillus subtilis subsp. subtilis str. AG1839
1302650	GCA_000349795.1	Bacillus subtilis subsp. subtilis str. BAB-1
1192196	GCA_000321395.1	Bacillus subtilis subsp. subtilis str. BSP1
1232554	GCA_000699465.1	Bacillus subtilis subsp. subtilis str. JH642
1404258	GCA_000706705.1	Bacillus subtilis subsp. subtilis str. OH 131.1
1052588	GCA_000227485.1	Bacillus subtilis subsp. subtilis str. RO-NN-1
1428	GCA_000833085.1	Bacillus thuringiensis 97-27
412694	GCA_000832885.1	Bacillus thuringiensis Al Hakam
714359	GCA_000092165.1	Bacillus thuringiensis BMB171
527021	GCA_000306745.1	Bacillus thuringiensis Bt407
1428	GCA_001455345.1	Bacillus thuringiensis CTC
1452729	GCA_000835025.1	Bacillus thuringiensis HD1002
1428	GCA_000832485.1	Bacillus thuringiensis HD1011
1428	GCA_000832825.1	Bacillus thuringiensis HD571
1428	GCA_000832925.1	Bacillus thuringiensis HD682
1218175	GCA_000292455.1	Bacillus thuringiensis HD-771
1217737	GCA_000292705.1	Bacillus thuringiensis HD-789
1428	GCA_001182785.1	Bacillus thuringiensis HS18-1
1195464	GCA_000300475.1	Bacillus thuringiensis MC28
1428	GCA_000774075.2	Bacillus thuringiensis XL6
529122	GCA_000497525.2	Bacillus thuringiensis YBT-1518
1428	GCA_001017635.1	Bacillus thuringiensis YC-10
1428	GCA_001420855.1	Bacillus thuringiensis YWC2-8
541229	GCA_000193355.1	Bacillus thuringiensis serovar chinensis CT-43
930170	GCA_000190515.1	Bacillus thuringiensis serovar finitimus YBT-020
29338	GCA_000803665.1	Bacillus thuringiensis serovar galleriae
180850	GCA_001183785.1	Bacillus thuringiensis serovar indiana
281309	GCA_000008505.1	[Bacillus thuringiensis] serovar konkukian str. 97-27
1261129	GCA_000717535.1	Bacillus thuringiensis serovar kurstaki str. HD-1
1279365	GCA_000338755.1	Bacillus thuringiensis serovar kurstaki str. HD73
570416	GCA_000688795.1	Bacillus thuringiensis serovar kurstaki str. YBT-1520
1441	GCA_000940785.1	Bacillus thuringiensis serovar morrisoni
1286404	GCA_000341665.1	Bacillus thuringiensis serovar thuringiensis str. IS5056
412694	GCA_000015065.1	Bacillus thuringiensis str. Al Hakam
1415784	GCA_000496285.1	Bacillus toyonensis BCT-7112

1225788	GCA_000319475.1	Bacillus velezensis AS43.3
1114958	GCA_000283695.1	Bacillus velezensis CAU B946
326423	GCA_000015785.1	Bacillus velezensis FZB42
492670	GCA_000987825.1	Bacillus velezensis JJ-D34
492670	GCA_000769555.1	Bacillus velezensis JS25R
1385727	GCA_000493375.1	Bacillus velezensis NAU-B3
1458206	GCA_000973585.1	Bacillus velezensis NJN-6
1423138	GCA_000685725.1	Bacillus velezensis SQR9
1449088	GCA_000583065.1	Bacillus velezensis TrigoCor1448
1338518	GCA_000455565.1	Bacillus velezensis UCMB5033
1150475	GCA_000341875.1	Bacillus velezensis UCMB5036
1150476	GCA_000455585.1	Bacillus velezensis UCMB5113
1155777	GCA_000284395.1	Bacillus velezensis YAU B9601-Y2
492670	GCA_000988345.1	Bacillus velezensis YJ11-1-4
315730	GCA_000018825.1	Bacillus weihenstephanensis KBAB4
86662	GCA_000775975.1	Bacillus weihenstephanensis WSBC10204
358681	GCA_000010165.1	Brevibacillus brevis NBRC 100599
1042163	GCA_000219535.3	Brevibacillus laterosporus LMG 15441
515622	GCA_000145035.1	Butyrivibrio proteoclasticus B316
273068	GCA_000007085.1	Caldanaerobacter subterraneus subsp. tengcongensis MB4
521460	GCA_000022325.1	Caldicellulosiruptor bescii DSM 6725
632292	GCA_000166355.1	Caldicellulosiruptor hydrothermalis 108
632335	GCA_000166695.1	Caldicellulosiruptor kristjanssonii I77R1B
632348	GCA_000166775.1	Caldicellulosiruptor kronotskyensis 2002
632516	GCA_000193435.3	Caldicellulosiruptor lactoaceticus 6A
608506	GCA_000145215.1	Caldicellulosiruptor obsidiansis OB47
632518	GCA_000166335.1	Caldicellulosiruptor owensensis OL
351627	GCA_000016545.1	Caldicellulosiruptor saccharolyticus DSM 8903
1029718	GCA_000270205.1	Candidatus Arthromitus sp. SFB-mouse-Japan
1508644	GCA_000709435.1	Candidatus Arthromitus sp. SFB-mouse-NL
1041809	GCA_000284435.1	Candidatus Arthromitus sp. SFB-mouse-Yit
1041504	GCA_000283555.1	Candidatus Arthromitus sp. SFB-rat-Yit
477974	GCA_000018425.1	Candidatus Desulforudis audaxviator MP104C
246194	GCA_000012865.1	Carboxydothermus hydrogenoformans Z-2901
1266845	GCA_000493735.1	Carnobacterium inhibens subsp. gilichinskyi
1234679	GCA_000317975.2	Carnobacterium maltaromaticum LMA28
208596	GCA_000195575.1	Carnobacterium sp. 17-4
1564681	GCA_001483965.1	Carnobacterium sp. CP1
84022	GCA_001042715.1	Clostridium aceticum DSM 1496
272562	GCA_000008765.1	Clostridium acetobutylicum ATCC 824
991791	GCA_000218855.1	Clostridium acetobutylicum DSM 1731
863638	GCA_000191905.1	Clostridium acetobutylicum EA 2018
1128398	GCA_000299355.1	[Clostridium] acidurici 9a
1341692	GCA_000484505.1	Clostridium autoethanogenum DSM 10061
1415775	GCA_000789395.1	Clostridium baratii str. Sullivan
864803	GCA_000767745.1	Clostridium beijerinckii ATCC 35702
290402	GCA_000016965.1	Clostridium beijerinckii NCIMB 8052
1216932	GCA_000577895.1	Clostridium bornimense M2/40
1491	GCA_000827935.1	Clostridium botulinum
1491	GCA_000829015.1	Clostridium botulinum 111
1415774	GCA_000789355.1	Clostridium botulinum 202F
536232	GCA_000022765.1	Clostridium botulinum A2 str. Kyoto
498214	GCA_000019545.1	Clostridium botulinum A3 str. Loch Maree
441770	GCA_000017025.1	Clostridium botulinum A str. ATCC 19397
413999	GCA_000063585.1	Clostridium botulinum A str. ATCC 3502
441771	GCA_000017045.1	Clostridium botulinum A str. Hall
498213	GCA_000019305.1	Clostridium botulinum B1 str. Okra
929506	GCA_000204565.1	Clostridium botulinum BKT015925
935198	GCA_000020165.1	Clostridium botulinum B str. Eklund 17B (NRP)
515621	GCA_000020345.1	Clostridium botulinum Ba4 str. 657
1408283	GCA_000817935.1	Clostridium botulinum CDC_1436
1408285	GCA_000816945.1	Clostridium botulinum CDC_297
508767	GCA_000020285.1	Clostridium botulinum E3 str. Alaska E43
758678	GCA_000092345.1	Clostridium botulinum F str. 230613
441772	GCA_000017065.1	Clostridium botulinum F str. Langeland
941968	GCA_000253195.1	Clostridium botulinum H04402 065
1491	GCA_000827955.1	Clostridium botulinum NCTC 8550
1492	GCA_001465175.1	Clostridium butyricum JKY6D1
1492	GCA_001456065.2	Clostridium butyricum KNU-L09
536227	GCA_001038625.1	Clostridium carboxidivorans P7
394503	GCA_000022065.1	[Clostridium] cellulolyticum H10
29343	GCA_000953215.1	[Clostridium] cellulosi
573061	GCA_000145275.1	Clostridium cellulovorans 743B
720554	GCA_000237085.1	[Clostridium] clariflavum DSM 19732
431943	GCA_000016505.1	Clostridium kluyveri DSM 555
583346	GCA_000010265.1	Clostridium kluyveri NBRC 12016
642492	GCA_000178835.2	Clostridium lentocellum DSM 5427
748727	GCA_000143685.1	Clostridium ljungdahlii DSM 13528
386415	GCA_000014125.1	Clostridium novyi NT
86416	GCA_000389635.1	Clostridium pasteurianum BC1
1262449	GCA_000807175.1	Clostridium pasteurianum DSM 525 = ATCC 6013
1428454	GCA_000506785.2	Clostridium pasteurianum NRRL B-598
195103	GCA_000013285.1	Clostridium perfringens ATCC 13124
1502	GCA_001304735.1	Clostridium perfringens FORC_003
289380	GCA_000013845.1	Clostridium perfringens SM101
195102	GCA_000009685.1	Clostridium perfringens str. 13
1345695	GCA_000473995.1	Clostridium saccharobutylicum DSM 13864
610130	GCA_000144625.1	[Clostridium] saccharolyticum WM1
931276	GCA_000340885.1	Clostridium saccharoperbutylacetonicum N1-4(HMT)
1548	GCA_000968375.1	Clostridium scatologenes ATCC 25775
755731	GCA_000244875.1	Clostridium sp. BNL1100
1042156	GCA_000270305.1	Clostridium sp. SY8519
1509	GCA_001020205.1	Clostridium sporogenes DSM 795
1509	GCA_000973705.1	Clostridium sporogenes NCIMB 10696

1121335	GCA_000331995.1	[Clostridium] stercorarium subsp. stercorarium DSM 8532
1511	GCA_000196455.1	[Clostridium] sticklandii DSM 519
1231072	GCA_000967115.1	Clostridium tetani 12124569
212717	GCA_000007625.1	Clostridium tetani E88
309798	GCA_000020945.1	Coprothermobacter proteolyticus DSM 5265
871738	GCA_000512895.1	Dehalobacter restrictus DSM 9455
1131462	GCA_000305815.1	Dehalobacter sp. CF
1147129	GCA_000305775.1	Dehalobacter sp. DCA
756499	GCA_000243155.3	Desulfitobacterium dehalogenans ATCC 51507
871963	GCA_000243135.3	Desulfitobacterium dichloroeliminans LMG P-21439
272564	GCA_000021925.1	Desulfitobacterium hafniense DCB-2
138119	GCA_000010045.1	Desulfitobacterium hafniense Y51
871968	GCA_000231405.3	Desulfitobacterium metallireducens DSM 15288
646529	GCA_000255115.3	Desulfosporosinus acidiphilus SJ4
768704	GCA_000231385.3	Desulfosporosinus meridiei DSM 13257
768706	GCA_000235605.1	Desulfosporosinus orientis DSM 765
485916	GCA_000024205.1	Desulfotomaculum acetoxidans DSM 771
767817	GCA_000233715.3	Desulfotomaculum gibsoniae DSM 7213
760568	GCA_000214705.1	Desulfotomaculum kuznetsovii DSM 6115
868595	GCA_000214435.1	Desulfotomaculum nigrificans CO-1-SRB
349161	GCA_000016165.1	Desulfotomaculum reducens MI-1
696281	GCA_000215085.1	Desulfotomaculum ruminis DSM 2154
565655	GCA_000157355.2	Enterococcus casseliflavus EC20
53345	GCA_001267865.1	Enterococcus durans KLDS 6.0930
53345	GCA_001267395.1	Enterococcus durans KLDS 6.0933
936153	GCA_000211255.1	Enterococcus faecalis 62
1201292	GCA_000742975.1	Enterococcus faecalis ATCC 29212
1206105	GCA_000281195.1	Enterococcus faecalis D32
1287066	GCA_000550745.1	Enterococcus faecalis DENG1
474186	GCA_000172575.2	Enterococcus faecalis OG1RF
226185	GCA_000007785.1	Enterococcus faecalis V583
1261557	GCA_000317915.1	Enterococcus faecalis str. Symbioflor 1
1352	GCA_001298485.1	Enterococcus faecium 64/3
1155766	GCA_000250945.1	Enterococcus faecium Aus0004
1305849	GCA_000444405.1	Enterococcus faecium Aus0085
333849	GCA_000174395.2	Enterococcus faecium DO
1104325	GCA_000336405.1	Enterococcus faecium NRRL B-2354
1344042	GCA_000737555.1	Enterococcus faecium T110
1352	GCA_001412695.1	Enterococcus faecium UW7606x64/3 TC1
768486	GCA_000271405.2	Enterococcus hirae ATCC 9790
1300150	GCA_000504125.1	Enterococcus mundtii QU 25
332949	GCA_001465115.1	Enterococcus silesiacus LMG 23085
1313290	GCA_000404205.1	Erysipelothrix rhusiopathiae SY1027
650150	GCA_000270085.1	Erysipelothrix rhusiopathiae str. Fujisawa
663278	GCA_000178115.2	Ethanoligenens harbinense YUAN-3
1286171	GCA_000597865.1	Eubacterium acidaminophilum DSM 3953
515620	GCA_000146185.1	[Eubacterium] eligens ATCC 27750
903814	GCA_000152245.2	Eubacterium limosum KIST612
1736	GCA_001481725.1	Eubacterium limosum SA11
515619	GCA_000020605.1	[Eubacterium rectale] ATCC 33656
888727	GCA_001189495.1	Eubacterium sulci ATCC 35585
1087448	GCA_000299435.1	Exiguobacterium antarcticum B7
262543	GCA_000019905.1	Exiguobacterium sibiricum 255-15
360911	GCA_000023045.1	Exiguobacterium sp. AT1b
1399115	GCA_000496635.1	Exiguobacterium sp. MH3
546269	GCA_000163895.2	Filifactor alocis ATCC 35896
334413	GCA_000010185.1	Finegoldia magna ATCC 29328
235909	GCA_000009785.1	Geobacillus kaustophilus HTA426
1629723	GCA_001028085.1	Geobacillus sp. 12AMOR1
691437	GCA_000092445.1	Geobacillus sp. C56-T3
1233873	GCA_000336445.1	Geobacillus sp. GHH01
1345697	GCA_000445995.2	Geobacillus sp. JF8
1519377	GCA_001191625.1	Geobacillus sp. LC300
471223	GCA_000023385.1	Geobacillus sp. WCH70
550542	GCA_000174795.2	Geobacillus sp. Y412MC52
544556	GCA_000024705.1	Geobacillus sp. Y412MC61
581103	GCA_000166075.1	Geobacillus sp. Y4.1MC1
272567	GCA_001274575.1	Geobacillus stearothermophilus 10
420246	GCA_000015745.1	Geobacillus thermodenitrificans NG80-2
634956	GCA_000178395.2	Geobacillus thermoglucosidasius C56-YS93
1426	GCA_001295365.1	Geobacillus thermoglucosidasius DSM 2542
1111068	GCA_000236605.1	Geobacillus thermoleovorans CCB_US3_UF5
656519	GCA_000166415.1	Halanaerobium hydrogeniformans missing
572479	GCA_000165465.1	Halanaerobium praevalens DSM 2228
866895	GCA_000284515.1	Halobacillus halophilus DSM 2266
748449	GCA_000328625.1	Halobacteroides halobius DSM 5150
373903	GCA_000020485.1	Halothermothrix orenii H 168
498761	GCA_000019165.1	Heliobacterium modesticaldum Ice1
1679721	GCA_001298655.2	Herbinix sp. SD1D
1297617	GCA_001454945.1	Intestinimonas butyriciproducens AF211
1508404	GCA_000818095.1	Jeotgalibacillus malaysiensis
1461582	GCA_000756715.2	Jeotgalicoccus sp. 13MG44_air
562970	GCA_000092905.1	Kyrpidia tusciae DSM 2912
357809	GCA_000018685.1	Lachnoclostridium phytofermentans ISDg
1600	GCA_001042405.1	Lactobacillus acetotolerans NBRC 13120
1604	GCA_000191545.1	Lactobacillus acidophilus 30SC
1579	GCA_000934625.1	Lactobacillus acidophilus FSI4
1314884	GCA_000389675.2	Lactobacillus acidophilus La-14
272621	GCA_000011985.1	Lactobacillus acidophilus NCFM
695562	GCA_000194115.1	Lactobacillus amylovorus GRL1118
387344	GCA_000014465.1	Lactobacillus brevis ATCC 367
1001583	GCA_000359625.1	Lactobacillus brevis KB290
1071400	GCA_000298115.2	Lactobacillus buchneri CD034
511437	GCA_000211375.1	Lactobacillus buchneri NRRL B-30929

1051650	GCA_000309565.2	Lactobacillus casei 12A
998820	GCA_000194765.1	Lactobacillus casei BD-II
543734	GCA_000026485.1	Lactobacillus casei BL23
999378	GCA_000194785.1	Lactobacillus casei LC2W
1318635	GCA_000418515.1	Lactobacillus casei LOCK919
1215914	GCA_000318035.1	Lactobacillus casei W56
498216	GCA_000019245.3	Lactobacillus casei str. Zhang
219334	GCA_000829055.1	Lactobacillus casei subsp. casei ATCC 393
1579	GCA_001469775.1	Lactobacillus delbrueckii subsp. bulgaricus
353496	GCA_000191165.1	Lactobacillus delbrueckii subsp. bulgaricus 2038
390333	GCA_000056065.1	Lactobacillus delbrueckii subsp. bulgaricus ATCC 11842 = JCM 1002
321956	GCA_000014405.1	Lactobacillus delbrueckii subsp. bulgaricus ATCC BAA-365
767455	GCA_000182835.1	Lactobacillus delbrueckii subsp. bulgaricus ND02
1381124	GCA_000466785.3	Lactobacillus fermentum 3872
712938	GCA_000210515.1	Lactobacillus fermentum CECT 5716
767453	GCA_000397165.1	Lactobacillus fermentum F-6
334390	GCA_000010145.1	Lactobacillus fermentum IFO 3956
52242	GCA_001314245.1	Lactobacillus gallinarum HFD4
1403312	GCA_000814885.1	Lactobacillus gasseri 130918
324831	GCA_000014425.1	Lactobacillus gasseri ATCC 33323 = JCM 1131
1007676	GCA_001050475.1	Lactobacillus ginsenosidimutans EMML 3041
1074467	GCA_000831645.3	Lactobacillus heilongjiangensis DSM 28069
1587	GCA_001308285.1	Lactobacillus helveticus CAUH18
326425	GCA_000422165.1	Lactobacillus helveticus CNRZ32
405566	GCA_000015385.1	Lactobacillus helveticus DPC 4571
767462	GCA_000189515.1	Lactobacillus helveticus H10
767456	GCA_000525715.1	Lactobacillus helveticus H9
1587	GCA_000961015.1	Lactobacillus helveticus KLDS1.8701
1587	GCA_001006025.1	Lactobacillus helveticus MB2-1
880633	GCA_000165775.3	Lactobacillus helveticus R0052
1291742	GCA_000829395.1	Lactobacillus hokkaidonensis JCM 18461
909954	GCA_000204985.1	Lactobacillus johnsonii DPC 6026
633699	GCA_000091405.1	Lactobacillus johnsonii FI9785
1408186	GCA_000498675.1	Lactobacillus johnsonii N6.2
257314	GCA_000008065.1	Lactobacillus johnsonii NCC 533
1033837	GCA_000214785.1	Lactobacillus kefiranofaciens ZW3
637971	GCA_001050435.1	Lactobacillus koreensis 26-25
148814	GCA_001314945.1	Lactobacillus kunkeei MP2
1130798	GCA_000248095.3	Lactobacillus mucosae LM1
321967	GCA_000014525.1	Lactobacillus paracasei ATCC 334
1597	GCA_001191565.1	Lactobacillus paracasei CAUH35
1597	GCA_001244395.1	Lactobacillus paracasei L9
1446494	GCA_000582665.1	Lactobacillus paracasei N1115
537973	GCA_000155515.2	Lactobacillus paracasei subsp. paracasei 8700:2
1226298	GCA_000829035.1	Lactobacillus paracasei subsp. paracasei JCM 8130
60520	GCA_001443645.1	Lactobacillus paraplantarum L-ZS9
1327988	GCA_000412205.1	Lactobacillus plantarum 16
1590	GCA_001278015.1	Lactobacillus plantarum 5-2
1590	GCA_000931425.1	Lactobacillus plantarum B21
1300221	GCA_000604105.1	Lactobacillus plantarum DOMLa
1590	GCA_001302645.1	Lactobacillus plantarum HFC8
644042	GCA_000023085.1	Lactobacillus plantarum JDM1
1590	GCA_001484005.1	Lactobacillus plantarum LZ95
220668	GCA_000203855.3	Lactobacillus plantarum WCFS1
1284663	GCA_000338115.2	Lactobacillus plantarum ZJ316
1590	GCA_001296095.1	Lactobacillus plantarum ZS2058
767468	GCA_000392485.2	Lactobacillus plantarum subsp. plantarum P-8
889932	GCA_000148815.2	Lactobacillus plantarum subsp. plantarum ST-III
557436	GCA_000016825.1	Lactobacillus reuteri DSM 20016
1340495	GCA_000410995.1	Lactobacillus reuteri I5007
1598	GCA_001046835.1	Lactobacillus reuteri IRT
557433	GCA_000010005.1	Lactobacillus reuteri JCM 1112
491077	GCA_000159455.2	Lactobacillus reuteri SD2112
1358027	GCA_000439275.1	Lactobacillus reuteri TD1
1088720	GCA_000233755.1	Lactobacillus rhamnosus ATCC 8530
568703	GCA_000026505.1	Lactobacillus rhamnosus GG
1316933	GCA_000418475.1	Lactobacillus rhamnosus LOCK900
1318634	GCA_000418495.1	Lactobacillus rhamnosus LOCK908
568704	GCA_000026525.1	Lactobacillus rhamnosus Lc 705
1069534	GCA_000224985.1	Lactobacillus ruminis ATCC 27782
314315	GCA_000026065.1	Lactobacillus sakei subsp. sakei 23K
712961	GCA_000143435.1	Lactobacillus salivarius CECT 5713
1624	GCA_000758365.1	Lactobacillus salivarius JCM1046
362948	GCA_000008925.1	Lactobacillus salivarius UCC118
1194971	GCA_001011095.1	Lactobacillus salivarius str. Ren
714313	GCA_000225325.1	Lactobacillus sanfranciscensis TMW 1.1304
1545702	GCA_000761135.1	Lactobacillus sp. wkB8
420889	GCA_000269925.1	Lactococcus garvieae ATCC 49156
420890	GCA_000269945.1	Lactococcus garvieae Lg2
1358	GCA_000761115.1	Lactococcus lactis AI06
1104322	GCA_000236475.1	Lactococcus lactis subsp. cremoris A76
1295826	GCA_000468955.1	Lactococcus lactis subsp. cremoris KW2
416870	GCA_000009425.1	Lactococcus lactis subsp. cremoris MG1363
746361	GCA_000143205.1	Lactococcus lactis subsp. cremoris NZ9000
272622	GCA_000014545.1	Lactococcus lactis subsp. cremoris SK11
1111678	GCA_000312685.1	Lactococcus lactis subsp. cremoris UC509.9
1360	GCA_000807375.1	Lactococcus lactis subsp. lactis
929102	GCA_000192705.1	Lactococcus lactis subsp. lactis CV56
1046624	GCA_000344575.1	Lactococcus lactis subsp. lactis IO-1
272623	GCA_000006865.1	Lactococcus lactis subsp. lactis Il1403
684738	GCA_000025045.1	Lactococcus lactis subsp. lactis KF147
1399116	GCA_000479375.2	Lactococcus lactis subsp. lactis KLDS 4.0325
1117941	GCA_000478255.2	Lactococcus lactis subsp. lactis NCDO 2118
297352	GCA_000981525.1	Lactococcus piscium MKFS47

1229758	GCA_000300135.1	Leuconostoc carnosum JB16
349519	GCA_000026405.1	Leuconostoc citreum KM20
1229756	GCA_000298875.1	Leuconostoc gelidum JB7
762550	GCA_000196855.1	Leuconostoc gelidum subsp. gasicomitatum LMG 18811
762051	GCA_000092505.1	Leuconostoc kimchii IMSNU 11154
427140	GCA_000512955.1	Leuconostoc mesenteroides KFRI-MG
33966	GCA_001047695.1	Leuconostoc mesenteroides subsp. dextranicum
203120	GCA_000014445.1	Leuconostoc mesenteroides subsp. mesenteroides ATCC 8293
1107880	GCA_000234825.3	Leuconostoc mesenteroides subsp. mesenteroides J18
979982	GCA_000219785.1	Leuconostoc sp. C2
1457190	GCA_000565155.1	Listeria ivanovii WSLC3009
202751	GCA_000763515.1	Listeria ivanovii subsp. ivanovii
881621	GCA_000252975.1	Listeria ivanovii subsp. ivanovii PAM 55
202752	GCA_000763475.1	Listeria ivanovii subsp. londoniensis
1639	GCA_000600015.1	Listeria monocytogenes
1126011	GCA_000258905.1	Listeria monocytogenes 07PF0776
653938	GCA_000093125.2	Listeria monocytogenes 08-5578
637381	GCA_000022925.1	Listeria monocytogenes 08-5923
393133	GCA_000168695.2	Listeria monocytogenes 10403S
1288295	GCA_000577745.1	Listeria monocytogenes 6179
882095	GCA_000307025.1	Listeria monocytogenes ATCC 19117
1639	GCA_000438605.1	Listeria monocytogenes C1-387
1639	GCA_000681515.1	Listeria monocytogenes CFSAN006122
1639	GCA_001005985.1	Listeria monocytogenes CFSAN007956
1639	GCA_001005925.1	Listeria monocytogenes CFSAN008100
1639	GCA_001047715.1	Listeria monocytogenes CFSAN023463
1334565	GCA_000582845.1	Listeria monocytogenes EGD
393126	GCA_000168575.2	Listeria monocytogenes FSL R2-561
393127	GCA_000168595.2	Listeria monocytogenes Finland 1998
552536	GCA_000021185.1	Listeria monocytogenes HCC23
393130	GCA_000168635.2	Listeria monocytogenes J0161
1639	GCA_000438665.1	Listeria monocytogenes J1-108
930781	GCA_000195395.4	Listeria monocytogenes J1-220
1639	GCA_000438705.2	Listeria monocytogenes J1776
930782	GCA_000195435.4	Listeria monocytogenes J1816
1639	GCA_000438725.2	Listeria monocytogenes J1817
1639	GCA_000438745.2	Listeria monocytogenes J1926
1639	GCA_000438645.1	Listeria monocytogenes J2-031
1639	GCA_000438625.1	Listeria monocytogenes J2-064
1639	GCA_001027085.1	Listeria monocytogenes L1846
1639	GCA_001027065.1	Listeria monocytogenes L2074
1639	GCA_001027165.1	Listeria monocytogenes L2624
1639	GCA_001027125.1	Listeria monocytogenes L2625
1639	GCA_001027245.1	Listeria monocytogenes L2626
1639	GCA_001027205.1	Listeria monocytogenes L2676
882094	GCA_000307085.1	Listeria monocytogenes L312
1639	GCA_001188655.1	Listeria monocytogenes LM850658
1639	GCA_000382925.1	Listeria monocytogenes La111
1639	GCA_000746625.1	Listeria monocytogenes Lm60
1639	GCA_001483425.1	Listeria monocytogenes Lm 3136
1639	GCA_001483405.1	Listeria monocytogenes Lm 3163
1639	GCA_001483445.1	Listeria monocytogenes Lm N1546
1030009	GCA_000218305.1	Listeria monocytogenes M7
1639	GCA_000438685.2	Listeria monocytogenes N1-011A
1639	GCA_000950775.1	Listeria monocytogenes N2306
1234142	GCA_000382945.1	Listeria monocytogenes N53-1
1639	GCA_000800335.1	Listeria monocytogenes NTSN
1639	GCA_000438585.1	Listeria monocytogenes R2-502
1437838	GCA_000613085.1	Listeria monocytogenes R479a
882097	GCA_000307065.1	Listeria monocytogenes SLCC2376
879088	GCA_000307615.1	Listeria monocytogenes SLCC2378
882020	GCA_000307005.1	Listeria monocytogenes SLCC2479
879089	GCA_000306905.1	Listeria monocytogenes SLCC2540
882096	GCA_000307045.1	Listeria monocytogenes SLCC5850
879090	GCA_000306985.1	Listeria monocytogenes SLCC7179
1457187	GCA_000568475.1	Listeria monocytogenes WSLC1001
1457188	GCA_000568935.1	Listeria monocytogenes WSLC1042
1639	GCA_001454845.1	Listeria monocytogenes WSLC 1018
1639	GCA_001454865.1	Listeria monocytogenes WSLC 1019
1639	GCA_001454885.1	Listeria monocytogenes WSLC 1020
1639	GCA_001454925.1	Listeria monocytogenes WSLC 1047
932919	GCA_000197755.2	Listeria monocytogenes serotype 1/2b str. SLCC2755
932920	GCA_000210815.2	Listeria monocytogenes serotype 1/2c str. SLCC2372
563174	GCA_000209755.1	Listeria monocytogenes serotype 4a str. L99
568819	GCA_000026705.1	Listeria monocytogenes serotype 4b str. CLIP 80459
265669	GCA_000008285.1	Listeria monocytogenes serotype 4b str. F2365
1230340	GCA_000318055.1	Listeria monocytogenes serotype 4b str. LL195
683837	GCA_000027145.1	Listeria seeligeri serovar 1/2b str. SLCC3954
386043	GCA_000060285.1	Listeria welshimeri serovar 6b str. SLCC5334
28031	GCA_000724775.3	Lysinibacillus fusiformis RB-21
444177	GCA_000017965.1	Lysinibacillus sphaericus C3-41
1145276	GCA_000600105.1	Lysinibacillus varians GY32
458233	GCA_000010585.1	Macrococcus caseolyticus JCSC5402
699246	GCA_000025225.2	Mageeibacillus indolicus UPII9-5
697281	GCA_000213255.1	Mahella australiensis 50-1 BON
1458465	GCA_001304715.1	Megasphaera elsdenii 14-14
1090974	GCA_000283975.1	Melissococcus plutonius DAT561
264732	GCA_000013105.1	Moorella thermoacetica ATCC 39073
1525	GCA_001267435.1	Moorella thermoacetica DSM 2955
1525	GCA_001267405.1	Moorella thermoacetica DSM 521
457570	GCA_000020005.1	Natranaerobius thermophilus JW/NM-WN-LF
221109	GCA_000011245.1	Oceanobacillus iheyensis HTE831
1045004	GCA_000241055.1	Oenococcus kitaharae DSM 17330
203123	GCA_000014385.1	Oenococcus oeni PSU-1
693746	GCA_000283575.1	Oscillibacter valericigenes Sjm18-20

1126833	GCA_000961095.1	Paenibacillus beijingensis DSM 24997
160799	GCA_000758665.1	Paenibacillus borealis DSM 13188
1616788	GCA_001421015.1	Paenibacillus bovis BD3526
1333534	GCA_000993825.1	Paenibacillus durus ATCC 35681
44251	GCA_000756615.1	Paenibacillus durus DSM 1735
189425	GCA_000758705.1	Paenibacillus graminis DSM 15220
697284	GCA_000511405.1	Paenibacillus larvae subsp. larvae DSM 25430
1116391	GCA_000250655.1	Paenibacillus mucilaginosus 3016
997761	GCA_000258535.2	Paenibacillus mucilaginosus K02
1036673	GCA_000218915.1	Paenibacillus mucilaginosus KNP414
162209	GCA_001465255.1	Paenibacillus naphthalenovorans 32O-Y
189426	GCA_000758725.1	Paenibacillus odorifer DSM 15391
59893	GCA_001272655.1	Paenibacillus peoriae HS311
1429244	GCA_000507205.2	Paenibacillus polymyxa CR1
349520	GCA_000146875.2	Paenibacillus polymyxa E681
1052684	GCA_000237325.1	Paenibacillus polymyxa M1
886882	GCA_000164985.2	Paenibacillus polymyxa SC2
1413214	GCA_000597985.1	Paenibacillus polymyxa SQR-21
1406	GCA_000819665.1	Paenibacillus polymyxa Sb3-1
1073571	GCA_000981585.1	Paenibacillus riograndensis SBR5
1268072	GCA_000612505.1	Paenibacillus sabinae T27
1695218	GCA_001465275.1	Paenibacillus sp. 32O-W
1536774	GCA_000758525.1	Paenibacillus sp. FSL H7-0357
1536775	GCA_000758545.1	Paenibacillus sp. FSL H7-0737
1536769	GCA_000758565.1	Paenibacillus sp. FSL P4-0081
1536770	GCA_000758585.1	Paenibacillus sp. FSL R5-0345
1536771	GCA_000758605.1	Paenibacillus sp. FSL R5-0912
1536772	GCA_000758625.1	Paenibacillus sp. FSL R7-0273
1536773	GCA_000758645.1	Paenibacillus sp. FSL R7-0331
1566358	GCA_000949425.1	Paenibacillus sp. IHBB 10380
867076	GCA_001447315.1	Paenibacillus sp. IHB B 3084
324057	GCA_000023585.1	Paenibacillus sp. JDR-2
481743	GCA_000024685.1	Paenibacillus sp. Y412MC10
169760	GCA_000758685.1	Paenibacillus stellifer DSM 14472
985665	GCA_000235585.1	Paenibacillus terrae HPL-003
33033	GCA_000800295.1	Parvimonas micra KCOM 1535 (=ChDC B708)
701521	GCA_000237995.2	Pediococcus claussenii ATCC BAA-344
278197	GCA_000014505.1	Pediococcus pentosaceus ATCC 25745
1408206	GCA_000496265.1	Pediococcus pentosaceus SL4
1192197	GCA_000271665.2	Pelosinus fermentans JBW45
484770	GCA_000725345.1	Pelosinus sp. UFO1
370438	GCA_000010565.1	Pelotomaculum thermopropionicum SI
272563	GCA_000009205.1	Peptoclostridium difficile 630
1496	GCA_000953275.1	Peptoclostridium difficile 630Derm
1121308	GCA_001077535.1	Peptoclostridium difficile ATCC 9689 = DSM 1296
699034	GCA_000211235.1	Peptoclostridium difficile BI1
645462	GCA_000085225.1	Peptoclostridium difficile CD196
699035	GCA_000210435.1	Peptoclostridium difficile M120
1496	GCA_001457575.1	Peptoclostridium difficile NCTC13307
1496	GCA_001447175.1	Peptoclostridium difficile Z31
875453	GCA_000952975.1	Peptoniphilus sp. 1-1
1374	GCA_001465835.1	Planococcus kocurii ATCC 43650
200991	GCA_001465795.1	Planococcus rifietoensis M8
1526927	GCA_000785555.1	Planococcus sp. PAMC 21323
585394	GCA_000225345.1	Roseburia hominis A2-183
203119	GCA_000015865.1	Ruminiclostridium thermocellum ATCC 27405
637887	GCA_000184925.1	Ruminiclostridium thermocellum DSM 1313
697329	GCA_000179635.2	Ruminococcus albus 7 = DSM 20455
1160721	GCA_000723465.1	Ruminococcus bicirculans 80/3
407035	GCA_001005905.1	Salinicoccus halodurans H3B36
927704	GCA_000284095.1	Selenomonas ruminantium subsp. lactilytica TAM6421
712538	GCA_001189555.1	Selenomonas sp. oral taxon 478
546271	GCA_000208405.1	Selenomonas sputigena ATCC 35185
1002809	GCA_000271325.1	Solibacillus silvestris StLB046
985762	GCA_001442815.1	Staphylococcus agnetis 908
985002	GCA_000236925.1	Staphylococcus argenteus MSHR1132
1280	GCA_001278745.1	Staphylococcus aureus
703339	GCA_000025145.2	Staphylococcus aureus 04-02981
1229492	GCA_000296595.1	Staphylococcus aureus 08BA02176
1280	GCA_000746505.1	Staphylococcus aureus 2395 USA500
1280	GCA_000597965.1	Staphylococcus aureus 502A
1280	GCA_000815045.1	Staphylococcus aureus ATCC BAA1680
1321369	GCA_000418345.1	Staphylococcus aureus Bmb9393
1280	GCA_001045795.2	Staphylococcus aureus CA12
1280	GCA_001021895.1	Staphylococcus aureus CA15
1323661	GCA_000412775.1	Staphylococcus aureus CA-347
1280	GCA_001045995.2	Staphylococcus aureus HUV05
1280	GCA_000953255.1	Staphylococcus aureus ILRI_Eymole1/1
1280	GCA_001021875.1	Staphylococcus aureus M121
1280	GCA_001457495.1	Staphylococcus aureus NCTC13435
1280	GCA_001457515.1	Staphylococcus aureus NCTC8532
1280	GCA_000626615.1	Staphylococcus aureus NRS100
273036	GCA_000009005.1	Staphylococcus aureus RF122
1280	GCA_001465635.1	Staphylococcus aureus RIVM1295
1280	GCA_001465675.1	Staphylococcus aureus RIVM1607
1280	GCA_001465755.1	Staphylococcus aureus RIVM3897
1280	GCA_001027045.1	Staphylococcus aureus RKI4
1280	GCA_001281145.1	Staphylococcus aureus SA564
1280	GCA_000695875.1	Staphylococcus aureus UA-S391_USA300
1458279	GCA_000568455.1	Staphylococcus aureus USA300-ISMMS1
1280	GCA_001046095.2	Staphylococcus aureus V2200
1280	GCA_000709475.1	Staphylococcus aureus XN108
1280	GCA_001444345.1	Staphylococcus aureus XQ
46170	GCA_000695215.1	Staphylococcus aureus subsp. aureus
1123523	GCA_000239235.1	Staphylococcus aureus subsp. aureus 11819-97

585143	GCA_000160335.2	Staphylococcus aureus subsp. aureus 55/2053
1392476	GCA_000462955.1	Staphylococcus aureus subsp. aureus 6850
1193576	GCA_000463055.1	Staphylococcus aureus subsp. aureus CN1
93062	GCA_000012045.1	Staphylococcus aureus subsp. aureus COL
1241616	GCA_001027105.1	Staphylococcus aureus subsp. aureus DSM 20231
889933	GCA_000253135.1	Staphylococcus aureus subsp. aureus ECT-R 2
685039	GCA_000210315.1	Staphylococcus aureus subsp. aureus ED133
681288	GCA_000024585.1	Staphylococcus aureus subsp. aureus ED98
1074252	GCA_000284535.1	Staphylococcus aureus subsp. aureus HO 5096 0412
359787	GCA_000017125.1	Staphylococcus aureus subsp. aureus JH1
359786	GCA_000016805.1	Staphylococcus aureus subsp. aureus JH9
869816	GCA_000144955.1	Staphylococcus aureus subsp. aureus JKD6159
985006	GCA_000237265.1	Staphylococcus aureus subsp. aureus LGA251
1118959	GCA_000237125.1	Staphylococcus aureus subsp. aureus M013
282459	GCA_000011525.1	Staphylococcus aureus subsp. aureus MRSA252
282458	GCA_000011505.1	Staphylococcus aureus subsp. aureus MSSA476
196620	GCA_000011265.1	Staphylococcus aureus subsp. aureus MW2
418127	GCA_000010445.1	Staphylococcus aureus subsp. aureus Mu3
158878	GCA_000009665.1	Staphylococcus aureus subsp. aureus Mu50
158879	GCA_000009645.1	Staphylococcus aureus subsp. aureus N315
93061	GCA_000013425.1	Staphylococcus aureus subsp. aureus NCTC 8325
1368166	GCA_000737615.1	Staphylococcus aureus subsp. aureus SA268
1194085	GCA_000470865.1	Staphylococcus aureus subsp. aureus SA40
1201010	GCA_000470845.1	Staphylococcus aureus subsp. aureus SA957
1074919	GCA_000382965.1	Staphylococcus aureus subsp. aureus ST228
523796	GCA_000009585.1	Staphylococcus aureus subsp. aureus ST398
1343064	GCA_000828035.1	Staphylococcus aureus subsp. aureus ST772-MRSA-V
1006543	GCA_000204665.1	Staphylococcus aureus subsp. aureus T0131
548473	GCA_000159535.2	Staphylococcus aureus subsp. aureus TCH60
663951	GCA_000027045.1	Staphylococcus aureus subsp. aureus TW20
451515	GCA_000013465.1	Staphylococcus aureus subsp. aureus USA300_FPR3757
451516	GCA_000017085.1	Staphylococcus aureus subsp. aureus USA300_TCH1516
1028799	GCA_000245495.1	Staphylococcus aureus subsp. aureus VC40
1406863	GCA_000485885.1	Staphylococcus aureus subsp. aureus Z172
546342	GCA_000145595.1	Staphylococcus aureus subsp. aureus str. JKD6008
426430	GCA_000010465.1	Staphylococcus aureus subsp. aureus str. Newman
72758	GCA_001028645.1	Staphylococcus capitis AYP1020
396513	GCA_000009405.1	Staphylococcus carnosus subsp. carnosus TM300
176280	GCA_000007645.1	Staphylococcus epidermidis ATCC 12228
1449752	GCA_000751035.1	Staphylococcus epidermidis PM221
176279	GCA_000011925.1	Staphylococcus epidermidis RP62A
1282	GCA_000759555.1	Staphylococcus epidermidis SEI
246432	GCA_001432245.1	Staphylococcus equorum KS1039
279808	GCA_000009865.1	Staphylococcus haemolyticus JCSC1435
1283	GCA_000972725.1	Staphylococcus haemolyticus Sh29/312/L2
1284	GCA_000816085.1	Staphylococcus hyicus ATCC 11249
698737	GCA_000025085.1	Staphylococcus lugdunensis HKU09-01
1034809	GCA_000270465.1	Staphylococcus lugdunensis N920143
1276282	GCA_000494875.1	Staphylococcus pasteuri SP1
1266717	GCA_000478385.1	Staphylococcus pseudintermedius E140
984892	GCA_000189495.1	Staphylococcus pseudintermedius ED99
937773	GCA_000185885.1	Staphylococcus pseudintermedius HKU10-03
342451	GCA_000010125.1	Staphylococcus saprophyticus subsp. saprophyticus ATCC 15305
1295	GCA_001188855.1	Staphylococcus schleiferi
1295	GCA_001188895.1	Staphylococcus schleiferi 2142-05
1295	GCA_001188915.1	Staphylococcus schleiferi 2317-03
1295	GCA_001188875.1	Staphylococcus schleiferi 5909-02
1194526	GCA_000332735.1	Staphylococcus warneri SG1
1288	GCA_000953575.1	Staphylococcus xylosus C2a
1288	GCA_000706685.1	Staphylococcus xylosus HKUOPL8
1288	GCA_000709415.1	Staphylococcus xylosus SMQ-121
1311	GCA_000831105.1	Streptococcus agalactiae
1309806	GCA_000427035.1	Streptococcus agalactiae 09mas018883
1417990	GCA_000599965.1	Streptococcus agalactiae 138P
1311	GCA_000636115.1	Streptococcus agalactiae 138spar
208435	GCA_000007265.1	Streptococcus agalactiae 2603V/R
205921	GCA_000012705.1	Streptococcus agalactiae A909
1427374	GCA_000782855.1	Streptococcus agalactiae CNCTC 10/84
342616	GCA_000689235.1	Streptococcus agalactiae COH1
1311	GCA_000967445.1	Streptococcus agalactiae Feb-22
1311	GCA_000831145.1	Streptococcus agalactiae GBS1-NY
1311	GCA_001266635.1	Streptococcus agalactiae GBS85147
1311	GCA_001448985.1	Streptococcus agalactiae GBS ST-1
1203670	GCA_000299135.1	Streptococcus agalactiae GD201008-001
1311	GCA_001190865.1	Streptococcus agalactiae GX026
1311	GCA_001190885.1	Streptococcus agalactiae H002
1311	GCA_001190805.1	Streptococcus agalactiae HN016
1309807	GCA_000427075.1	Streptococcus agalactiae ILRI005
1318615	GCA_000427055.1	Streptococcus agalactiae ILRI112
1311	GCA_000730215.2	Streptococcus agalactiae NGBS061
1311	GCA_000730255.1	Streptococcus agalactiae NGBS572
1231389	GCA_000302475.2	Streptococcus agalactiae SA20-06
1311	GCA_001275545.2	Streptococcus agalactiae SG-M1
1311	GCA_001026925.1	Streptococcus agalactiae SS1
1328	GCA_000831165.1	Streptococcus anginosus
862970	GCA_000463465.1	Streptococcus anginosus C1051
862971	GCA_000463505.1	Streptococcus anginosus C238
1328	GCA_001412635.1	Streptococcus anginosus J4211
1353243	GCA_000478925.1	Streptococcus anginosus subsp. whileyi MAS624
862969	GCA_000463425.1	Streptococcus constellatus subsp. pharyngis C1050
696216	GCA_000463395.1	Streptococcus constellatus subsp. pharyngis C232
862968	GCA_000463445.1	Streptococcus constellatus subsp. pharyngis C818
1247189	GCA_000493775.1	Streptococcus dysgalactiae subsp. equisimilis 167
759913	GCA_000317855.1	Streptococcus dysgalactiae subsp. equisimilis AC-2713
663954	GCA_000188715.1	Streptococcus dysgalactiae subsp. equisimilis ATCC 12394

486410	GCA_000010705.1	Streptococcus dysgalactiae subsp. equisimilis GGS_124
617121	GCA_000307185.1	Streptococcus dysgalactiae subsp. equisimilis RE378
553482	GCA_000026585.1	Streptococcus equi subsp. equi 4047
40041	GCA_000026605.1	Streptococcus equi subsp. zooepidemicus
1051072	GCA_000219765.1	Streptococcus equi subsp. zooepidemicus ATCC 35246
1403449	GCA_000696505.1	Streptococcus equi subsp. zooepidemicus CY
552526	GCA_000020765.1	Streptococcus equi subsp. zooepidemicus MGCS10565
315405	GCA_001477575.1	Streptococcus gallolyticus ICDDRB-NRC-S1
637909	GCA_000027185.1	Streptococcus gallolyticus UCN34
981539	GCA_000270145.1	Streptococcus gallolyticus subsp. gallolyticus ATCC 43143
990317	GCA_000203195.1	Streptococcus gallolyticus subsp. gallolyticus ATCC BAA-2069
1302	GCA_001281105.1	Streptococcus gordonii KCOM 1506 (= ChDC B679)
467705	GCA_000017005.1	Streptococcus gordonii str. Challis substr. CH1
102684	GCA_001477615.1	Streptococcus infantarius ICDDRB-NRC-S5
1069533	GCA_000246835.1	Streptococcus infantarius subsp. infantarius CJ18
1346	GCA_000648525.1	Streptococcus iniae ISET0901
1346	GCA_000648555.1	Streptococcus iniae ISNO
1318633	GCA_000403625.1	Streptococcus iniae SF1
1346	GCA_000831485.1	Streptococcus iniae YSFST01-82
862967	GCA_000463355.1	Streptococcus intermedius B196
862966	GCA_000463385.1	Streptococcus intermedius C270
591365	GCA_000306805.1	Streptococcus intermedius JTH08
1338	GCA_001296205.1	Streptococcus intermedius KCOM 1545
1076934	GCA_000441535.1	Streptococcus lutetiensis 033
1116231	GCA_000283635.1	Streptococcus macedonicus ACA-DC 198
365659	GCA_000027165.1	Streptococcus mitis B6
28037	GCA_001281025.1	Streptococcus mitis KCOM 1350 (= ChDC B183)
1198676	GCA_000271865.1	Streptococcus mutans GS-5
1155071	GCA_000284575.1	Streptococcus mutans LJ23
511691	GCA_000091645.1	Streptococcus mutans NN2025
210007	GCA_000007465.2	Streptococcus mutans UA159
1437447	GCA_000817065.1	Streptococcus mutans UA159-FR
1302863	GCA_000385925.1	Streptococcus oligofermentans AS 1.3089
927666	GCA_000253155.1	Streptococcus oralis Uo5
760570	GCA_000164675.2	Streptococcus parasanguinis ATCC 15912
1114965	GCA_000262145.1	Streptococcus parasanguinis FW213
936154	GCA_000213825.1	Streptococcus parauberis KCTC 11537
873447	GCA_000187935.2	Streptococcus parauberis NCFD 2020
981540	GCA_000270165.1	Streptococcus pasteurianus ATCC 43144
189423	GCA_000147095.1	Streptococcus pneumoniae 670-6B
488221	GCA_000018965.1	Streptococcus pneumoniae 70585
574093	GCA_000146975.1	Streptococcus pneumoniae AP200
561276	GCA_000026665.1	Streptococcus pneumoniae ATCC 700669
516950	GCA_000019985.1	Streptococcus pneumoniae CGSP14
373153	GCA_000014365.1	Streptococcus pneumoniae D39
512566	GCA_000019825.1	Streptococcus pneumoniae G54
487214	GCA_000019265.1	Streptococcus pneumoniae Hungary19A-6
869269	GCA_000210975.1	Streptococcus pneumoniae INV104
869216	GCA_000210935.1	Streptococcus pneumoniae INV200
488222	GCA_000018985.1	Streptococcus pneumoniae JJA
1313	GCA_001457635.1	Streptococcus pneumoniae NCTC7465
1313	GCA_000817005.1	Streptococcus pneumoniae NT_110_58
869215	GCA_000210955.1	Streptococcus pneumoniae OXC141
488223	GCA_000019005.1	Streptococcus pneumoniae P1031
1159083	GCA_000348705.1	Streptococcus pneumoniae PCS8235
171101	GCA_000007045.1	Streptococcus pneumoniae R6
869311	GCA_000211075.1	Streptococcus pneumoniae SPN032672
869312	GCA_000211095.1	Streptococcus pneumoniae SPN033038
869303	GCA_000210995.1	Streptococcus pneumoniae SPN034156
869304	GCA_000211015.1	Streptococcus pneumoniae SPN034183
869306	GCA_000211035.2	Streptococcus pneumoniae SPN994038
869307	GCA_000211055.2	Streptococcus pneumoniae SPN994039
869309	GCA_000180515.2	Streptococcus pneumoniae SPNA45
1130804	GCA_000251085.2	Streptococcus pneumoniae ST556
525381	GCA_000196595.1	Streptococcus pneumoniae TCH8431/19A
170187	GCA_000006885.1	Streptococcus pneumoniae TIGR4
487213	GCA_000019025.1	Streptococcus pneumoniae Taiwan19F-14
697283	GCA_000299015.1	Streptococcus pneumoniae gamPNI0373
1054460	GCA_000221985.1	Streptococcus pseudopneumoniae IS7493
1314	GCA_001039695.2	Streptococcus pyogenes
1314	GCA_000772245.1	Streptococcus pyogenes 1E1
1314	GCA_001021955.1	Streptococcus pyogenes 5448
1314	GCA_000767505.1	Streptococcus pyogenes 7F7
1235829	GCA_000307535.1	Streptococcus pyogenes A20
1314	GCA_000993765.1	Streptococcus pyogenes AP1
1314	GCA_000743015.1	Streptococcus pyogenes ATCC 19615
487215	GCA_000230295.1	Streptococcus pyogenes Alab49
1314	GCA_001014305.1	Streptococcus pyogenes D471
1314	GCA_000772185.1	Streptococcus pyogenes HKU360
1314	GCA_001051095.1	Streptococcus pyogenes HKU488
1048264	GCA_000275625.1	Streptococcus pyogenes HKU QMH11M0907901
1336746	GCA_000422045.1	Streptococcus pyogenes HSC5
1150773	GCA_001014285.1	Streptococcus pyogenes JRS4
1207470	GCA_000349925.2	Streptococcus pyogenes M1 476
160490	GCA_000006785.2	Streptococcus pyogenes M1 GAS
1314	GCA_000756485.1	Streptococcus pyogenes M23ND
1314	GCA_001020185.1	Streptococcus pyogenes M28PF1
370552	GCA_000013505.1	Streptococcus pyogenes MGAS10270
286636	GCA_000011665.1	Streptococcus pyogenes MGAS10394
370554	GCA_000013545.1	Streptococcus pyogenes MGAS10750
798300	GCA_000250905.1	Streptococcus pyogenes MGAS15252
1010840	GCA_000250925.1	Streptococcus pyogenes MGAS1882
370553	GCA_000013525.1	Streptococcus pyogenes MGAS2096
198466	GCA_000007425.1	Streptococcus pyogenes MGAS315
293653	GCA_000011765.2	Streptococcus pyogenes MGAS5005

319701	GCA_000012165.1	Streptococcus pyogenes MGAS6180
186103	GCA_000007285.1	Streptococcus pyogenes MGAS8232
370551	GCA_000013485.1	Streptococcus pyogenes MGAS9429
1314	GCA_001267805.1	Streptococcus pyogenes NGAS322
1314	GCA_001019695.1	Streptococcus pyogenes NGAS327
1314	GCA_001019675.1	Streptococcus pyogenes NGAS596
1314	GCA_001267845.1	Streptococcus pyogenes NGAS638
1314	GCA_001019635.1	Streptococcus pyogenes NGAS743
471876	GCA_000018125.1	Streptococcus pyogenes NZ131
193567	GCA_000011285.1	Streptococcus pyogenes SSI-1
1314	GCA_001023495.1	Streptococcus pyogenes STAB10015
1440772	GCA_000732385.1	Streptococcus pyogenes STAB901
1437007	GCA_000732425.1	Streptococcus pyogenes STAB902
160491	GCA_000009385.1	Streptococcus pyogenes str. Manfredo
1046629	GCA_000305335.1	Streptococcus salivarius 57.I
1048332	GCA_000253335.1	Streptococcus salivarius CCHSS3
347253	GCA_000253315.1	Streptococcus salivarius JIM8777
1304	GCA_000785515.1	Streptococcus salivarius NCTC 8618
388919	GCA_000014205.1	Streptococcus sanguinis SK36
1156431	GCA_000479335.1	Streptococcus sp. I-G2
1156433	GCA_000479315.1	Streptococcus sp. I-P16
1419814	GCA_000688775.1	Streptococcus sp. VT 162
672190	GCA_000168355.3	Streptococcus suis 05HAS68
391295	GCA_000014305.1	Streptococcus suis 05ZYH33
1214179	GCA_000732355.1	Streptococcus suis 6407
391296	GCA_000014325.1	Streptococcus suis 98HAH33
993512	GCA_000233575.1	Streptococcus suis A7
568814	GCA_000026745.1	Streptococcus suis BM407
1004952	GCA_000231905.1	Streptococcus suis D12
1005042	GCA_000231885.1	Streptococcus suis D9
423211	GCA_000018185.1	Streptococcus suis GZ1
945704	GCA_000186405.1	Streptococcus suis JS14
1307	GCA_001272635.1	Streptococcus suis NSUI002
218494	GCA_000091905.1	Streptococcus suis P1/7
1184252	GCA_000294495.1	Streptococcus suis S735
1246365	GCA_000344765.1	Streptococcus suis SC070731
568813	GCA_000026725.1	Streptococcus suis SC84
1005041	GCA_000231865.1	Streptococcus suis SS12
1004951	GCA_000231925.1	Streptococcus suis ST1
1007064	GCA_000204625.1	Streptococcus suis ST3
1340847	GCA_000494895.1	Streptococcus suis T15
1276647	GCA_000390245.1	Streptococcus suis TL13
1380773	GCA_000471985.1	Streptococcus suis YB51
1307	GCA_000993745.1	Streptococcus suis ZY05719
1408178	GCA_000698885.1	Streptococcus thermophilus ASCC 1275
299768	GCA_000011845.1	Streptococcus thermophilus CNRZ1066
1051074	GCA_000253395.1	Streptococcus thermophilus JIM 8232
322159	GCA_000014485.1	Streptococcus thermophilus LMD-9
264199	GCA_000011825.1	Streptococcus thermophilus LMG 18311
1308	GCA_001280285.1	Streptococcus thermophilus MN-BM-A01
1308	GCA_001008015.1	Streptococcus thermophilus MN-BM-A02
1187956	GCA_000262675.1	Streptococcus thermophilus MN-ZLW-002
767463	GCA_000182875.1	Streptococcus thermophilus ND03
1308	GCA_000971665.1	Streptococcus thermophilus SMQ-301
218495	GCA_000009545.1	Streptococcus uberis 0140J
679936	GCA_000237975.1	Sulfobacillus acidophilus DSM 10332
1051632	GCA_000219855.1	Sulfobacillus acidophilus TPY
292459	GCA_000009905.1	Symbiobacterium thermophilum IAM 14863
645991	GCA_000190635.1	Syntrophobotulus glycolicus DSM 8271
335541	GCA_000014725.1	Syntrophomonas wolfei subsp. wolfei str. Goettingen G311
643648	GCA_000092405.1	Syntrophothermus lipocalidus DSM 12680
1209989	GCA_000213235.1	Tepidanaerobacter acetatoxydans Re1
586416	GCA_000725365.1	Terribacillus aidingensis MP602
945021	GCA_000283615.1	Tetragenococcus halophilus NBRC 12172
1089553	GCA_000305935.1	Thermacetogenium phaeum DSM 12270
644966	GCA_000184705.1	Thermaerobacter marianensis DSM 12885
635013	GCA_000092945.1	Thermincola potens JR
509193	GCA_000175295.2	Thermoanaerobacter brockii subsp. finnii Ako-1
580331	GCA_000025645.1	Thermoanaerobacter italicus Ab9
2325	GCA_000763575.1	Thermoanaerobacter kivui LKT-1
583358	GCA_000092965.1	Thermoanaerobacter mathranii subsp. mathranii str. A3
340099	GCA_000019085.1	Thermoanaerobacter pseudethanolicus ATCC 33223
573062	GCA_000148425.1	Thermoanaerobacter sp. X513
399726	GCA_000019065.1	Thermoanaerobacter sp. X514
697303	GCA_000147695.3	Thermoanaerobacter wiegelii Rt8.B1
1094508	GCA_000307585.1	Thermoanaerobacterium saccharolyticum JW/SL-YS485
580327	GCA_000145615.1	Thermoanaerobacterium thermosaccharolyticum DSM 571
698948	GCA_000328545.1	Thermoanaerobacterium thermosaccharolyticum M0795
858215	GCA_000189775.3	Thermoanaerobacterium xylanolyticum LX-11
717605	GCA_000227705.3	Thermobacillus composti KWC4
747365	GCA_000212395.1	Thermodesulfobium narugense DSM 14796
555079	GCA_000144645.1	Thermosediminibacter oceani DSM 16646
479436	GCA_000024945.1	Veillonella parvula DSM 2008
403957	GCA_000725285.1	Virgibacillus sp. SK37
759620	GCA_000732905.1	Weissella ceti WS08
759620	GCA_000750535.1	Weissella ceti WS105
759620	GCA_000750515.1	Weissella ceti WS74
137591	GCA_001308145.1	Weissella cibaria CH2
1045854	GCA_000219805.1	Weissella koreensis KACC 15510

.2 Génomes et protéomes complets de *Firmicutes* après échantillonnage taxonomique.

Identifiant taxonomique	Numéro d'assemblage	Nom de la souche
931626	GCA_000247605.1	Acetobacterium woodii DSM 1030
574087	GCA_000144695.1	Acetohalobium arabaticum DSM 5501
591001	GCA_000025305.1	Acidaminococcus fermentans DSM 20731
568816	GCA_000230275.1	Acidaminococcus intestini RyC-MR95
866775	GCA_000193205.1	Aerococcus urinae ACS-120-V-Col10a
1048834	GCA_000219875.1	Alicyclobacillus acidocaldarius subsp. acidocaldarius Tc-4-1
293826	GCA_000016985.1	Alkaliphilus metalliredigens QYMF
350688	GCA_000018325.1	Alkaliphilus oremlandii OhILAs
429009	GCA_000024605.1	Ammonifex degensii KC4
698758	GCA_000307165.1	Amphibacillus xylanus NBRC 15112
525919	GCA_000024105.1	Anaerococcus prevotii DSM 20548
491915	GCA_000019045.1	Anoxybacillus flavithermus WK1
198467	GCA_001187595.1	Anoxybacillus gonensis G2
1126211	GCA_000262385.1	Bacillus amyloliquefaciens Y2
261591	GCA_000742895.1	Bacillus anthracis str. Vollum
1452	GCA_000830075.1	Bacillus atrophaeus NRS 1221A
1330043	GCA_000831065.1	Bacillus bombysepticus str. Wang
649639	GCA_000177235.2	Bacillus cellulosilyticus DSM 2522
637380	GCA_000143605.1	Bacillus cereus biovar anthracis str. CI
66692	GCA_000009825.1	Bacillus clausii KSM-K16
1398	GCA_001039495.1	Bacillus coagulans S-lac
315749	GCA_000017425.1	Bacillus cytotoxicus NVH 391-98
135735	GCA_000972245.2	Bacillus endophyticus Hbe603
272558	GCA_000011145.1	Bacillus halodurans C-125
1367477	GCA_000473245.1	Bacillus infantis NRRL B-14911
1246626	GCA_000706725.1	Bacillus lehensis G1
279010	GCA_000008425.1	Bacillus licheniformis DSM 13 = ATCC 14580
1006007	GCA_000225265.1	Bacillus megaterium WSH-002
796606	GCA_000724485.1	Bacillus methanolicus MGA3
1405	GCA_000832605.1	Bacillus mycoides ATCC 6462
1648923	GCA_000876525.1	Bacillus paralicheniformis BL-09
398511	GCA_000005825.2	Bacillus pseudofirmus OF4
1408	GCA_000972685.1	Bacillus pumilus W3
439292	GCA_000093085.1	[Bacillus] selenitireducens MLS10
1479	GCA_001050115.1	Bacillus smithii DSM 4216
1574141	GCA_000877815.1	Bacillus sp. YP1
224308	GCA_000009045.1	Bacillus subtilis subsp. subtilis str. 168
412694	GCA_000015065.1	Bacillus thuringiensis str. Al Hakam
1415784	GCA_000496285.1	Bacillus toyonensis BCT-7112
492670	GCA_000988345.1	Bacillus velezensis YJ11-1-4
86662	GCA_000775975.1	Bacillus weihenstephanensis WSBC10204
358681	GCA_000010165.1	Brevibacillus brevis NBRC 100599
1042163	GCA_000219535.3	Brevibacillus laterosporus LMG 15441
515622	GCA_000145035.1	Butyrivibrio proteoclasticus B316
273068	GCA_000007085.1	Caldanaerobacter subterraneus subsp. tengcongensis MB4
521460	GCA_000022325.1	Caldicellulosiruptor bescii DSM 6725
632292	GCA_000166355.1	Caldicellulosiruptor hydrothermalis 108
632335	GCA_000166695.1	Caldicellulosiruptor kristjanssonii I77R1B
632348	GCA_000166775.1	Caldicellulosiruptor kronotskyensis 2002
632516	GCA_000193435.3	Caldicellulosiruptor lactoaceticus 6A
608506	GCA_000145215.1	Caldicellulosiruptor obsidiansis OB47
632518	GCA_000166335.1	Caldicellulosiruptor owensensis OL
351627	GCA_000016545.1	Caldicellulosiruptor saccharolyticus DSM 8903
1041504	GCA_000283555.1	Candidatus Arthromitus sp. SFB-rat-Yit
477974	GCA_000018425.1	Candidatus Desulforudis audaxviator MP104C
246194	GCA_000012865.1	Carboxydothermus hydrogenoformans Z-2901
1266845	GCA_000493735.1	Carnobacterium inhibens subsp. gilichinskyi
1234679	GCA_000317975.2	Carnobacterium maltaromaticum LMA28
1564681	GCA_001483965.1	Carnobacterium sp. CP1
84022	GCA_001042715.1	Clostridium aceticum DSM 1496
863638	GCA_000191905.1	Clostridium acetobutylicum EA 2018
1128398	GCA_000299355.1	[Clostridium] acidurici 9a
1341692	GCA_000484505.1	Clostridium autoethanogenum DSM 10061
1415775	GCA_000789395.1	Clostridium baratii str. Sullivan
290402	GCA_000016965.1	Clostridium beijerinckii NCIMB 8052
1216932	GCA_000577895.1	Clostridium bornimense M2/40

1491	GCA_000827955.1	Clostridium botulinum NCTC 8550
1492	GCA_001456065.2	Clostridium butyricum KNU-L09
536227	GCA_001038625.1	Clostridium carboxidivorans P7
394503	GCA_000022065.1	[Clostridium] cellulolyticum H10
29343	GCA_000953215.1	[Clostridium] cellulosi
573061	GCA_000145275.1	Clostridium cellulovorans 743B
720554	GCA_000237085.1	[Clostridium] clariflavum DSM 19732
583346	GCA_000010265.1	Clostridium kluyveri NBRC 12016
642492	GCA_000178835.2	Clostridium lentocellum DSM 5427
748727	GCA_000143685.1	Clostridium ljungdahlii DSM 13528
386415	GCA_000014125.1	Clostridium novyi NT
1428454	GCA_000506785.2	Clostridium pasteurianum NRRL B-598
195102	GCA_000009685.1	Clostridium perfringens str. 13
1345695	GCA_000473995.1	Clostridium saccharobutylicum DSM 13864
610130	GCA_000144625.1	[Clostridium] saccharolyticum WM1
931276	GCA_000340885.1	Clostridium saccharoperbutylacetonicum N1-4(HMT)
1548	GCA_000968375.1	Clostridium scatologenes ATCC 25775
1042156	GCA_000270305.1	Clostridium sp. SY8519
1509	GCA_000973705.1	Clostridium sporogenes NCIMB 10696
1121335	GCA_000331995.1	[Clostridium] stercorarium subsp. stercorarium DSM 8532
1511	GCA_000196455.1	[Clostridium] sticklandii DSM 519
212717	GCA_000007625.1	Clostridium tetani E88
871738	GCA_000512895.1	Dehalobacter restrictus DSM 9455
1147129	GCA_000305775.1	Dehalobacter sp. DCA
756499	GCA_000243155.3	Desulfitobacterium dehalogenans ATCC 51507
871963	GCA_000243135.3	Desulfitobacterium dichloroeliminans LMG P-21439
138119	GCA_000010045.1	Desulfitobacterium hafniense Y51
871968	GCA_000231405.3	Desulfitobacterium metallireducens DSM 15288
646529	GCA_000255115.3	Desulfosporosinus acidiphilus SJ4
768704	GCA_000231385.3	Desulfosporosinus meridiei DSM 13257
768706	GCA_000235605.1	Desulfosporosinus orientis DSM 765
485916	GCA_000024205.1	Desulfotomaculum acetoxidans DSM 771
767817	GCA_000233715.3	Desulfotomaculum gibsoniae DSM 7213
760568	GCA_000214705.1	Desulfotomaculum kuznetsovii DSM 6115
868595	GCA_000214435.1	Desulfotomaculum nigrificans CO-1-SRB
349161	GCA_000016165.1	Desulfotomaculum reducens MI-1
696281	GCA_000215085.1	Desulfotomaculum ruminis DSM 2154
565655	GCA_000157355.2	Enterococcus casseliflavus EC20
53345	GCA_001267395.1	Enterococcus durans KLDS 6.0933
1261557	GCA_000317915.1	Enterococcus faecalis str. Symbioflor 1
1352	GCA_001412695.1	Enterococcus faecium UW7606x64/3 TC1
768486	GCA_000271405.2	Enterococcus hirae ATCC 9790
1300150	GCA_000504125.1	Enterococcus mundtii QU 25
332949	GCA_001465115.1	Enterococcus silesiacus LMG 23085
650150	GCA_000270085.1	Erysipelothrix rhusiopathiae str. Fujisawa
663278	GCA_000178115.2	Ethanoligenens harbinense YUAN-3
1286171	GCA_000597865.1	Eubacterium acidaminophilum DSM 3953
515620	GCA_000146185.1	[Eubacterium] eligens ATCC 27750
1736	GCA_001481725.1	Eubacterium limosum SA11
515619	GCA_000020605.1	[Eubacterium rectale] ATCC 33656
888727	GCA_001189495.1	Eubacterium sulci ATCC 35585
1087448	GCA_000299435.1	Exiguobacterium antarcticum B7
262543	GCA_000019905.1	Exiguobacterium sibiricum 255-15
1399115	GCA_000496635.1	Exiguobacterium sp. MH3
546269	GCA_000163895.2	Filifactor alocis ATCC 35896
334413	GCA_000010185.1	Finegoldia magna ATCC 29328
235909	GCA_000009785.1	Geobacillus kaustophilus HTA426
581103	GCA_000166075.1	Geobacillus sp. Y4.1MC1
272567	GCA_001274575.1	Geobacillus stearothermophilus 10
420246	GCA_000015745.1	Geobacillus thermodenitrificans NG80-2
1426	GCA_001295365.1	Geobacillus thermoglucosidasius DSM 2542
1111068	GCA_000236605.1	Geobacillus thermoleovorans CCB_US3_UF5
656519	GCA_000166415.1	Halanaerobium hydrogeniformans missing
572479	GCA_000165465.1	Halanaerobium praevalens DSM 2228
866895	GCA_000284515.1	Halobacillus halophilus DSM 2266
748449	GCA_000328625.1	Halobacteroides halobius DSM 5150
373903	GCA_000020485.1	Halothermothrix orenii H 168
498761	GCA_000019165.1	Heliobacterium modesticaldum Ice1
1679721	GCA_001298655.2	Herbinix sp. SD1D

1297617	GCA_001454945.1	Intestinimonas butyriciproducens AF211
1508404	GCA_000818095.1	Jeotgalibacillus malaysiensis
1461582	GCA_000756715.2	Jeotgalicoccus sp. 13MG44_air
562970	GCA_000092905.1	Kyrpidia tusciae DSM 2912
357809	GCA_000018685.1	Lachnoclostridium phytofermentans ISDg
1600	GCA_001042405.1	Lactobacillus acetotolerans NBRC 13120
272621	GCA_000011985.1	Lactobacillus acidophilus NCFM
695562	GCA_000194115.1	Lactobacillus amylovorus GRL1118
1001583	GCA_000359625.1	Lactobacillus brevis KB290
511437	GCA_000211375.1	Lactobacillus buchneri NRRL B-30929
219334	GCA_000829055.1	Lactobacillus casei subsp. casei ATCC 393
767455	GCA_000182835.1	Lactobacillus delbrueckii subsp. bulgaricus ND02
334390	GCA_000010145.1	Lactobacillus fermentum IFO 3956
52242	GCA_001314245.1	Lactobacillus gallinarum HFD4
324831	GCA_000014425.1	Lactobacillus gasseri ATCC 33323 = JCM 1131
1007676	GCA_001050475.1	Lactobacillus ginsenosidimutans EMML 3041
1074467	GCA_000831645.3	Lactobacillus heilongjiangensis DSM 28069
880633	GCA_000165775.3	Lactobacillus helveticus R0052
1291742	GCA_000829395.1	Lactobacillus hokkaidonensis JCM 18461
257314	GCA_000008065.1	Lactobacillus johnsonii NCC 533
1033837	GCA_000214785.1	Lactobacillus kefiranofaciens ZW3
637971	GCA_001050435.1	Lactobacillus koreensis 26-25
148814	GCA_001314945.1	Lactobacillus kunkeei MP2
1130798	GCA_000248095.3	Lactobacillus mucosae LM1
1226298	GCA_000829035.1	Lactobacillus paracasei subsp. paracasei JCM 8130
60520	GCA_001443645.1	Lactobacillus paraplantarum L-ZS9
889932	GCA_000148815.2	Lactobacillus plantarum subsp. plantarum ST-III
1358027	GCA_000439275.1	Lactobacillus reuteri TD1
568704	GCA_000026525.1	Lactobacillus rhamnosus Lc 705
1069534	GCA_000224985.1	Lactobacillus ruminis ATCC 27782
314315	GCA_000026065.1	Lactobacillus sakei subsp. sakei 23K
1194971	GCA_001011095.1	Lactobacillus salivarius str. Ren
714313	GCA_000225325.1	Lactobacillus sanfranciscensis TMW 1.1304
1545702	GCA_000761135.1	Lactobacillus sp. wkB8
420890	GCA_000269945.1	Lactococcus garvieae Lg2
1117941	GCA_000478255.2	Lactococcus lactis subsp. lactis NCDO 2118
297352	GCA_000981525.1	Lactococcus piscium MKFS47
1229758	GCA_000300135.1	Leuconostoc carnosum JB16
349519	GCA_000026405.1	Leuconostoc citreum KM20
762550	GCA_000196855.1	Leuconostoc gelidum subsp. gasicomitatum LMG 18811
762051	GCA_000092505.1	Leuconostoc kimchii IMSNU 11154
1107880	GCA_000234825.3	Leuconostoc mesenteroides subsp. mesenteroides J18
979982	GCA_000219785.1	Leuconostoc sp. C2
202752	GCA_000763475.1	Listeria ivanovii subsp. londoniensis
1230340	GCA_000318055.1	Listeria monocytogenes serotype 4b str. LL195
683837	GCA_000027145.1	Listeria seeligeri serovar 1/2b str. SLCC3954
386043	GCA_000060285.1	Listeria welshimeri serovar 6b str. SLCC5334
28031	GCA_000724775.3	Lysinibacillus fusiformis RB-21
444177	GCA_000017965.1	Lysinibacillus sphaericus C3-41
1145276	GCA_000600105.1	Lysinibacillus varians GY32
458233	GCA_000010585.1	Macrococcus caseolyticus JCSC5402
699246	GCA_000025225.2	Mageeibacillus indolicus UPII9-5
697281	GCA_000213255.1	Mahella australiensis 50-1 BON
1458465	GCA_001304715.1	Megasphaera elsdenii 14-14
1090974	GCA_000283975.1	Melissococcus plutonius DAT561
1525	GCA_001267405.1	Moorella thermoacetica DSM 521
457570	GCA_000020005.1	Natranaerobius thermophilus JW/NM-WN-LF
221109	GCA_000011245.1	Oceanobacillus iheyensis HTE831
1045004	GCA_000241055.1	Oenococcus kitaharae DSM 17330
203123	GCA_000014385.1	Oenococcus oeni PSU-1
693746	GCA_000283575.1	Oscillibacter valericigenes Sjm18-20
1126833	GCA_000961095.1	Paenibacillus beijingensis DSM 24997
160799	GCA_000758665.1	Paenibacillus borealis DSM 13188
1616788	GCA_001421015.1	Paenibacillus bovis BD3526
44251	GCA_000756615.1	Paenibacillus durus DSM 1735
189425	GCA_000758705.1	Paenibacillus graminis DSM 15220
697284	GCA_000511405.1	Paenibacillus larvae subsp. larvae DSM 25430
1036673	GCA_000218915.1	Paenibacillus mucilaginosus KNP414
162209	GCA_001465255.1	Paenibacillus naphthalenovorans 32O-Y

189426	GCA_000758725.1	Paenibacillus odorifer DSM 15391
59893	GCA_001272655.1	Paenibacillus peoriae HS311
1406	GCA_000819665.1	Paenibacillus polymyxa Sb3-1
1073571	GCA_000981585.1	Paenibacillus riograndensis SBR5
1268072	GCA_000612505.1	Paenibacillus sabinae T27
481743	GCA_000024685.1	Paenibacillus sp. Y412MC10
169760	GCA_000758685.1	Paenibacillus stellifer DSM 14472
985665	GCA_000235585.1	Paenibacillus terrae HPL-003
33033	GCA_000800295.1	Parvimonas micra KCOM 1535 (=ChDC B708)
701521	GCA_000237995.2	Pediococcus claussenii ATCC BAA-344
1408206	GCA_000496265.1	Pediococcus pentosaceus SL4
1192197	GCA_000271665.2	Pelosinus fermentans JBW45
484770	GCA_000725345.1	Pelosinus sp. UFO1
370438	GCA_000010565.1	Pelotomaculum thermopropionicum SI
1496	GCA_001447175.1	Peptoclostridium difficile Z31
875453	GCA_000952975.1	Peptoniphilus sp. 1-1
1374	GCA_001465835.1	Planococcus kocurii ATCC 43650
200991	GCA_001465795.1	Planococcus rifietoensis M8
1526927	GCA_000785555.1	Planococcus sp. PAMC 21323
585394	GCA_000225345.1	Roseburia hominis A2-183
637887	GCA_000184925.1	Ruminiclostridium thermocellum DSM 1313
697329	GCA_000179635.2	Ruminococcus albus 7 = DSM 20455
1160721	GCA_000723465.1	Ruminococcus bicirculans 80/3
407035	GCA_001005905.1	Salinicoccus halodurans H3B36
927704	GCA_000284095.1	Selenomonas ruminantium subsp. lactilytica TAM6421
712538	GCA_001189555.1	Selenomonas sp. oral taxon 478
546271	GCA_000208405.1	Selenomonas sputigena ATCC 35185
1002809	GCA_000271325.1	Solibacillus silvestris StLB046
985762	GCA_001442815.1	Staphylococcus agnetis 908
985002	GCA_000236925.1	Staphylococcus argenteus MSHR1132
93061	GCA_000013425.1	Staphylococcus aureus subsp. aureus NCTC 8325
72758	GCA_001028645.1	Staphylococcus capitis AYP1020
396513	GCA_000009405.1	Staphylococcus carnosus subsp. carnosus TM300
1282	GCA_000759555.1	Staphylococcus epidermidis SEI
246432	GCA_001432245.1	Staphylococcus equorum KS1039
1283	GCA_000972725.1	Staphylococcus haemolyticus Sh29/312/L2
1284	GCA_000816085.1	Staphylococcus hyicus ATCC 11249
1034809	GCA_000270465.1	Staphylococcus lugdunensis N920143
1276282	GCA_000494875.1	Staphylococcus pasteuri SP1
937773	GCA_000185885.1	Staphylococcus pseudintermedius HKU10-03
342451	GCA_000010125.1	Staphylococcus saprophyticus subsp. saprophyticus ATCC 15305
1295	GCA_001188875.1	Staphylococcus schleiferi 5909-02
1194526	GCA_000332735.1	Staphylococcus warneri SG1
1288	GCA_000709415.1	Staphylococcus xylosus SMQ-121
1311	GCA_001026925.1	Streptococcus agalactiae SS1
1353243	GCA_000478925.1	Streptococcus anginosus subsp. whileyi MAS624
862968	GCA_000463445.1	Streptococcus constellatus subsp. pharyngis C818
617121	GCA_000307185.1	Streptococcus dysgalactiae subsp. equisimilis RE378
552526	GCA_000020765.1	Streptococcus equi subsp. zooepidemicus MGCS10565
990317	GCA_000203195.1	Streptococcus gallolyticus subsp. gallolyticus ATCC BAA-2069
467705	GCA_000017005.1	Streptococcus gordonii str. Challis substr. CH1
1069533	GCA_000246835.1	Streptococcus infantarius subsp. infantarius CJ18
1346	GCA_000831485.1	Streptococcus iniae YSFST01-82
1338	GCA_001296205.1	Streptococcus intermedius KCOM 1545
1076934	GCA_000441535.1	Streptococcus lutetiensis 033
1116231	GCA_000283635.1	Streptococcus macedonicus ACA-DC 198
28037	GCA_001281025.1	Streptococcus mitis KCOM 1350 (= ChDC B183)
1437447	GCA_000817065.1	Streptococcus mutans UA159-FR
1302863	GCA_000385925.1	Streptococcus oligofermentans AS 1.3089
927666	GCA_000253155.1	Streptococcus oralis Uo5
1114965	GCA_000262145.1	Streptococcus parasanguinis FW213
873447	GCA_000187935.2	Streptococcus parauberis NCFD 2020
981540	GCA_000270165.1	Streptococcus pasteurianus ATCC 43144
171101	GCA_000007045.1	Streptococcus pneumoniae R6
1054460	GCA_000221985.1	Streptococcus pseudopneumoniae IS7493
160491	GCA_000009385.1	Streptococcus pyogenes str. Manfredo
1304	GCA_000785515.1	Streptococcus salivarius NCTC 8618
388919	GCA_000014205.1	Streptococcus sanguinis SK36
1419814	GCA_000688775.1	Streptococcus sp. VT 162

1307	GCA_000993745.1	Streptococcus suis ZY05719
1308	GCA_000971665.1	Streptococcus thermophilus SMQ-301
218495	GCA_000009545.1	Streptococcus uberis 0140J
1051632	GCA_000219855.1	Sulfobacillus acidophilus TPY
292459	GCA_000009905.1	Symbiobacterium thermophilum IAM 14863
645991	GCA_000190635.1	Syntrophobotulus glycolicus DSM 8271
335541	GCA_000014725.1	Syntrophomonas wolfei subsp. wolfei str. Goettingen G311
643648	GCA_000092405.1	Syntrophothermus lipocalidus DSM 12680
1209989	GCA_000213235.1	Tepidanaerobacter acetatoxydans Re1
586416	GCA_000725365.1	Terribacillus aidingensis MP602
945021	GCA_000283615.1	Tetragenococcus halophilus NBRC 12172
1089553	GCA_000305935.1	Thermacetogenium phaeum DSM 12270
644966	GCA_000184705.1	Thermaerobacter marianensis DSM 12885
635013	GCA_000092945.1	Thermincola potens JR
509193	GCA_000175295.2	Thermoanaerobacter brockii subsp. finnii Ako-1
580331	GCA_000025645.1	Thermoanaerobacter italicus Ab9
2325	GCA_000763575.1	Thermoanaerobacter kivui LKT-1
583358	GCA_000092965.1	Thermoanaerobacter mathranii subsp. mathranii str. A3
340099	GCA_000019085.1	Thermoanaerobacter pseudethanolicus ATCC 33223
399726	GCA_000019065.1	Thermoanaerobacter sp. X514
697303	GCA_000147695.3	Thermoanaerobacter wiegelii Rt8.B1
1094508	GCA_000307585.1	Thermoanaerobacterium saccharolyticum JW/SL-YS485
698948	GCA_000328545.1	Thermoanaerobacterium thermosaccharolyticum M0795
858215	GCA_000189775.3	Thermoanaerobacterium xylanolyticum LX-11
717605	GCA_000227705.3	Thermobacillus composti KWC4
555079	GCA_000144645.1	Thermosediminibacter oceani DSM 16646
479436	GCA_000024945.1	Veillonella parvula DSM 2008
403957	GCA_000725285.1	Virgibacillus sp. SK37
759620	GCA_000750515.1	Weissella ceti WS74
137591	GCA_001308145.1	Weissella cibaria CH2
1045854	GCA_000219805.1	Weissella koreensis KACC 15510

.3 Protéomes utilisés pour le groupe externe.

Identifiant taxonomique	Numéro d'assemblage	Nom de la souche
700015	GCA_000195315.1	Coriobacterium glomerans PW2
525909	GCA_000023265.1	Acidimicrobium ferrooxidans DSM 10331
1313172	GCA_000348785.1	Ilumatobacter coccineus YM16-304
292563	GCA_000317655.1	Cyanobacterium stanieri PCC 7202
551115	GCA_000196515.1	'Nostoc azollae' 0708
59919	GCA_000011465.1	Prochlorococcus marinus subsp. pastoris str. CCMP1986
1140	GCA_000012525.1	Synechococcus elongatus PCC 7942
469383	GCA_000025265.1	Conexibacter woesei DSM 14684
266117	GCA_000014185.1	Rubrobacter xylanophilus DSM 9941

.4 Nombre de protéines ribosomiques identifiées par souche.

Souche	Nombre de protéines ribosomiques identifiées et sélectionnées
Carboxydothermus hydrogenoformans Z-2901	53
Finegoldia magna ATCC 29328	53
Desulfitobacterium dehalogenans ATCC 51507	53
Ruminiclostridium thermocellum DSM 1313	53
Carnobacterium inhibens subsp. gilichinskyi	53
Mahella australiensis 50-1 BON	53
Candidatus Arthromitus sp. SFB-rat-Yit	53
Desulfotomaculum reducens MI-1	53
Desulfotomaculum ruminis DSM 2154	53
Clostridium tetani E88	53
Clostridium sp. SY8519	53
Paenibacillus larvae subsp. larvae DSM 25430	53
Aerococcus urinae ACS-120-V-Col10a	53
Desulfotomaculum acetoxidans DSM 771	53
Desulfotomaculum kuznetsovii DSM 6115	53
Peptoniphilus sp. 1-1	53
Thermincola potens JR	53
Candidatus Desulforudis audaxviator MP104C	53
Acetobacterium woodii DSM 1030	53
Lachnoclostridium phytofermentans ISDg	53
Paenibacillus mucilaginosus KNP414	53
Natranaerobius thermophilus JW/NM-WN-LF	53
Thermobacillus composti KWC4	53
Desulfitobacterium dichloroeliminans LMG P-21439	53
Desulfotomaculum gibsoniae DSM 7213	53
Clostridium lentocellum DSM 5427	53
Pelotomaculum thermopropionicum SI	53
Ammonifex degensii KC4	53
Kyrpidia tusciae DSM 2912	53
Heliobacterium modesticaldum Ice1	53
Desulfitobacterium hafniense Y51	53
Melissococcus plutonius DAT561	53
Anaerococcus prevotii DSM 20548	53
[Clostridium] saccharolyticum WM1	53
Syntrophomonas wolfei subsp. wolfei str. Goettingen G311	53
Pelosinus sp. UFO1	52
Acidaminococcus intestini RyC-MR95	52
Ethanoligenens harbinense YUAN-3	52
Lactobacillus sp. wkB8	52
Ruminococcus bicirculans 80/3	52
Lactobacillus delbrueckii subsp. bulgaricus ND02	52
Geobacillus kaustophilus HTA426	52
Geobacillus stearothermophilus 10	52
Desulfosporosinus acidiphilus SJ4	52
Clostridium scatologenes ATCC 25775	52
Desulfitobacterium metallireducens DSM 15288	52
Acidimicrobium ferrooxidans DSM 10331	52
Weissella ceti WS74	52
Parvimonas micra KCOM 1535 (=ChDC B708)	52
Syntrophothermus lipocalidus DSM 12680	52
Leuconostoc gelidum subsp. gasicomitatum LMG 18811	52
Leuconostoc sp. C2	52
Planococcus kocurii ATCC 43650	52
Lactobacillus ruminis ATCC 27782	52
Weissella koreensis KACC 15510	52
Planococcus rifietoensis M8	52
Ruminococcus albus 7 = DSM 20455	52
Desulfosporosinus orientis DSM 765	52
Geobacillus thermoleovorans CCB_US3_UF5	52
Cyanobacterium stanieri PCC 7202	52
Geobacillus thermodenitrificans NG80-2	52
Brevibacillus brevis NBRC 100599	52
Desulfotomaculum nigrificans CO-1-SRB	52
Clostridium aceticum DSM 1496	52
Paenibacillus naphthalenovorans 32O-Y	52
Veillonella parvula DSM 2008	52
Acidaminococcus fermentans DSM 20731	52
Jeotgalicoccus sp. 13MG44_air	52
Amphibacillus xylanus NBRC 15112	52
Thermoanaerobacterium thermosaccharolyticum M0795	52
[Eubacterium] eligens ATCC 27750	52
Desulfosporosinus meridiei DSM 13257	52
Thermoanaerobacterium saccharolyticum JW/SL-YS485	52
Geobacillus thermoglucosidasius DSM 2542	52
Tepidanaerobacter acetatoxydans Re1	52
Leuconostoc kimchii IMSNU 11154	52
Alicyclobacillus acidocaldarius subsp. acidocaldarius Tc-4-1	52

Anoxybacillus flavithermus WK1	52
Bacillus endophyticus Hbe603	52
Staphylococcus equorum KS1039	52
Clostridium botulinum NCTC 8550	52
Paenibacillus peoriae HS311	52
Geobacillus sp. Y4.1MC1	52
Paenibacillus beijingensis DSM 24997	52
[Clostridium] stercorarium subsp. stercorarium DSM 8532	52
[Clostridium] clariflavum DSM 19732	52
Enterococcus durans KLDS 6.0933	52
Thermosediminibacter oceani DSM 16646	52
Anoxybacillus gonensis G2	52
Clostridium cellulovorans 743B	52
Leuconostoc citreum KM20	52
[Clostridium] cellulosi	52
Staphylococcus agnetis 908	52
Caldicellulosiruptor obsidiansis OB47	52
Enterococcus silesiacus LMG 23085	52
Sulfobacillus acidophilus TPY	52
Exiguobacterium sibiricum 255-15	52
Oenococcus oeni PSU-1	52
Eubacterium acidaminophilum DSM 3953	52
Moorella thermoacetica DSM 521	52
Virgibacillus sp. SK37	52
Thermaerobacter marianensis DSM 12885	52
Enterococcus faecium UW7606x64/3 TC1	52
Exiguobacterium antarcticum B7	52
Staphylococcus capitis AYP1020	52
Lysinibacillus sphaericus C3-41	52
Rubrobacter xylanophilus DSM 9941	52
Brevibacillus laterosporus LMG 15441	52
Clostridium carboxidivorans P7	52
Solibacillus silvestris StLB046	52
Clostridium ljungdahlii DSM 13528	52
Mageeibacillus indolicus UPII9-5	52
Thermacetogenium phaeum DSM 12270	52
Carnobacterium sp. CP1	52
Butyrivibrio proteoclasticus B316	52
Exiguobacterium sp. MH3	52
[Eubacterium rectale] ATCC 33656	52
Leuconostoc mesenteroides subsp. mesenteroides J18	52
Bacillus coagulans S-lac	52
Clostridium butyricum KNU-L09	52
Clostridium pasteurianum NRRL B-598	52
Roseburia hominis A2-183	52
Selenomonas sputigena ATCC 35185	52
Bacillus smithii DSM 4216	52
Thermoanaerobacterium xylanolyticum LX-11	52
Staphylococcus schleiferi 5909-02	52
Clostridium baratii str. Sullivan	51
Lactobacillus mucosae LM1	51
Staphylococcus argenteus MSHR1132	51
Staphylococcus pasteuri SP1	51
Bacillus clausii KSM-K16	51
Thermoanaerobacter brockii subsp. finnii Ako-1	51
Coprothermobacter proteolyticus DSM 5265	51
Caldicellulosiruptor owensensis OL	51
Streptococcus mitis KCOM 1350 (= ChDC B183)	51
Bacillus thuringiensis str. Al Hakam	51
Staphylococcus haemolyticus Sh29/312/L2	51
Lactobacillus casei subsp. casei ATCC 393	51
Staphylococcus aureus subsp. aureus str. Newman	51
Streptococcus sp. VT 162	51
Staphylococcus xylosus SMQ-121	51
Intestinimonas butyriciproducens AF211	51
Caldicellulosiruptor lactoaceticus 6A	51
Bacillus cellulosilyticus DSM 2522	51
Lactobacillus gasseri ATCC 33323 = JCM 1131	51
Bacillus weihenstephanensis WSBC10204	51
Conexibacter woesei DSM 14684	51
Lactococcus garvieae Lg2	51
Lactobacillus sanfranciscensis TMW 1.1304	51
Halobacillus halophilus DSM 2266	51
Lactobacillus plantarum subsp. plantarum ST-III	51
Pediococcus claussenii ATCC BAA-344	51
Clostridium sporogenes NCIMB 10696	51
Alkaliphilus oremlandii OhILAs	51
Dehalobacter sp. DCA	51
Lactobacillus hokkaidonensis JCM 18461	51
Streptococcus anginosus subsp. whileyi MAS624	51

Lactobacillus acidophilus NCFM	51
Enterococcus hirae ATCC 9790	51
Lactobacillus kunkeei MP2	51
Staphylococcus lugdunensis N920143	51
Coriobacterium glomerans PW2	51
Eubacterium sulci ATCC 35585	51
Thermoanaerobacter kivui LKT-1	51
Thermodesulfobium narugense DSM 14796	51
Planococcus sp. PAMC 21323	51
Streptococcus sanguinis SK36	51
Streptococcus pseudopneumoniae IS7493	51
Bacillus lehensis G1	51
'Nostoc azollae' 0708	51
Lactobacillus paracasei subsp. paracasei JCM 8130	51
Clostridium perfringens str. 13	51
Enterococcus faecalis str. Symbioflor 1	51
Pelosinus fermentans JBW45	51
[Clostridium] sticklandii DSM 519	51
Bacillus mycoides ATCC 6462	51
Bacillus cytotoxicus NVH 391-98	51
Clostridium bornimense M2/40	51
Selenomonas ruminantium subsp. lactilytica TAM6421	51
Syntrophobotulus glycolicus DSM 8271	51
Lactobacillus helveticus R0052	51
Salinicoccus halodurans H3B36	51
Streptococcus parasanguinis FW213	51
Leuconostoc carnosum JB16	51
Oscillibacter valericigenes Sjm18-20	51
Clostridium kluyveri NBRC 12016	51
Lysinibacillus varians GY32	51
Lactobacillus fermentum IFO 3956	51
Staphylococcus warneri SG1	51
Bacillus bombysepticus str. Wang	51
Lactobacillus rhamnosus Lc 705	51
Lactobacillus gallinarum HFD4	51
Lactobacillus acetotolerans NBRC 13120	51
Streptococcus pneumoniae R6	51
[Clostridium] cellulolyticum H10	51
Pediococcus pentosaceus SL4	51
Thermoanaerobacter sp. X514	51
Staphylococcus hyicus ATCC 11249	51
Clostridium acetobutylicum EA 2018	51
Synechococcus elongatus PCC 7942	51
Lactobacillus brevis KB290	51
Bacillus halodurans C-125	51
Staphylococcus pseudintermedius HKU10-03	51
Lactobacillus salivarius str. Ren	51
Weissella cibaria CH2	51
Thermoanaerobacter italicus Ab9	51
Clostridium beijerinckii NCIMB 8052	51
Tetragenococcus halophilus NBRC 12172	51
Streptococcus thermophilus SMQ-301	51
Macrococcus caseolyticus JCSC5402	51
Caldicellulosiruptor kristjanssonii I77R1B	51
Lactobacillus amylovorus GRL1118	51
Caldicellulosiruptor bescii DSM 6725	51
Bacillus anthracis str. Vollum	51
Selenomonas sp. oral taxon 478	51
Lactobacillus sakei subsp. sakei 23K	51
[Clostridium] acidurici 9a	51
Lactobacillus johnsonii NCC 533	51
Terribacillus aidingensis MP602	51
Lysinibacillus fusiformis RB-21	51
Staphylococcus saprophyticus subsp. saprophyticus ATCC 15305	51
Clostridium autoethanogenum DSM 10061	51
Caldicellulosiruptor hydrothermalis 108	51
Bacillus methanolicus MGA3	51
Thermoanaerobacter pseudethanolicus ATCC 33223	51
Streptococcus gordonii str. Challis substr. CH1	51
Symbiobacterium thermophilum IAM 14863	51
Staphylococcus carnosus subsp. carnosus TM300	51
Streptococcus oralis Uo5	51
[Bacillus] selenitireducens MLS10	51
Caldicellulosiruptor kronotskyensis 2002	51
Caldanaerobacter subterraneus subsp. tengcongensis MB4	51
Clostridium saccharobutylicum DSM 13864	51
Staphylococcus epidermidis SEI	51
Streptococcus salivarius NCTC 8618	51
Bacillus toyonensis BCT-7112	51
Thermoanaerobacter wiegelii Rt8.B1	51

Alkaliphilus metalliredigens QYMF	51
Halobacteroides halobius DSM 5150	50
Streptococcus pasteurianus ATCC 43144	50
Streptococcus infantarius subsp. infantarius CJ18	50
Caldicellulosiruptor saccharolyticus DSM 8903	50
Bacillus sp. YP1	50
Dehalobacter restrictus DSM 9455	50
Streptococcus intermedius KCOM 1545	50
Bacillus subtilis subsp. subtilis str. RO-NN-1	50
Bacillus pumilus W3	50
Paenibacillus sabinae T27	50
Listeria welshimeri serovar 6b str. SLCC5334	50
Paenibacillus riograndensis SBR5	50
Jeotgalibacillus malaysiensis	50
Streptococcus oligofermentans AS 1.3089	50
Lactobacillus buchneri NRRL B-30929	50
Streptococcus iniae YSFST01-82	50
Lactobacillus reuteri TD1	50
Clostridium saccharoperbutylacetonicum N1-4(HMT)	50
Halanaerobium hydrogeniformans missing	50
Lactobacillus heilongjiangensis DSM 28069	50
Streptococcus pyogenes str. Manfredo	50
Enterococcus mundtii QU 25	50
Streptococcus constellatus subsp. pharyngis C818	50
Streptococcus parauberis NCFD 2020	50
Streptococcus dysgalactiae subsp. equisimilis RE378	50
Bacillus pseudofirmus OF4	50
Oceanobacillus iheyensis HTE831	50
Lactobacillus kefiranofaciens ZW3	50
Streptococcus mutans UA159-FR	50
Enterococcus casseliflavus EC20	50
Clostridium novyi NT	50
Acetohalobium arabaticum DSM 5501	50
Paenibacillus stellifer DSM 14472	50
Streptococcus equi subsp. zooepidemicus MGCS10565	50
Lactobacillus ginsenosidimutans EMML 3041	50
Streptococcus uberis 0140J	50
Bacillus cereus biovar anthracis str. CI	50
Bacillus atrophaeus NRS 1221A	50
Paenibacillus polymyxa Sb3-1	50
Streptococcus agalactiae SS1	50
Paenibacillus terrae HPL-003	50
Streptococcus suis ZY05719	50
Ilumatobacter coccineus YM16-304	50
Bacillus licheniformis DSM 13 = ATCC 14580	50
Streptococcus lutetiensis 033	50
Herbinix sp. SD1D	50
Listeria ivanovii subsp. londoniensis	50
Megasphaera elsdenii 14-14	50
Lactococcus lactis subsp. lactis NCDO 2118	50
Listeria monocytogenes serotype 4b str. LL195	50
Prochlorococcus marinus subsp. pastoris str. CCMP1986	50
Listeria seeligeri serovar 1/2b str. SLCC3954	50
Bacillus infantis NRRL B-14911	50
Paenibacillus durus DSM 1735	50
Streptococcus macedonicus ACA-DC 198	50
Paenibacillus bovis BD3526	50
Paenibacillus odorifer DSM 15391	49
Lactobacillus koreensis 26-25	49
Streptococcus gallolyticus subsp. gallolyticus ATCC BAA-2069	49
Paenibacillus borealis DSM 13188	49
Paenibacillus graminis DSM 15220	49
Peptoclostridium difficile Z31	49
Filifactor alocis ATCC 35896	49
Halanaerobium praevalens DSM 2228	49
Bacillus amyloliquefaciens Y2	49
Lactococcus piscium MKFS47	49
Bacillus velezensis YJ11-1-4	49
Carnobacterium maltaromaticum LMA28	49
Bacillus megaterium WSH-002	49
Erysipelothrix rhusiopathiae str. Fujisawa	49
Oenococcus kitaharae DSM 17330	49
Eubacterium limosum SA11	49
Bacillus paralicheniformis BL-09	48
Paenibacillus sp. Y412MC10	48
Lactobacillus paraplantarum L-ZS9	48
Halothermothrix orenii H 168	48
Thermoanaerobacter mathranii subsp. mathranii str. A3	44

.5 Nombre de séquences et de positions par alignement de protéines ribosomiques.

315 protéomes sont représentés.

Protéine ribosomique	nombre de séquences	nombre de positions
bL12	313	103
bL17	314	104
bL19	314	108
bL20	313	118
bL21	315	99
bL25	167	163
bL27	315	84
bL28	312	60
bL31	291	58
bL32	261	49
bL33	197	46
bL34	303	44
bL35	315	62
bL36	310	37
bL9	305	147
bS16	315	79
bS18	313	67
bS20	315	84
bS21	300	48
bS6	315	87
uL1	313	227
uL10	314	156
uL11	313	139
uL13	314	141
uL14	315	122
uL15	315	136
uL16	315	139
uL18	315	118
uL2	315	275
uL22	315	105
uL23	314	83
uL24	315	88
uL29	314	55
uL3	315	195
uL30	280	56
uL4	314	195
uL5	315	176
uL6	315	170
uS10	314	100
uS11	315	122
uS12	313	123
uS13	315	121
uS14	226	61
uS15	315	88
uS17	315	80
uS19	315	89
uS2	315	223
uS3	315	196
uS4	278	187
uS5	315	156
uS7	313	156
uS8	315	126
uS9	311	127

.6 Récapitulatif des familles reconstruites.

Les séquences en gras ont été prouvées expérimentalement selon Uniprot, les séquences en italique ont été revues selon Uniprot, les autres sont seulement inférées. SP : *Streptoccoccus pneumoniae* R6, SPD : *Streptoccoccus pneumoniae* D39, BS : *Bacillus subtilis*, EC : *Escherichia coli*, Lf : Lactobacillus fermentum, Efm : *Enterococcus faecium*, EF : *Enterococcus faecalis*, PA : *Pseudomonas aeruginosa*, PP : *Pseudomonas putida*, Ssp : *Streptomyces sp.*, CA : *Clostridium acetobutylicum*, AF : *Anoxybacillus flavithermus* , BPL : *Bacillus paralicheniformis*, MeThe : *Methanothermobacter thermautotrophicus*, MX : *Myxococcus xanthus*, SA : *Staphylococcus aureus*. ajout de FN : des séquences rejetées par la procédure ont été ajoutées au jeu de données. nettoyage : des séquences ont été enlevées de l'alignement car considérées comme non- homologues. hF/bF/W/B : cf matériel et méthodes. Les familles multigéniques sont présentées dans des cases grisées tandis que familles/sous-familles monogéniques sont dans des cases blanches.

Familles multigéniques		Construction des familles						
Famille	Sous-familles (orthologues)	Graines	détection des homologues	Homologues	Pseudogènes	Indice de confiance	Incertitude au niveau de la topologie	Informations discriminants supplémentaires
1 AlaS		**NP_417177.1(EC)**	défaut	2372	10			
	AlaS, LovB			940	10	2		Contexte (RuvA)
2 Alr PCDP6		*NP_359133.1(SP)* **NP_389646.1(BS)** NP_359102.1(SP)	défaut	2485	10			
	Alr, DadX(EC)			1185	3	2		Contexte (PemK, AcpS)
	Alr-like			468	2	0	groupes éparses, incohérences taxonomiques	
	PCDP6			832	5	2		Contexte (FtsZ, FtsA, SepF, ...)
3 Aslfm		**WP_002294441.1(EF)**	défaut + ajout de FN	1329	2			
	Aslfm			168	2	1	1 séquence douteuse	
4 CDP1		**WP_000363002.1(SPD)**	défaut	1574	6			
	CDP1			647	6	1	1 groupe basal exclu peut être à tord	Contexte (TrmB)
5 CDP4		NP_358993.1(SP)	défaut+ ajout de FN + recherche par profil HMM+ ajout de FN + nettoyage	726	4			
	CDP4			587	4	2		Contexte (AspB, AsnC)
6 CozE		**WP_000489841.1(SP)**	défaut	3293	1			
	CozE1			436	1	1	1 groupe basal exclu peut être à tord	
	CozE2			232	0	1	manque de signal phylogénétique compensé par élagage	
	CozE3			709	0	1	manque de signal phylogénétique compensé par élagage	
	CozE4			382	0	1	manque de signal phylogénétique compensé par élagage	Contexte (AlaS)
7 CpsD ParA MinD		**WP_001142534.1(SP) NP_390677.1(BS) NP_391977.1(BS)**	défaut	3928	1			
	CpsD			915	0	1	quelques séquences basales douteuses	Contexte (CpsB, CpsC)
	MinD			545	1	1	1 groupe basal douteux	Contexte (MinC)
	ParA, Soj			618	0	1	quelques séquences basales douteuses	Contexte (ParB)
8 DacA DacB DacF		**NP_387891.1(BS) NP_390200.1(BS)**	défaut	2732	45			
	DacA(BS), PBP5(BS), DacC(EC), PBP6(EC)			1388	34	2		
	DacB(BS), PBP5*(BS), DacD(EC), PBP6b(EC)			387	2	0	manque de signal phylogénétique compensé par élagage	Contexte (ScpA, ScpB) (synténie partielle)
	DacF(BS), DacA(EC), PBP5(EC)			404	7	1	1 séquence basale douteuse	Contexte (XerD)
9 DdI		*NP_359108.1(SP)*	défaut	6829	2			
	DdI			882	2	1	1 groupe basal douteux	
10 DivIC FtsL		*NP_357897.1(SP)* NP_357602.1(SP)	défaut + nettoyage + recherche par profil HMM + ajout de FN	1985	5			
	DivIC(BS), FtsB(EC)			1049	2	1	manque de signal phylogénétique	Contexte (Mfd, TilS, FtsH, hpt)
	FtsL, MraR			936	3	1	manque de signal phylogénétique	Contexte (MraW, MraZ, ...)
11 DivIVA GpsB		**NP_359098.1(SP)** *NP_357926.1(SP)*	défaut	1668	12			
	DivIVA			919	9	2		Contexte (IleS, FtsA, FtsZ, ...)
	GpsB			749	3	2		Contexte (RecU, PBPA3, PCDP10)
12 DnaA DnaI		*NP_357595.1(SP)* **NP_391924.2(BS)**	défaut	2934	13			
	DnaA			952	2	1	quelques séquences basales douteuses	Contexte (DnaN, RecF, GyrB , ...)
	DnaI			772	4	2		Contexte (DnaB(BS))
13 DnaB		**NP_418476.1(EC)**	défaut	1285	4			
	DnaB1(EC), DnaC(BS)			964	11	2		Contexte (RplL)
	DnaB2			182	0	1	incohérence taxonomique	
14 DnaD		NP_390116.1(BS)	défaut	1587				
	DnaD			1010	0	2		
15 DnaG		NP_417538.1(EC)	défaut	1053				
	DnaG			942	12	2	monogénique ?	
16 DnaN		**NP_418156.1(EC)**	défaut	1090	1			
	DnaN			1072	1	2	monogénique ?	Contexte (DnaA, RecF, GyrB , ...)
17 FtsA Mbl MreB MreBH		NP_359104.1(SP) *NP_390681.2(BS)*	défaut	3707	5			
	FtsA			865	2	1	quelques séquences basales douteuses	Contexte (FtsZ, ...)
	Mbl			594	1	1	manque de signal phylogénétique compensé par élagage	Contexte (MurA SpoIIID)
	MreB			601	2	1	manque de signal phylogénétique compensé par élagage	Contexte (MreC, MreD)
	MreBH			194	0	1	manque de signal phylogénétique compensé par élagage	
18 FtsE		NP_358260.1(SP)	1 blastp	744	0			
	FtsE			744	0	2		Contexte (FtsX)
19 FtsH ClpX SpoVK		*NP_357606.1(SP)* *NP_359020.1(SP)*	défaut	5671	28			
	ClpX, LocP			955	9	2		Contexte (Tig, ClpP)
	FtsH1, HflB, MrsC, Std, TolZ			942	13	2		Contexte (TilS, Hpt)
	FtsH2			97	3	2		
	FtsH3			173	2	2		
	SpoVK, SpoVJ			423	10	1	1 groupe basal douteux	
20 FtsJ		*NP_390306.2(BS)*	défaut	1828	8			
	FtsJ, MrsF, RrmJ			840	8	2		Contexte (RecN, AhrC)
21 FtsK		**NP_358375.1(SP) WP_003245350.1(BS)**	défaut	2515	27			
	FtsK1(EC), SpoIIIE(BS)			957	11	2		

	FtsK2(EC), SftA(BS)			437	16	2		Contexte (MurC, tRNA binding protein)
22 FtsW RodA SpoVE (SEDS)		NP_358306.1(SP) NP_358567.1(SP)	défaut	2910	27			
	FtsW			1062	12	1	manque de signal phylogénétique compensé par élagage	Contexte (Ppc, Tuf)
	RodA, MrdB			1240	11	1	manque de signal phylogénétique compensé par élagage	
	RodA2			177	0	1	manque de signal phylogénétique compensé par élagage	Contexte (LstR)
	SpoVE			431	4	2		Contexte (MurD, MurG, MraY)
23 FtsX		NP_358261.1(SP)	défaut	3030	3			
	FtsX, ftsS			760	3	2		Contexte (FtsE)
24 FtsY		**NP_417921.1(EC)**	défaut	2325	8			
	FtsY			940	8	2		Contexte (Smc)
25 GatD		**WP_010876345.1(MeThe)**	défaut	1012	4			
	GatD1			561	3	2		
	GatD2			115	0	1	incohérence taxonomique	
	GatD3			106	0	1		Contexte (PrsW)
	GatD4			228	1	1	incohérence taxonomique	
26 GidA		**NP_418197.1(EC)**	défaut	1816	6			
	GidA, MnmG, TrmF			920	6	2		Contexte (MnmE, gidB, Noc, ParA, ...)
27 GlmM		**NP_359010.1(SP)**	défaut	2432	5			
	GlmM, MrsA, FemD(SA)			937	5	2		Contexte (GlmS)
28 GlmS		*NP_357839.1(SP)*	défaut	3495	4			
	GlmS			944	4	2		Contexte (GlmM)
29 GlmU		**NP_358485.1(SP)**	défaut	9133	4			
	GlmU			936	4	1	groupe basal exclu peut être à tord	Contexte (Prs)
30 IleS ValS LeuS		**NP_414567.1(EC)**	défaut	4955	21			
	IleS1			809	8	2		Contexte (DivIVA, FtsZ, FtsA, ...)
	IleS2			239	0	2		
	LeuS			939	7	2		Contexte (MurJ)
	ValS			941	6	2		Contexte (FolC, MreC, MreD) (Synténie partielle)
31 LysA		**NP_417315.1(EC)**	défaut	1130	1			
	LysA1			688	1	2		
	LysA2			165	0	1	incohérence taxonomique	
	LysA3			119	0	1	incohérence taxonomique	
32 LytB(SP)		**NP_358461.1(SP)**	recherche domaine PASTA/Pkinase	35	0			
	LytB(SP)			35	0	2		Domaine (CW_binding_1, Glucosaminidase
33 MraW		*NP_357896.1(SP)*	défaut	1929	5			
	MraW, RsmH			953	5	2		Contexte (MraZ, FtsL, ...)
34 MraY WecA		*NP_357899.1(SP)* **NP_418231.1(EC)**	défaut+ nettoyage+recher che par profil HMM+recherche par profil HMM	1951	9			
	MraY, MurX			962	6	2		Contexte (MurE, murF, MurD, ...)
	WecA, Rfe			928	3	2		
35 Mtf		*NP_359172.1(SP)*	défaut	2039	6			
	Mtf, Fmt			940	6	2		Contexte (StkP, PhpP, SunL, ...)
36 MurA MurZ		*NP_359373.1(SP)* *NP_358583.1(SP)*	défaut	2756	11			
	MurA1			853	5	2		Contexte (SpoIIID)
	MurA2			129	1	2		Contexte (MurG, FtsQ, ftsA, ...)
	MurZ			870	5	2		
37 MurB		*NP_358840.1(SP)*	défaut	2164	3			
	MurB1			795	2	1	groupe basal inclus peut être à tord	
	MurB2			247	1	1	groupe basal exclu peut être à tord	Contexte (MurG, MurC, MurA, ...)
38 MurC MurD MurE MurF MurT		*NP_358966.1(SP)* *NP_358197.1(SP)* *NP_358977.1(SP)* NP_359107.1(SP) NP_359036.1(SP)	défaut	5829	34			
	MurC			939	4	2		
	MurD			939	3	2		Contexte (MurG, MraY, ...)
	MurE			1090	15	1	groupe interne inclus peut être à tord	Contexte (MurF, MraY, ...)
	MurF			1034	8	2		Contexte (Ddl/MurE, MraY,...)
	MurT			557	4	2		Contexte (Pfp1)
39 MurG		*NP_358198.1(SP)*	défaut	1844	7			
	MurG1			549	5	2		Contexte (MurD, FtsQ, FtsA, ...)
	MurG2			532	2	2		Contexte (MurD, FtsQ, FtsA, ...)
40 MurI		*NP_359288.1(SP)*	défaut	2626	11			
	Asr			596	5	1	incohérence taxonomique	
	MurI			1074	6	2		
41 MurJ YabM YkvU SpoVB		**NP_415587.1(EC)**	défaut	3090	6			
	MurJ, MviN			992	6	0	Sous-familles encastrées	
	YabM			851	0	1	Sous-familles encastrées	Contexte (Mfd)
	YkvU			95	0	0	Sous-familles encastrées	
	SpoVB, SpoIIIF			399	0	0	Sous-familles encastrées	
42 MurK		**NP_346825.1(CA)**	défaut	309	6			
	MurK			285	6	2		
43 MurN MurM FemA FemB FemX fmh		NP_358134.1(SP) NP_358135.1(SP)	défaut	1037	32			
	FemA			95	0	2		Contexte (FemB, fmh)
	FemB			95	0	2		Contexte (FemA, fmh)
	FemX			161	1	2		
	Fmh			161	7	2		Contexte (FemA, FemB)
	MurM			264	12	2		Contexte (MurN)
	MurN			261	12	2		Contexte (MurM)
44 MurQ		**NP_416923.1(EC)**	défaut	884	1			
	MurQ			572	1	2		
45 NagA		**NP_415203.1(EC)**	défaut	4780	5			
	NagA			1010	5	2		Contexte (NagB)
46 NagB		**NP_415204.1(EC)**	défaut	1201	3			
	NagB, GlmD			896	3	2		Contexte (NagA)

47 Noc ParB		NP_359613.1(SP) **NP_417715.1(EC)** **NP_390679.1(BS)**	défaut	1897	11			
	Noc			617	1	2		Contexte (GidB, ParA, ...)
	ParB, Spo0J			937	10	2		Contexte (GidB, Noc, ParA, ...)
48 NudF		NP_358486.1(SP)	défaut	6439	1			
	NudF			924	1	1	manque de signal phylogénétique compensé par élagage	Contexte (Pfs/XerD)
49 PBPs		*NP_388295.1(BS)* **NP_357923.1(SP)** NP_359500.1(SP) NP_359415.1(SP) **NP_359110.1(SP)** **NP_357898.1(SP)**	défaut	6425	92			
	A3, PBP1a(SP), PBP1(BS)			814	12	2		Contexte (RecU, PCDp10, GpsB)
	A4, PBP2a(SP), PBP2c(BS)			763	13	0	manque de signal phylogénétique compensé par élagage, Sous-familles encastrées	
	A5, PBP1b,(SP), PBP4(BS)			116	1	0	manque de signal phylogénétique compensé par élagage, Sous-familles encastrées	
	A6, PBP1c(EC)			286	5	0	manque de signal phylogénétique compensé par élagage, Sous-familles encastrées	
	AX1			628	22	0	manque de signal phylogénétique compensé par élagage, Sous-familles encastrées	
	B1, PBP3(BS)			450	4	2		
	B3, SpoVD(BS), PBP3(EC)			253	4	2		Contexte (MraZ, MraW, FtsL, B4)
	B4, PBP2x(SP), PBP2b(BS)			1085	18	2		Contexte (MraZ, MraW, FtsL)
	B5, PBP2b(SP), PBP2a(BS), PBPH(BS)			1340	5	2		
	B6, PBP4b(BS)			331	3	2		
	BX1			126	1	1	peut être inclus dans un autre groupe	Contexte (FtsW)
	MGT1			94	1	2		Contexte (RecX)
	MGT2			89	3	2		Contexte (TrxA)
50 Pfs		NP_358488.1(SP)	défaut	2231	2			
	Pfs, MtnM			966	2	1	incohérence taxonomique	Contexte (MacP)
51 PhpP		*NP_359170.1(SP)*	défaut	1102	2			
	PhpP, PppL			936	2	2		Contexte (StkP, SunL, ...)
52 PriA, Mfd, RecG		NP_359173.1(SP)	défaut	7733	27			
	Mfd			936	13	2		Contexte (Pth)
	RecG			947	6	2		Contexte (RecN)
	PriA			942	8	2		Contexte (Fmt, SunL, StkP, ...)
53 RacA		**NP_391584.1(BS)**	défaut	473	1			
	RacA			211	1	1	groupe basal exclu peut être à tord	
54 Smc RecN RecF SbcC		NP_358719.1(SP)	défaut	3664	19			
	Smc			941	19	2		Contexte (FtsY)
	RecN			938	10	2		Contexte (FtsJ, AhrC)
	RecF			937	6	2		Contexte (DnaA, GyrB , ...)
	SbcC1			442	3	2		Contexte (SbcD)
	SbcC2			210	2	2		Contexte (SbcD)
55 SpoIID LytB(BS)		*NP_391556.1(BS)*	défaut	684	4			
	SpoIID, SpoIIC			402	2	2		Contexte (MurA Mbl)
	LytB(BS), CwbA			239	2	2		
56 SpoIIE		**NP_387945.1(BS)**	défaut	1260	1			
	SpoIIE, SpoIIH			401	1	2		Contexte (TilS, FtsH, ...)
57 SpoIIIJ		**NP_391984.1(BS)**	défaut	1636	11			
	SpoIIIJ1, MisCA			854	6	2		Contexte (RnpA, Jag, GidA, ...)
	SpoIIIJ2, MisCB			752	5	2		Contexte (AcyP)
58 StkP		**NP_359169.1(SP)**	défaut	1870	21			
	StkP(SP), PknB(MT), Stk1(SA), PrkC(BS)			939	21	2		Contexte (PhpP, SunL, ...) Domaine (PASTA, Pkinase)
59 SunL		NP_359171.1(SP)	défaut	2560	11			
	SunL			934	11	2		Contexte (StkP,P hpP, ...)
60 TilS		**NP_414730.1(EC)**	défaut	2105	15			
	TilS, MesJ			876	15	2		Contexte (FtsH, ...)
61 TolQ		**NP_415265.1(EC)**	blastp + recherche par profil HMM	715	0			
	TolQ			32	0	2		Contexte (TolR)
62 WalH WalI		WP_020453618.1(BPL) YP_002317174.1(AF)	hF+bF+W+B/défaut + recherche par profil HMM + ajout de FN + nettoyage + ajout de FN	1190	8			
	WalH			595	6	2		Contexte (WalI, WalK, ...)
	WalI			595	2	2		Contexte (WalH, WalK, ..)
63 WalJ		WP_003177960.1(BPL)	défaut	6750	9			
	WalJ			903	9	1	groupe basal exclu peut être à tord	Contexte (WalH, WalI, WalK, ...)
64 WalK		**WP_000871607.1(SA)**	défaut	1818	0			
	WalK, VicK			800	5	2		Contexte (WalR)
65 WalR		**WP_000101976.1(SA)**	défaut	8565	0			
	WalR			775	0	2		Contexte (WalK)
66 XerC XerD XerS		**NP_358640.1(SP)**	défaut	6627	14			
	XerC			553	3	2		Contexte (TrmF0)
	XerD, XprB, RipX			701	6	2		
	XerS	809		458	5	1	manque de signal phylogénétique compensé par élagage	

Familles monogéniques		**Construction des familles**						
	Famille (orthologue)	**Graines**	**détection des homologues**	**Homologues**	**Pseudogènes**	**Indice de confiance**		
	AnmK	**NP_416157.1(EC)**	défaut	130	6	2		
	CDP3(RocS)	NP_358489.1(SP)	bF + hF + sélection des copies chromosomiques + nettoyage	282	0	2		
	CpsB, Cap1B, Wzb	**WP_000567603.1(SPD)**	défaut	727	11	2		

	CpsC, Wzd	**WP_000664148.1(SPD)**	défaut + nettoyage	1032	3	2		
	DacC(BS), PBP4a(BS), DacB(EC), PBP4(EC)	**NP_389717.1(BS)**	défaut + nettoyage	101	0	2		
	DnaB(BS)	**NP_390777.1(BS)**	défaut	762	3	2		
	EzrA	**NP_358310.1(SP)**	défaut	725	4	2		
	FtsQ(EC), DivIB(BS)	NP_358199.1(SP) **NP_389407.1(BS)**	défaut	937	16	2		
	FtsZ	**NP_359103.1(SP)**	défaut	954	9	2		
	GidB, RsmG	**NP_418196.1(EC)**	défaut + nettoyage	938	9	2		
	Jag, EloR	*NP_391983.1(BS)*	défaut	782	8	2		
	MacP	NP_358487.1(SP) WP_003565192.1(SP)	défaut + nettoyage + recherche par profil HMM + nettoyage + recherche par profil HMM + nettoyage + recherche par profil HMM + nettoyage	642	1	2		
	Maf	**NP_390683.1(BS)**	défaut	500	2	2		
	MapZ, LocZ	**NP_357928.1(SP)**	défaut	222	3	2		
	MinC	**NP_390678.1(BS)**	défaut	549	1	2		
	MinE	**NP_415692.1(EC)**	défaut + nettoyage	159	1	2		
	MinJ	**NP_391402.1(BS)**	défaut + nettoyage	473	13	2		
	MraZ	*NP_389396.1(BS)*	défaut + nettoyage	591	7	2		
	MreC	NP_359614.1(SP)	défaut	865	12	2		
	MreD	NP_359613.1(SP) **NP_417715.1(EC)** **NP_390679.1(BS)**	défaut + nettoyage + recherche par profil HMM_reject	852	6	2		
	PCDP10	NP_357925.1(SP)	défaut	769	2	2		
	PCDP4	NP_359105.1(SP)	défaut	429	3	2		
	PCDP7	NP_359100.1(SP)	défaut	898	15	2		
	PCDP8	NP_359099.1(SP)	défaut + nettoyage	913	16	2		
	RadC	NP_390682.1(BS)	défaut	1042	16	2		
	RecA, LexB, RecH, RnmB	**NP_389576.2(BS)**	défaut	996	56	2		
	RecR, RecM	NP_359109.1(SP)	défaut	938	1	2		
	RecU, PrfA	NP_357924.1(SP)	défaut + nettoyage	819	5	2		
	RodZ	WP_000137488.1(BTH) **NP_417011.1(EC)**	défaut + nettoyage + nettoyage + nettoyage	809	9	2		
	Rtp	**NP_389731.1(BS)**	défaut	123	0	2		
	ScpA	NP_359283.1(SP)	défaut	928	4	2		
	ScpB	**NP_359282.1(SP)**	défaut + nettoyage	930	5	2		
	SepF	NP_359101.1(SP) WP_014479735.1(BS)	bF + hF	941	5	2		
	Spo0M	**NP_388756.2(BS)**	défaut	255	4	2		
	SpoIIGA	**NP_389414.1(BS)**	défaut	400	4	2		
	SpoIIID	**NP_391523.1(BS)**	défaut	416	1	2		
	SpoVG	**NP_387930.1(BS)**	défaut + nettoyage	686	1	2		
	TolR	**NP_415266.1(EC)**	défaut + nettoyage	35	0	2		
	ZapA	NP_357960.1(SP) **NP_390739.1(BS)**	défaut	865	4	2		
Familles absentes		**Construction des familles**						
	Protéine	**Graines**	**détection des homologues**	**E-value la plus faible**	**Famille détectée**	**Observation**		
	AmgK	**CSHYR5_9ACTN(PP)**	défaut	2,50E-001	hypothetical protein	Non significatif		
	AmpG	**NP_414967.1(EC)**	défaut	2,00E-007	multidrug transporter	Homologues non orthologue		
	FtsN	**NP_418368.1(EC)**	défaut	4,00E-005	N-acetylmuramoyl-L-alanine amidase	Non significatif		
	LpoB	**NP_415623.1(EC)**	défaut	0,019	YSIRK signal domain protein	Non significatif		
	MatP	**NP_415476.1(EC)**	défaut	-	-	Pas d'homologue		
	MipZ	**NP_420968.1(CC)**	défaut	2,00E-008	ParA	Homologues non orthologue		
	MpaA	**NP_415842.2(EC)**	défaut	2,00E-005	peptidase M14	Non significatif		
	Mpl	**NP_418654.1(EC)**	défaut	1,00E-056	MurE	Homologues non orthologue		
	MukB	**NP_415444.1(EC)**	défaut	2,00E-006	Smc	Homologues non orthologue		
	MukE	**NP_415443.2(EC)**	défaut	-	-	Pas d'homologue		
	MukF	**NP_415442.1(EC)**	défaut	-	-	Pas d'homologue		
	MurU	**NP_742572.1(PP)**	défaut	7,00E-031	nucleotidyltransferase	Homologues non orthologue		
	Pal	**NP_415269.1(EC)**	défaut	9,00E-009	OmpA/MotB domain-containing protein	Homologues non orthologue		
	PomZ	**WP_011550766.1(MX)**	défaut	1,00E-043	ParA	Homologues non orthologue		
	SeqA	**NP_415213.1(EC)**	défaut	-	-	Pas d'homologue		
	SlmA	**NP_418098.4(EC)**	défaut	3,00E-012	TetR family regulator	Homologues non orthologue		
	SsgB	**NP_625820.1(SC)**	défaut	-	-	Pas d'homologue		
	SulA	**NP_415478.1(EC)**	défaut	-	-	Pas d'homologue		
	TolA	**NP_415267.1(EC)**	défaut	-	-	Pas d'homologue		
	Tus	**NP_416127.1(EC)**	défaut	-	-	Pas d'homologue		
	ZapB	**NP_418363.1(EC)**	défaut	0,084	transporter	Non significatif		
	ZapC	**NP_415466.4(EC)**	défaut	-	-	Pas d'homologue		
	ZapD	**NP_414644.1(EC)**	défaut	-	-	Pas d'homologue		
	ZipA	**NP_416907.1(EC)**	défaut	2,00E-004	zinc metalloprotease	Non significatif		

.7 Rôles biologiques des familles reconstruites.

SP : *Streptoccoccus pneumoniae* R6, SPD : *Streptoccoccus pneumoniae* D39, BS : *Bacillus subtilis*, EC : *Escherichia coli*, Lf : Lactobacillus fermentum, Efm : *Enterococcus faecium*, EF : *Enterococcus faecalis*, PA : *Pseudomonas aeruginosa*, PP : *Pseudomonas putida*, Ssp : *Streptomyces sp.*, CA : *Clostridium acetobutylicum*, AF : *Anoxybacillus flavithermus* , BPL : *Bacillus paralicheniformis*, MeThe : *Methanothermobacter thermautotrophicus*, MX : *Myxococcus xanthus*, SA : *Staphylococcus aureus*. Les rôles biologiques avec un « ? » n'ont jamais été prouvées expérimentalement. Les familles multigéniques sont présentées dans des cases grisées tandis que familles/sous-familles monogéniques sont dans des cases blanches.

Familles multigéniques		Rôle biologique		
Famille	**Sous-familles (orthologues)**	**Inférence du rôle dans le cycle cellulaire**	**Activité biochimique**	**Processus cellulaire**
1 AlaS				
	AlaS, LovB	prouvé expérimentalement (e.g. EC)	Alanine-tRNA synthetase	Traduction
2 Alr PCDP6				
	Alr, DadX(EC)	prouvé expérimentalement (e.g. BS)	Alanine racémase	Synthèse du peptidoglycane
	Alr-like	par homologie	Alanine racémase ?	Synthèse du peptidoglycane ?
	PCDP6	contexte génomique	Alanine racémase ?	Synthèse du peptidoglycane ?
3 Aslfm				
	Aslfm	prouvé expérimentalement (e.g. EF)	D-Aspartate ligase	Synthèse du peptidoglycane
4 CDP1				
	CDP1	prouvé expérimentalement (e.g. SP) NON PUBLIE		Couplage ségrégation-division
5 CDP4				
	CDP4	prouvé expérimentalement (e.g. SP) NON PUBLIE		Régulation de l'activité PBP
6 CozE				
	CozE1	prouvé expérimentalement (e.g. SP)	Transporteur?	Régulation de l'activité PBP, Élongation
	CozE2	prouvé expérimentalement (e.g. SA)	Transporteur?	Régulation de l'activité PBP, Élongation
	CozE3	prouvé expérimentalement (e.g. SA)	Transporteur?	Régulation de l'activité PBP, Élongation
	CozE4	par homologie	Transporteur?	Régulation de l'activité PBP, Élongation ?
7 CpsD ParA MinD				
	CpsD	prouvé expérimentalement (e.g. SP)	Tyrosine kinase	Régulation synthèse capsule
	MinD	prouvé expérimentalement (e.g. BS)	ATPase	Positionnement de l'anneau Z
	ParA, Soj	prouvé expérimentalement (e.g. SP,BS)	ATPase	Ségrégation chromosome
8 DacA DacB DacF				
	DacA(BS), PBP5(BS), DacC(EC), PBP6(EC)	prouvé expérimentalement (e.g. BS)	DD-Carboxypeptidase	Réticulation du peptidoglycane
	DacB(BS), PBP5*(BS), DacD(EC), PBP6b(EC)	prouvé expérimentalement (e.g. BS)	DD-Carboxypeptidase	Réticulation du peptidoglycane
	DacF(BS), DacA(EC), PBP5(EC)	prouvé expérimentalement (e.g. BS)	DD-Carboxypeptidase	Réticulation du peptidoglycane
9 Ddl				
	Ddl	prouvé expérimentalement (e.g. BA)	D-Ala-D-Ala ligase	Synthèse du peptidoglycane
10 DivIC FtsL				
	DivIC(BS), FtsB(EC)	prouvé expérimentalement (e.g. BS)		Division
	FtsL, MraR	prouvé expérimentalement (e.g. BS)	Zinc transporteur	Division
11 DivIVA GpsB				
	DivIVA	prouvé expérimentalement (e.g. SP,BS)		Positionnement de l'anneau Z, Couplage division-élongation
	GpsB	prouvé expérimentalement (e.g. SP,BS)		Couplage division-élongation
12 DnaA DnaI				
	DnaA	prouvé expérimentalement (e.g. BS)	ATPase	Initiation de la réplication
	DnaI	prouvé expérimentalement (e.g. BS)	ATPase	Réamorcage de la réplication
13 DnaB				
	DnaB1(EC), DnaC(BS)	prouvé expérimentalement (e.g. BS)	ATPase, hélicase	Initiation de la réplication
	DnaB2	par homologie	ATPase, hélicase ?	Initiation de la réplication ?
14 DnaD				
	DnaD	prouvé expérimentalement (e.g. BS)		Élongation de la réplication
15 DnaG				
	DnaG	prouvé expérimentalement (e.g. EC)		Élongation de la réplication
16 DnaN				
	DnaN	prouvé expérimentalement (e.g. BS)		Élongation de la réplication
17 FtsA Mbl MreB MreBH				
	FtsA	prouvé expérimentalement (e.g. BS)	ATPase	Ancrage à la membrane de l'anneau Z
	Mbl	prouvé expérimentalement (e.g. BS)	ATPase	Échafaudage de l'élongasome
	MreB	prouvé expérimentalement (e.g. BS)	ATPase	Échafaudage de l'élongasome
	MreBH	prouvé expérimentalement (e.g. BS)	ATPase	Échafaudage de l'élongasome
18 FtsE				
	FtsE	prouvé expérimentalement (e.g. BS)	sous unité ABC transporteur	Division
19 FtsH ClpX SpoVK				
	ClpX, LocP	prouvé expérimentalement (e.g. BS)	sous unité fixant l'ATP de la protease ClpXP	Régulation assemblage de l'anneau Z
	FtsH1, HflB, MrsC, Std, TolZ	prouvé expérimentalement (e.g. BS)	metalloprotease ATP-dependante	Division
	FtsH2	par homologie	metalloprotease ATP-dependante?	Division?
	FtsH3	par homologie	metalloprotease ATP-dependante?	Division?
	SpoVK, SpoVJ	prouvé expérimentalement (e.g. BS)	ATPase	Sporulation
20 FtsJ				
	FtsJ, MrsF, RrmJ	prouvé expérimentalement (e.g. EC)	rRNA-methyltransferase	Division
21 FtsK				
	FtsK1(EC), SpoIIIE(BS)	prouvé expérimentalement (e.g. BS)	Translocase	Ségrégation du chromosome
	FtsK2(EC), SftA(BS)	prouvé expérimentalement (e.g. BS)	Translocase	Ségrégation du chromosome?
22 FtsW RodA SpoVE (SEDS)				
	FtsW	prouvé expérimentalement (e.g. BS)	Lipide II Flippase/Peptidoglycane polymérase	Synthèse du peptidoglycane
	RodA, MrdB	prouvé expérimentalement (e.g. BS)	Peptidoglycane polymérase	Synthèse du peptidoglycane
	RodA2	par homologie	Lipide II Flippase/Peptidoglycane polymérase ?	Synthèse du peptidoglycane ?
	SpoVE	prouvé expérimentalement (e.g. BS)	Lipide II Flippase/Peptidoglycane polymérase	Synthèse du peptidoglycane
23 FtsX				
	FtsX, ftsS	prouvé expérimentalement (e.g. BS)	sous unité ABC transporteur	Division
24 FtsY				
	FtsY	prouvé expérimentalement (e.g. BS)	RNA-methyltransferase	Division
25 GatD				
	GatD1	prouvé expérimentalement (e.g. MeThe, SA)	Glutamyl-tRNA(Gln) amidotransferase	Synthèse du peptidoglycane
	GatD2	par homologie	Glutamyl-tRNA(Gln) amidotransferase ?	Synthèse du peptidoglycane ?

	GatD3	par homologie	Glutamyl-tRNA(Gln) amidotransferase ?	Synthèse du peptidoglycane ?
	GatD4	par homologie	Glutamyl-tRNA(Gln) amidotransferase ?	Synthèse du peptidoglycane ?
26 GidA				
	GidA, MnmG, TrmF	prouvé expérimentalement (e.g. EC)	tRNA uridine 5-carboxymethylaminomethyl modification	Traduction/Division
27 GlmM				
	GlmM, MrsA, FemD(SA)	prouvé expérimentalement (e.g. BS)	phosphoglucosamine mutase	Synthèse du peptidoglycane
28 GlmS				
	GlmS	prouvé expérimentalement (e.g. BS)	glutamine-fructose6P aminotransferase	Synthèse du peptidoglycane
29 GlmU				
	GlmU	prouvé expérimentalement (e.g. EC)	UDP-N-acetylglucosamine pyrophosphorylase+Glucosamine -1-phosphate N-acetyltransferase	Synthèse du peptidoglycane
30 IleS ValS LeuS				
	IleS1	prouvé expérimentalement (e.g. EC), contexte génomique	Isoleucine-tRNA synthetase	Traduction
	IleS2	par homologie	Isoleucine-tRNA synthetase ?	Traduction ?
	LeuS	prouvé expérimentalement (e.g. EC), contexte génomique	Leucine-tRNA synthetase	Traduction
	ValS	prouvé expérimentalement (e.g. EC), contexte génomique	Valine-tRNA synthetase	Traduction
31 LysA				
	LysA1	prouvé expérimentalement (e.g. BS)	Diaminopimelate decarboxylase	Synthèse du peptidoglycane
	LysA2	par homologie	Diaminopimelate decarboxylase ?	Synthèse du peptidoglycane ?
	LysA3	par homologie	Diaminopimelate decarboxylase ?	Synthèse du peptidoglycane ?
32 LytB(SP)				
	LytB(SP)	prouvé expérimentalement (e.g. SP)	Glucosaminidase	Remodelage du peptidoglycane
33 MraW				
	MraW, RsmH	prouvé expérimentalement (e.g. EC), contexte génomique	rRNA methyltransferase H	Traduction
34 MraY WecA				
	MraY, MurX	prouvé expérimentalement (e.g. BS)	Phospho-N-acetylmuramoyl-pentapeptide-transferase	Synthèse du peptidoglycane
	WecA, Rfe	prouvé expérimentalement (e.g. PS)	N-acetylglucosamine-1-phosphate transferase	Synthèse de la capsule
35 Mtf				
	Mtf, Fmt	prouvé expérimentalement (e.g. BS), contexte génomique	methionyl-tRNA formyltransferase	Traduction
36 MurA MurZ				
	MurA1	prouvé expérimentalement (e.g. BA)	UDP-N-acetylglucosamine enolpyruvyle transferase	Synthèse du peptidoglycane
	MurA2	par homologie	UDP-N-acetylglucosamine enolpyruvyle transferase ?	Synthèse du peptidoglycane ?
	MurZ	prouvé expérimentalement (e.g. BA)	UDP-N-acetylglucosamine enolpyruvyle transferase	Synthèse du peptidoglycane
37 MurB				
	MurB1	prouvé expérimentalement (e.g. EC)	UDP-N-acetylenolpyruvoylglucosamine reductase	Synthèse du peptidoglycane
	MurB2	par homologie	UDP-N-acetylenolpyruvoylglucosamine reductase ?	Synthèse du peptidoglycane ?
38 MurC MurD MurE MurF MurT				
	MurC	prouvé expérimentalement (e.g. EC)	UDP-N-acetylmuramate--L-alanine ligase	Synthèse du peptidoglycane
	MurD	prouvé expérimentalement (e.g. EC)	UDP-N-acetylmuramoylalanine--D-glutamate ligase	Synthèse du peptidoglycane
	MurE	prouvé expérimentalement (e.g. EC)	UDP-N-acetylmuramoyl-L-alanyl-D-glutamate--2,6-diaminopimelate ligase	Synthèse du peptidoglycane
	MurF	prouvé expérimentalement (e.g. EC)	UDP-N-acetylmuramoyl-tripeptide--D-alanyl-D-alanine ligase	Synthèse du peptidoglycane
	MurT	prouvé expérimentalement (e.g. SA)	Undecaprenyl-PP-N-acetylmuramoyl-pentapeptide Glutamine aminase	Synthèse du peptidoglycane
39 MurG				
	MurG1	prouvé expérimentalement (e.g. EC)	UDP-N-acetylglucosamine--N-acetylmuramyl-(pentapeptide) pyrophosphoryl-undecaprenol N-acetylglucosamine transferase	Synthèse du peptidoglycane
	MurG2	par homologie	UDP-N-acetylglucosamine--N-acetylmuramyl-(pentapeptide) pyrophosphoryl-undecaprenol N-acetylglucosamine transferase ?	Synthèse du peptidoglycane ?
40 MurI				
	Asr	prouvé expérimentalement (e.g. EFm)	Aspartate racémase	Synthèse du peptidoglycane
	MurI	prouvé expérimentalement (e.g. LF)	glutamate racemase	Synthèse du peptidoglycane
41 MurJ YabM YkvU SpoVB				
	MurJ, MviN	prouvé expérimentalement (e.g. BS)	Lipide II Flippase	Synthèse du peptidoglycane
	YabM	par homologie	Polysaccharide synthase?	?
	YkvU	prouvé expérimentalement (e.g. BS)	Polysaccharide synthase?	Sporulation
	SpoVB, SpoIIIF	prouvé expérimentalement (e.g. BS)	Polysaccharide synthase?	Sporulation
42 MurK				
	MurK	prouvé expérimentalement (e.g. CA)	N-acetylmuramic acid/N-acetylglucosamine kinase	Recyclage du peptidoglycane
43 MurN MurM FemA FemB FemX fmh				
	FemA	prouvé expérimentalement (e.g. SA)	Undecaprenyl-PP-N-acetylmuramoyl-pentapeptide Glycine ligase	Synthèse du peptidoglycane
	FemB	prouvé expérimentalement (e.g. SA)	Undecaprenyl-PP-N-acetylmuramoyl-pentapeptide Glycine ligase	Synthèse du peptidoglycane
	FemX	prouvé expérimentalement (e.g. SA)	Undecaprenyl-PP-N-acetylmuramoyl-pentapeptide Glycine ligase	Synthèse du peptidoglycane
	Fmh	prouvé expérimentalement (e.g. SA)	Undecaprenyl-PP-N-acetylmuramoyl-pentapeptide Glycine ligase	Synthèse du peptidoglycane
	MurM	prouvé expérimentalement (e.g. SP)	Undecaprenyl-PP-N-acetylmuramoyl-pentapeptide Alanine ligase	Synthèse du peptidoglycane
	MurN	prouvé expérimentalement (e.g. SP)	Undecaprenyl-PP-N-acetylmuramoyl-pentapeptide Alanine ligase	Synthèse du peptidoglycane
44 MurQ				
	MurQ	prouvé expérimentalement (e.g. EC)	N-acetylmuramic acid 6-phosphate hydrolase	Recyclage du peptidoglycane
45 NagA				
	NagA	prouvé expérimentalement (e.g. BS)	N-acetylglucosamine-6-phosphate deacetylase	Recyclage du peptidoglycane
46 NagB				
	NagB, GlmD	prouvé expérimentalement (e.g. BS)	Glucosamine-6-phosphate deaminase	Synthèse du peptidoglycane
47 Noc ParB				
	Noc	prouvé expérimentalement (e.g. BS)		Positionnement de l'anneau Z
	ParB, Spo0J	prouvé expérimentalement (e.g. BS)		Ségrégation chromosomique

48 NudF				
	NudF	prouvé expérimentalement (e.g. EC), contexte génomique	ADP ribose diphosphatase	Détoxification
49 PBPs				
	A3, PBP1a(SP), PBP1(BS)	prouvé expérimentalement (e.g. SP,BS)	Transpeptidase/Glycosyltransferase	Réticulation du peptidoglycane
	A4, PBP2a(SP), PBP2c(BS)	prouvé expérimentalement (e.g. SP,BS)	Transpeptidase/Glycosyltransferase	Réticulation du peptidoglycane
	A5, PBP1b,(SP), PBP4(BS)	prouvé expérimentalement (e.g. SP,BS)	Transpeptidase/Glycosyltransferase	Réticulation du peptidoglycane
	A6, PBP1c(EC)	prouvé expérimentalement (e.g. EC)	Transpeptidase/Glycosyltransferase	Réticulation du peptidoglycane
	AX1	par homologie	Transpeptidase/Glycosyltransferase?	Réticulation du peptidoglycane ?
	B1, PBP3(BS)	prouvé expérimentalement (e.g. BS)	Transpeptidase	Réticulation du peptidoglycane
	B3, SpoVD(BS), PBP3(EC)	prouvé expérimentalement (e.g. BS)	Transpeptidase	Réticulation du peptidoglycane
	B4, PBP2x(SP), PBP2b(BS)	prouvé expérimentalement (e.g. SP,BS)	Transpeptidase	Réticulation du peptidoglycane
	B5, PBP2b(SP), PBP2a(BS), PBPH(BS)	prouvé expérimentalement (e.g. SP,BS)	Transpeptidase	Réticulation du peptidoglycane
	B6, PBP4b(BS)	prouvé expérimentalement (e.g. BS)	Transpeptidase	Réticulation du peptidoglycane
	BX1	par homologie	Transpeptidase?	Réticulation du peptidoglycane ?
	MGT1	prouvé expérimentalement (e.g. SA)	Glycosyltransferase	Réticulation du peptidoglycane
	MGT2	par homologie	Glycosyltransferase ?	Réticulation du peptidoglycane ?
50 Pfs				
	Pfs, MtnM	prouvé expérimentalement (e.g. EC), contexte génomique	5'-methylthioadenosine/S-adenosylhomocysteine nucleosidase	Methylation des ARN
51 PhpP				
	PhpP, PppL	prouvé expérimentalement (e.g. SP)	Sérine/Thréhonine phosphatase	Couplage division-élongation
52 PriA, Mfd, RecG				
	Mfd	prouvé expérimentalement (e.g. BS), par homologie, contexte génomique		Couplage transcription-réparation
	RecG	prouvé expérimentalement (e.g. BS)		Recombinaison
	PriA	prouvé expérimentalement (e.g. BS)	Hélicase	Réamorçage de la réplication
53 RacA				
	RacA	prouvé expérimentalement (e.g. BS)		Sporulation
54 Smc RecN RecF SbcC				
	Smc	prouvé expérimentalement (e.g. BS)		Ségrégation du chromosome
	RecN	prouvé expérimentalement (e.g. BS)		Recombinaison
	RecF	prouvé expérimentalement (e.g. BS)		Recombinaison
	SbcC1	prouvé expérimentalement (e.g. SA)	Nuclease	Recombinaison/Réplication
	SbcC2	par homologie	Nuclease ?	Recombinaison/Réplication ?
55 SpoIID LytB(BS)				
	SpoIID, SpoIIC	prouvé expérimentalement (e.g. BS)		Sporulation
	LytB(BS), CwbA	prouvé expérimentalement (e.g. BS)		Sporulation
56 SpoIIE				
	SpoIIE, SpoIIH	prouvé expérimentalement (e.g. BS)		Sporulation
57 SpoIIIJ				
	SpoIIIJ1, MisCA	prouvé expérimentalement (e.g. BS)	Insertase	Sporulation
	SpoIIIJ2, MisCB	prouvé expérimentalement (e.g. BS)	Insertase	Sporulation
58 StkP				
	StkP(SP), PknB(MT), Stk1(SA), PrkC(BS)	prouvé expérimentalement (e.g. SP)	Sérine/Thréhonine kinase	Couplage division-élongation
59 SunL				
	SunL	contexte génomique	rRNA methyltransferase ?	Traduction ?
60 TilS				
	TilS, MesJ	prouvé expérimentalement (e.g. EC), contexte génomique	tRNA(Ile)-lysidine synthase	Traduction
61 TolQ				
	TolQ	prouvé expérimentalement (e.g. EC)		Division
62 WalH Wall				
	WalH	prouvé expérimentalement (e.g. SA)		Régulation de la synthèse du peptidoglycane
	Wall	prouvé expérimentalement (e.g. SA)		Régulation de la synthèse du peptidoglycane
63 WalJ				
	WalJ	prouvé expérimentalement (e.g. SA)		Régulation de la synthèse du peptidoglycane
64 WalK				
	WalK, VicK	prouvé expérimentalement (e.g. SA)	Histidine kinase	Régulation de la synthèse du peptidoglycane
65 WalR				
	WalR	prouvé expérimentalement (e.g. SA)	Régulateur transcriptionnel	Régulation de la synthèse du peptidoglycane
66 XerC XerD XerS				
	XerC	prouvé expérimentalement (e.g. BS)	Recombinase	Recombinaison
	XerD, XprB, RipX	prouvé expérimentalement (e.g. BS)	Recombinase	Recombinaison
	XerS	prouvé expérimentalement (e.g. SP)	Recombinase	Recombinaison
Familles monogéniques		**Rôle biologique**		
	Protéine (orthologue)	Inférence du rôle dans le cycle cellulaire	**Activité biochimique**	**Processus cellulaire**
	AnmK	prouvé expérimentalement (e.g. EC)	Anhydro-N-acetylmuramic acid kinase	Recyclage du peptidoglycane
	CDP3(RocS)	prouvé expérimentalement (e.g. SP) NON PUBLIE		Couplage ségrégation-division-capsule
	CpsB, Cap1B, Wzb	prouvé expérimentalement (e.g. SP)	Tyrosine phosphatase	Régulation de la synthèse de la capsule
	CpsC, Wzd	prouvé expérimentalement (e.g. SP)		Régulation de la synthèse de la capsule
	DacC(BS), PBP4a(BS), DacB(EC), PBP4(EC)	prouvé expérimentalement (e.g. BS)	Endopeptidase	Remodelage du peptidoglycane
	DnaB(BS)	prouvé expérimentalement (e.g. BS)		Élongation de la réplication
	EzrA	prouvé expérimentalement (e.g. BS,SP)		Régulation assemblage de l'anneau Z
	FtsQ(EC), DivIB(BS)	prouvé expérimentalement (e.g. BS)		Division
	FtsZ	prouvé expérimentalement (e.g. BS)	GTPase	Échafaudage du divisome
	GidB, RsmG	prouvé expérimentalement (e.g. BS)	rNA-methyltransferase	Traduction/Division?
	Jag, EloR	prouvé expérimentalement (e.g. SP)		Traduction
	MacP	prouvé expérimentalement (e.g. SP)		Régulation de l'activité des PBPs
	Maf	prouvé expérimentalement (e.g. BS)	Nucleotide pyrophosphatase	Division

	MapZ, LocZ	prouvé expérimentalement (e.g. SP)		Positionnement de l'anneau Z
	MinC	prouvé expérimentalement (e.g. BS)		Positionnement de l'anneau Z
	MinE	prouvé expérimentalement (e.g. EC)		Positionnement de l'anneau Z
	MinJ	prouvé expérimentalement (e.g. BS)		Positionnement de l'anneau Z
	MraZ	prouvé expérimentalement (e.g. EC), contexte génomique	Régulateur transcriptionnel	?
	MreC	prouvé expérimentalement (e.g. BS)		Élongation
	MreD	prouvé expérimentalement (e.g. BS)		Élongation
	PCDP10	prouvé expérimentalement (e.g. SP) NON PUBLIE, contexte génomique		?
	PCDP4	contexte génomique		?
	PCDP7	contexte génomique		?
	PCDP8	contexte génomique		?
	RadC	contexte génomique		Recombinaison ?
	RecA, LexB, RecH, RnmB	prouvé expérimentalement (e.g. BS)	ATPase	Recombinaison
	RecR, RecM	prouvé expérimentalement (e.g. EC)		Recombinaison
	RecU, PrfA	prouvé expérimentalement (e.g. BS)	Résolvase de jonction holliday	Recombinaison
	RodZ	prouvé expérimentalement (e.g. BS)		Élongation
	Rtp	prouvé expérimentalement (e.g. BS)		Terminaison de la réplication
	ScpA	prouvé expérimentalement (e.g. BS)		Ségrégation du chromosome
	ScpB	prouvé expérimentalement (e.g. BS)		Ségrégation du chromosome
	SepF	prouvé expérimentalement (e.g. BS)		Ancrage à la membrane de l'anneau Z
	Spo0M	prouvé expérimentalement (e.g. BS)		Sporulation
	SpoIIGA	prouvé expérimentalement (e.g. BS)	Peptidase	Sporulation
	SpoIIID	prouvé expérimentalement (e.g. BS)		Sporulation
	SpoVG	prouvé expérimentalement (e.g. BS)		Sporulation
	TolR	prouvé expérimentalement (e.g. EC)		Division
	ZapA	prouvé expérimentalement (e.g. BS)		Régulation assemblage de l'anneau Z

Familles absentes			**Rôle biologique**	
	Protéine	Inférence du rôle dans le cycle cellulaire	**Activité biochimique**	**Processus cellulaire**
	AmgK	prouvé expérimentalement (e.g. EC)	N-acetylmuramate/N-acetylglucosamine kinase	Recyclage du peptidoglycane
	AmpG	prouvé expérimentalement (e.g. EC)	Transporteur	Recyclage du peptidoglycane
	FtsN	prouvé expérimentalement (e.g. EC)		Division
	LpoB	prouvé expérimentalement (e.g. EC)		Régulation de l'activité des PBPs
	MatP	prouvé expérimentalement (e.g. EC)		Compaction du chromosome
	MipZ	prouvé expérimentalement (e.g. CC)	ATPase	Positionnement de l'anneau Z
	MpaA	prouvé expérimentalement (e.g. EC)		Recyclage du peptidoglycane
	Mpl	prouvé expérimentalement (e.g. EC)	UDP-N-acetylmuramate--L-alanyl-gamma-D-glutamyl-meso-2,6-diaminoheptandioate ligase	Recyclage du peptidoglycane
	MukB	prouvé expérimentalement (e.g. EC)		Ségrégation du chromosome
	MukE	prouvé expérimentalement (e.g. EC)		Ségrégation du chromosome
	MukF	prouvé expérimentalement (e.g. EC)		Ségrégation du chromosome
	MurU	prouvé expérimentalement (e.g. PP)	N-acetylmuramate alpha-1-phosphate uridylyltransferase	Recyclage du peptidoglycane
	Pal	prouvé expérimentalement (e.g. EC)		Division
	PomZ	prouvé expérimentalement (e.g. MX)	ATPase	Positionnement de l'anneau Z
	SeqA	prouvé expérimentalement (e.g. EC)		Initiation de la réplication
	SlmA	prouvé expérimentalement (e.g. EC)		Positionnement de l'anneau Z
	SsgB	prouvé expérimentalement (e.g. SC)		Positionnement de l'anneau Z
	SulA	prouvé expérimentalement (e.g. EC)		Régulation assemblage de l'anneau Z
	TolA	prouvé expérimentalement (e.g. EC)		Division
	Tus	prouvé expérimentalement (e.g. EC)		Terminaison de la réplication
	ZapB	prouvé expérimentalement (e.g. EC)		Régulation assemblage de l'anneau Z
	ZapC	prouvé expérimentalement (e.g. EC)		Régulation assemblage de l'anneau Z
	ZapD	prouvé expérimentalement (e.g. EC)		Régulation assemblage de l'anneau Z
	ZipA	prouvé expérimentalement (e.g. EC)		Ancrage à la membrane de l'anneau Z

Familles non traitées			**Rôle biologique**	
	Protéine			
	Acd			Remodelage du peptidoglycane
	AcmA			Remodelage du peptidoglycane
	AcmB			Remodelage du peptidoglycane
	AcmC			Remodelage du peptidoglycane
	AcmD			Remodelage du peptidoglycane
	AepA			Remodelage du peptidoglycane
	AltA			Remodelage du peptidoglycane
	AmiA			Remodelage du peptidoglycane
	AmiB			Remodelage du peptidoglycane
	AmiC			Remodelage du peptidoglycane
	AmiD			Remodelage du peptidoglycane
	AmpH			Remodelage du peptidoglycane
	AppC			Recyclage du peptidoglycane
	BacA			Recyclage du peptidoglycane
	BofA			Sporulation
	chargeur de pinces			Elongation de la réplication
	CpsA			Synthèse de la capsule
	CpsE			Synthèse de la capsule
	CpsF			Synthèse de la capsule
	CpsG			Synthèse de la capsule
	CpsH			Synthèse de la capsule

	CpsI			Synthèse de la capsule
	CpsJ			Synthèse de la capsule
	CpsK			Synthèse de la capsule
	CpsL			Synthèse de la capsule
	CpsM			Synthèse de la capsule
	CpsN			Synthèse de la capsule
	CpsO			Synthèse de la capsule
	CpsP			Synthèse de la capsule
	CwlC			Sporulation
	CwlP			Remodelage du peptidoglycane
	DnaQ			Elongation de la réplication
	GerE			Sporulation
	GerM			Sporulation
	LdcB			Remodelage du peptidoglycane
	LytC			Remodelage du peptidoglycane
	LytD			Remodelage du peptidoglycane
	LytH			Remodelage du peptidoglycane
	MapA			Remodelage du peptidoglycane
	MdcA			Remodelage du peptidoglycane
	MltA			Remodelage du peptidoglycane
	MltB			Remodelage du peptidoglycane
	MltD			Remodelage du peptidoglycane
	MltE			Remodelage du peptidoglycane
	MltF			Remodelage du peptidoglycane
	MltC			Remodelage du peptidoglycane
	MupP			Recyclage du peptidoglycane
	MurP			Recyclage du peptidoglycane
	MurR			Recyclage du peptidoglycane
	MutT			Réplication
	NagE			Recyclage du peptidoglycane
	NagK			Recyclage du peptidoglycane
	NagZ			Remodelage du peptidoglycane
	OppD			Recyclage du peptidoglycane
	PbpE			Remodelage du peptidoglycane
	PcsB			Remodelage du peptidoglycane
	PepD			Recyclage du peptidoglycane
	PgdA			Remodelage du peptidoglycane
	PMP23			Remodelage du peptidoglycane
	polymerase III			Elongation de la réplication
	RecB			Recombinaison
	RecC			Recombinaison
	RecD			Recombinaison
	RecE			Recombinaison
	RecJ			Recombinaison
	RecO			Recombinaison
	RecQ			Recombinaison
	RecR			Recombinaison
	RecT			Recombinaison
	Rnase H			Elongation de la réplication
	RsfA			Sporulation
	RuvA			Recombinaison
	RuvB			Recombinaison
	RuvC			Recombinaison
	SigE			Sporulation
	SigF			Sporulation
	SigG			Sporulation
	SigH			Sporulation
	SigK			Sporulation
	Slt70			Remodelage du peptidoglycane
	Spo0A			Sporulation
	Spo0B			Sporulation
	Spo0E			Sporulation
	Spo0F			Sporulation
	SpoIIAA			Sporulation
	SpoIIAB			Sporulation
	SpoIID			Sporulation
	SpoIIE			Sporulation
	SpoIIGA			Sporulation
	SpoIIIA			Sporulation
	SpoIIIAH			Sporulation
	SpoIIID			Sporulation
	SpoIIIE			Sporulation
	SpoIIM			Sporulation
	SpoIIP			Sporulation
	SpoIIR			Sporulation
	SpoIISA			Sporulation
	SpoIISB			Sporulation
	SpoIVA			Sporulation
	SpoIVB			Sporulation
	SpoIVFA			Sporulation
	SpoIVR			Sporulation
	SpoVT			Sporulation
	VanX			Remodelage du peptidoglycane
	VanY			Remodelage du peptidoglycane
	YcjG			Recyclage du peptidoglycane

.8 Protéomes de référence utilisés dans la recherche par profil HMM.

Identifiant taxonomique	Numéro d'assemblage	Nom de la souche
441770	GCA_000017025.1	Clostridium botulinum A str. ATCC 19397
572479	GCA_000165465.1	Halanaerobium praevalens DSM 2228
457570	GCA_000020005.1	Natranaerobius thermophilus JW/NM-WN-LF
546271	GCA_000208405.1	Selenomonas sputigena ATCC 35185
543734	GCA_000026485.1	Lactobacillus casei BL23
171101	GCA_000007045.1	Streptococcus pneumoniae R6
1280	GCA_001278745.1	Staphylococcus aureus
83333	GCA_000800765.1	Escherichia coli K-12
1279007	GCA_000496605.2	Pseudomonas aeruginosa PA1
224308	GCA_000009045.1	Bacillus subtilis subsp. subtilis str. 168

.9 Récapitulatif des phylogénies des familles de gènes.

Les phylogénies construites avec des alignements nettoyés de moins de 100 positions sont représentées en rouge. Les familles multigéniques sont présentées dans des cases grisées tandis que familles/sous-familles monogéniques sont dans des cases blanches.

Familles multigéniques		Phylogénie			
Famille	Sous-familles (orthologues)	Nombre de séquences	Nombre de positions d'acide aminés	Modèle évolutif	Réconciliation qualitative
1 AlaS					
	AlaS, LovB	305	786	LG+F+I+G4	vertical
2 Alr PCDP6					
	Alr, DadX(EC)	362	247	LG+I+G4	vertical
	Alr-like	155	144	LG+G4	multiples transferts/ paralogies cachées
	PCDP6	260	189	LG+I+G4	vertical
3 Aslfm					
	Aslfm	59	365	LG+F+G4	multiples transferts/ paralogies cachées
4 CDP1					
	CDP1	123	199	LG+I+G4	vertical
5 CDP4					
	CDP4	157	63	LG+F+G4	vertical
6 CozE					
	CozE1	90	333	LG+F+G4	vertical
	CozE2	92	294	LG+F+I+G4	Vertical + quelques transferts (Clostridia)
	CozE3	152	316	LG+F+I+G4	Vertical + quelques transferts (Enterococcaceae)
	CozE4	154	230	LG+F+I+G4	vertical
7 CpsD ParA MinD					
	CpsD	278	162	LG+G4	Vertical + quelques transferts (Streptococcaceae)
	MinD	206	245	LG+I+G4	vertical
	ParA, Soj	240	240	LG+I+G4	vertical
8 DacA DacB DacF					
	DacA(BS), PBP5(BS), DacC(EC), PBP6(EC)	275	214	LG+F+I+G4	vertical
	DacB(BS), PBP5*(BS), DacD(EC), PBP6b(EC)	157	242	LG+I+G4	vertical
	DacF(BS), DacA(EC), PBP5(EC)	168	298	LG+I+G4	vertical
9 Ddl					
	Ddl	263	268	LG+I+G4	Vertical + quelques transferts (Lactobacillaceae,Leuconostocaceae)
10 DivIC FtsL					
	DivIC(BS), FtsB(EC)	314	31	LG+F+G4	vertical
	FtsL, MraR	304	25		multiples transferts/ paralogies cachées
11 DivIVA GpsB					
	DivIVA	284	141	LG+G4	Vertical + quelques transferts (Lactobacillaceae)
	GpsB	172	81	LG+I+G4	Vertical + quelques transferts (Paenibacillaceae)
12 DnaA DnaI					
	DnaA	312	342	LG+I+G4	vertical
	DnaI	183	240	LG+F+I+G4	vertical
13 DnaB					
	DnaB1(EC), DnaC(BS)	304	387	LG+F+I+G4	vertical
	DnaB2	29	379	LG+I+G4	multiples transferts/ paralogies cachées
14 DnaD					
	DnaD	293	92	LG+G4	Vertical + quelques transferts (Streptococcaceae/Enterococcaceae)
15 DnaG					
	DnaG	305	346	LG+I+G4	vertical
16 DnaN					
	DnaN	322	273	LG+F+I+G4	vertical
17 FtsA Mbl MreB MreBH					
	FtsA	244	335	LG+I+G4	vertical
	Mbl	229	311	LG+I+G4	vertical
	MreB	231	315	LG+I+G4	vertical
	MreBH	42	330	LG+G4	Vertical + quelques transferts (Clostridia)
18 FtsE					
	FtsE	243	226	LG+I+G4	vertical
19 FtsH ClpX SpoVK					

	ClpX, LocP	316	363	LG+I+G4	vertical
	FtsH1, HflB, MrsC, Std, TolZ	312	515	LG+I+G4	vertical
	FtsH2	57	553	LG+I+G4	vertical
	FtsH3	108	413	LG+I+G4	vertical
	SpoVK, SpoVJ	179	215	LG+I+G4	Vertical +nombreux transferts basaux
20 FtsJ					
	FtsJ, MrsF, RrmJ	284	231	LG+I+G4	vertical
21 FtsK					
	FtsK1(EC), SpoIIIE(BS)	321	487	LG+I+G4	vertical
	FtsK2(EC), SftA(BS)	90	477	LG+I+G4	vertical
22 FtsW RodA SpoVE (SEDS)					
	FtsW	298	258	LG+F+I+G4	vertical
	RodA, MrdB	371	278	LG+F+I+G4	vertical
	RodA2	42	295	LG+F+G4	multiples transferts/ paralogies cachées
	SpoVE	189	321	LG+F+I+G4	vertical
23 FtsX					
	FtsX, ftsS	253	219	LG+F+I+G4	vertical
24 FtsY					
	FtsY	304	284	LG+I+G4	vertical
25 GatD					
	GatD1	138	277	LG+I+G4	vertical
	GatD2	34	316	LG+G4	multiples transferts/ paralogies cachées
	GatD3	28	305	LG+I+G4	vertical
	GatD4	62	263	LG+I+G4	multiples transferts/ paralogies cachées
26 GidA					
	GidA, MnmG, TrmF	293	602	LG+I+G4	vertical
27 GlmM					
	GlmM, MrsA, FemD(SA)	303	416	LG+I+G4	vertical
28 GlmS					
	GlmS	307	536	LG+I+G4	vertical
29 GlmU					
	GlmU	300	425	LG+I+G4	vertical
30 IleS ValS LeuS					
	IleS1	232	856	LG+F+I+G4	vertical
	IleS2	91	966	LG+F+I+G4	vertical
	LeuS	304	744	LG+I+G4	vertical
	ValS	305	808	LG+I+G4	vertical
31 LysA					
	LysA1	227	386	LG+I+G4	Vertical + quelques transferts (Staphylococcaceae,Enterococcaceae)
	LysA2	62	410	LG+I+G4	multiples transferts/ paralogies cachées
	LysA3	26	318	LG+I+G4	multiples transferts/ paralogies cachées
32 LytB(SP)					
	LytB(SP)	5	604	JTT+F+I	vertical
33 MraW					
	MraW, RsmH	319	278	LG+F+I+G4	Vertical +nombreux transferts basaux
34 MraY WecA					
	MraY, MurX	310	253	LG+F+I+G4	vertical
	WecA, Rfe	284	296	LG+F+I+G4	Vertical + quelques transferts (Paenibacillaceae)
35 Mtf					
	Mtf, Fmt	305	282	LG+I+G4	vertical
36 MurA MurZ					
	MurA1	265	383	LG+I+G4	vertical
	MurA2	97	390	LG+I+G4	vertical
	MurZ	251	382	LG+I+G4	vertical
37 MurB					
	MurB1	248	258	LG+I+G4	vertical
	MurB2	72	284	LG+I+G4	vertical
38 MurC MurD MurE MurF MurT					
	MurC	304	346	LG+I+G4	Vertical + quelques transferts (Paenibacillaceae)
	MurD	305	343	LG+I+G4	vertical

	MurE	360	326	LG+I+G4	multiples transferts/ paralogies cachées
	MurF	335	295	LG+G4	Vertical +nombreux transferts basaux
	MurT	186	343	LG+I+G4	Vertical + quelques transferts (Staphylococcaceae,Lactobacillales)
39 MurG					
	MurG1	200	312	LG+I+G4	vertical
	MurG2	123	318	LG+I+G4	vertical
40 MurI					
	Asr	159	155	LG+I+G4	multiples transferts/ paralogies cachées
	MurI	329	184	LG+I+G4	Vertical +nombreux transferts basaux
41 MurJ YabM YkvU SpoVB					
	MurJ, MviN	201	426	LG+F+I+G4	vertical
	YabM	289	363	LG+F+G4	vertical
	YkvU	18	438	LG+F+G4	vertical
	SpoVB, SpoIIIF	158	422	LG+F+I+G4	vertical
42 MurK					
	MurK	99	153	LG+I+G4	multiples transferts/ paralogies cachées
43 MurN MurM FemA FemB FemX fmh					
	FemA	19	418	LG+F+I+G4	vertical
	FemB	19	419	LG+G4	vertical
	FemX	67	298	LG+I+G4	multiples transferts/ paralogies cachées
	Fmh	26	406	LG+F+G4	vertical
	MurM	39	396	LG+I+G4	vertical
	MurN	62	351	LG+I+G4	Vertical + quelques transferts (Clostridia)
44 MurQ					
	MurQ	157	283	LG+I+G4	multiples transferts/ paralogies cachées
45 NagA					
	NagA	277	268	LG+I+G4	multiples transferts/ paralogies cachées
46 NagB					
	NagB, GlmD	244	214	LG+I+G4	multiples transferts/ paralogies cachées
47 Noc ParB					
	Noc	191	228	LG+I+G4	Vertical + quelques transferts (Staphylococcaceae)
	ParB, Spo0J	304	220	LG+I+G4	vertical
48 NudF					
	NudF	297	157	LG+I+G4	Vertical + quelques transferts (Staphylococcaceae)
49 PBPs					
	A3, PBP1a(SP), PBP1(BS)	176	518	LG+F+I+G4	vertical
	A4, PBP2a(SP), PBP2c(BS)	203	478	LG+F+I+G4	multiples transferts/ paralogies cachées
	A5, PBP1b,(SP), PBP4(BS)	38	526	LG+F+I+G4	multiples transferts/ paralogies cachées
	A6, PBP1c(EC)	114	525	LG+I+G4	Vertical + quelques transferts (Brevibacillaceae)
	AX1	213	448	LG+F+I+G4	vertical
	B1, PBP3(BS)	88	543	LG+F+I+G4	vertical
	B3, SpoVD(BS), PBP3(EC)	65	635	LG+I+G4	vertical
	B4, PBP2x(SP), PBP2b(BS)	357	445	LG+F+I+G4	vertical
	B5, PBP2b(SP), PBP2a(BS), PBPH(BS)	372	327	LG+F+I+G4	vertical
	B6, PBP4b(BS)	106	335	LG+I+G4	vertical
	BX1	85	361	LG+I+G4	vertical
	MGT1	18	261	LG+G4	vertical
	MGT2	8	258	LG+G4	vertical
50 Pfs					
	Pfs, MtnM	252	192	LG+I+G4	multiples transferts/ paralogies cachées
51 PhpP					
	PhpP, PppL	304	152	LG+I+G4	vertical
52 PriA, Mfd, RecG					
	Mfd	304	879	LG+I+G4	vertical
	RecG	303	536	LG+I+G4	vertical

	PriA	305	563	LG+I+G4	vertical
53 RacA					
	RacA	40	117	LG+G4	vertical
54 Smc RecN RecF SbcC					
	Smc	298	643	LG+F+I+G4	vertical
	RecN	304	442	LG+F+I+G4	Vertical + quelques transferts (Staphylococcaceae, Lactobacillaceae, Streptococcaceae)
	RecF	304	289	LG+I+G4	vertical
	SbcC1	126	340	LG+F+I+G4	multiples transferts/ paralogies cachées
	SbcC2	75	662	LG+F+I+G4	vertical
55 SpoIID LytB(BS)					
	SpoIID, SpoIIC	167	209	LG+I+G4	vertical
	LytB(BS), CwbA	126	189	LG+I+G4	multiples transferts/ paralogies cachées
56 SpoIIE					
	SpoIIE, SpoIIH	165	473	LG+F+I+G4	vertical
57 SpoIIIJ					
	SpoIIIJ1, MisCA	292	157	LG+F+I+G4	vertical
	SpoIIIJ2, MisCB	172	200	LG+I+G4	vertical
58 StkP					
	StkP(SP), PknB(MT), Stk1(SA), PrkC(BS)	307	325	LG+I+G4	vertical
59 SunL					
	SunL	302	330	LG+F+I+G4	vertical
60 TilS					
	TilS, MesJ	300	194	LG+I+G4	Vertical + quelques transferts (Paenibacillaceae)
61 TolQ					
	TolQ	29	156	LG+I+G4	vertical
62 WalH WalI					
	WalH	184	78	LG+G4	vertical
	WalI	184	68	LG+I+G4	vertical
63 WalJ					
	WalJ	278	234	LG+I+G4	vertical
64 WalK					
	WalK, VicK	216	454	LG+F+I+G4	vertical
65 WalR					
	WalR	195	216	LG+I+G4	vertical
66 XerC XerD XerS					
	XerC	165	255	LG+I+G4	vertical
	XerD, XprB, RipX	246	258	LG+I+G4	Vertical + quelques transferts (Paenibacillaceae)
	XerS	117	272	LG+F+I+G4	vertical
Familles monogéniques		**Phylogénie**			
	Protéine (orthologue)	**Nombre de séquences**	**Nombre de positions d'acide aminés**		
	AnmK	28	364	LG+G4	multiples transferts/ paralogies cachées
	CDP3(RocS)	75	99	LG+G4	Vertical +nombreux transferts basaux
	CpsB, Cap1B, Wzb	227	148	LG+G4	Vertical + quelques transferts (Streptococcaceae)
	CpsC, Wzd	338	143	LG+F+G4	Vertical + quelques transferts (Streptococcaceae, Lactobacillaceae)
	DacC(BS), PBP4a(BS), DacB(EC), PBP4(EC)	24	451	LG+I+G4	vertical
	DnaB(BS)	181	193	LG+I+G4	vertical
	EzrA	164	406	LG+F+G4	vertical
	FtsQ(EC), DivIB(BS)	304	72	LG+F+G4	Vertical + quelques transferts (Staphylococcaceae, Listeriaceae)
	FtsZ	316	292	LG+G4	vertical
	GidB, RsmG	305	190	LG+I+G4	vertical
	Jag, EloR	255	158	LG+I+G4	vertical
	MacP	134	25	LG+G4	vertical
	Maf	205	150	LG+I+G4	Vertical + quelques transferts (Paenibacillaceae)
	MapZ, LocZ	39	349	LG+F+I+G4	vertical
	MinC	209	130	LG+I+G4	vertical
	MinE	106	64	LG+G4	vertical
	MinJ	154	257	LG+F+G4	vertical

	MraZ	239	135	LG+I+G4	Vertical + quelques transferts (Leuconostocaceae)
	MreC	296	159	LG+F+I+G4	vertical
	MreD	290	76	LG+F+G4	Vertical + quelques transferts (Lactobacillaceae)
	PCDP10	183	140	LG+I+G4	vertical
	PCDP4	84	131	VT+F+G4	multiples transferts/ paralogies cachées
	PCDP7	273	60	LG+F+G4	Vertical + quelques transferts (Listeriaceae)
	PCDP8	287	161	LG+I+G4	vertical
	RadC	353	174	LG+I+G4	multiples transferts/ paralogies cachées
	RecA, LexB, RecH, RnmB	318	311	LG+I+G4	vertical
	RecR, RecM	304	193	LG+I+G4	vertical
	RecU, PrfA	196	164	LG+G4	Vertical + quelques transferts (Clostridia)
	RodZ	243	102	LG+F+I+G4	Vertical + quelques transferts (Lactobacillaceae)
	Rtp	30	112	LG+G4	vertical
	ScpA	297	153	LG+I+G4	Vertical + quelques transferts (Lactobacillaceae, Staphylococcaceae)
	ScpB	297	138	LG+I+G4	vertical
	SepF	306	82	LG+F+I+G4	vertical
	Spo0M	80	111	LG+G4	Vertical + quelques transferts (Clostridia)
	SpoIIGA	162	152	LG+F+I+G4	vertical
	SpoIIID	168	77	LG+G4	vertical
	SpoVG	236	76	LG+G4	Vertical + quelques transferts (Paenibacillaceae)
	TolR	30	111	LG+I+G4	vertical
	ZapA	275	57	LG+G4	vertical

.10 Nombre de gènes du cycle cellulaire analysés pour chaque souche.

Souche	Nombre de gènes du cycle cellulaire présents
Bacillus weihenstephanensis WSBC10204	146/179
Bacillus infantis NRRL B-14911	146/179
Bacillus mycoides ATCC 6462	145/179
Bacillus thuringiensis str. Al Hakam	145/179
Bacillus bombysepticus str. Wang	145/179
Bacillus amyloliquefaciens Y2	144/179
Bacillus megaterium WSH-002	144/179
Bacillus cereus biovar anthracis str. CI	144/179
Bacillus anthracis str. Vollum	144/179
Bacillus pumilus W3	143/179
Bacillus atrophaeus NRS 1221A	143/179
Bacillus subtilis subsp. subtilis str. 168	143/179
Bacillus velezensis YJ11-1-4	143/179
Bacillus sp. YP1	143/179
Bacillus toyonensis BCT-7112	141/179
Bacillus paralicheniformis BL-09	141/179
Bacillus licheniformis DSM 13 = ATCC 14580	141/179
Bacillus endophyticus Hbe603	139/179
Halobacillus halophilus DSM 2266	138/179
Bacillus clausii KSM-K16	137/179
Oceanobacillus iheyensis HTE831	137/179
Brevibacillus brevis NBRC 100599	136/179
Bacillus lehensis G1	135/179
Terribacillus aidingensis MP602	135/179
Virgibacillus sp. SK37	135/179
Bacillus cellulosilyticus DSM 2522	134/179
Lysinibacillus fusiformis RB-21	134/179
Bacillus methanolicus MGA3	134/179
Bacillus pseudofirmus OF4	134/179
Paenibacillus sp. Y412MC10	134/179
Lysinibacillus varians GY32	134/179
Bacillus cytotoxicus NVH 391-98	134/179
Paenibacillus borealis DSM 13188	134/179
Lysinibacillus sphaericus C3-41	134/179
Paenibacillus sabinae T27	133/179
Paenibacillus peoriae HS311	133/179
Jeotgalibacillus malaysiensis	133/179
Bacillus coagulans S-lac	133/179
Geobacillus thermodenitrificans NG80-2	133/179
Brevibacillus laterosporus LMG 15441	133/179
Paenibacillus beijingensis DSM 24997	132/179
Geobacillus thermoleovorans CCB_US3_UF5	132/179
Geobacillus kaustophilus HTA426	132/179
Solibacillus silvestris StLB046	132/179
Paenibacillus terrae HPL-003	132/179
Bacillus halodurans C-125	131/179
Paenibacillus graminis DSM 15220	131/179
Paenibacillus stellifer DSM 14472	131/179
Paenibacillus durus DSM 1735	131/179
Paenibacillus riograndensis SBR5	131/179
Geobacillus thermoglucosidasius DSM 2542	131/179
Paenibacillus odorifer DSM 15391	131/179
Paenibacillus polymyxa Sb3-1	131/179
Paenibacillus mucilaginosus KNP414	131/179
Geobacillus sp. Y4.1MC1	131/179
Geobacillus stearothermophilus 10	130/179

Bacillus smithii DSM 4216	129/179
Amphibacillus xylanus NBRC 15112	128/179
Paenibacillus larvae subsp. larvae DSM 25430	127/179
Anoxybacillus gonensis G2	126/179
Anoxybacillus flavithermus WK1	126/179
Planococcus sp. PAMC 21323	125/179
Paenibacillus naphthalenovorans 32O-Y	125/179
Thermobacillus composti KWC4	124/179
Planococcus rifietoensis M8	124/179
[Bacillus] selenitireducens MLS10	123/179
Planococcus kocurii ATCC 43650	121/179
Carnobacterium maltaromaticum LMA28	121/179
Caldanaerobacter subterraneus subsp. tengcongensis MB4	121/179
Carnobacterium inhibens subsp. gilichinskyi	120/179
Lactobacillus paracasei subsp. paracasei JCM 8130	120/179
Lactobacillus rhamnosus Lc 705	120/179
Carnobacterium sp. CP1	119/179
Alkaliphilus metalliredigens QYMF	119/179
Enterococcus silesiacus LMG 23085	118/179
Lactobacillus casei subsp. casei ATCC 393	118/179
Lactobacillus brevis KB290	117/179
Exiguobacterium sibiricum 255-15	117/179
Lactobacillus koreensis 26-25	117/179
Enterococcus hirae ATCC 9790	117/179
Exiguobacterium antarcticum B7	117/179
Thermoanaerobacter wiegelii Rt8.B1	117/179
Lactobacillus plantarum subsp. plantarum ST-III	116/179
Enterococcus faecium UW7606x64/3 TC1	116/179
Thermoanaerobacter mathranii subsp. mathranii str. A3	116/179
Exiguobacterium sp. MH3	116/179
Thermosediminibacter oceani DSM 16646	116/179
Enterococcus durans KLDS 6.0933	116/179
Enterococcus casseliflavus EC20	116/179
Enterococcus mundtii QU 25	116/179
Lactobacillus salivarius str. Ren	116/179
Thermoanaerobacterium thermosaccharolyticum M0795	116/179
Paenibacillus bovis BD3526	115/179
Listeria welshimeri serovar 6b str. SLCC5334	115/179
Clostridium aceticum DSM 1496	115/179
Lactobacillus paraplantarum L-ZS9	115/179
Listeria seeligeri serovar 1/2b str. SLCC3954	115/179
Thermoanaerobacter kivui LKT-1	115/179
Clostridium acetobutylicum EA 2018	115/179
Thermoanaerobacterium xylanolyticum LX-11	115/179
Listeria ivanovii subsp. londoniensis	115/179
Pelosinus fermentans JBW45	115/179
Listeria monocytogenes serotype 4b str. LL195	115/179
Staphylococcus haemolyticus Sh29/312/L2	114/179
Lactobacillus sakei subsp. sakei 23K	114/179
Thermoanaerobacterium saccharolyticum JW/SL-YS485	114/179
Clostridium butyricum KNU-L09	114/179
Thermoanaerobacter sp. X514	113/179
Staphylococcus argenteus MSHR1132	113/179
Enterococcus faecalis str. Symbioflor 1	113/179
Staphylococcus agnetis 908	113/179
Tetragenococcus halophilus NBRC 12172	113/179
Macrococcus caseolyticus JCSC5402	113/179
Lactobacillus ruminis ATCC 27782	113/179
Staphylococcus hyicus ATCC 11249	112/179

Staphylococcus lugdunensis N920143	112/179
Staphylococcus carnosus subsp. carnosus TM300	112/179
Clostridium beijerinckii NCIMB 8052	112/179
Thermoanaerobacter italicus Ab9	112/179
Clostridium sporogenes NCIMB 10696	112/179
Alkaliphilus oremlandii OhILAs	112/179
Desulfosporosinus orientis DSM 765	112/179
Clostridium saccharoperbutylacetonicum N1-4(HMT)	112/179
Thermoanaerobacter pseudethanolicus ATCC 33223	112/179
Pelosinus sp. UFO1	112/179
Thermoanaerobacter brockii subsp. finnii Ako-1	112/179
Clostridium baratii str. Sullivan	111/179
Clostridium scatologenes ATCC 25775	111/179
Clostridium carboxidivorans P7	111/179
Lactobacillus buchneri NRRL B-30929	111/179
Staphylococcus saprophyticus subsp. saprophyticus ATCC 15305	111/179
[Clostridium] acidurici 9a	111/179
Clostridium pasteurianum NRRL B-598	111/179
Clostridium cellulovorans 743B	111/179
Staphylococcus aureus subsp. aureus NCTC 8325	111/179
Lactobacillus heilongjiangensis DSM 28069	110/179
Lactobacillus ginsenosidimutans EMML 3041	110/179
Salinicoccus halodurans H3B36	110/179
Desulfosporosinus meridiei DSM 13257	110/179
Desulfosporosinus acidiphilus SJ4	110/179
Staphylococcus schleiferi 5909-02	110/179
Clostridium saccharobutylicum DSM 13864	110/179
Tepidanaerobacter acetatoxydans Re1	110/179
Clostridium perfringens str. 13	110/179
Streptococcus oligofermentans AS 1.3089	110/179
Lactococcus piscium MKFS47	110/179
Desulfitobacterium hafniense Y51	110/179
Clostridium autoethanogenum DSM 10061	110/179
Desulfitobacterium metallireducens DSM 15288	110/179
Clostridium ljungdahlii DSM 13528	110/179
Staphylococcus pasteuri SP1	110/179
Streptococcus gordonii str. Challis substr. CH1	109/179
Staphylococcus pseudintermedius HKU10-03	109/179
Intestinimonas butyriciproducens AF211	109/179
Streptococcus sanguinis SK36	109/179
Clostridium botulinum NCTC 8550	109/179
Alicyclobacillus acidocaldarius subsp. acidocaldarius Tc-4-1	109/179
Clostridium novyi NT	109/179
Symbiobacterium thermophilum IAM 14863	109/179
Desulfitobacterium dehalogenans ATCC 51507	109/179
Lactococcus lactis subsp. lactis NCDO 2118	109/179
Staphylococcus equorum KS1039	109/179
[Clostridium] cellulolyticum H10	109/179
Streptococcus intermedius KCOM 1545	109/179
Streptococcus sp. VT 162	109/179
Streptococcus parasanguinis FW213	109/179
Staphylococcus warneri SG1	109/179
Melissococcus plutonius DAT561	109/179
[Clostridium] clariflavum DSM 19732	109/179
Streptococcus mitis KCOM 1350 (= ChDC B183)	109/179
Streptococcus salivarius NCTC 8618	108/179
Streptococcus thermophilus SMQ-301	108/179
Clostridium bornimense M2/40	108/179
Jeotgalicoccus sp. 13MG44_air	108/179

Staphylococcus capitis AYP1020	108/179
Ruminiclostridium thermocellum DSM 1313	108/179
Streptococcus oralis Uo5	108/179
Lactobacillus johnsonii NCC 533	108/179
Lactococcus garvieae Lg2	108/179
Lactobacillus helveticus R0052	107/179
Kyrpidia tusciae DSM 2912	107/179
Lactobacillus kunkeei MP2	107/179
Weissella cibaria CH2	107/179
Lactobacillus gasseri ATCC 33323 = JCM 1131	107/179
Desulfotomaculum reducens MI-1	107/179
Lactobacillus kefiranofaciens ZW3	107/179
Clostridium tetani E88	107/179
Streptococcus uberis 0140J	107/179
Streptococcus parauberis NCFD 2020	107/179
Clostridium lentocellum DSM 5427	107/179
Staphylococcus epidermidis SEI	107/179
Streptococcus macedonicus ACA-DC 198	107/179
Streptococcus suis ZY05719	107/179
Streptococcus iniae YSFST01-82	107/179
Mahella australiensis 50-1 BON	107/179
Lactobacillus hokkaidonensis JCM 18461	107/179
Lactobacillus acidophilus NCFM	106/179
Streptococcus pasteurianus ATCC 43144	106/179
Streptococcus lutetiensis 033	106/179
Lactobacillus reuteri TD1	106/179
Streptococcus infantarius subsp. infantarius CJ18	106/179
Thermincola potens JR	106/179
Streptococcus gallolyticus subsp. gallolyticus ATCC BAA-2069	106/179
Staphylococcus xylosus SMQ-121	106/179
Streptococcus anginosus subsp. whileyi MAS624	106/179
Lactobacillus amylovorus GRL1118	106/179
Lactobacillus gallinarum HFD4	106/179
Desulfotomaculum acetoxidans DSM 771	105/179
Streptococcus pseudopneumoniae IS7493	105/179
Desulfotomaculum ruminis DSM 2154	105/179
Aerococcus urinae ACS-120-V-Col10a	105/179
Streptococcus mutans UA159-FR	105/179
Streptococcus pneumoniae R6	105/179
Oscillibacter valericigenes Sjm18-20	105/179
[Clostridium] stercorarium subsp. stercorarium DSM 8532	105/179
Desulfitobacterium dichloroeliminans LMG P-21439	105/179
Lactobacillus acetotolerans NBRC 13120	105/179
Streptococcus constellatus subsp. pharyngis C818	105/179
Natranaerobius thermophilus JW/NM-WN-LF	105/179
Lachnoclostridium phytofermentans ISDg	105/179
Syntrophobotulus glycolicus DSM 8271	105/179
[Clostridium] saccharolyticum WM1	105/179
Streptococcus equi subsp. zooepidemicus MGCS10565	104/179
Thermacetogenium phaeum DSM 12270	104/179
Lactobacillus delbrueckii subsp. bulgaricus ND02	104/179
Caldicellulosiruptor kristjanssonii I77R1B	104/179
Lactobacillus sanfranciscensis TMW 1.1304	104/179
Caldicellulosiruptor owensensis OL	104/179
Desulfotomaculum kuznetsovii DSM 6115	104/179
Desulfotomaculum nigrificans CO-1-SRB	103/179
Caldicellulosiruptor kronotskyensis 2002	103/179
Pediococcus claussenii ATCC BAA-344	103/179
Leuconostoc mesenteroides subsp. mesenteroides J18	103/179

Acetohalobium arabaticum DSM 5501	103/179
Lactobacillus sp. wkB8	103/179
Clostridium kluyveri NBRC 12016	103/179
Pediococcus pentosaceus SL4	102/179
Leuconostoc citreum KM20	102/179
Caldicellulosiruptor bescii DSM 6725	102/179
Streptococcus agalactiae SS1	102/179
Dehalobacter restrictus DSM 9455	102/179
Desulfotomaculum gibsoniae DSM 7213	102/179
Dehalobacter sp. DCA	102/179
Lactobacillus mucosae LM1	102/179
Heliobacterium modesticaldum Ice1	102/179
Herbinix sp. SD1D	101/179
Halothermothrix orenii H 168	101/179
Caldicellulosiruptor lactoaceticus 6A	101/179
Lactobacillus fermentum IFO 3956	101/179
Peptoclostridium difficile Z31	101/179
Leuconostoc gelidum subsp. gasicomitatum LMG 18811	101/179
Pelotomaculum thermopropionicum SI	101/179
Eubacterium acidaminophilum DSM 3953	101/179
Caldicellulosiruptor hydrothermalis 108	101/179
Streptococcus dysgalactiae subsp. equisimilis RE378	100/179
Moorella thermoacetica DSM 521	100/179
Weissella koreensis KACC 15510	100/179
Syntrophomonas wolfei subsp. wolfei str. Goettingen G311	100/179
Roseburia hominis A2-183	100/179
Caldicellulosiruptor saccharolyticus DSM 8903	100/179
Caldicellulosiruptor obsidiansis OB47	100/179
Leuconostoc kimchii IMSNU 11154	99/179
Weissella ceti WS74	99/179
Thermaerobacter marianensis DSM 12885	99/179
Leuconostoc carnosum JB16	99/179
Halobacteroides halobius DSM 5150	99/179
[Eubacterium] eligens ATCC 27750	99/179
Syntrophothermus lipocalidus DSM 12680	99/179
Streptococcus pyogenes str. Manfredo	98/179
Leuconostoc sp. C2	98/179
Carboxydothermus hydrogenoformans Z-2901	97/179
[Clostridium] cellulosi	97/179
Selenomonas ruminantium subsp. lactilytica TAM6421	96/179
Oenococcus kitaharae DSM 17330	96/179
Ethanoligenens harbinense YUAN-3	96/179
[Eubacterium rectale] ATCC 33656	95/179
Oenococcus oeni PSU-1	94/179
[Clostridium] sticklandii DSM 519	94/179
Butyrivibrio proteoclasticus B316	94/179
Candidatus Arthromitus sp. SFB-rat-Yit	93/179
Candidatus Desulforudis audaxviator MP104C	93/179
Sulfobacillus acidophilus TPY	93/179
Halanaerobium hydrogeniformans missing	93/179
Eubacterium sulci ATCC 35585	92/179
Selenomonas sp. oral taxon 478	91/179
Halanaerobium praevalens DSM 2228	91/179
Ruminococcus albus 7 = DSM 20455	89/179
Ruminococcus bicirculans 80/3	88/179
Eubacterium limosum SA11	88/179
Selenomonas sputigena ATCC 35185	88/179
Acidaminococcus fermentans DSM 20731	87/179
Filifactor alocis ATCC 35896	87/179

Erysipelothrix rhusiopathiae str. Fujisawa	87/179
Finegoldia magna ATCC 29328	86/179
Ammonifex degensii KC4	86/179
Megasphaera elsdenii 14-14	86/179
Clostridium sp. SY8519	85/179
Acidaminococcus intestini RyC-MR95	84/179
Veillonella parvula DSM 2008	83/179
Peptoniphilus sp. 1-1	82/179
Acetobacterium woodii DSM 1030	82/179
Parvimonas micra KCOM 1535 (=ChDC B708)	81/179
Anaerococcus prevotii DSM 20548	81/179
Mageeibacillus indolicus UPII9-5	70/179

.11 Phylogénies non racinées des familles multigéniques.

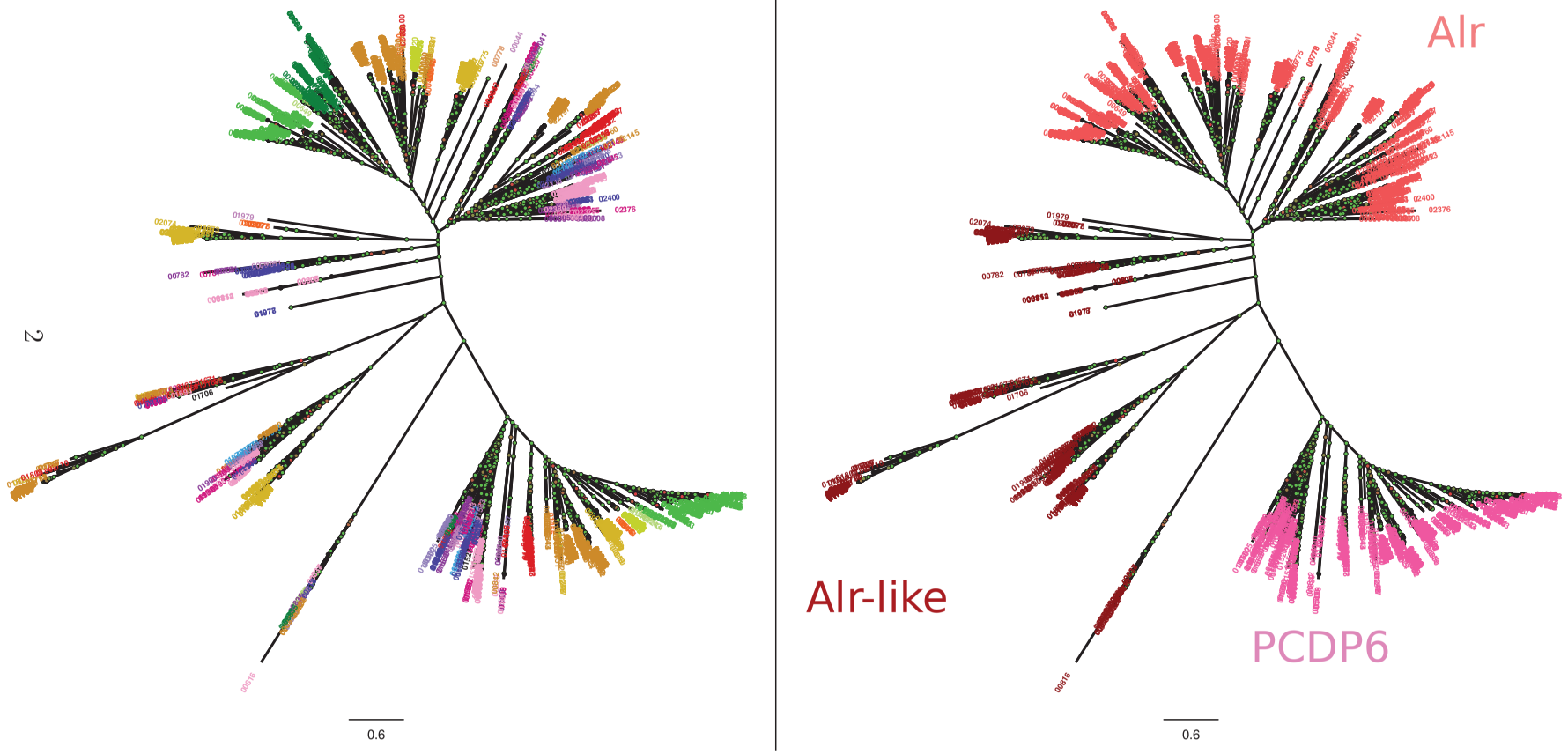
1. AlaS. 2372 séquences, 1390 positions sans nettoyage, FastTree, WAG+G4
2. Alr PCDP6. 2485 séquences, 1401 positions sans nettoyage, FastTree, WAG+G4
3. Aslfm. 1329 séquences, 248 positions, FastTree, WAG+G4
4. CDP1 (Ccrz). 1574 séquences, 1810 positions sans nettoyage, FastTree, WAG+G4
5. CozE. 3293 séquences, 149 positions, FastTree, WAG+G4
6. CpsD ParA MinD. 3928 séquences, 4132 positions sans nettoyage, FastTree, WAG+G4
7. DacA DacB DacF. 2732 séquences, 2760 positions sans nettoyage, FastTree, WAG+G4
8. Ddl. 6829 séquences, 5359 positions sans nettoyage, FastTree, WAG+G4
9. DivIC FtsL. 1985 séquences, 659 positions sans nettoyage, FastTree, WAG+G4
10. DivIVA GpsB. 1668 séquences, 768 positions sans nettoyage, FastTree, WAG+G4
11. DnaA DnaI. 2934 séquences, 1737 positions sans nettoyage, FastTree, WAG+G4
12. DnaB. 1285 séquences, 374 positions, FastTree, WAG+G4
13. DnaD. 1587 séquences, 1853 positions sans nettoyage, FastTree, WAG+G4
14. DnaG. 1053 séquences, 222 positions, FastTree, WAG+G4
15. DnaN. 1090 séquences, 804 positions, FastTree, WAG+G4
16. FtsA Mbl MreB MreBH. 3707 séquences, 128 positions, FastTree, WAG+G4
17. FtsH ClpX SpoVK. 5671 séquences, 4884 positions sans nettoyage, FastTree, WAG+G4
18. FtsJ. 1828 séquences, 151 positions, FastTree, WAG+G4
19. FtsK. 2515 séquences, 135 positions, FastTree, WAG+G4
20. FtsW RodA SpoVE (SEDS). 2910 séquences, 193 positions, FastTree, WAG+G4
21. FtsX. 3030 séquences, 8261 positions sans nettoyage, FastTree, WAG+G4
22. FtsY. 2325 séquences, 228 positions, FastTree, WAG+G4
23. GatD. 1012 séquences, 239 positions, FastTree, WAG+G4
24. GidA. 1816 séquences, 304 positions, FastTree, WAG+G4

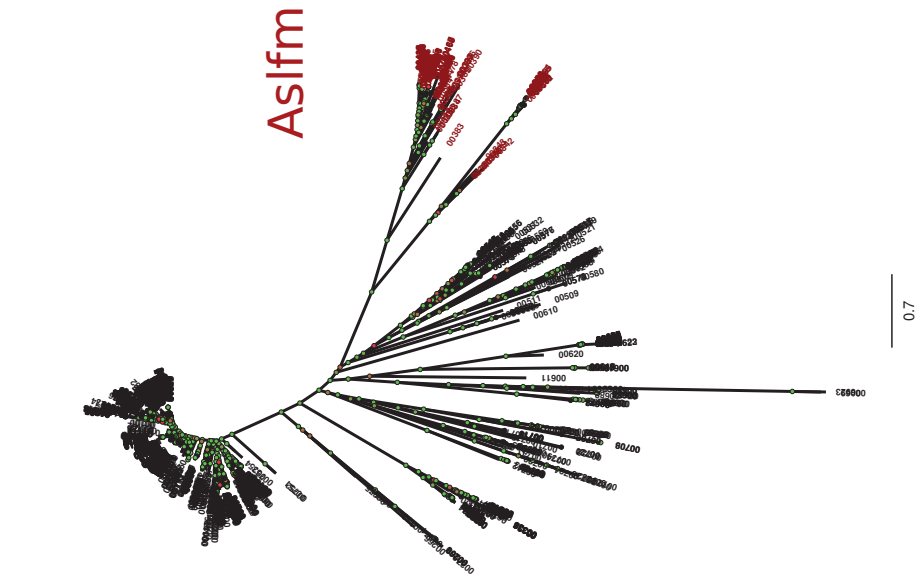
25. GlmM. 2432 séquences, 300 positions, FastTree, WAG+G4
26. GlmS. 3495 séquences, 64 positions, FastTree, WAG+G4
27. GlmU. 9133 séquences, 4787 positions sans nettoyage, FastTree, WAG+G4
28. IleS ValS LeuS. 4955 séquences, 236 positions, FastTree, WAG+G4
29. LysA. 1130 séquences, 297 positions, FastTree, WAG+G4
30. MraW. 1929 séquences, 133 positions, FastTree, WAG+G4
31. MraY WecA. 1951 séquences, 227 positions, FastTree, WAG+G4
32. Mtf. 2039 séquences, 132 positions, FastTree, WAG+G4
33. MurA MurZ. 2756 séquences, 273 positions, FastTree, WAG+G4
34. MurB. 2164 séquences, 134 positions, FastTree, WAG+G4
35. MurC MurD MurE MurF MurT. 5829 séquences, 88 positions, FastTree, WAG+G4
36. MurG. 1844 séquences, 120 positions, FastTree, WAG+G4
37. MurI Asr. 2626 séquences, 57 positions, FastTree, WAG+G4
38. MurJ YabM YkvU SpoVB. 3090 séquences, 158 positions, FastTree, WAG+G4
39. MurK. 309 séquences, 154 positions, FastTree, WAG+G4
40. MurN MurM FemA FemB FemX fmh. 1037 séquences, 265 positions, FastTree, WAG+G4
41. MurQ. 884 séquences, 175 positions, FastTree, WAG+G4
42. NagA. 4780 séquences, 3730 positions sans nettoyage, FastTree, WAG+G4
43. NagB. 1201 séquences, 189 positions, FastTree, WAG+G4
44. Noc ParB. 1897 séquences, 1914 positions sans nettoyage, FastTree, WAG+G4
45. PBPs. 6425 séquences, 7816 positions sans nettoyage, FastTree, WAG+G4
46. NudF. 6439 séquences, 2500 positions sans nettoyage, FastTree, WAG+G4
47. Pfs. 2231 séquences, 1930 positions sans nettoyage, FastTree, WAG+G4
48. PhpP. 1102 séquences, 131 positions, FastTree, WAG+G4
49. PriA, Mfd, RecG. 7733 séquences, 120 positions, FastTree, WAG+G4
50. RacA. 473 séquences, 84 positions, FastTree, WAG+G4

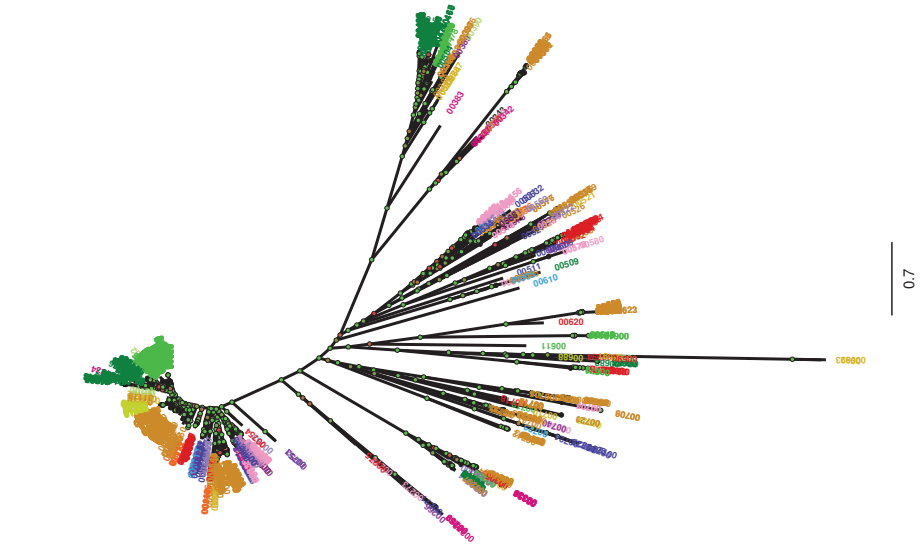
51. Smc RecN RecF SbcC. 3664 séquences, 87 positions, FastTree, WAG+G4
52. SpoIID LytB(BS). 684 séquences, 129 positions, FastTree, WAG+G4
53. SpoIIE. 1260 séquences, 89 positions, FastTree, WAG+G4
54. SpoIIIJ. 1636 séquences, 1359 positions sans nettoyage, FastTree, WAG+G4
55. CDP4 (Spr1400). 726 séquences, 65 positions, FastTree, WAG+G4
56. StkP. 1870 séquences, 68 positions, FastTree, WAG+G4
57. SunL. 2560 séquences, 55 positions, FastTree, WAG+G4
58. TilS. 2105 séquences, 105 positions, FastTree, WAG+G4
59. WalH WalI. 1190 séquences, 1445 positions sans nettoyage, FastTree, WAG+G4
60. WalJ. 6750 séquences, 4522 positions sans nettoyage, FastTree, WAG+G4
61. WalK. 1818 séquences, 326 positions, FastTree, WAG+G4
62. WalR. 8565 séquences, 123 positions, FastTree, WAG+G4
63. XerC XerD XerS. 6627 séquences, 4078 positions sans nettoyage, FastTree, WAG+G4

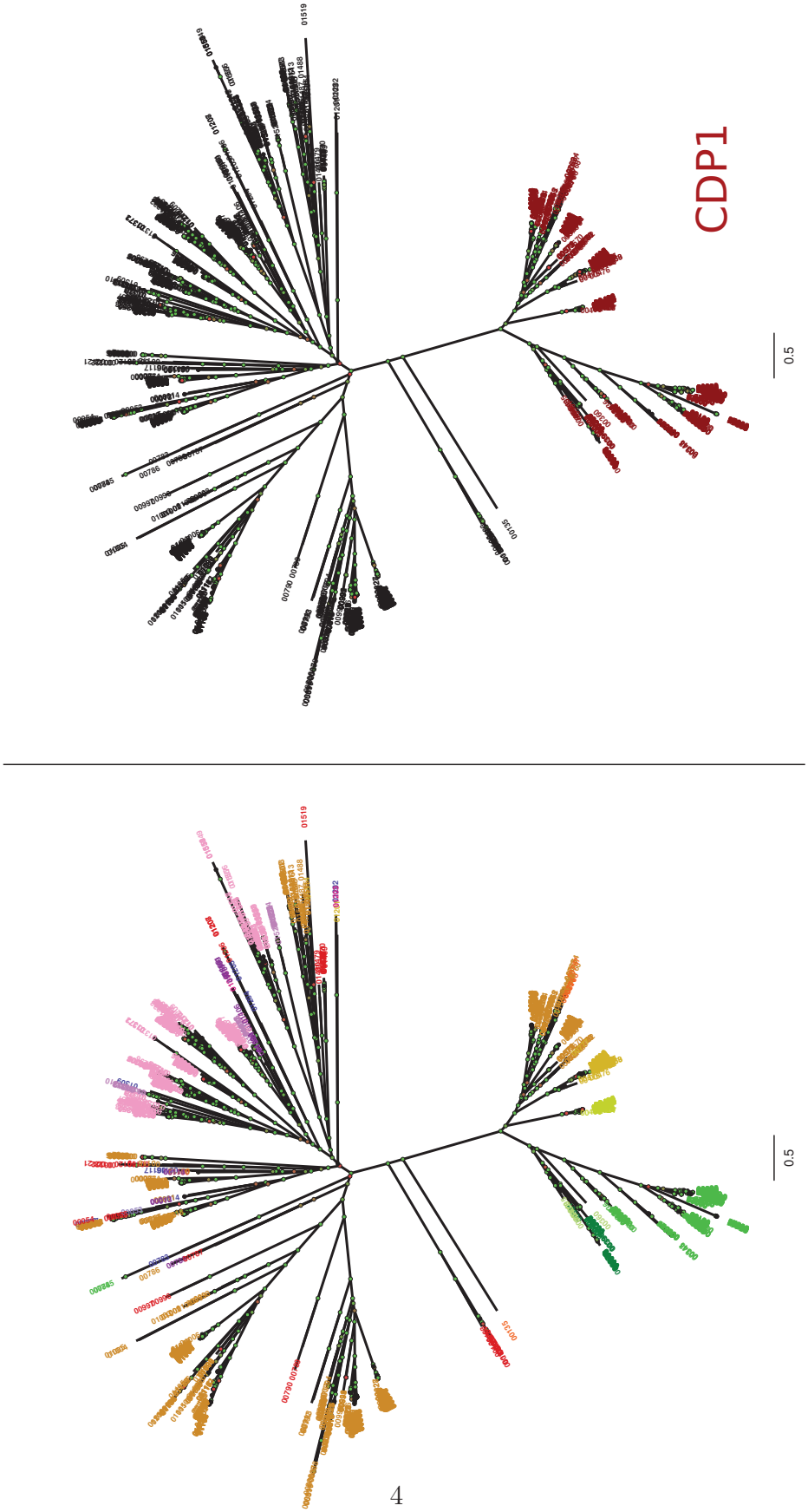
AlaS
0.7

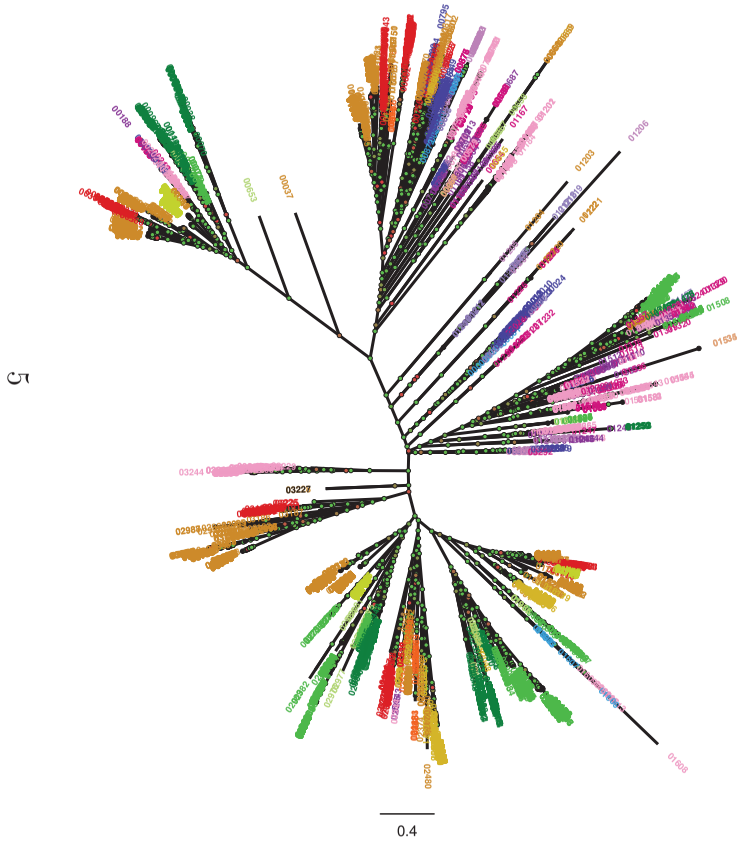
0.7

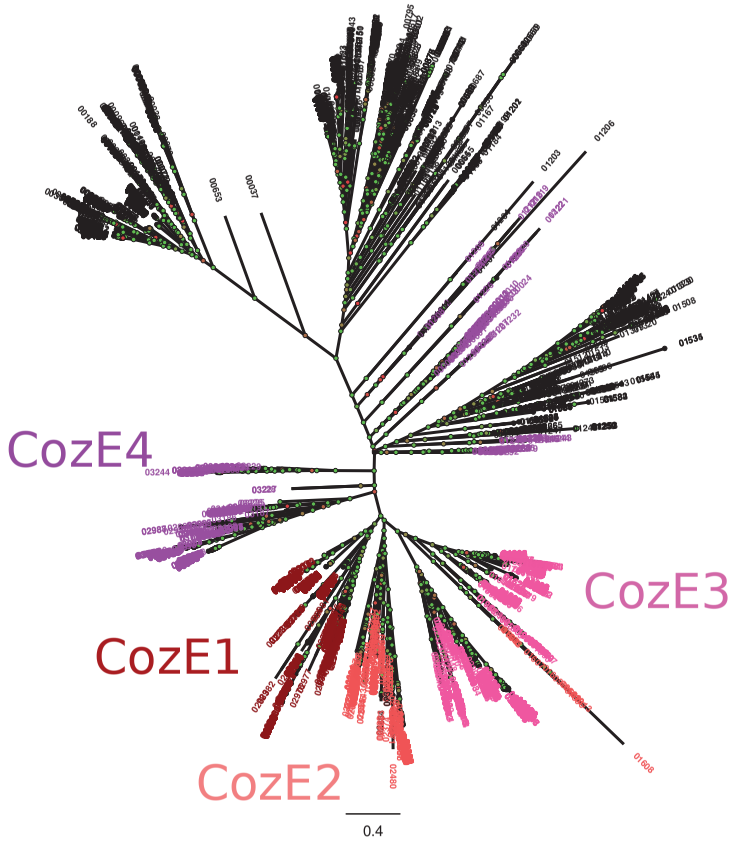

2
0.6
Alr
Alr-like
PCDP6
0.6

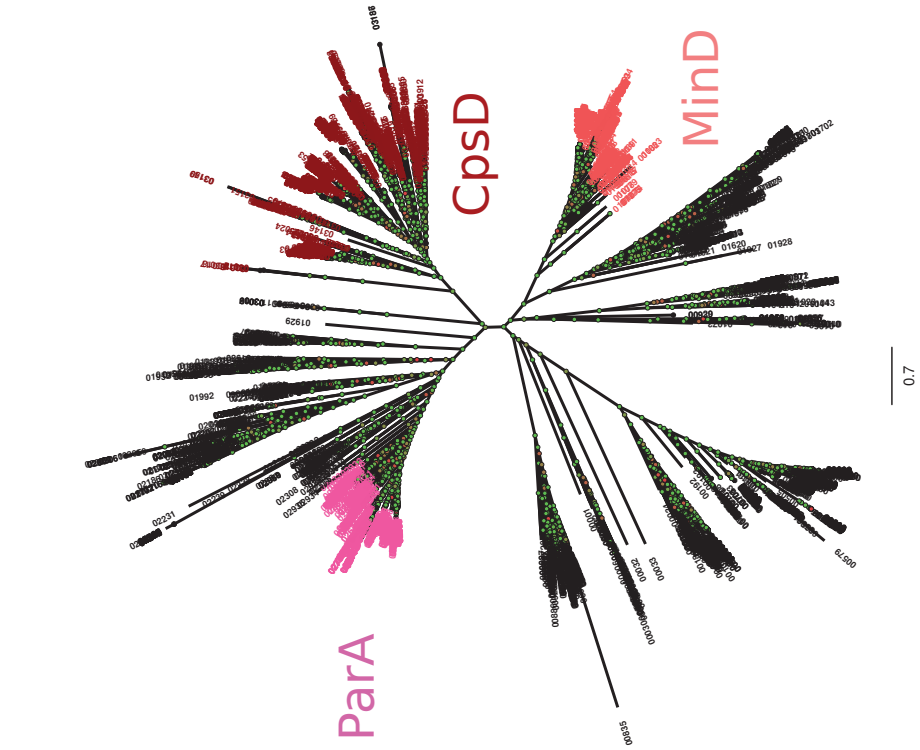

Aslfm
0.7

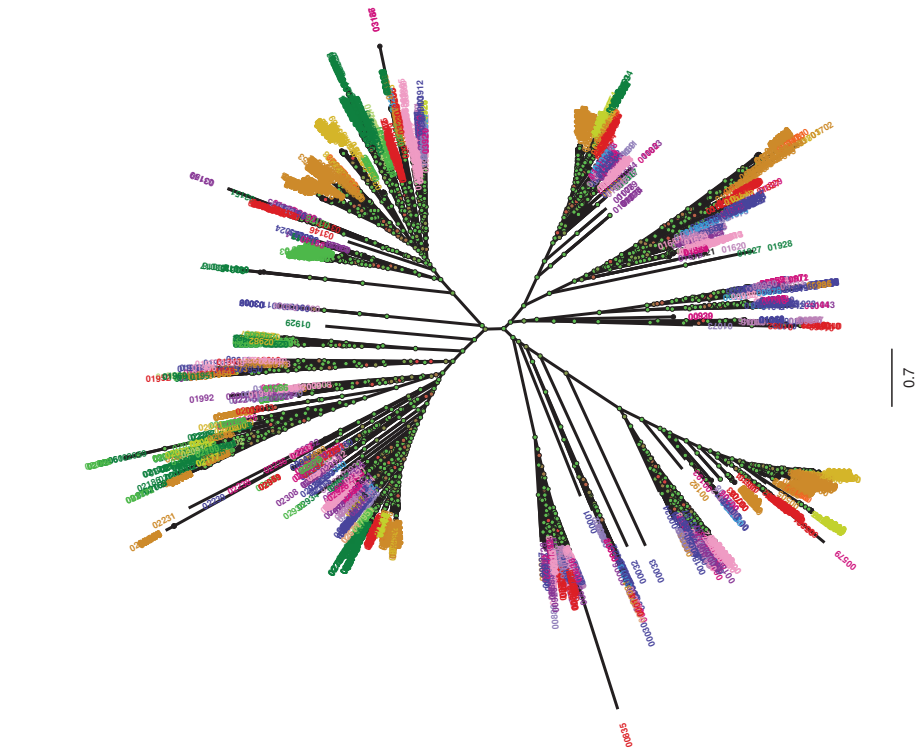

0.7

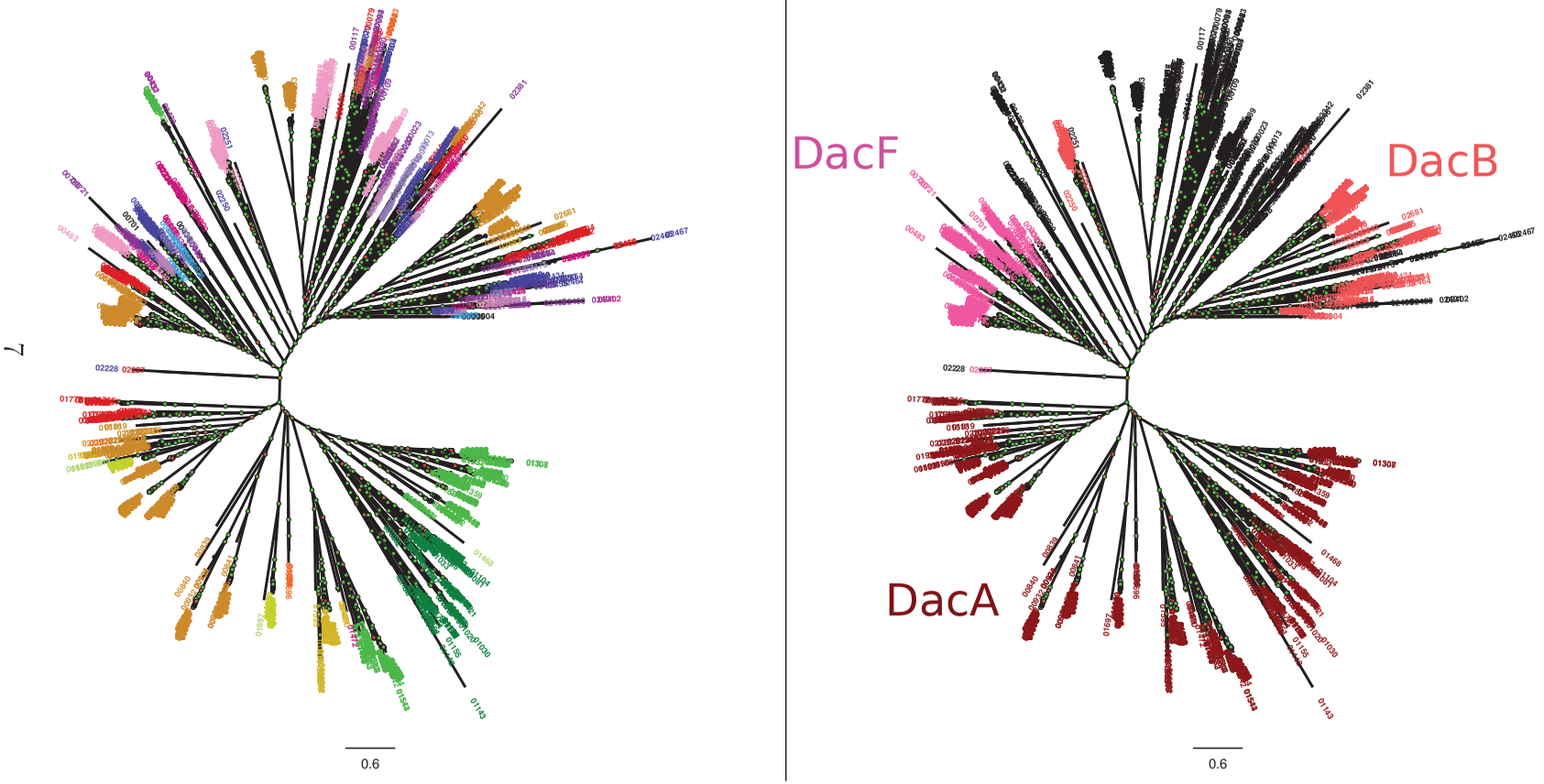

CDP1
0.5
0.5

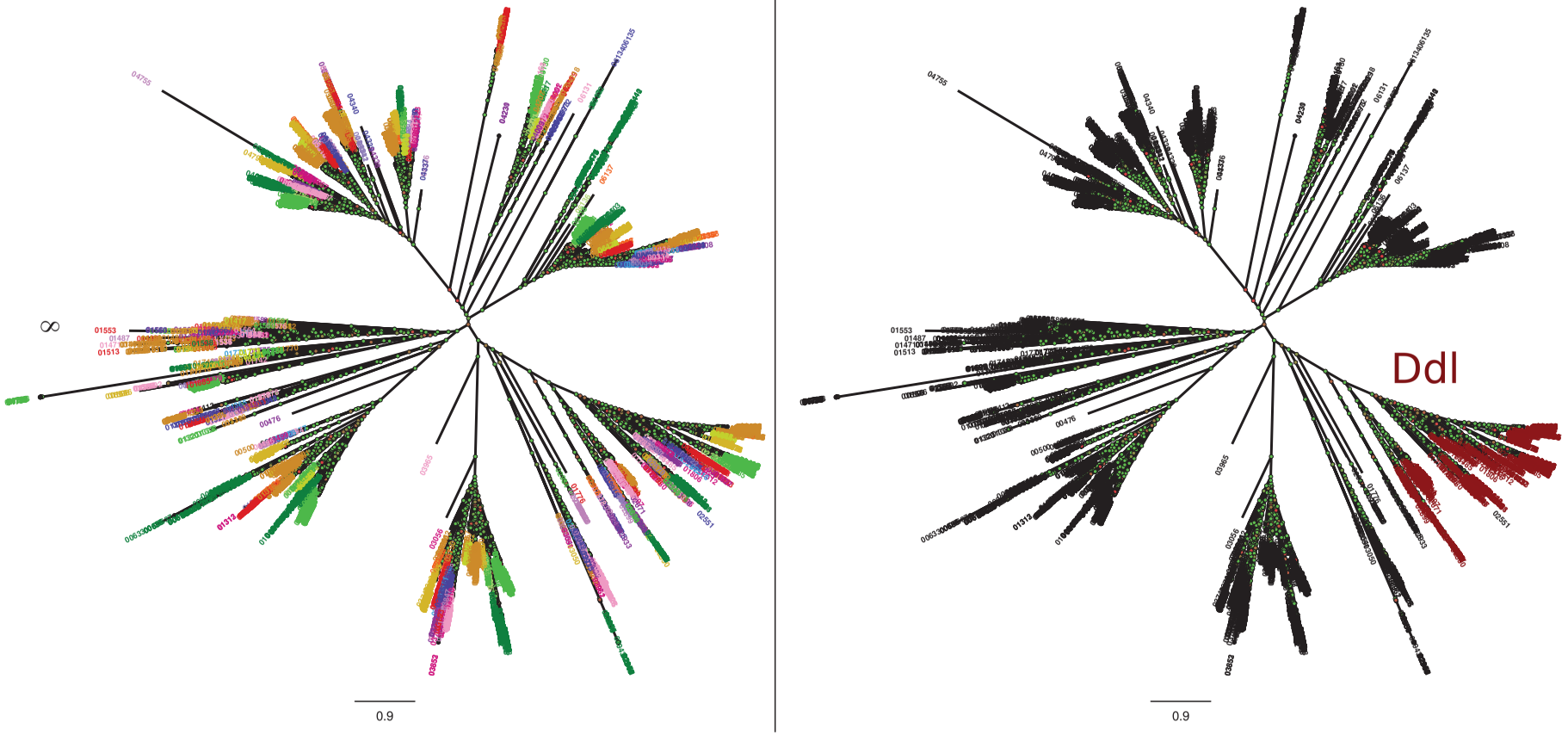

0.4

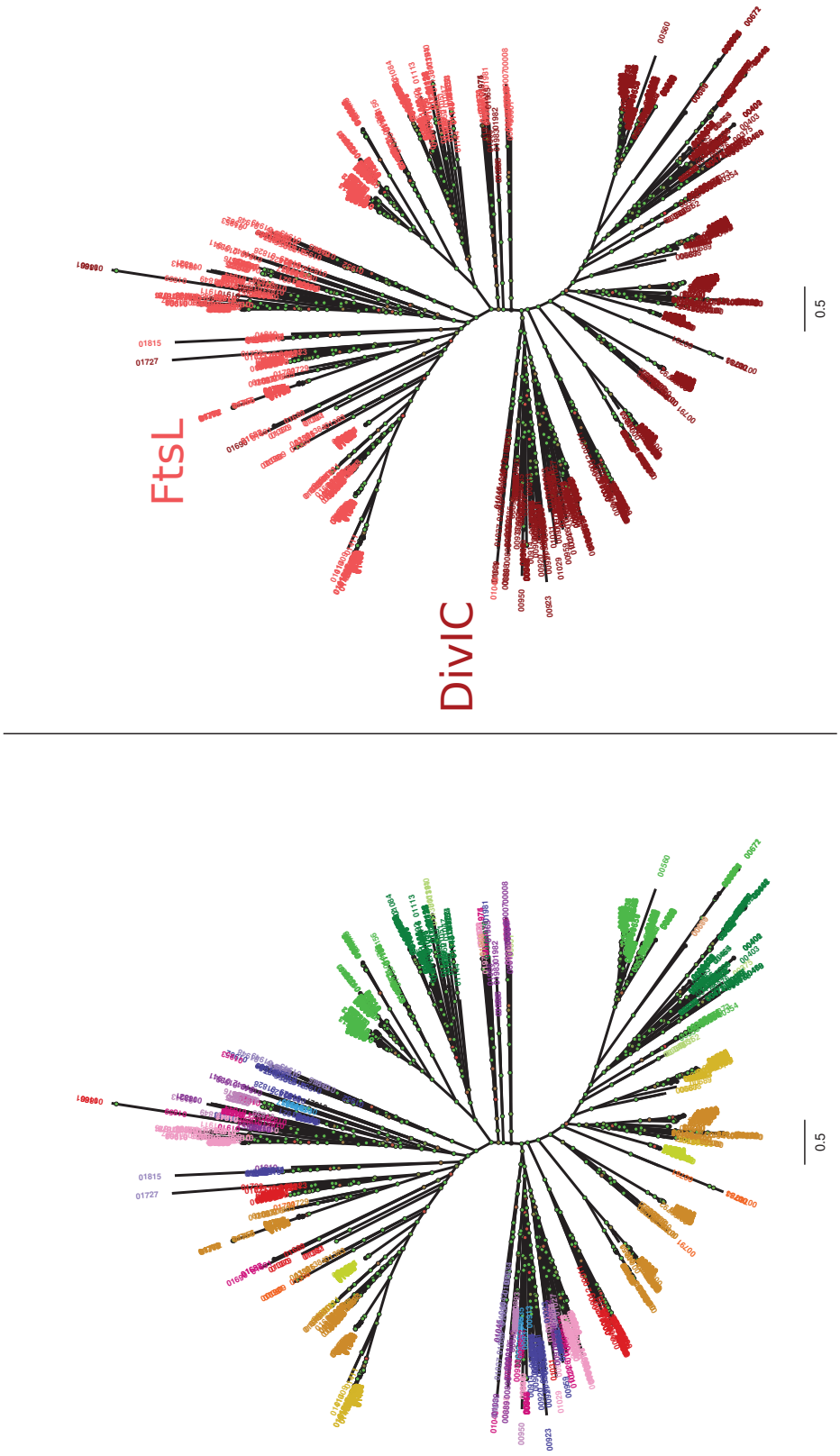

CozE4
CozE3
CozE1
CozE2
0.4

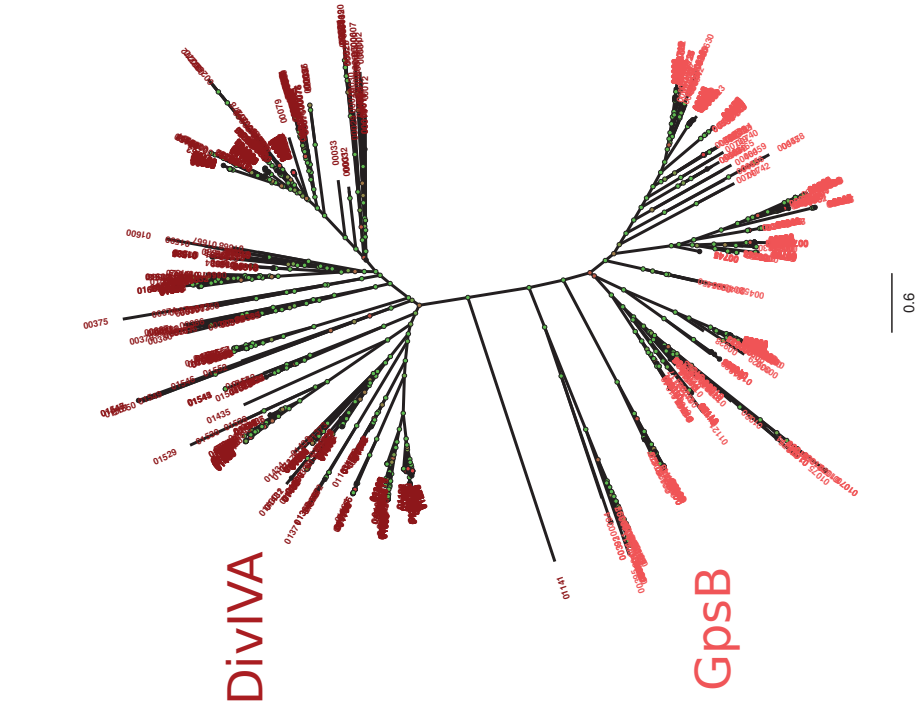

CpsD
MinD
ParA
0.7

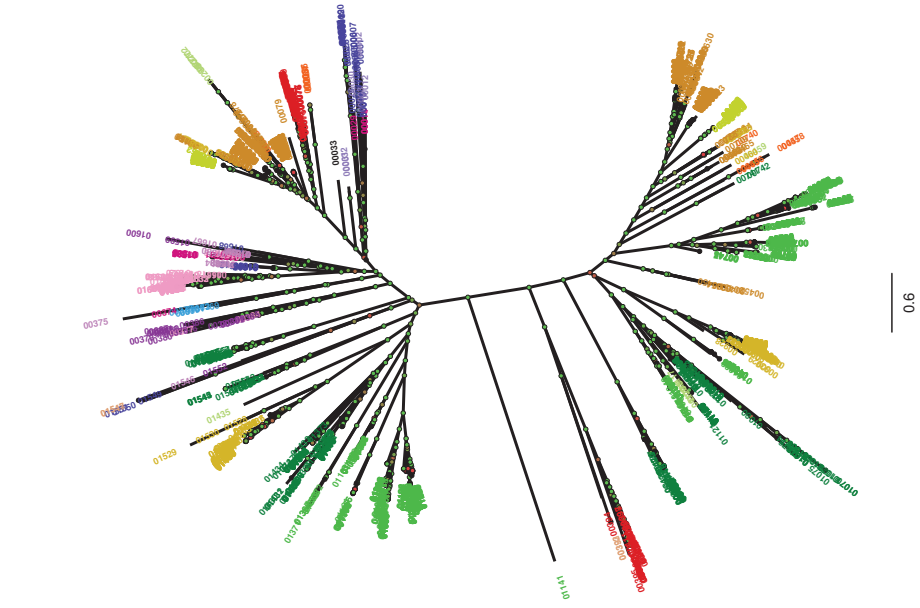

0.7

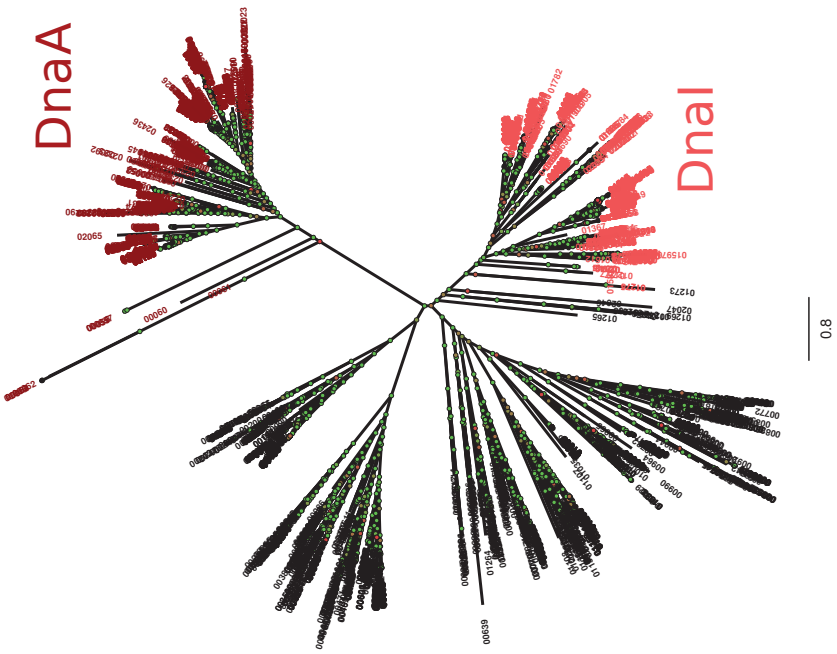

7
DacF
DacB
DacA
0.6
0.6

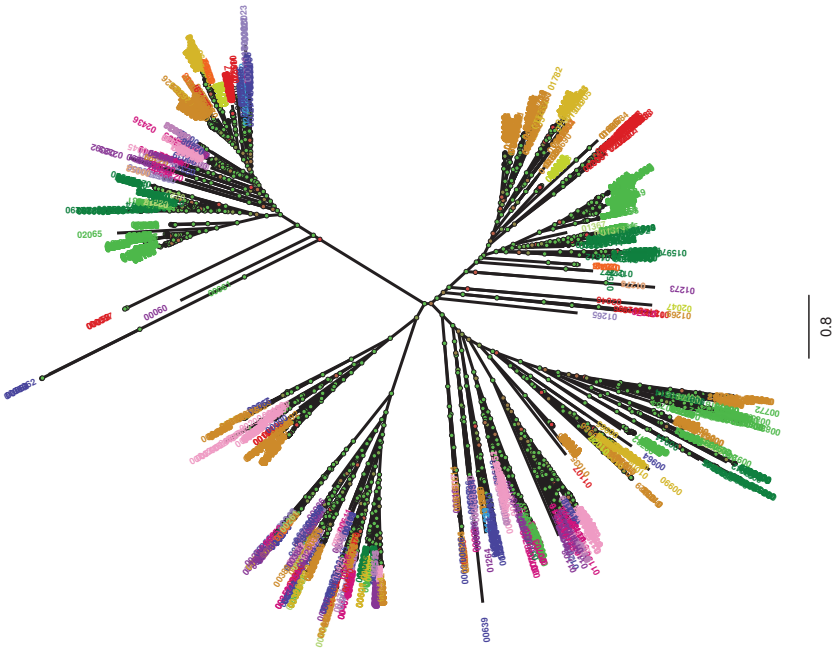

∞
0.9
Ddl
0.9

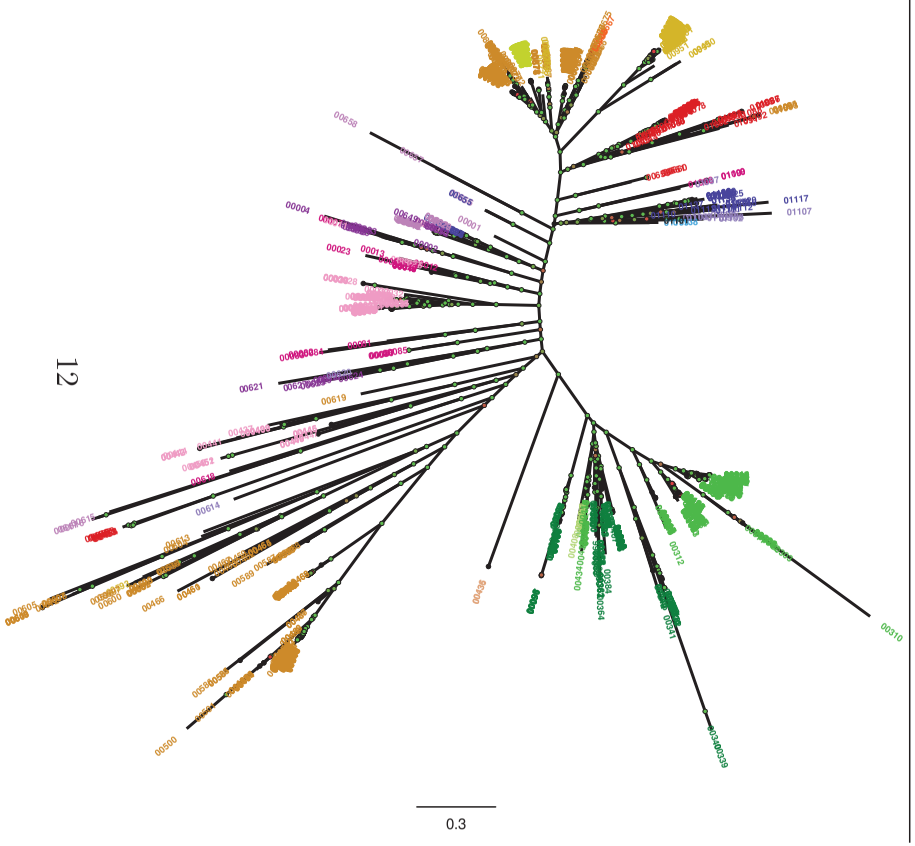

FtsL
DivIC
0.5
0.5

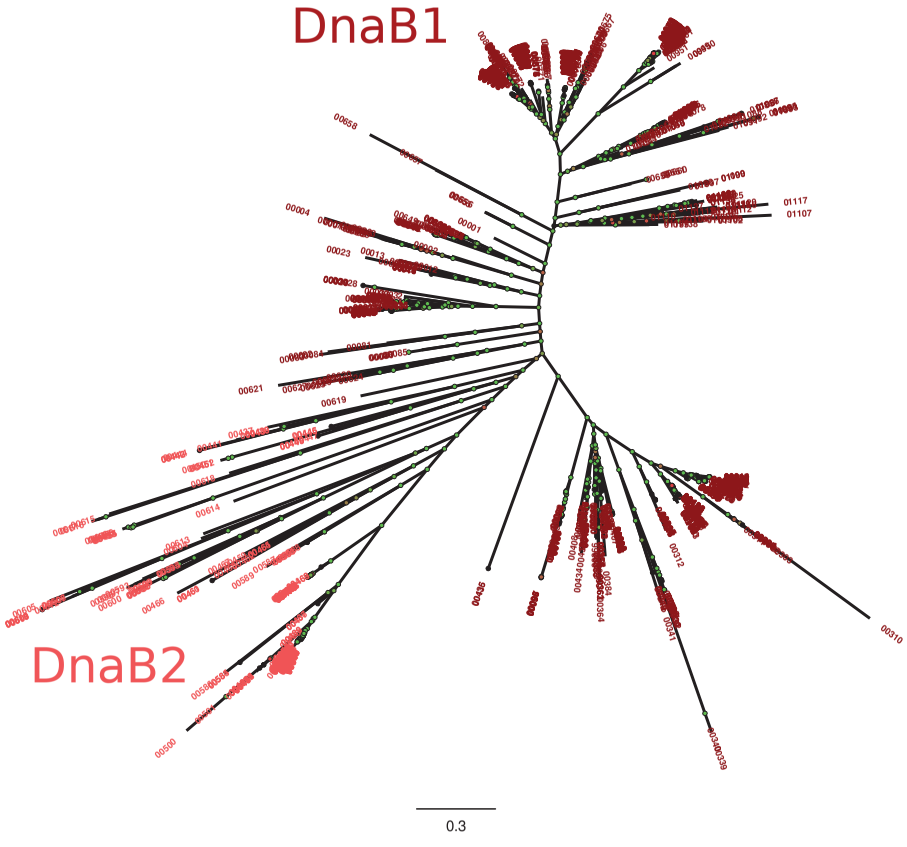

DivIVA
GpsB
0.6

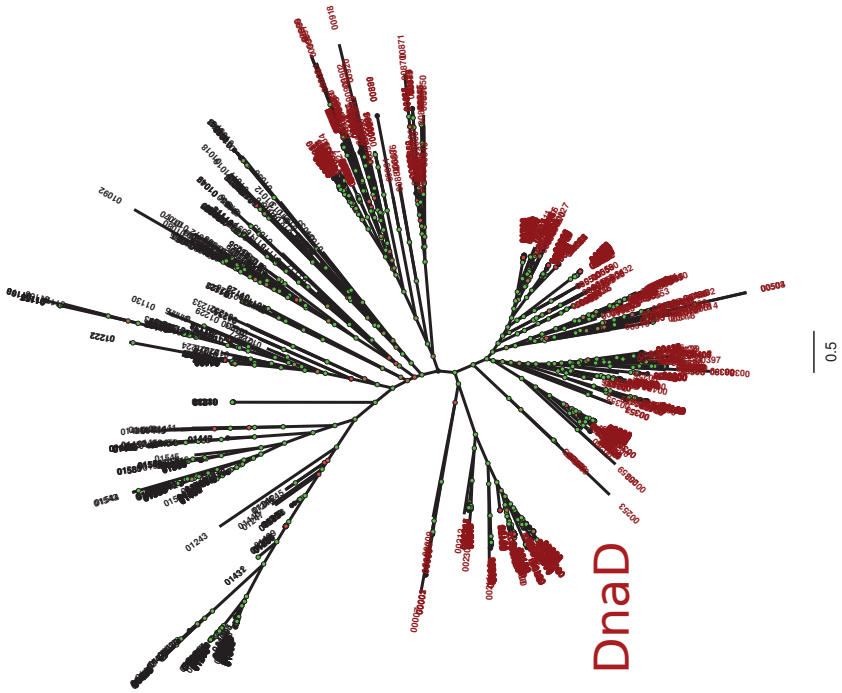

0.6

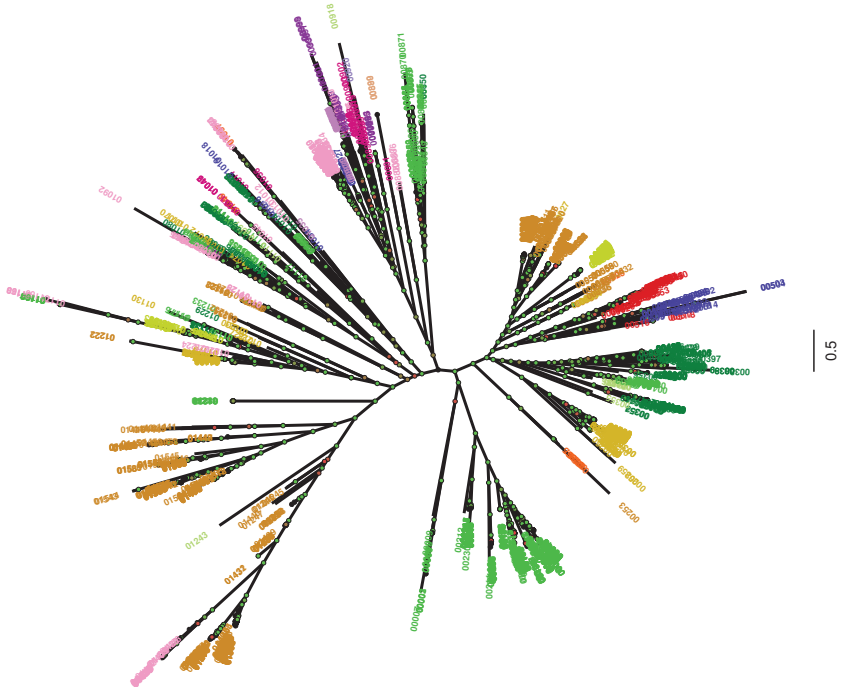

DnaA
DnaI
0.8

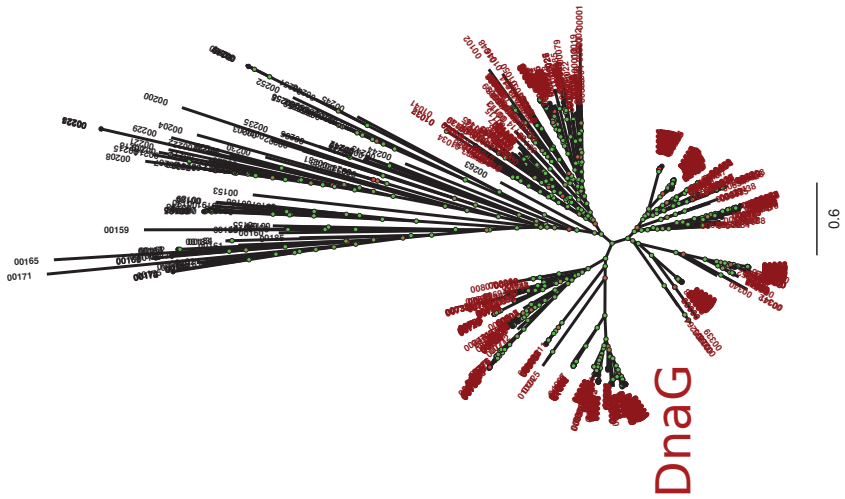

0.8

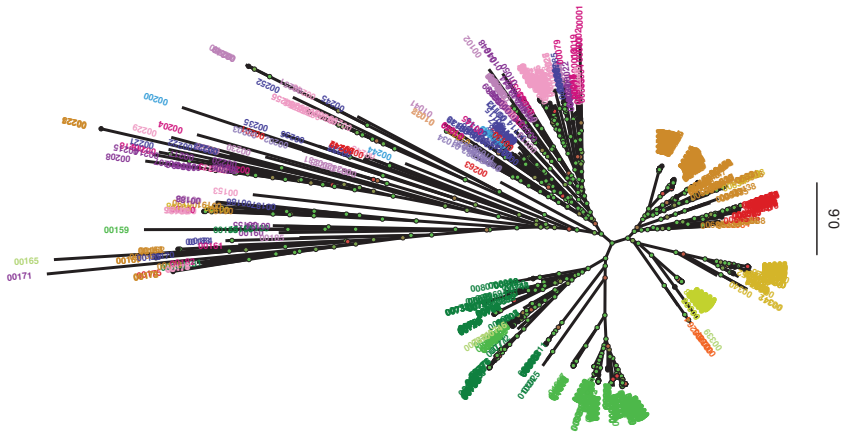

0.3

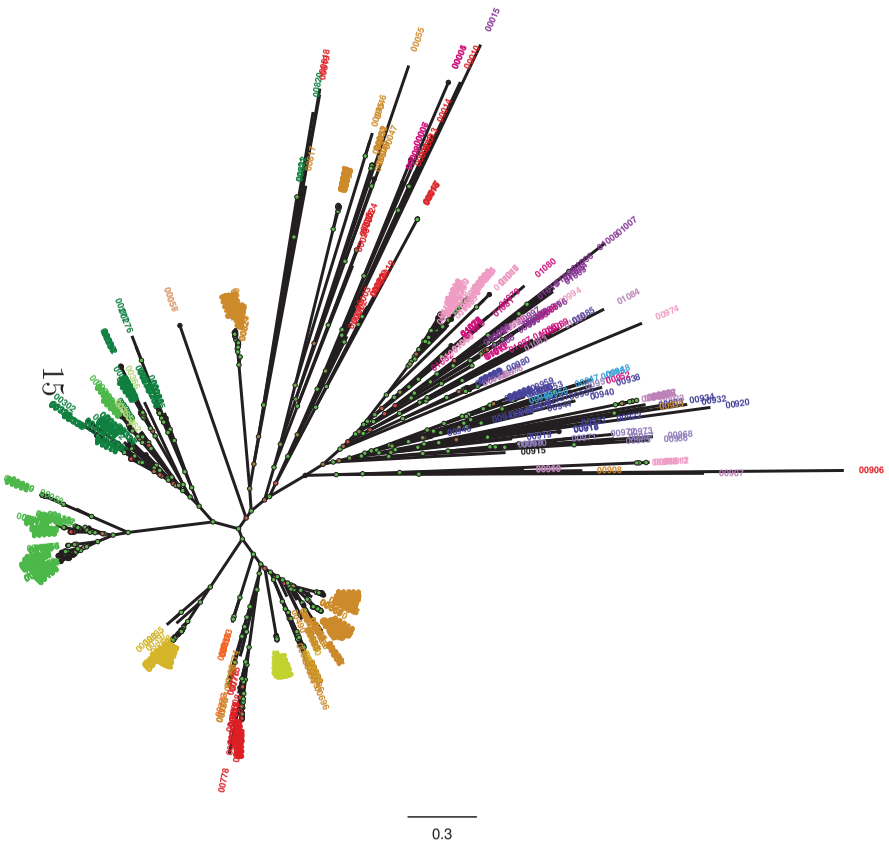

DnaB1
DnaB2
0.3

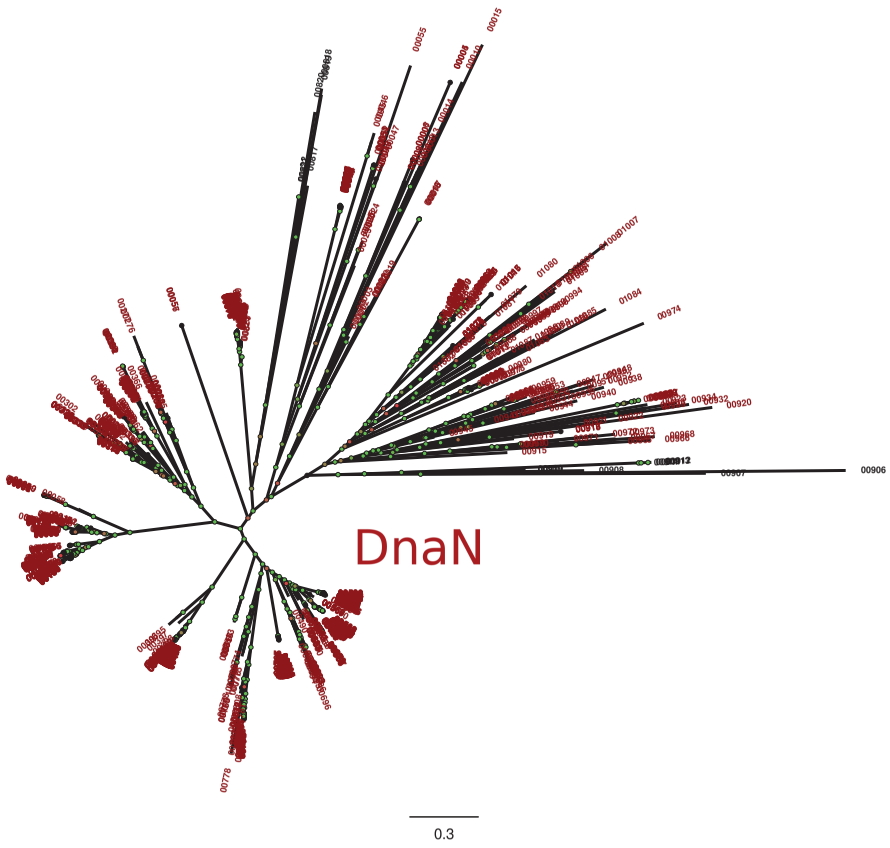

DnaD
0.5

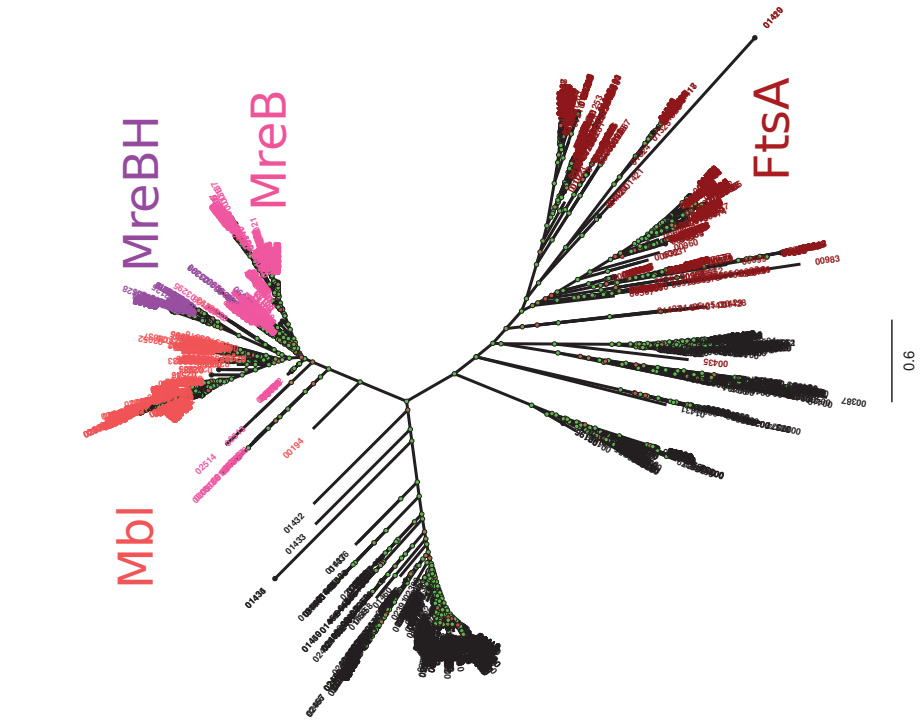

0.5

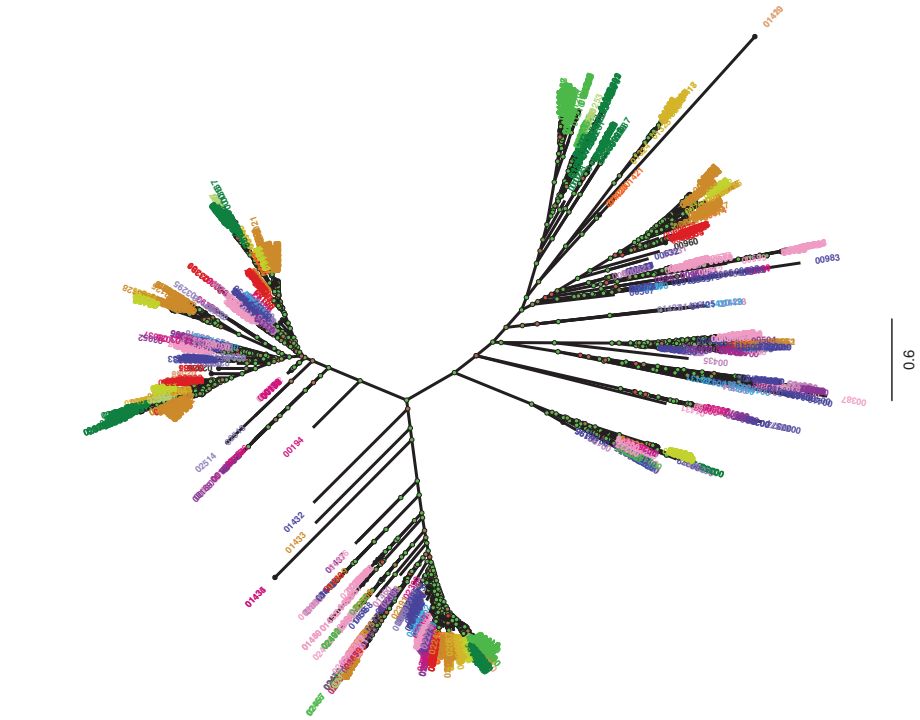

DnaG
0.6

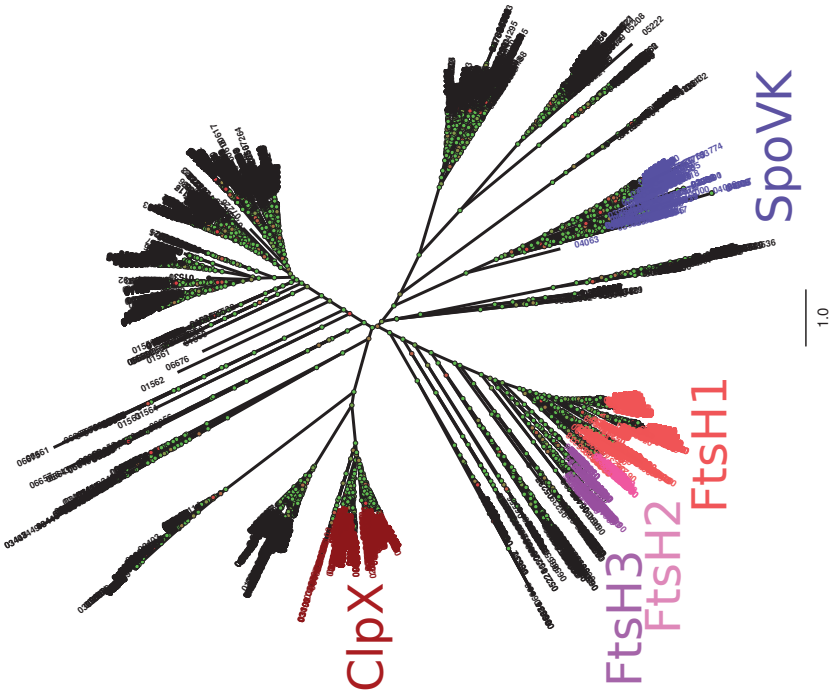

0.6

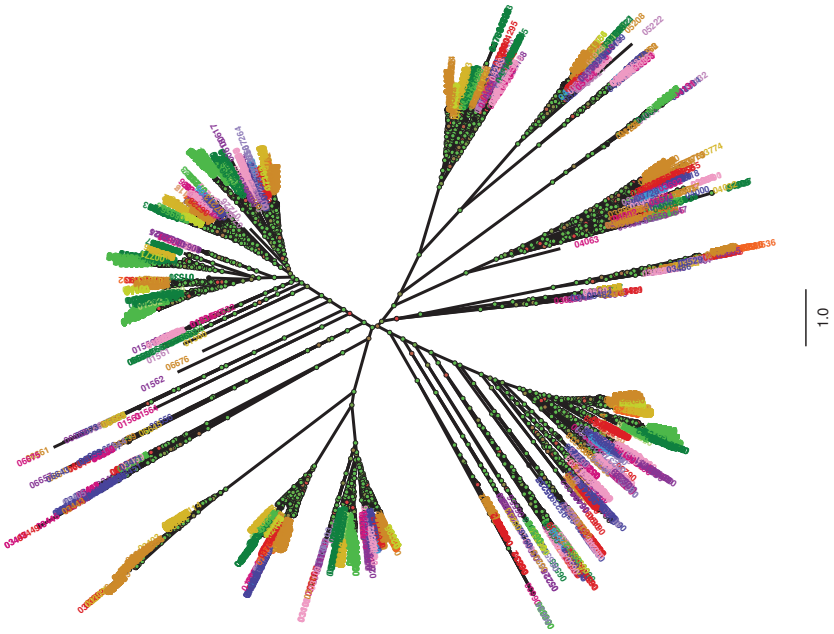

0.3

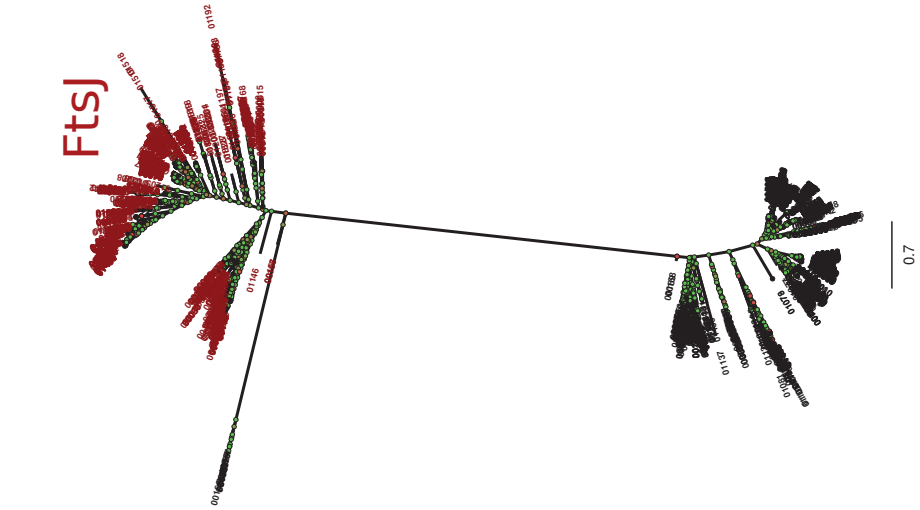

DnaN
0.3

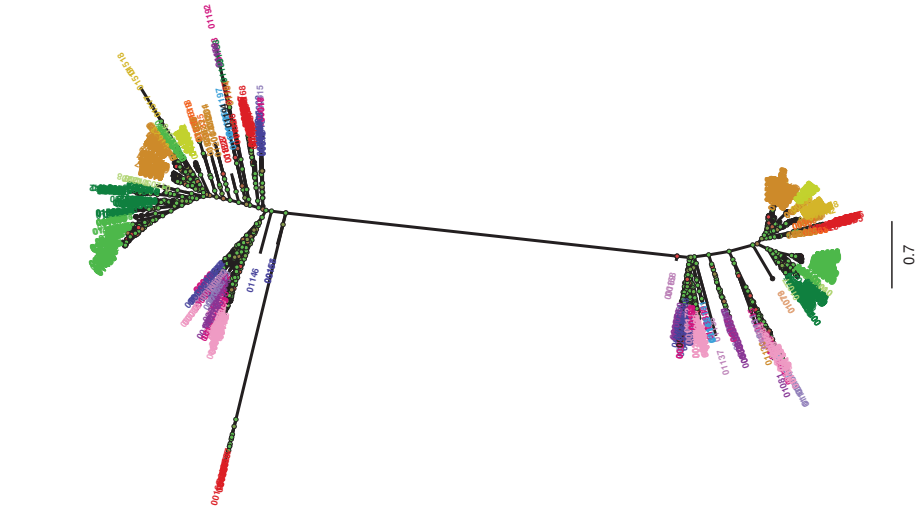

MreBH
MreB
FtsA
Mbl
0.6

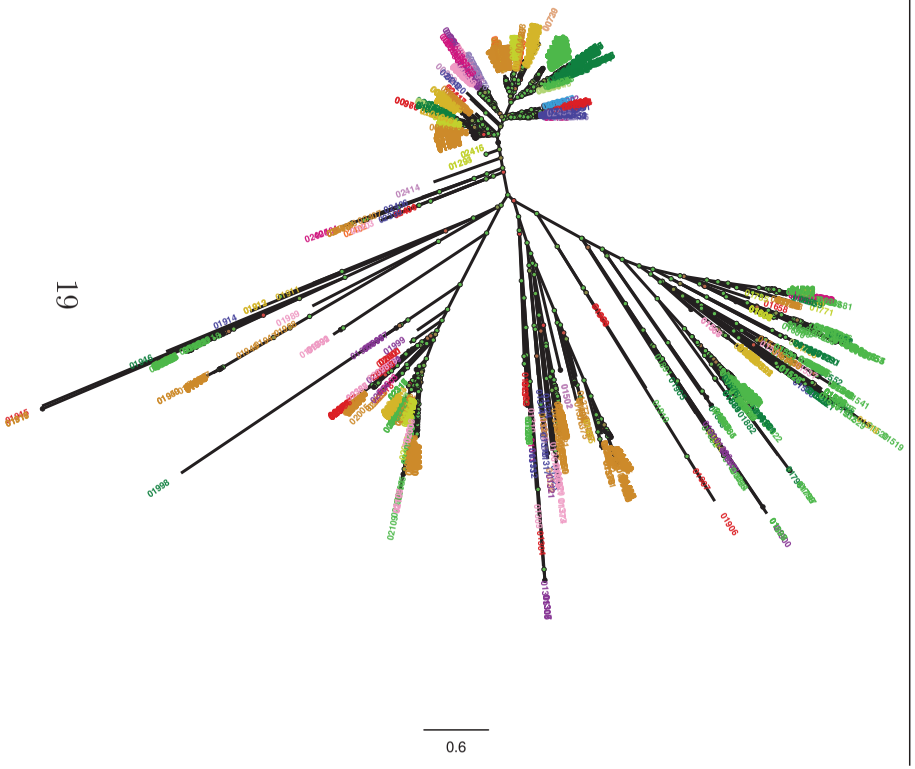

0.6

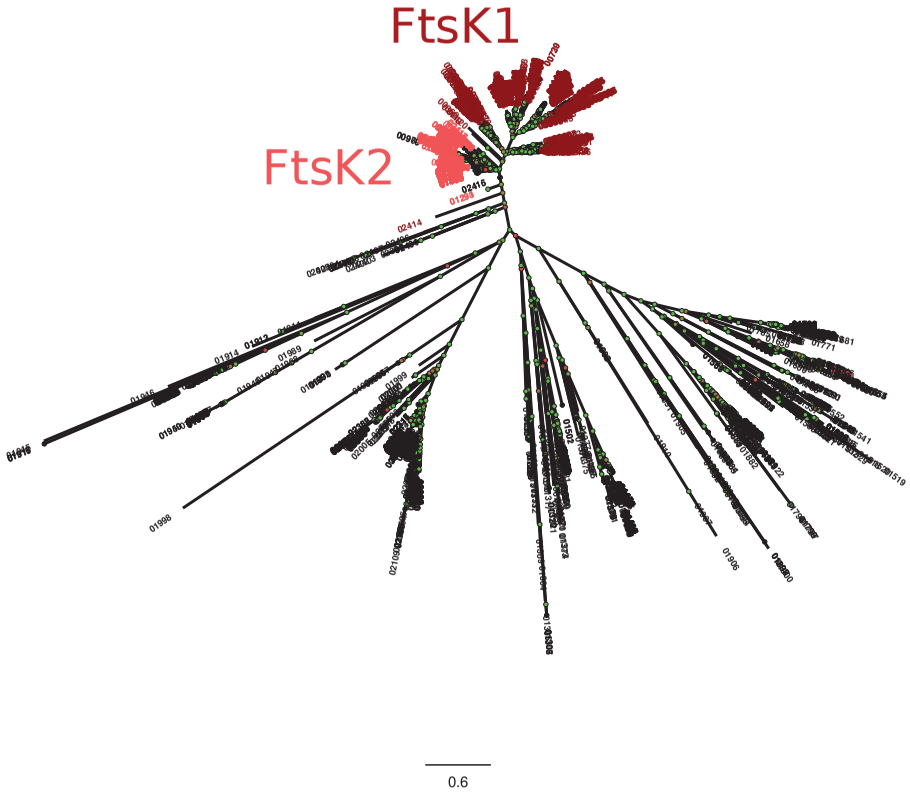

SpoVK
FtsH1
FtsH2
FtsH3
ClpX
1.0

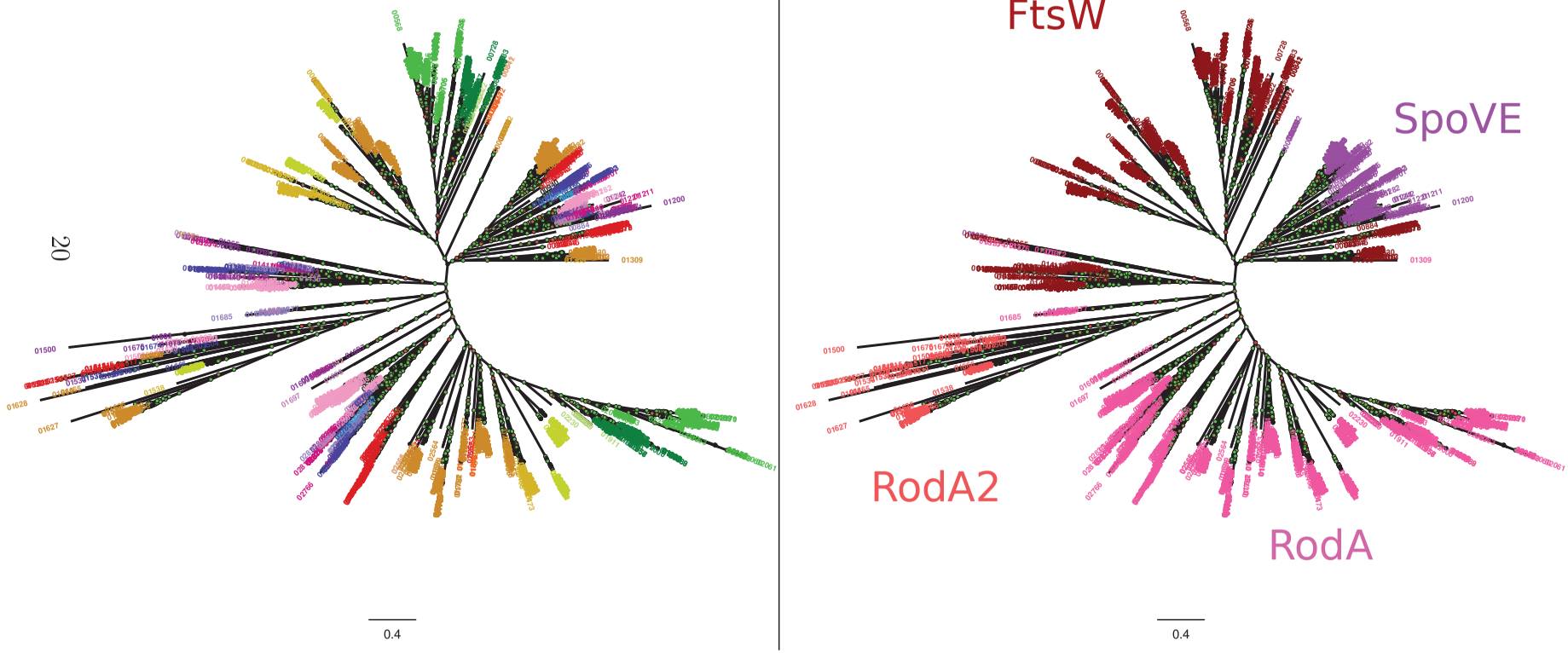

1.0

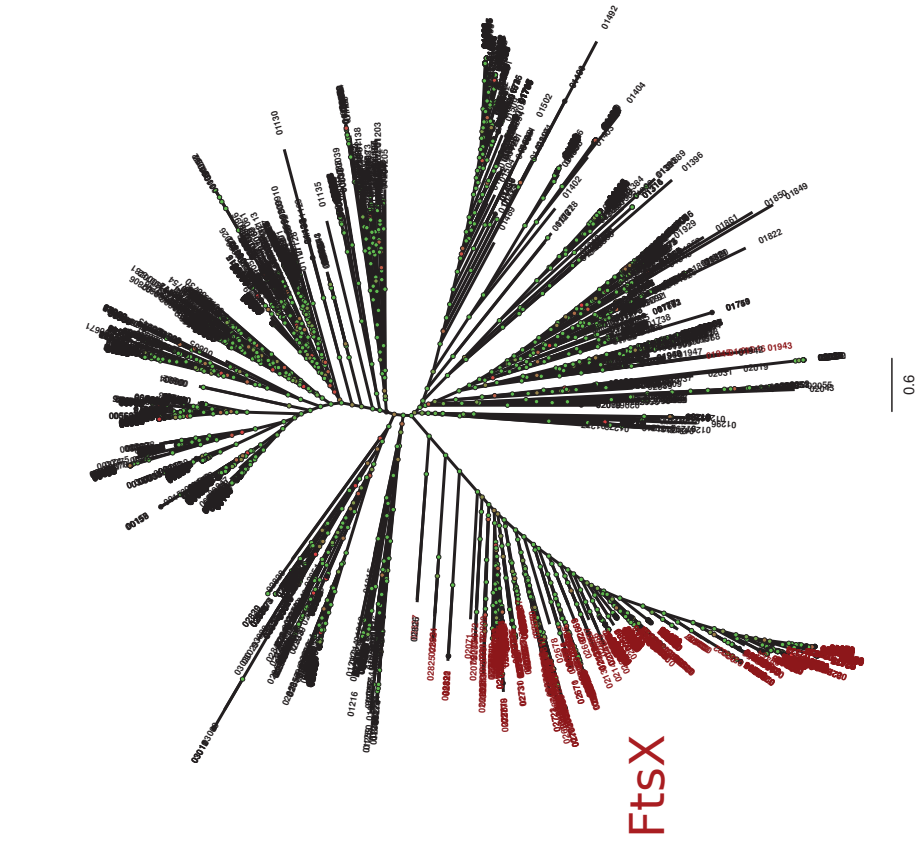

FtsJ
0.7

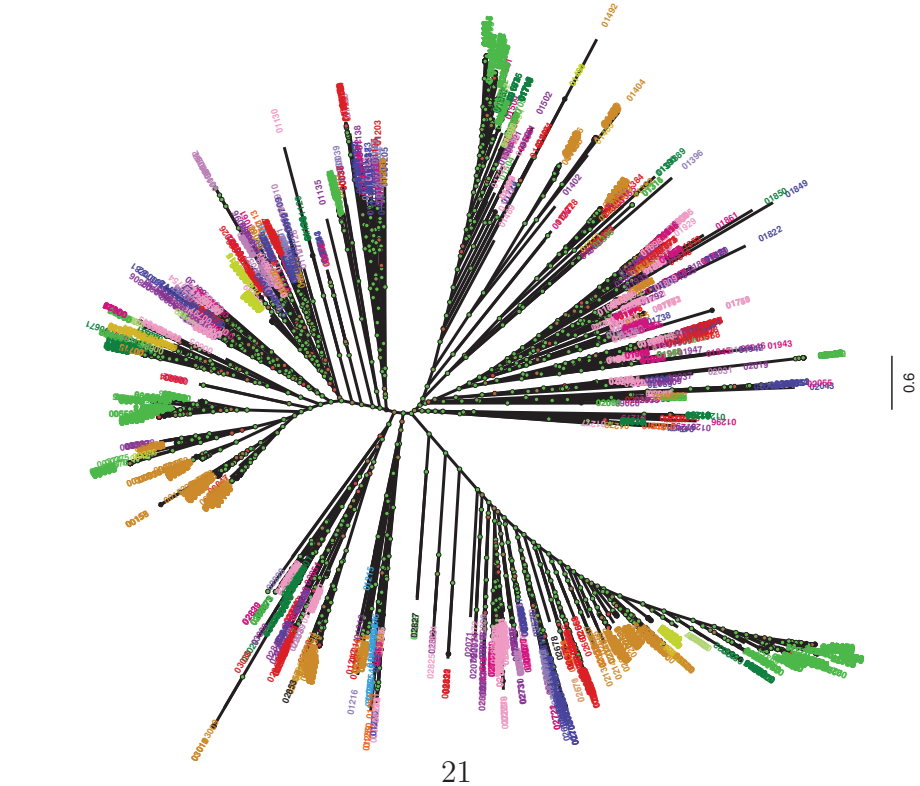

0.7

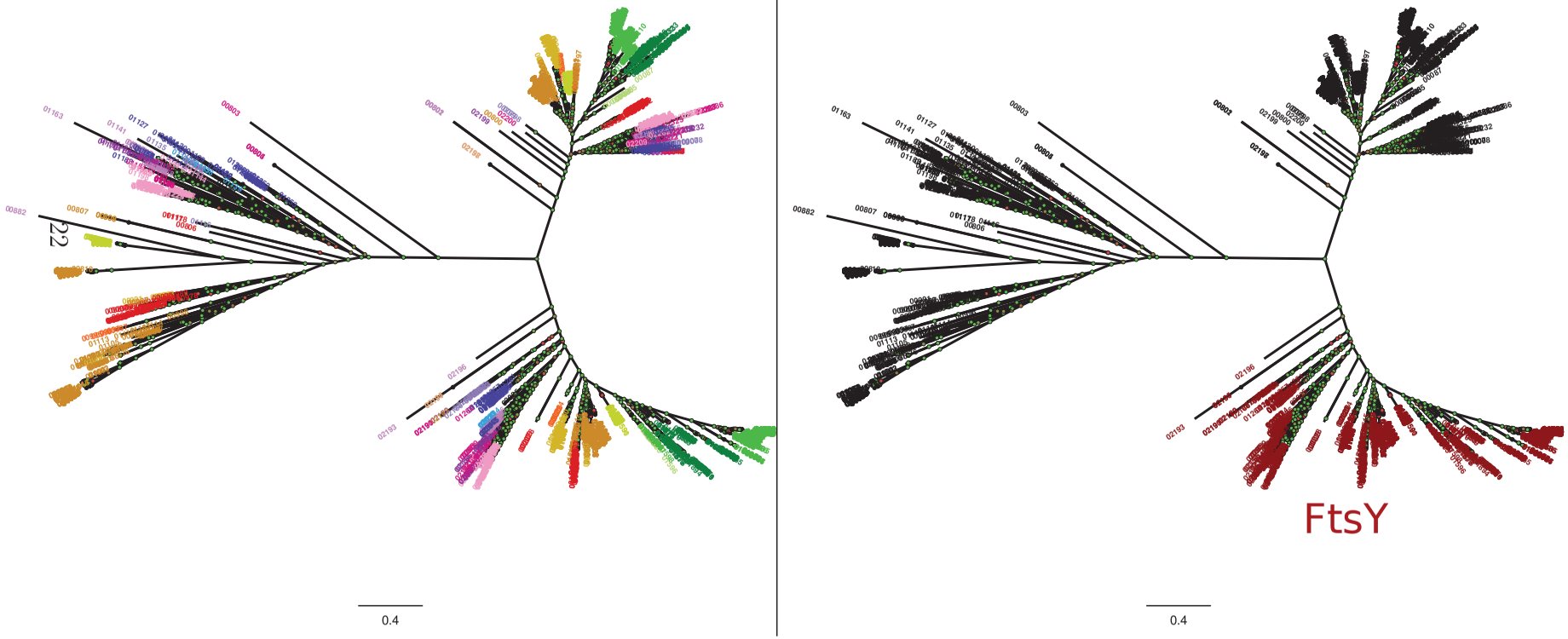

0.6

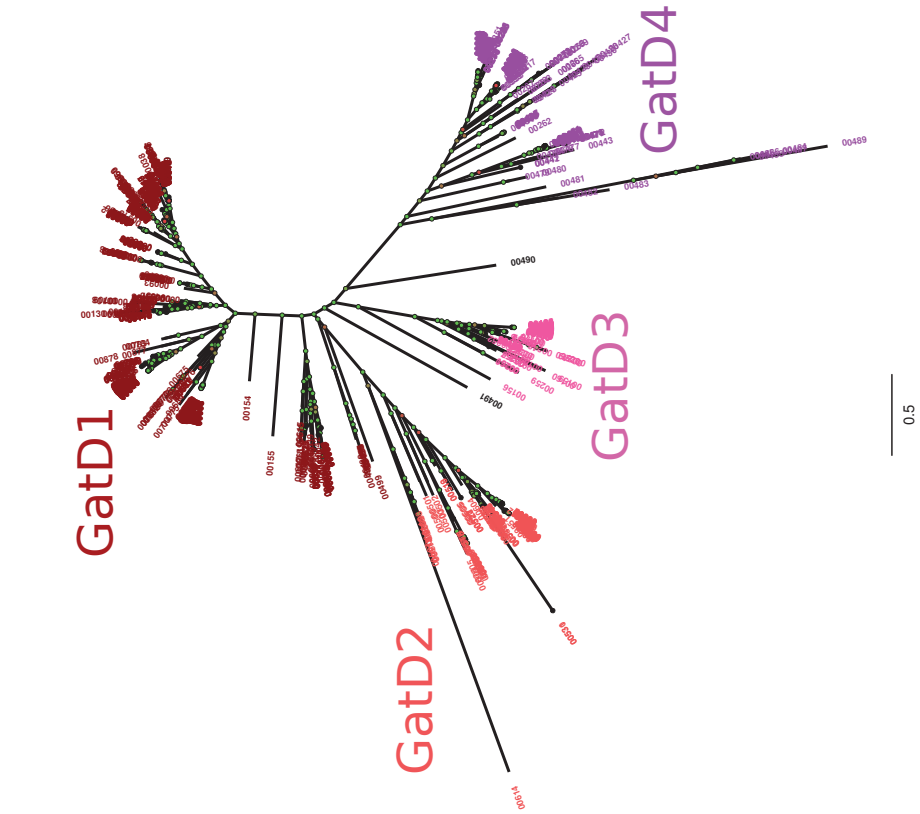

FtsK1
FtsK2
0.6

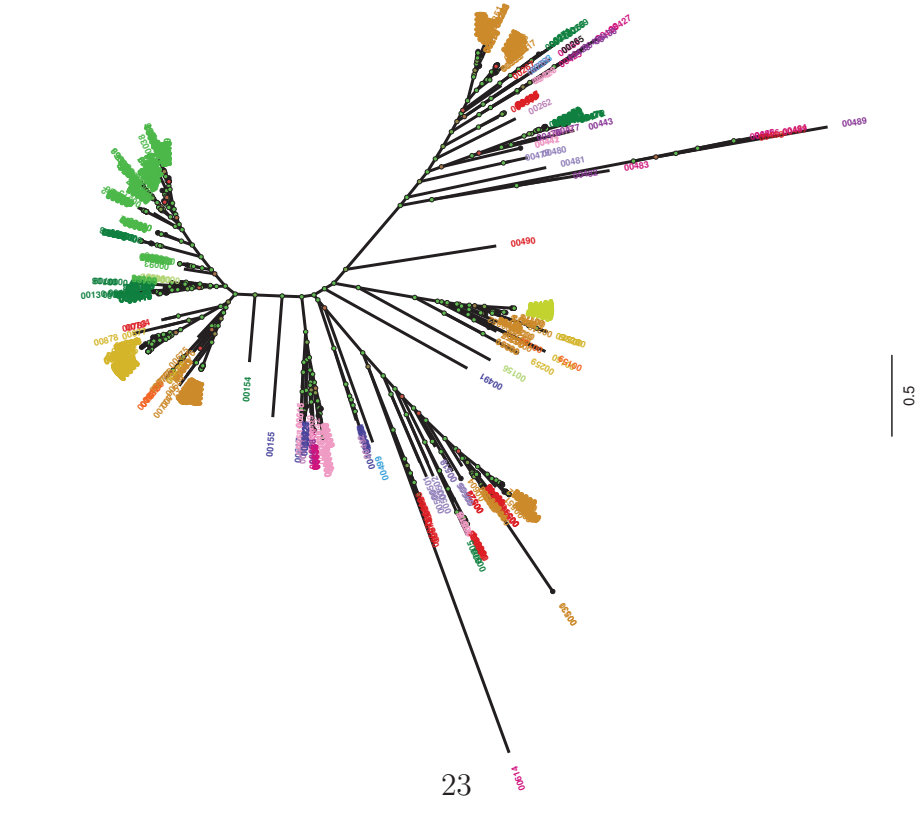

20
0.4
FtsW
SpoVE
RodA2
RodA
0.4

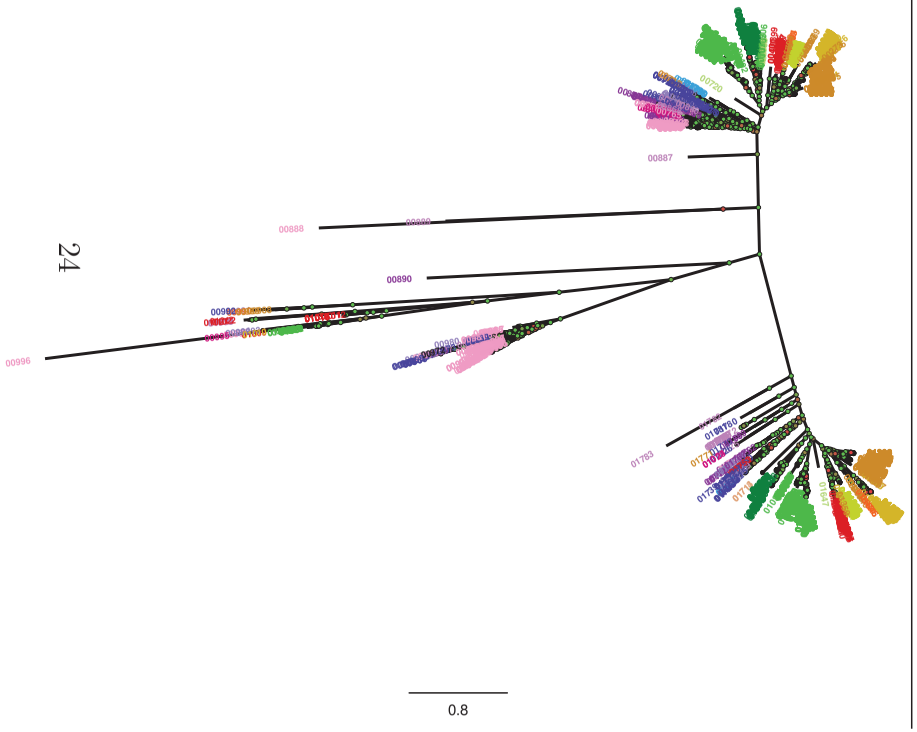

FtsX
0.6


0.6

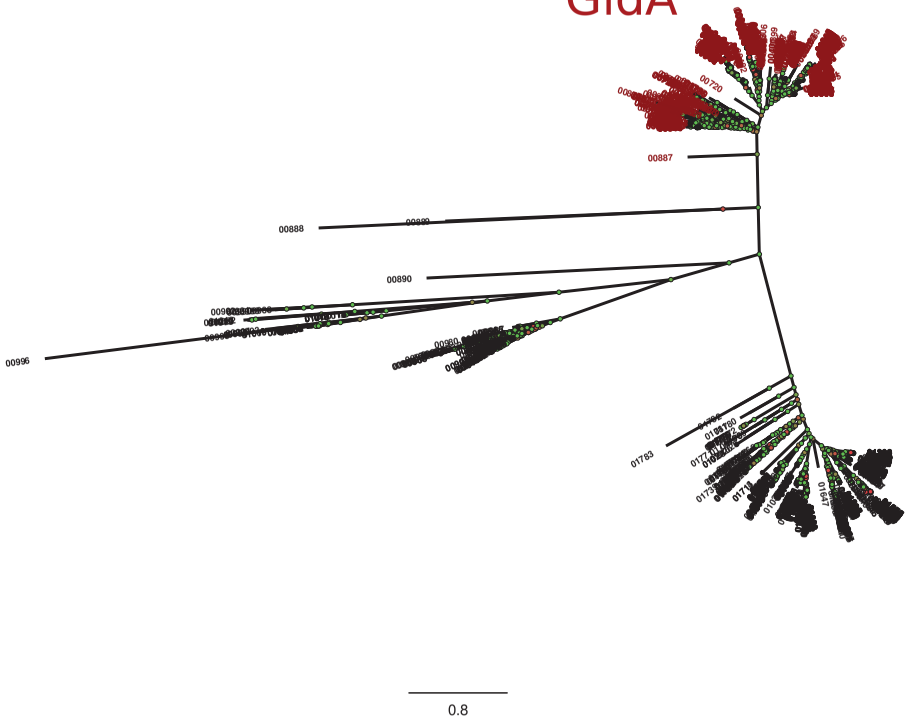

22
0.4
FtsY
0.4

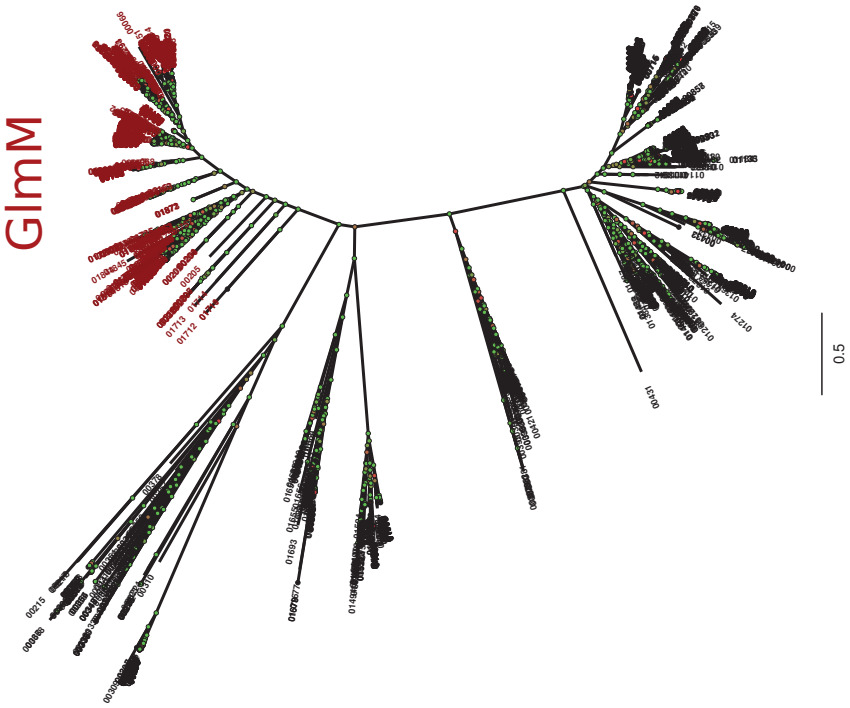

GatD1
GatD2
GatD3
GatD4
0.5

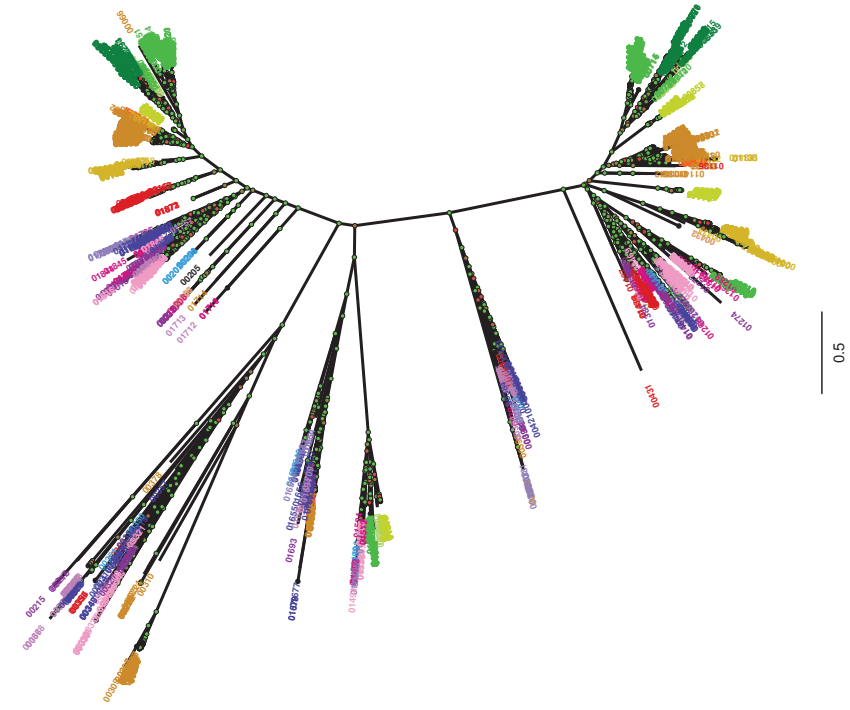

0.5

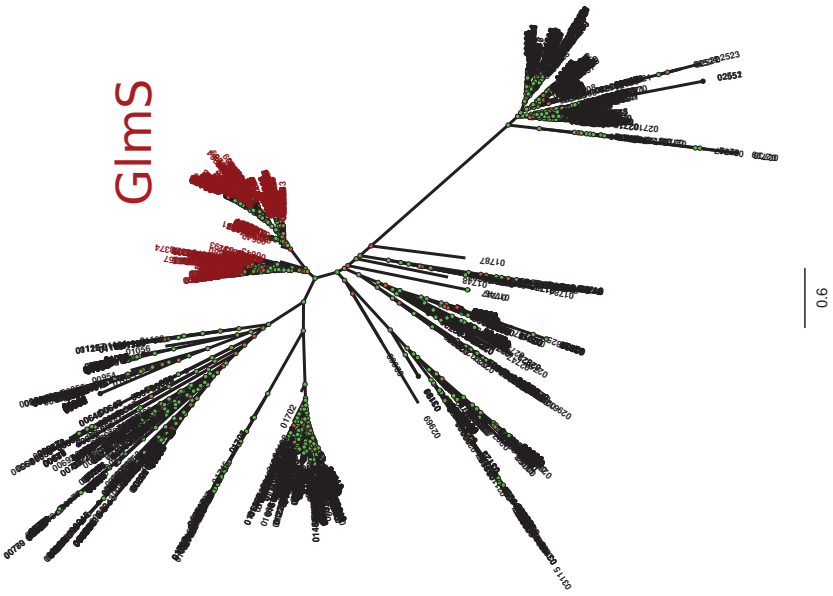

00887
00888
00890
00996
01763
0.8

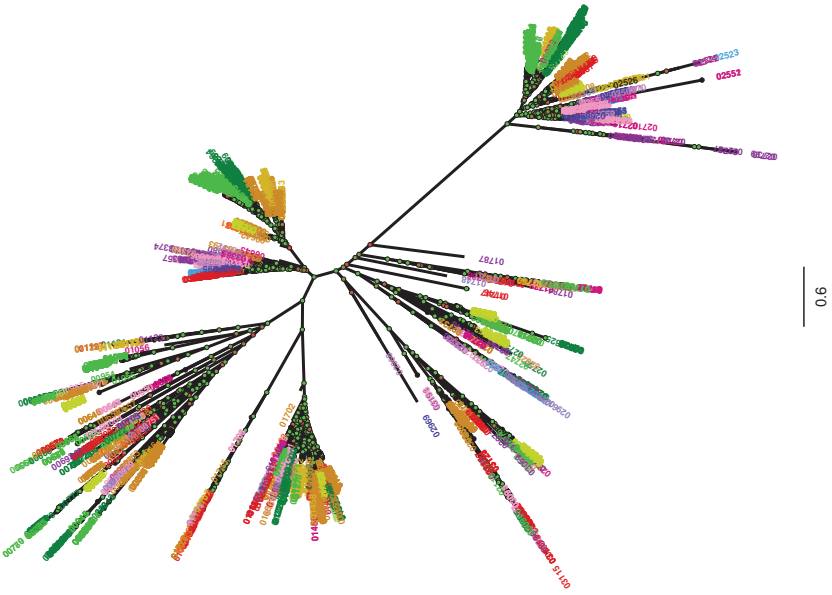
GidA

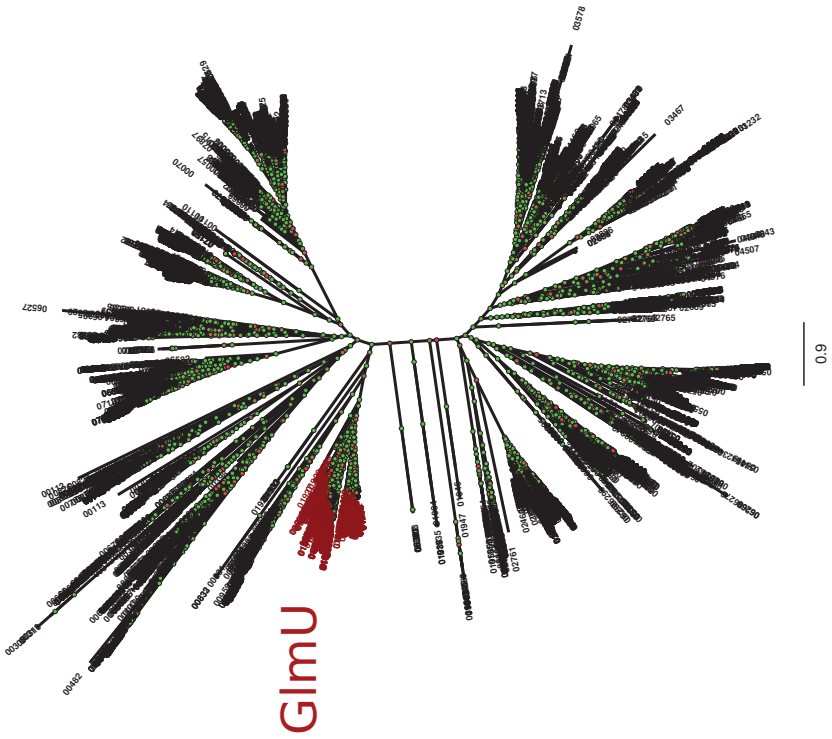

00887
00888
00890
00996
01763
0.8

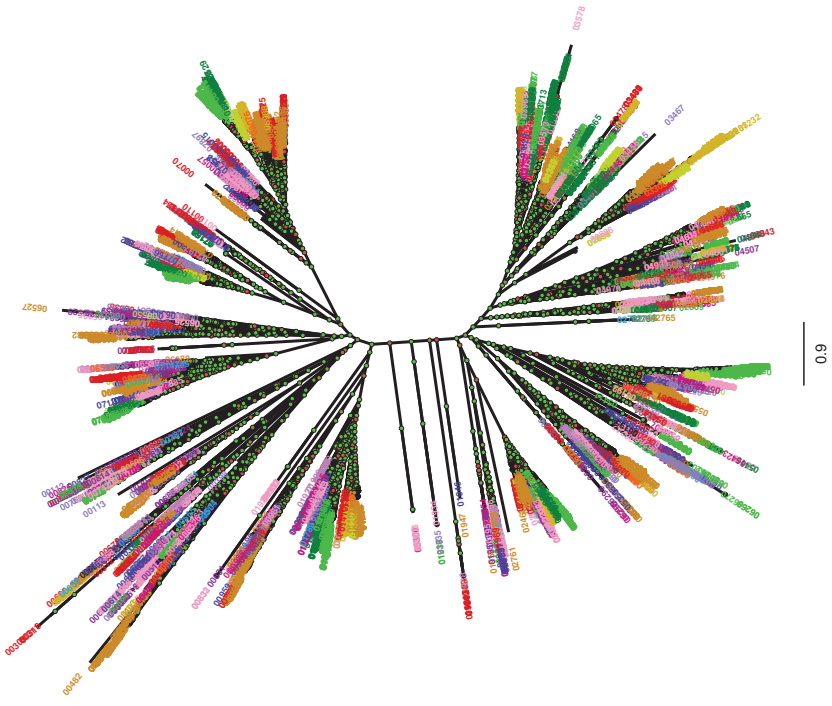

GlmM
0.5

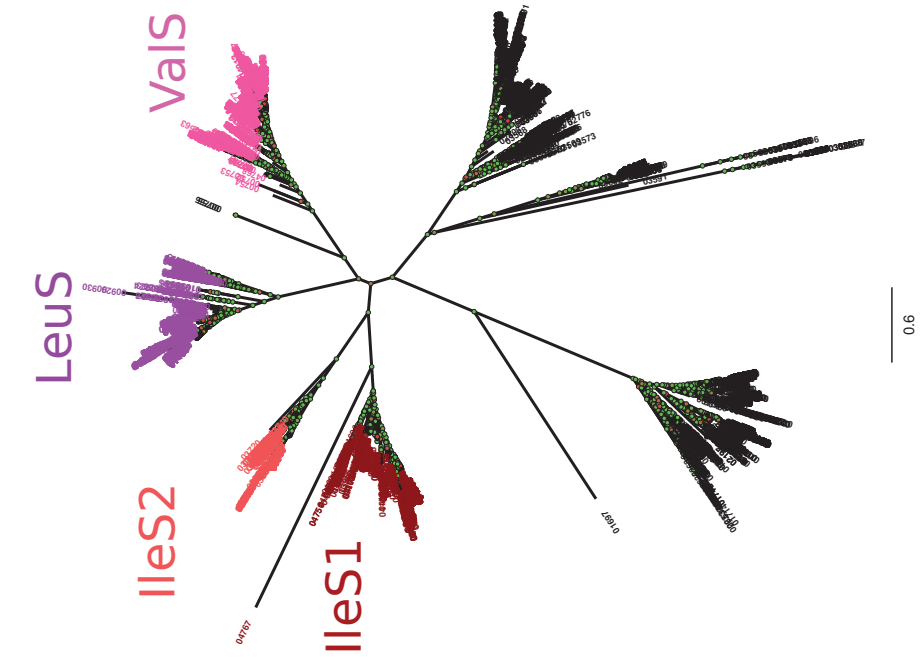

0.5

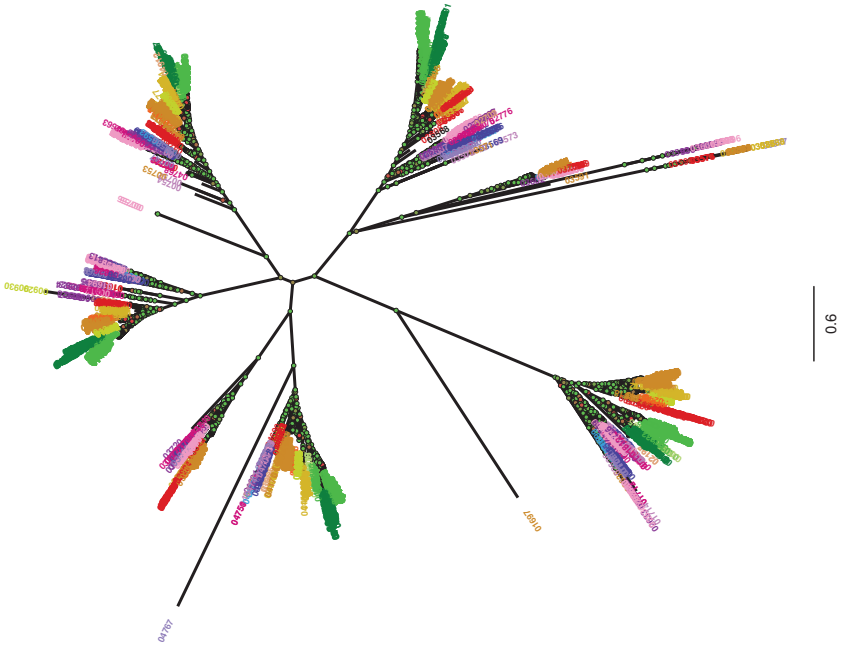

GlmS
0.6

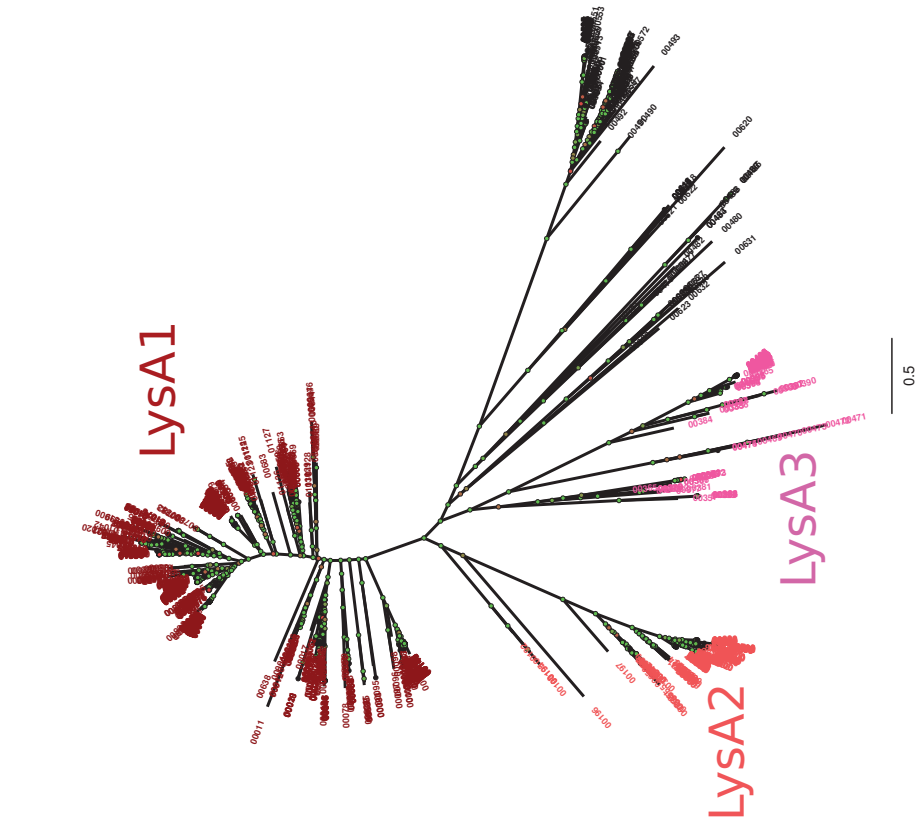

0.6


GlmU
0.9

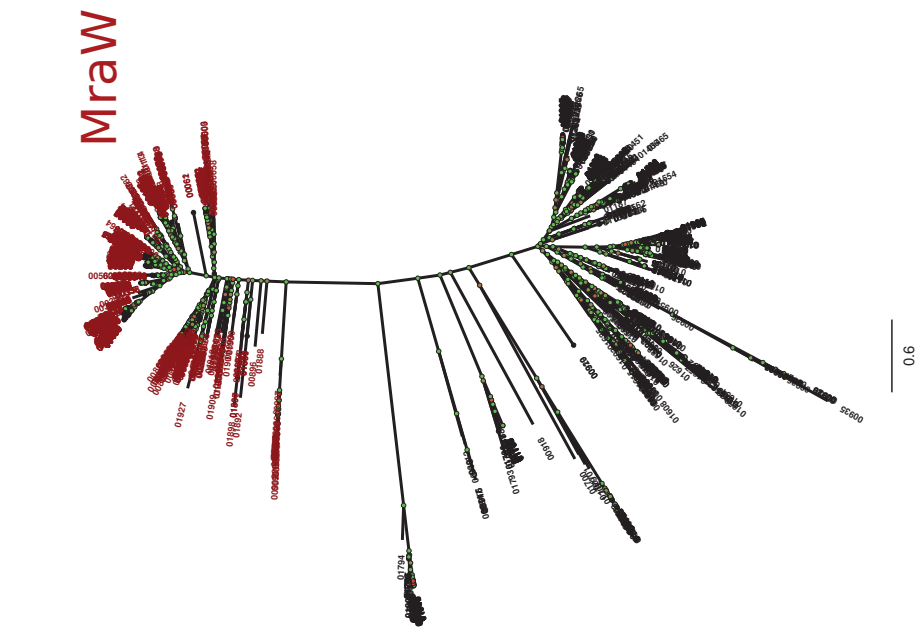

0.9


ValS
LeuS
IleS2
IleS1
0.6

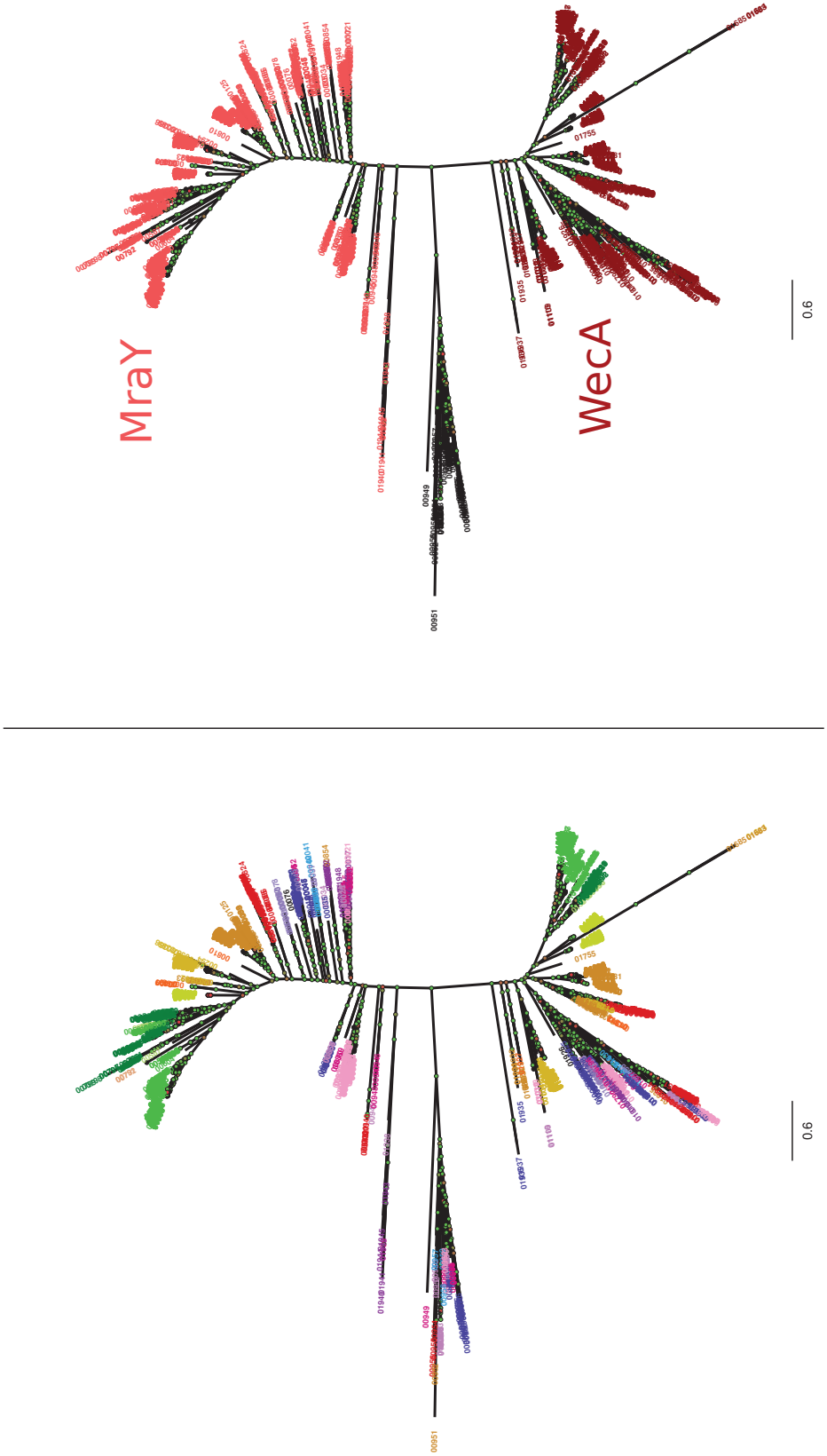

0.6

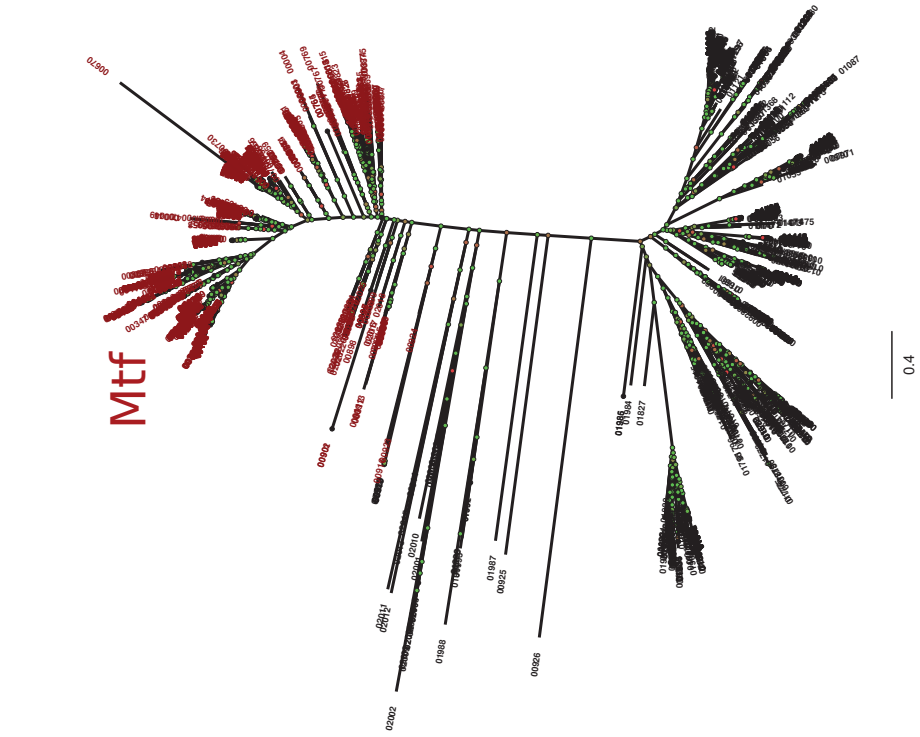

LysA1
LysA2
LysA3
0.5

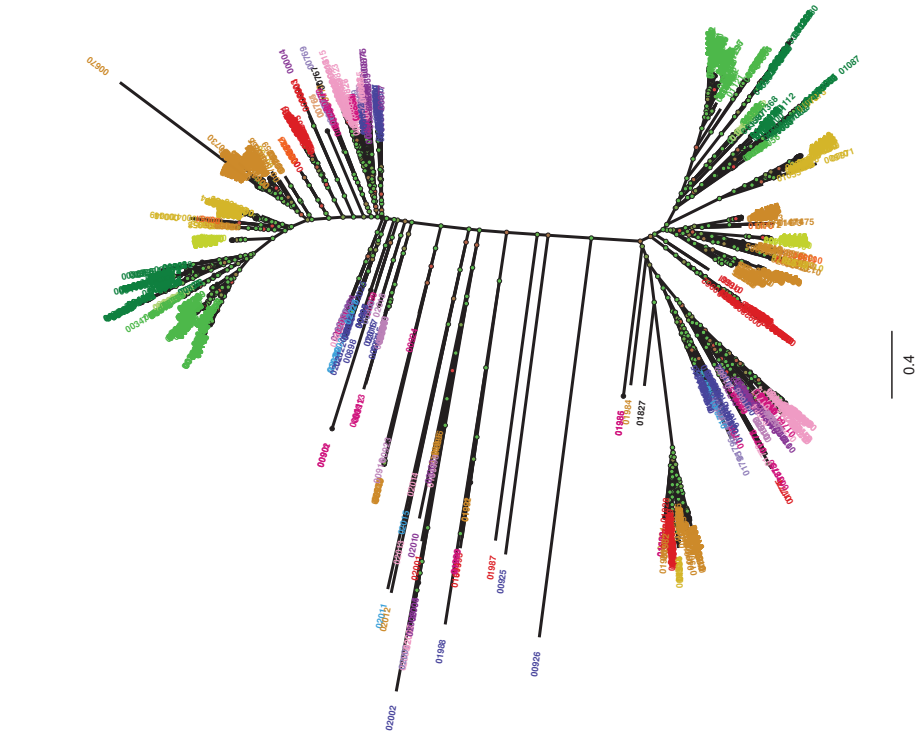
0.5

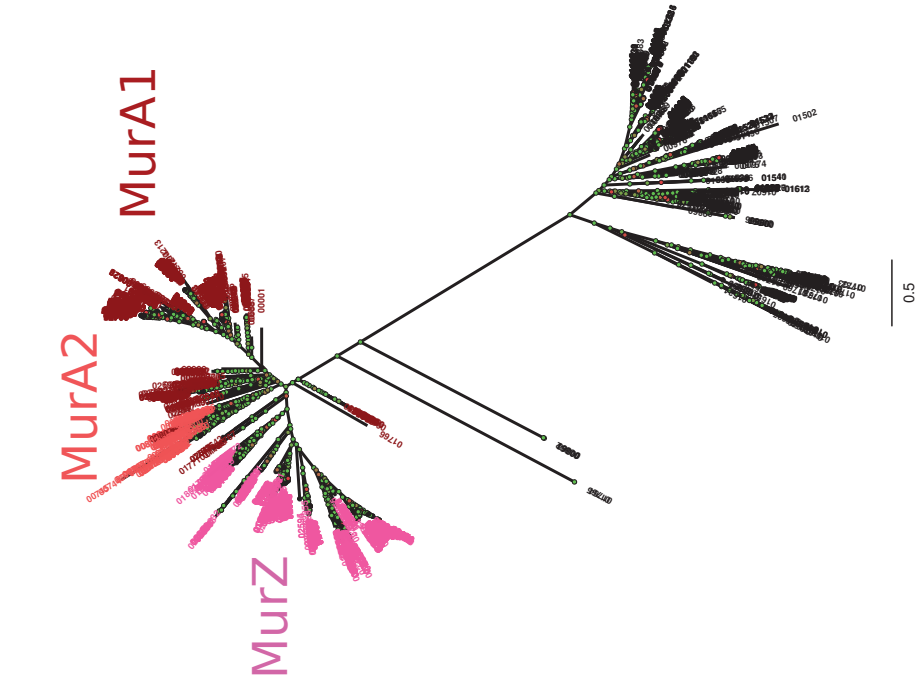

MraW
0.6

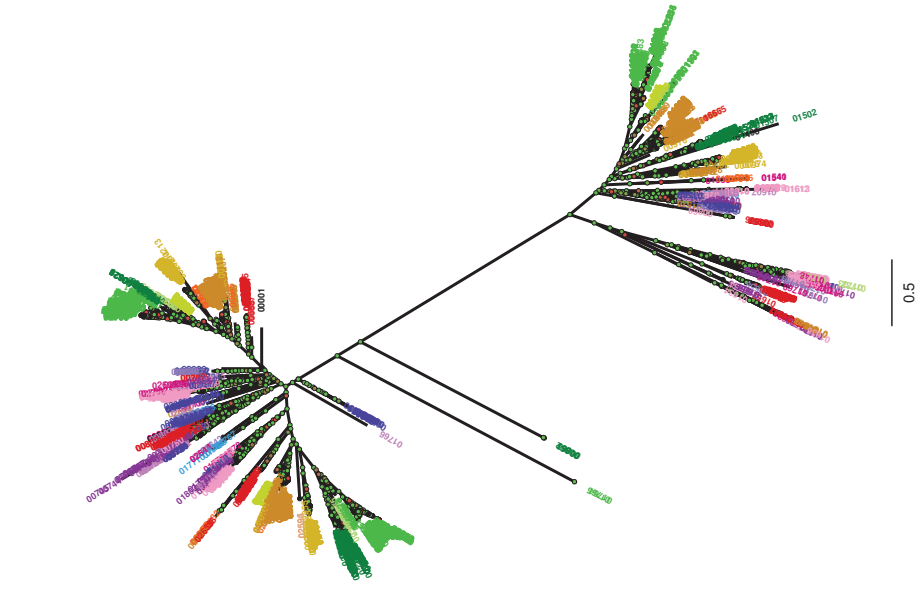
0.6

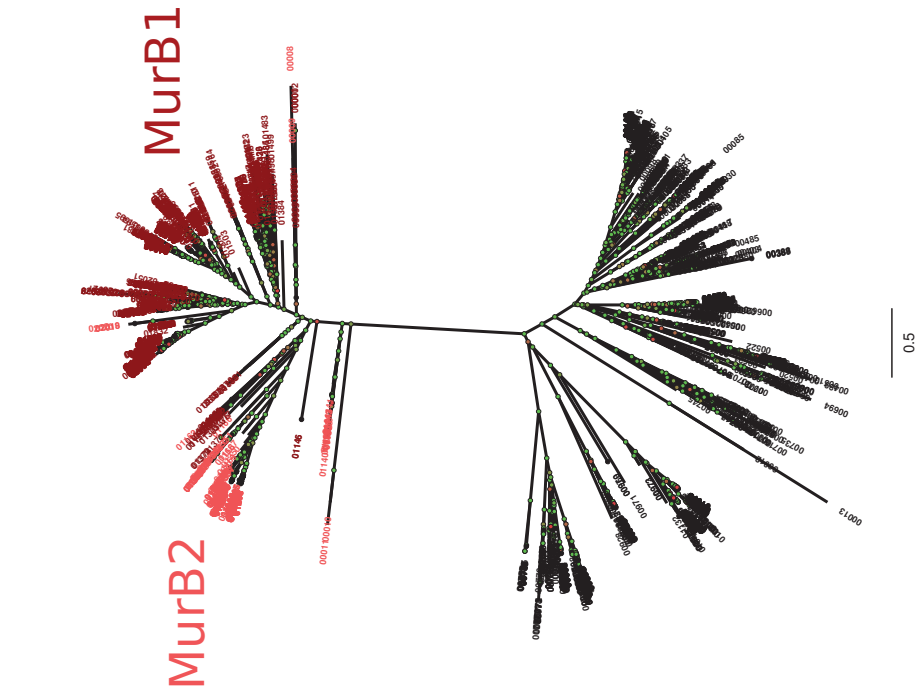

MraY
WecA
0.6
0.6

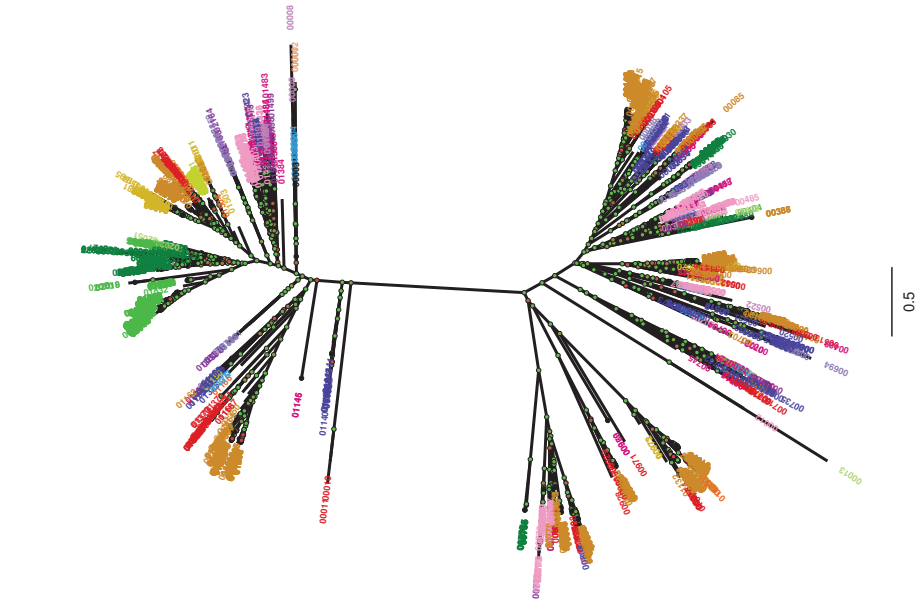

Mtf
0.4

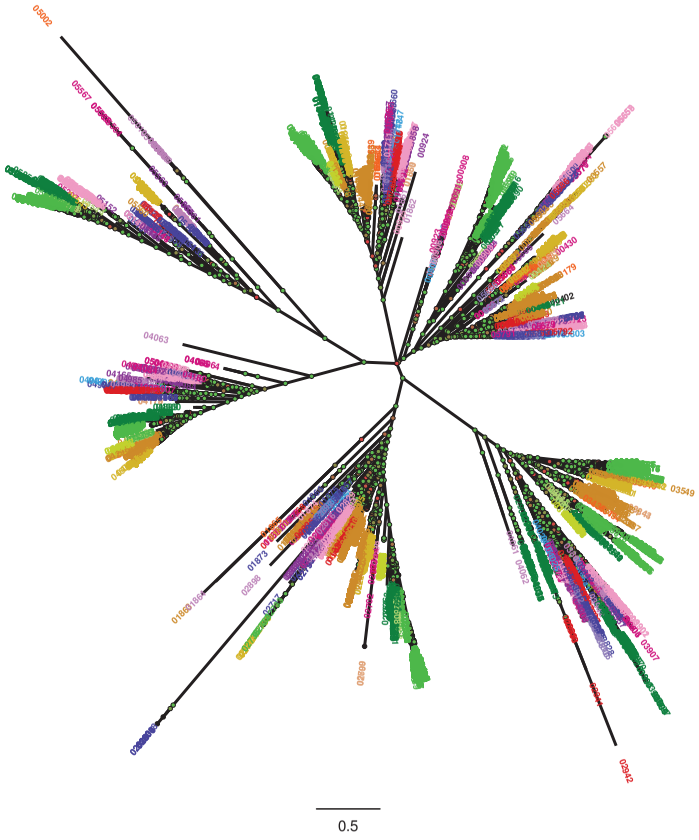

0.4

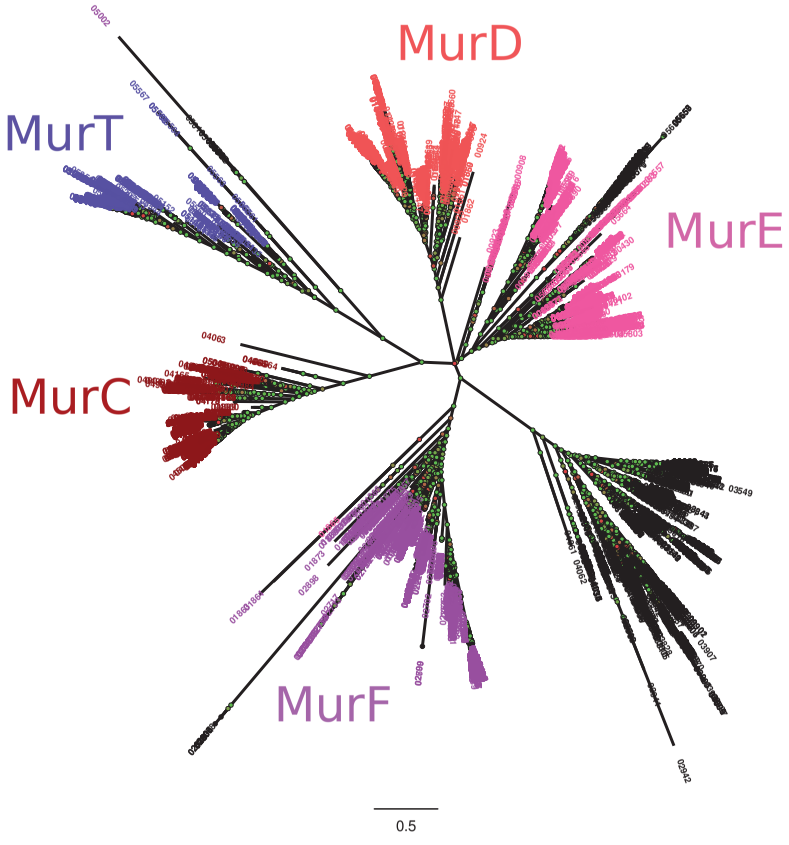

MurA1
MurA2
MurZ
0.5

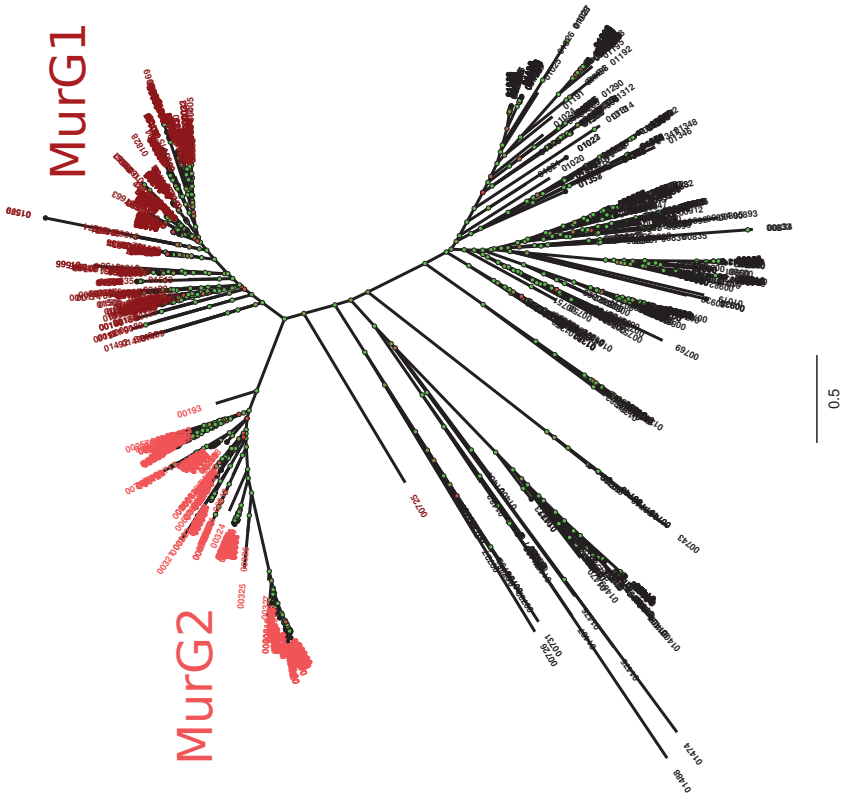

0.5

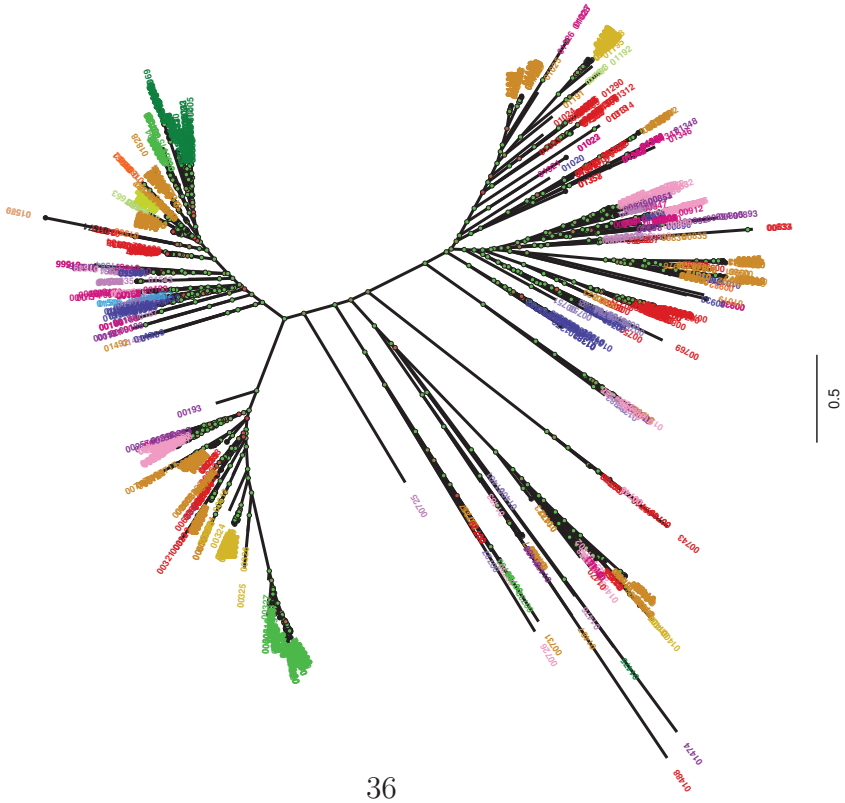

MurB1
MurB2
0.5

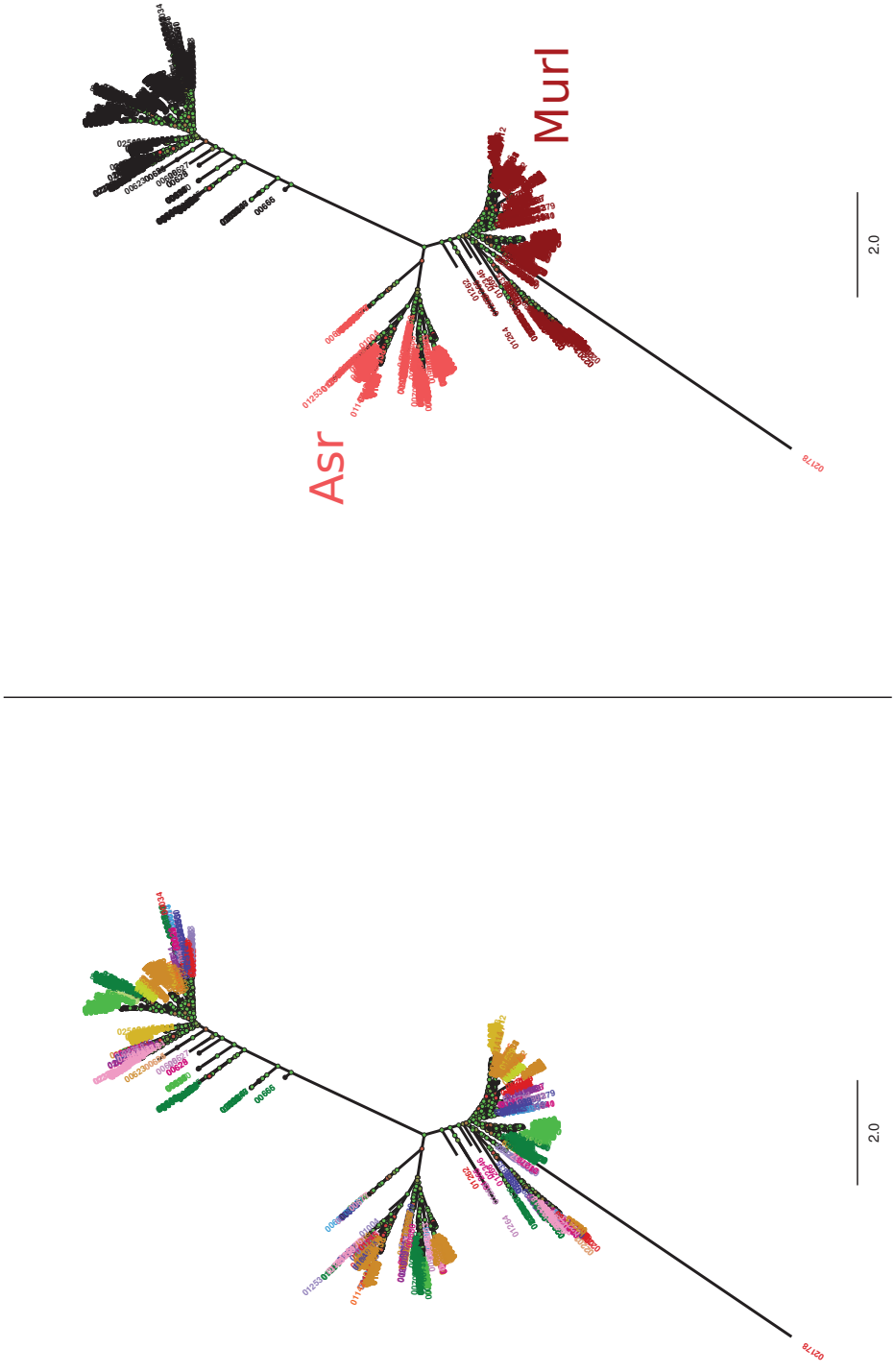

0.5


0.5

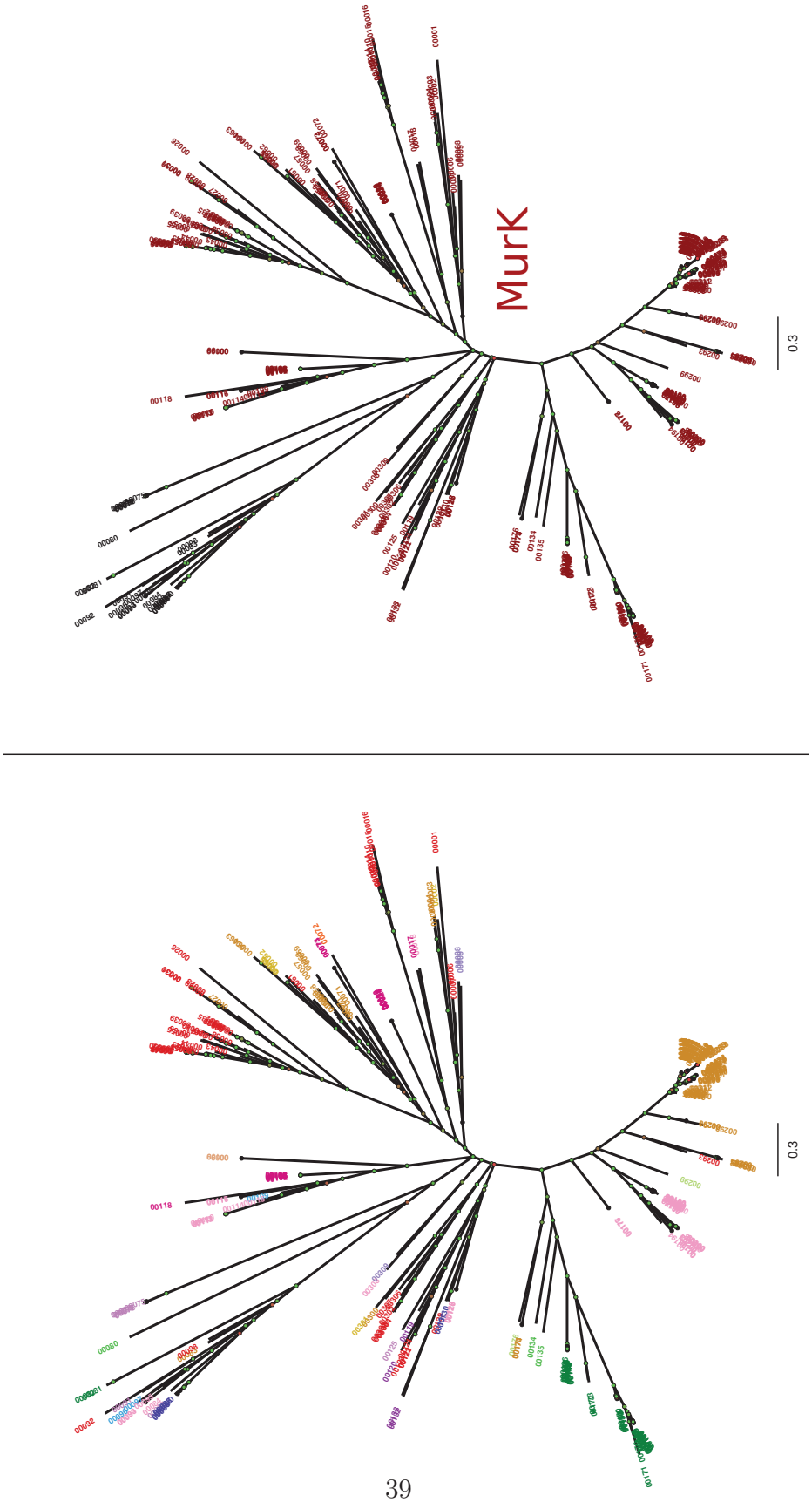

MurD
MurT
MurE
MurC
MurF
0.5

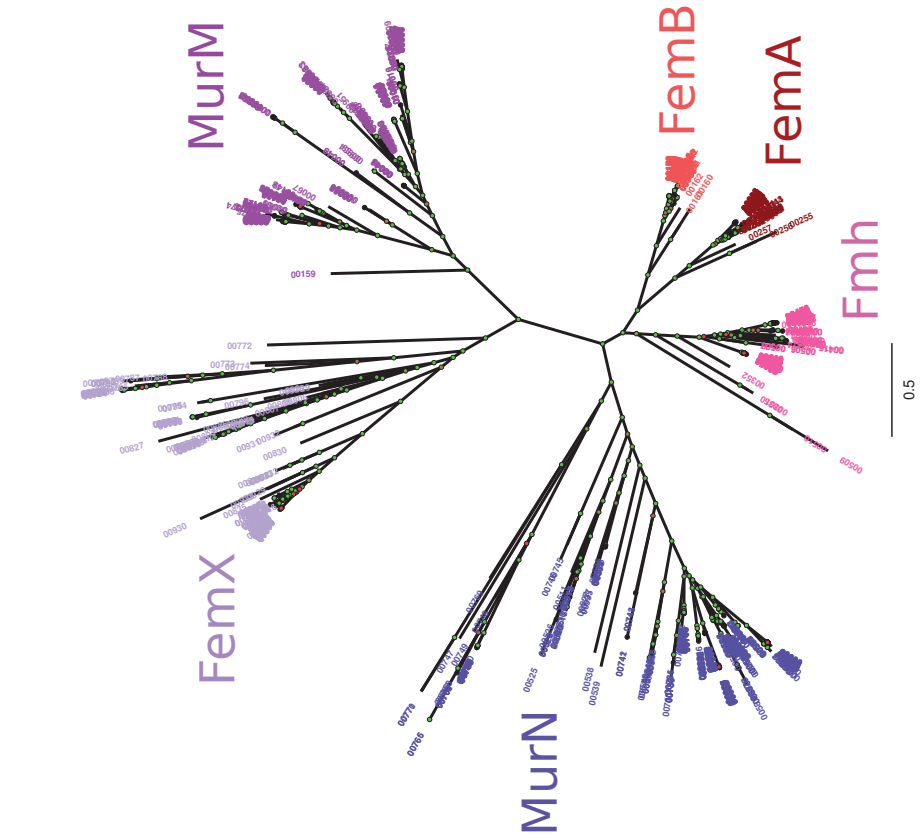

MurG1
MurG2
0.5

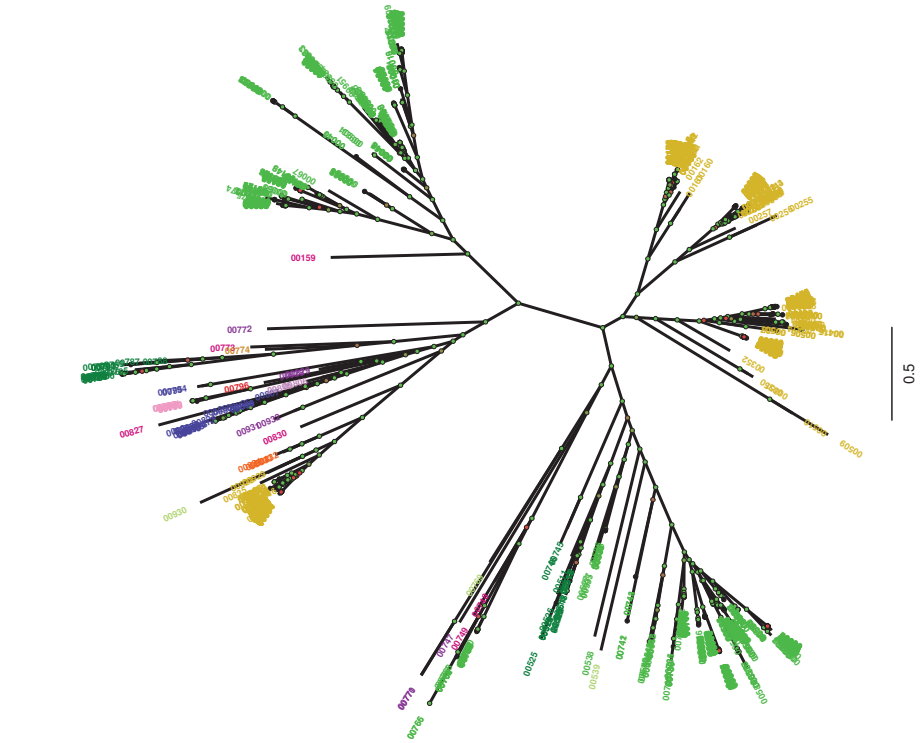

0.5

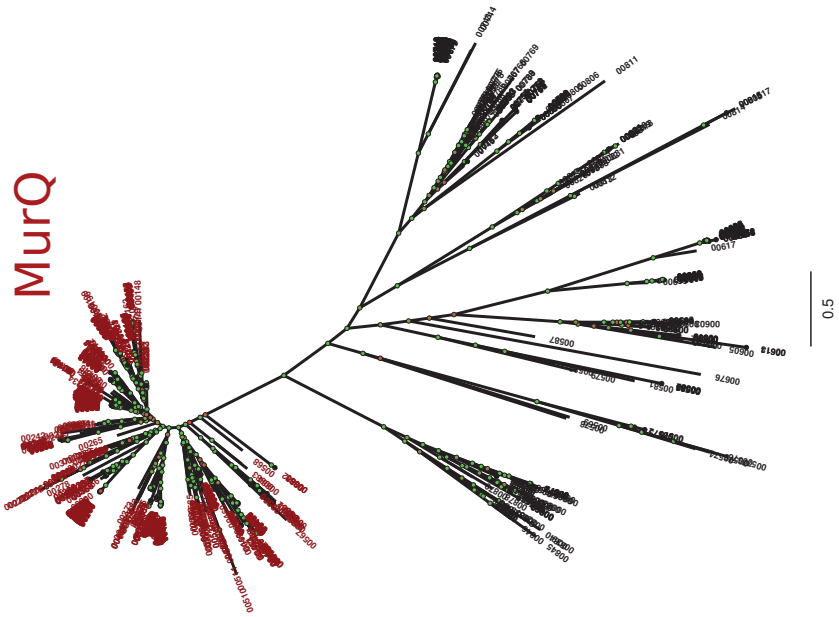

MurI
Asr
2.0
2.0

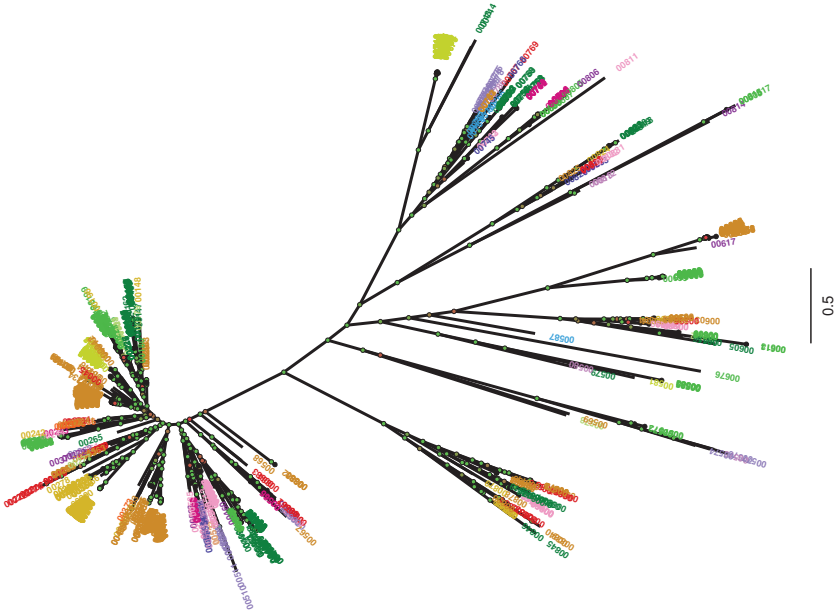
YkvU
SpoVB
MurJ
YabM
0.6
0.6

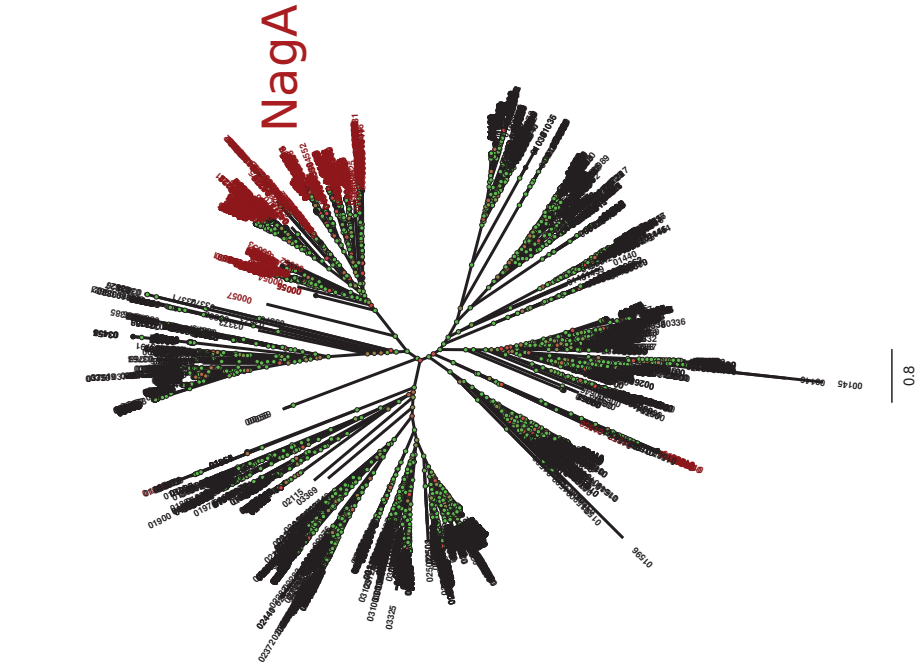

MurK
0.3
MurK
0.3

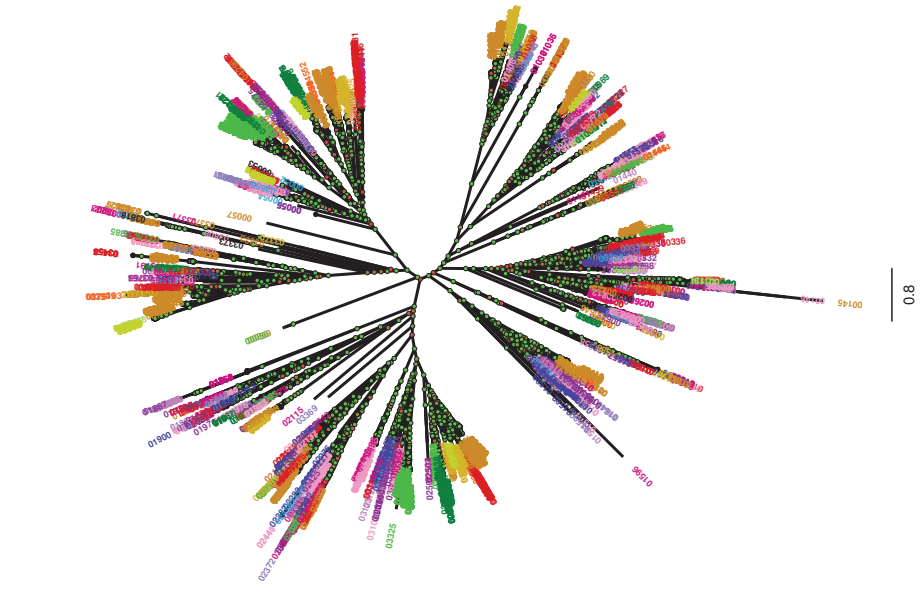

MurM
FemB
FemA
Fmh
FemX
MurN
0.5

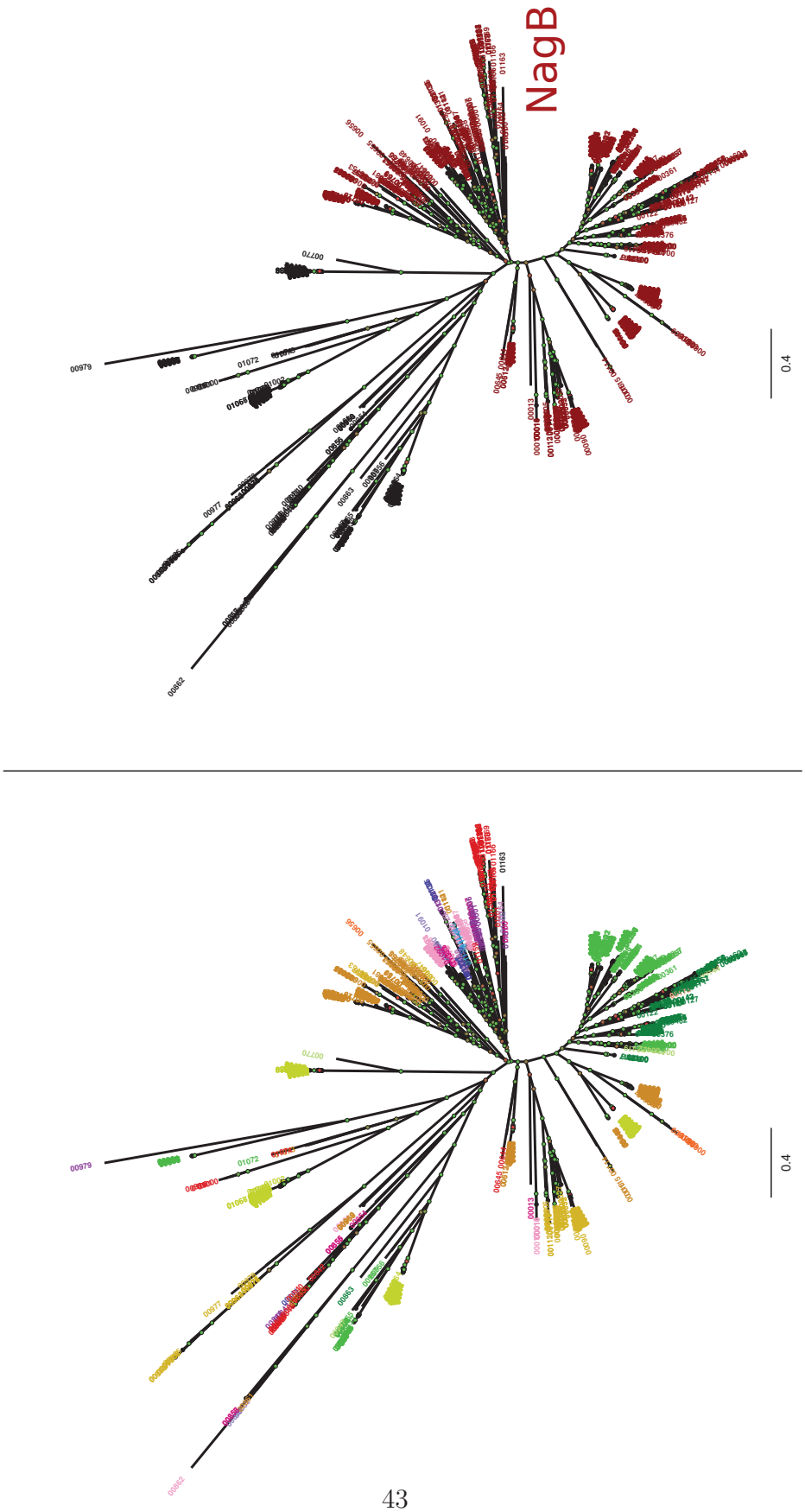

0.5

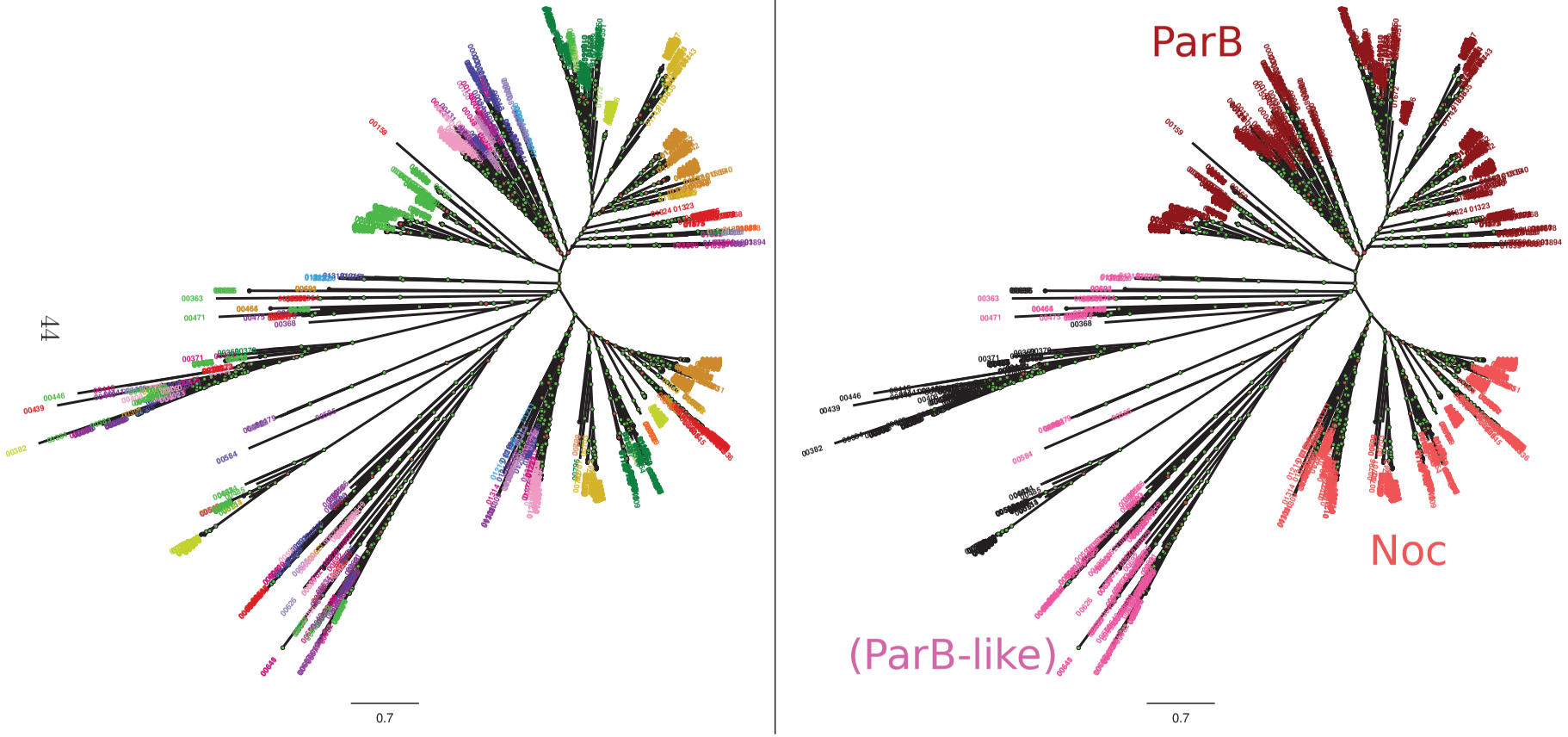

MurQ
0.5

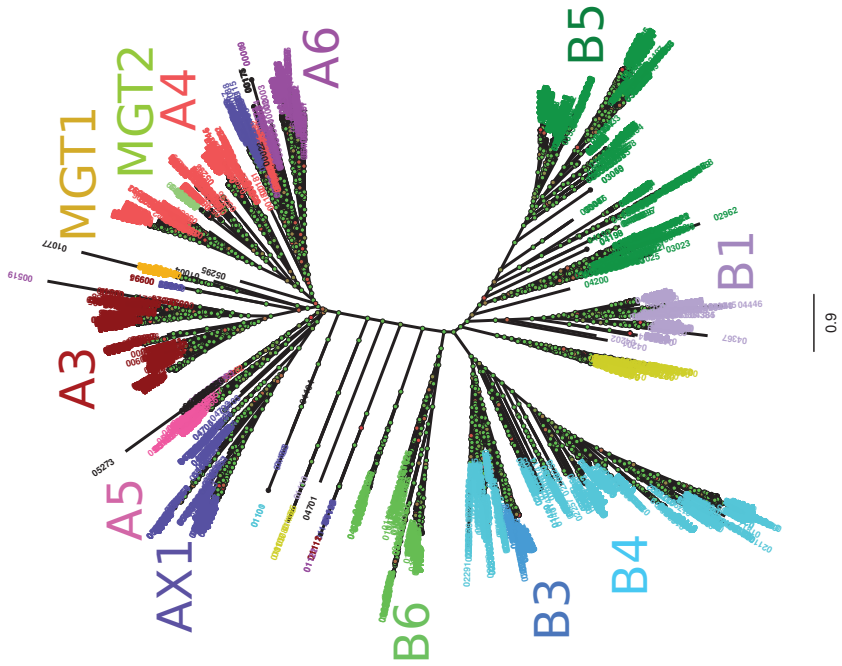

0.5

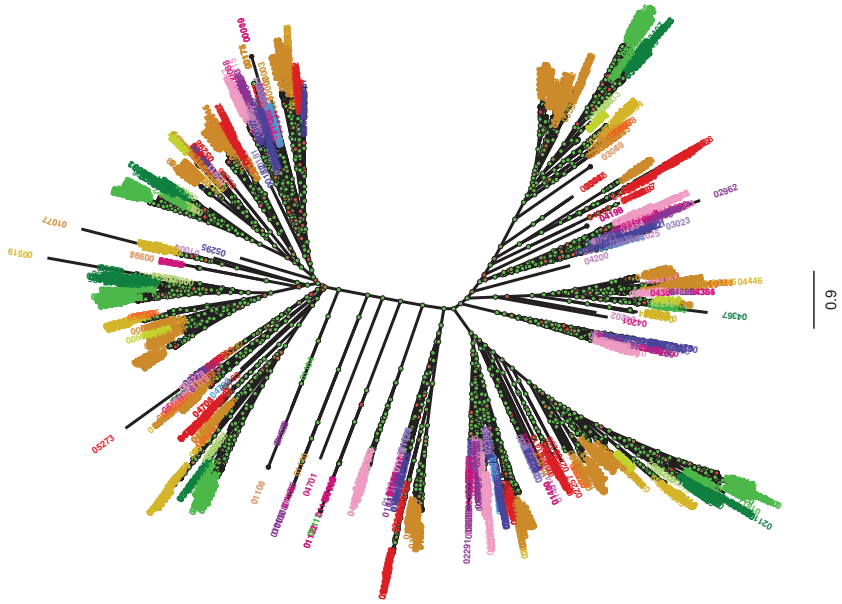

NagA
0.8

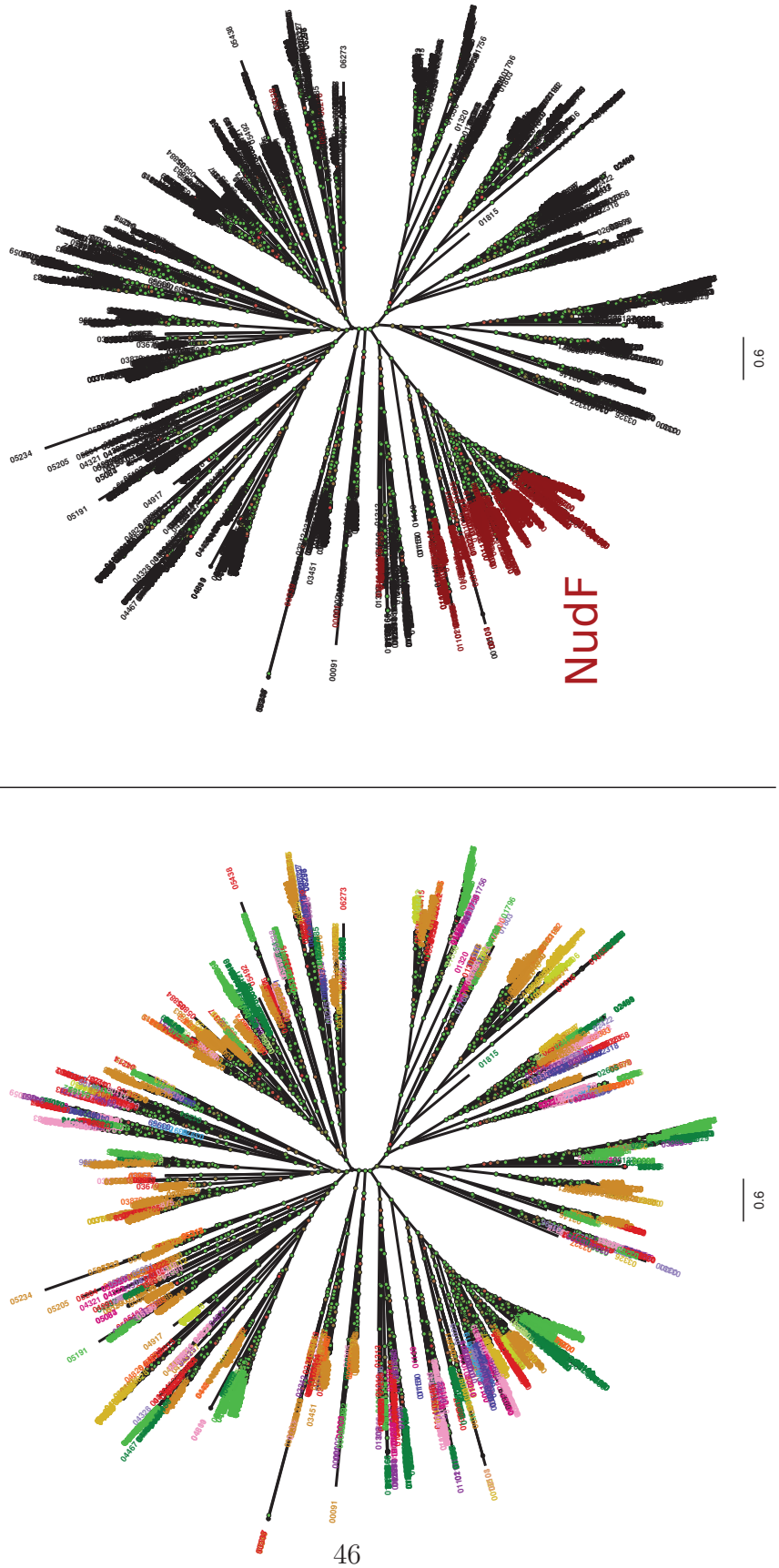

0.8

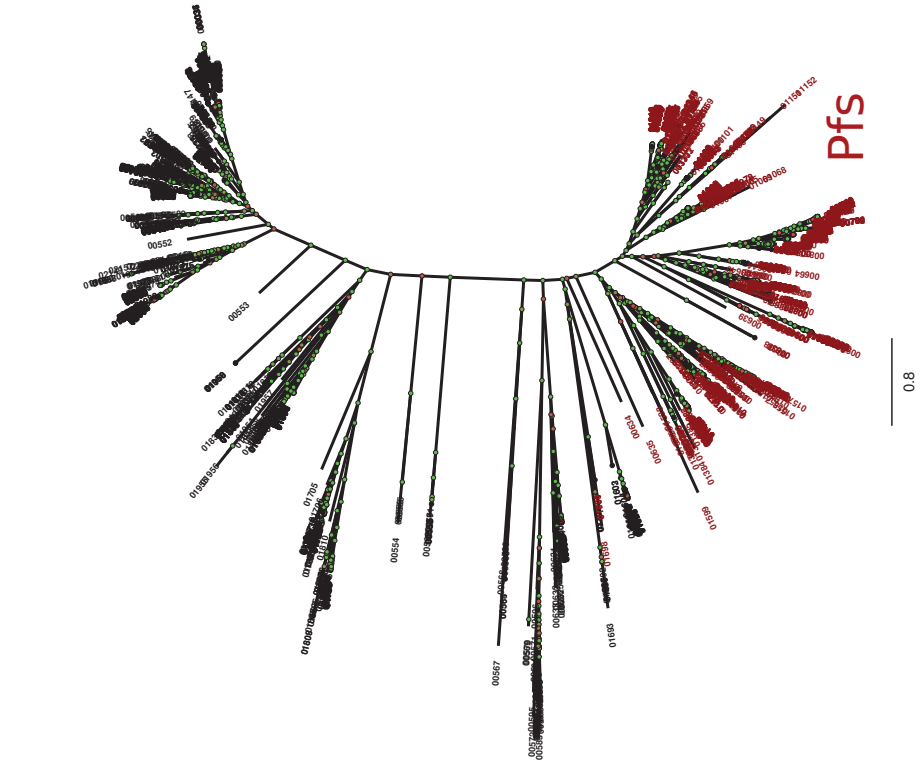

NagB
0.4
0.4

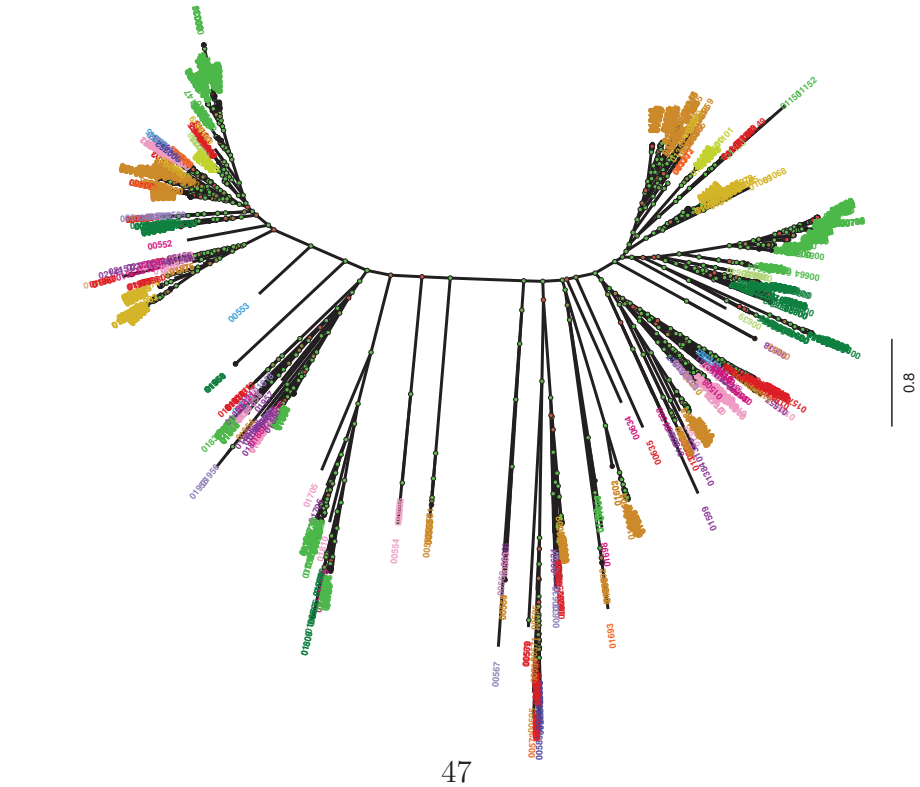

44
0.7
ParB
Noc
(ParB-like)
0.7

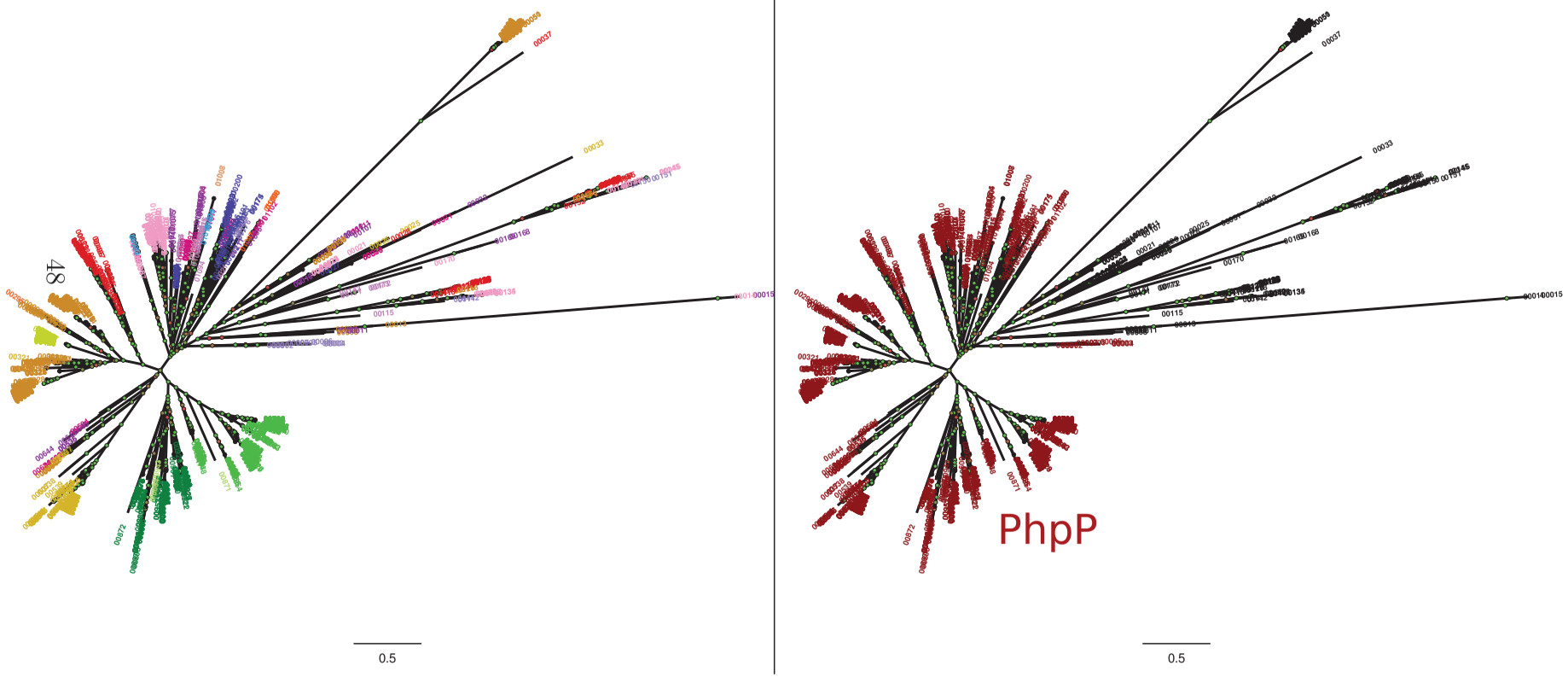

A6
A4
MGT2
MGT1
B5
B1
A3
A5
AX1
B6
B3
B4
0.9

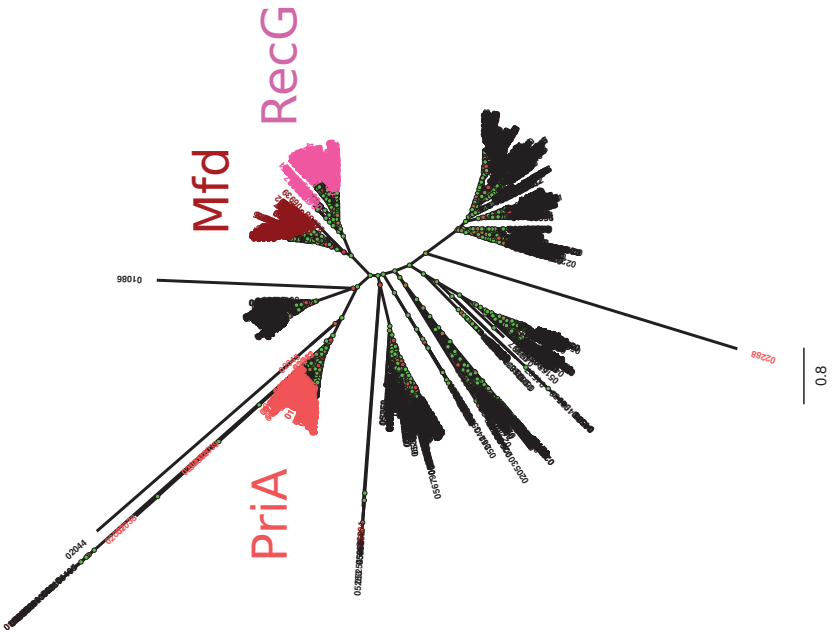

0.9

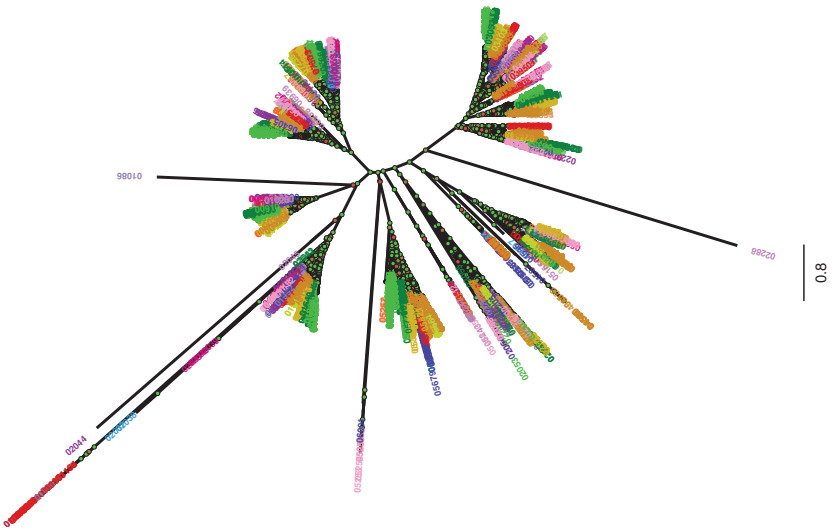

NudF
0.6
0.6

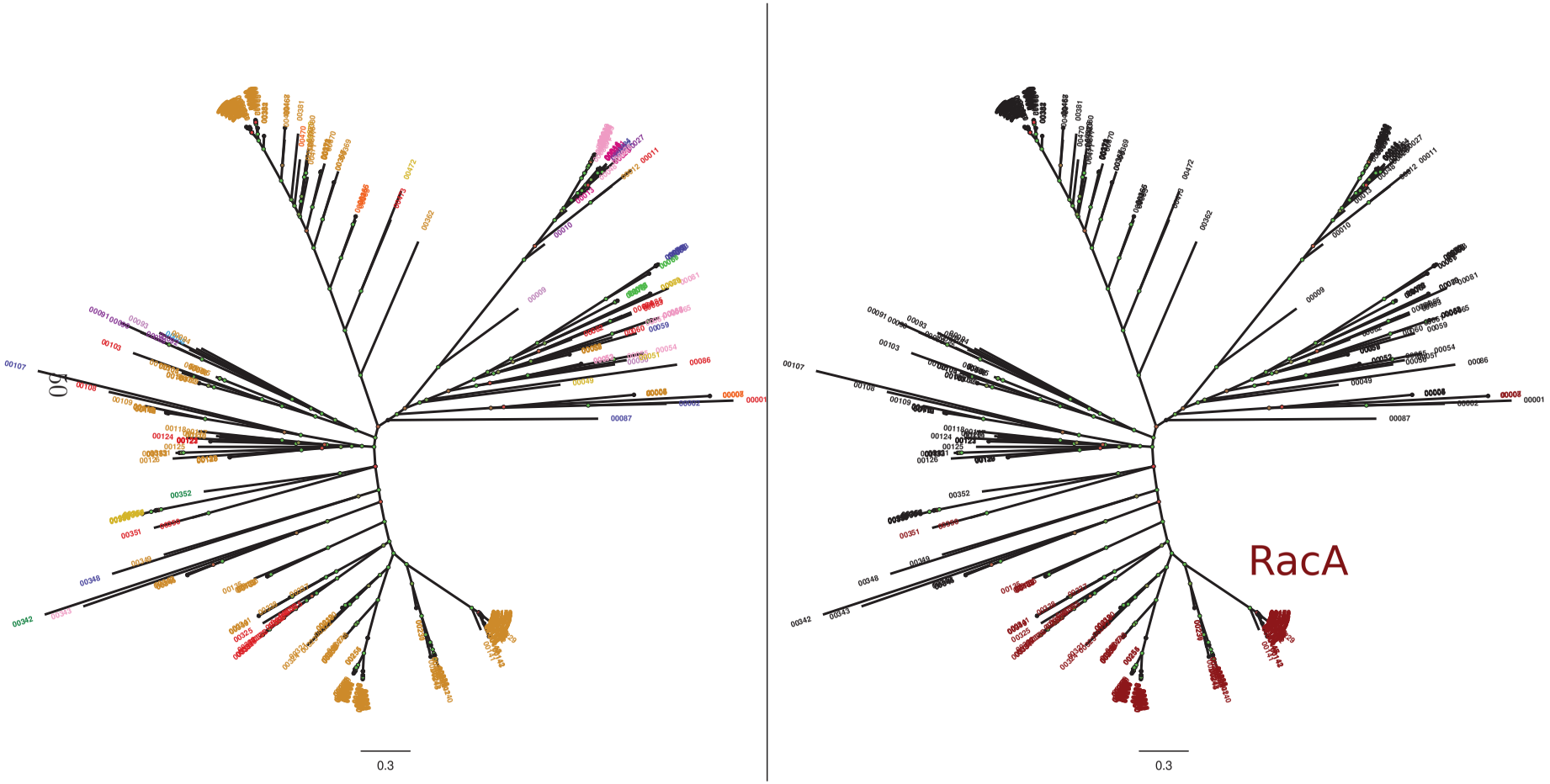

Pfs
0.8

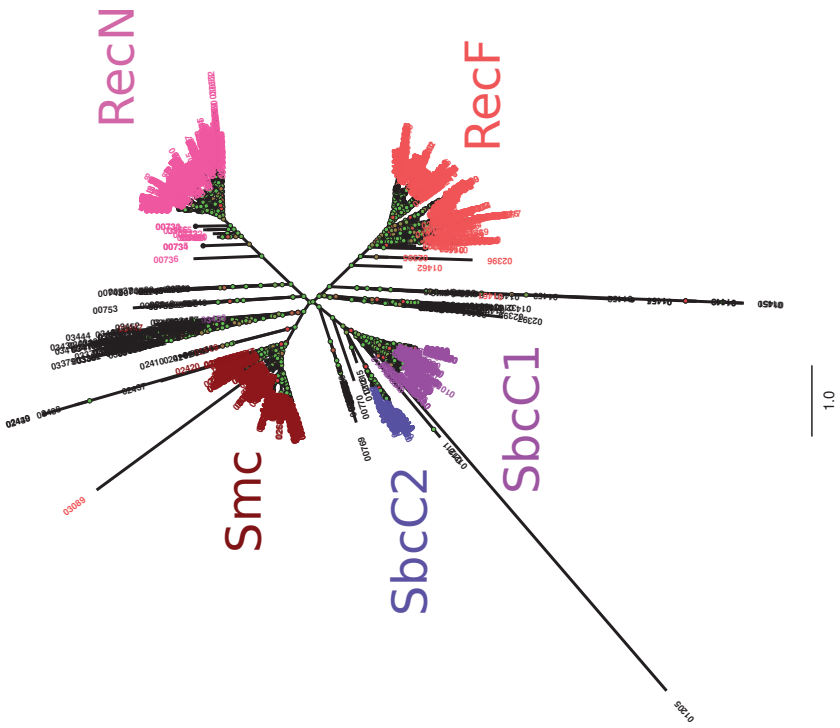

0.8

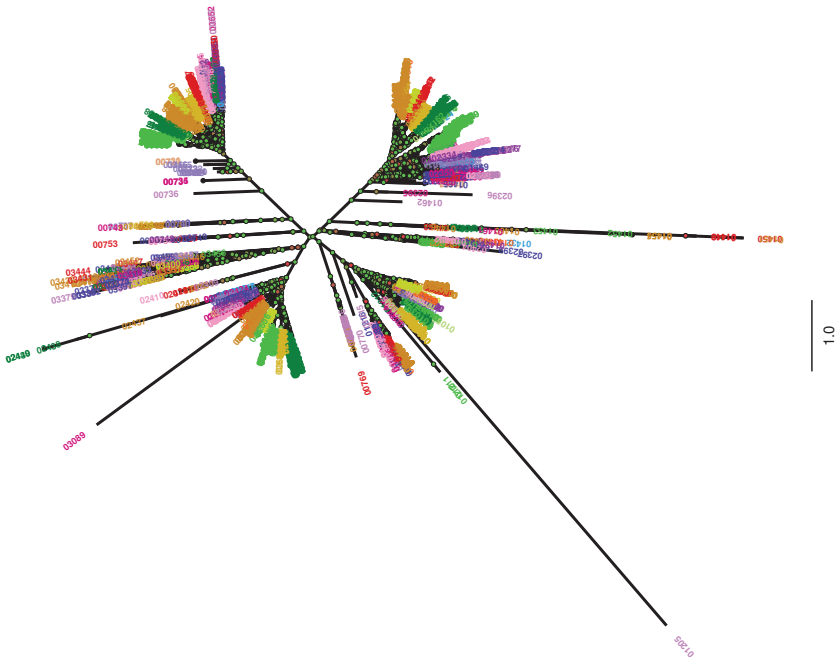

48
0.5
PhpP
0.5

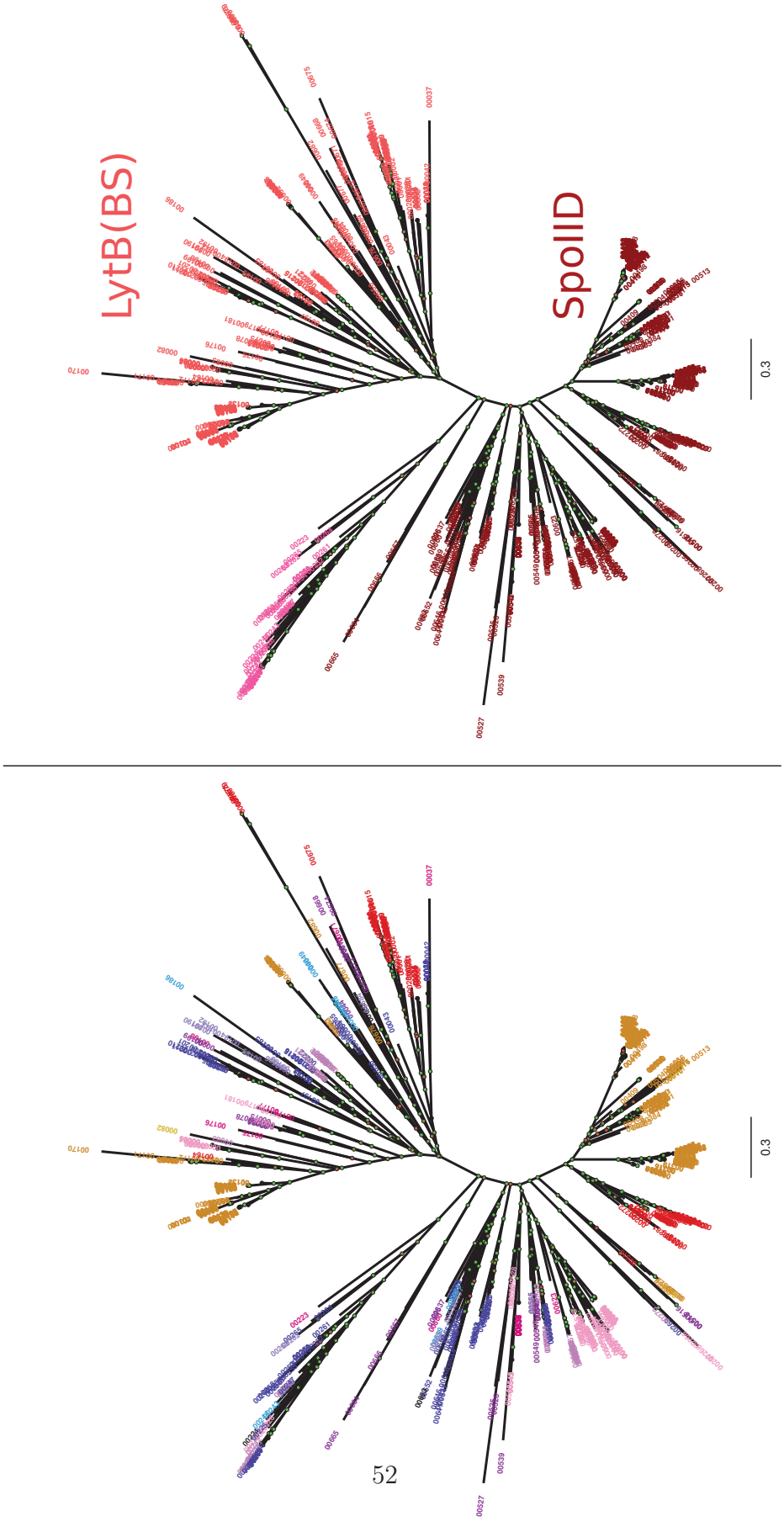

Mfd
RecG
PriA
0.8

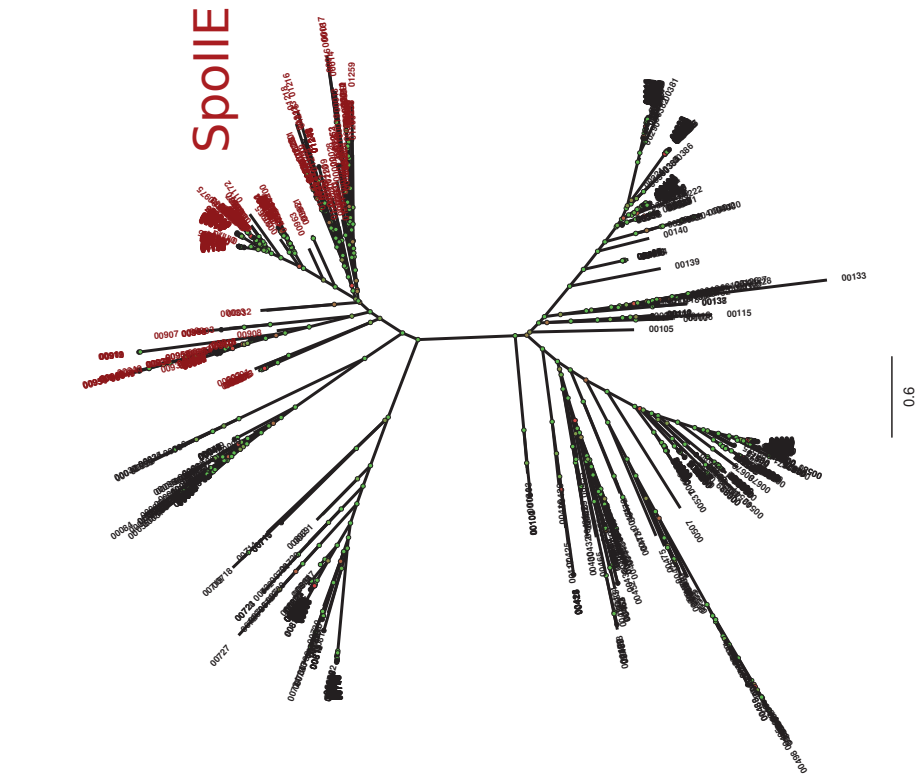

0.8

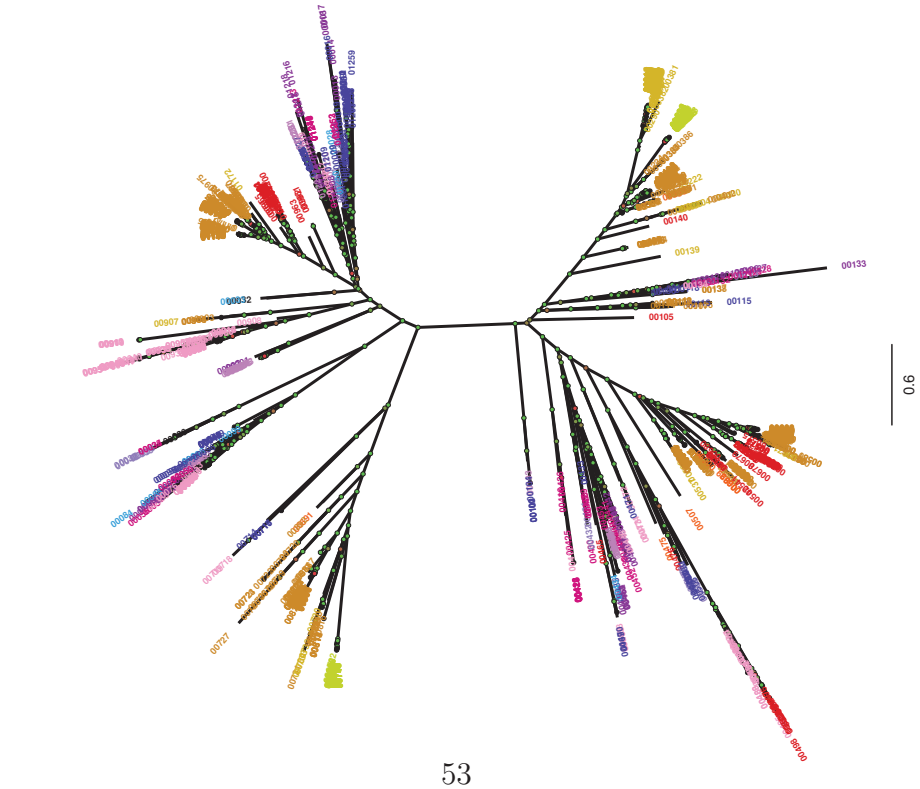

50
0.3
RacA
0.3

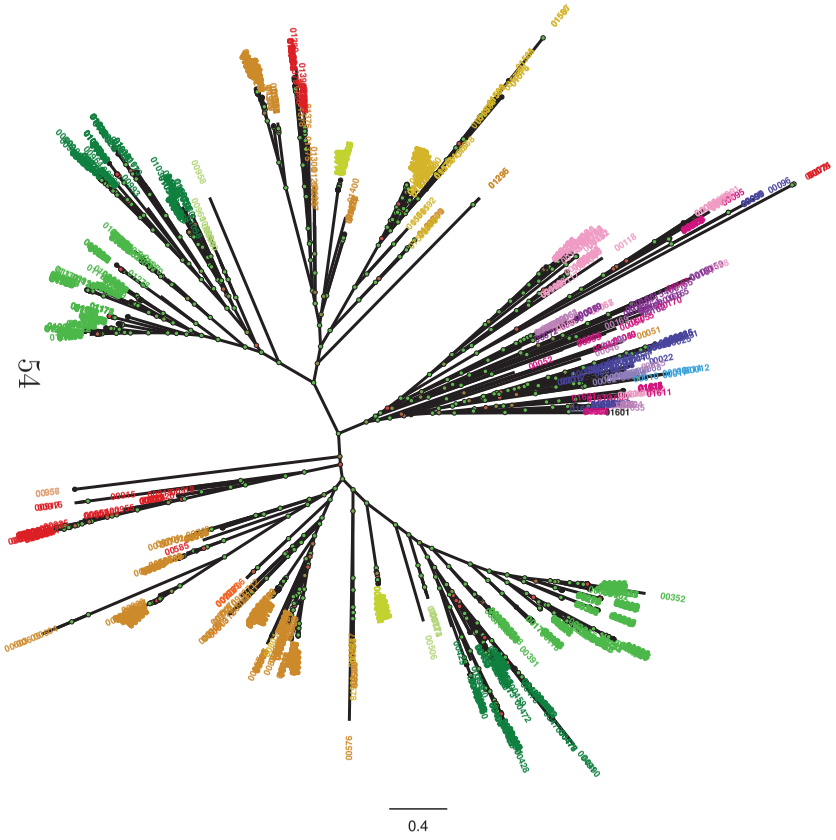

RecN
RecF
Smc
SbcC2
SbcC1
1.0

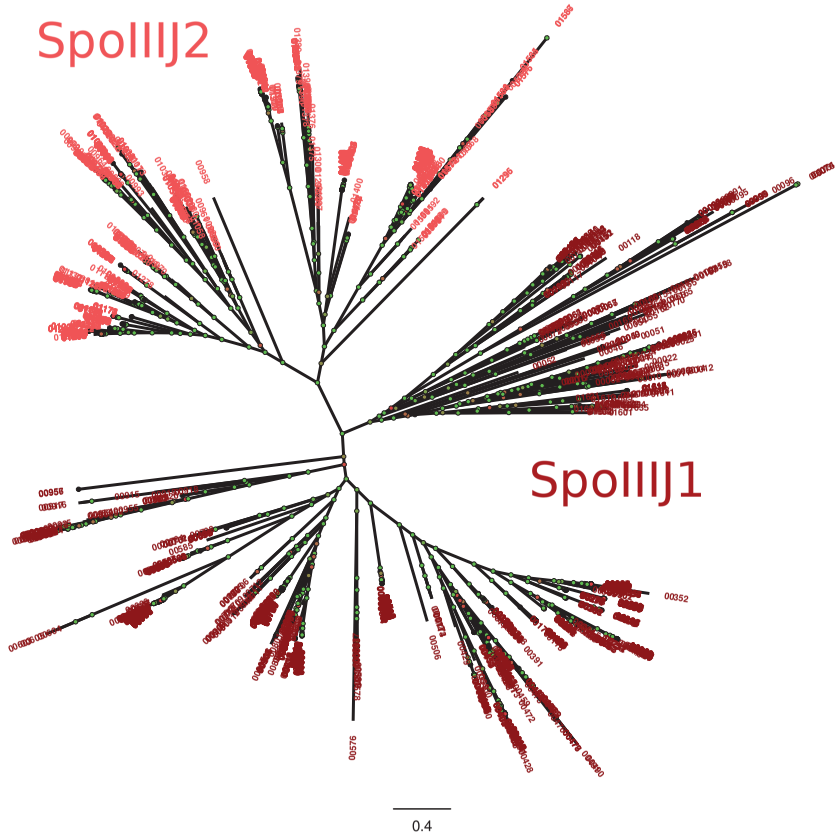

1.0

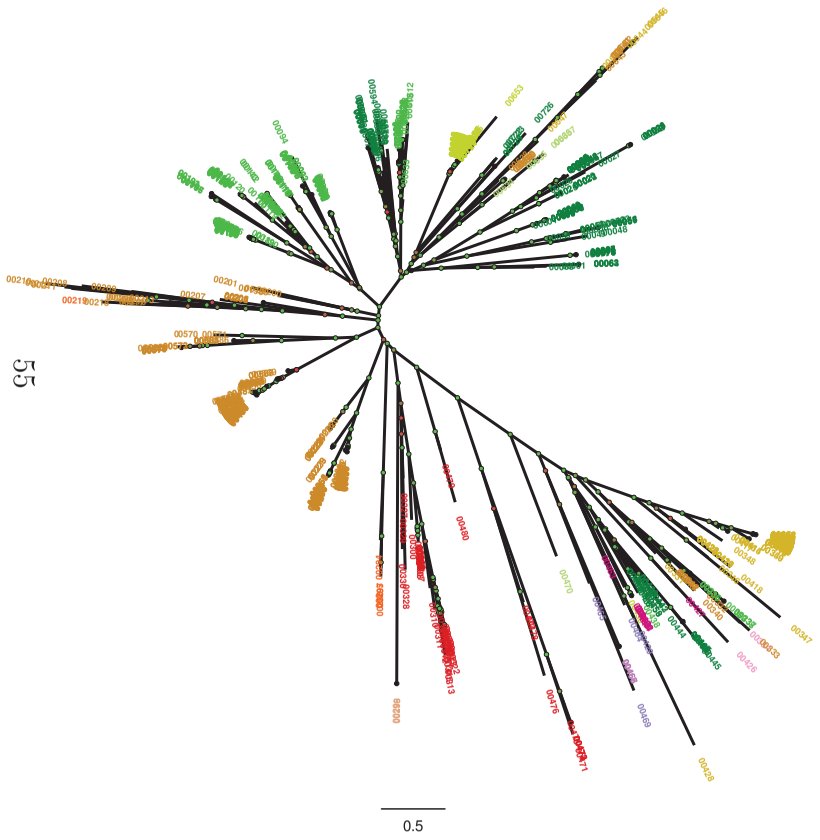

LytB(BS)
SpoIID
0.3
0.3

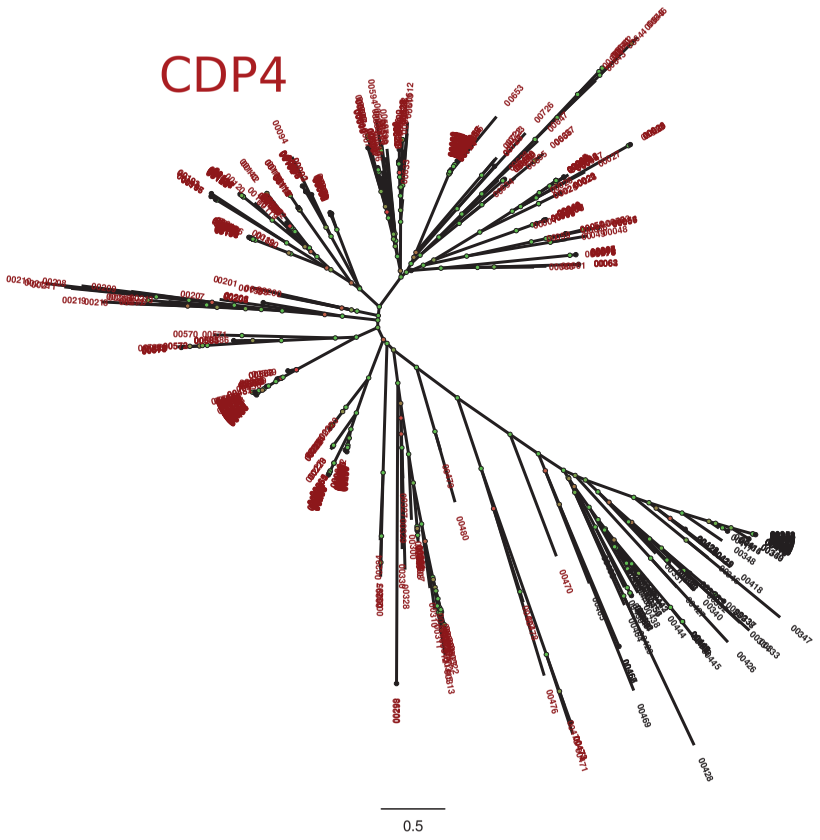

SpoIIE
0.6


0.6

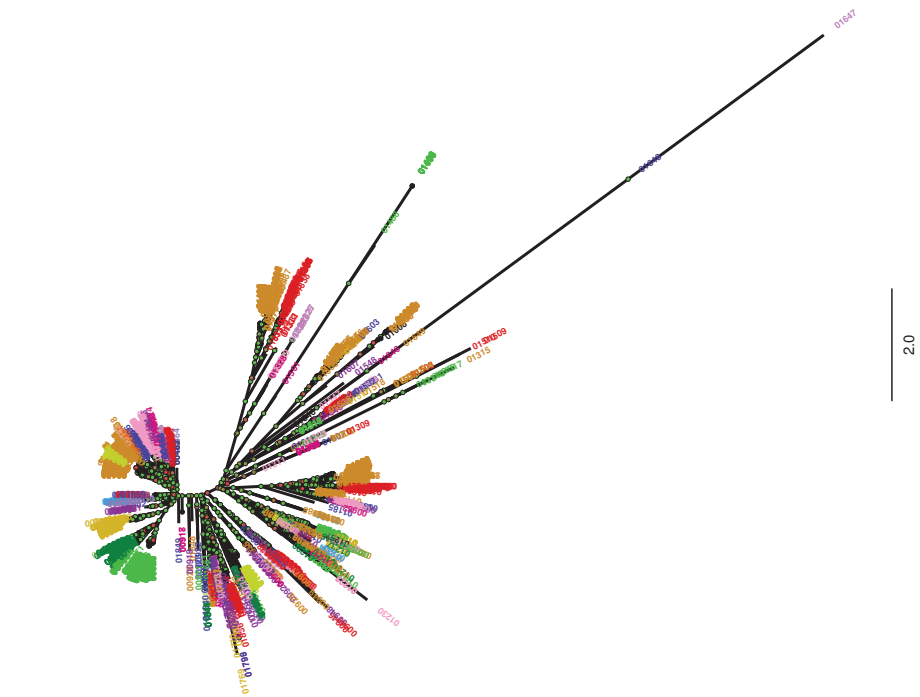

0.4

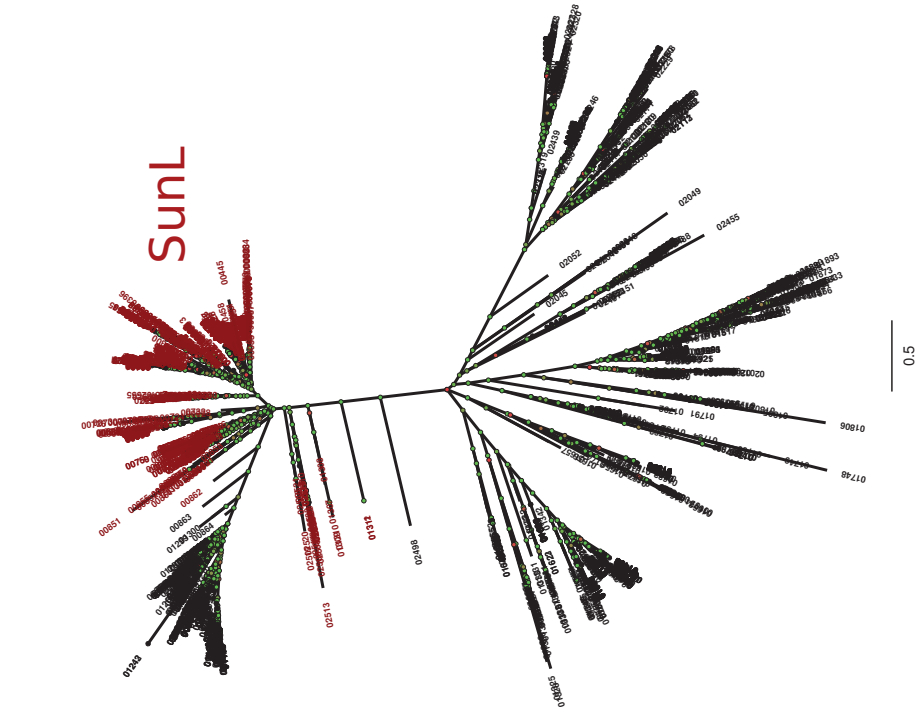

SpoIIIJ2
SpoIIIJ1
0.4

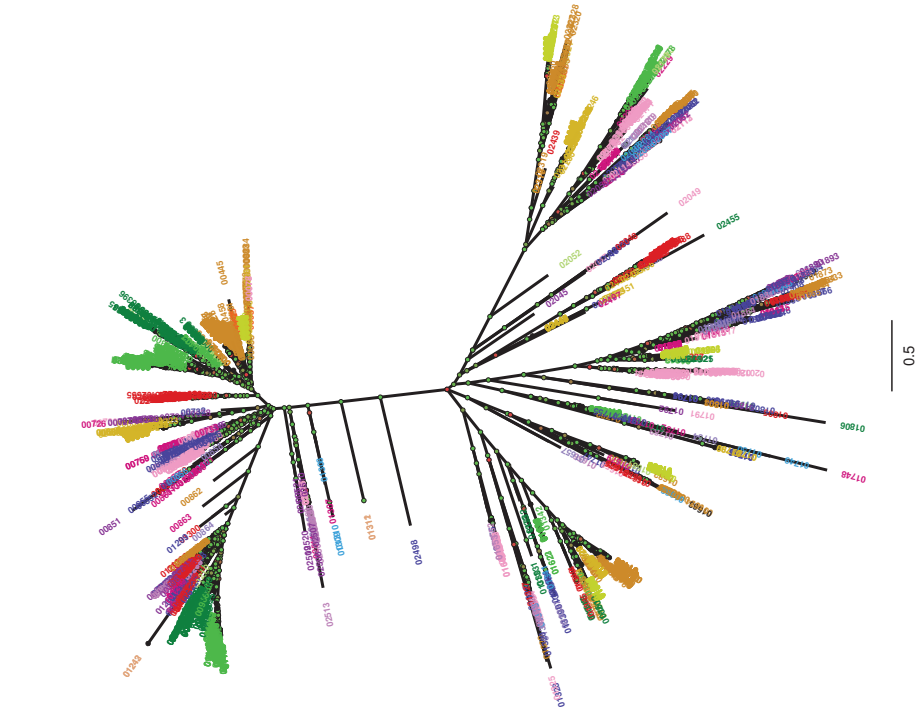

0.5

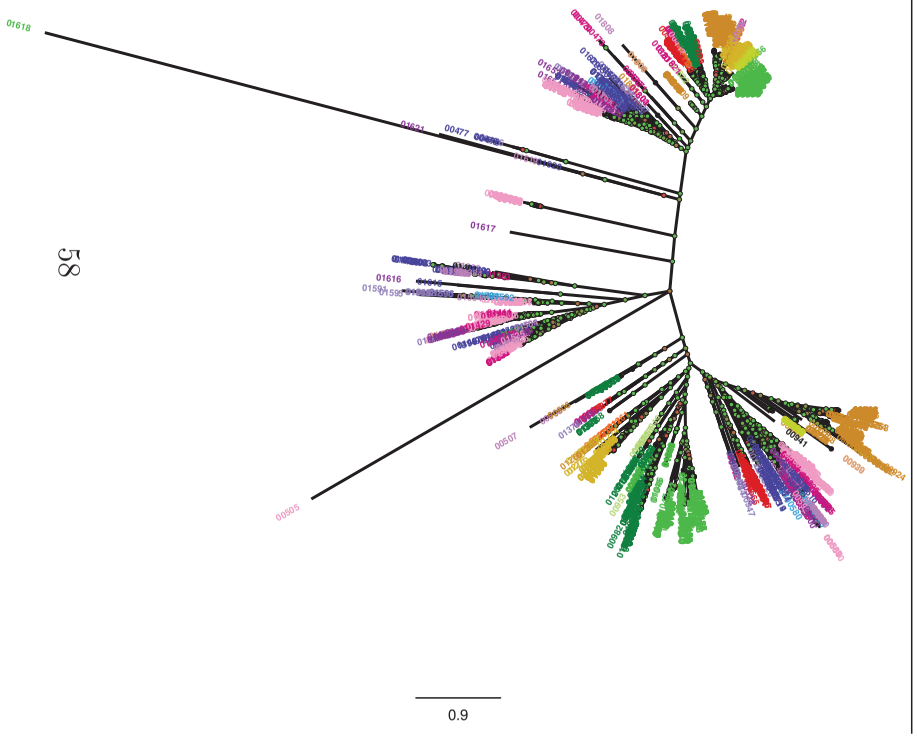

CDP4
0.5

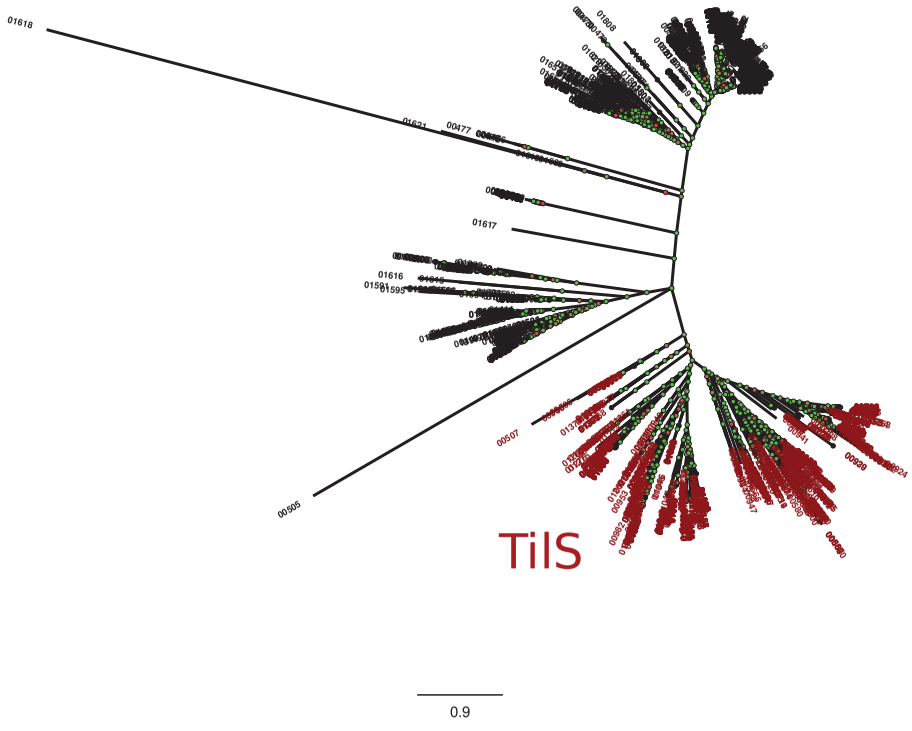
StkP
01647
2.0

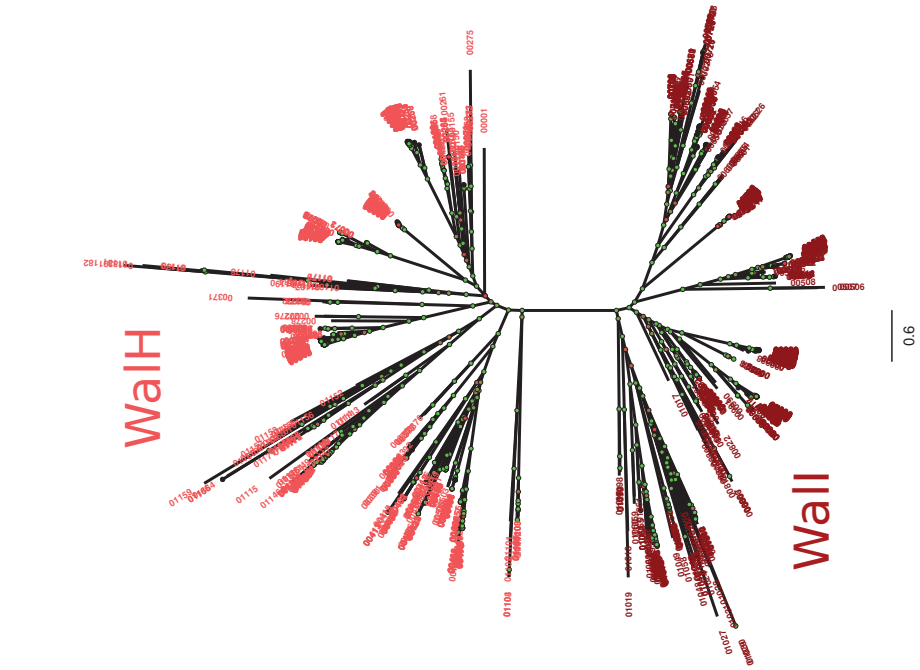

01647
2.0

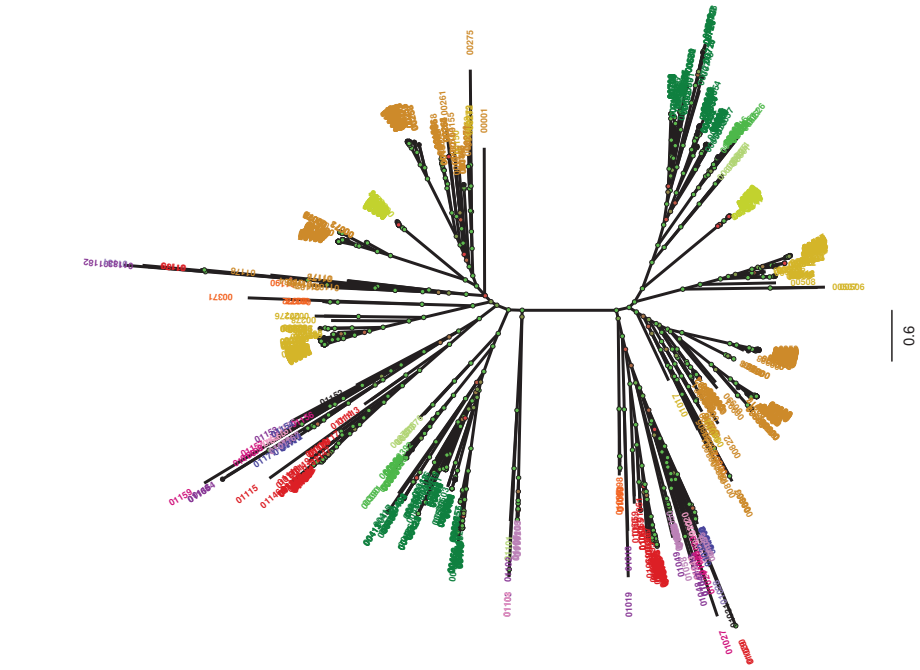

SunL
0.5

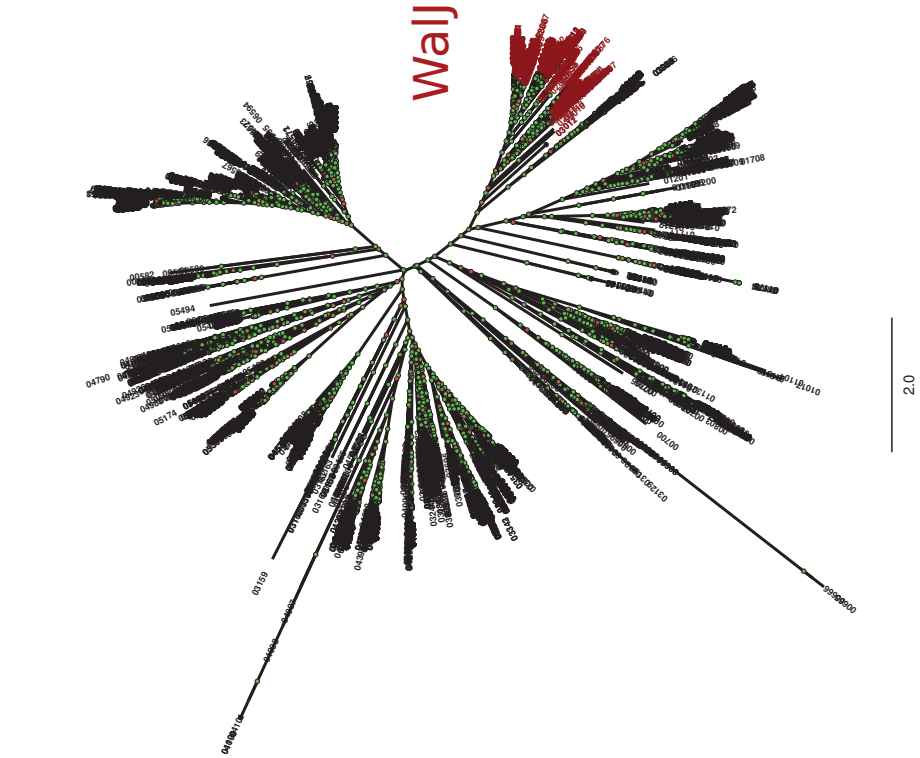

0.5

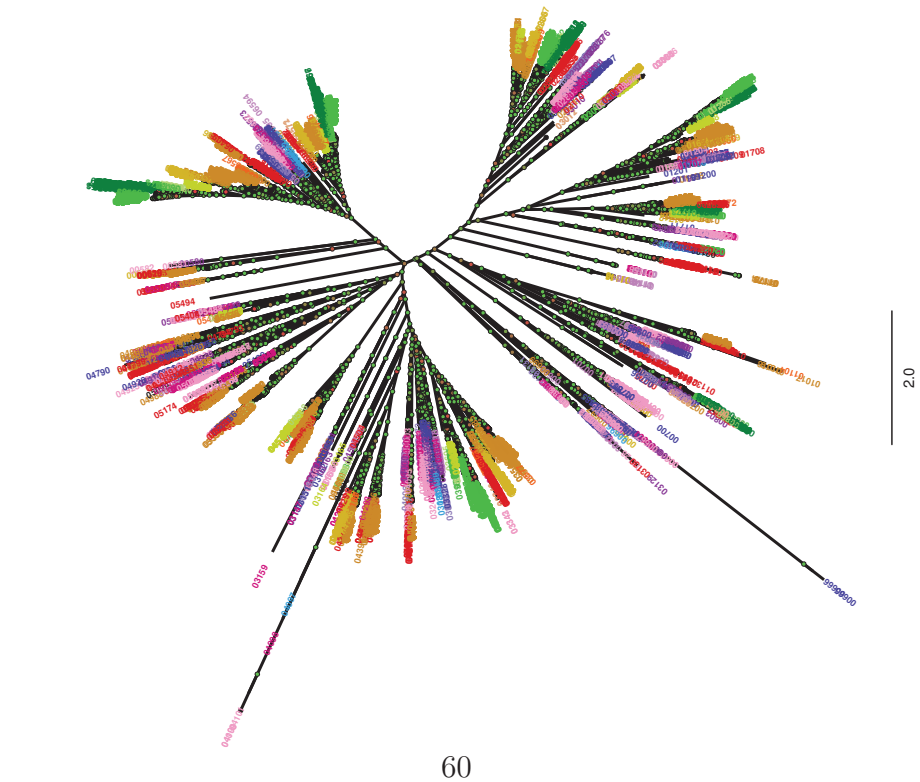

01618
01621
00477
01617
01616
01593
00505
00507
00941
00929
00924
0.9

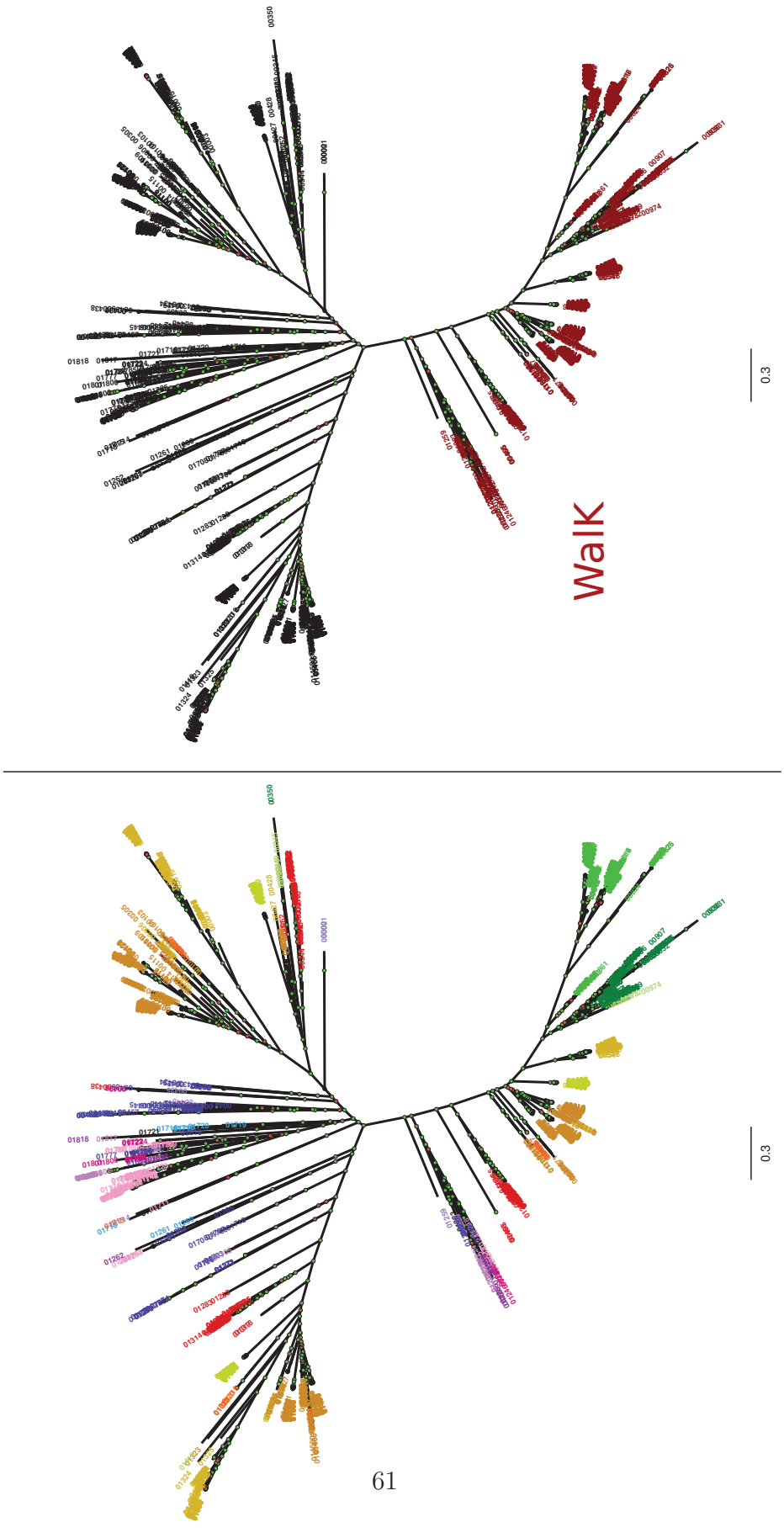

01618
01621
00477
01617
01616
01593
00505
00507
00941
00929
00924
TilS
0.9

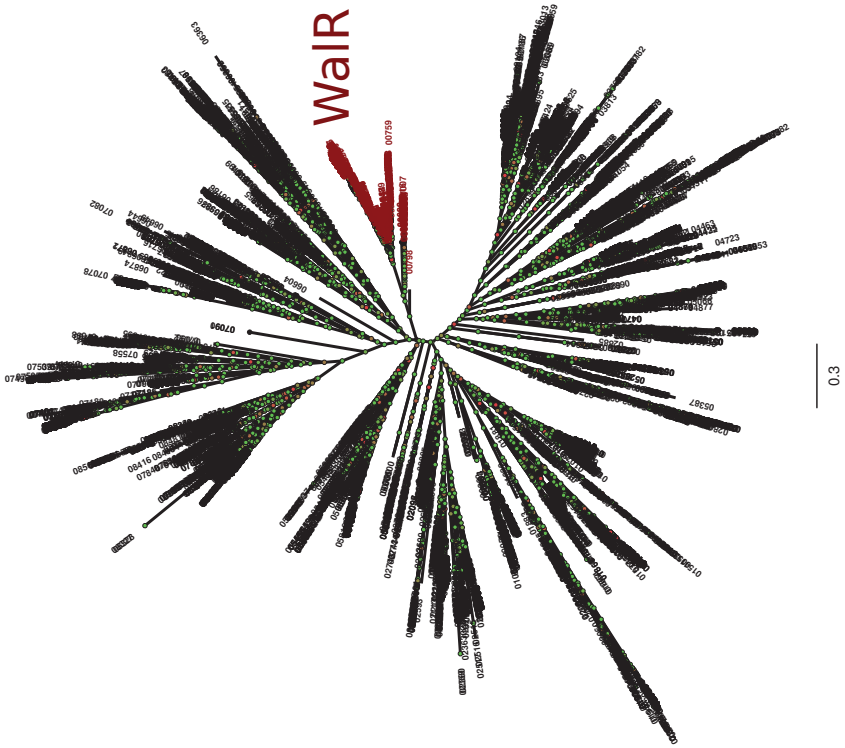

WalH
WalI
0.6

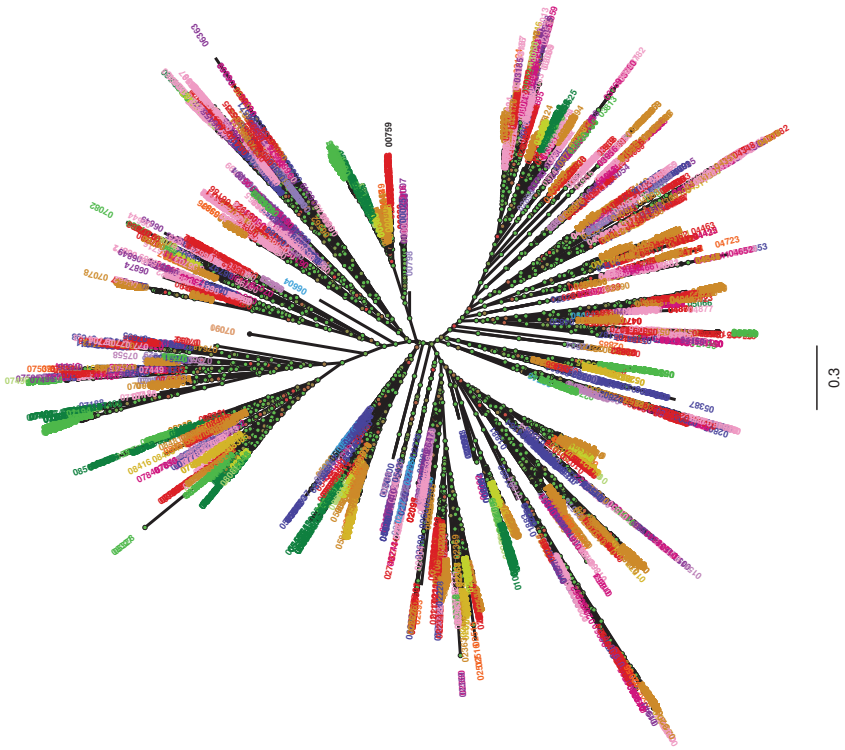

0.6

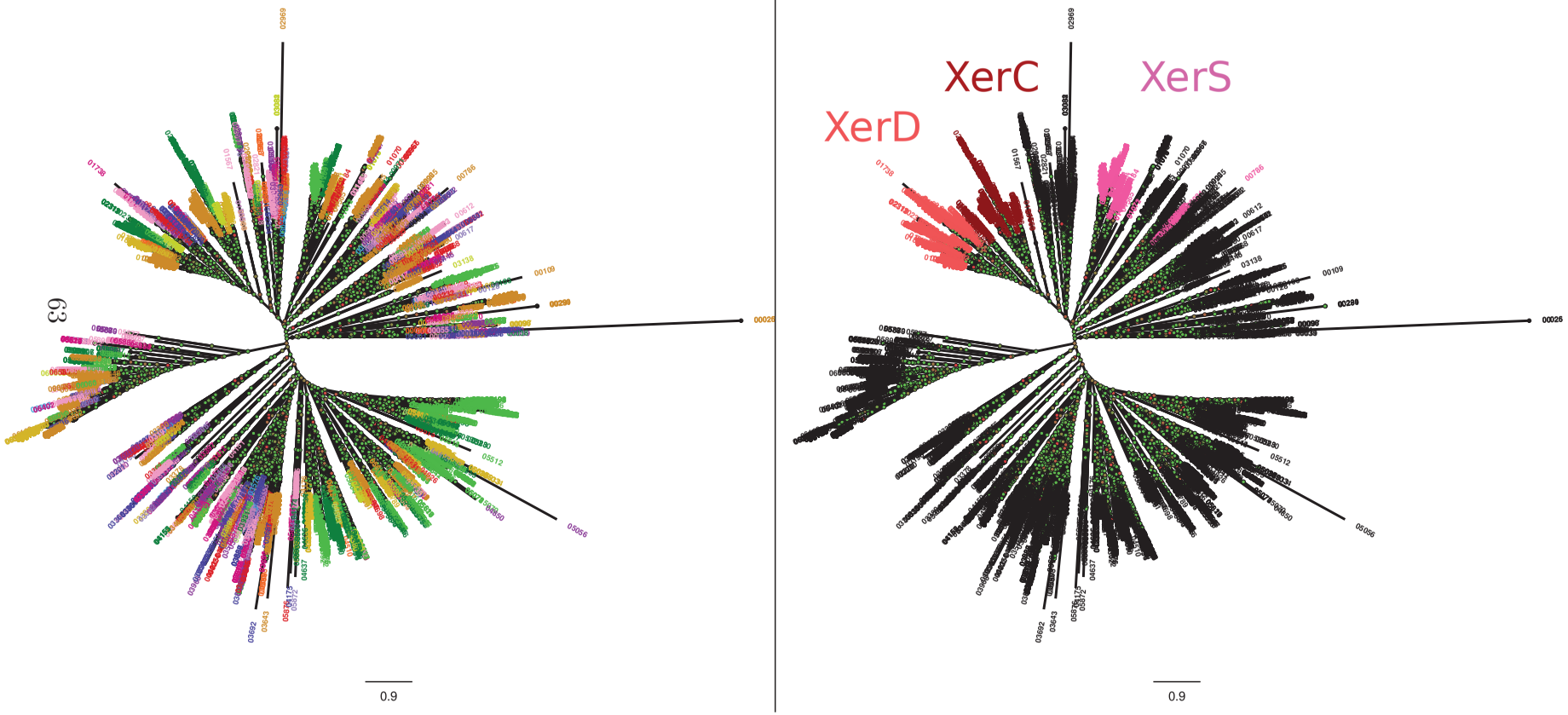

WalJ
2.0


2.0


WalK
0.3
0.3


WalR
0.3


0.3


63
XerC
XerS
XerD
0.9
0.9

.12 Phylogénies non racinées des familles/sous-familles de gènes.

Liste des arbres phylogénétiques non racinés des familles/sous-familles de gènes. Les barres d'echelle représentent le nombre moyen de substitutions par site. La couleur des feuilles correspond à la taxonomie. Les cercles aux noeuds correspondent aux valeurs de boostrap (gradietn de 0% à 100%, rouge : 0%, vert :100%).

1. A3, PBP1a(SP), PBP1(BS). 176 séquences, 518 positions, RAxML, LG+F+I+G4
2. A4, PBP2a(SP), PBP2c(BS). 203 séquences, 478 positions, RAxML, LG+F+I+G4
3. A5, PBP1b,(SP), PBP4(BS). 38 séquences, 526 positions, RAxML, LG+F+I+G4
4. A6, PBP1c(EC). 114 séquences, 525 positions, RAxML, LG+I+G4
5. AlaS, LovB. 305 séquences, 786 positions, RAxML, LG+F+I+G4
6. Alr-like. 155 séquences, 144 positions, RAxML, LG+G4
7. Alr, DadX(EC). 362 séquences, 247 positions, RAxML, LG+I+G4
8. AnmK. 28 séquences, 364 positions, RAxML, LG+G4
9. Aslfm. 59 séquences, 365 positions, RAxML, LG+F+G4
10. Asr. 159 séquences, 155 positions, RAxML, LG+I+G4
11. AX1. 213 séquences, 448 positions, RAxML, LG+F+I+G4
12. B1, PBP3(BS). 88 séquences, 543 positions, RAxML, LG+F+I+G4
13. B3, SpoVD(BS), PBP3(EC). 65 séquences, 635 positions, RAxML, LG+I+G4
14. B4, PBP2x(SP), PBP2b(BS). 357 séquences, 445 positions, RAxML, LG+F+I+G4
15. B5, PBP2b(SP), PBP2a(BS), PBPH(BS). 372 séquences, 327 positions, RAxML, LG+F+I+G4
16. B6, PBP4b(BS). 106 séquences, 335 positions, RAxML, LG+I+G4
17. BX1. 85 séquences, 361 positions, RAxML, LG+I+G4
18. CDP1. 123 séquences, 199 positions, RAxML, LG+I+G4
19. CDP3(RocS). 75 séquences, 99 positions, RAxML, LG+G4

20. CDP4. 157 séquences, 63 positions, RAxML, LG+F+G4
21. ClpX, LocP. 316 séquences, 363 positions, RAxML, LG+I+G4
22. CozE1. 90 séquences, 333 positions, RAxML, LG+F+G4
23. CozE2. 92 séquences, 294 positions, RAxML, LG+F+I+G4
24. CozE3. 152 séquences, 316 positions, RAxML, LG+F+I+G4
25. CozE4. 154 séquences, 230 positions, RAxML, LG+F+I+G4
26. CpsB, Cap1B, Wzb. 227 séquences, 148 positions, RAxML, LG+G4
27. CpsC, Wzd. 338 séquences, 143 positions, RAxML, LG+F+G4
28. CpsD. 278 séquences, 162 positions, RAxML, LG+G4
29. DacA(BS), PBP5(BS), DacC(EC), PBP6(EC). 275 séquences, 214 positions, RAxML, LG+F+I+G4
30. DacB(BS), PBP5*(BS), DacD(EC), PBP6b(EC). 157 séquences, 242 positions, RAxML, LG+I+G4
31. DacC(BS), PBP4a(BS), DacB(EC), PBP4(EC). 24 séquences, 451 positions, RAxML, LG+I+G4
32. DacF(BS), DacA(EC), PBP5(EC). 168 séquences, 298 positions, RAxML, LG+I+G4
33. Ddl. 263 séquences, 268 positions, RAxML, LG+I+G4
34. DivIC(BS), FtsB(EC). 314 séquences, 31 positions, RAxML, LG+F+G4
35. DivIVA. 284 séquences, 141 positions, RAxML, LG+G4
36. DnaA. 312 séquences, 342 positions, RAxML, LG+I+G4
37. DnaB(BS). 181 séquences, 193 positions, RAxML, LG+I+G4
38. DnaB1(EC), DnaC(BS). 304 séquences, 387 positions, RAxML, LG+F+I+G4
39. DnaB2. 29 séquences, 379 positions, RAxML, LG+I+G4
40. DnaD. 293 séquences, 92 positions, RAxML, LG+G4
41. DnaG. 305 séquences, 346 positions, RAxML, LG+I+G4
42. DnaI. 183 séquences, 240 positions, RAxML, LG+F+I+G4

43. DnaN. 322 séquences, 273 positions, RAxML, LG+F+I+G4
44. EzrA. 164 séquences, 406 positions, RAxML, LG+F+G4
45. FemA. 19 séquences, 418 positions, RAxML, LG+F+I+G4
46. FemB. 19 séquences, 419 positions, RAxML, LG+G4
47. FemX. 67 séquences, 298 positions, RAxML, LG+I+G4
48. Fmh. 26 séquences, 406 positions, RAxML, LG+F+G4
49. FtsA. 244 séquences, 335 positions, RAxML, LG+I+G4
50. FtsE. 243 séquences, 226 positions, RAxML, LG+I+G4
51. FtsH1, HflB, MrsC, Std, TolZ. 312 séquences, 515 positions, RAxML, LG+I+G4
52. FtsH2. 57 séquences, 553 positions, RAxML, LG+I+G4
53. FtsH3. 108 séquences, 413 positions, RAxML, LG+I+G4
54. FtsJ, MrsF, RrmJ. 284 séquences, 231 positions, RAxML, LG+I+G4
55. FtsK1(EC), SpoIIIE(BS). 321 séquences, 487 positions, RAxML, LG+I+G4
56. FtsK2(EC), SftA(BS). 90 séquences, 477 positions, RAxML, LG+I+G4
57. FtsL, MraR. 304 séquences, 25 positions, RAxML,
58. FtsQ(EC), DivIB(BS). 304 séquences, 72 positions, RAxML, LG+F+G4
59. FtsW. 298 séquences, 258 positions, RAxML, LG+F+I+G4
60. FtsX, ftsS. 253 séquences, 219 positions, RAxML, LG+F+I+G4
61. FtsY. 304 séquences, 284 positions, RAxML, LG+I+G4
62. FtsZ. 316 séquences, 292 positions, RAxML, LG+G4
63. GatD1. 138 séquences, 277 positions, RAxML, LG+I+G4
64. GatD2. 34 séquences, 316 positions, RAxML, LG+G4
65. GatD3. 28 séquences, 305 positions, RAxML, LG+I+G4
66. GatD4. 62 séquences, 263 positions, RAxML, LG+I+G4
67. GidA, MnmG, TrmF. 293 séquences, 602 positions, RAxML, LG+I+G4
68. GidB, RsmG. 305 séquences, 190 positions, RAxML, LG+I+G4

69. GlmM, MrsA, FemD(SA). 303 séquences, 416 positions, RAxML, LG+I+G4
70. GlmS. 307 séquences, 536 positions, RAxML, LG+I+G4
71. GlmU. 300 séquences, 425 positions, RAxML, LG+I+G4
72. GpsB. 172 séquences, 81 positions, RAxML, LG+I+G4
73. IleS1. 232 séquences, 856 positions, RAxML, LG+F+I+G4
74. IleS2. 91 séquences, 966 positions, RAxML, LG+F+I+G4
75. Jag, EloR. 255 séquences, 158 positions, RAxML, LG+I+G4
76. LeuS. 304 séquences, 744 positions, RAxML, LG+I+G4
77. LysA1. 227 séquences, 386 positions, RAxML, LG+I+G4
78. LysA2. 62 séquences, 410 positions, RAxML, LG+I+G4
79. LysA3. 26 séquences, 318 positions, RAxML, LG+I+G4
80. LytB(BS), CwbA. 126 séquences, 189 positions, RAxML, LG+I+G4
81. LytB(SP). 5 séquences, 604 positions, RAxML, JTT+F+I
82. MacP. 134 séquences, 25 positions, RAxML, LG+G4
83. Maf. 205 séquences, 150 positions, RAxML, LG+I+G4
84. MapZ, LocZ. 39 séquences, 349 positions, RAxML, LG+F+I+G4
85. Mbl. 229 séquences, 311 positions, RAxML, LG+I+G4
86. Mfd. 304 séquences, 879 positions, RAxML, LG+I+G4
87. MGT1. 18 séquences, 261 positions, RAxML, LG+G4
88. MGT2. 8 séquences, 258 positions, RAxML, LG+G4
89. MinC. 209 séquences, 130 positions, RAxML, LG+I+G4
90. MinD. 206 séquences, 245 positions, RAxML, LG+I+G4
91. MinE. 106 séquences, 64 positions, RAxML, LG+G4
92. MinJ. 154 séquences, 257 positions, RAxML, LG+F+G4
93. MraW, RsmH. 319 séquences, 278 positions, RAxML, LG+F+I+G4
94. MraY, MurX. 310 séquences, 253 positions, RAxML, LG+F+I+G4

95. MraZ. 239 séquences, 135 positions, RAxML, LG+I+G4
96. MreB. 231 séquences, 315 positions, RAxML, LG+I+G4
97. MreBH. 42 séquences, 330 positions, RAxML, LG+G4
98. MreC. 296 séquences, 159 positions, RAxML, LG+F+I+G4
99. MreD. 290 séquences, 76 positions, RAxML, LG+F+G4
100. Mtf, Fmt. 305 séquences, 282 positions, RAxML, LG+I+G4
101. MurA1. 265 séquences, 383 positions, RAxML, LG+I+G4
102. MurA2. 97 séquences, 390 positions, RAxML, LG+I+G4
103. MurB1. 248 séquences, 258 positions, RAxML, LG+I+G4
104. MurB2. 72 séquences, 284 positions, RAxML, LG+I+G4
105. MurC. 304 séquences, 346 positions, RAxML, LG+I+G4
106. MurD. 305 séquences, 343 positions, RAxML, LG+I+G4
107. MurE. 360 séquences, 326 positions, RAxML, LG+I+G4
108. MurF. 335 séquences, 295 positions, RAxML, LG+G4
109. MurG1. 200 séquences, 312 positions, RAxML, LG+I+G4
110. MurG2. 123 séquences, 318 positions, RAxML, LG+I+G4
111. MurI. 329 séquences, 184 positions, RAxML, LG+I+G4
112. MurJ, MviN. 201 séquences, 426 positions, RAxML, LG+F+I+G4
113. MurK. 99 séquences, 153 positions, RAxML, LG+I+G4
114. MurM. 39 séquences, 396 positions, RAxML, LG+I+G4
115. MurN. 62 séquences, 351 positions, RAxML, LG+I+G4
116. MurQ. 157 séquences, 283 positions, RAxML, LG+I+G4
117. MurT. 186 séquences, 343 positions, RAxML, LG+I+G4
118. MurZ. 251 séquences, 382 positions, RAxML, LG+I+G4
119. NagA. 277 séquences, 268 positions, RAxML, LG+I+G4
120. NagB, GlmD. 244 séquences, 214 positions, RAxML, LG+I+G4

121. Noc. 191 séquences, 228 positions, RAxML, LG+I+G4
122. NudF. 297 séquences, 157 positions, RAxML, LG+I+G4
123. ParA, Soj. 240 séquences, 240 positions, RAxML, LG+I+G4
124. ParB, Spo0J. 304 séquences, 220 positions, RAxML, LG+I+G4
125. PCDP10. 183 séquences, 140 positions, RAxML, LG+I+G4
126. PCDP4. 84 séquences, 131 positions, RAxML, VT+F+G4
127. PCDP6. 260 séquences, 189 positions, RAxML, LG+I+G4
128. PCDP7. 273 séquences, 60 positions, RAxML, LG+F+G4
129. PCDP8. 287 séquences, 161 positions, RAxML, LG+I+G4
130. Pfs, MtnM. 252 séquences, 192 positions, RAxML, LG+I+G4
131. PhpP, PppL. 304 séquences, 152 positions, RAxML, LG+I+G4
132. PriA. 305 séquences, 563 positions, RAxML, LG+I+G4
133. RacA. 40 séquences, 117 positions, RAxML, LG+G4
134. RadC. 353 séquences, 174 positions, RAxML, LG+I+G4
135. RecA, LexB, RecH, RnmB. 318 séquences, 311 positions, RAxML, LG+I+G4
136. RecF. 304 séquences, 289 positions, RAxML, LG+I+G4
137. RecG. 303 séquences, 536 positions, RAxML, LG+I+G4
138. RecN. 304 séquences, 442 positions, RAxML, LG+F+I+G4
139. RecR, RecM. 304 séquences, 193 positions, RAxML, LG+I+G4
140. RecU, PrfA. 196 séquences, 164 positions, RAxML, LG+G4
141. RodA, MrdB. 371 séquences, 278 positions, RAxML, LG+F+I+G4
142. RodA2. 42 séquences, 295 positions, RAxML, LG+F+G4
143. RodZ. 243 séquences, 102 positions, RAxML, LG+F+I+G4
144. Rtp. 30 séquences, 112 positions, RAxML, LG+G4
145. SbcC1. 126 séquences, 340 positions, RAxML, LG+F+I+G4
146. SbcC2. 75 séquences, 662 positions, RAxML, LG+F+I+G4

147. ScpA. 297 séquences, 153 positions, RAxML, LG+I+G4
148. ScpB. 297 séquences, 138 positions, RAxML, LG+I+G4
149. SepF. 306 séquences, 82 positions, RAxML, LG+F+I+G4
150. Smc. 298 séquences, 643 positions, RAxML, LG+F+I+G4
151. Spo0M. 80 séquences, 111 positions, RAxML, LG+G4
152. SpoIID, SpoIIC. 167 séquences, 209 positions, RAxML, LG+I+G4
153. SpoIIE, SpoIIH. 165 séquences, 473 positions, RAxML, LG+F+I+G4
154. SpoIIGA. 162 séquences, 152 positions, RAxML, LG+F+I+G4
155. SpoIIID. 168 séquences, 77 positions, RAxML, LG+G4
156. SpoIIIJ1, MisCA. 292 séquences, 157 positions, RAxML, LG+F+I+G4
157. SpoIIIJ2, MisCB. 172 séquences, 200 positions, RAxML, LG+I+G4
158. SpoVB, SpoIIIF. 158 séquences, 422 positions, RAxML, LG+F+I+G4
159. SpoVE. 189 séquences, 321 positions, RAxML, LG+F+I+G4
160. SpoVG. 236 séquences, 76 positions, RAxML, LG+G4
161. SpoVK, SpoVJ. 179 séquences, 215 positions, RAxML, LG+I+G4
162. StkP(SP), PknB(MT), Stk1(SA), PrkC(BS). 307 séquences, 325 positions, RAxML, LG+I+G4
163. SunL. 302 séquences, 330 positions, RAxML, LG+F+I+G4
164. TilS, MesJ. 300 séquences, 194 positions, RAxML, LG+I+G4
165. TolQ. 29 séquences, 156 positions, RAxML, LG+I+G4
166. TolR. 30 séquences, 111 positions, RAxML, LG+I+G4
167. ValS. 305 séquences, 808 positions, RAxML, LG+I+G4
168. WalH. 184 séquences, 78 positions, RAxML, LG+G4
169. WalI. 184 séquences, 68 positions, RAxML, LG+I+G4
170. WalJ. 278 séquences, 234 positions, RAxML, LG+I+G4
171. WalK, VicK. 216 séquences, 454 positions, RAxML, LG+F+I+G4
172. WalR. 195 séquences, 216 positions, RAxML, LG+I+G4

173. WecA, Rfe. 284 séquences, 296 positions, RAxML, LG+F+I+G4

174. XerC. 165 séquences, 255 positions, RAxML, LG+I+G4

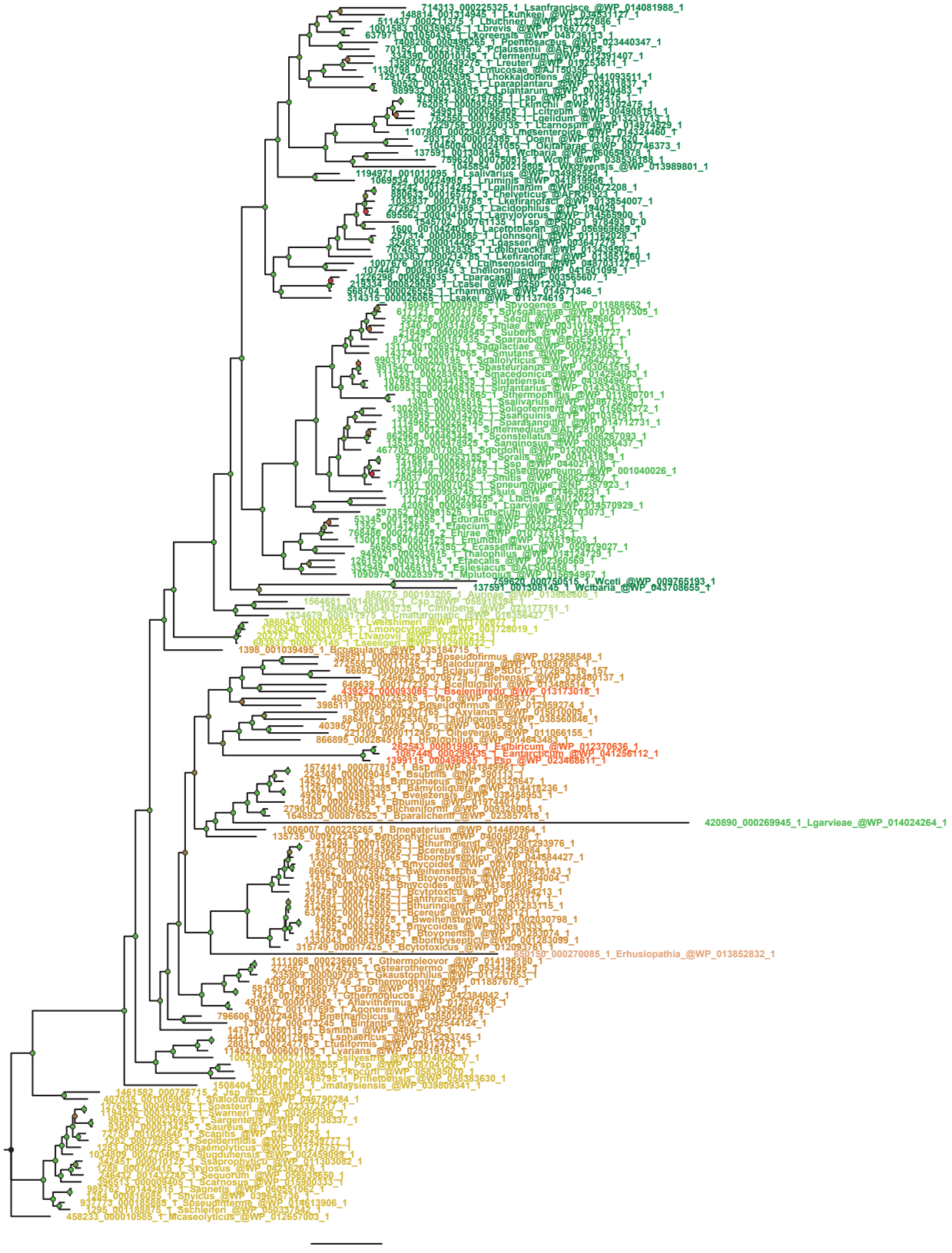
175. XerD, XprB, RipX. 246 séquences, 258 positions, RAxML, LG+I+G4

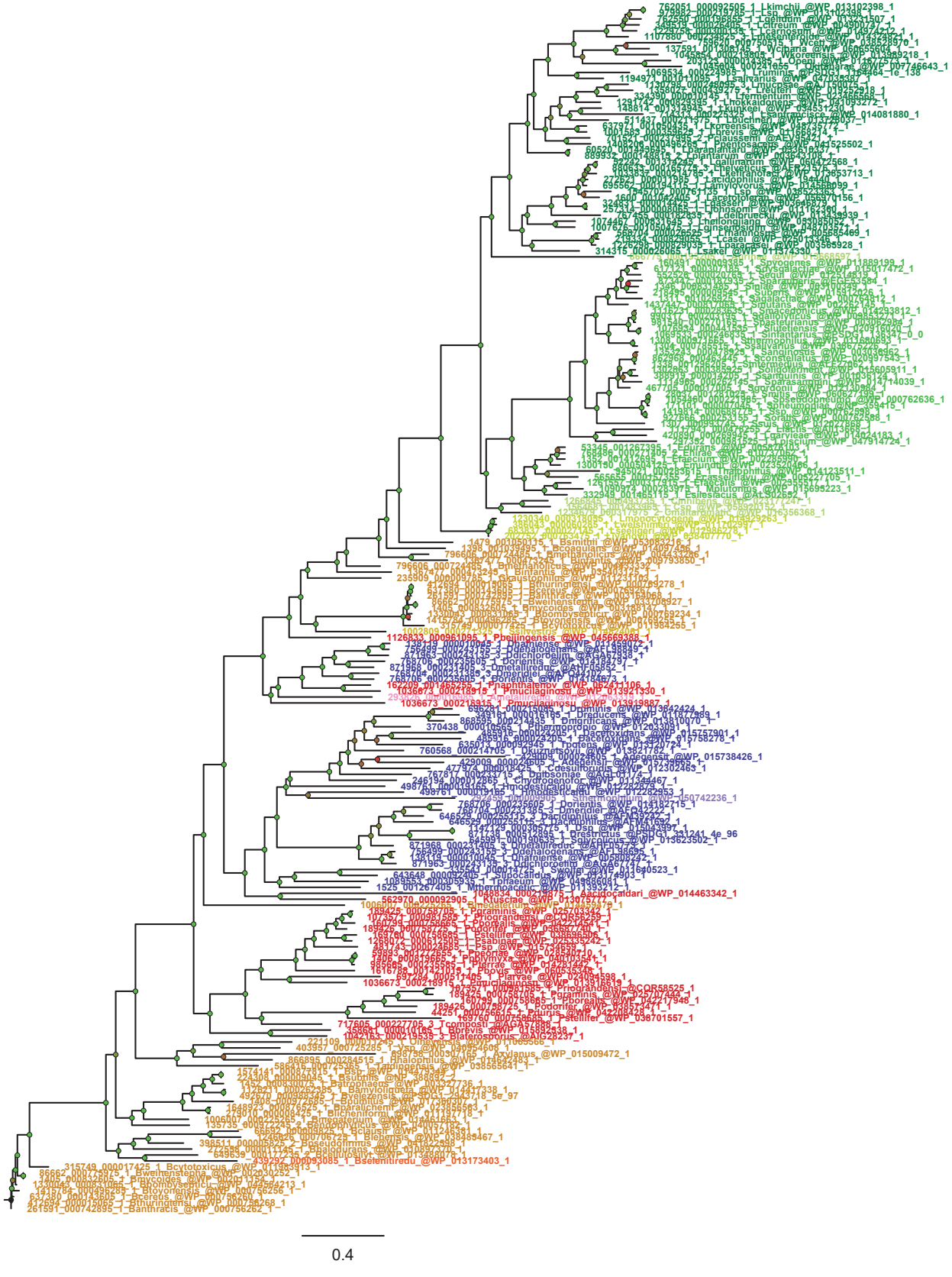
176. XerS. 117 séquences, 272 positions, RAxML, LG+F+I+G4

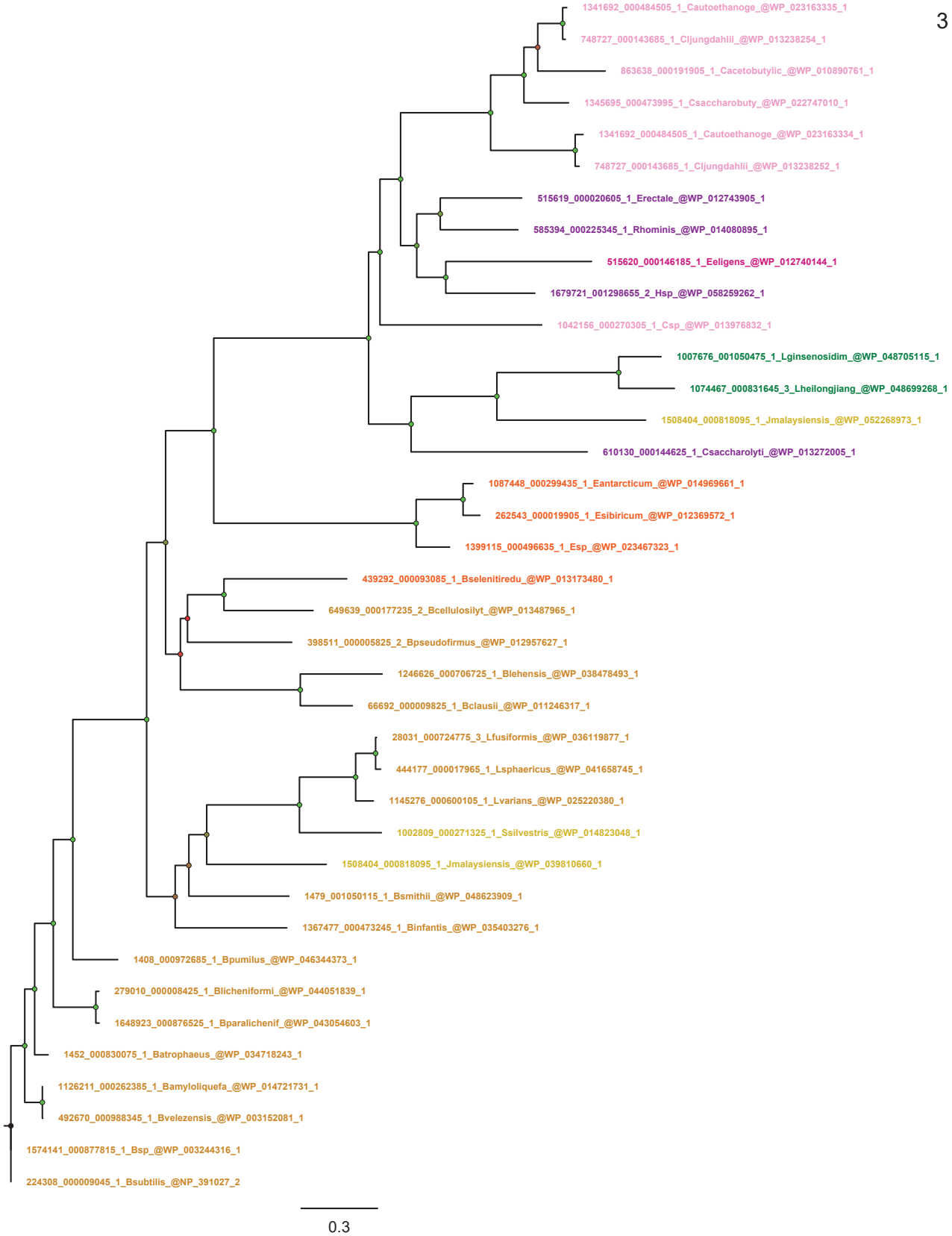
177. YabM. 289 séquences, 363 positions, RAxML, LG+F+G4

178. YkvU. 18 séquences, 438 positions, RAxML, LG+F+G4

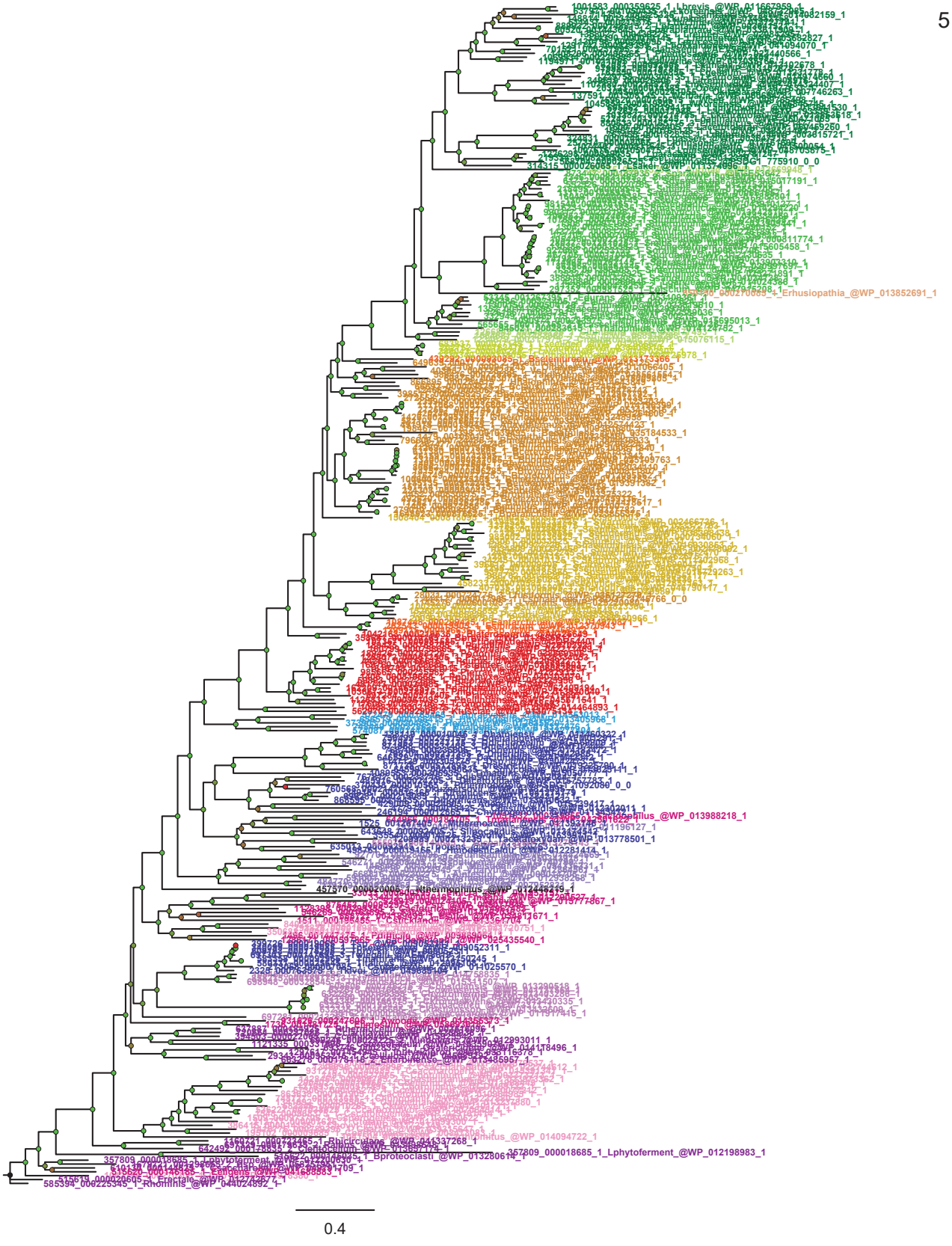
179. ZapA. 275 séquences, 57 positions, RAxML, LG+G4

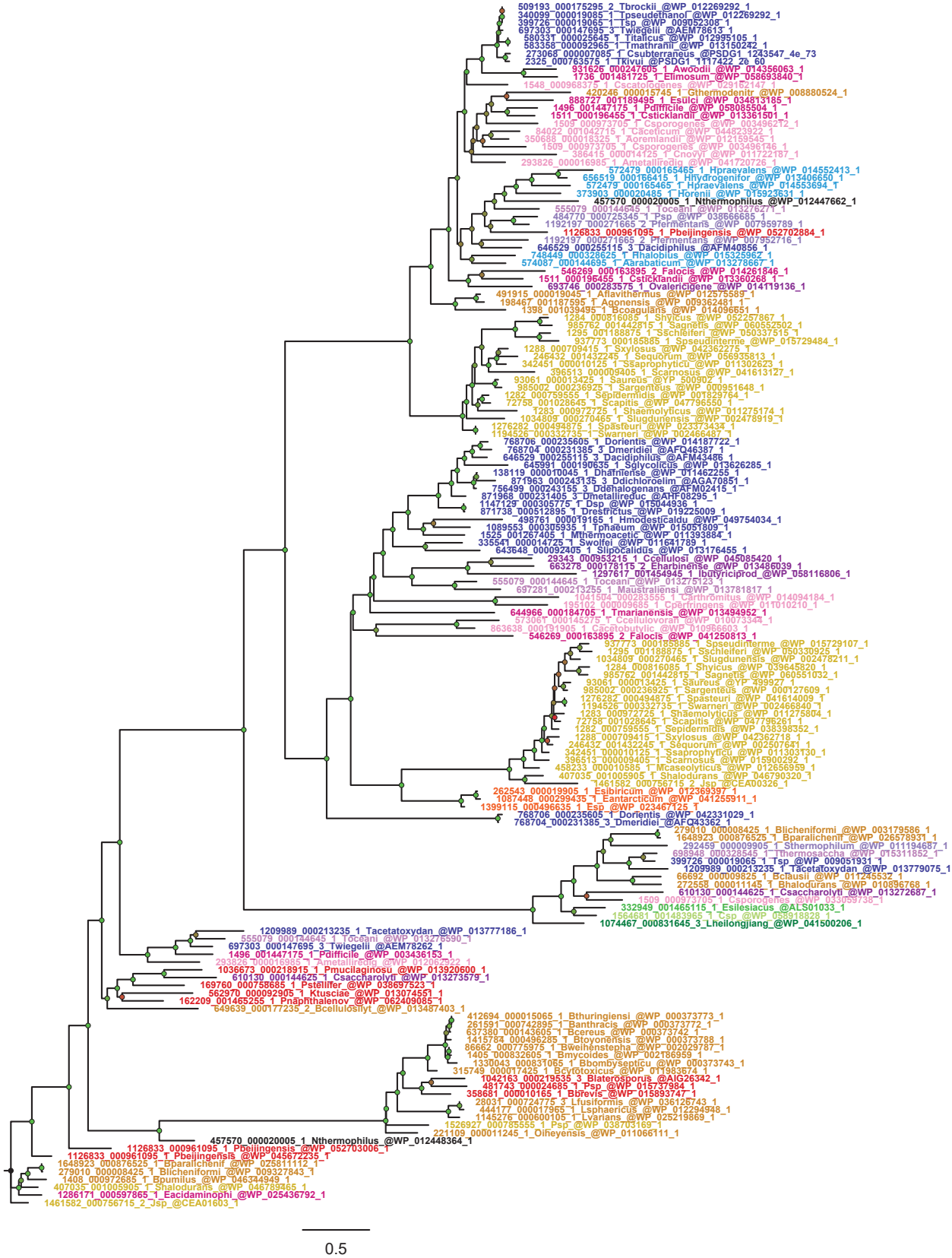

714313_000225325_1_Lsanfranciscе_@WP_014081988_1
1489114_001719845_1_Lkunkeei_@WP_013451127_1
511437_000211375_1_Lbuchneri_@WP_013747886_1
1001583_000359625_1_Lbrevis_@WP_011667779_1
637971_001050435_1_Lkoreensis_@WP_048736113_1
1408206_000496269_1_Ppentosaceus_@WP_023440347_1
701941_000237895_2_Pclaussenii_@AEV95285_1
1042390_000101145_1_Lfermentum_@WP_012391407_1
1358027_000439275_1_Lreuteri_@WP_019255611_1
1130798_000248095_1_Lmucosae_@AJT50358_1
1251742_000829395_1_Lhokkaidonens_@WP_041093511_1
60520_001443645_1_Lparaplantaru_@WP_033671837_1
889932_001438815_2_Lplantarum_@WP_003640483_1
872982_000219789_1_Lsp_@WP_011102475_1
762051_000092505_1_Lkimchii_@WP_013102475_1
342519_000096405_1_Lcrispatus_@WP_004908151_1
342520_000198855_1_Lgeilidum_@WP_013231713_1
1229758_000300135_1_Lcarnosum_@WP_014974529_1
1107880_000234925_3_Lmesenteroide_@WP_014324460_1
208123_000014385_1_Loeni_@WP_011677620_1
1045004_000241055_1_Okitaharae_@WP_007746373_1
137591_001308145_1_Wcibaria_@WP_036685878_1
759620_000750515_1_Wceti_@WP_038536188_1
1045854_000218605_1_Wkoreensis_@WP_013989801_1
1194971_001011095_1_Lsalivarius_@WP_034982554_1
1069534_000224965_1_Lruminis_@WP_041819366_1
5721_001314245_1_Lgallinarum_@WP_006472208_1
880433_000167775_3_Lhelveticus_@AFK21921_1
1023837_000214785_1_Lkefiranofact_@WP_013854007_1
270621_000111955_1_Lacidophilus_@WP_11940241_1
695562_000134115_1_Lamylovorus_@WP_014683900_1
1545702_000761135_1_Lsp_@PS5G1_976493_0_0
1600_001044205_1_Lacetotoleran_@WP_056858689_1
257314_000008085_1_Ljohnsonii_@WP_011162028_1
224831_000014425_1_Lgasseri_@WP_003641279_1
767465_000182835_1_Ldelbrueckii_@WP_013439802_1
1023837_000214785_1_Lkefiranofact_@WP_013851260_1
1007676_001050475_1_Lsmenosidim_@WP_038703127_1
1074467_000831945_3_Lhellingiang_@WP_041501099_1
1226298_000829035_1_Lparacasei_@WP_003565607_1
218534_000329655_1_Lcasei_@WP_011672331_1
568704_000225525_1_Lrhamnosus_@WP_014571346_1
314315_000026065_1_Lsakei_@WP_011374619_1
1004971_000093385_1_Spyogenes_@WP_011888662_1
617421_000077195_1_Sdysgalactiae_@WP_015017305_1
552526_000020765_1_Sagal_@WP_041755680_1
7346_000311455_1_Shiae_@WP_003107794_1
219495_000009545_1_Suberis_@WP_015931727_1
873477_000187935_2_Sparauberis_@EJE54501_1
1311_001026925_1_Sagalactiae_@WP_000628365_1
1437447_000817065_1_Smutans_@WP_002263053_1
980317_000203135_1_Sgallolyticus_@WP_015627975_1
981540_000701165_1_Spasteurianus_@WP_003063575_1
1116211_000283635_1_Smacedonicus_@WP_014294053_1
1076933_000241935_1_Slutetiensis_@WP_043854967_1
1076933_000241935_1_Sinfantarius_@WP_014354388_1
1308_000971665_1_Sthermophilus_@WP_011680701_1
1704_000755515_1_Ssalivarius_@WP_013967252_1
1302883_000388915_1_Soligofermen_@WP_015053372_1
388919_000014205_1_Ssanguinis_@YP_001035791_1
1741965_000624145_1_Sparasanguin_@WP_014712731_1
1338_001298505_1_Smitermedius_@AEF28100_1
862868_000463445_1_Sconstellatus_@WP_006267033_1
862323_000478925_1_Sanginosus_@WP_003063437_1
467705_000017005_1_Sgordonii_@WP_012000082_1
927689_000233155_1_Sgralis_@WP_001041839_1
1141818_000886775_1_Ssp_@WP_044021318_1
1054460_000221965_1_Spseudopneumo_@WP_001040026_1
28037_001028515_1_Smitis_@WP_003627837_1
171101_000010045_1_Spneumoniae_@NP_357923_1
1307_000993745_1_Ssuis_@WP_014736231_1
417941_000478285_2_Llactis_@AII74822_1
420890_000269945_1_Lgarvieae_@WP_014570929_1
297352_000981525_1_Lplantarum_@WP_050703073_1
53345_001167395_1_Edurans_@WP_005878838_1
7352_001412695_1_Efaecium_@WP_002428422_1
768486_000271405_2_Ehirae_@WP_010731513_1
1300150_000504125_1_Emundtii_@WP_023619603_1
565655_001147355_2_Ecasseliflavu_@WP_050979027_1
945021_000283615_1_Thalophilus_@WP_013134729_1
1261317_001317915_1_Efaecalis_@WP_002360569_1
332948_001465115_1_Esileslacus_@AFS00458_1
1090974_000283975_1_Mplutonius_@WP_015934967_1
759620_000750515_1_Wceti_@WP_009765193_1
137591_001308145_1_Wcibaria_@WP_043708655_1
866776_000193205_1_Auricac_@WP_013668605_1
1564681_001483965_1_Csp_@WP_058933937_1
1266845_000493735_1_Cinnibens_@WP_023177751_1
1234678_000317975_2_Cmaltaromatic_@WP_013556427_1
386043_000060285_1_Lwelshimeri_@WP_017702677_1
1290340_000316055_1_Lmonocytogene_@WP_003728019_1
202782_000763475_1_Livanovii_@WP_009720214_1
683837_000027145_1_Lseeligeri_@WP_012956022_1
1398_001039495_1_Bcoagulans_@WP_035184715_1
398511_000005825_2_Bpseudofirmus_@WP_012958548_1
272558_000011145_1_Bhalodurans_@WP_010897863_1
66692_000009825_1_Bclausii_@PS5G1_2172693_19_157
246626_000709725_1_Blehensis_@WP_039480137_1
649639_000177235_2_Bcellulosilyt_@WP_013489514_1
438291_000093085_1_Bsubtilis_@WP_013173018_1
403957_000725285_1_Vsp_@WP_040955474_1
398511_000005825_2_Bpseudofirmus_@WP_012959274_1
5987754_000007765_1_Axylanus_@WP_015100005_1
586416_000725365_1_Taidingensis_@WP_038560846_1
403957_000725285_1_Vsp_@WP_040955474_1
421109_000071245_1_Chevensis_@WP_011066155_1
866895_000284515_1_Hhalophilus_@WP_014643483_1
262593_000198905_1_Esibiricum_@WP_012370636_1
1087448_000293435_1_Eantarcticum_@WP_041256112_1
1399115_000498635_1_Esp_@WP_023468611_1
1574141_000877815_1_Bsp_@WP_015959973_1
224308_000009045_1_Bsubtilis_@NP_390713_1
1452_000830075_1_Batrophaeus_@WP_003325647_1
1126211_000782385_1_Bamyloliquefa_@WP_014413236_1
492870_000988345_1_Bvelezensis_@WP_038458953_1
1408_000972685_1_Bpumilus_@WP_019744017_1
279010_000008425_1_Blichenformi_@WP_003283005_1
1648923_000876525_1_Bparalicheni_@WP_023857418_1
420890_000269945_1_Lgarvieae_@WP_014024264_1
1006007_000225265_1_Bmegaterium_@WP_014460964_1
135735_000972245_2_Bendophyticus_@WP_040058248_1
414984_000015065_1_Bthurmgiensi_@WP_001293976_1
637300_001143805_1_Bcereus_@WP_001283984_1
1330043_000831065_1_Bbombysepticu_@WP_044584427_1
1405_000832805_1_Bmycoides_@WP_001189071_1
86662_000775975_1_Bweihenstepha_@WP_038626143_1
1415784_000496285_1_Btoyonensis_@WP_001294004_1
1405_000832805_1_Bmycoides_@WP_041988005_1
415749_000171425_1_Bcytotoxicus_@WP_012094213_1
261591_000742895_1_Banthracis_@WP_001283117_1
414984_000015065_1_Bthuringiens_@WP_001283115_1
637300_001143805_1_Bcereus_@WP_001283121_1
86662_000775975_1_Bweihenstepha_@WP_002030798_1
1405_000832805_1_Bmycoides_@WP_001283133_1
1415784_000496285_1_Btoyonensis_@WP_001283074_1
1330043_000831065_1_Bbombysepticu_@WP_001283099_1
315749_000171425_1_Bcytotoxicus_@WP_012094761_1
1111068_000236605_1_Gthermoleovor_@WP_014196180_1
272557_001473575_1_Gstearothermo_@WP_014174695_1
235909_000009785_1_Gkaustophilus_@WP_011231653_1
420246_000815745_1_Ghaloduranic_@WP_011887678_1
581103_000166075_1_Gsp_@WP_013009529_1
1428_001293365_1_Gthermoglucos_@WP_042384042_1
491915_000019045_1_Altiaythermus_@WP_012974768_1
198467_001187595_1_Agonensis_@WP_035066992_1
796606_000724485_1_Bmethanolicus_@WP_038502205_1
1357477_000473245_1_Bintermis_@WP_022544124_1
1479_001050115_1_Bsmithii_@WP_048623543_1
444417_000011965_1_Lspheericus_@WP_012293745_1
280031_000742775_1_Efusiformis_@WP_038124731_1
1145276_000800105_1_Lvarians_@WP_025218152_1
1002868_000721525_1_Ssilvestris_@WP_014972767_1
1262927_000785555_1_Psp_@WP_038704126_1
1374_001468935_1_Pkpcurri_@WP_058385070_1
280991_001485795_1_Prifletoensis_@WP_058833630_1
1508404_000818095_1_Jmalaysiensis_@WP_039809341_1
1461582_000756715_2_Jsp_@CEA00824_1
407038_001005905_1_Shaladurans_@WP_046790284_1
1216282_000494915_1_Spasteuri_@WP_023375517_1
1184526_000332735_1_Swarneri_@WP_002466808_1
985002_000235925_1_Sargenteus_@WP_001738337_1
83051_000013425_1_Saureus_@YP_499885_1
72758_001026645_1_Scapitis_@WP_033855251_1
1282_000759555_1_Sepidermidis_@WP_002439777_1
1263_000974725_1_Shaemolyticus_@WP_011275757_1
1034809_000270485_1_Slugudunensis_@WP_002456899_1
74245T_000010125_1_Ssaprophyticu_@WP_011303082_1
1288_000709415_1_Sxylosus_@WP_012362878_1
246432_001432245_1_Sequorum_@WP_058936650_1
496513_000008405_1_Scarnosus_@WP_015800333_1
985762_001442815_1_Sgenetis_@WP_060551062_1
1284_000816085_1_Shyicus_@WP_033845736_1
937273_000185885_1_Spseudintermm_@WP_014613906_1
1245_001188875_1_Sschleiferi_@WP_050397540_1
458233_000010585_1_Mcaseolyticus_@WP_012657003_1
0.4

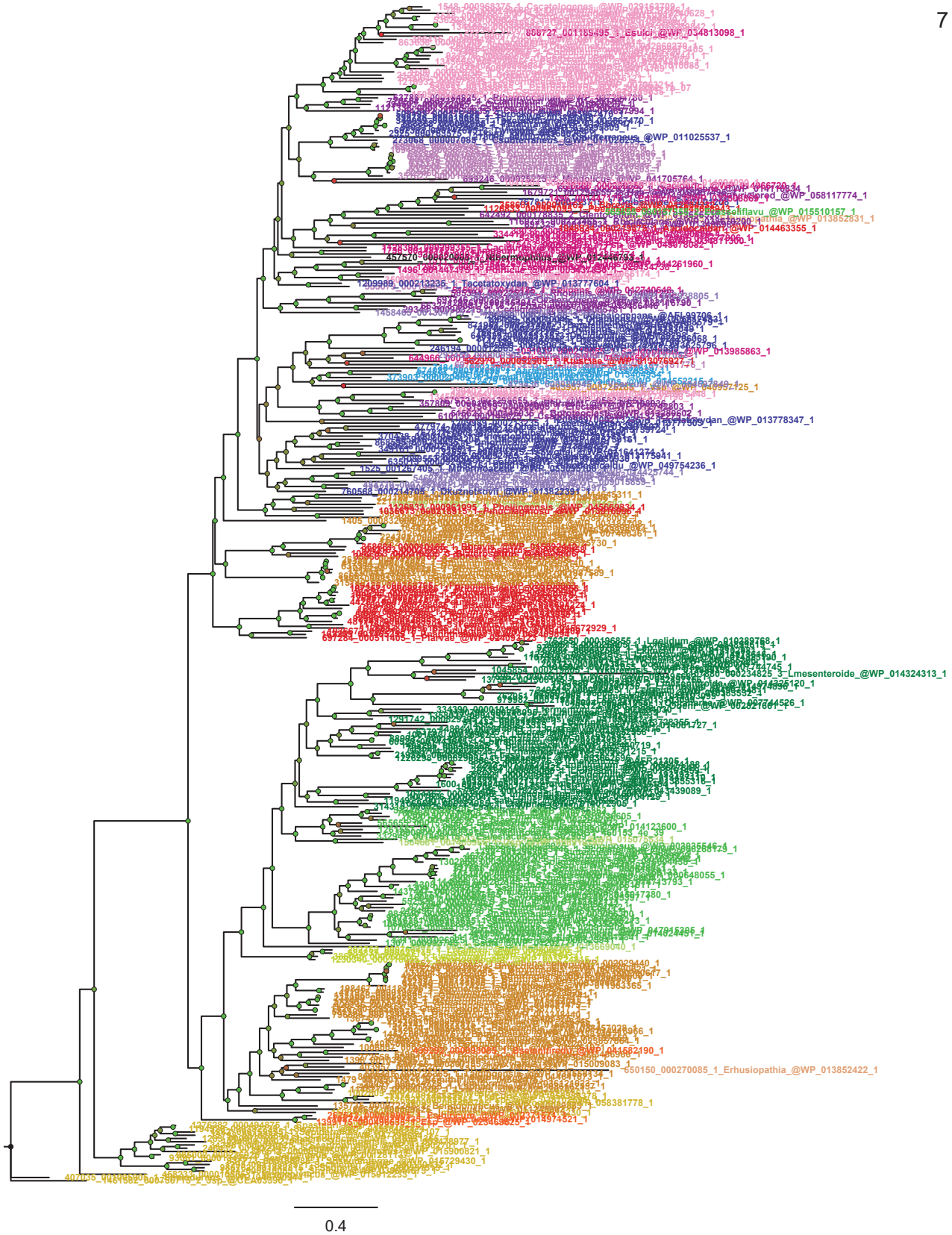

2
0.4

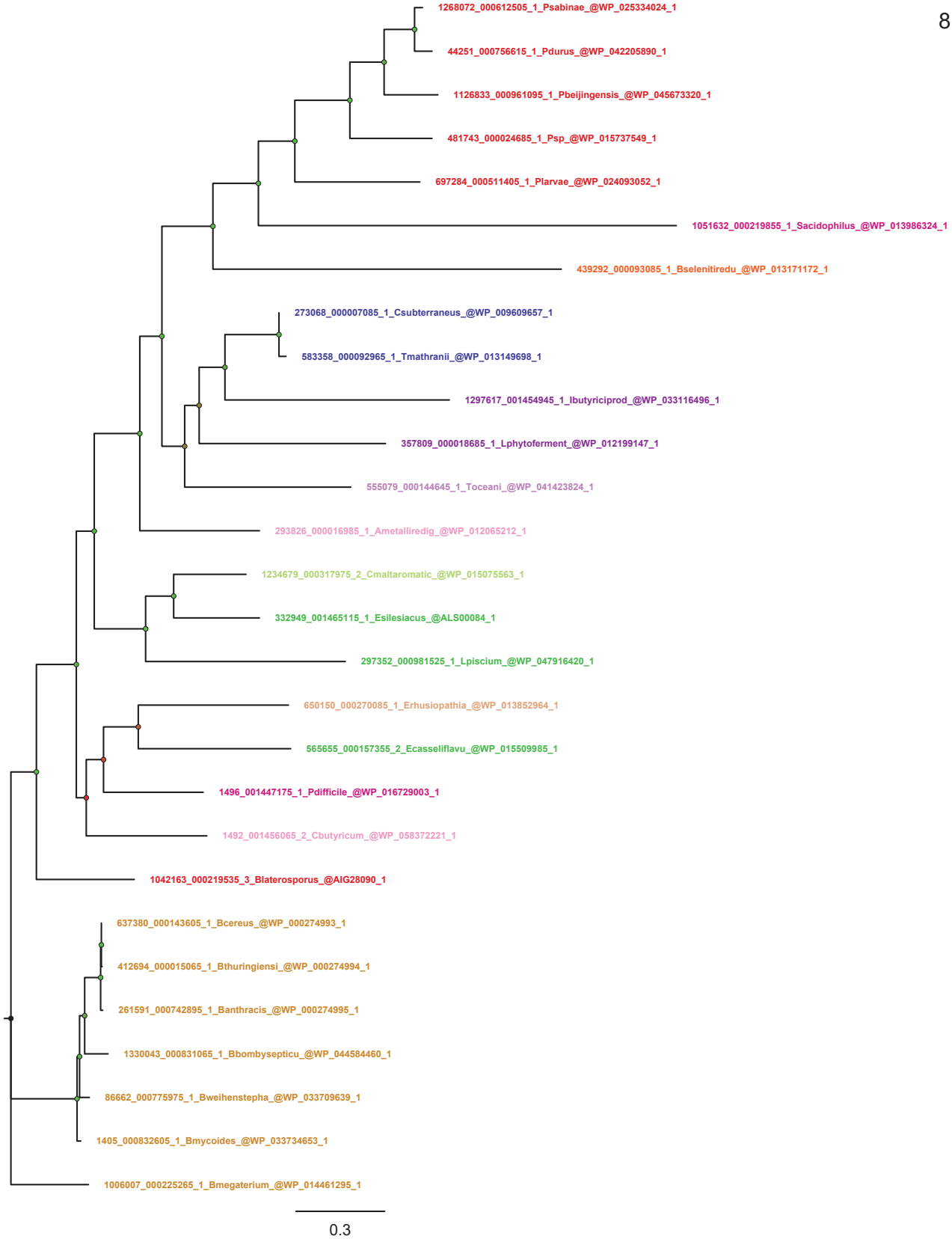

1341692_000484505_1_Cautoethanoge_@WP_023163335_1
748727_000143685_1_Cljungdahlii_@WP_013238254_1
863638_000191905_1_Cacetobutylic_@WP_010890761_1
1345695_000473995_1_Csaccharobuty_@WP_022747010_1
1341692_000484505_1_Cautoethanoge_@WP_023163334_1
748727_000143685_1_Cljungdahlii_@WP_013238252_1
515619_000020605_1_Erectale_@WP_012743905_1
585394_000225345_1_Rhominis_@WP_014080895_1
515620_000146185_1_Eeligens_@WP_012740144_1
1679721_001298655_2_Hsp_@WP_058259262_1
1042156_000270305_1_Csp_@WP_013976832_1
1007676_001050475_1_Lginsenosidim_@WP_048705115_1
1074467_000831645_3_Lheilongjiang_@WP_048699268_1
1508404_000818095_1_Jmalaysiensis_@WP_052268973_1
610130_000144625_1_Csaccharolyti_@WP_013272005_1
1087448_000299435_1_Eantarcticum_@WP_014969661_1
262543_000019905_1_Esibiricum_@WP_012369572_1
1399115_000496635_1_Esp_@WP_023467323_1
439292_000093085_1_Bselenitiredu_@WP_013173480_1
649639_000177235_2_Bcellulosilyt_@WP_013487965_1
398511_000005825_2_Bpseudofirmus_@WP_012957627_1
1246626_000706725_1_Blehensis_@WP_038478493_1
66692_000009825_1_Bclausii_@WP_011246317_1
28031_000724775_3_Lfusiformis_@WP_036119877_1
444177_000017965_1_Lsphaericus_@WP_041658745_1
1145276_000600105_1_Lvarians_@WP_025220380_1
1002809_000271325_1_Ssilvestris_@WP_014823048_1
1508404_000818095_1_Jmalaysiensis_@WP_039810660_1
1479_001050115_1_Bsmithii_@WP_048623909_1
1367477_000473245_1_Binfantis_@WP_035403276_1
1408_000972685_1_Bpumilus_@WP_046344373_1
279010_000008425_1_Blicheniformi_@WP_044051839_1
1648923_000876525_1_Bparalichenif_@WP_043054603_1
1452_000830075_1_Batrophaeus_@WP_034718243_1
1126211_000262385_1_Bamyloliquefa_@WP_014721731_1
492670_000988345_1_Bvelezensis_@WP_003152081_1
1574141_000877815_1_Bsp_@WP_003244316_1
224308_000009045_1_Bsubtilis_@NP_391027_2
0.3

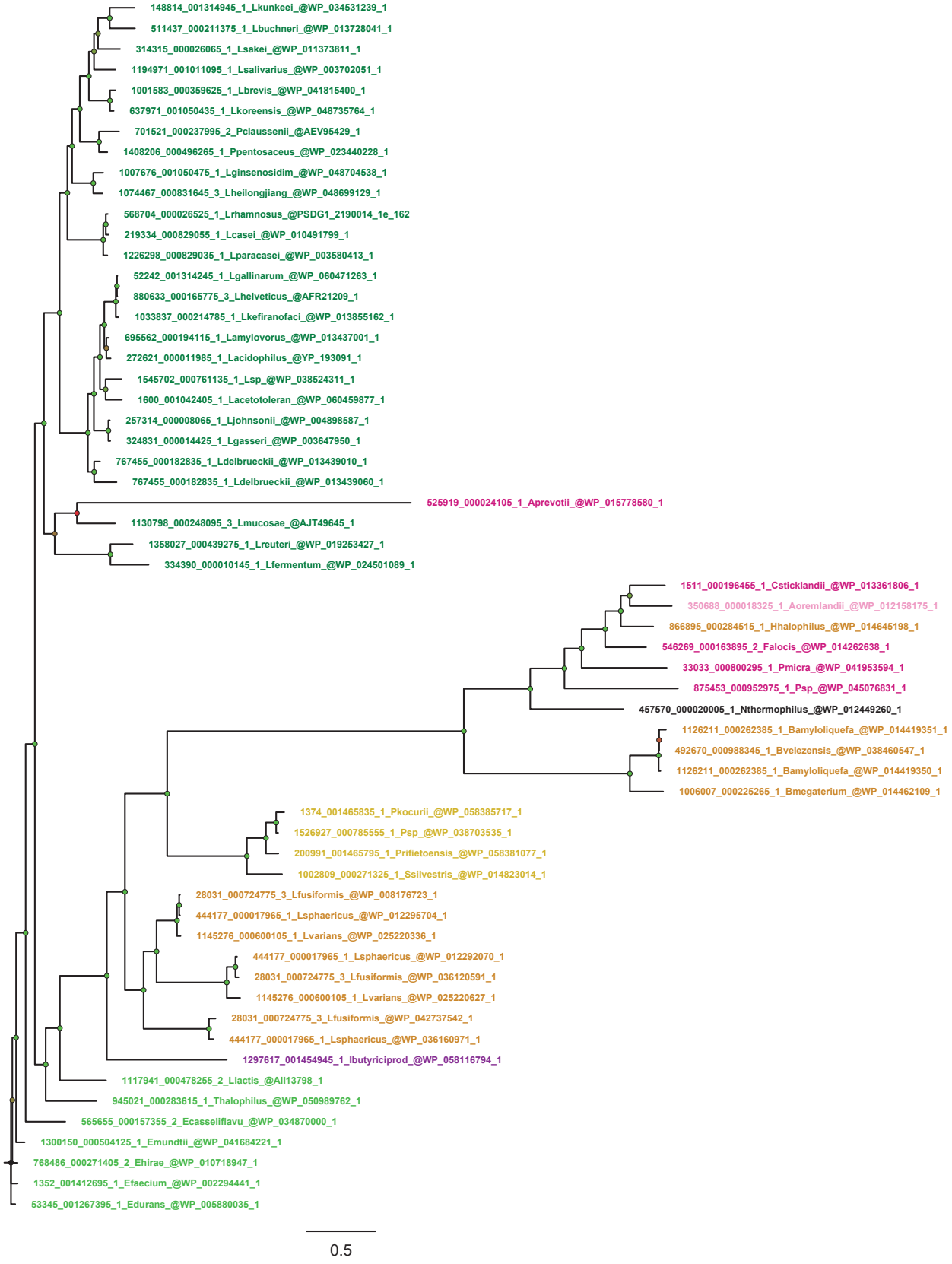
546271_000208405_1_Ssputigena_@WP_006192435_1
927704_000284095_1_Sruminantium_@WP_014425107_1
712538_001189555_1_Ssp_@WP_050343953_1
484770_000725345_1_Psp_@WP_038670251_1
1192197_000271665_2_Pfermentans_@WP_007961506_1
479436_000024945_1_Vparvula_@WP_012864473_1
1458465_001304715_1_Melsdenii_@WP_054335836_1
591001_000025305_1_Afermentans_@WP_012938708_1
568816_000230275_1_Aintestini_@WP_009014761_1
572479_000165465_1_Hpraevalens_@WP_014553011_1
656519_000166415_1_Hhydrogenifor_@WP_013405970_1
373903_000020485_1_Horenii_@WP_012635514_1
574087_000144695_1_Aarabaticum_@WP_013277398_1
748449_000328625_1_Hhalobius_@WP_015327680_1
292459_000009905_1_Sthermophilum_@WP_011195253_1
292459_000009905_1_Sthermophilum_@WP_011194571_1
644966_000184705_1_Tmarianensis_@WP_013495260_1
1209989_000213235_1_Tacetatoxydan_@WP_013777793_1
555079_000144645_1_Toceani_@WP_013276177_1
1042163_000219535_3_Blaterosporus_@AIG25058_1
697284_000511405_1_Plarvae_@WP_024093018_1
358681_000010165_1_Bbrevis_@WP_012686271_1
1192197_000271665_2_Pfermentans_@WP_007956588_1
484770_000725345_1_Psp_@WP_038670992_1
927704_000284095_1_Sruminantium_@WP_014425150_1
546271_000208405_1_Ssputigena_@WP_013740757_1
712538_001189555_1_Ssp_@WP_050344267_1
484770_000725345_1_Psp_@WP_038667829_1
1192197_000271665_2_Pfermentans_@WP_007961580_1
1458465_001304715_1_Melsdenii_@WP_022498085_1
479436_000024945_1_Vparvula_@WP_012864816_1
1051632_000219855_1_Sacidophilus_@WP_013986119_1
1048834_000219875_1_Aacidocaldari_@WP_014464381_1
591001_000025305_1_Afermentans_@WP_012937979_1
568816_000230275_1_Aintestini_@WP_009016373_1
1051632_000219855_1_Sacidophilus_@WP_013986168_1
1051632_000219855_1_Sacidophilus_@WP_013986151_1
644966_000184705_1_Tmarianensis_@WP_013496715_1
1051632_000219855_1_Sacidophilus_@WP_013988640_1
696281_000215085_1_Druminis_@WP_013841711_1
868595_000214435_1_Dnigrificans_@WP_013810055_1
349161_000016165_1_Dreducens_@WP_011878423_1
760568_000214705_1_Dkuznetsovii_@WP_013822554_1
485916_000024205_1_Dacetoxidans_@WP_015756897_1
370438_000010565_1_Pthermopropio_@WP_012032184_1
635013_000092945_1_Tpotens_@WP_013121757_1
643648_000092405_1_Slipocalidus_@WP_013175537_1
335541_000014725_1_Swolfei_@WP_011640342_1
1048834_000219875_1_Aacidocaldari_@WP_041695531_1
562970_000092905_1_Ktusciae_@WP_013077170_1
189425_000758705_1_Pgraminis_@WP_042265826_1
1073571_000981585_1_Priograndensi_@CQR51257_1
160799_000758665_1_Pborealis_@WP_042209979_1
189426_000758725_1_Podorifer_@WP_038568066_1
1268072_000612505_1_Psabinae_@WP_025332650_1
169760_000758685_1_Pstellifer_@WP_038692950_1
44251_000756615_1_Pdurus_@WP_042204622_1
59893_001272655_1_Ppeoriae_@WP_053324351_1
1406_000819665_1_Ppolymyxa_@WP_040102423_1
985665_000235585_1_Pterrae_@WP_014279114_1
481743_000024685_1_Psp_@WP_012818359_1
162209_001465255_1_Pnaphthalenov_@WP_062410304_1
1036673_000218915_1_Pmucilaginosu_@WP_013921293_1
697284_000511405_1_Plarvae_@WP_042119861_1
1126833_000961095_1_Pbeijingensis_@WP_045670424_1
717605_000227705_3_Tcomposti_@AGA59590_1
358681_000010165_1_Bbrevis_@WP_015893677_1
1042163_000219535_3_Blaterosporus_@AIG28745_1
1508404_000818095_1_Jmalaysiensis_@WP_052268762_1
272558_000011145_1_Bhalodurans_@WP_010899937_1
398511_000005825_2_Bpseudofirmus_@WP_012960757_1
66692_000009825_1_Bclausii_@PSDG1_4058501_1e_165
1246626_000706725_1_Blehensis_@WP_051667684_1
649639_000177235_2_Bcellulosilyt_@WP_013490679_1
28031_000724775_3_Lfusiformis_@WP_036122944_1
444177_000017965_1_Lsphaericus_@WP_012292711_1
1145276_000600105_1_Lvarians_@WP_025218563_1
1002809_000271325_1_Ssilvestris_@WP_014824750_1
1508404_000818095_1_Jmalaysiensis_@WP_052268763_1
866895_000284515_1_Hhalophilus_@WP_014644874_1
698758_000307165_1_Axylanus_@WP_015010791_1
586416_000725365_1_Taidingensis_@WP_038564014_1
221109_000011245_1_Oiheyensis_@WP_011067413_1
403957_000725285_1_Vsp_@WP_040956582_1
1111068_000236605_1_Gthermoleovor_@WP_014196978_1
235909_000009785_1_Gkaustophilus_@WP_011232873_1
272567_001274575_1_Gstearothermo_@WP_053414187_1
420246_000015745_1_Gthermodenitr_@WP_008880740_1
581103_000166075_1_Gsp_@WP_013401861_1
1426_001295365_1_Gthermoglucos_@WP_042384314_1
198467_001187595_1_Agonensis_@WP_035064428_1
491915_000019045_1_Aflavithermus_@WP_041639168_1
637380_000143605_1_Bcereus_@WP_000372992_1
261591_000742895_1_Banthracis_@WP_000372993_1
412694_000015065_1_Bthuringiensi_@WP_000372996_1
1415784_000496285_1_Btoyonensis_@PSDG1_2504724_0_0
1330043_000831065_1_Bbombysepticu_@WP_000372962_1
1405_000832605_1_Bmycoides_@WP_002184502_1
86662_000775975_1_Bweihenstepha_@WP_002034921_1
315749_000017425_1_Bcytotoxicus_@WP_012096333_1
796606_000724485_1_Bmethanolicus_@WP_003346888_1
1367477_000473245_1_Binfantis_@WP_022544514_1
1398_001039495_1_Bcoagulans_@WP_035183221_1
1479_001050115_1_Bsmithii_@WP_048624263_1
1006007_000225265_1_Bmegaterium_@WP_014457735_1
135735_000972245_2_Bendophyticus_@WP_040060369_1
1408_000972685_1_Bpumilus_@WP_017368234_1
279010_000008425_1_Blicheniformi_@WP_003186083_1
1648923_000876525_1_Bparalichenif_@WP_023857049_1
1574141_000877815_1_Bsp_@WP_043940226_1
224308_000009045_1_Bsubtilis_@NP_391631_2
1452_000830075_1_Batrophaeus_@WP_003326485_1
1126211_000262385_1_Bamyloliquefa_@WP_014419133_1
492670_000988345_1_Bvelezensis_@WP_038464365_1
0.4

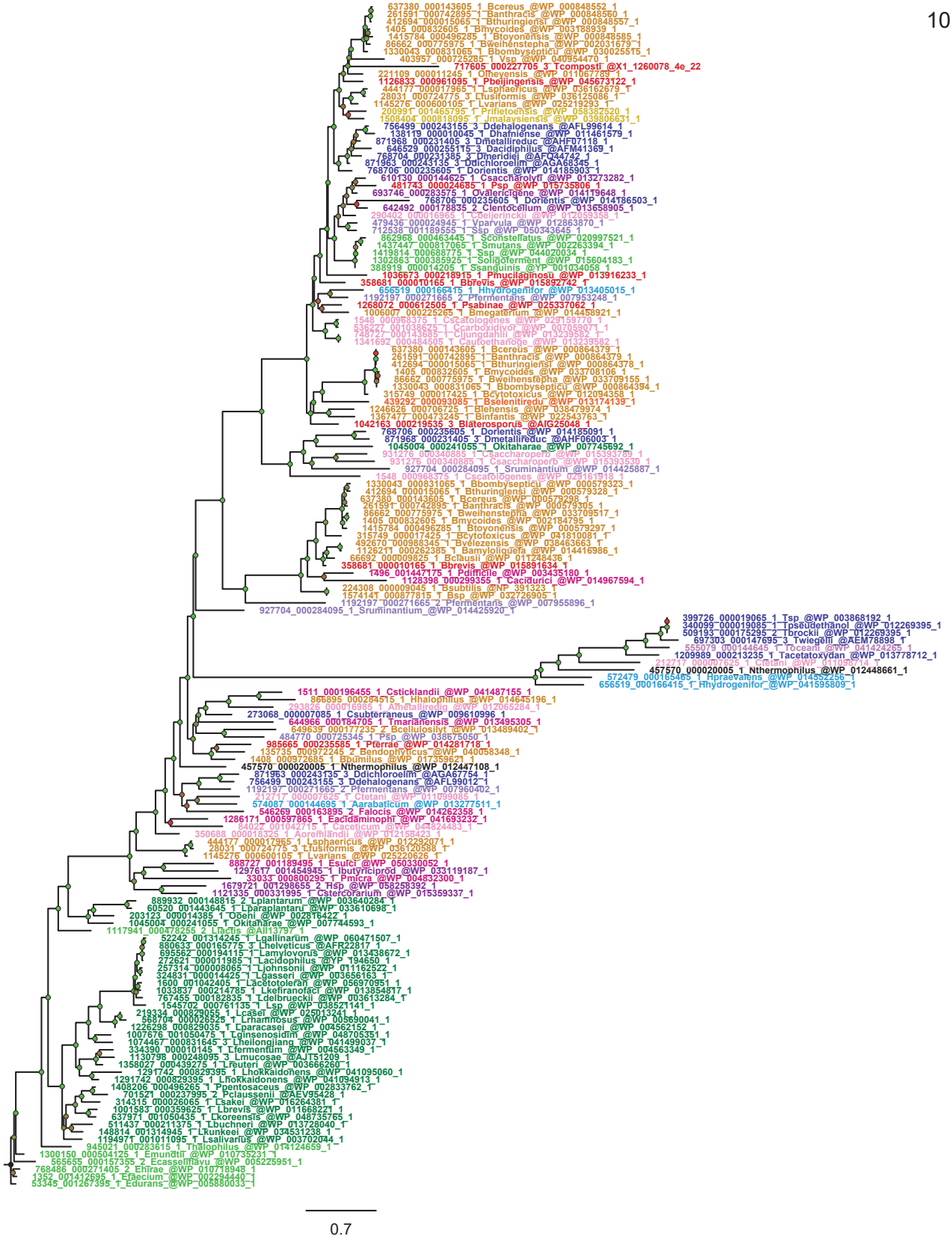

Erhusiopathia_@WP_013852691_1
357809_000018685_1_Lphytoferment_@WP_012198983_1
0.4

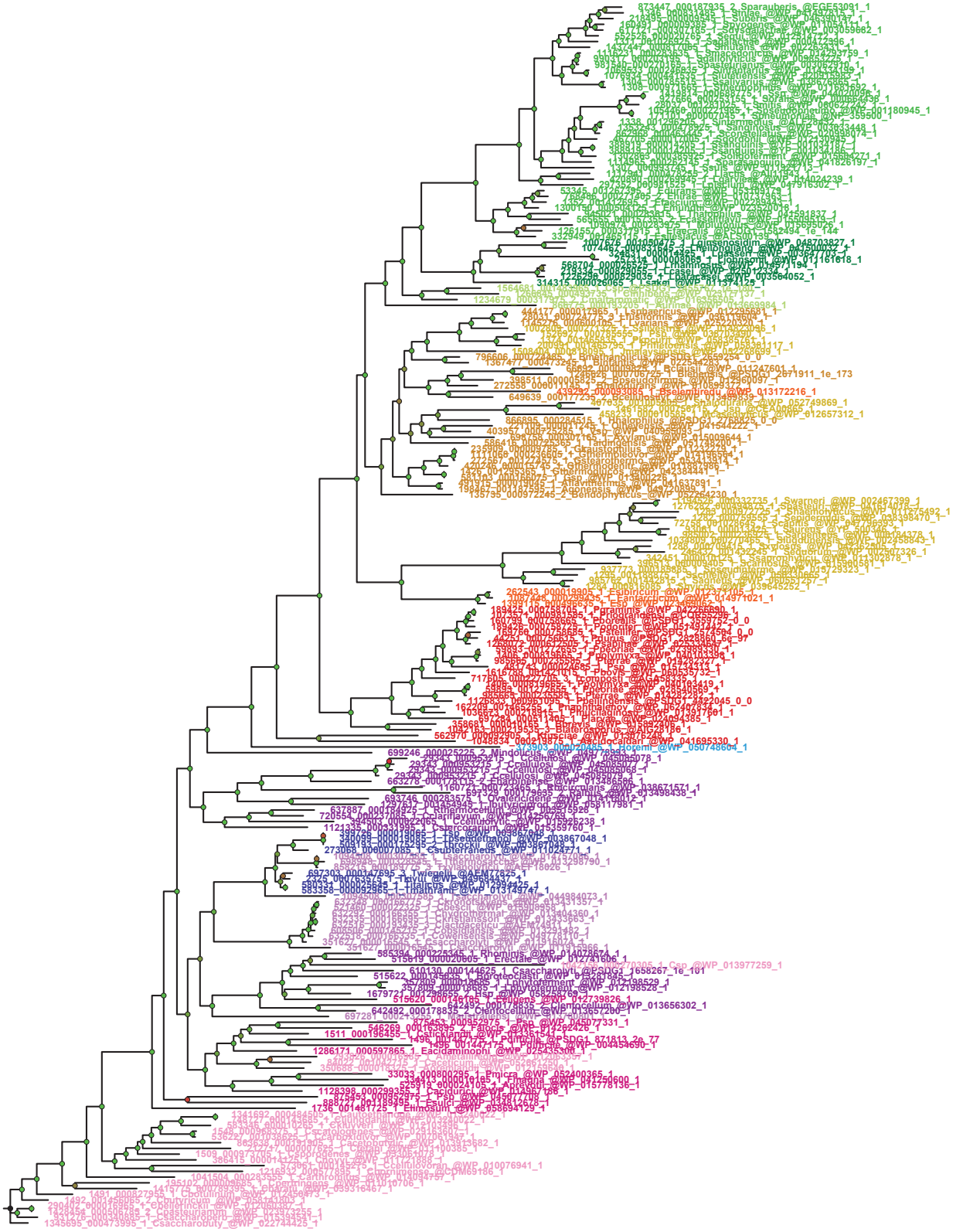

509193_000175295_2_Tbrockii_@WP_012269292_1
340099_000019085_1_Tpseudethanol_@WP_012269292_1
399726_000019065_1_Tsp_@WP_009052308_1
697303_000147695_3_Twiegelii_@AEM78613_1
580331_000025645_1_Titalicus_@WP_012995105_1
583356_000092965_1_Tmathranii_@WP_013150242_1
273068_000007085_1_Csubterraneus_@PSDG1_1243547_4e-73
2325_000763575_1_Tkivui_@PSDG1_1117422_2e-60
931626_000247605_1_Awoodii_@WP_014356063_1
1736_001481725_1_Elimosum_@WP_058693840_1
1548_000968375_1_Cscatalogenes_@WP_029152147_1
420246_000015745_1_Gthermodenitr_@WP_008880524_1
888727_001189495_1_Esulci_@WP_034813185_1
1496_001447175_1_Pdifficile_@WP_058085504_1
1511_000196455_1_Csticklandii_@WP_013361501_1
1509_000973705_1_Csporogenes_@WP_003496746_1
84022_001042715_1_Caceticum_@WP_044823922_1
350688_000018325_1_Agremiandii_@WP_012159545_1
386415_000014125_1_Cnovyi_@WP_011722187_1
293826_000016985_1_Ametalliredig_@WP_041720726_1
572479_000165465_1_Hpraevalens_@WP_014552413_1
656519_000166415_1_Hhydrogenifor_@WP_013406650_1
572479_000165465_1_Hpraevalens_@WP_014553694_1
373903_000020485_1_Horenii_@WP_015923831_1
457570_000020005_1_Nthermophilus_@WP_012447662_1
555079_000144645_1_Toceani_@WP_013276271_1
484770_000725345_1_Psp_@WP_038666685_1
1192197_000271665_2_Ptermentans_@WP_007959789_1
1126833_000961095_1_Pbeijingensis_@WP_052702884_1
1192197_000271665_2_Ptermentans_@WP_007952716_1
646529_000255115_3_Dacidiphilus_@AFM40856_1
748449_000328625_1_Rhalobius_@WP_015325962_1
574087_000144695_1_Aarabaticum_@WP_013278667_1
546269_000163895_2_Falocis_@WP_014261846_1
1511_000196455_1_Csticklandii_@WP_013360268_1
693746_000283575_1_Ovalericigene_@WP_014119136_1
491915_000019045_1_Aflavithermus_@WP_012575589_1
198467_001187595_1_Agonensis_@WP_009362481_1
1398_001039495_1_Bcoagulans_@WP_014096651_1
1284_000816085_1_Shyicus_@WP_052257867_1
985762_001442815_1_Sagnetis_@WP_060552502_1
1295_001188875_1_Sschleiferi_@WP_050337515_1
937773_000185885_1_Spseuditerme_@WP_015729484_1
1288_000709415_1_Sxylosus_@WP_042362279_1
246432_001432245_1_Sequorum_@WP_059935813_1
342451_000010125_1_Ssaprophyticu_@WP_011302623_1
396513_000009405_1_Scarnosus_@WP_041613127_1
93061_000013425_1_Saureus_@YP_500902_1
985002_000236925_1_Sargenteus_@WP_000951648_1
1282_000759555_1_Sepidermidis_@WP_001829764_1
72758_001028645_1_Scapitis_@WP_047796550_1
1283_000972725_1_Shaemolyticus_@WP_011275174_1
1034809_000270465_1_Slugdunensis_@WP_002478919_1
1276282_000494875_1_Spasteuri_@WP_023734434_1
1194526_000332735_1_Swarneri_@WP_002466487_1
768706_000235605_1_Dorientis_@WP_014187722_1
768704_000231385_3_Dmeridiei_@AFQ43387_1
646529_000255115_3_Dacidiphilus_@AFM43486_1
645991_000190635_1_Sglycolicus_@WP_013626285_1
138119_000010045_1_Dhafniense_@WP_011462255_1
871963_000243135_3_Ddichloroelim_@AGA70851_1
756499_000243155_3_Ddehalogenans_@AFM02415_1
871968_000231405_3_Dmetallireduc_@AHF08295_1
1147129_000305775_1_Dsp_@WP_015044936_1
871738_000512895_1_Dreductus_@WP_019225009_1
498761_000019165_1_Hmodesticaldu_@WP_049754034_1
1089553_000305935_1_Tphaeum_@WP_015051809_1
1525_001267405_1_Mthermoacetic_@WP_011393884_1
335541_000014725_1_Swolfei_@WP_011641789_1
643648_000092405_1_Slipocalidus_@WP_013176455_1
29343_000953215_1_Ccellulosi_@WP_045085420_1
663278_000171115_2_Eharbinense_@WP_013486039_1
1297617_001454945_1_Ibutyriciprod_@WP_058116806_1
555079_000144645_1_Toceani_@WP_013275123_1
697281_000213255_1_Maustraliensi_@WP_013781817_1
1041604_000283555_1_Carthromitus_@WP_014094184_1
195102_000009685_1_Cperfringens_@WP_011010210_1
644966_000184705_1_Tmarianensis_@WP_013494952_1
57306?_000145275_1_Ccellulovoran_@WP_010073344_1
863638_000191905_1_Cacetobutylic_@WP_010986603_1
546269_000163895_2_Falocis_@WP_041250813_1
937773_000185885_1_Spseuditerme_@WP_015729107_1
1295_001188875_1_Sschleiferi_@WP_050330925_1
1034809_000270465_1_Slugdunensis_@WP_002478211_1
1284_000816085_1_Shyicus_@WP_039645820_1
985762_001442815_1_Sagnetis_@WP_060551032_1
93061_000013425_1_Saureus_@YP_499927_1
985002_000236925_1_Sargenteus_@WP_000127609_1
1276282_000494875_1_Spasteuri_@WP_041614009_1
1194526_000332735_1_Swarneri_@WP_002466849_1
1283_000972725_1_Shaemolyticus_@WP_011275804_1
72758_001028645_1_Scapitis_@WP_047796261_1
1282_000759555_1_Sepidermidis_@WP_038398352_1
1288_000709415_1_Sxylosus_@WP_042362718_1
246432_001432245_1_Sequorum_@WP_002507641_1
342451_000010125_1_Ssaprophyticu_@WP_011303130_1
396513_000009405_1_Scarnosus_@WP_015900292_1
458233_000010585_1_Mcasseolyticus_@WP_012656959_1
407035_001005905_1_Shalodurans_@WP_046790320_1
1461582_000756715_2_Jsp_@CEA01026_1
262543_000019905_1_Esibiricum_@WP_012369397_1
1087448_000299435_1_Eantarcticum_@WP_041255911_1
1399115_000496635_1_Esp_@WP_023467125_1
768706_000235605_1_Dorientis_@WP_042331029_1
768704_000231385_3_Dmeridiei_@AFQ43362_1
279010_000008425_1_Blicheniformi_@WP_003179586_1
1648923_000876525_1_Bparalicheni_@WP_026578931_1
292459_000009905_1_Sthermophilum_@WP_011194687_1
698948_000328545_1_Tthermosaccha_@WP_015311852_1
399726_000019065_1_Tsp_@WP_009051931_1
1209989_000213235_1_Tacetatoxydan_@WP_013779075_1
66692_000009825_1_Bclausii_@WP_011245532_1
272558_000011145_1_Bhalodurans_@WP_010896768_1
610130_000144625_1_Csaccharolyti_@WP_013272687_1
1509_000973705_1_Csporogenes_@WP_033059738_1
332949_001465115_1_Esilesiacus_@ALS01033_1
1564681_001483965_1_Csp_@WP_058818828_1
1074467_000831645_3_Lheilongjiang_@WP_041500206_1
1209989_000213235_1_Tacetatoxydan_@WP_013777186_1
555079_000144645_1_Toceani_@WP_013275590_1
697303_000147695_3_Twiegelii_@AEM78262_1
1496_001447175_1_Pdifficile_@WP_003436153_1
293826_000116915_1_Ametalliredig_@WP_041720722_1
1036673_000218915_1_Pmucilaginosu_@WP_013920600_1
610130_000144625_1_Csaccharolyti_@WP_013273579_1
169760_000758685_1_Pstellifer_@WP_038697523_1
562970_000092905_1_Ktusciae_@WP_013074551_1
162209_001465255_1_Pnaphthalenov_@WP_062409085_1
649639_000177235_2_Bcellulosilyt_@WP_013487403_1
412694_000015065_1_Bthuringiensi_@WP_000373773_1
261591_000742895_1_Banthracis_@WP_000373772_1
637380_000143605_1_Bcereus_@WP_000373742_1
1415784_000496285_1_Btoyonensis_@WP_000373788_1
86662_000775975_1_Bweihenstepha_@WP_002029787_1
1405_000832605_1_Bmycoides_@WP_002166959_1
1330043_000831065_1_Bbombysepticu_@WP_000373743_1
315749_000017425_1_Bcytotoxicus_@WP_011983674_1
1042163_000219535_3_Blatersporus_@AIG26342_1
481743_000024685_1_Psp_@WP_015737984_1
358681_000010165_1_Bbrevis_@WP_015893747_1
28031_000724775_3_Lfusiformis_@WP_036126743_1
444177_000017965_1_Lsphaericus_@WP_012294948_1
1145276_000600105_1_Lvariians_@WP_025219869_1
1526927_000785555_1_Psp_@WP_038703169_1
221109_000011245_1_Oiheyensis_@WP_011066111_1
457570_000020005_1_Nthermophilus_@WP_012448364_1
1126833_000961095_1_Pbeijingensis_@WP_052703006_1
1126833_000961095_1_Pbeijingensis_@WP_045672235_1
1648923_000876525_1_Bparalicheni_@WP_025811112_1
279010_000008425_1_Blicheniformi_@WP_009327843_1
1408_000972685_1_Bpumilus_@WP_046344949_1
407035_001005905_1_Shalodurans_@WP_046789465_1
1286171_000597865_1_Eacidaminophi_@WP_025436792_1
1461582_000756715_2_Jsp_@CEA01603_1
0.5

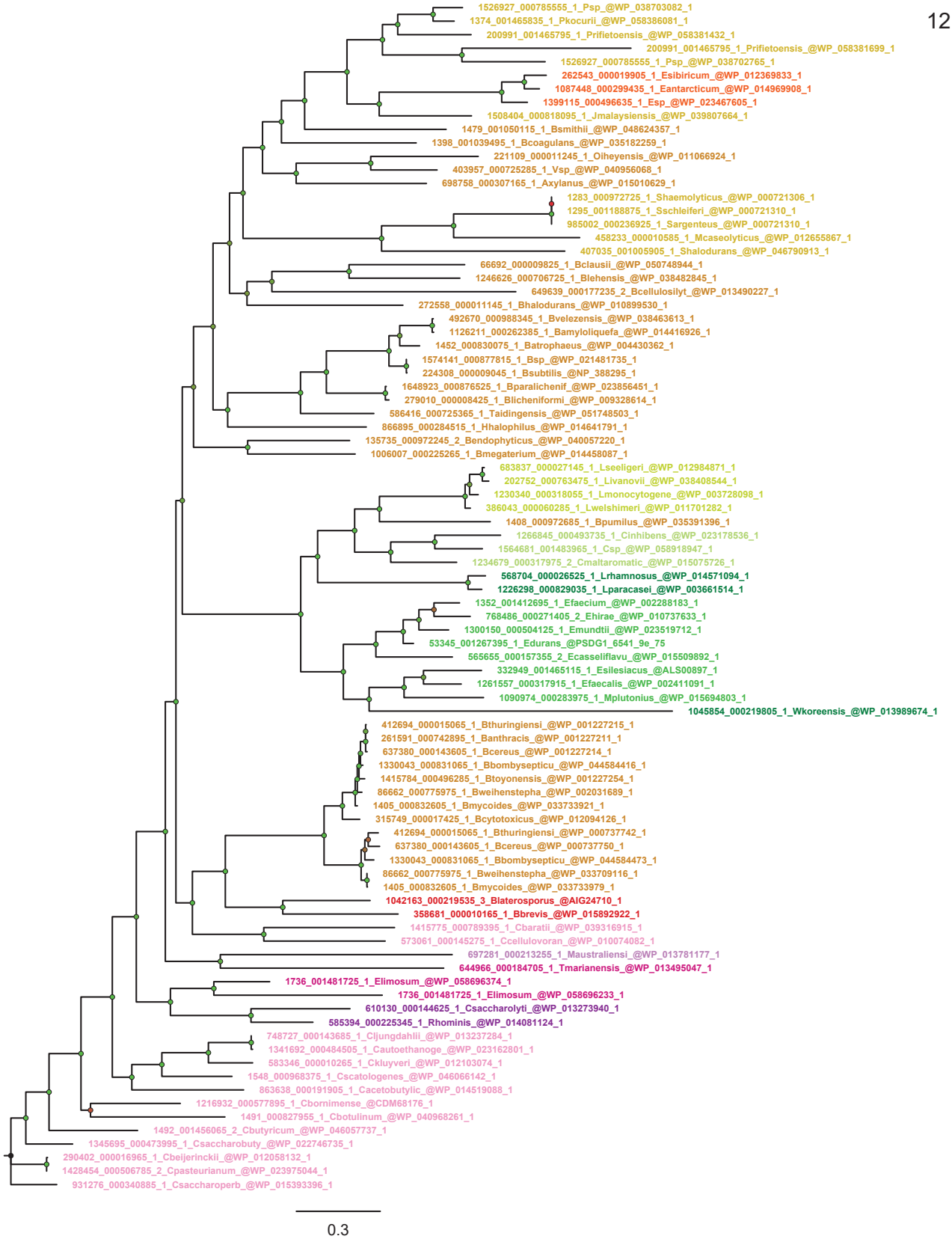

0.4

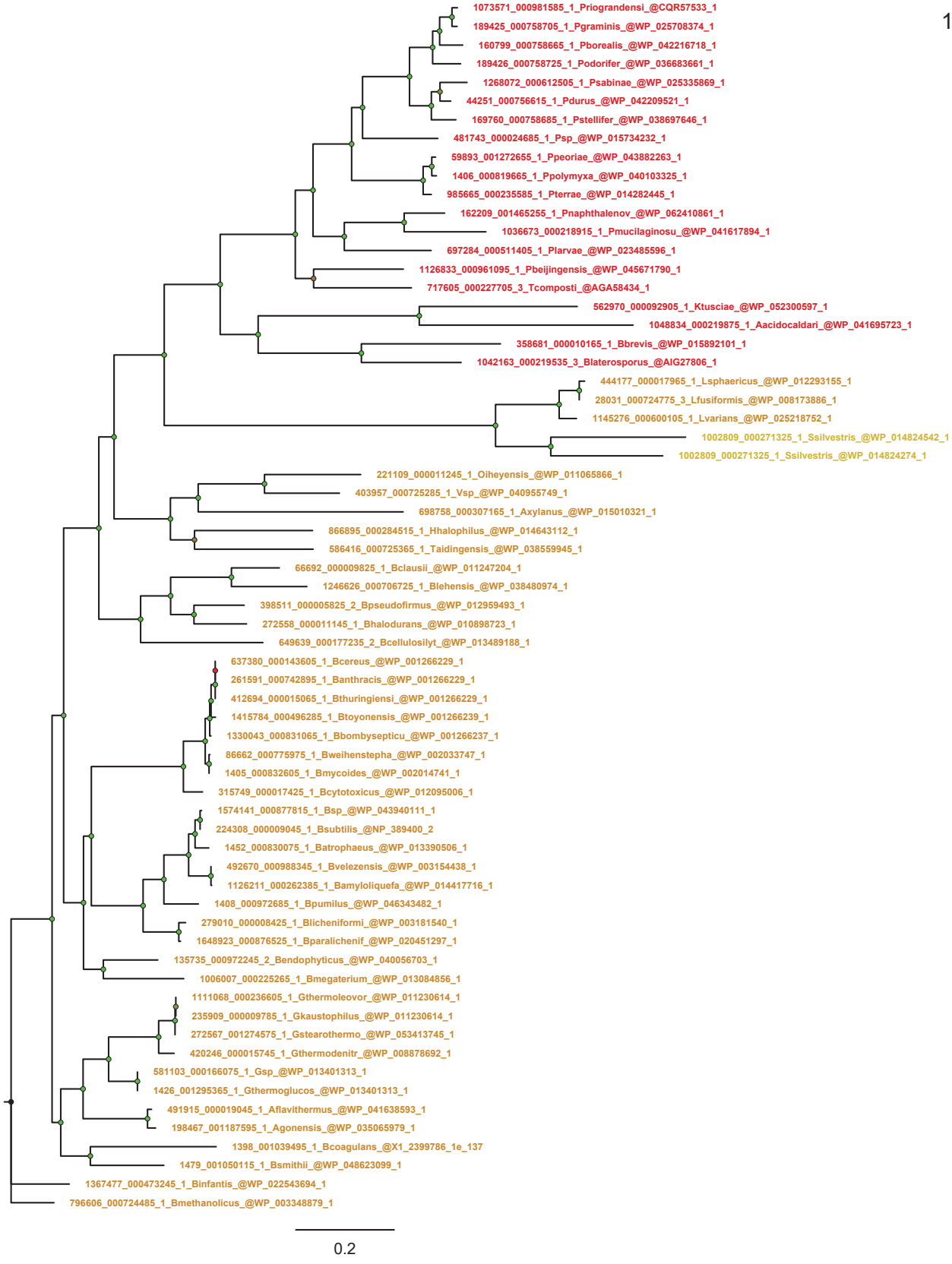

1268072_000612505_1_Psabinae_@WP_025334024_1
44251_000756615_1_Pdurus_@WP_042205890_1
1126833_000961095_1_Pbeijingensis_@WP_045673320_1
481743_000024685_1_Psp_@WP_015737549_1
697284_000511405_1_Plarvae_@WP_024093052_1
1051632_000219855_1_Sacidophilus_@WP_013986324_1
439292_000093085_1_Bselenitiredu_@WP_013171172_1
273068_000007085_1_Csubterraneus_@WP_009609657_1
583358_000092965_1_Tmathranii_@WP_013149698_1
1297617_001454945_1_Ibutyriciprod_@WP_033116496_1
357809_000018685_1_Lphytoferment_@WP_012199147_1
555079_000144645_1_Toceani_@WP_041423824_1
293826_000016985_1_Ametalliredig_@WP_012065212_1
1234679_000317975_2_Cmaltaromatic_@WP_015075563_1
332949_001465115_1_Esilesiacus_@ALS00084_1
297352_000981525_1_Lpiscium_@WP_047916420_1
650150_000270085_1_Erhusiopathia_@WP_013852964_1
565655_000157355_2_Ecasseliflavu_@WP_015509985_1
1496_001447175_1_Pdifficile_@WP_016729003_1
1492_001456065_2_Cbutyricum_@WP_058372221_1
1042163_000219535_3_Blaterosporus_@AIG28090_1
637380_000143605_1_Bcereus_@WP_000274993_1
412694_000015065_1_Bthuringiensi_@WP_000274994_1
261591_000742895_1_Banthracis_@WP_000274995_1
1330043_000831065_1_Bbombysepticu_@WP_044584460_1
86662_000775975_1_Bweihenstepha_@WP_033709639_1
1405_000832605_1_Bmycoides_@WP_033734653_1
1006007_000225265_1_Bmegaterium_@WP_014461295_1
0.3

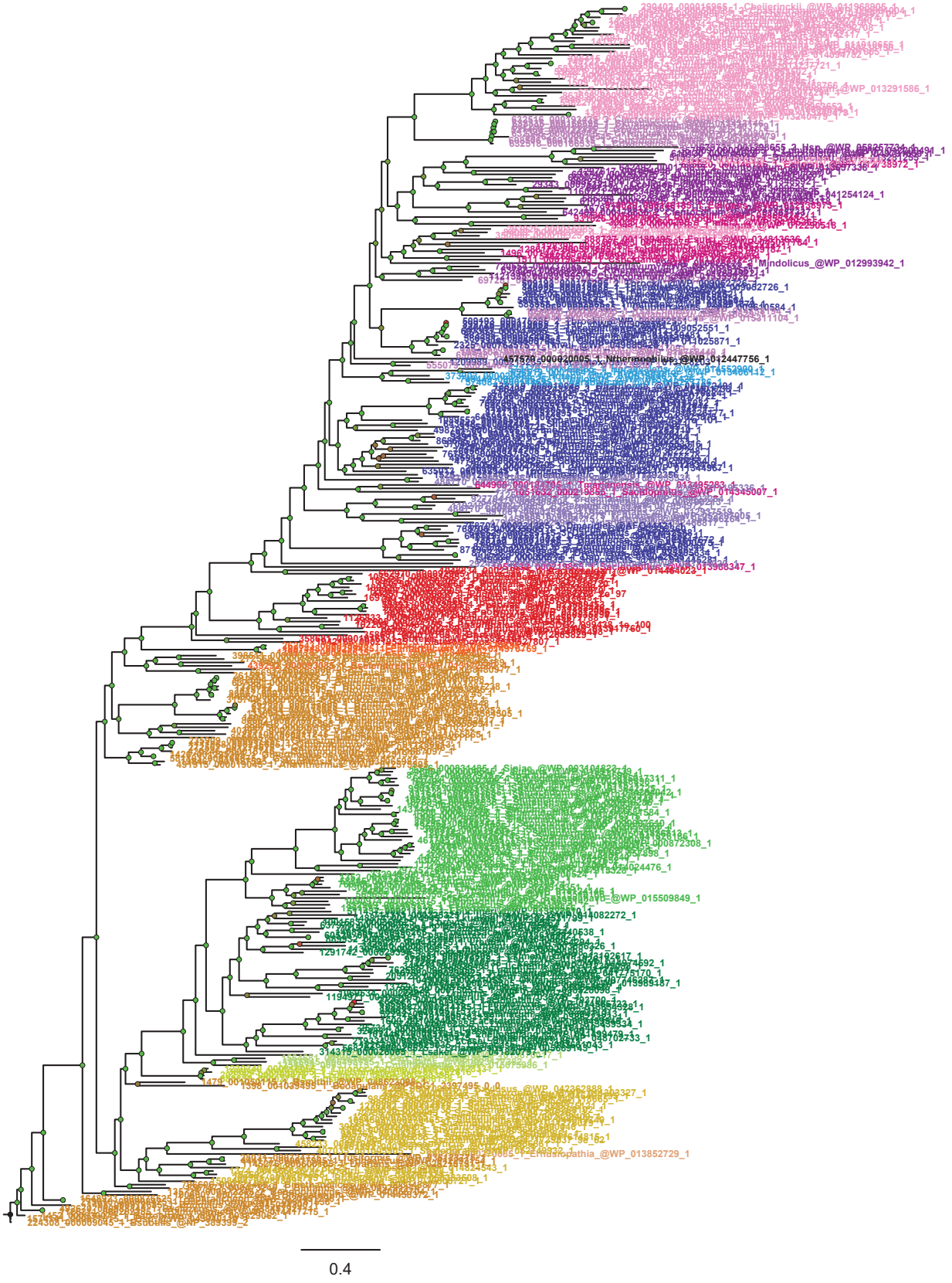

148814_001314945_1_Lkunkeei_@WP_034531239_1
511437_000211375_1_Lbuchneri_@WP_013728041_1
314315_000026065_1_Lsakei_@WP_011373811_1
1194971_001011095_1_Lsalivarius_@WP_003702051_1
1001583_000359625_1_Lbrevis_@WP_041815400_1
637971_001050435_1_Lkoreensis_@WP_048735764_1
701521_000237995_2_Pclaussenii_@AEV95429_1
1408206_000496265_1_Ppentosaceus_@WP_023440228_1
1007676_001050475_1_Lginsenosidim_@WP_048704538_1
1074467_000831645_3_Lheilongjiang_@WP_048699129_1
568704_000026525_1_Lrhamnosus_@PSDG1_2190014_1e_162
219334_000829055_1_Lcasei_@WP_010491799_1
1226298_000829035_1_Lparacasei_@WP_003580413_1
52242_001314245_1_Lgallinarum_@WP_060471263_1
880633_000165775_3_Lhelveticus_@AFR21209_1
1033837_000214785_1_Lkefiranofaci_@WP_013855162_1
695562_000194115_1_Lamylovorus_@WP_013437001_1
272621_000011985_1_Lacidophilus_@YP_193091_1
1545702_000761135_1_Lsp_@WP_038524311_1
1600_001042405_1_Lacetotoleran_@WP_060459877_1
257314_000008065_1_Ljohnsonii_@WP_004898587_1
324831_000014425_1_Lgasseri_@WP_003647950_1
767455_000182835_1_Ldelbrueckii_@WP_013439010_1
767455_000182835_1_Ldelbrueckii_@WP_013439060_1
525919_000024105_1_Aprevotii_@WP_015778580_1
1130798_000248095_3_Lmucosae_@AJT49645_1
1358027_000439275_1_Lreuteri_@WP_019253427_1
334390_000010145_1_Lfermentum_@WP_024501089_1
1511_000196455_1_Csticklandii_@WP_013361806_1
350688_000018325_1_Aoremlandii_@WP_012158175_1
866895_000284515_1_Hhalophilus_@WP_014645198_1
546269_000163895_2_Falocis_@WP_014262638_1
33033_000800295_1_Pmicra_@WP_041953594_1
875453_000952975_1_Psp_@WP_045076831_1
457570_000020005_1_Nthermophilus_@WP_012449260_1
1126211_000262385_1_Bamyloliquefa_@WP_014419351_1
492670_000988345_1_Bvelezensis_@WP_038460547_1
1126211_000262385_1_Bamyloliquefa_@WP_014419350_1
1006007_000225265_1_Bmegaterium_@WP_014462109_1
1374_001465835_1_Pkocurii_@WP_058385717_1
1526927_000785555_1_Psp_@WP_038703535_1
200991_001465795_1_Prifietoensis_@WP_058381077_1
1002809_000271325_1_Ssilvestris_@WP_014823014_1
28031_000724775_3_Lfusiformis_@WP_008176723_1
444177_000017965_1_Lsphaericus_@WP_012295704_1
1145276_000600105_1_Lvarians_@WP_025220336_1
444177_000017965_1_Lsphaericus_@WP_012292070_1
28031_000724775_3_Lfusiformis_@WP_036120591_1
1145276_000600105_1_Lvarians_@WP_025220627_1
28031_000724775_3_Lfusiformis_@WP_042737542_1
444177_000017965_1_Lsphaericus_@WP_036160971_1
1297617_001454945_1_Ibutyriciprod_@WP_058116794_1
1117941_000478255_2_Llactis_@AII13798_1
945021_000283615_1_Thalophilus_@WP_050989762_1
565655_000157355_2_Ecasseliflavu_@WP_034870000_1
1300150_000504125_1_Emundtii_@WP_041684221_1
768486_000271405_2_Ehirae_@WP_010718947_1
1352_001412695_1_Efaecium_@WP_002294441_1
53345_001267395_1_Edurans_@WP_005880035_1
0.5

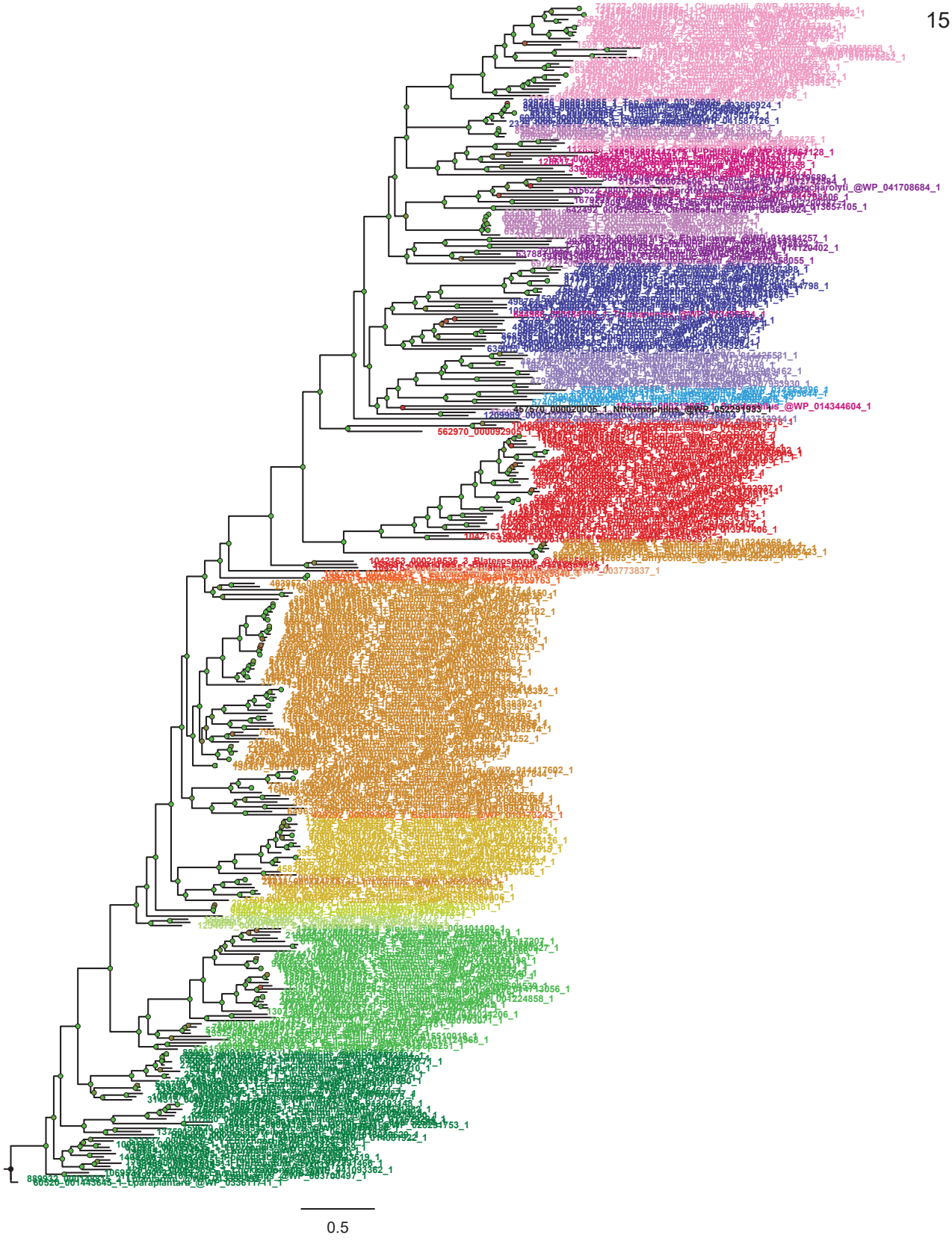

637380_000143605_1_Bcereus_@WP_000848552_1
261591_000742895_1_Banthracis_@WP_000848560_1
412694_000015065_1_Bthuringiensi_@WP_000848557_1
1405_000832605_1_Bmycoides_@WP_003788939_1
1415784_000496285_1_Btoyonensis_@WP_000848585_1
86662_000775975_1_Bweihenstephа_@WP_002031679_1
1330043_000831065_1_Bbombysepticu_@WP_030025515_1
403957_000725285_1_Vsp_@WP_040954470_1
717605_000227705_3_Tcompostl_@X1_1260078_4e-22
221109_000011245_1_Otheyensis_@WP_011067782_1
1126833_000961095_1_Pbeijingensis_@WP_045673122_1
444177_000017965_1_Lsphaericus_@WP_036162679_1
28031_000724775_3_Lfusiformis_@WP_036126086_1
1145276_000600105_1_Lvarians_@WP_025219293_1
200991_001465795_1_Prifetoensis_@WP_058382520_1
1508404_000818085_1_Jmalaysiensis_@WP_039806631_1
756499_000243155_3_Ddehalogenans_@AFL99614_1
138119_000010045_1_Dhafniense_@WP_011461579_1
871968_000231405_3_Dmetallireduc_@AHF07118_1
646529_000255115_3_Dacidiphilus_@AFM41369_1
768704_000231385_3_Dmeridiei_@AFQ04742_1
871963_000243135_3_Ddichloroelim_@AGA68345_1
768706_000235605_1_Dorientis_@WP_014185903_1
610130_000144625_1_Csaccharolyti_@WP_013273282_1
481743_000024685_1_Psp_@WP_015735806_1
693746_000283575_1_Ovalericigene_@WP_014119648_1
768706_000235605_1_Dorientis_@WP_014186503_1
642492_000178835_2_Clentocellum_@WP_013658905_1
290402_000016965_1_Cbeijerinckii_@WP_012059358_1
479436_000024945_1_Vparvula_@WP_012863870_1
712538_001189555_1_Ssp_@WP_050343645_1
862968_000463445_1_Sconstellatus_@WP_020997521_1
1437447_000817065_1_Smutans_@WP_002263394_1
1419814_000688775_1_Ssp_@WP_044020034_1
1302863_000385925_1_Soligoferment_@WP_015604183_1
388919_000014205_1_Ssanguinis_@YP_001034058_1
1036573_000219915_1_Pmucilaginosu_@WP_013916233_1
358681_000010165_1_Bbrevis_@WP_015892742_1
656519_000166415_1_Hhydrogenifor_@WP_013405015_1
1192197_000271665_2_Pfermentans_@WP_007953248_1
1268072_000612505_1_Psabinae_@WP_025337062_1
1006007_000225265_1_Bmegaterium_@WP_014458921_1
1548_000958375_1_Cscatalogenes_@WP_029159770_1
527447_001139675_1_Ccarboxidivor_@WP_007059071_1
748727_001143885_1_Cjungdahlii_@WP_013239552_1
1341692_000484505_1_Cautoethanoge_@WP_013239582_1
637380_000143605_1_Bcereus_@WP_000864379_1
261591_000742895_1_Banthracis_@WP_000864379_1
412694_000015065_1_Bthuringiensi_@WP_000864378_1
1405_000832605_1_Bmycoides_@WP_033708106_1
86662_000775975_1_Bweihenstephа_@WP_033709155_1
1330043_000831065_1_Bbombysepticu_@WP_000864394_1
315749_000017425_1_Bcytotoxicus_@WP_012094358_1
439292_000093085_1_Bselenitiredu_@WP_013174139_1
1246626_000706725_1_Blehensis_@WP_038479974_1
1367477_000473245_1_Binfantis_@WP_022543763_1
1042163_000219535_3_Blaterosporus_@AIG25048_1
768706_000235605_1_Dorientis_@WP_014185091_1
871968_000231405_3_Dmetallireduc_@AHF06003_1
1045004_000241055_1_Okitaharae_@WP_007745892_1
931276_000340885_1_Csaccharoperb_@WP_015393539_1
931276_000340885_1_Csaccharoperb_@WP_015393530_1
927704_000284095_1_Sruminantium_@WP_014425887_1
1548_000968375_1_Cscatalogenes_@WP_029161918_1
1330043_000831065_1_Bbombysepticu_@WP_000579323_1
412694_000015065_1_Bthuringiensi_@WP_000579328_1
637380_000143605_1_Bcereus_@WP_000579298_1
261591_000742895_1_Banthracis_@WP_000579305_1
86662_000775975_1_Bweihenstephа_@WP_033709517_1
1405_000832605_1_Bmycoides_@WP_002184795_1
1415784_000496285_1_Btoyonensis_@WP_000579297_1
315749_000017425_1_Bcytotoxicus_@WP_041810081_1
492670_000988345_1_Bvelezensis_@WP_038463663_1
1126211_000262385_1_Bamyloliquefa_@WP_014416986_1
66692_000009825_1_Bclausii_@WP_011284436_1
358681_000010165_1_Bbrevis_@WP_015891634_1
1496_001447175_1_Pdifficile_@WP_003435180_1
1128398_000299355_1_Caciduric_@WP_014967594_1
224308_000009045_1_Bsubtilis_@NP_391323_1
1574141_000877815_1_Bsp_@WP_032726905_1
1192197_000271665_2_Pfermentans_@WP_007955896_1
927704_000284095_1_Sruminantium_@WP_014425920_1
399726_000019065_1_Tsp_@WP_003868192_1
340099_000019085_1_Tpseudethanol_@WP_012269395_1
509193_000175295_2_Tbrockii_@WP_012269395_1
697303_000147695_3_Twiegelii_@AEM78898_1
555079_000144645_1_Toceani_@WP_041424265_1
1209989_000213235_1_Tacetatoxydan_@WP_013778712_1
212717_000007625_1_Ctetani_@WP_011039085_1
457570_000020005_1_Nthermophilus_@WP_012448661_1
572479_000165465_1_Hpraevalens_@WP_014552256_1
656519_000166415_1_Hhydrogenifor_@WP_041595809_1
1511_000196455_1_Csticklandii_@WP_041487155_1
866895_000284515_1_Hhalophilus_@WP_014645196_1
263525_000016985_1_Ametallirubi_@WP_012065284_1
273068_000007085_1_Csubterraneus_@WP_009610996_1
644966_000184705_1_Tmarianensis_@WP_013495305_1
649639_000177235_2_Bcellulosilyt_@WP_013489402_1
484770_000725345_1_Psp_@WP_038675050_1
985665_000235585_1_Pterrae_@WP_014281718_1
135735_000972245_2_Bendophyticus_@WP_040058348_1
1408_000972685_1_Bpumilus_@WP_017359621_1
457570_000020005_1_Nthermophilus_@WP_012447108_1
871963_000243135_3_Ddichloroelim_@AGA67754_1
756499_000243155_3_Ddehalogenans_@AFL99012_1
1192197_000271665_2_Pfermentans_@WP_007960402_1
212717_000007625_1_Ctetani_@WP_011039085_1
574087_000144695_1_Aarabaticum_@WP_013277511_1
546269_000163895_2_Falocis_@WP_014262358_1
1286171_000597865_1_Eacidaminophi_@WP_041693232_1
84027_001042715_1_Caceticum_@WP_044924483_1
350688_000018325_1_Aoremlandii_@WP_012158423_1
444177_000017965_1_Lsphaericus_@WP_012292071_1
28031_000724775_3_Lfusiformis_@WP_036126598_1
1145276_000600105_1_Lvarians_@WP_025220626_1
888727_001189495_1_Esulci_@WP_050330052_1
1297617_001454945_1_Ibutyriciprod_@WP_033119187_1
33033_000800295_1_Pmicra_@WP_004832300_1
1679721_001298655_2_Hsp_@WP_058258392_1
1121335_000331995_1_Cstercorarium_@WP_015359337_1
889932_000148815_2_Lplantarum_@WP_003640284_1
60520_001443645_1_Lparaplantaru_@WP_033610698_1
203123_000014385_1_Ooeni_@WP_002816422_1
1045004_000241055_1_Okitaharae_@WP_007744593_1
1117941_000478255_2_Llactis_@AII13797_1
52242_001314245_1_Lgallinarum_@WP_060471507_1
880633_000165775_3_Lhelveticus_@AFR22817_1
695562_000194115_1_Lamylovorus_@WP_013438672_1
272621_000011985_1_Lacidophilus_@YP_194650_1
257314_000008065_1_Ljohnsonii_@WP_011162522_1
324831_000014425_1_Lgasseri_@WP_003656163_1
1600_001042405_1_Lacetotolera_@WP_056970951_1
1033837_000214785_1_Lkefiranofac_@WP_013854817_1
767485_000182935_1_Ldelbrueckii_@WP_003613284_1
1545702_000761135_1_Lsp_@WP_038521141_1
219334_000829055_1_Lcase_@WP_025013241_1
56700_000026525_1_Lrhamnosus_@WP_055890041_1
1226298_000629035_1_Lparacasei_@WP_004562152_1
1007676_001050475_1_Lginsenosidim_@WP_048705351_1
1074467_000831645_3_Lheilongjiang_@WP_041499037_1
334390_000010145_1_Lferrenstim_@WP_004563349_1
1130798_000248095_3_Lmucosae_@AJT51209_1
1358027_000439275_1_Lreuteri_@WP_003666260_1
1291742_000829395_1_Lhokkaidonens_@WP_041095060_1
1291742_000829395_1_Lhokkaidonens_@WP_041094913_1
1408206_000496265_1_Ppentosaceus_@WP_002833762_1
701521_000237995_2_Pclaussenii_@AEV95428_1
314315_000026065_1_Lsakei_@WP_011264381_1
1001583_000359625_1_Lbrevis_@WP_011668221_1
637971_001050435_1_Lkoreensis_@WP_048735765_1
511437_000211375_1_Lbuchneri_@WP_013728040_1
148814_001314945_1_Lkunkeei_@WP_034531238_1
1194971_001011095_1_Lsalivarius_@WP_003702044_1
945021_000283615_1_Thalophilus_@WP_014124659_1
1300150_000504125_1_Emundtii_@WP_010735231_1
565655_000157355_2_Ecasseliflavu_@WP_005225951_1
768486_000271405_2_Ehirae_@WP_010718948_1
1352_001416995_1_Efaecium_@WP_002294440_1
53345_001267395_1_Edurans_@WP_005880033_1
0.7

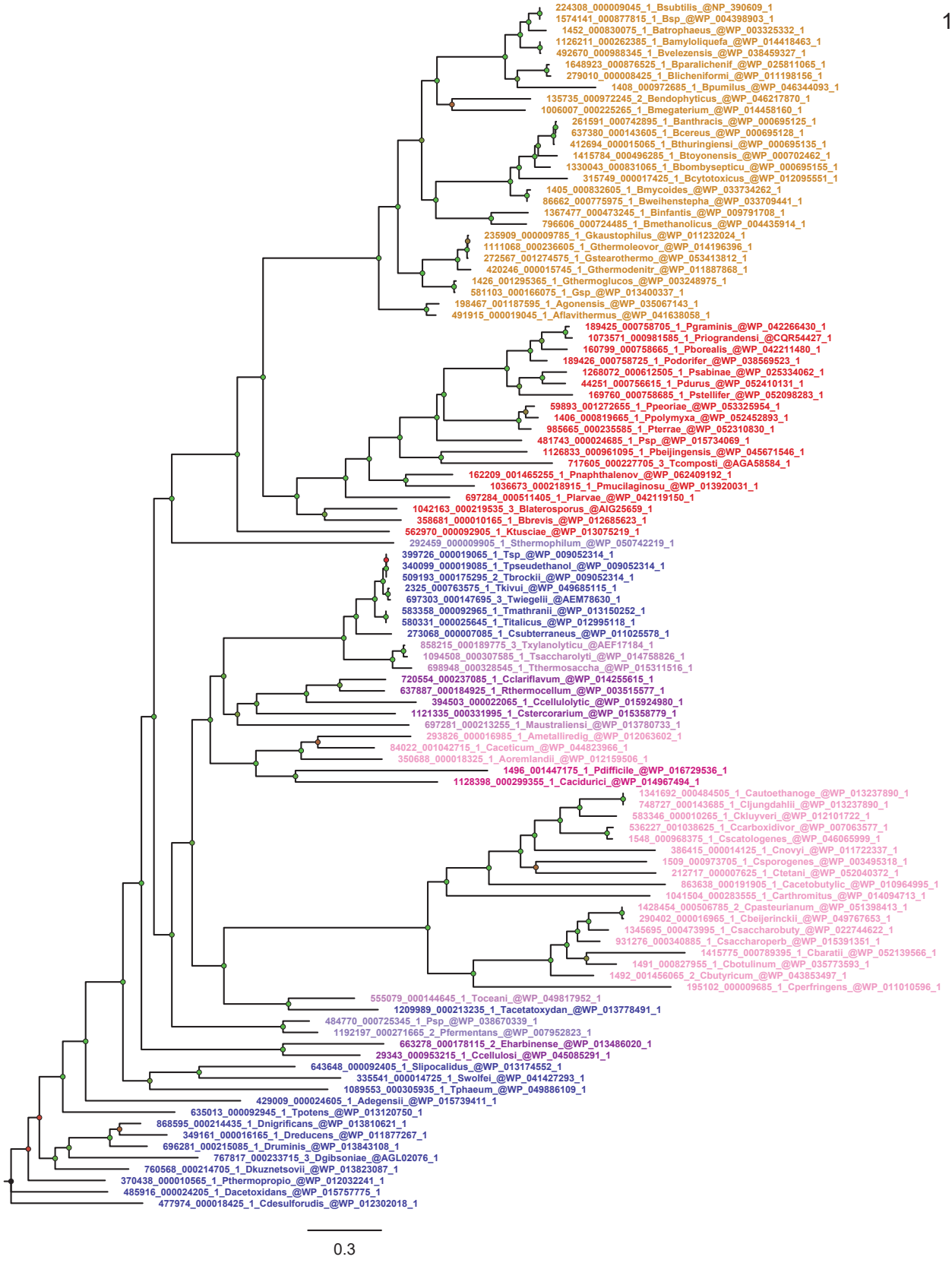

0.4

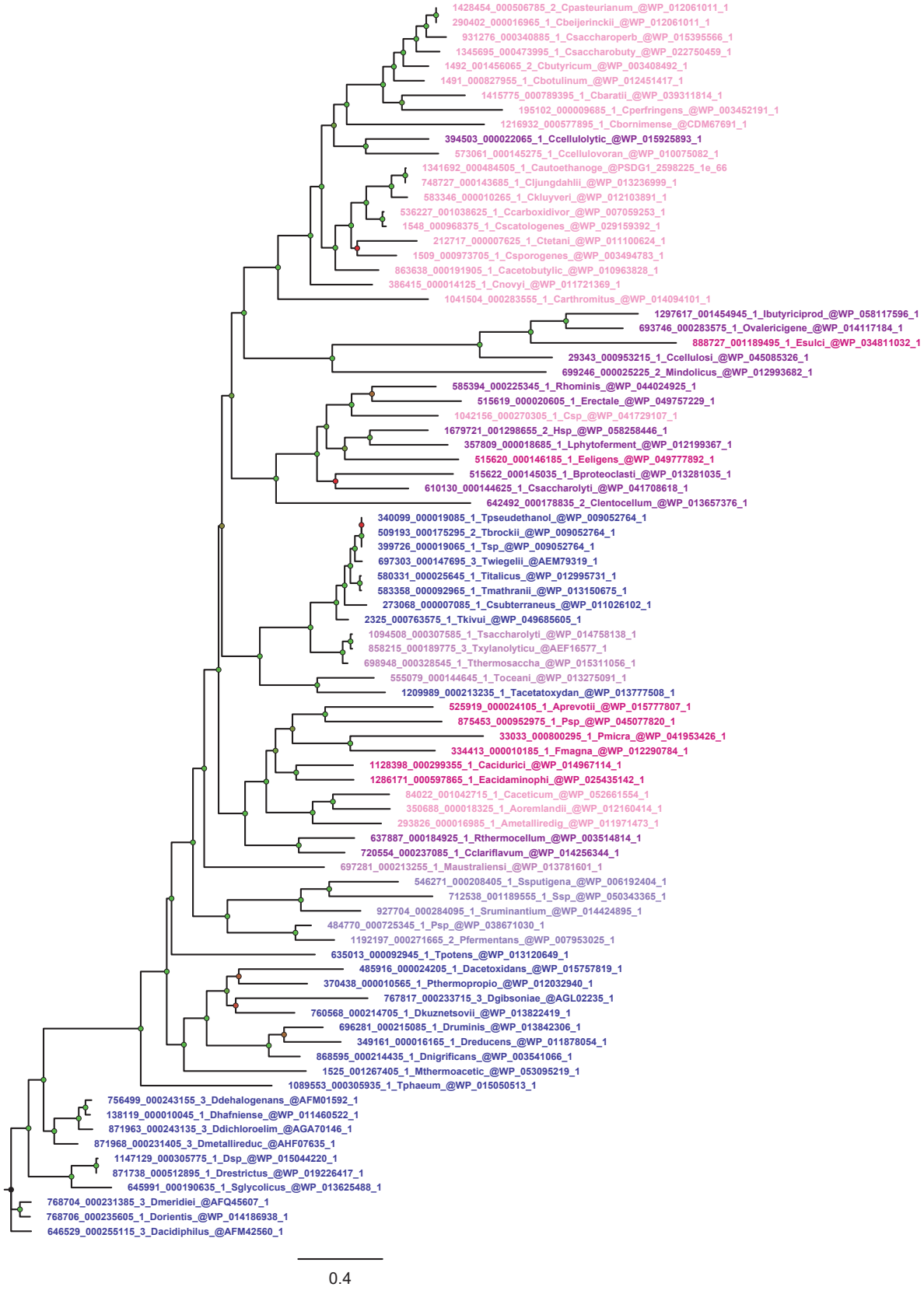

1526927_000785555_1_Psp_@WP_038703082_1
1374_001465835_1_Pkocurii_@WP_058386081_1
200991_001465795_1_Prifietoensis_@WP_058381432_1
200991_001465795_1_Prifietoensis_@WP_058381699_1
1526927_000785555_1_Psp_@WP_038702765_1
262543_000019905_1_Esibiricum_@WP_012369833_1
1087448_000299435_1_Eantarcticum_@WP_014969908_1
1399115_000496635_1_Esp_@WP_023467605_1
1508404_000818095_1_Jmalaysiensis_@WP_039807664_1
1479_001050115_1_Bsmithii_@WP_048624357_1
1398_001039495_1_Bcoagulans_@WP_035182259_1
221109_000011245_1_Oiheyensis_@WP_011066924_1
403957_000725285_1_Vsp_@WP_040956068_1
698758_000307165_1_Axylanus_@WP_015010629_1
1283_000972725_1_Shaemolyticus_@WP_000721306_1
1295_001188875_1_Sschleiferi_@WP_000721310_1
985002_000236925_1_Sargenteus_@WP_000721310_1
458233_000010585_1_Mcaseolyticus_@WP_012655867_1
407035_001005905_1_Shalodurans_@WP_046790913_1
66692_000009825_1_Bclausii_@WP_050748944_1
1246626_000706725_1_Blehensis_@WP_038482845_1
649639_000177235_2_Bcellulosilyt_@WP_013490227_1
272558_000011145_1_Bhalodurans_@WP_010899530_1
492670_000988345_1_Bvelezensis_@WP_038463613_1
1126211_000262385_1_Bamyloliquefa_@WP_014416926_1
1452_000830075_1_Batrophaeus_@WP_004430362_1
1574141_000877815_1_Bsp_@WP_021481735_1
224308_000009045_1_Bsubtilis_@NP_388295_1
1648923_000876525_1_Bparalichenif_@WP_023856451_1
279010_000008425_1_Blicheniformi_@WP_009328614_1
586416_000725365_1_Taidingensis_@WP_051748503_1
866895_000284515_1_Hhalophilus_@WP_014641791_1
135735_000972245_2_Bendophyticus_@WP_040057220_1
1006007_000225265_1_Bmegaterium_@WP_014458087_1
683837_000027145_1_Lseeligeri_@WP_012984871_1
202752_000763475_1_Livanovii_@WP_038408544_1
1230340_000318055_1_Lmonocytogene_@WP_003728098_1
386043_000060285_1_Lwelshimeri_@WP_011701282_1
1408_000972685_1_Bpumilus_@WP_035391396_1
1266845_000493735_1_Cinhibens_@WP_023178536_1
1564681_001483965_1_Csp_@WP_058918947_1
1234679_000317975_2_Cmaltaromatic_@WP_015075726_1
568704_000026525_1_Lrhamnosus_@WP_014571094_1
1226298_000829035_1_Lparacasei_@WP_003661514_1
1352_001412695_1_Efaecium_@WP_002288183_1
768486_000271405_2_Ehirae_@WP_010737633_1
1300150_000504125_1_Emundtii_@WP_023519712_1
53345_001267395_1_Edurans_@PSDG1_6541_9e_75
565655_000157355_2_Ecasseliflavu_@WP_015509892_1
332949_001465115_1_Esilesiacus_@ALS00897_1
1261557_000317915_1_Efaecalis_@WP_002411091_1
1090974_000283975_1_Mplutonius_@WP_015694803_1
1045854_000219805_1_Wkoreensis_@WP_013989674_1
412694_000015065_1_Bthuringiensi_@WP_001227215_1
261591_000742895_1_Banthracis_@WP_001227211_1
637380_000143605_1_Bcereus_@WP_001227214_1
1330043_000831065_1_Bbombysepticu_@WP_044584416_1
1415784_000496285_1_Btoyonensis_@WP_001227254_1
86662_000775975_1_Bweihenstepha_@WP_002031689_1
1405_000832605_1_Bmycoides_@WP_033733921_1
315749_000017425_1_Bcytotoxicus_@WP_012094126_1
412694_000015065_1_Bthuringiensi_@WP_000737742_1
637380_000143605_1_Bcereus_@WP_000737750_1
1330043_000831065_1_Bbombysepticu_@WP_044584473_1
86662_000775975_1_Bweihenstepha_@WP_033709116_1
1405_000832605_1_Bmycoides_@WP_033733979_1
1042163_000219535_3_Blaterosporus_@AIG24710_1
358681_000010165_1_Bbrevis_@WP_015892922_1
1415775_000789395_1_Cbaratii_@WP_039316915_1
573061_000145275_1_Ccellulovoran_@WP_010074082_1
697281_000213255_1_Maustraliensi_@WP_013781177_1
644966_000184705_1_Tmarianensis_@WP_013495047_1
1736_001481725_1_Elimosum_@WP_058696374_1
1736_001481725_1_Elimosum_@WP_058696233_1
610130_000144625_1_Csaccharolyti_@WP_013273940_1
585394_000225345_1_Rhominis_@WP_014081124_1
748727_000143685_1_Cljungdahlii_@WP_013237284_1
1341692_000484505_1_Cautoethanoge_@WP_023162801_1
583346_000010265_1_Ckluyveri_@WP_012103074_1
1548_000968375_1_Cscatologenes_@WP_046066142_1
863638_000191905_1_Cacetobutylic_@WP_014519088_1
1216932_000577895_1_Cbornimense_@CDM68176_1
1491_000827955_1_Cbotulinum_@WP_040968261_1
1492_001456065_2_Cbutyricum_@WP_046057737_1
1345695_000473995_1_Csaccharobuty_@WP_022746735_1
290402_000016965_1_Cbeijerinckii_@WP_012058132_1
1428454_000506785_2_Cpasteurianum_@WP_023975044_1
931276_000340885_1_Csaccharoperb_@WP_015393396_1
0.3

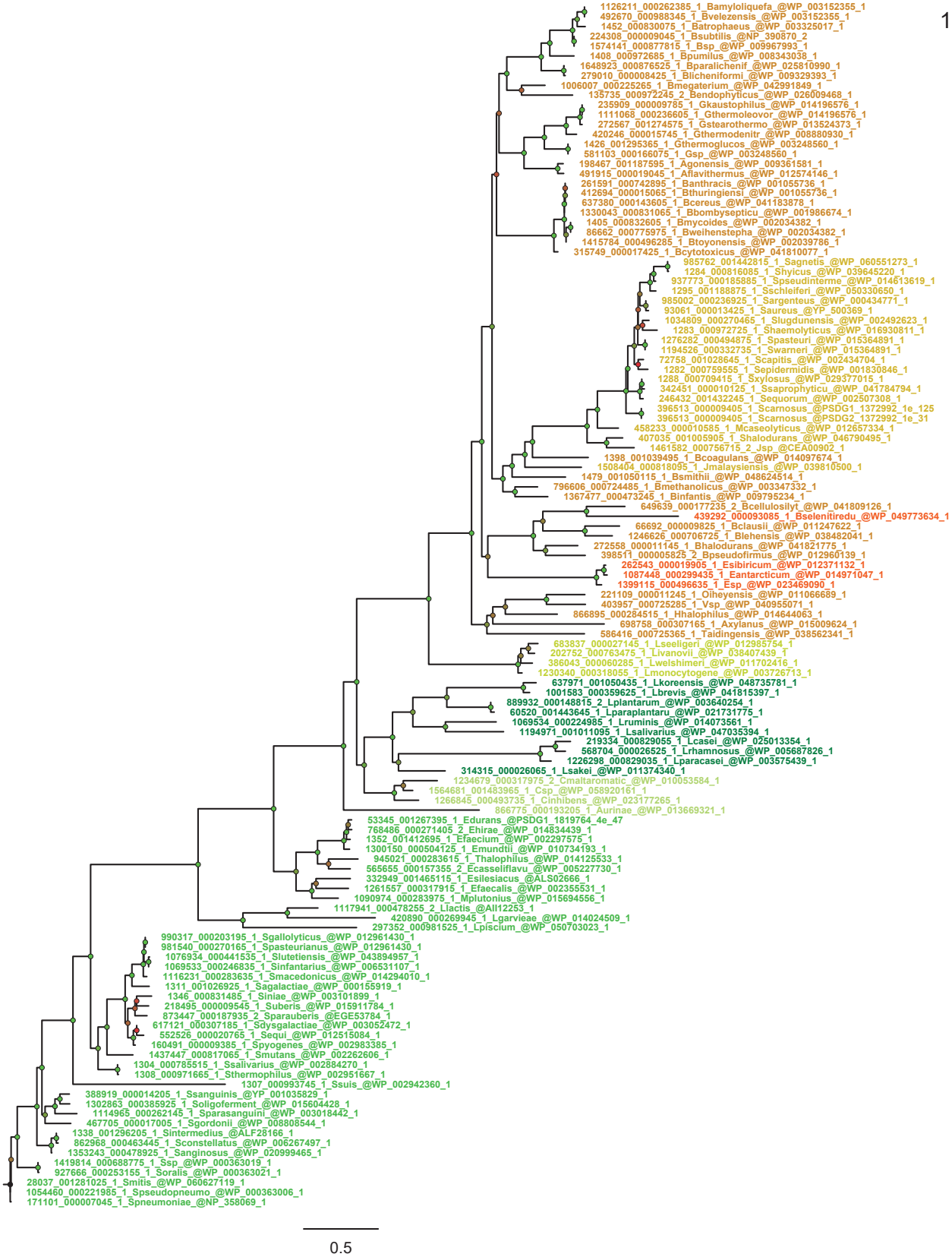

1073571_000981585_1_Priograndensi_@CQR57533_1
189425_000758705_1_Pgraminis_@WP_025708374_1
160799_000758665_1_Pborealis_@WP_042216718_1
189426_000758725_1_Podorifer_@WP_036683661_1
1268072_000612505_1_Psabinae_@WP_025335869_1
44251_000756615_1_Pdurus_@WP_042209521_1
169760_000758685_1_Pstellifer_@WP_038697646_1
481743_000024685_1_Psp_@WP_015734232_1
59893_001272655_1_Ppeoriae_@WP_043882263_1
1406_000819665_1_Ppolymyxa_@WP_040103325_1
985665_000235585_1_Pterrae_@WP_014282445_1
162209_001465255_1_Pnaphthalenov_@WP_062410861_1
1036673_000218915_1_Pmucilaginosu_@WP_041617894_1
697284_000511405_1_Plarvae_@WP_023485596_1
1126833_000961095_1_Pbeijingensis_@WP_045671790_1
717605_000227705_3_Tcomposti_@AGA58434_1
562970_000092905_1_Ktusciae_@WP_052300597_1
1048834_000219875_1_Aacidocaldari_@WP_041695723_1
358681_000010165_1_Bbrevis_@WP_015892101_1
1042163_000219535_3_Blaterosporus_@AIG27806_1
444177_000017965_1_Lsphaericus_@WP_012293155_1
28031_000724775_3_Lfusiformis_@WP_008173886_1
1145276_000600105_1_Lvarians_@WP_025218752_1
1002809_000271325_1_Ssilvestris_@WP_014824542_1
1002809_000271325_1_Ssilvestris_@WP_014824274_1
221109_000011245_1_Oiheyensis_@WP_011065866_1
403957_000725285_1_Vsp_@WP_040955749_1
698758_000307165_1_Axylanus_@WP_015010321_1
866895_000284515_1_Hhalophilus_@WP_014643112_1
586416_000725365_1_Taidingensis_@WP_038559945_1
66692_000009825_1_Bclausii_@WP_011247204_1
1246626_000706725_1_Blehensis_@WP_038480974_1
398511_000005825_2_Bpseudofirmus_@WP_012959493_1
272558_000011145_1_Bhalodurans_@WP_010898723_1
649639_000177235_2_Bcellulosilyt_@WP_013489188_1
637380_000143605_1_Bcereus_@WP_001266229_1
261591_000742895_1_Banthracis_@WP_001266229_1
412694_000015065_1_Bthuringiensi_@WP_001266229_1
1415784_000496285_1_Btoyonensis_@WP_001266239_1
1330043_000831065_1_Bbombysepticu_@WP_001266237_1
86662_000775975_1_Bweihenstepha_@WP_002033747_1
1405_000832605_1_Bmycoides_@WP_002014741_1
315749_000017425_1_Bcytotoxicus_@WP_012095006_1
1574141_000877815_1_Bsp_@WP_043940111_1
224308_000009045_1_Bsubtilis_@NP_389400_2
1452_000830075_1_Batrophaeus_@WP_013390506_1
492670_000988345_1_Bvelezensis_@WP_003154438_1
1126211_000262385_1_Bamyloliquefa_@WP_014417716_1
1408_000972685_1_Bpumilus_@WP_046343482_1
279010_000008425_1_Blicheniformi_@WP_003181540_1
1648923_000876525_1_Bparalichenif_@WP_020451297_1
135735_000972245_2_Bendophyticus_@WP_040056703_1
1006007_000225265_1_Bmegaterium_@WP_013084856_1
1111068_000236605_1_Gthermoleovor_@WP_011230614_1
235909_000009785_1_Gkaustophilus_@WP_011230614_1
272567_001274575_1_Gstearothermo_@WP_053413745_1
420246_000015745_1_Gthermodenitr_@WP_008878692_1
581103_000166075_1_Gsp_@WP_013401313_1
1426_001295365_1_Gthermoglucos_@WP_013401313_1
491915_000019045_1_Aflavithermus_@WP_041638593_1
198467_001187595_1_Agonensis_@WP_035065979_1
1398_001039495_1_Bcoagulans_@X1_2399786_1e_137
1479_001050115_1_Bsmithii_@WP_048623099_1
1367477_000473245_1_Binfantis_@WP_022543694_1
796606_000724485_1_Bmethanolicus_@WP_003348879_1
0.2

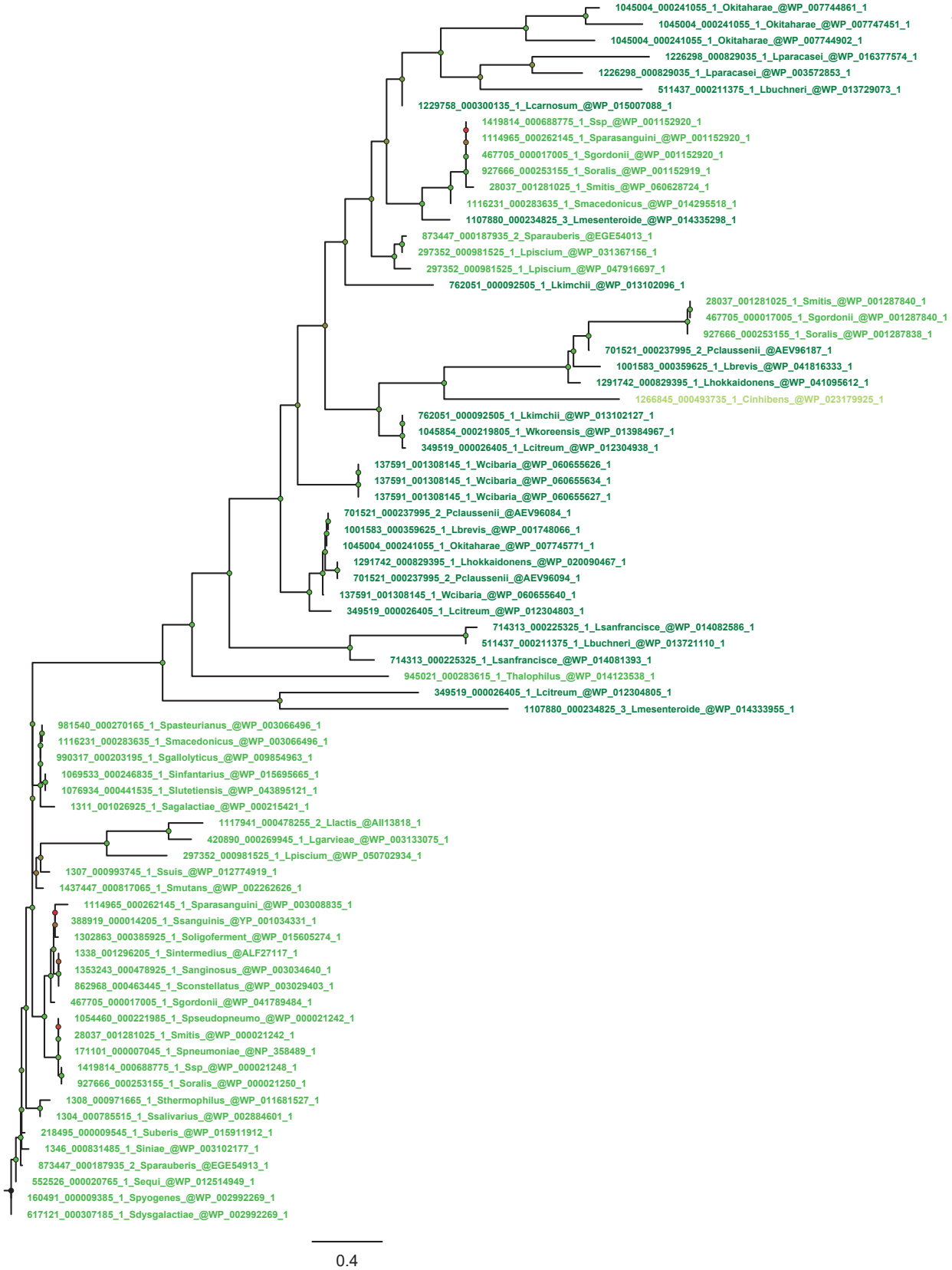

0.4

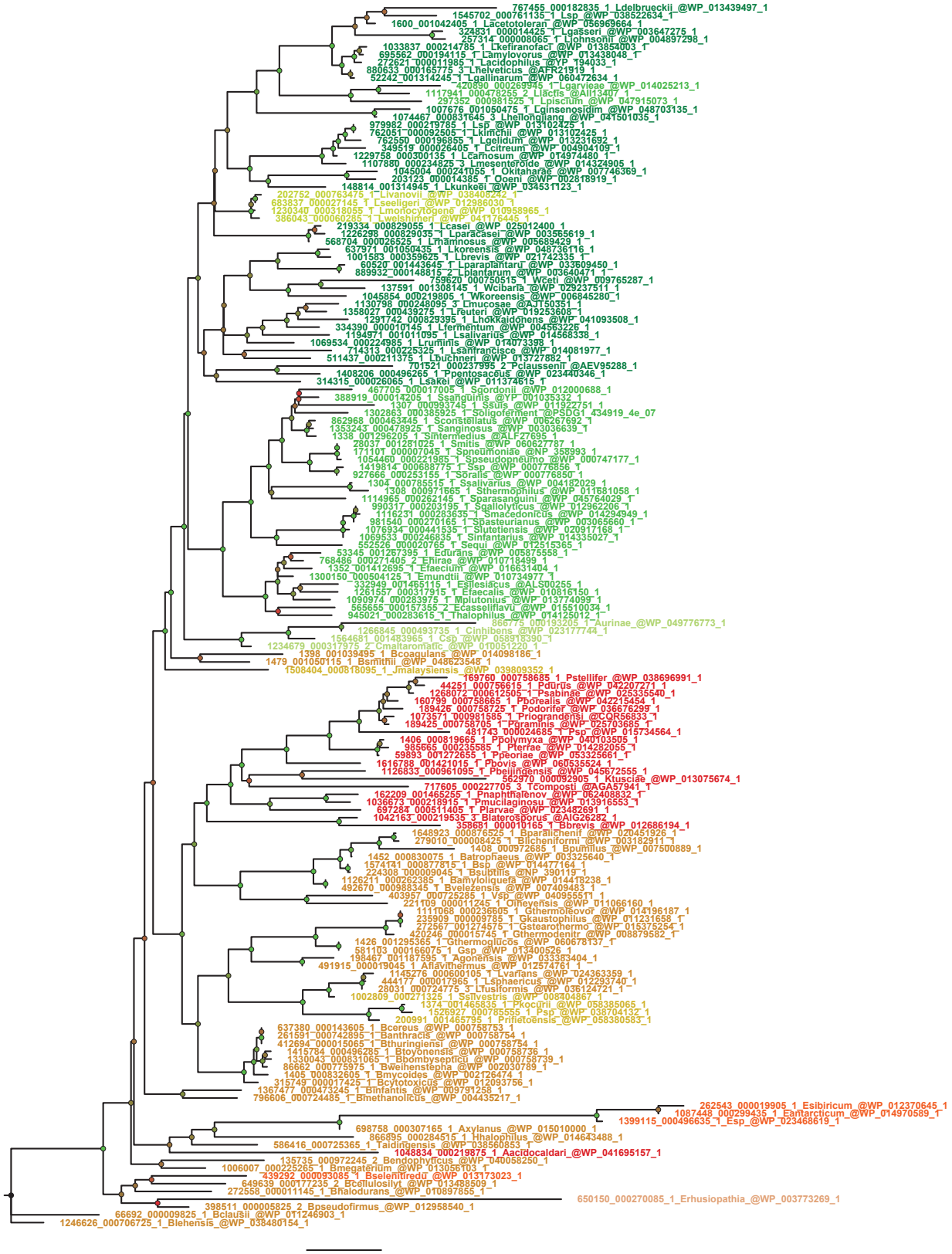

0.5

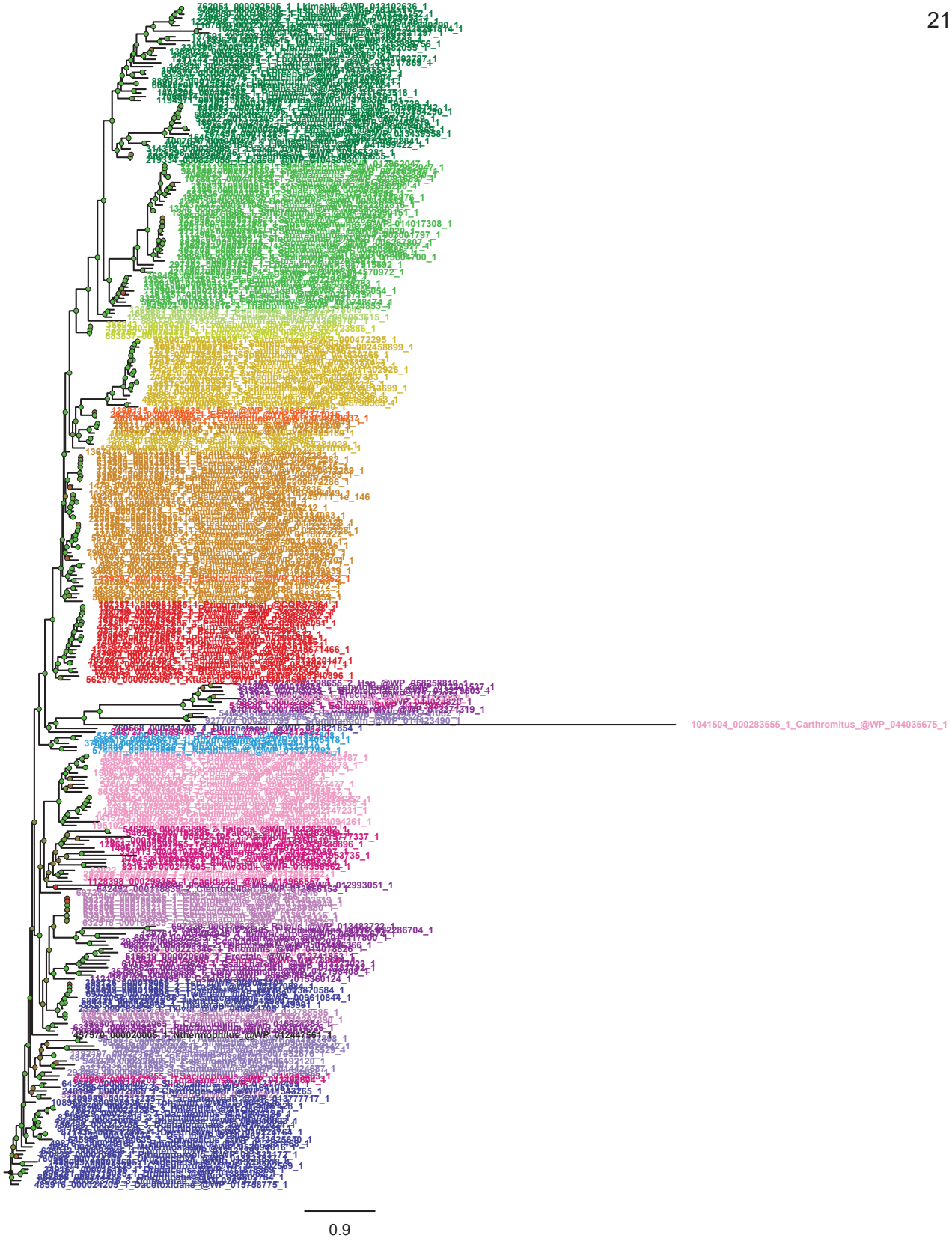

224308_000009045_1_Bsubtilis_@NP_390609_1
1574141_000877815_1_Bsp_@WP_004398903_1
1452_000830075_1_Batrophaeus_@WP_003325332_1
1126211_000262385_1_Bamyloliquefa_@WP_014418463_1
492670_000988345_1_Bvelezensis_@WP_038459327_1
1648923_000876525_1_Bparalichenif_@WP_025811065_1
279010_000008425_1_Blicheniformi_@WP_011198156_1
1408_000972685_1_Bpumilus_@WP_046344093_1
135735_000972245_2_Bendophyticus_@WP_046217870_1
1006007_000225265_1_Bmegaterium_@WP_014458160_1
261591_000742895_1_Banthracis_@WP_000695125_1
637380_000143605_1_Bcereus_@WP_000695128_1
412694_000015065_1_Bthuringiensi_@WP_000695135_1
1415784_000496285_1_Btoyonensis_@WP_000702462_1
1330043_000831065_1_Bbombysepticu_@WP_000695155_1
315749_000017425_1_Bcytotoxicus_@WP_012095551_1
1405_000832605_1_Bmycoides_@WP_033734262_1
86662_000775975_1_Bweihenstepha_@WP_033709441_1
1367477_000473245_1_Binfantis_@WP_009791708_1
796606_000724485_1_Bmethanolicus_@WP_004435914_1
235909_000009785_1_Gkaustophilus_@WP_011232024_1
1111068_000236605_1_Gthermoleovor_@WP_014196396_1
272567_001274575_1_Gstearothermo_@WP_053413812_1
420246_000015745_1_Gthermodenitr_@WP_011887868_1
1426_001295365_1_Gthermoglucos_@WP_003248975_1
581103_000166075_1_Gsp_@WP_013400337_1
198467_001187595_1_Agonensis_@WP_035067143_1
491915_000019045_1_Aflavithermus_@WP_041638058_1
189425_000758705_1_Pgraminis_@WP_042266430_1
1073571_000981585_1_Priograndensi_@CQR54427_1
160799_000758665_1_Pborealis_@WP_042211480_1
189426_000758725_1_Podorifer_@WP_038569523_1
1268072_000612505_1_Psabinae_@WP_025334062_1
44251_000756615_1_Pdurus_@WP_052410131_1
169760_000758685_1_Pstellifer_@WP_052098283_1
59893_001272655_1_Ppeoriae_@WP_053325954_1
1406_000819665_1_Ppolymyxa_@WP_052452893_1
985665_000235585_1_Pterrae_@WP_052310830_1
481743_000024685_1_Psp_@WP_015734069_1
1126833_000961095_1_Pbeijingensis_@WP_045671546_1
717605_000227705_3_Tcomposti_@AGA58584_1
162209_001465255_1_Pnaphthalenov_@WP_062409192_1
1036673_000218915_1_Pmucilaginosu_@WP_013920031_1
697284_000511405_1_Plarvae_@WP_042119150_1
1042163_000219535_3_Blaterosporus_@AIG25659_1
358681_000010165_1_Bbrevis_@WP_012685623_1
562970_000092905_1_Ktusciae_@WP_013075219_1
292459_000009905_1_Sthermophilum_@WP_050742219_1
399726_000019065_1_Tsp_@WP_009052314_1
340099_000019085_1_Tpseudethanol_@WP_009052314_1
509193_000175295_2_Tbrockii_@WP_009052314_1
2325_000763575_1_Tkivui_@WP_049685115_1
697303_000147695_3_Twiegelii_@AEM78630_1
583358_000092965_1_Tmathranii_@WP_013150252_1
580331_000025645_1_Titalicus_@WP_012995118_1
273068_000007085_1_Csubterraneus_@WP_011025578_1
858215_000189775_3_Txylanolyticu_@AEF17184_1
1094508_000307585_1_Tsaccharolyti_@WP_014758826_1
698948_000328545_1_Tthermosaccha_@WP_015311516_1
720554_000237085_1_Cclariflavum_@WP_014255615_1
637887_000184925_1_Rthermocellum_@WP_003515577_1
394503_000022065_1_Ccellulolytic_@WP_015924980_1
1121335_000331995_1_Cstercorarium_@WP_015358779_1
697281_000213255_1_Maustraliensi_@WP_013780733_1
293826_000016985_1_Ametalliredig_@WP_012063602_1
84022_001042715_1_Caceticum_@WP_044823966_1
350688_000018325_1_Aoremlandii_@WP_012159506_1
1496_001447175_1_Pdifficile_@WP_016729536_1
1128398_000299355_1_Cacidurici_@WP_014967494_1
1341692_000484505_1_Cautoethanoge_@WP_013237890_1
748727_000143685_1_Cljungdahlii_@WP_013237890_1
583346_000010265_1_Ckluyveri_@WP_012101722_1
536227_001038625_1_Ccarboxidivor_@WP_007063577_1
1548_000968375_1_Cscatologenes_@WP_046065999_1
386415_000014125_1_Cnovyi_@WP_011722337_1
1509_000973705_1_Csporogenes_@WP_003495318_1
212717_000007625_1_Ctetani_@WP_052040372_1
863638_000191905_1_Cacetobutylic_@WP_010964995_1
1041504_000283555_1_Carthromitus_@WP_014094713_1
1428454_000506785_2_Cpasteurianum_@WP_051398413_1
290402_000016965_1_Cbeijerinckii_@WP_049767653_1
1345695_000473995_1_Csaccharobuty_@WP_022744622_1
931276_000340885_1_Csaccharoperb_@WP_015391351_1
1415775_000789395_1_Cbaratii_@WP_052139566_1
1491_000827955_1_Cbotulinum_@WP_035773593_1
1492_001456065_2_Cbutyricum_@WP_043853497_1
195102_000009685_1_Cperfringens_@WP_011010596_1
555079_000144645_1_Toceani_@WP_049817952_1
1209989_000213235_1_Tacetatoxydan_@WP_013778491_1
484770_000725345_1_Psp_@WP_038670339_1
1192197_000271665_2_Pfermentans_@WP_007952823_1
663278_000178115_2_Eharbinense_@WP_013486020_1
29343_000953215_1_Ccellulosi_@WP_045085291_1
643648_000092405_1_Slipocalidus_@WP_013174552_1
335541_000014725_1_Swolfei_@WP_041427293_1
1089553_000305935_1_Tphaeum_@WP_049886109_1
429009_000024605_1_Adegensii_@WP_015739411_1
635013_000092945_1_Tpotens_@WP_013120750_1
868595_000214435_1_Dnigrificans_@WP_013810621_1
349161_000016165_1_Dreducens_@WP_011877267_1
696281_000215085_1_Druminis_@WP_013843108_1
767817_000233715_3_Dgibsoniae_@AGL02076_1
760568_000214705_1_Dkuznetsovii_@WP_013823087_1
370438_000010565_1_Pthermopropio_@WP_012032241_1
485916_000024205_1_Dacetoxidans_@WP_015757775_1
477974_000018425_1_Cdesulforudis_@WP_012302018_1
0.3

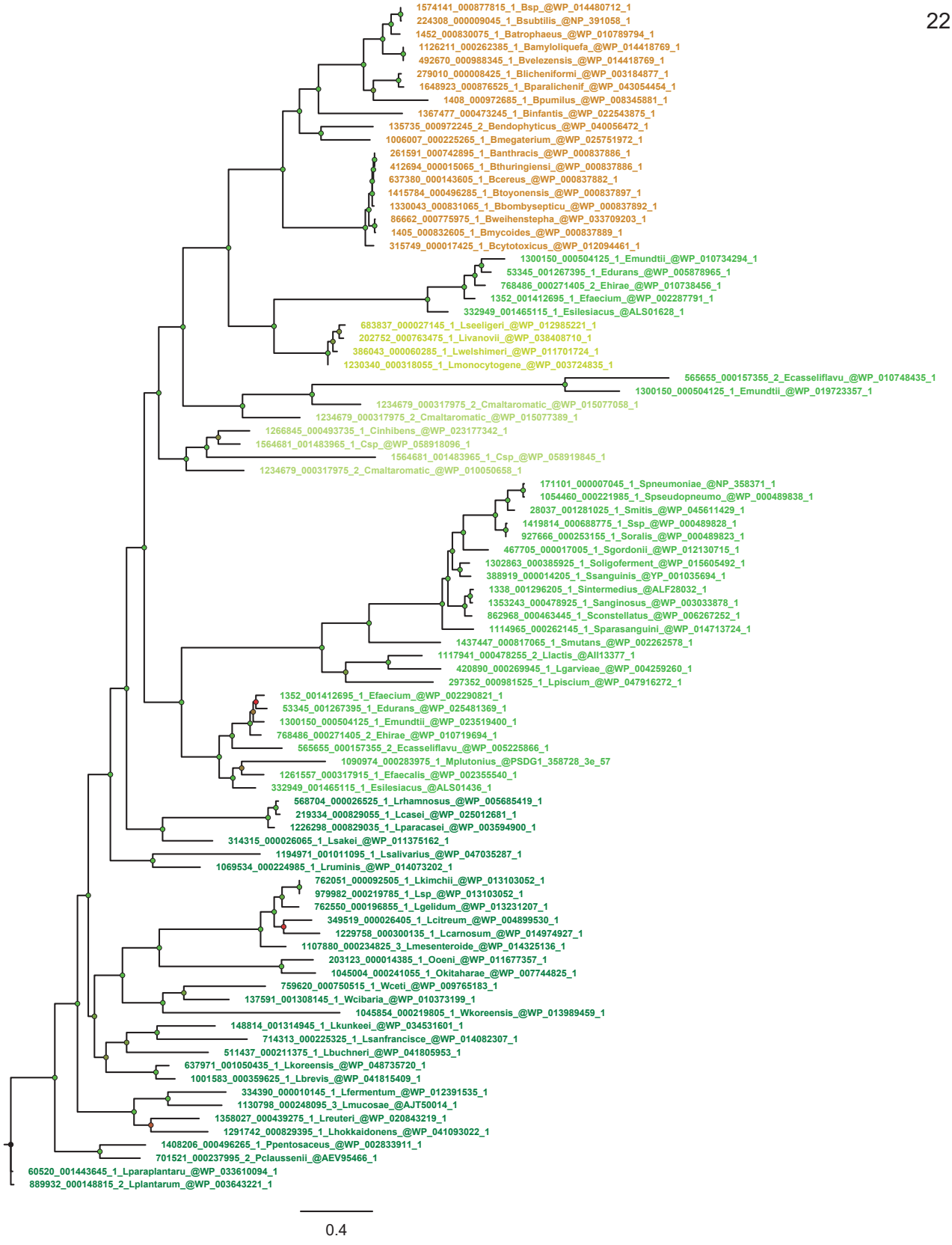

1428454_000506785_2_Cpasteurianum_@WP_012061011_1
290402_000016965_1_Cbeijerinckii_@WP_012061011_1
931276_000340885_1_Csaccharoperb_@WP_015395566_1
1345695_000473995_1_Csaccharobuty_@WP_022750459_1
1492_001456065_2_Cbutyricum_@WP_003408492_1
1491_000827955_1_Cbotulinum_@WP_012451417_1
1415775_000789395_1_Cbaratii_@WP_039311814_1
195102_000009685_1_Cperfringens_@WP_003452191_1
1216932_000577895_1_Cbornimense_@CDM67691_1
394503_000022065_1_Ccellulolytic_@WP_015925893_1
573061_000145275_1_Ccellulovoran_@WP_010075082_1
1341692_000484505_1_Cautoethanoge_@PSDG1_2598225_1e_66
748727_000143685_1_Cljungdahlii_@WP_013236999_1
583346_000010265_1_Ckluyveri_@WP_012103891_1
536227_001038625_1_Ccarboxidivor_@WP_007059253_1
1548_000968375_1_Cscatologenes_@WP_029159392_1
212717_000007625_1_Ctetani_@WP_011100624_1
1509_000973705_1_Csporogenes_@WP_003494783_1
863638_000191905_1_Cacetobutylic_@WP_010963828_1
386415_000014125_1_Cnovyi_@WP_011721369_1
1041504_000283555_1_Carthromitus_@WP_014094101_1
1297617_001454945_1_Ibutyriciprod_@WP_058117596_1
693746_000283575_1_Ovalericigene_@WP_014117184_1
888727_001189495_1_Esulci_@WP_034811032_1
29343_000953215_1_Ccellulosi_@WP_045085326_1
699246_000025225_2_Mindolicus_@WP_012993682_1
585394_000225345_1_Rhominis_@WP_044024925_1
515619_000020605_1_Erectale_@WP_049757229_1
1042156_000270305_1_Csp_@WP_041729107_1
1679721_001298655_2_Hsp_@WP_058258446_1
357809_000018685_1_Lphytoferment_@WP_012199367_1
515620_000146185_1_Eeligens_@WP_049777892_1
515622_000145035_1_Bproteoclasti_@WP_013281035_1
610130_000144625_1_Csaccharolyti_@WP_041708618_1
642492_000178835_2_Clentocellum_@WP_013657376_1
340099_000019085_1_Tpseudethanol_@WP_009052764_1
509193_000175295_2_Tbrockii_@WP_009052764_1
399726_000019065_1_Tsp_@WP_009052764_1
697303_000147695_3_Twiegelii_@AEM79319_1
580331_000025645_1_Titalicus_@WP_012995731_1
583358_000092965_1_Tmathranii_@WP_013150675_1
273068_000007085_1_Csubterraneus_@WP_011026102_1
2325_000763575_1_Tkivui_@WP_049685605_1
1094508_000307585_1_Tsaccharolyti_@WP_014758138_1
858215_000189775_3_Txylanolyticu_@AEF16577_1
698948_000328545_1_Tthermosaccha_@WP_015311056_1
555079_000144645_1_Toceani_@WP_013275091_1
1209989_000213235_1_Tacetatoxydan_@WP_013777508_1
525919_000024105_1_Aprevotii_@WP_015777807_1
875453_000952975_1_Psp_@WP_045077820_1
33033_000800295_1_Pmicra_@WP_041953426_1
334413_000010185_1_Fmagna_@WP_012290784_1
1128398_000299355_1_Cacidurici_@WP_014967114_1
1286171_000597865_1_Eacidaminophi_@WP_025435142_1
84022_001042715_1_Caceticum_@WP_052661554_1
350688_000018325_1_Aoremlandii_@WP_012160414_1
293826_000016985_1_Ametalliredig_@WP_011971473_1
637887_000184925_1_Rthermocellum_@WP_003514814_1
720554_000237085_1_Cclariflavum_@WP_014256344_1
697281_000213255_1_Maustraliensi_@WP_013781601_1
546271_000208405_1_Ssputigena_@WP_006192404_1
712538_001189555_1_Ssp_@WP_050343365_1
927704_000284095_1_Sruminantium_@WP_014424895_1
484770_000725345_1_Psp_@WP_038671030_1
1192197_000271665_2_Pfermentans_@WP_007953025_1
635013_000092945_1_Tpotens_@WP_013120649_1
485916_000024205_1_Dacetoxidans_@WP_015757819_1
370438_000010565_1_Pthermopropio_@WP_012032940_1
767817_000233715_3_Dgibsoniae_@AGL02235_1
760568_000214705_1_Dkuznetsovii_@WP_013822419_1
696281_000215085_1_Druminis_@WP_013842306_1
349161_000016165_1_Dreducens_@WP_011878054_1
868595_000214435_1_Dnigrificans_@WP_003541066_1
1525_001267405_1_Mthermoacetic_@WP_053095219_1
1089553_000305935_1_Tphaeum_@WP_015050513_1
756499_000243155_3_Ddehalogenans_@AFM01592_1
138119_000010045_1_Dhafniense_@WP_011460522_1
871963_000243135_3_Ddichloroelim_@AGA70146_1
871968_000231405_3_Dmetallireduc_@AHF07635_1
1147129_000305775_1_Dsp_@WP_015044220_1
871738_000512895_1_Drestrictus_@WP_019226417_1
645991_000190635_1_Sglycolicus_@WP_013625488_1
768704_000231385_3_Dmeridiei_@AFQ45607_1
768706_000235605_1_Dorientis_@WP_014186938_1
646529_000255115_3_Dacidiphilus_@AFM42560_1
0.4

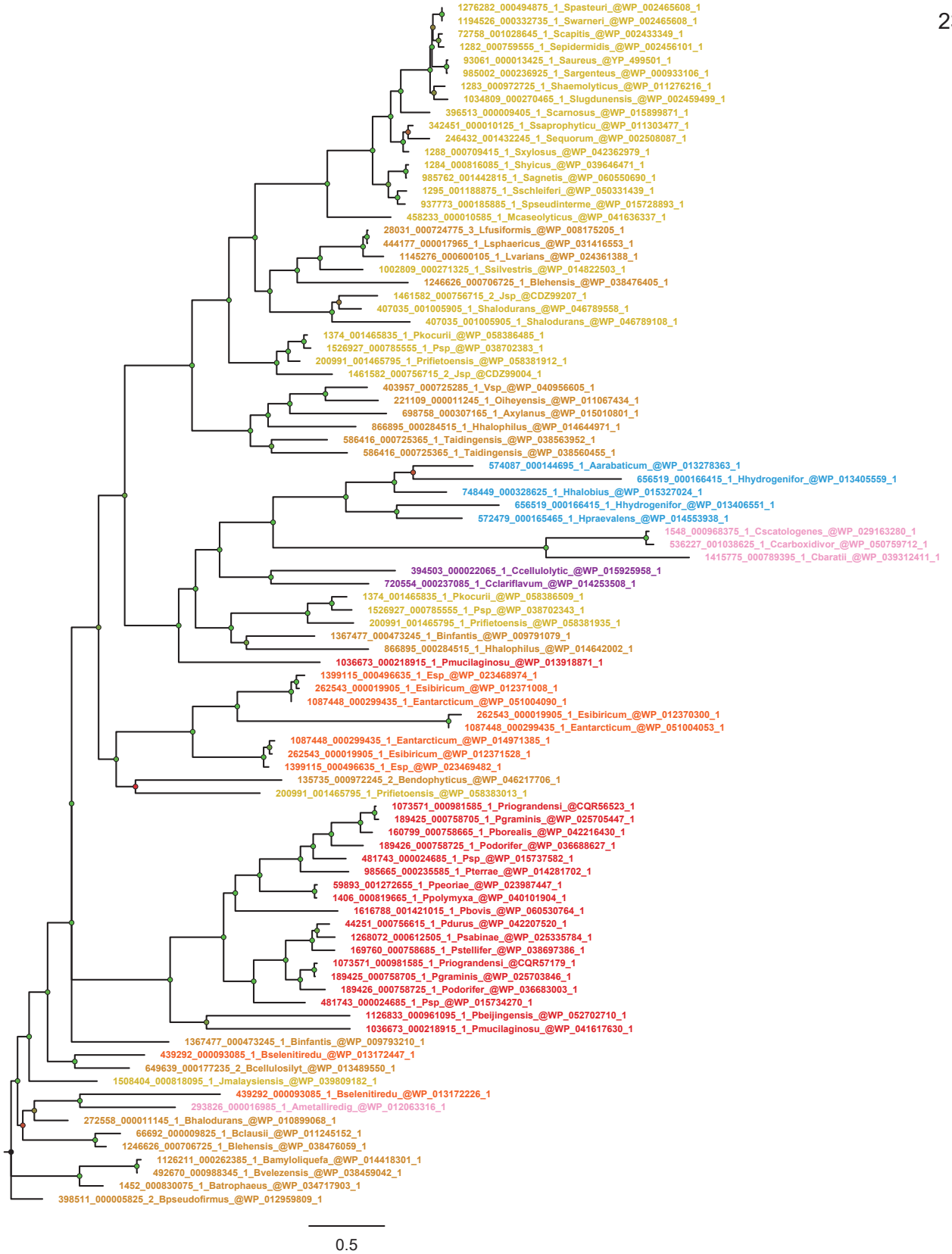

1126211_000262385_1_Bamyloliquefa_@WP_003152355_1
492670_000988345_1_Bvelezensis_@WP_003152355_1
1452_000830075_1_Batrophaeus_@WP_003325017_1
224308_000009045_1_Bsubtilis_@NP_390870_2
1574141_000877815_1_Bsp_@WP_009967993_1
1408_000972685_1_Bpumilus_@WP_008343038_1
1648923_000876525_1_Bparalichenif_@WP_025810990_1
279010_000008425_1_Blicheniformi_@WP_009329393_1
1006007_000225265_1_Bmegaterium_@WP_042991849_1
135735_000972245_2_Bendophyticus_@WP_026009468_1
235909_000009785_1_Gkaustophilus_@WP_014196576_1
1111068_000236605_1_Gthermoleovor_@WP_014196576_1
272567_001274575_1_Gstearothermo_@WP_013524373_1
420246_000015745_1_Gthermodenitr_@WP_008880930_1
1426_001295365_1_Gthermoglucos_@WP_003248560_1
581103_000166075_1_Gsp_@WP_003248560_1
198467_001187595_1_Agonensis_@WP_009361581_1
491915_000019045_1_Aflavithermus_@WP_012574146_1
261591_000742895_1_Banthracis_@WP_001055736_1
412694_000015065_1_Bthuringiensi_@WP_001055736_1
637380_000143605_1_Bcereus_@WP_041183878_1
1330043_000831065_1_Bbombysepticu_@WP_001986674_1
1405_000832605_1_Bmycoides_@WP_002034382_1
86662_000775975_1_Bweihenstepha_@WP_002034382_1
1415784_000496285_1_Btoyonensis_@WP_002039786_1
315749_000017425_1_Bcytotoxicus_@WP_041810077_1
985762_001442815_1_Sagnetis_@WP_060551273_1
1284_000816085_1_Shyicus_@WP_039645220_1
937773_000185885_1_Spseudinterme_@WP_014613619_1
1295_001188875_1_Sschleiferi_@WP_050330650_1
985002_000236925_1_Sargenteus_@WP_000434771_1
93061_000013425_1_Saureus_@YP_500369_1
1034809_000270465_1_Slugdunensis_@WP_002492623_1
1283_000972725_1_Shaemolyticus_@WP_016930811_1
1276282_000494875_1_Spasteuri_@WP_015364891_1
1194526_000332735_1_Swarneri_@WP_015364891_1
72758_001028645_1_Scapitis_@WP_002434704_1
1282_000759555_1_Sepidermidis_@WP_001830846_1
1288_000709415_1_Sxylosus_@WP_029377015_1
342451_000010125_1_Ssaprophyticu_@WP_041784794_1
246432_001432245_1_Sequorum_@WP_002507308_1
396513_000009405_1_Scarnosus_@PSDG1_1372992_1e_125
396513_000009405_1_Scarnosus_@PSDG2_1372992_1e_31
458233_000010585_1_Mcaseolyticus_@WP_012657334_1
407035_001005905_1_Shalodurans_@WP_046790495_1
1461582_000756715_2_Jsp_@CEA00902_1
1398_001039495_1_Bcoagulans_@WP_014097674_1
1508404_000818095_1_Jmalaysiensis_@WP_039810500_1
1479_001050115_1_Bsmithii_@WP_048624514_1
796606_000724485_1_Bmethanolicus_@WP_003347332_1
1367477_000473245_1_Binfantis_@WP_009795234_1
649639_000177235_2_Bcellulosilyt_@WP_041809126_1
439292_000093085_1_Bselenitiredu_@WP_049773634_1
66692_000009825_1_Bclausii_@WP_011247622_1
1246626_000706725_1_Blehensis_@WP_038482041_1
272558_000011145_1_Bhalodurans_@WP_041821775_1
398511_000005825_2_Bpseudofirmus_@WP_012960139_1
262543_000019905_1_Esibiricum_@WP_012371132_1
1087448_000299435_1_Eantarcticum_@WP_014971047_1
1399115_000496635_1_Esp_@WP_023469090_1
221109_000011245_1_Oiheyensis_@WP_011066689_1
403957_000725285_1_Vsp_@WP_040955071_1
866895_000284515_1_Hhalophilus_@WP_014644063_1
698758_000307165_1_Axylanus_@WP_015009624_1
586416_000725365_1_Taidingensis_@WP_038562341_1
683837_000027145_1_Lseeligeri_@WP_012985754_1
202752_000763475_1_Livanovii_@WP_038407439_1
386043_000060285_1_Lwelshimeri_@WP_011702416_1
1230340_000318055_1_Lmonocytogene_@WP_003726713_1
637971_001050435_1_Lkoreensis_@WP_048735781_1
1001583_000359625_1_Lbrevis_@WP_041815397_1
889932_000148815_2_Lplantarum_@WP_003640254_1
60520_001443645_1_Lparaplantaru_@WP_021731775_1
1069534_000224985_1_Lruminis_@WP_014073561_1
1194971_001011095_1_Lsalivarius_@WP_047035394_1
219334_000829055_1_Lcasei_@WP_025013354_1
568704_000026525_1_Lrhamnosus_@WP_005687826_1
1226298_000829035_1_Lparacasei_@WP_003575439_1
314315_000026065_1_Lsakei_@WP_011374340_1
1234679_000317975_2_Cmaltaromatic_@WP_010053584_1
1564681_001483965_1_Csp_@WP_058920161_1
1266845_000493735_1_Cinhibens_@WP_023177265_1
886775_000193205_1_Aurinae_@WP_013669321_1
53345_001267395_1_Edurans_@PSDG1_1819764_4e_47
768486_000271405_2_Ehirae_@WP_014834439_1
1352_001412695_1_Efaecium_@WP_002297575_1
1300150_000504125_1_Emundtii_@WP_010734193_1
945021_000283615_1_Thalophilus_@WP_014125533_1
565655_000157355_2_Ecasseliflavu_@WP_005227730_1
332949_001465115_1_Esilesiacus_@ALS02666_1
1261557_000317915_1_Efaecalis_@WP_002355531_1
1090974_000283975_1_Mplutonius_@WP_015694556_1
1117941_000478255_2_Llactis_@AII12253_1
420890_000269945_1_Lgarvieae_@WP_014024509_1
297352_000981525_1_Lpiscium_@WP_050703023_1
990317_000203195_1_Sgallolyticus_@WP_012961430_1
981540_000270165_1_Spasteurianus_@WP_012961430_1
1076934_000441535_1_Sluteticnsis_@WP_043894957_1
1069533_000246835_1_Sinfantarius_@WP_006531107_1
1116231_000283635_1_Smacedonicus_@WP_014294010_1
1311_001026925_1_Sagalactiae_@WP_000155919_1
1346_000831485_1_Siniae_@WP_003101899_1
218495_000009545_1_Suberis_@WP_015911784_1
873447_000187935_2_Sparauberis_@EGE53784_1
617121_000307185_1_Sdysgalactiae_@WP_003052472_1
552526_000020765_1_Sequi_@WP_012515084_1
160491_000009385_1_Spyogenes_@WP_002983385_1
1437447_000817065_1_Smutans_@WP_002262606_1
1304_000785515_1_Ssalivarius_@WP_002884270_1
1308_000971665_1_Sthermophilus_@WP_002951667_1
1307_000993745_1_Ssuis_@WP_002942360_1
388919_000014205_1_Ssanguinis_@YP_001035829_1
1302863_000385925_1_Soligoferment_@WP_015604428_1
1114965_000262145_1_Sparasanguini_@WP_003018442_1
467705_000017005_1_Sgordonii_@WP_008808544_1
1338_001296205_1_Sintermedius_@ALF28166_1
862968_000463445_1_Sconstellatus_@WP_006267497_1
1353243_000478925_1_Sanginosus_@WP_020999465_1
1419814_000688775_1_Ssp_@WP_000363019_1
927666_000253155_1_Soralis_@WP_000363021_1
28037_001281025_1_Smitis_@WP_060627119_1
1054460_000221985_1_Spseudopneumo_@WP_000363006_1
171101_000007045_1_Spneumoniae_@NP_358069_1
0.5


1045004_000241055_1_Okitaharae_@WP_007744861_1
1045004_000241055_1_Okitaharae_@WP_007747451_1
1045004_000241055_1_Okitaharae_@WP_007744902_1
1226298_000829035_1_Lparacasei_@WP_016377574_1
1226298_000829035_1_Lparacasei_@WP_003572853_1
511437_000211375_1_Lbuchneri_@WP_013729073_1
1229758_000300135_1_Lcarnosum_@WP_015007088_1
1419814_000688775_1_Ssp_@WP_001152920_1
1114965_000262145_1_Sparasanguini_@WP_001152920_1
467705_000017005_1_Sgordonii_@WP_001152920_1
927666_000253155_1_Soralis_@WP_001152919_1
28037_001281025_1_Smitis_@WP_060628724_1
1116231_000283635_1_Smacedonicus_@WP_014295518_1
1107880_000234825_3_Lmesenteroide_@WP_014335298_1
873447_000187935_2_Sparauberis_@EGE54013_1
297352_000981525_1_Lpiscium_@WP_031367156_1
297352_000981525_1_Lpiscium_@WP_047916697_1
762051_000092505_1_Lkimchii_@WP_013102096_1
28037_001281025_1_Smitis_@WP_001287840_1
467705_000017005_1_Sgordonii_@WP_001287840_1
927666_000253155_1_Soralis_@WP_001287838_1
701521_000237995_2_Pclaussenii_@AEV96187_1
1001583_000359625_1_Lbrevis_@WP_041816333_1
1291742_000829395_1_Lhokkaidonens_@WP_041095612_1
1266845_000493735_1_Cinhibens_@WP_023179925_1
762051_000092505_1_Lkimchii_@WP_013102127_1
1045854_000219805_1_Wkoreensis_@WP_013984967_1
349519_000026405_1_Lcitreum_@WP_012304938_1
137591_001308145_1_Wcibaria_@WP_060655626_1
137591_001308145_1_Wcibaria_@WP_060655634_1
137591_001308145_1_Wcibaria_@WP_060655627_1
701521_000237995_2_Pclaussenii_@AEV96084_1
1001583_000359625_1_Lbrevis_@WP_001748066_1
1045004_000241055_1_Okitaharae_@WP_007745771_1
1291742_000829395_1_Lhokkaidonens_@WP_020090467_1
701521_000237995_2_Pclaussenii_@AEV96094_1
137591_001308145_1_Wcibaria_@WP_060655640_1
349519_000026405_1_Lcitreum_@WP_012304803_1
714313_000225325_1_Lsanfrancisce_@WP_014082586_1
511437_000211375_1_Lbuchneri_@WP_013721110_1
714313_000225325_1_Lsanfrancisce_@WP_014081393_1
945021_000283615_1_Thalophilus_@WP_014123538_1
349519_000026405_1_Lcitreum_@WP_012304805_1
1107880_000234825_3_Lmesenteroide_@WP_014333955_1
981540_000270165_1_Spasteurianus_@WP_003066496_1
1116231_000283635_1_Smacedonicus_@WP_003066496_1
990317_000203195_1_Sgallolyticus_@WP_009854963_1
1069533_000246835_1_Sinfantarius_@WP_015695665_1
1076934_000441535_1_Slutetiensis_@WP_043895121_1
1311_001026925_1_Sagalactiae_@WP_000215421_1
1117941_000478255_2_Llactis_@AII13818_1
420890_000269945_1_Lgarvieae_@WP_003133075_1
297352_000981525_1_Lpiscium_@WP_050702934_1
1307_000993745_1_Ssuis_@WP_012774919_1
1437447_000817065_1_Smutans_@WP_002262626_1
1114965_000262145_1_Sparasanguini_@WP_003008835_1
388919_000014205_1_Ssanguinis_@YP_001034331_1
1302863_000385925_1_Soligoferment_@WP_015605274_1
1338_001296205_1_Sintermedius_@ALF27117_1
1353243_000478925_1_Sanginosus_@WP_003034640_1
862968_000463445_1_Sconstellatus_@WP_003029403_1
467705_000017005_1_Sgordonii_@WP_041789484_1
1054460_000221985_1_Spseudopneumo_@WP_000021242_1
28037_001281025_1_Smitis_@WP_000021242_1
171101_000007045_1_Spneumoniae_@NP_358489_1
1419814_000688775_1_Ssp_@WP_000021248_1
927666_000253155_1_Soralis_@WP_000021250_1
1308_000971665_1_Sthermophilus_@WP_011681527_1
1304_000785515_1_Ssalivarius_@WP_002884601_1
218495_000009545_1_Suberis_@WP_015911912_1
1346_000831485_1_Siniae_@WP_003102177_1
873447_000187935_2_Sparauberis_@EGE54913_1
552526_000020765_1_Sequi_@WP_012514949_1
160491_000009385_1_Spyogenes_@WP_002992269_1
617121_000307185_1_Sdysgalactiae_@WP_002992269_1
0.4

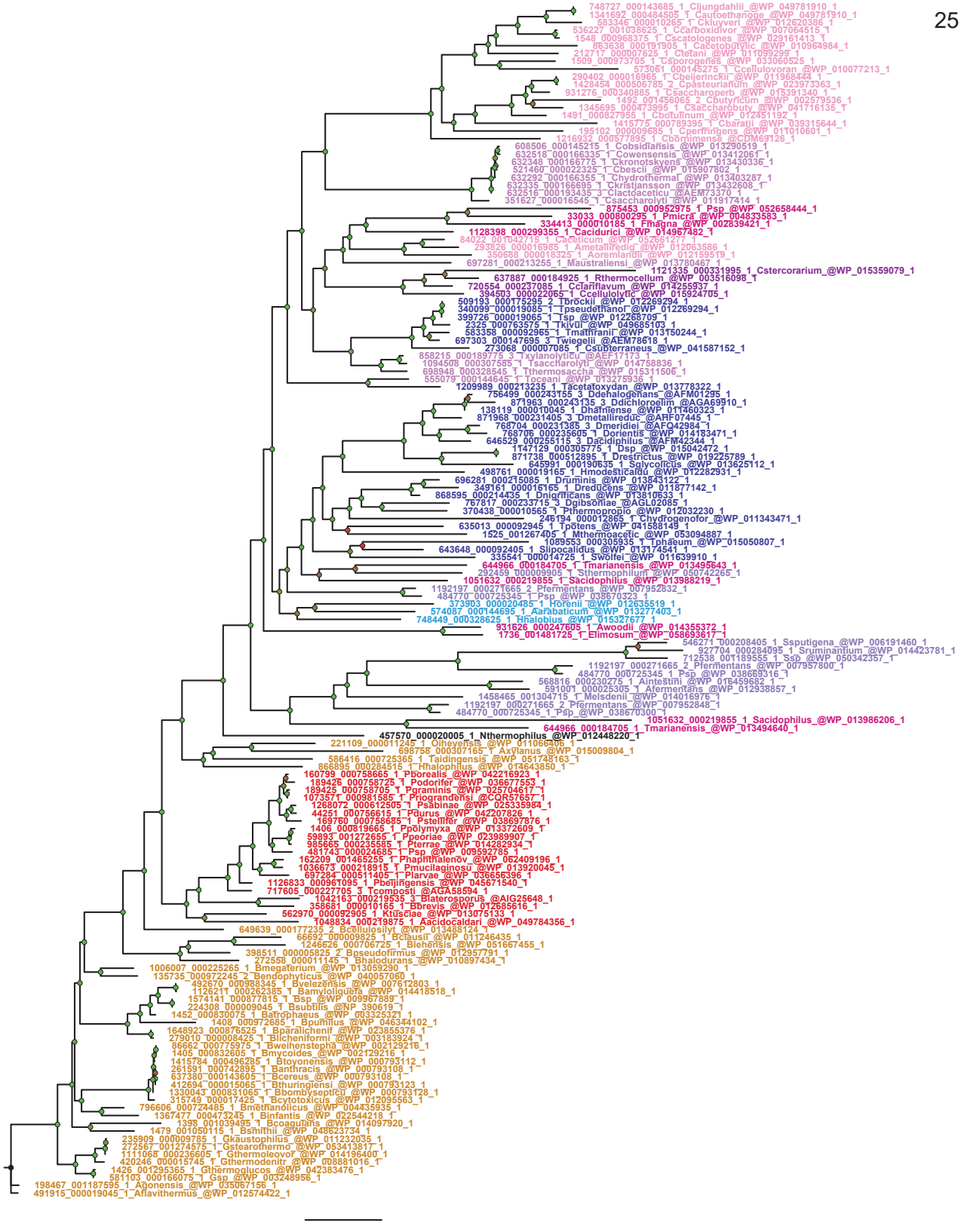

767455_000182835_1_Ldelbrueckii_@WP_013439497_1
1545702_000761135_1_Lsp_@WP_038522634_1
1600_001042405_1_Lacetotolerans_@WP_056969664_1
324831_000014425_1_Lgasseri_@WP_003647275_1
257314_000008065_1_Llohnsonii_@WP_004897298_1
1033837_000214795_1_Lkefiranofaci_@WP_013854003_1
695562_000194115_1_Lamylovorus_@WP_013438048_1
272621_000011985_1_Lacidophilus_@YP_194033_1
880633_000165775_3_Lhelveticus_@AFR21919_1
52242_001314245_1_Lgallinarum_@WP_060472634_1
420890_000269945_1_Lgarvieae_@WP_014025213_1
1117941_000478255_2_Llactis_@AII13407_1
297352_000981525_1_Lpiscium_@WP_047915073_1
1007676_001050475_1_Lginsenosidim_@WP_048703135_1
1074467_000831645_3_Lhelongjiang_@WP_041501035_1
979982_000219785_1_Lsp_@WP_013102425_1
762551_000092505_1_Llimchii_@WP_013102425_1
762550_000196855_1_Lgelidum_@WP_013231692_1
349519_000026405_1_Lcitreum_@WP_004904109_1
1229758_000300135_1_Lcamosum_@WP_014974480_1
1107980_000234825_3_Lmesenteroide_@WP_014324905_1
1045004_000241055_1_Okitaharae_@WP_007746369_1
203123_000014385_1_Oeeni_@WP_002818919_1
148814_001314945_1_Llunkeii_@WP_034531123_1
202752_000763475_1_Livanovii_@WP_035405242_1
683837_000027145_1_Lseeligeri_@WP_012986030_1
1230340_000318055_1_Lmonocytogene_@WP_010958965_1
386043_000060285_1_Lwelshimeri_@WP_041176445_1
219334_000829055_1_Lcasei_@WP_025012400_1
1226298_000829035_1_Lparacasei_@WP_003565619_1
568704_000026525_1_Lhamnosus_@WP_005689429_1
637971_001050435_1_Lkoreensis_@WP_048736116_1
1001583_000359625_1_Lbrevis_@WP_021742335_1
60520_001443645_1_Lparaplantaru_@WP_033609450_1
889932_000148815_2_Lplantarum_@WP_003640471_1
759620_000750515_1_Wceti_@WP_009765287_1
137591_001308145_1_Wcibaria_@WP_029237511_1
1045854_000219805_1_Wkoreensis_@WP_006845280_1
1130798_000248095_3_Lmucosae_@ATT50351_1
1358027_000439275_1_Lreuteri_@WP_019253608_1
1291742_000829395_1_Lhokkaidonens_@WP_041093508_1
334390_000010145_1_Lfermentum_@WP_004563226_1
1194971_001011095_1_Lsalivariu_@WP_014568338_1
1069534_000224985_1_Lrumenis_@WP_014073398_1
714313_000225325_1_Lsanfrancisce_@WP_014081977_1
511437_000211375_1_Lbuchneri_@WP_013727882_1
701521_000237995_2_Pclaussenii_@AEV95288_1
1408206_000496265_1_Ppentosaceus_@WP_023440346_1
314315_000026065_1_Lsakei_@WP_011374615_1
467705_000017005_1_Sgordonii_@WP_012000688_1
388919_000014205_1_Ssanguinis_@YP_001035332_1
1307_000993745_1_Ssuis_@WP_011922751_1
1302863_000385925_1_Soligofermen_@PSDG1_434919_4e_07
862968_000463445_1_Sconstellatus_@WP_006267692_1
1353243_000478925_1_Sanginosus_@WP_003036639_1
1338_001296205_1_Sintermedius_@KLE27695_1
28037_001281025_1_Smitis_@WP_060627787_1
171101_000007045_1_Spneumoniae_@NP_358993_1
1054460_000221985_1_Spseudopneumo_@WP_000747177_1
1419814_000688775_1_Ssp_@WP_000778856_1
927666_000253155_1_Soralis_@WP_000776850_1
1304_000785515_1_Ssalivarius_@WP_004182029_1
1308_000971655_1_Sthermophilus_@WP_011681058_1
1114965_000262145_1_Sparasanguini_@WP_045764029_1
990317_000203195_1_Sgallolyticus_@WP_012962206_1
1116231_000283635_1_Smacedonicus_@WP_014294949_1
981540_000270165_1_Spasteurianus_@WP_003065660_1
1076934_000441535_1_Slutetiensis_@WP_020917168_1
1069533_000246835_1_Sinfantarius_@WP_014335027_1
552526_000020765_1_Sequi_@WP_012515365_1
53345_001267395_1_Edurans_@WP_005875558_1
768486_000271405_2_Ehirae_@WP_010718499_1
1352_001412695_1_Efaecium_@WP_016631404_1
1300150_000504125_1_Emundtii_@WP_010734977_1
332949_001465115_1_Esilesiacus_@ALS00255_1
1261557_000317915_1_Efaecalis_@WP_010816150_1
1090974_000283975_1_Mplutonius_@WP_013774099_1
565655_000157355_2_Ecasseliflavu_@WP_015510034_1
945021_000283615_1_Thalophilus_@WP_014125012_1
866775_000193205_1_Aurinae_@WP_049776773_1
1266845_000493735_1_Cinhibens_@WP_023177744_1
1564681_001483965_1_Csp_@WP_058913390_1
1234679_000317975_2_Cmaltaromatic_@WP_010051220_1
1398_001039495_1_Bcoagulans_@WP_014098186_1
1479_001050115_1_Bsmithii_@WP_048623548_1
1508404_000818095_1_Jmalaysiensis_@WP_039809352_1
169760_000758685_1_Pstellifer_@WP_038696991_1
44251_000756615_1_Pdurus_@WP_042207271_1
1268072_000612505_1_Psabinae_@WP_025335540_1
160799_000758665_1_Pbozealis_@WP_042Z15454_1
189426_000758725_1_Podorifer_@WP_036676299_1
1073571_000981585_1_Priograndensi_@CQR56833_1
189425_000758705_1_Pgramnis_@WP_025703685_1
481743_000024685_1_Psp_@WP_015734564_1
1406_000819665_1_Ppolymyxa_@WP_040103505_1
985665_000235585_1_Pterrae_@WP_014282055_1
59893_001272655_1_Ppeoriae_@WP_053325661_1
1616788_001421015_1_Pbovis_@WP_060535524_1
1126833_000961095_1_Pbeijingensis_@WP_045672555_1
562970_000092905_1_Ktusciae_@WP_013075674_1
717605_000227705_3_Tcomposti_@AGA57941_1
162209_001465255_1_Pnaphthalenov_@WP_062408832_1
103673_000218915_1_Pmucilaginosu_@WP_013916553_1
697284_000511405_1_Plarvae_@WP_023482691_1
1042163_000219535_3_Blaterosporus_@AIG26282_1
35801_000010165_1_Bbrevis_@WP_012686194_1
1648923_000876525_1_Bparabrevi_@WP_020451926_1
279010_000008425_1_Bhalodurans_@WP_003182911_1
1408_000972685_1_Bpumilus_@WP_007500889_1
1452_000830075_1_Batrophaeus_@WP_003325640_1
1574141_000877815_1_Bsp_@WP_014477164_1
224308_000009045_1_Bsubtilis_@NP_390119_1
1126211_000262385_1_Bamyloliquefa_@WP_014418238_1
492670_000988345_1_Bvelezensis_@WP_007409483_1
403957_000725285_1_Vsp_@WP_040955511_1
221109_000011245_1_Olleyensis_@WP_011066160_1
1117068_000236605_1_Gthermodenov_@WP_014196187_1
235909_000009785_1_Gkaustophilus_@WP_011231658_1
272567_001274575_1_Gstearothermo_@WP_015375254_1
420246_000015745_1_Gthermodenitr_@WP_008879582_1
1426_001295365_1_Gthermoglucos_@WP_060678137_1
581103_000166075_1_Gsp_@WP_013400526_1
198467_001187595_1_Agonensis_@WP_033383404_1
491915_000019045_1_Alfalvithermus_@WP_012574761_1
1145276_000600105_1_Lvarians_@WP_024363359_1
444177_000017965_1_Lsphaericus_@WP_012293740_1
28031_000724775_3_Lfusiformis_@WP_036124721_1
1002809_000271325_1_Sslivestris_@WP_008304867_1
1374_001465835_1_Pkocurii_@WP_058385065_1
1526927_000785555_1_Psp_@WP_038704132_1
200991_001465795_1_Prifietoensis_@WP_058380583_1
637380_000143605_1_Bcereus_@WP_000758753_1
261591_000742895_1_Banthracis_@WP_000758754_1
412694_000015065_1_Bthuringiensi_@WP_000758754_1
1415784_000496285_1_Bcoyonensis_@WP_000758736_1
1330043_000831065_1_Bbombysepticu_@WP_000758739_1
86662_000775975_1_Bweihenstepha_@WP_002030785_1
1405_000832605_1_Bmycoides_@WP_002126474_1
315749_000017425_1_Bpyrotoxicus_@WP_012093756_1
1367477_000473245_1_Bnhatnis_@WP_009791258_1
796606_000724485_1_Bmethanolicus_@WP_004435217_1
262543_000019905_1_Esibiricum_@WP_012370645_1
1087448_000299435_1_Eantarcticum_@WP_014970589_1
1399115_000496635_1_Esp_@WP_023468619_1
698758_000307165_1_Axylanus_@WP_015010000_1
866895_000284515_1_Hhalophilus_@WP_014643488_1
586416_000725365_1_Taidingensis_@WP_038560853_1
1048834_000219875_1_Aacidocaldari_@WP_041695157_1
135735_000972245_2_Bendophyticus_@WP_040058250_1
1006007_000225265_1_Bmegaterium_@WP_013058703_1
439292_000093085_1_Bselenitiredu_@WP_013173023_1
649639_000177235_2_Bcellulosilyt_@WP_013488509_1
272558_000011145_1_Bhalodurans_@WP_010897855_1
650150_000270085_1_Erhusiopathia_@WP_003773269_1
398511_000005825_2_Bpseudofirmus_@WP_012958540_1
66692_000009825_1_Bclausii_@WP_011246903_1
1246626_000706725_1_Blehensis_@WP_038480154_1
0.5

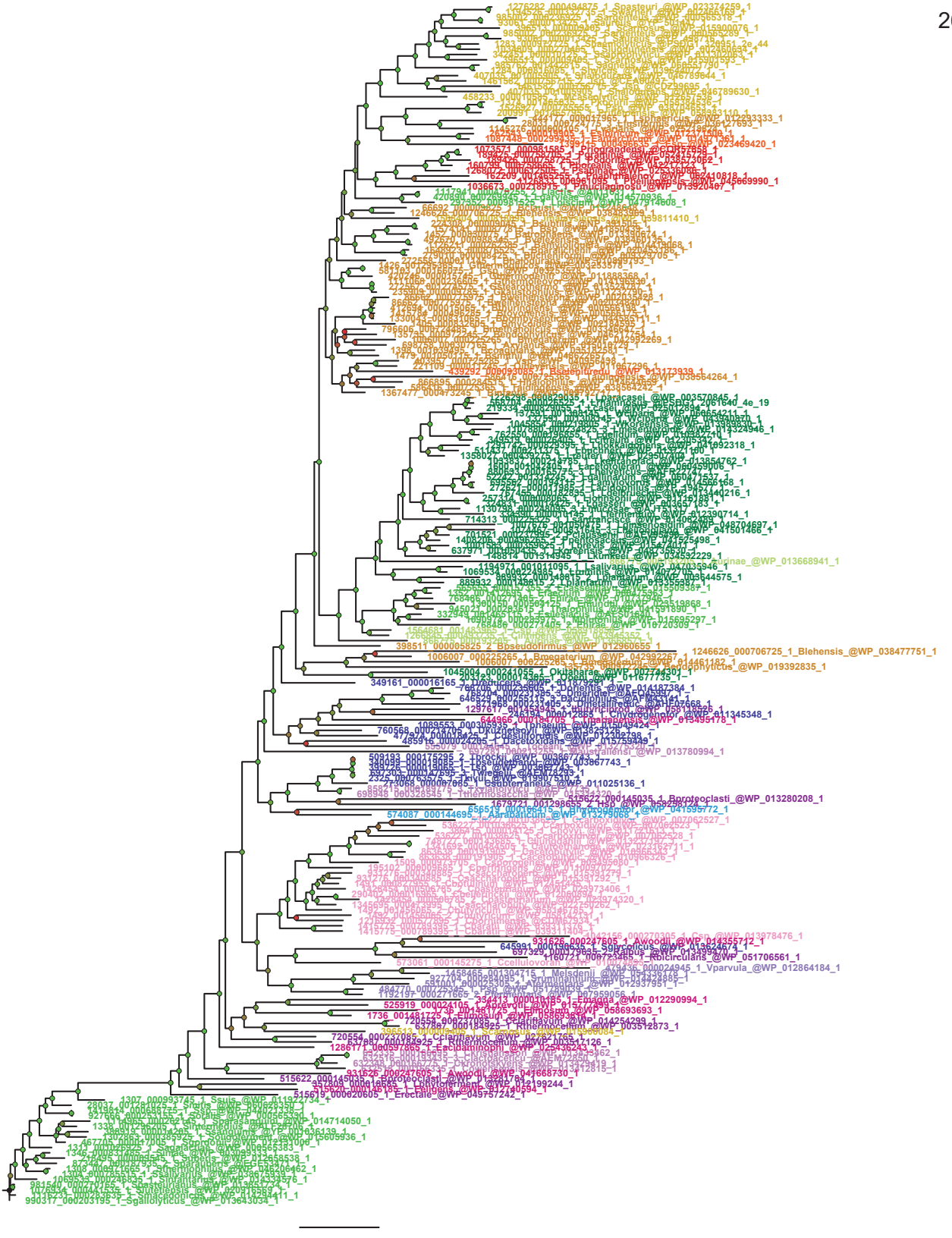

1041504_000283555_1_Carthromitus_@WP_044035675_1
0.9

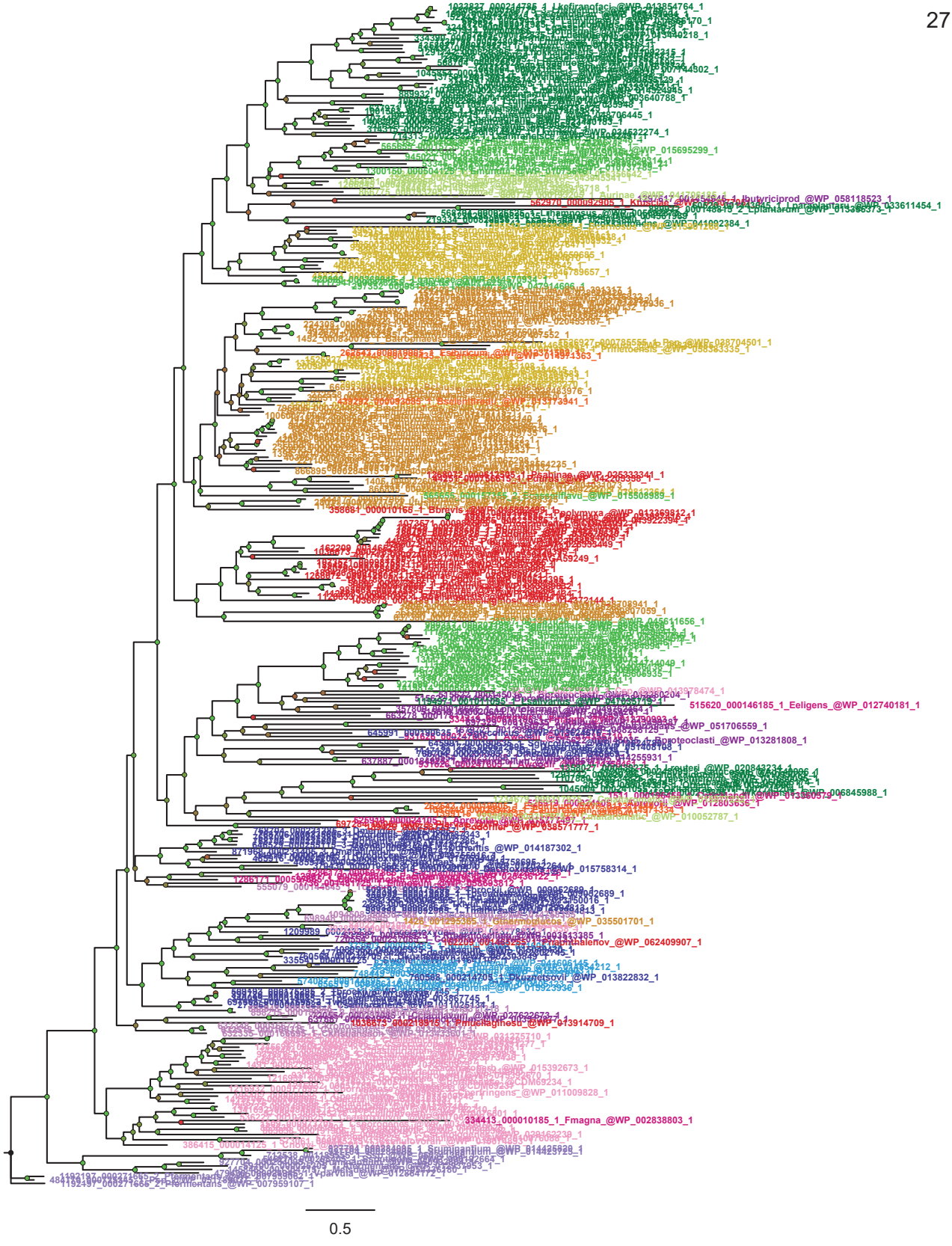

1574141_000877815_1_Bsp_@WP_014480712_1
224308_000009045_1_Bsubtilis_@NP_391058_1
1452_000830075_1_Batrophaeus_@WP_010789794_1
1126211_000262385_1_Bamyloliquefa_@WP_014418769_1
492670_000988345_1_Bvelezensis_@WP_014418769_1
279010_000008425_1_Blicheniformi_@WP_003184877_1
1648923_000876525_1_Bparalichenif_@WP_043054454_1
1408_000972685_1_Bpumilus_@WP_008345881_1
1367477_000473245_1_Binfantis_@WP_022543875_1
135735_000972245_2_Bendophyticus_@WP_040056472_1
1006007_000225265_1_Bmegaterium_@WP_025751972_1
261591_000742895_1_Banthracis_@WP_000837886_1
412694_000015065_1_Bthuringiensi_@WP_000837886_1
637380_000143605_1_Bcereus_@WP_000837882_1
1415784_000496285_1_Btoyonensis_@WP_000837897_1
1330043_000831065_1_Bbombysepticu_@WP_000837892_1
86662_000775975_1_Bweihenstepha_@WP_033709203_1
1405_000832605_1_Bmycoides_@WP_000837889_1
315749_000017425_1_Bcytotoxicus_@WP_012094461_1
1300150_000504125_1_Emundtii_@WP_010734294_1
53345_001267395_1_Edurans_@WP_005878965_1
768486_000271405_2_Ehirae_@WP_010738456_1
1352_001412695_1_Efaecium_@WP_002287791_1
332949_001465115_1_Esilesiacus_@ALS01628_1
683837_000027145_1_Lseeligeri_@WP_012985221_1
202752_000763475_1_Livanovii_@WP_038408710_1
386043_000060285_1_Lwelshimeri_@WP_011701724_1
1230340_000318055_1_Lmonocytogene_@WP_003724835_1
565655_000157355_2_Ecasseliflavu_@WP_010748435_1
1300150_000504125_1_Emundtii_@WP_019723357_1
1234679_000317975_2_Cmaltaromatic_@WP_015077058_1
1234679_000317975_2_Cmaltaromatic_@WP_015077389_1
1266845_000493735_1_Cinhibens_@WP_023177342_1
1564681_001483965_1_Csp_@WP_058918096_1
1564681_001483965_1_Csp_@WP_058919845_1
1234679_000317975_2_Cmaltaromatic_@WP_010050658_1
171101_000007045_1_Spneumoniae_@NP_358371_1
1054460_000221985_1_Spseudopneumo_@WP_000489838_1
28037_001281025_1_Smitis_@WP_045611429_1
1419814_000688775_1_Ssp_@WP_000489828_1
927666_000253155_1_Soralis_@WP_000489823_1
467705_000017005_1_Sgordonii_@WP_012130715_1
1302863_000385925_1_Soligoferment_@WP_015605492_1
388919_000014205_1_Ssanguinis_@YP_001035694_1
1338_001296205_1_Sintermedius_@ALF28032_1
1353243_000478925_1_Sanginosus_@WP_003033878_1
862968_000463445_1_Sconstellatus_@WP_006267252_1
1114965_000262145_1_Sparasanguini_@WP_014713724_1
1437447_000817065_1_Smutans_@WP_002262578_1
1117941_000478255_2_Llactis_@AII13377_1
420890_000269945_1_Lgarvieae_@WP_004259260_1
297352_000981525_1_Lpiscium_@WP_047916272_1
1352_001412695_1_Efaecium_@WP_002290821_1
53345_001267395_1_Edurans_@WP_025481369_1
1300150_000504125_1_Emundtii_@WP_023519400_1
768486_000271405_2_Ehirae_@WP_010719694_1
565655_000157355_2_Ecasseliflavu_@WP_005225866_1
1090974_000283975_1_Mplutonius_@PSDG1_358728_3e_57
1261557_000317915_1_Efaecalis_@WP_002355540_1
332949_001465115_1_Esilesiacus_@ALS01436_1
568704_000026525_1_Lrhamnosus_@WP_005685419_1
219334_000829055_1_Lcasei_@WP_025012681_1
1226298_000829035_1_Lparacasei_@WP_003594900_1
314315_000026065_1_Lsakei_@WP_011375162_1
1194971_001011095_1_Lsalivarius_@WP_047035287_1
1069534_000224985_1_Lruminis_@WP_014073202_1
762051_000092505_1_Lkimchii_@WP_013103052_1
979982_000219785_1_Lsp_@WP_013103052_1
762550_000196855_1_Lgelidum_@WP_013231207_1
349519_000026405_1_Lcitreum_@WP_004899530_1
1229758_000300135_1_Lcarnosum_@WP_014974927_1
1107880_000234825_3_Lmesenteroide_@WP_014325136_1
203123_000014385_1_Ooeni_@WP_011677357_1
1045004_000241055_1_Okitaharae_@WP_007744825_1
759620_000750515_1_Wceti_@WP_009765183_1
137591_001308145_1_Wcibaria_@WP_010373199_1
1045854_000219805_1_Wkoreensis_@WP_013989459_1
148814_001314945_1_Lkunkeei_@WP_034531601_1
714313_000225325_1_Lsanfrancisce_@WP_014082307_1
511437_000211375_1_Lbuchneri_@WP_041805953_1
637971_001050435_1_Lkoreensis_@WP_048735720_1
1001583_000359625_1_Lbrevis_@WP_041815409_1
334390_000010145_1_Lfermentum_@WP_012391535_1
1130798_000248095_3_Lmucosae_@AJT50014_1
1358027_000439275_1_Lreuteri_@WP_020843219_1
1291742_000829395_1_Lhokkaidonens_@WP_041093022_1
1408206_000496265_1_Ppentosaceus_@WP_002833911_1
701521_000237995_2_Pclaussenii_@AEV95466_1
60520_001443645_1_Lparaplantaru_@WP_033610094_1
889932_000148815_2_Lplantarum_@WP_003643221_1
0.4

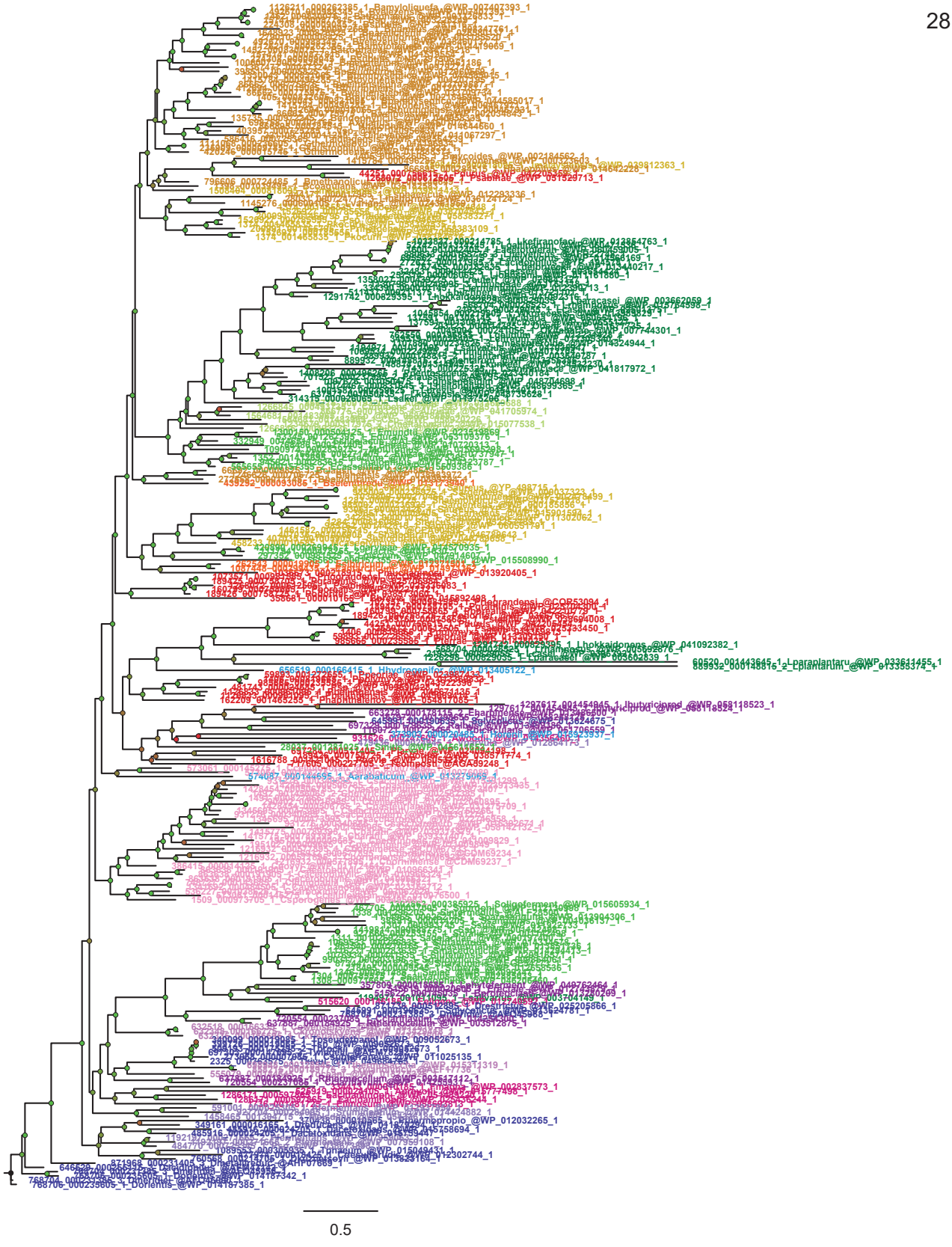

1276282_000494875_1_Spasteuri_@WP_002465608_1
1194526_000332735_1_Swarneri_@WP_002465608_1
72758_001028645_1_Scapitis_@WP_002433349_1
1282_000759555_1_Sepidermidis_@WP_002456101_1
93061_000013425_1_Saureus_@YP_499501_1
985002_000236925_1_Sargenteus_@WP_000933106_1
1283_000972725_1_Shaemolyticus_@WP_011276216_1
1034809_000270465_1_Slugdunensis_@WP_002459499_1
396513_000009405_1_Scarnosus_@WP_015899871_1
342451_000010125_1_Ssaprophyticu_@WP_011303477_1
246432_001432245_1_Sequorum_@WP_002508087_1
1288_000709415_1_Sxylosus_@WP_042362979_1
1284_000816085_1_Shyicus_@WP_039646471_1
985762_001442815_1_Sagnetis_@WP_060550690_1
1295_001188875_1_Sschleiferi_@WP_050331439_1
937773_000185885_1_Spseudinterme_@WP_015728893_1
458233_000010585_1_Mcaseolyticus_@WP_041636337_1
28031_000724775_3_Lfusiformis_@WP_008175205_1
444177_000017965_1_Lsphaericus_@WP_031416553_1
1145276_000600105_1_Lvarians_@WP_024361388_1
1002809_000271325_1_Ssilvestris_@WP_014822503_1
1246626_000706725_1_Blehensis_@WP_038476405_1
1461582_000756715_2_Jsp_@CDZ99207_1
407035_001005905_1_Shalodurans_@WP_046789558_1
407035_001005905_1_Shalodurans_@WP_046789108_1
1374_001465835_1_Pkocurii_@WP_058386485_1
1526927_000785555_1_Psp_@WP_038702383_1
200991_001465795_1_Prifietoensis_@WP_058381912_1
1461582_000756715_2_Jsp_@CDZ99004_1
403957_000725285_1_Vsp_@WP_040956605_1
221109_000011245_1_Oiheyensis_@WP_011067434_1
698758_000307165_1_Axylanus_@WP_015010801_1
866895_000284515_1_Hhalophilus_@WP_014644971_1
586416_000725365_1_Taidingensis_@WP_038563952_1
586416_000725365_1_Taidingensis_@WP_038560455_1
574087_000144695_1_Aarabaticum_@WP_013278363_1
656519_000166415_1_Hhydrogenifor_@WP_013405559_1
748449_000328625_1_Hhalobius_@WP_015327024_1
656519_000166415_1_Hhydrogenifor_@WP_013406551_1
572479_000165465_1_Hpraevalens_@WP_014553938_1
1548_000968375_1_Cscatologenes_@WP_029163280_1
536227_001038625_1_Ccarboxidivor_@WP_050759712_1
1415775_000789395_1_Cbaratii_@WP_039312411_1
394503_000022065_1_Ccellulolytic_@WP_015925958_1
720554_000237085_1_Cclariflavum_@WP_014253508_1
1374_001465835_1_Pkocurii_@WP_058386509_1
1526927_000785555_1_Psp_@WP_038702343_1
200991_001465795_1_Prifietoensis_@WP_058381935_1
1367477_000473245_1_Binfantis_@WP_009791079_1
866895_000284515_1_Hhalophilus_@WP_014642002_1
1036673_000218915_1_Pmucilaginosu_@WP_013918871_1
1399115_000496635_1_Esp_@WP_023468974_1
262543_000019905_1_Esibiricum_@WP_012371008_1
1087448_000299435_1_Eantarcticum_@WP_051004090_1
262543_000019905_1_Esibiricum_@WP_012370300_1
1087448_000299435_1_Eantarcticum_@WP_051004053_1
1087448_000299435_1_Eantarcticum_@WP_014971385_1
262543_000019905_1_Esibiricum_@WP_012371528_1
1399115_000496635_1_Esp_@WP_023469482_1
135735_000972245_2_Bendophyticus_@WP_046217706_1
200991_001465795_1_Prifietoensis_@WP_058383013_1
1073571_000981585_1_Priograndensi_@CQR56523_1
189425_000758705_1_Pgraminis_@WP_025705447_1
160799_000758665_1_Pborealis_@WP_042216430_1
189426_000758725_1_Podorifer_@WP_036688627_1
481743_000024685_1_Psp_@WP_015737582_1
985665_000235585_1_Pterrae_@WP_014281702_1
59893_001272655_1_Ppeoriae_@WP_023987447_1
1406_000819665_1_Ppolymyxa_@WP_040101904_1
1616788_001421015_1_Pbovis_@WP_060530764_1
44251_000756615_1_Pdurus_@WP_042207520_1
1268072_000612505_1_Psabinae_@WP_025335784_1
169760_000758685_1_Pstellifer_@WP_038697386_1
1073571_000981585_1_Priograndensi_@CQR57179_1
189425_000758705_1_Pgraminis_@WP_025703846_1
189426_000758725_1_Podorifer_@WP_036683003_1
481743_000024685_1_Psp_@WP_015734270_1
1126833_000961095_1_Pbeijingensis_@WP_052702710_1
1036673_000218915_1_Pmucilaginosu_@WP_041617630_1
1367477_000473245_1_Binfantis_@WP_009793210_1
439292_000093085_1_Bselenitiredu_@WP_013172447_1
649639_000177235_2_Bcellulosilyt_@WP_013489550_1
1508404_000818095_1_Jmalaysiensis_@WP_039809182_1
439292_000093085_1_Bselenitiredu_@WP_013172226_1
293826_000016985_1_Ametalliredig_@WP_012063316_1
272558_000011145_1_Bhalodurans_@WP_010899068_1
66692_000009825_1_Bclausii_@WP_011245152_1
1246626_000706725_1_Blehensis_@WP_038476059_1
1126211_000262385_1_Bamyloliquefa_@WP_014418301_1
492670_000988345_1_Bvelezensis_@WP_038459042_1
1452_000830075_1_Batrophaeus_@WP_034717903_1
398511_000005825_2_Bpseudofirmus_@WP_012959809_1
0.5

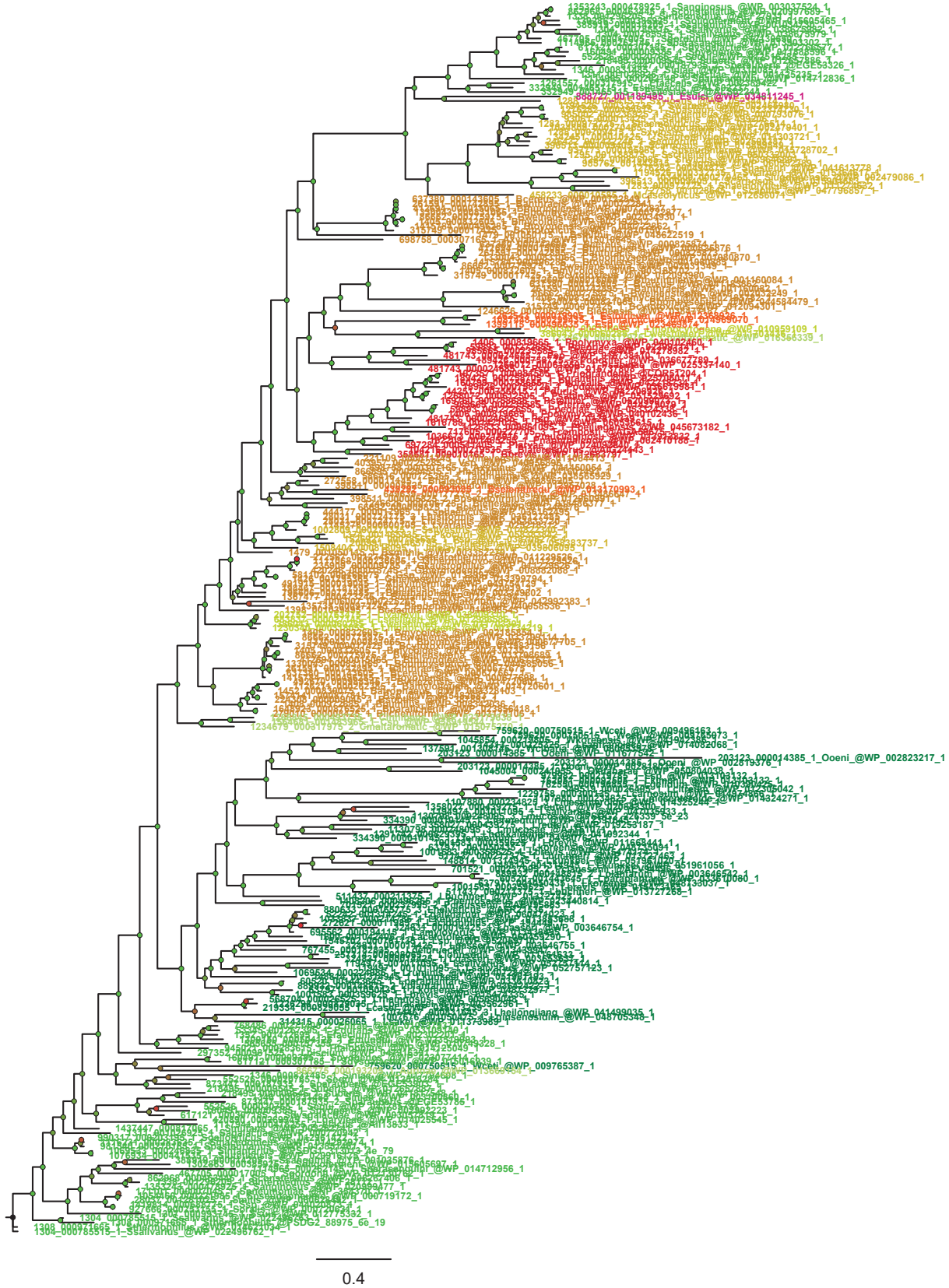
0.5

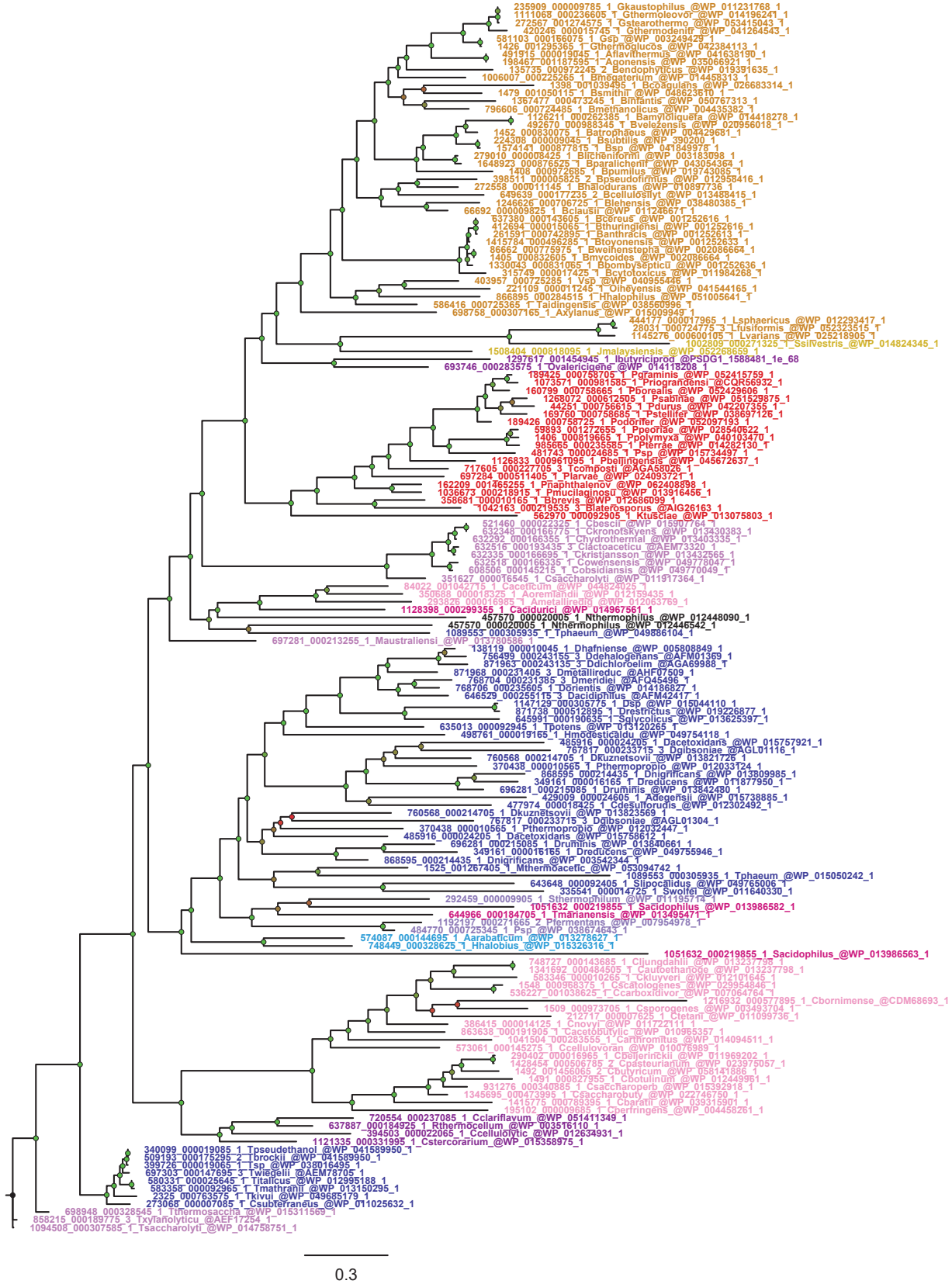

748727_000143685_1_Cljungdahlii_@WP_049781910_1
1341692_000484505_1_Cautoethanoge_@WP_049787910_1
583346_000010265_1_Ckluyveri_@WP_012620386_1
536227_001038625_1_Ccarboxidivor_@WP_007064515_1
1548_000968375_1_Cscatologenes_@WP_029161413_1
863638_000191905_1_Cacetobutylic_@WP_010964984_1
212717_000007625_1_Ctetani_@WP_011099295_1
1509_000973705_1_Csporogenes_@WP_033060525_1
573061_000145275_1_Ccellulovoran_@WP_010077213_1
290402_000016965_1_Cbeijerinckii_@WP_011969444_1
1428454_000506785_2_Cpasteurianum_@WP_023973363_1
931276_000340885_1_Csaccharoperb_@WP_015391340_1
1492_001456065_2_Cbutyricum_@WP_002579536_1
1345695_000473995_1_Csaccharobuty_@WP_041716135_1
1491_000827955_1_Cbotulinum_@WP_012451192_1
1415775_000789395_1_Cbaratii_@WP_039315644_1
195102_000009685_1_Cperfringens_@WP_011010601_1
1216932_000577895_1_Cborrilumens_@CDM69126_1
608506_000145215_1_Cobsidiansis_@WP_013290519_1
632518_000166335_1_Cowensensis_@WP_013412061_1
632348_000166775_1_Ckronotskyens_@WP_013430336_1
521460_000022325_1_Cbescii_@WP_015907802_1
632292_000166355_1_Chydrothermal_@WP_013403287_1
632335_000166695_1_Ckristjansson_@WP_013432608_1
632516_000193435_1_Clactoaceticu_@AEM73370_1
351627_000016545_1_Csaccharolyti_@WP_011917414_1
875453_000952975_1_Psp_@WP_052658444_1
33033_000800295_1_Pmicra_@WP_004833583_1
334413_000010185_1_Fnucina_@WP_002839421_1
1128398_000299355_1_Cacidurici_@WP_014957482_1
84022_001042715_1_Cacticum_@WP_052661277_1
293826_000016985_1_Ametalliredig_@WP_012063586_1
350688_000018325_1_Aoremlandii_@WP_012159519_1
697281_000213255_1_Maustraliensi_@WP_013780467_1
1121335_000331995_1_Cstercorarium_@WP_015359079_1
637887_000184925_1_Rthermocellum_@WP_003516098_1
720554_000237085_1_Cclariflavum_@WP_014255937_1
394503_000022065_1_Ccellulolytic_@WP_015924705_1
509193_000175295_2_Tbrockii_@WP_012269294_1
340099_000019085_1_Tpseudethanol_@WP_012269294_1
399726_000019065_1_Tsp_@WP_012268709_1
2325_000763575_1_Tkivui_@WP_049685103_1
583358_000092965_1_Tmathranii_@WP_013150244_1
697303_000147695_3_Twiegelii_@AEM78618_1
273066_000007085_1_Csubterraneus_@WP_041587152_1
858215_000189775_3_Txylanolyticu_@AEF17173_1
1094508_000307585_1_Tsaccharolyti_@WP_014758836_1
698948_000328545_1_Tthermosaccha_@WP_015311506_1
555079_000144645_1_Toceani_@WP_013275936_1
1209989_000213235_1_Tacetatoxydan_@WP_013778322_1
756499_000243155_3_Ddehalogenans_@AFM01295_1
871963_000243135_3_Ddichloroelim_@AGA69910_1
138119_000010045_1_Dhafniense_@WP_011460323_1
871968_000231405_3_Dmetallireduc_@ARF07445_1
768704_000231385_3_Dmeridiei_@AFQ42984_1
768705_000235605_1_Dorientis_@WP_014183471_1
646529_000255115_3_Dacidiphilus_@AFM42344_1
1147129_000305775_1_Dsp_@WP_015042472_1
871738_000512895_1_Drestrictus_@WP_019225789_1
645991_000190635_1_Sglycolicus_@WP_013625112_1
498761_000019165_1_Hmodesticaldu_@WP_012282931_1
696281_000215085_1_Druminis_@WP_013843122_1
349161_000016165_1_Dreducens_@WP_011877142_1
868595_000214435_1_Dniacificans_@WP_013810633_1
767817_000233715_3_Dgibsoniae_@AGL02085_1
370438_000010565_1_Pthermopropio_@WP_012032230_1
245194_000012865_1_Chydrogenofor_@WP_011343471_1
635013_000092945_1_Tpotens_@WP_041588149_1
1525_001267405_1_Mthermoacetic_@WP_053094887_1
1089553_000305935_1_Tphaeum_@WP_015050807_1
643648_000092405_1_Slipocalidus_@WP_013174541_1
335541_000014725_1_Swolfei_@WP_011639910_1
644966_000184705_1_Tmarianensis_@WP_013495643_1
292459_000009805_1_Sthermophilum_@WP_050742265_1
1051632_000219855_1_Sacidophilus_@WP_013988219_1
1192197_000271665_2_Pfermentans_@WP_007952832_1
484770_000725345_1_Psp_@WP_038670323_1
373903_000020485_1_Hforeni_@WP_012633519_1
574087_000144695_1_Aarabaticum_@WP_013277403_1
748449_000328625_1_Thalobius_@WP_015327677_1
931526_000247505_1_Awoodii_@WP_014355372_1
1736_001481725_1_Elimosum_@WP_058693617_1
546271_000208405_1_Ssputigena_@WP_006191460_1
927704_000284095_1_Sruminantium_@WP_014423781_1
712538_001189555_1_Ssp_@WP_050342357_1
1192197_000271665_2_Pfermentans_@WP_007957800_1
484770_000725345_1_Psp_@WP_038669316_1
568816_000230275_1_Aintestini_@WP_016459682_1
591001_000025305_1_Afermentans_@WP_012938857_1
1458465_001304715_1_Meisdenii_@WP_014016976_1
1192197_000271665_2_Pfermentans_@WP_007952848_1
484770_000725345_1_Psp_@WP_038670300_1
1051632_000219855_1_Sacidophilus_@WP_013986206_1
644966_000184705_1_Tmarianensis_@WP_013494640_1
457570_000020005_1_Nthermophilus_@WP_012448220_1
221109_000011245_1_Clneyensis_@WP_011066406_1
698758_000307165_1_Axylanus_@WP_015009804_1
586416_000725365_1_Taidingensis_@WP_051748163_1
866895_000284515_1_Hhalophilus_@WP_013643850_1
160799_000758665_1_Pborealis_@WP_042216923_1
189426_000758725_1_Podorifer_@WP_036677553_1
189425_000758705_1_Pgraminis_@WP_025704617_1
1073571_000981585_1_Priograndensi_@CQR57657_1
1268072_000612505_1_Psabinae_@WP_025335984_1
44251_000756615_1_Pdurus_@WP_042207826_1
169760_000758685_1_Pstellifer_@WP_038697876_1
1406_000819665_1_Ppolymyxa_@WP_013372609_1
58932_001272655_1_Ppeoriae_@WP_023989907_1
985665_000235585_1_Pterrae_@WP_014282934_1
481743_000024685_1_Psp_@WP_009592785_1
162209_001465255_1_Phaphthalenov_@WP_062409196_1
1036673_000218915_1_Pmucilaginosu_@WP_013920045_1
697284_000511405_1_Plarvae_@WP_036656396_1
1126833_000967095_1_Pbeijingensis_@WP_045671540_1
717605_000227705_3_Tcomposti_@AGA58594_1
1042163_000219535_3_Blaterosporus_@AIG25648_1
358681_000010165_1_Bbrevis_@WP_012685616_1
562970_000092905_1_Ktusciae_@WP_013075133_1
1048834_000219875_1_Aacidocaldari_@WP_049784356_1
649639_000177235_2_Bcellulosilyt_@WP_013488124_1
66692_000009825_1_Bclausii_@WP_011246435_1
1246626_000706725_1_Blehensis_@WP_051667455_1
398511_000005825_2_Bpseudofirmus_@WP_012957791_1
272558_000011145_1_Bhalodurans_@WP_010897434_1
1006007_000225265_1_Bmegaterium_@WP_013059290_1
135735_000972245_2_Bendophyticus_@WP_040057060_1
492670_000988345_1_Bvelezensis_@WP_007612803_1
1126211_000262385_1_Bamyloliquefa_@WP_014418518_1
1574141_000877815_1_Bsp_@WP_009967889_1
224308_000009045_1_Bsubtilis_@NP_390619_1
1452_000830075_1_Batrophaeus_@WP_003325321_1
1408_000972685_1_Bpumilus_@WP_046344102_1
1648923_000876525_1_Bparalichenif_@WP_023855376_1
279010_000008425_1_Blicheniformi_@WP_003183924_1
96662_000775975_1_Bweihenstepha_@WP_002129216_1
1405_000832605_1_Bmycoides_@WP_002129216_1
1415784_000496285_1_Btoyonensis_@WP_000793112_1
261591_000742895_1_Banthracis_@WP_000793108_1
637380_000143605_1_Bcereus_@WP_000793108_1
412694_000015065_1_Bthuringiensi_@WP_000793123_1
1330043_000831065_1_Bbombysepticu_@WP_000793128_1
315749_000017425_1_Bcytotoxicus_@WP_012095563_1
796606_000724485_1_Bmethanolicus_@WP_004435935_1
1367477_000473245_1_Binfantis_@WP_022544218_1
1398_001039495_1_Bcoagulans_@WP_014097920_1
1479_001050115_1_Bsmithii_@WP_048623734_1
235909_000009785_1_Gkaustophilus_@WP_011232035_1
272567_001274575_1_Gstearothermo_@WP_053413817_1
1111068_000236605_1_Gthermoleovor_@WP_014196400_1
420246_000015745_1_Gthermodenitr_@WP_008881016_1
1426_001295365_1_Gthermoglucos_@WP_042383476_1
581103_000166075_1_Gsp_@WP_003248956_1
198467_001187595_1_Agonensis_@WP_035067156_1
491915_000019045_1_Aflavithermus_@WP_012574422_1
0.6

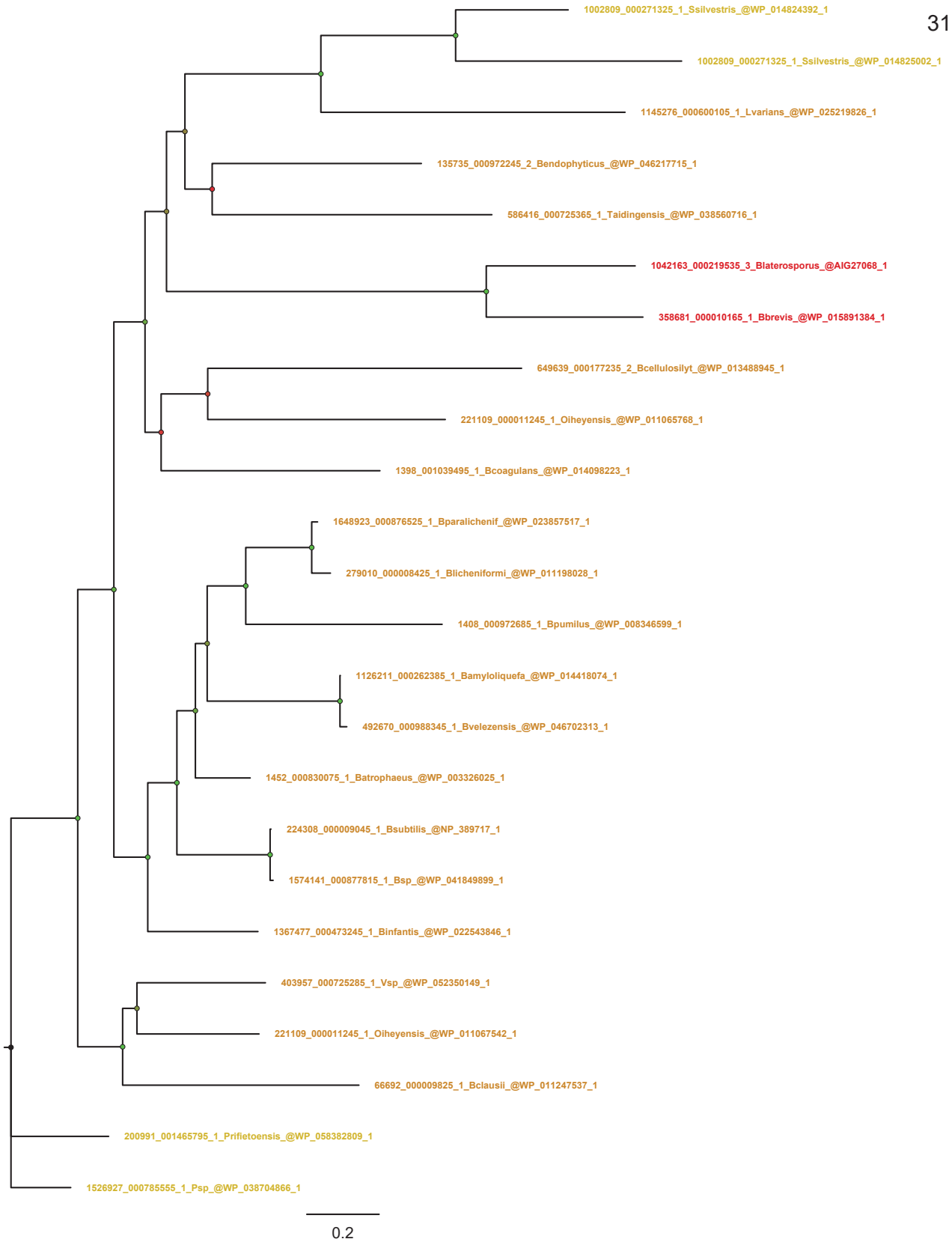

0.6

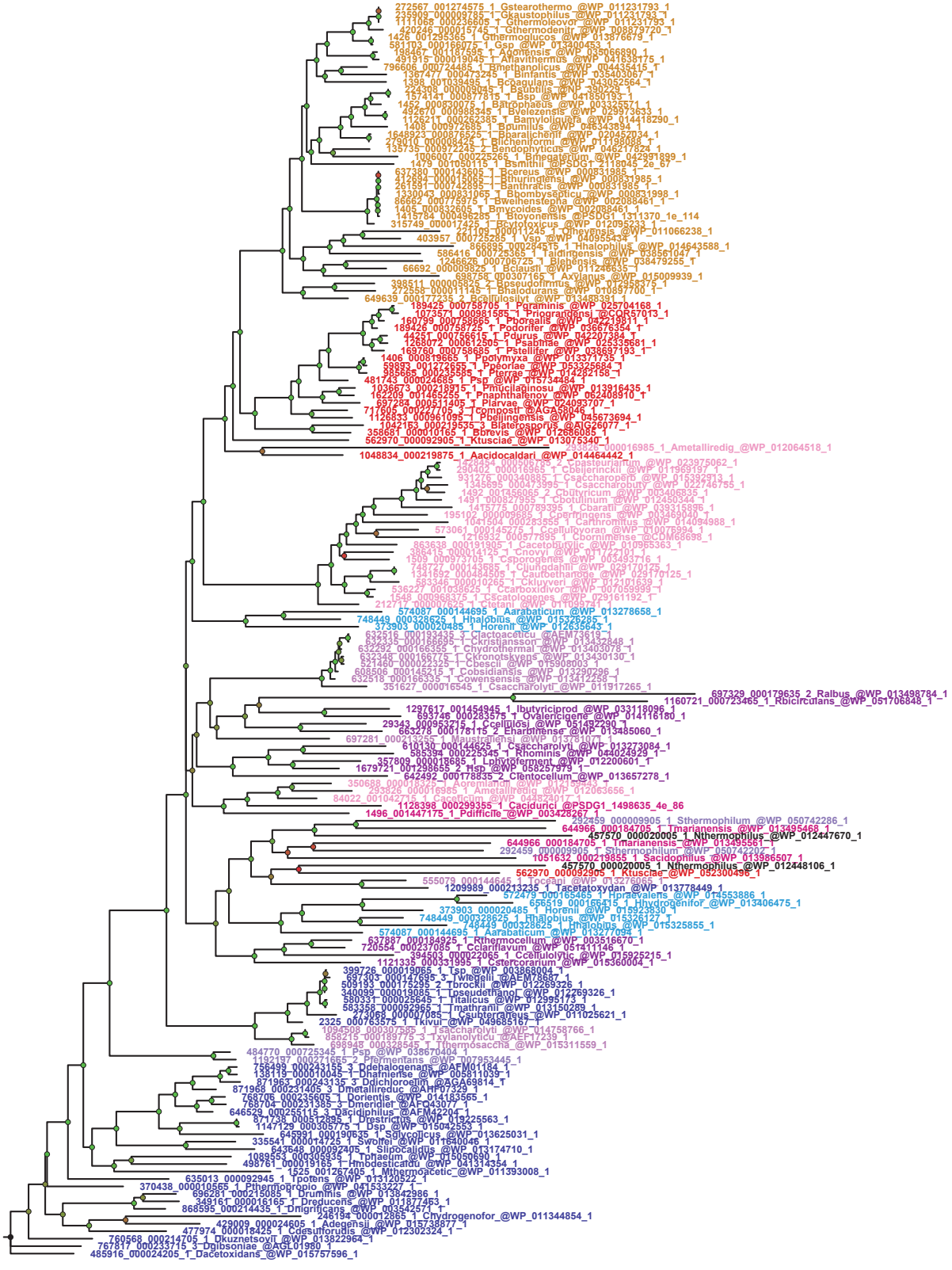

515620_000146185_1_Eeligens_@WP_012740181_1
334413_000010185_1_Fmagna_@WP_002838803_1
0.5

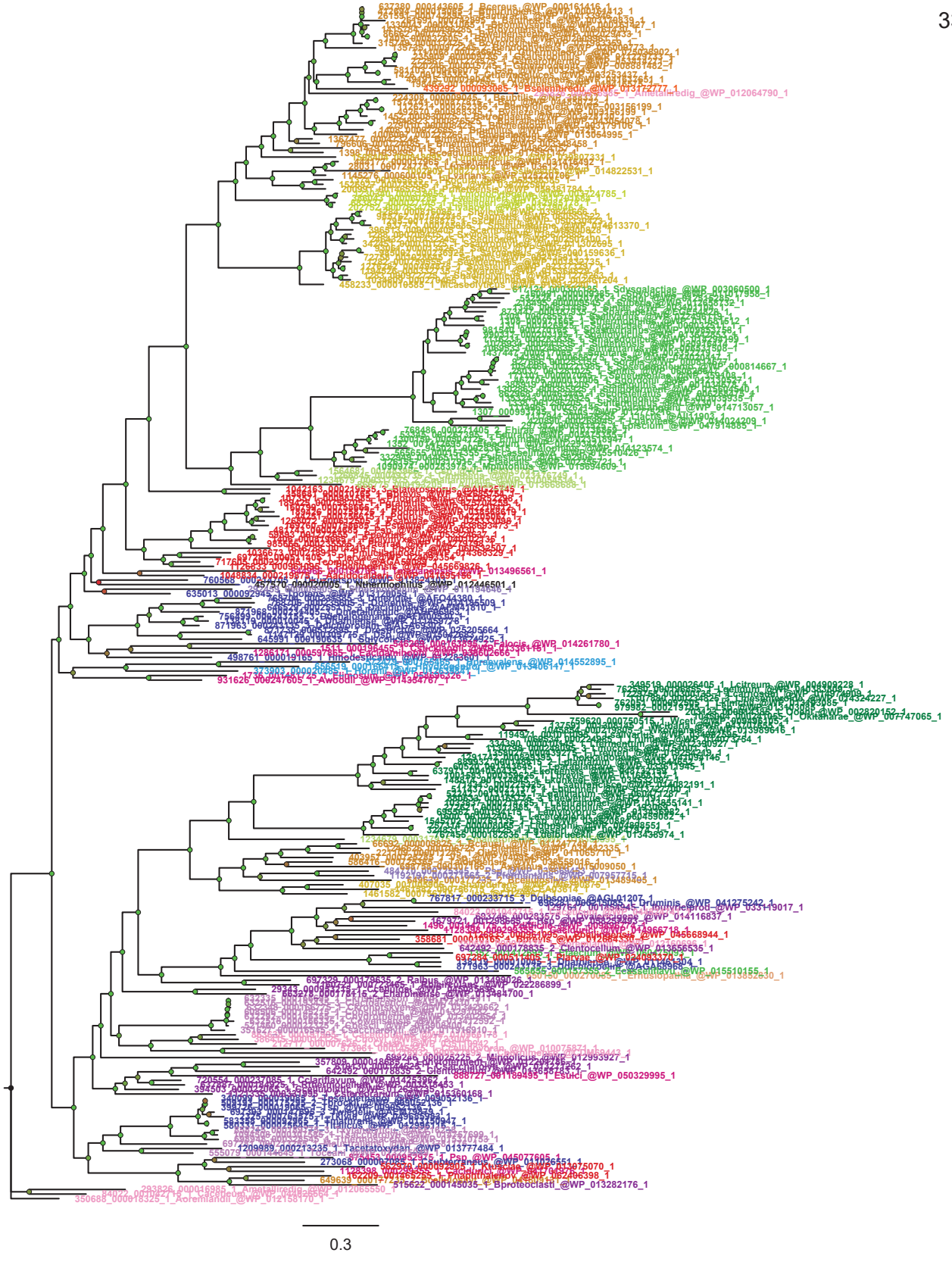

0.5

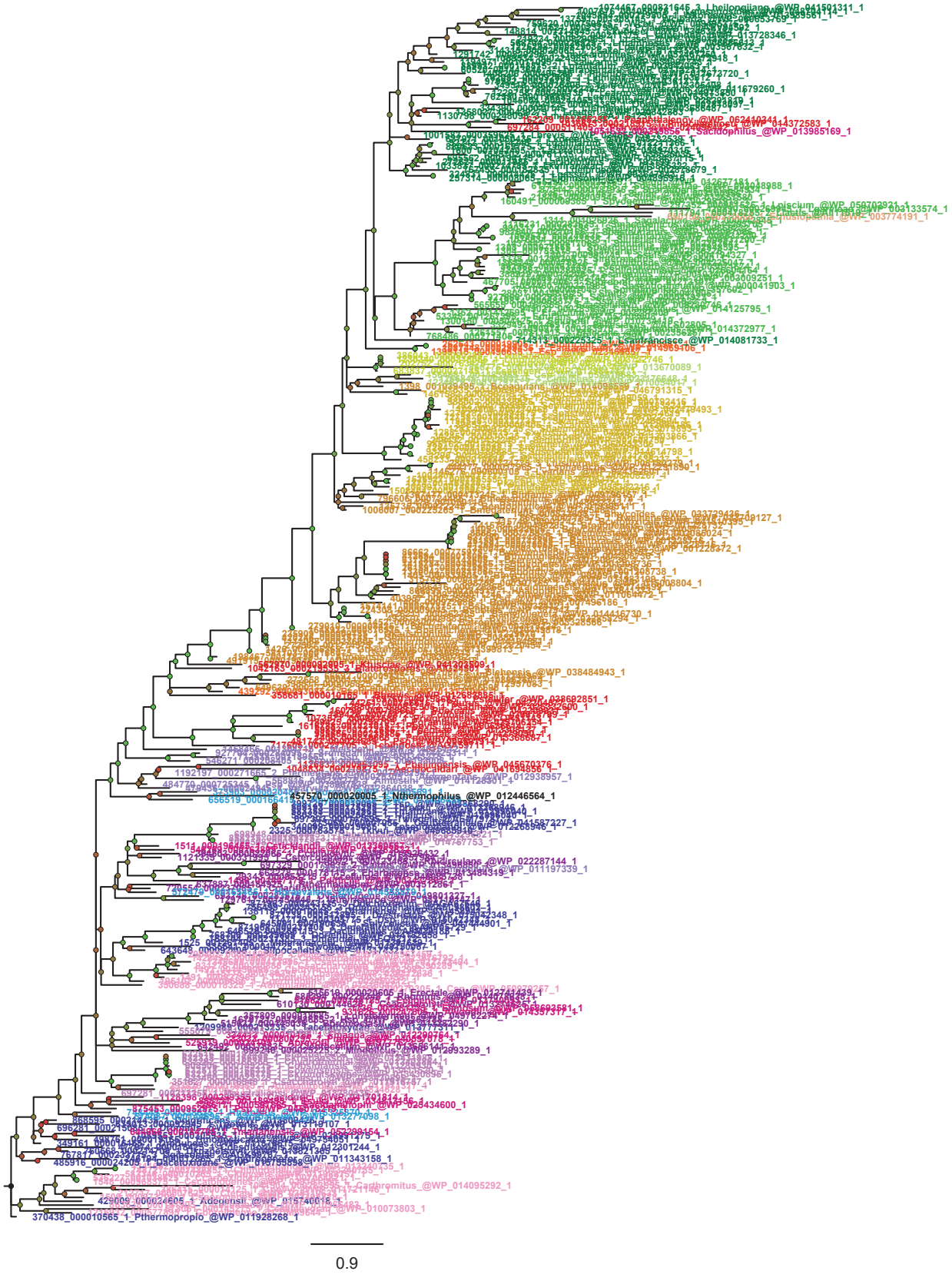

0.4

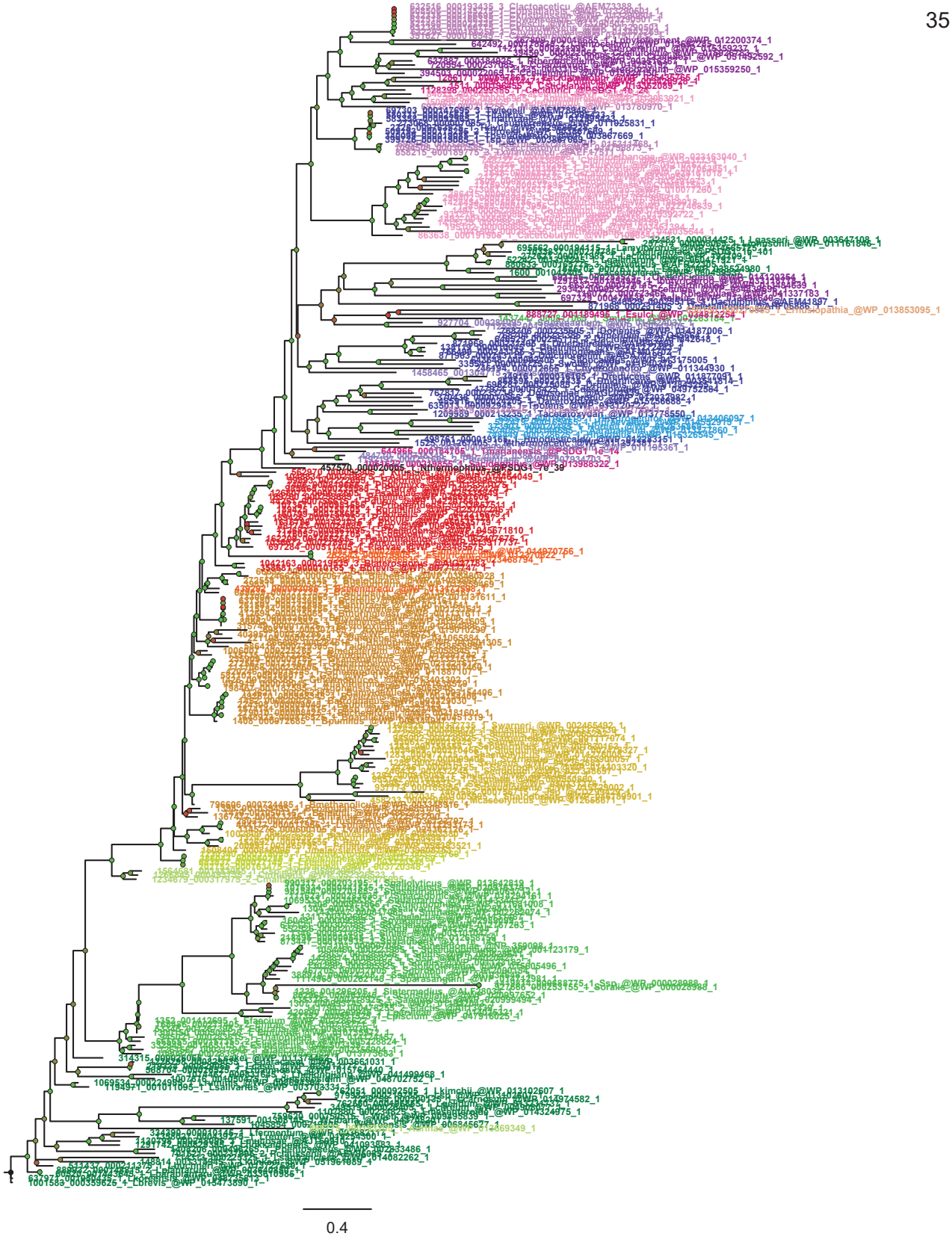

235909_000009785_1_Gkaustophilus_@WP_011231768_1
1111068_000236605_1_Gthermoleovor_@WP_014196241_1
272567_001274575_1_Gstearothermo_@WP_053415043_1
420246_000015745_1_Gthermodenitr_@WP_041264543_1
581103_000186075_1_Gsp_@WP_003249429_1
1428_001295365_1_Gthermoglucos_@WP_042384113_1
491915_000019045_1_Aflavithermus_@WP_041638190_1
198467_001187595_1_Agonensis_@WP_035066921_1
135735_000972245_2_Bendophyticus_@WP_019391635_1
1006007_000225265_1_Bmegaterium_@WP_014458313_1
1398_001039495_1_Bcoagulans_@WP_026683314_1
1479_001050115_1_Bsmithii_@WP_048623610_1
1367477_000473245_1_Binfantis_@WP_050767313_1
796606_000724485_1_Bmethanolicus_@WP_004435382_1
1126211_000262385_1_Bamyloliquefa_@WP_014418278_1
492670_000988345_1_Bvelezensis_@WP_020956018_1
1452_000830075_1_Batrophaeus_@WP_004429681_1
224308_000009045_1_Bsubtilis_@NP_390200_1
1574141_000877815_1_Bsp_@WP_041849978_1
279010_000008425_1_Blicheniformi_@WP_003183098_1
1648923_000876525_1_Bparalicheni_@WP_041305464_1
1408_000972685_1_Bpumilus_@WP_019743085_1
398511_000005825_2_Bpseudofirmus_@WP_012958416_1
272558_000011145_1_Bhalodurans_@WP_010897736_1
649639_000177235_2_Bcellulosilyt_@WP_013488415_1
1246626_000706725_1_Blehensis_@WP_038480385_1
66692_000009825_1_Bclausii_@WP_011246671_1
537380_000743605_1_Bcereus_@WP_001212616_1
412694_000015065_1_Bthuringiensi_@WP_001252616_1
261591_000742895_1_Banthracis_@WP_001252613_1
1415794_000496295_1_Btoyonensis_@WP_001252633_1
86662_000775975_1_Bweihenstepha_@WP_002086664_1
1405_000832605_1_Bmycoides_@WP_002086664_1
1330043_000831065_1_Bbombysepticu_@WP_012958636_1
315749_000017425_1_Bcytotoxicus_@WP_011984268_1
403957_000725285_1_Vsp_@WP_040955446_1
221109_000011245_1_Oihevensis_@WP_041544165_1
866895_000284515_1_Hhalophilus_@WP_051005641_1
586416_000725365_1_Taidingensis_@WP_038560996_1
698758_000307165_1_Axylanus_@WP_015009949_1
444177_000017965_1_Lsphaericus_@WP_012293417_1
28031_000724775_3_Lfusiformis_@WP_052323515_1
1145276_000600105_1_Lvarians_@WP_025218905_1
1002809_000271325_1_Lsilvestris_@WP_014824345_1
1508404_000818095_1_Jmalaysiensis_@WP_052268659_1
1297617_001454945_1_Ibutyriciprod_@PSDG1_1588481_1e-68
693746_000283575_1_Ovalericigene_@WP_014118208_1
189425_000758705_1_Pgaminis_@WP_052415759_1
1073571_000981585_1_Priograndensi_@CQR56932_1
160799_000758665_1_Pborealis_@WP_052429606_1
1268072_000612505_1_Psabinae_@WP_051529875_1
44251_000756615_1_Pdurus_@WP_042207355_1
169760_000758685_1_Pstellifer_@WP_038697126_1
189426_000758725_1_Podorifer_@WP_052097193_1
59893_001272655_1_Ppeoriae_@WP_028540622_1
1408_000819665_1_Ppolymyxa_@WP_040103470_1
985665_000235585_1_Pterrae_@WP_014282130_1
481743_000024685_1_Psp_@WP_015734497_1
1126833_000961095_1_Pbeijingensis_@WP_045672637_1
717605_000227705_3_Tcompositi_@AGA58026_1
697284_000511405_1_Plarvae_@WP_024093721_1
162209_001465255_1_Phaphthalenov_@WP_062408898_1
1036673_000218915_1_Pmucilaginosu_@WP_013916456_1
358661_000010165_1_Bbrevis_@WP_012668099_1
1042163_000219535_3_Blateroporus_@AIG26163_1
562970_000092905_1_Ktusciae_@WP_013075803_1
521460_000022325_1_Cbescii_@WP_015907764_1
632348_000166775_1_Ckronofskyens_@WP_013430383_1
632292_000166355_1_Chydrothermal_@WP_013403335_1
632516_000193435_3_Clactoaceticu_@AEM73320_1
632335_000166895_1_Ckristjansson_@WP_013432565_1
632518_000166335_1_Cowensensis_@WP_049778047_1
608506_000145215_1_Cobsidiansis_@WP_049770049_1
351627_000016545_1_Csaccharolyti_@WP_011917364_1
84022_001042775_1_Cacetobut_@WP_044822025_1
350688_000018325_1_Aoremlandii_@WP_012159435_1
1003605_000016985_1_Ametalliredg_@WP_012063769_1
1128398_000299355_1_Caciduric_@WP_014967561_1
457570_000020005_1_Nthermophilus_@WP_012448090_1
457570_000020005_1_Nthermophilus_@WP_012446542_1
1089553_000305935_1_Tphaeum_@WP_049586104_1
697281_000213255_1_Maustraliensi_@WP_015780585_1
138119_000010045_1_Dhafniense_@WP_005808849_1
756499_000243155_3_Ddehalogenans_@AFM01369_1
871963_000243135_3_Ddichloroelim_@AGA69988_1
871968_000231405_3_Dmetallireduc_@AHF07509_1
768704_000231385_3_Dmeridiei_@AFQ45496_1
768706_000235605_1_Dorientis_@WP_014186827_1
646529_000255115_3_Dacidiphilus_@AFM42417_1
1147129_000305775_1_Dsp_@WP_015044110_1
871738_000512895_1_Drestrictus_@WP_019226877_1
645991_000190635_1_Solvocuprus_@WP_013625397_1
635013_000092945_1_Tpotens_@WP_013120265_1
498761_000019165_1_Hmodesticaldu_@WP_049754118_1
485916_000024205_1_Dacetoxidans_@WP_015757921_1
767817_000233715_3_Dglibsoniae_@AGL01116_1
760568_000214705_1_Dkuznetsovii_@WP_013821726_1
370438_000010565_1_Pthermopropio_@WP_012033124_1
868595_000214435_1_Dhighfican_@WP_013809985_1
349161_000016165_1_Dreducens_@WP_011877950_1
696281_000215085_1_Druminis_@WP_013842480_1
429009_000024605_1_Adegensii_@WP_015738885_1
477974_000018525_1_Cdesulforudis_@WP_012302492_1
760568_000214705_1_Dkuznetsovii_@WP_013823569_1
767817_000233715_3_Dglibsoniae_@AGL01304_1
370438_000010565_1_Pthermopropio_@WP_012032447_1
485916_000024205_1_Dacetoxidans_@WP_015758612_1
696281_000215085_1_Druminis_@WP_013840661_1
349161_000016165_1_Dreducens_@WP_049755946_1
868595_000214435_1_Dnigrificans_@WP_003542344_1
1525_001267405_1_Mthermoacetic_@WP_053094742_1
1089553_000305935_1_Tphaeum_@WP_015050242_1
643648_000092405_1_Slipocalidus_@WP_049765006_1
335541_000014725_1_Swolfei_@WP_011640330_1
292459_000009905_1_Sthermophilum_@WP_011195713_1
1051632_000219855_1_Sacidophilus_@WP_013986582_1
644966_000184705_1_Tmarianensis_@WP_013495471_1
1192197_000271665_2_Ptermentans_@WP_007954978_1
484770_000725345_1_Psp_@WP_038674643_1
574087_000144695_1_Aarabaticum_@WP_013278627_1
748449_000328625_1_Hhalobius_@WP_015326316_1
1051632_000219855_1_Sacidophilus_@WP_013986563_1
748727_000143685_1_Cljungdahlii_@WP_013237798_1
1341692_000484505_1_Cautoethanole_@WP_013237798_1
583346_000010265_1_Ckluyveri_@WP_012101645_1
1548_000968375_1_Cscatologenes_@WP_029954846_1
536227_001038625_1_Ccarboxidivor_@WP_007064764_1
1216932_000577895_1_Cbornimense_@CDM68693_1
1509_000973705_1_Csporogenes_@WP_003493704_1
212717_000007625_1_Ctetani_@WP_011099736_1
386415_000014125_1_Cnovyi_@WP_011722111_1
863638_000191905_1_Cacetobutylic_@WP_010965357_1
1041504_000283555_1_Carthromitus_@WP_014094511_1
573061_000145275_1_Ccellulovoran_@WP_010076989_1
290402_000016965_1_Cbeijerinckii_@WP_011969202_1
1428454_000506785_2_Cpasteurianum_@WP_023975057_1
1492_001456065_2_Cbutyricum_@WP_058141886_1
1491_000827955_1_Cbotulinum_@WP_012449961_1
931276_000340885_1_Csaccharoperb_@WP_015392918_1
1345695_000473995_1_Csaccharobuty_@WP_022746750_1
1415775_000789395_1_Cbaratii_@WP_039315901_1
195102_000009685_1_Cperfringens_@WP_004458261_1
720554_000237085_1_Cclariflavum_@WP_051411349_1
637887_000184925_1_Rthermocellum_@WP_003516110_1
394503_000022065_1_Ccellulolytic_@WP_012634931_1
1121335_000331995_1_Cstercorarium_@WP_015388975_1
340099_000019085_1_Tpseudethanol_@WP_041589950_1
509193_000175295_2_Tbrockii_@WP_041589950_1
398726_000019065_1_Tsp_@WP_038016495_1
697303_000147695_3_Twiegelii_@AEM76705_1
580331_000025645_1_Titalicus_@WP_012995188_1
583358_000092965_1_Tmathranii_@WP_013150295_1
2325_000763575_1_Tkivui_@WP_049685179_1
273068_000007085_1_Csubterraneus_@WP_011025632_1
698948_000328545_1_Tthermosacchа_@WP_015311569_1
858215_000189775_3_Txylanolyticu_@AEF17254_1
1094508_000307585_1_Tsaccharolyti_@WP_014758751_1
0.3

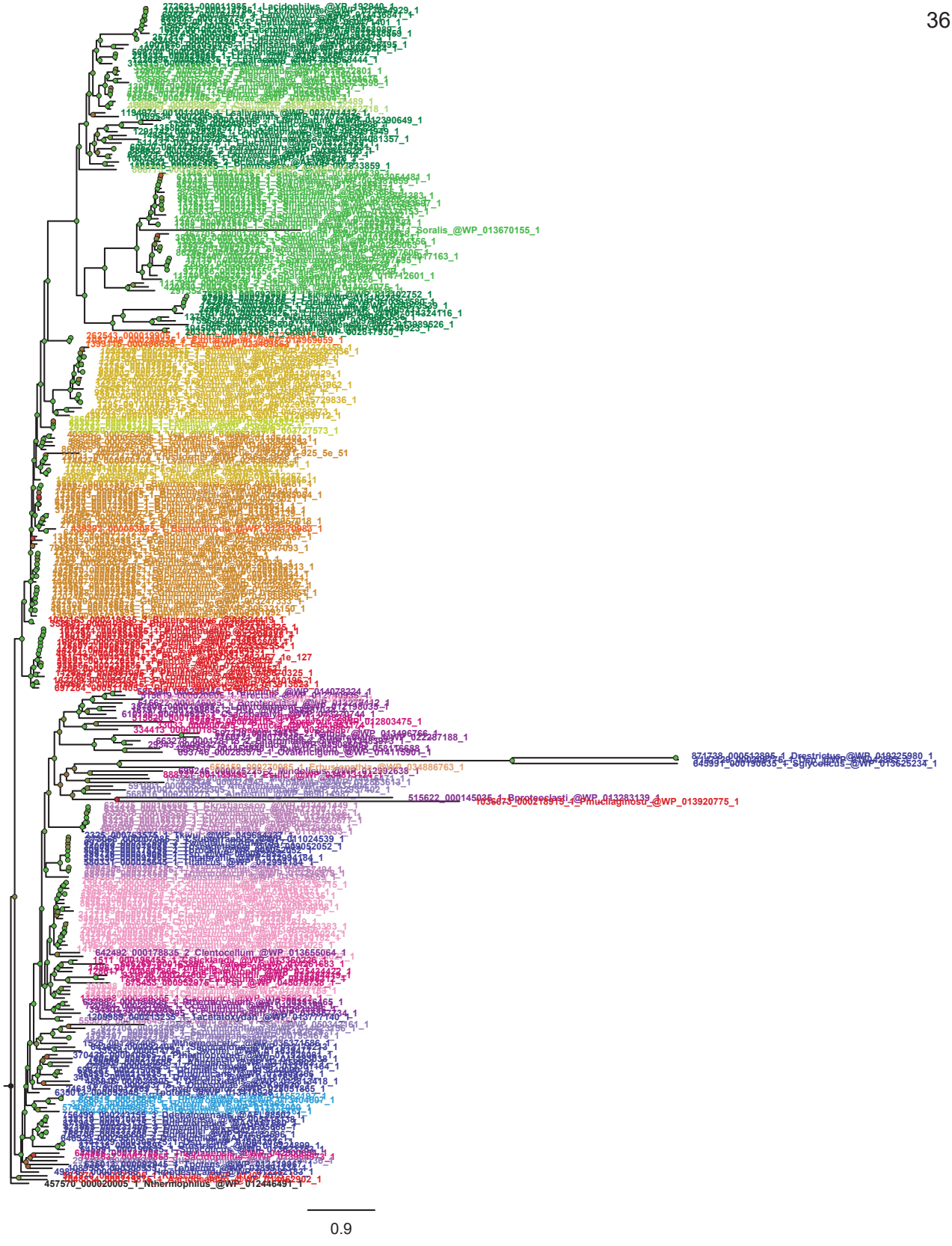

1002809_000271325_1_Ssilvestris_@WP_014824392_1
1002809_000271325_1_Ssilvestris_@WP_014825002_1
1145276_000600105_1_Lvarians_@WP_025219826_1
135735_000972245_2_Bendophyticus_@WP_046217715_1
586416_000725365_1_Taidingensis_@WP_038560716_1
1042163_000219535_3_Blaterosporus_@AIG27068_1
358681_000010165_1_Bbrevis_@WP_015891384_1
649639_000177235_2_Bcellulosilyt_@WP_013488945_1
221109_000011245_1_Oiheyensis_@WP_011065768_1
1398_001039495_1_Bcoagulans_@WP_014098223_1
1648923_000876525_1_Bparalichenif_@WP_023857517_1
279010_000008425_1_Blicheniformi_@WP_011198028_1
1408_000972685_1_Bpumilus_@WP_008346599_1
1126211_000262385_1_Bamyloliquefa_@WP_014418074_1
492670_000988345_1_Bvelezensis_@WP_046702313_1
1452_000830075_1_Batrophaeus_@WP_003326025_1
224308_000009045_1_Bsubtilis_@NP_389717_1
1574141_000877815_1_Bsp_@WP_041849899_1
1367477_000473245_1_Binfantis_@WP_022543846_1
403957_000725285_1_Vsp_@WP_052350149_1
221109_000011245_1_Oiheyensis_@WP_011067542_1
66692_000009825_1_Bclausii_@WP_011247537_1
200991_001465795_1_Prifietoensis_@WP_058382809_1
1526927_000785555_1_Psp_@WP_038704866_1
0.2

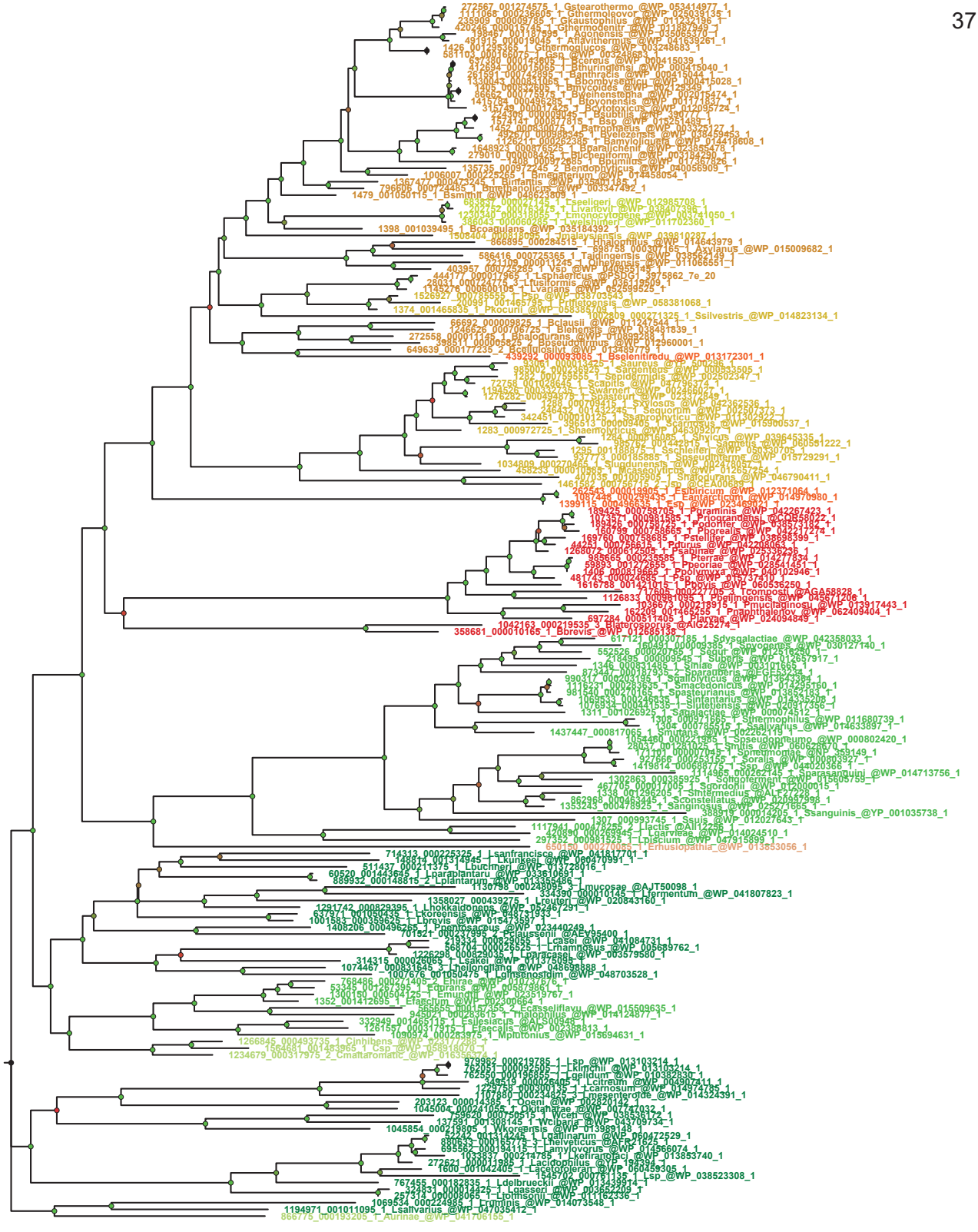

0.3

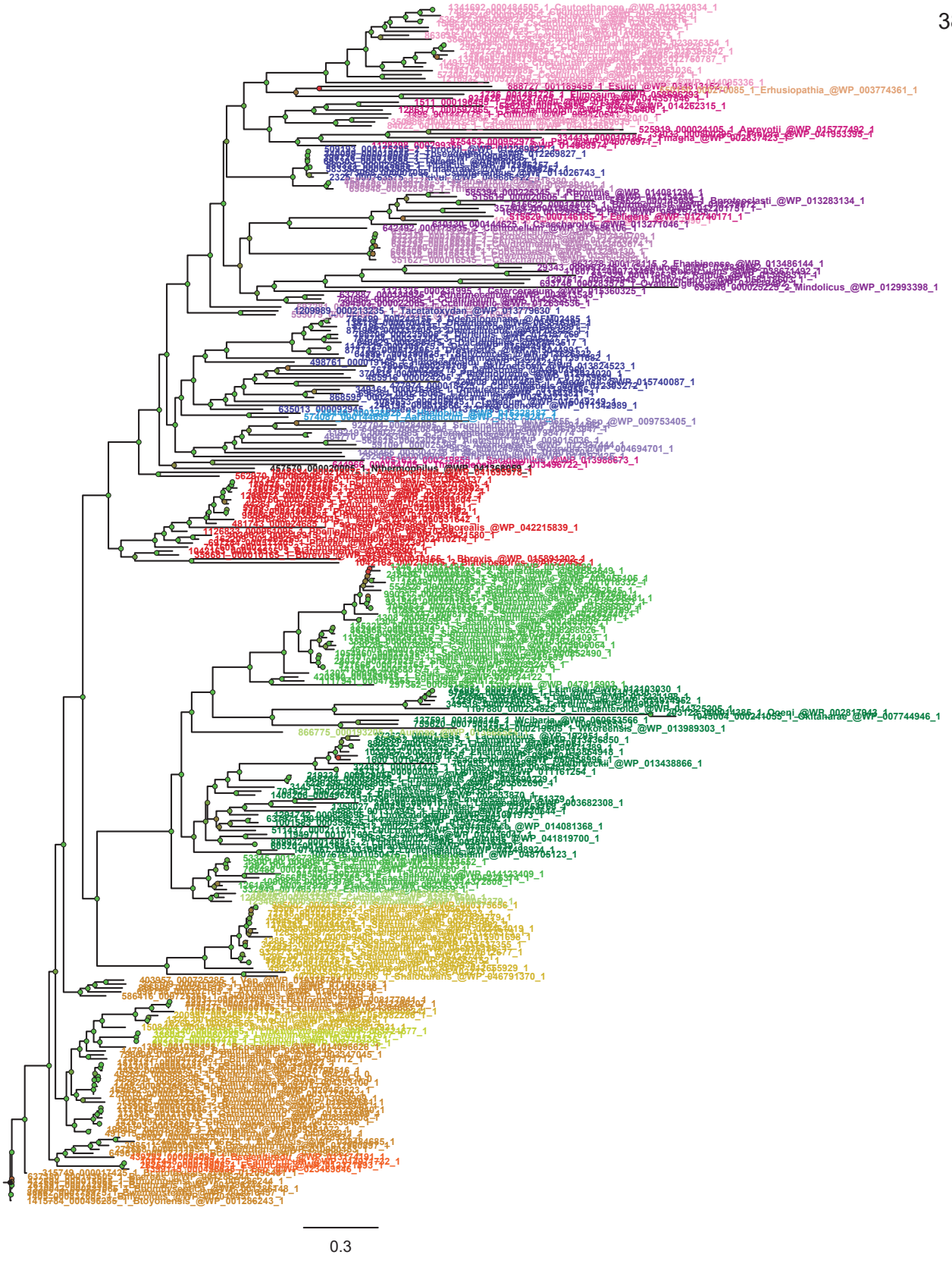

0.3

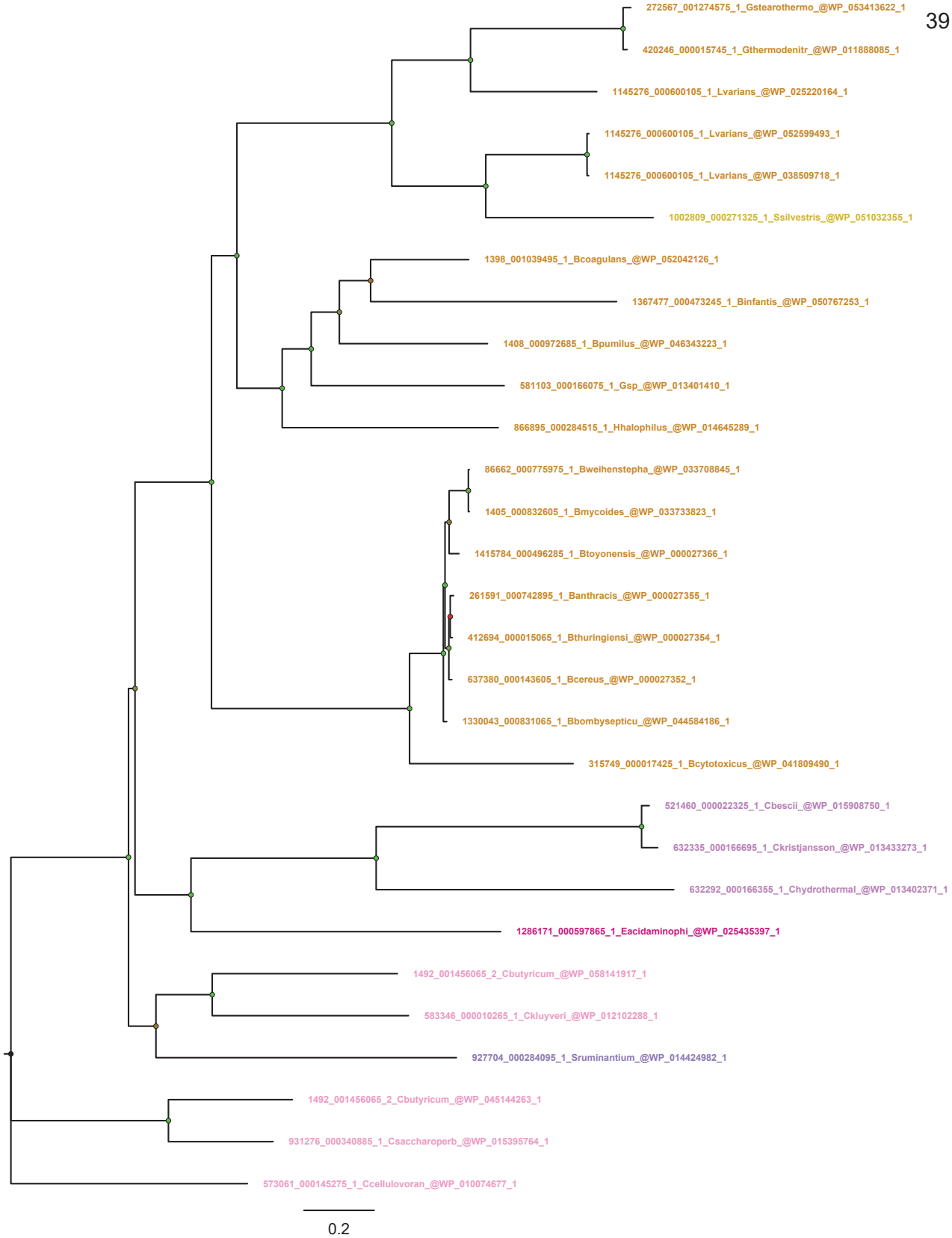

457570_000020005_1_Nthermophilus_@WP_012446564_1
429009_000024605_1_Addegensii_@WP_015740018_1
370438_000010565_1_Pthermopropio_@WP_011928268_1
0.9

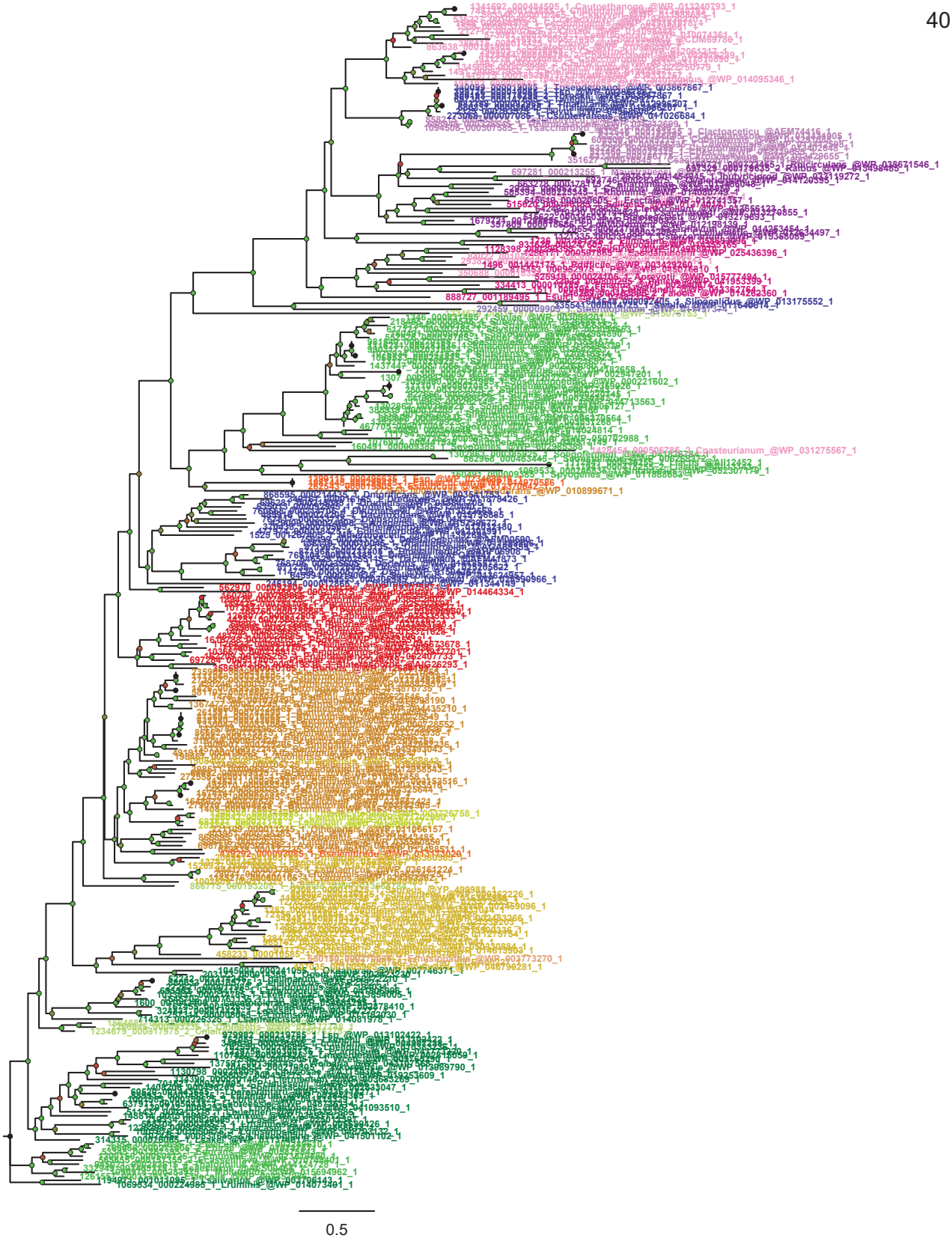

0.4

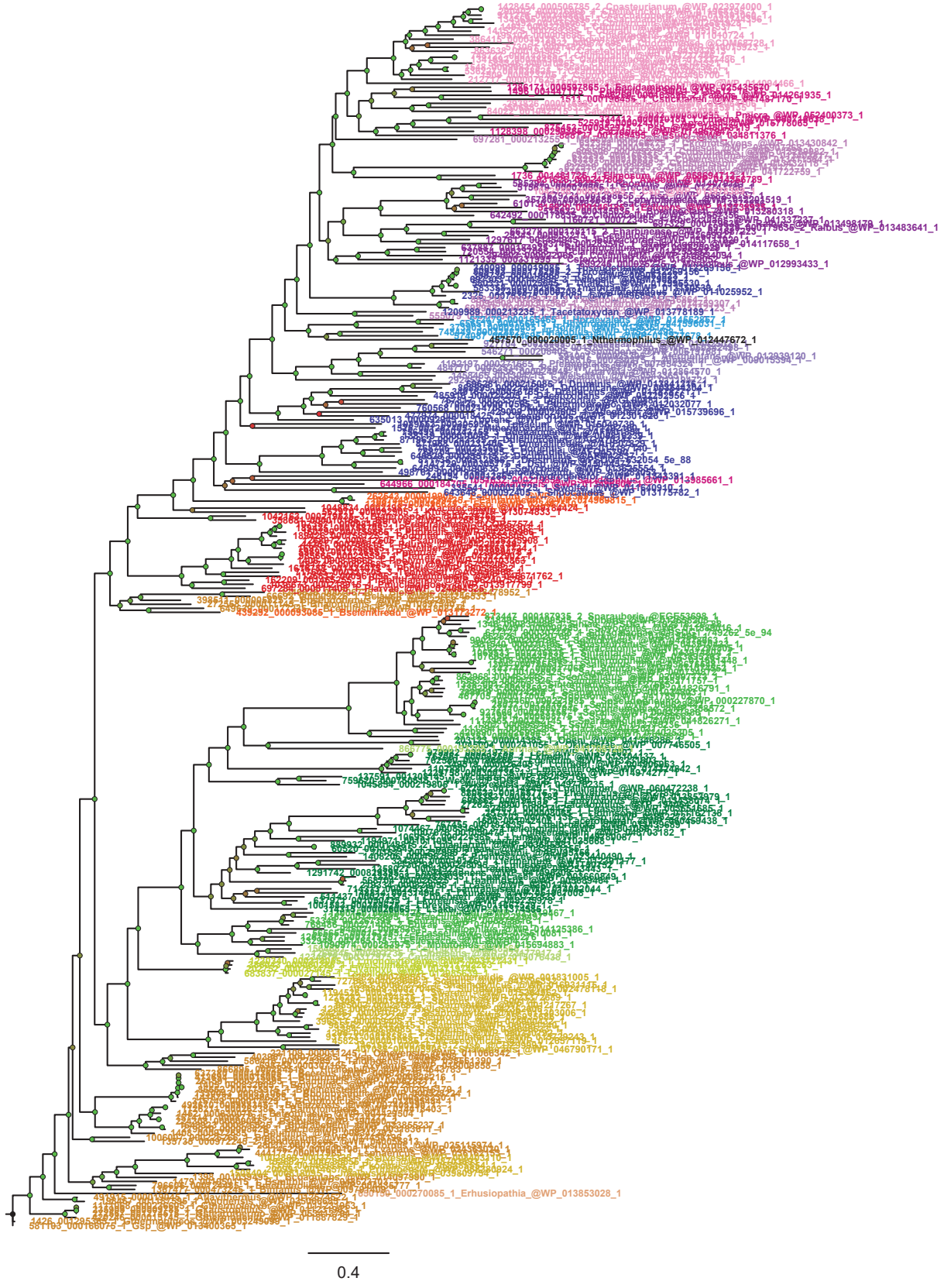

0.9

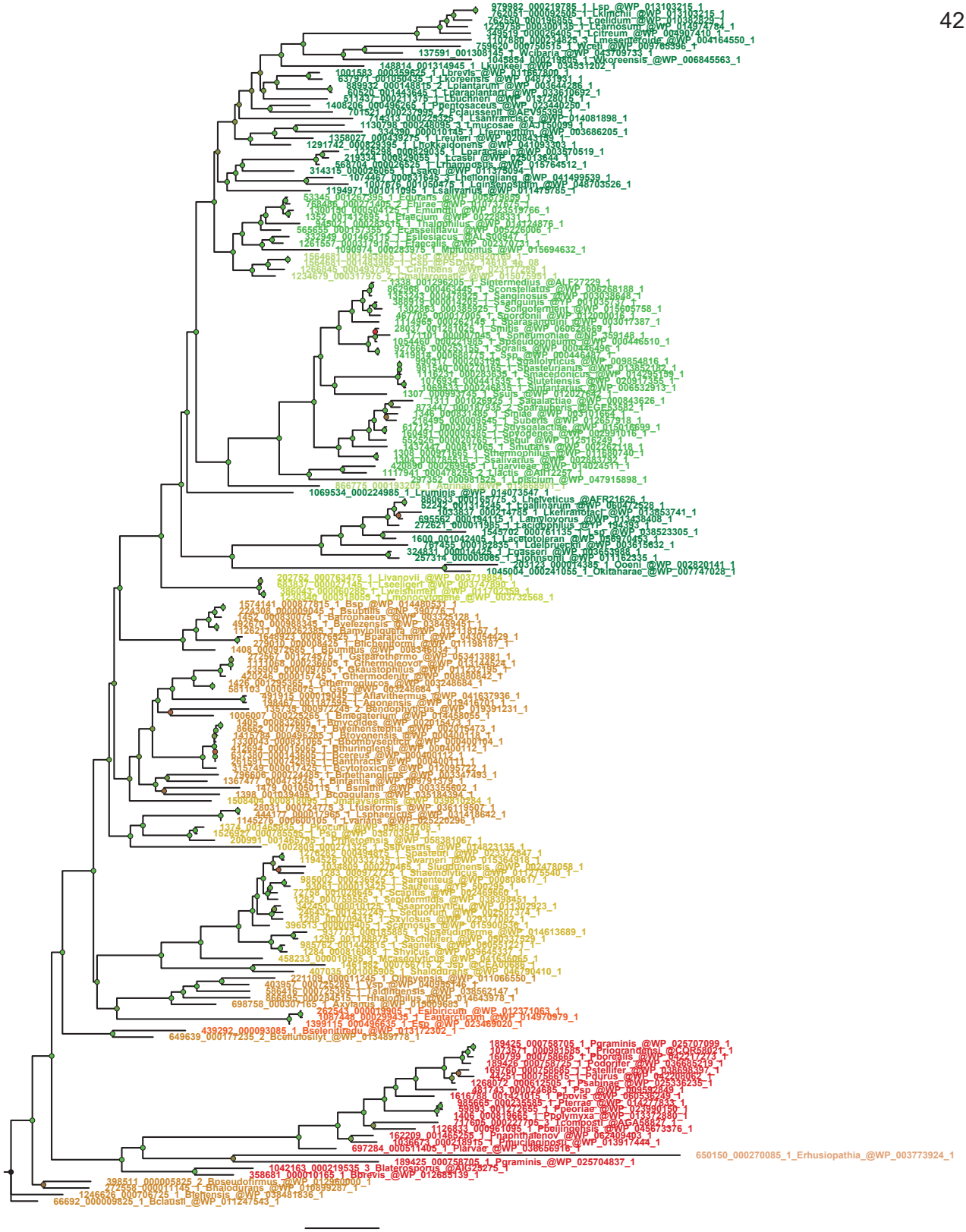

0.2

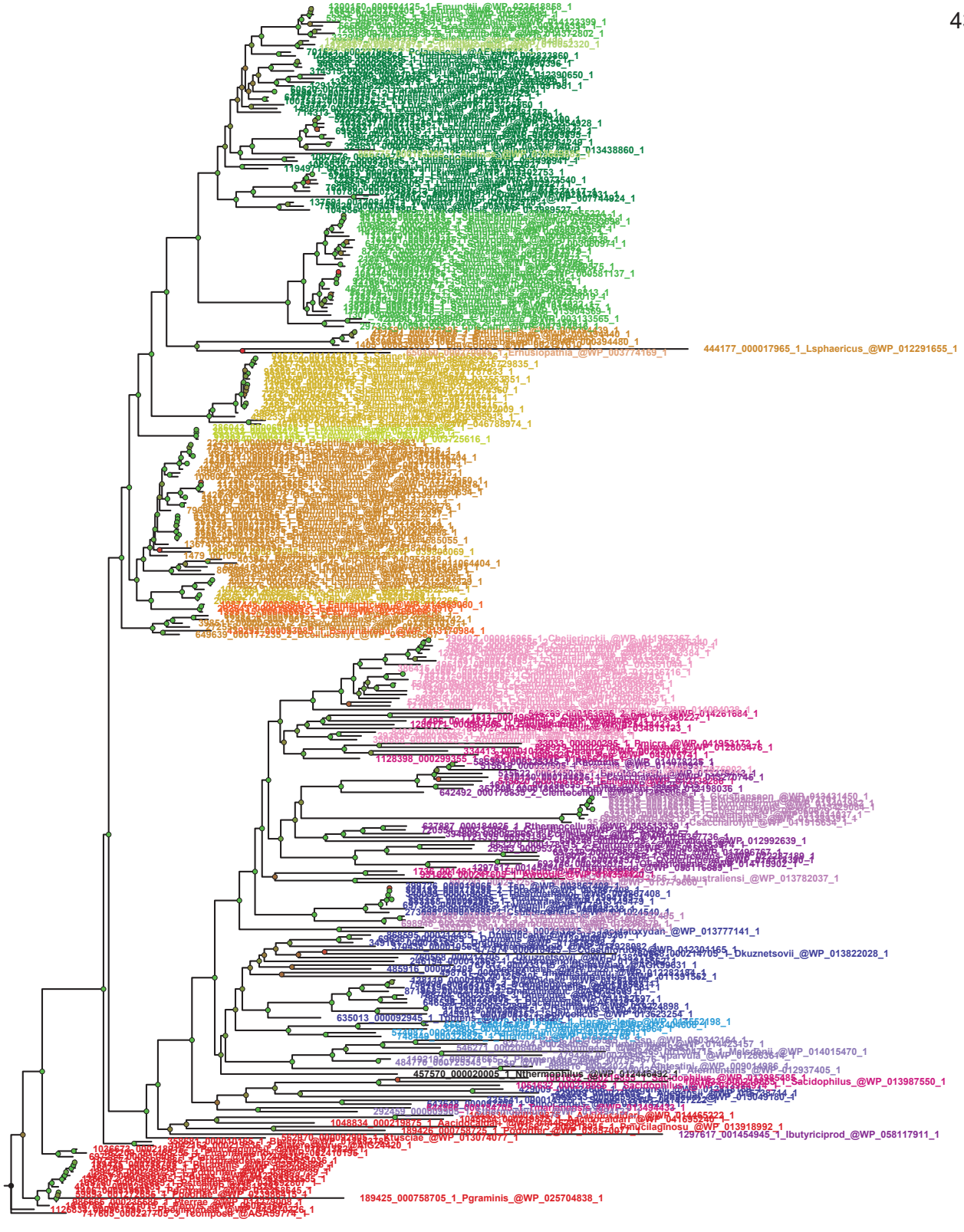

0.3

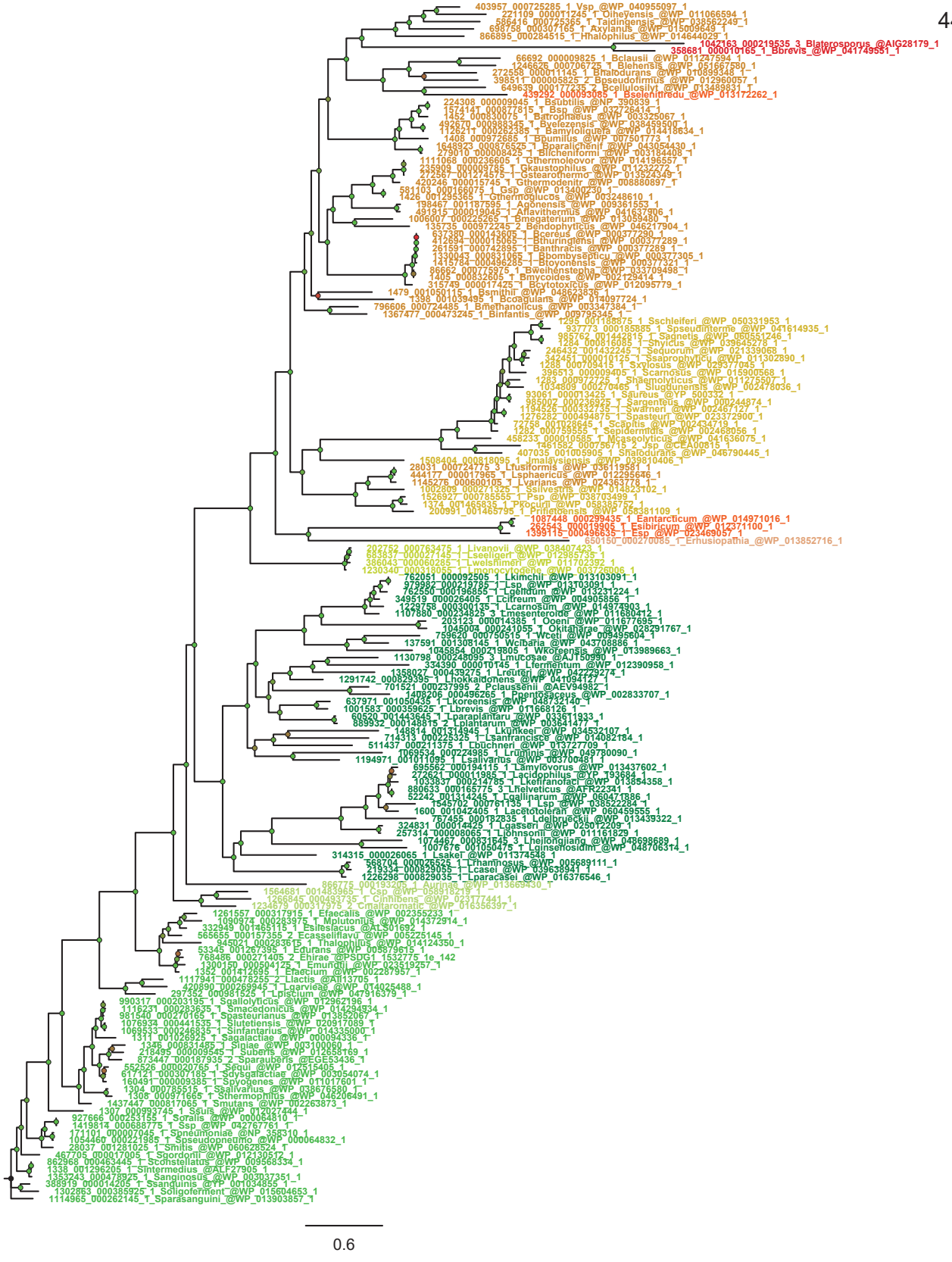

272567_001274575_1_Gstearothermo_@WP_053413622_1
420246_000015745_1_Gthermodenitr_@WP_011888085_1
1145276_000600105_1_Lvarians_@WP_025220164_1
1145276_000600105_1_Lvarians_@WP_052599493_1
1145276_000600105_1_Lvarians_@WP_038509718_1
1002809_000271325_1_Ssilvestris_@WP_051032355_1
1398_001039495_1_Bcoagulans_@WP_052042126_1
1367477_000473245_1_Binfantis_@WP_050767253_1
1408_000972685_1_Bpumilus_@WP_046343223_1
581103_000166075_1_Gsp_@WP_013401410_1
866895_000284515_1_Hhalophilus_@WP_014645289_1
86662_000775975_1_Bweihenstepha_@WP_033708845_1
1405_000832605_1_Bmycoides_@WP_033733823_1
1415784_000496285_1_Btoyonensis_@WP_000027366_1
261591_000742895_1_Banthracis_@WP_000027355_1
412694_000015065_1_Bthuringiensi_@WP_000027354_1
637380_000143605_1_Bcereus_@WP_000027352_1
1330043_000831065_1_Bbombysepticu_@WP_044584186_1
315749_000017425_1_Bcytotoxicus_@WP_041809490_1
521460_000022325_1_Cbescii_@WP_015908750_1
632335_000166695_1_Ckristjansson_@WP_013433273_1
632292_000166355_1_Chydrothermal_@WP_013402371_1
1286171_000597865_1_Eacidaminophi_@WP_025435397_1
1492_001456065_2_Cbutyricum_@WP_058141917_1
583346_000010265_1_Ckluyveri_@WP_012102288_1
927704_000284095_1_Sruminantium_@WP_014424982_1
1492_001456065_2_Cbutyricum_@WP_045144263_1
931276_000340885_1_Csaccharoperb_@WP_015395764_1
573061_000145275_1_Ccellulovoran_@WP_010074677_1
0.2

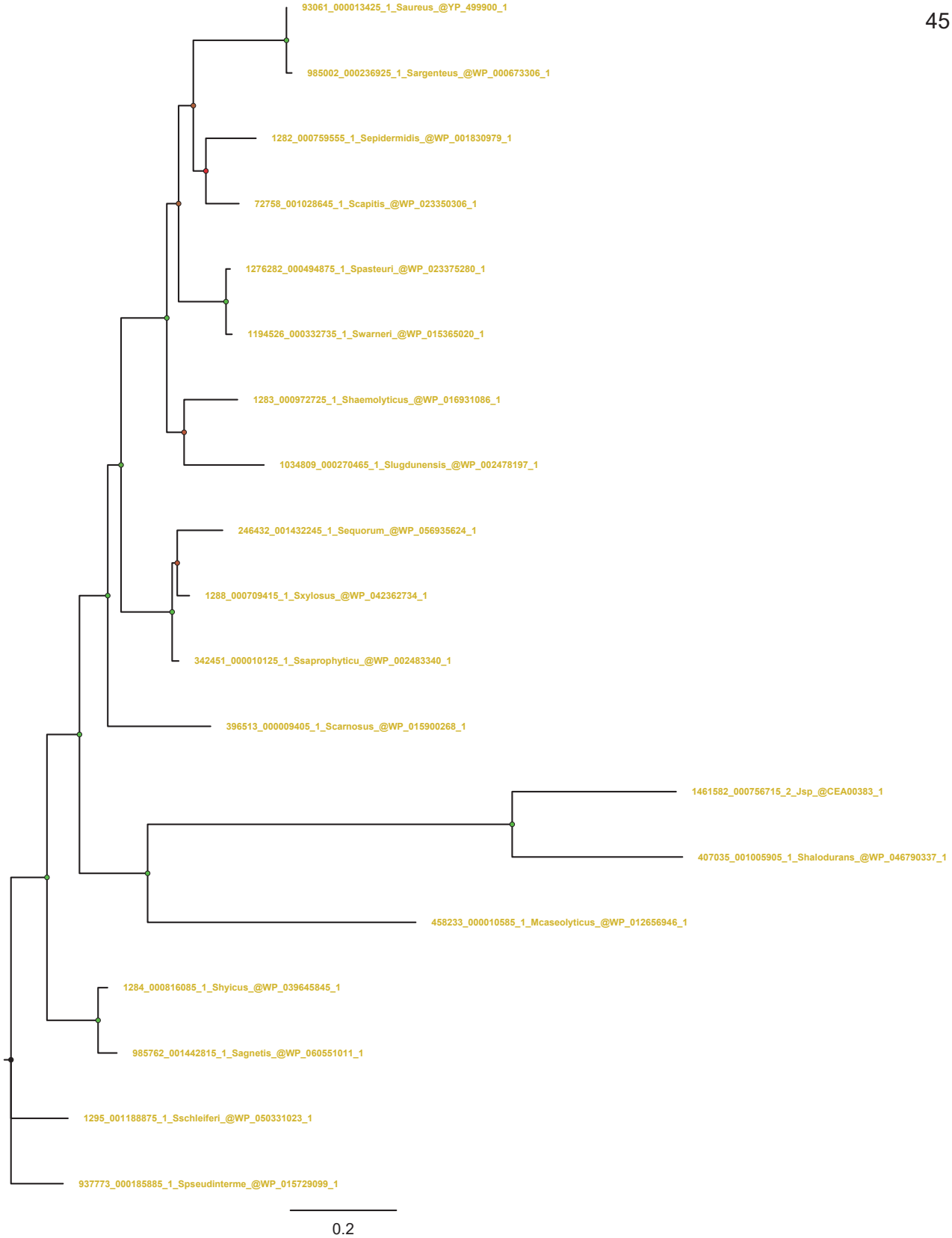

0.5

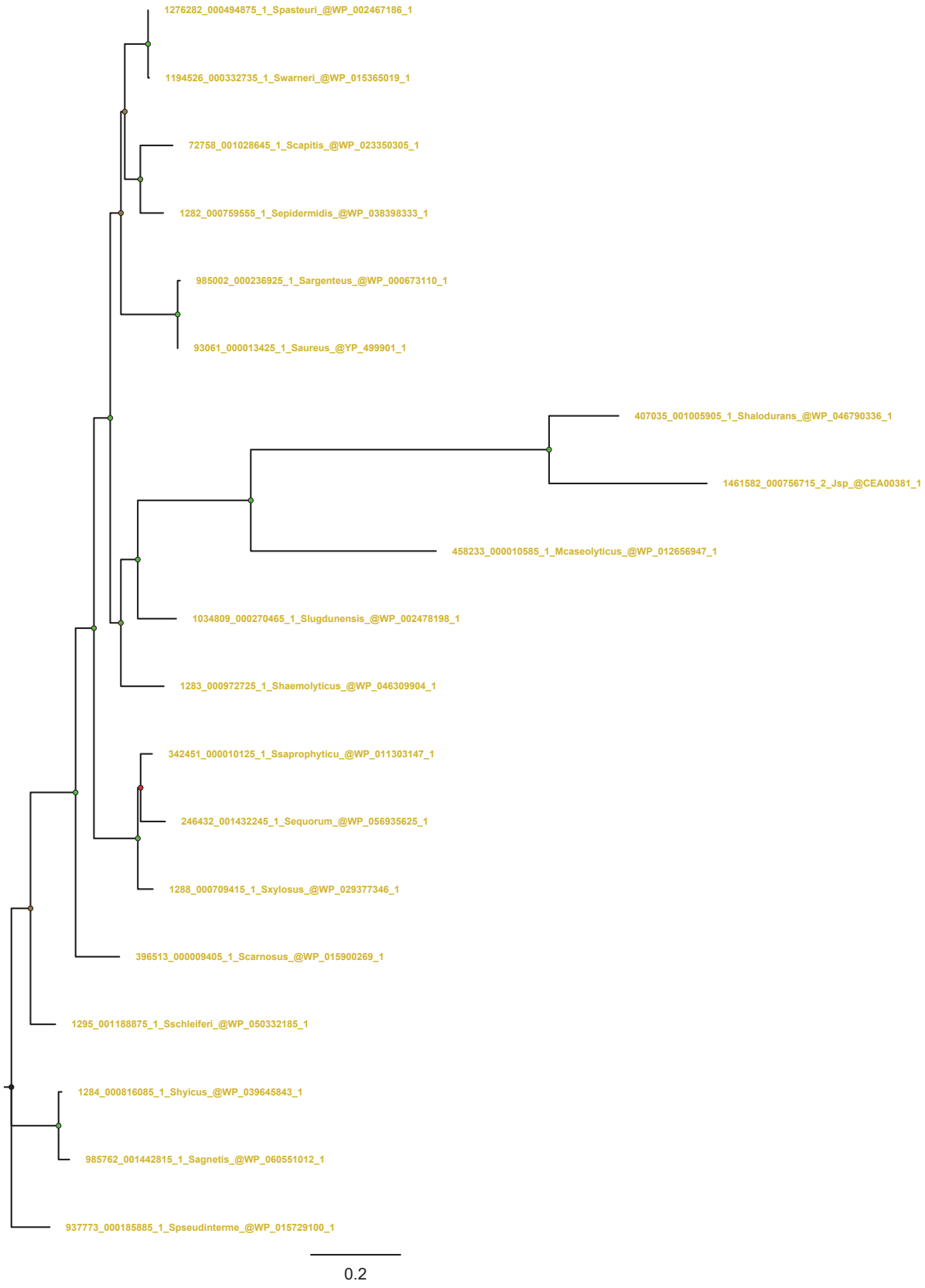

0.4

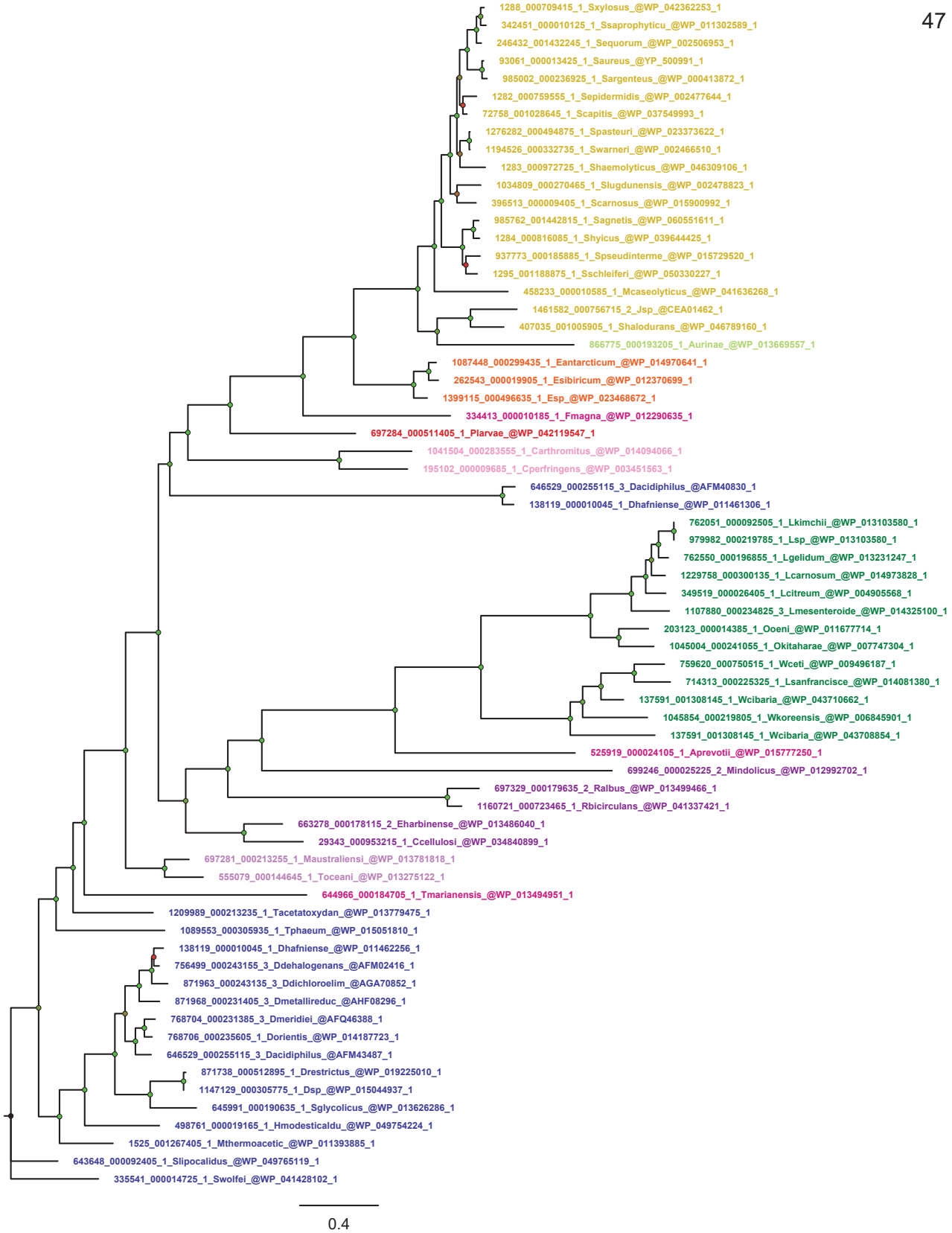

650150_000270085_1_Erhusiopathia_@WP_003773924_1
0.4

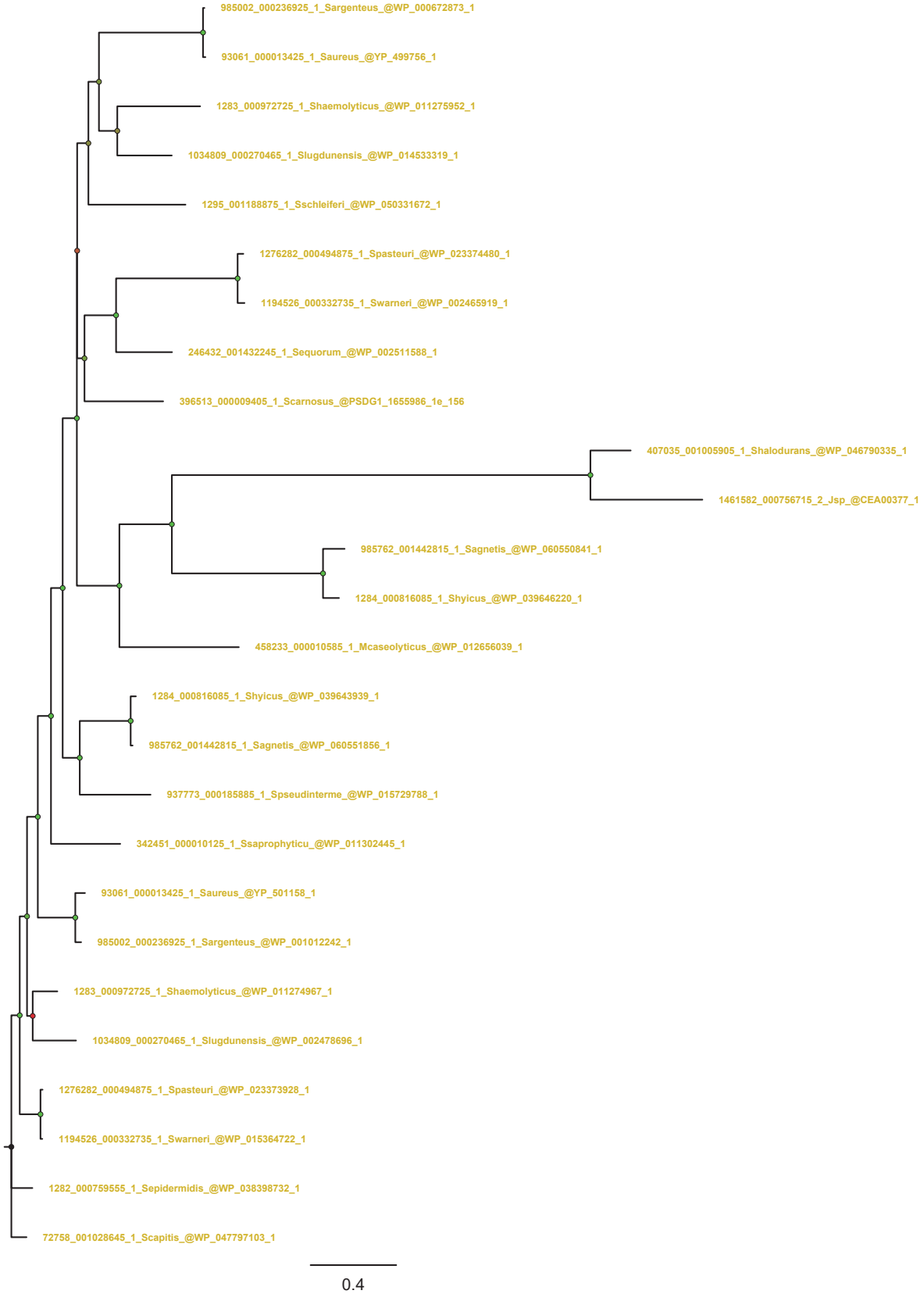

444177_000017965_1_Lsphaericus_@WP_012291655_1
189425_000758705_1_Pgraminis_@WP_025704838_1
457570_000020005_1_Nthermophilus_@WP_011448432_1
635013_000092945_1_Tmaritima_@WP_013623254_1
0.4

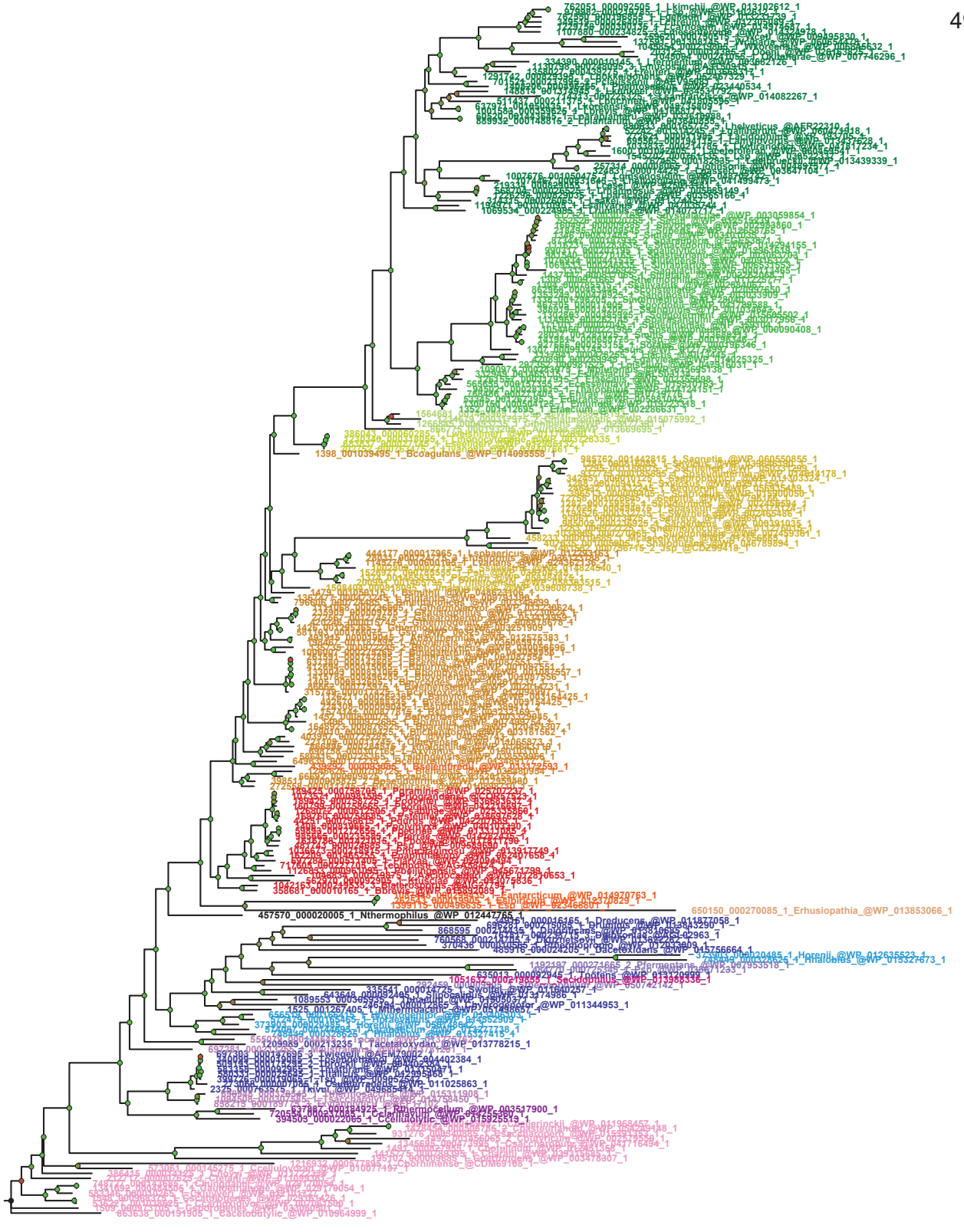

403957_000725285_1_Vsp_@WP_040955097_1
221109_000011245_1_Olheyensis_@WP_011066594_1
586416_000725265_1_Taidingensis_@WP_038562249_1
698758_000307165_1_Axylanus_@WP_015009649_1
866895_000284515_1_Hhalophilus_@WP_014644029_1
1042163_000219535_3_Blateroporus_@AIG28179_1
358681_000010165_1_Bbrevis_@WP_041749551_1
66692_000009825_1_Bclausii_@WP_011247594_1
124626_000706725_1_Bhemicellulosilyt_@WP_051667590_1
272558_000011145_1_Bhalodurans_@WP_010899348_1
398511_000005825_2_Bpseudofirmus_@WP_012960057_1
649679_000177255_2_Bcellulosilyt_@WP_014499931_1
439292_000093085_1_Bselenitiredu_@WP_013172262_1
224308_000009045_1_Bsubtilis_@NP_390839_1
1374147_000877815_1_Bsp_@WP_027728414_1
1452_000830075_1_Batrophaeus_@WP_003325067_1
492670_000988345_1_Bvelezensis_@WP_038459500_1
112621_000262385_1_Bamyloliquefa_@WP_014418634_1
1408_000972685_1_Bpumilus_@WP_007501773_1
1648923_000876525_1_Bparalicheni_@WP_043054430_1
279010_000008425_1_Blicheniformi_@WP_003184408_1
1111068_000236605_1_Gthermoleovor_@WP_014196557_1
235909_000009785_1_Gkaustophilus_@WP_011232272_1
272567_001274575_1_Gstearothermo_@WP_013924349_1
420246_000015745_1_Gthermodenitr_@WP_008880897_1
581103_000166075_1_Gsp_@WP_013400230_1
1426_001295365_1_Gthermoglucos_@WP_003248610_1
199467_001187595_1_Agonensis_@WP_009361553_1
491915_000019045_1_Aflavithermus_@WP_041637806_1
1006007_000225265_1_Bmegaterium_@WP_013059480_1
135735_000972245_2_Bendophyticus_@WP_046217904_1
637380_000143605_1_Bcereus_@WP_000377290_1
412694_000015065_1_Bthuringiensi_@WP_000377289_1
261591_000742895_1_Banthracis_@WP_000377289_1
1330043_000831065_1_Bbombysepticu_@WP_000377305_1
1415784_000496285_1_Btoyonensis_@WP_000377321_1
86662_000775975_1_Bweihenstepha_@WP_033709498_1
1405_000832605_1_Bmycoides_@WP_002129414_1
315749_000017425_1_Bcytotoxicus_@WP_012095779_1
1479_001050115_1_Bsmithii_@WP_048623836_1
1398_001039495_1_Bcoagulans_@WP_014097724_1
796606_000724485_1_Bmethanolicus_@WP_003347384_1
1367477_000473245_1_Binfantis_@WP_009795345_1
1295_001108575_1_Sschleiferi_@WP_050331953_1
937773_000185885_1_Spseudintermе_@WP_041614935_1
985762_001442815_1_Sagnetis_@WP_060551246_1
1384_000816085_1_Shyicus_@WP_039642716_1
246432_001432245_1_Sequorum_@WP_021339068_1
342451_000010125_1_Ssaprophyticu_@WP_011302890_1
1248_000709415_1_Sxylosus_@WP_029377945_1
396513_000009405_1_Scarnosus_@WP_015900568_1
1283_000972725_1_Shaemolyticus_@WP_011275507_1
1034809_000270465_1_Slugdunensis_@WP_002478036_1
93061_000013425_1_Saureus_@YP_500352_1
985007_000236925_1_Sargenteus_@WP_000244874_1
1194526_000332735_1_Swarneri_@WP_002467127_1
1276282_000494875_1_Spasteurl_@WP_023372900_1
72758_001028645_1_Scapitis_@WP_002434718_1
1282_000759555_1_Sepidermidis_@WP_002468056_1
458233_000010585_1_Mgasseolyticus_@WP_041636075_1
1461582_000756715_2_Jsp_@CEA00515_1
407035_001005905_1_Shalodurans_@WP_046790445_1
1508404_000818095_1_Jmalaysiensis_@WP_039810406_1
28031_000724775_3_Lfusiformis_@WP_036115981_1
444177_000017965_1_Lsphaericus_@WP_012295646_1
1145278_000600105_1_Lvarians_@WP_024363778_1
1002809_000271325_1_Ssilvestris_@WP_014823102_1
1526927_000785555_1_Psp_@WP_038703499_1
1374_001465935_1_Pkocurii_@WP_058385752_1
200991_001465795_1_Pmicroensis_@WP_058381109_1
1087448_000299435_1_Eantarcticum_@WP_014971016_1
262543_000019905_1_Esibiricum_@WP_012371100_1
1399115_000498635_1_Esp_@WP_023469057_1
650150_000270085_1_Ehusipathia_@WP_013852716_1
202752_000763475_1_Livanovii_@WP_038407423_1
685634_000027145_1_Lseeliger_@WP_012985715_1
386043_000060285_1_Lweihenstephi_@WP_011702592_1
1230340_000318055_1_Lmonocytogenes_@WP_003726006_1
762051_000092505_1_Lkimchii_@WP_013103091_1
979982_000219795_1_Lsp_@WP_013103091_1
762550_000196855_1_Lgendtii_@WP_013231224_1
349519_000026405_1_Lcitreum_@WP_004905856_1
1229758_000300135_1_Lcamosum_@WP_014473903_1
1107880_000234825_3_Lmesenteroide_@WP_011680412_1
203123_000014385_1_Ogeni_@WP_011677695_1
1045004_000241055_1_Okitaharae_@WP_028291767_1
759620_000750515_1_Wceti_@WP_009495604_1
137591_001308145_1_Wcibaria_@WP_043708886_1
1045854_000219805_1_Wkoreensis_@WP_013989663_1
1130798_000248995_3_Lmucosae_@AJT50990_1
334390_000010145_1_Lfermentum_@WP_012390958_1
1358027_000439275_1_Lreuteri_@WP_042229274_1
1291742_000829395_1_Lhokkaidonens_@WP_041094127_1
701521_000237995_2_Pclaussenii_@AEV94982_1
1408206_000496265_1_Pnentosaceus_@WP_002833707_1
637971_001050435_1_Lkoreensis_@WP_048732140_1
1001583_000359625_1_Lbrevis_@WP_011668126_1
60520_001443645_1_Lparaplantaru_@WP_033671933_1
889352_000148815_2_Lplantarum_@WP_003641477_1
148814_001314945_1_Lkunkeei_@WP_034532107_1
714313_000225325_1_Lsanfrancisce_@WP_014082184_1
511437_000211375_1_Lbuchneri_@WP_013727709_1
1069534_000224985_1_Lruminis_@WP_049730090_1
1194971_001011095_1_Lsalivarius_@WP_003700481_1
695562_000194115_1_Lamylovorus_@WP_013437602_1
272621_000011985_1_Lacidophilus_@YP_193681_1
1033837_000214785_1_Lkefiranofaci_@WP_013854358_1
880633_000165775_3_Lhelveticus_@AFR22741_1
52242_001314245_1_Lgallinarum_@WP_060413586_1
1545702_000761135_1_Lsp_@WP_038522284_1
1600_001042405_1_Lacetotoleran_@WP_060459555_1
767455_000182835_1_Ldelbrueckii_@WP_013439322_1
324831_000014425_1_Lgasseri_@WP_025012209_1
257314_000008065_1_Ljohnsonii_@WP_011161829_1
1074467_000315545_3_Lhelsingborg_@WP_049698689_1
1007676_001050475_1_Lgroningosidm_@WP_048706314_1
314315_000026065_1_Lsakei_@WP_011374548_1
568704_000025525_1_Lfarraginis_@WP_005689111_1
219334_000829055_1_Lcasei_@WP_013963841_1
1226298_000829035_1_Lparacasei_@WP_016376546_1
886775_000193205_1_Aurinae_@WP_013669430_1
685634_000027145_1_Lsp_@WP_058918219_1
1266845_000493735_1_Cjhildens_@WP_023777441_1
1224878_000017375_1_Cmaltaromati_@WP_016356397_1
1261557_000317915_1_Efaecalis_@WP_002355233_1
1090974_000283975_1_Mplutonius_@WP_014372914_1
332949_001465115_1_Esiliaclacus_@ALS01692_1
565655_000157355_2_Ecasseliflavu_@WP_005225145_1
945021_000283615_1_Thalophilus_@WP_014124350_1
53345_001267395_1_Edurans_@WP_005879615_1
768486_000271405_2_Ehirae_@PSDS1_1532775_1e_142
1300150_000504125_1_Emundtii_@WP_023519257_1
1352_001412695_1_Efaecium_@WP_002287957_1
1117941_000478255_2_Llactis_@AII13705_1
420890_000269945_1_Lgarviae_@WP_014025488_1
297352_000981525_1_Lplantaru_@WP_047916379_1
990317_000203195_1_Sgallolyticus_@WP_012962196_1
1116231_000283635_1_Smacedonicus_@WP_014294334_1
981540_000270165_1_Spasteurianus_@WP_013852067_1
1076934_000441535_1_Slutetiensis_@WP_020917089_1
1069533_000246835_1_Sinfantarius_@WP_014335000_1
1311_001026925_1_Sagalactiae_@WP_000094336_1
1346_000831485_1_Siniae_@WP_003100060_1
218495_000009545_1_Suberis_@WP_012658169_1
873447_000187935_2_Sparauberis_@EGE53436_1
552826_000020765_1_Sequi_@WP_012515405_1
647121_000307195_1_Sdysgalactiae_@WP_003054074_1
160491_000009385_1_Spyogenes_@WP_011017601_1
1304_000785515_1_Ssalivarius_@WP_038676580_1
1308_000971685_1_Sthermophilus_@WP_046206491_1
14374447_000817065_1_Smutans_@WP_002263873_1
1307_000993745_1_Ssuis_@WP_012027444_1
927666_000253755_1_Sgrails_@WP_000064810_1
1419814_000688775_1_Ssp_@WP_042787761_1
171101_000007045_1_Spneumoniae_@NP_358310_1
1054460_000221985_1_Spseudopneumo_@WP_000064832_1
28037_001281025_1_Smitis_@WP_060628524_1
467705_000917005_1_Sgordonii_@WP_012130512_1
862968_000463445_1_Sconstellatus_@WP_009568334_1
1338_001296205_1_Sintermedius_@ALE27905_1
1353243_000478925_1_Sanginosus_@WP_003037351_1
388919_000014205_1_Ssanguinis_@YP_001034855_1
1302853_000385925_1_Sgilliervmint_@WP_015604653_1
1114965_000262145_1_Sparasanguini_@WP_013903857_1
0.6

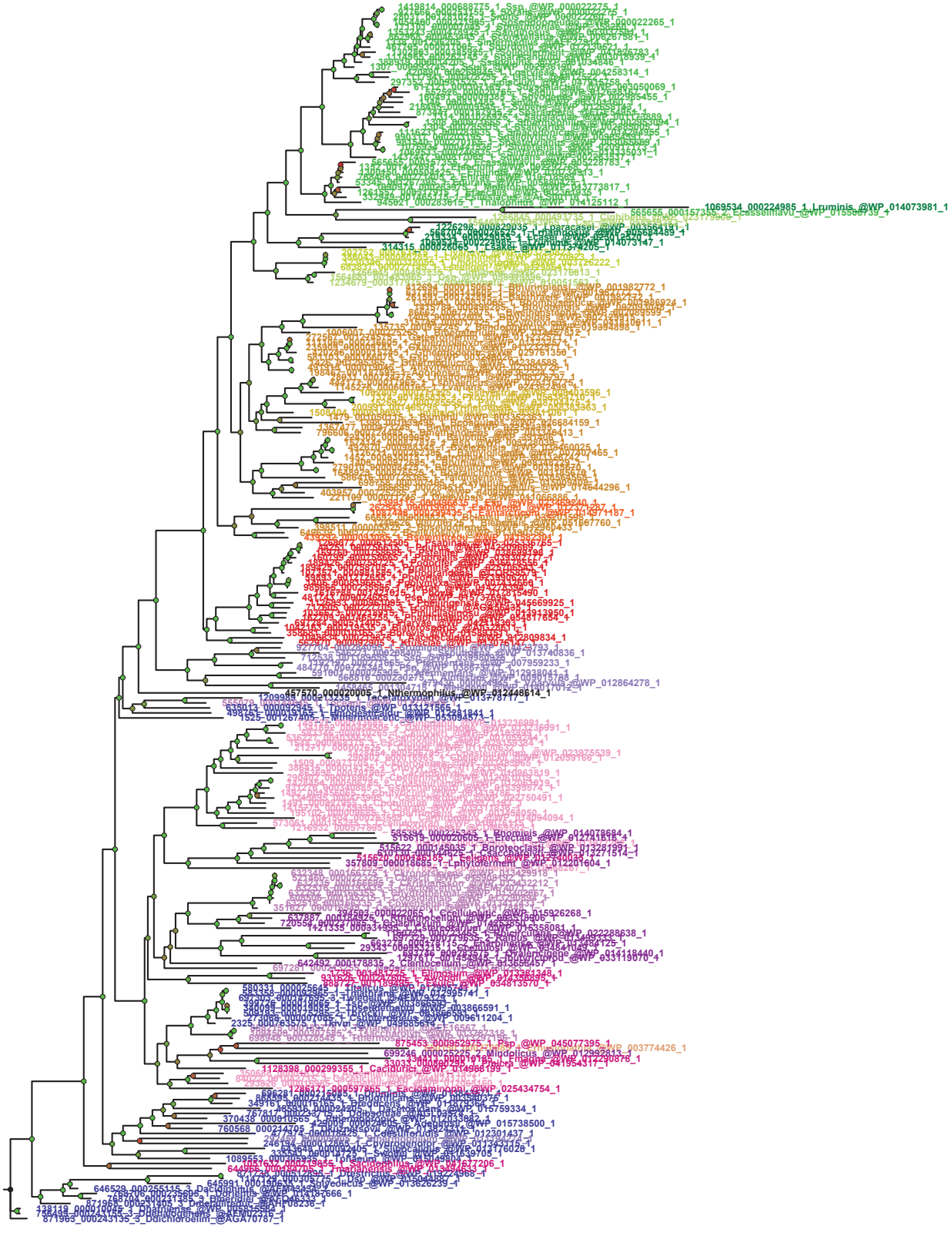

93061_000013425_1_Saureus_@YP_499900_1
985002_000236925_1_Sargenteus_@WP_000673306_1
1282_000759555_1_Sepidermidis_@WP_001830979_1
72758_001028645_1_Scapitis_@WP_023350306_1
1276282_000494875_1_Spasteuri_@WP_023375280_1
1194526_000332735_1_Swarneri_@WP_015365020_1
1283_000972725_1_Shaemolyticus_@WP_016931086_1
1034809_000270465_1_Slugdunensis_@WP_002478197_1
246432_001432245_1_Sequorum_@WP_056935624_1
1288_000709415_1_Sxylosus_@WP_042362734_1
342451_000010125_1_Ssaprophyticu_@WP_002483340_1
396513_000009405_1_Scarnosus_@WP_015900268_1
1461582_000756715_2_Jsp_@CEA00383_1
407035_001005905_1_Shalodurans_@WP_046790337_1
458233_000010585_1_Mcaseolyticus_@WP_012656946_1
1284_000816085_1_Shyicus_@WP_039645845_1
985762_001442815_1_Sagnetis_@WP_060551011_1
1295_001188875_1_Sschleiferi_@WP_050331023_1
937773_000185885_1_Spseudinterme_@WP_015729099_1
0.2

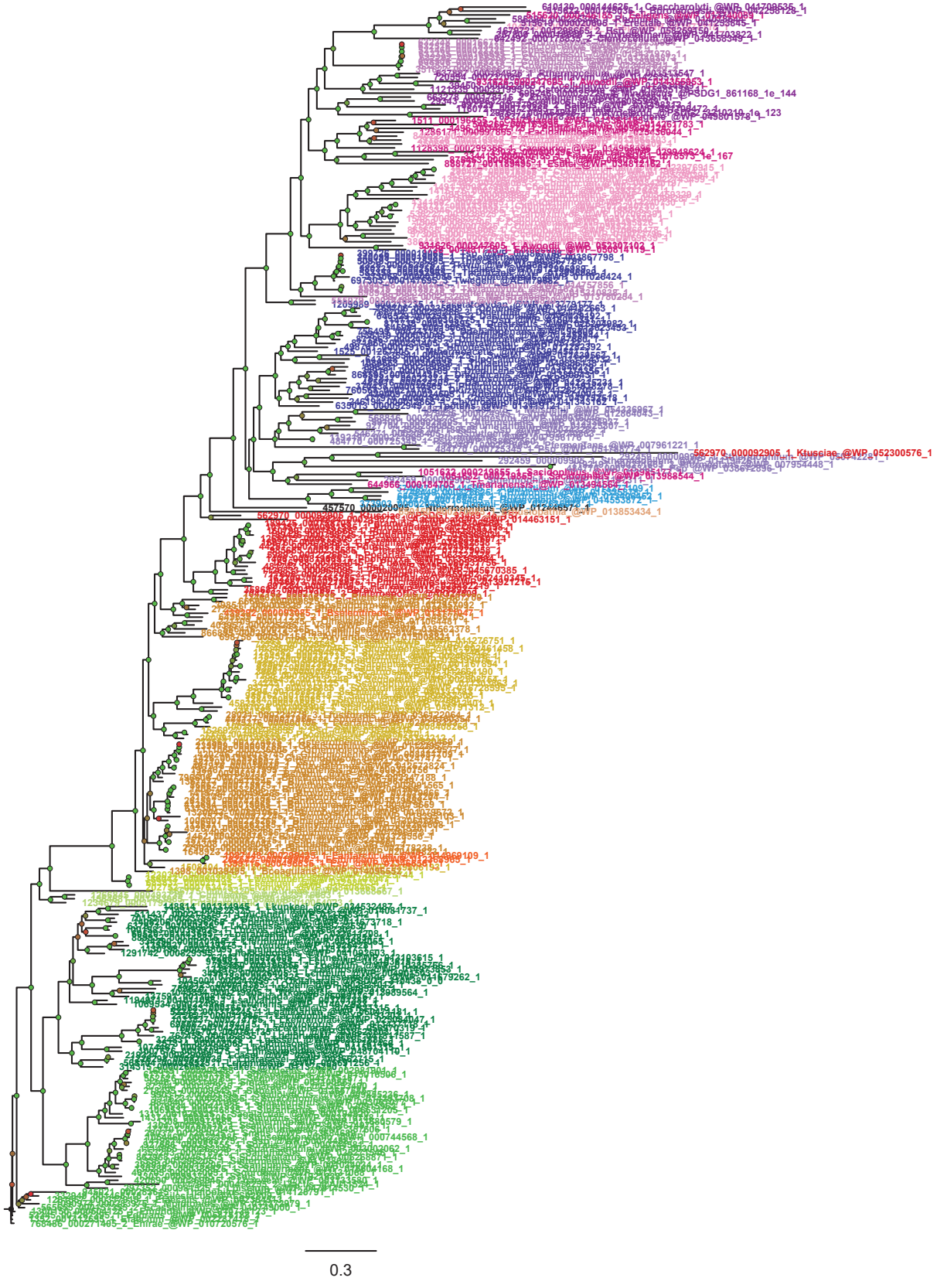

1276282_000494875_1_Spasteuri_@WP_002467186_1
1194526_000332735_1_Swarneri_@WP_015365019_1
72758_001028645_1_Scapitis_@WP_023350305_1
1282_000759555_1_Sepidermidis_@WP_038398333_1
985002_000236925_1_Sargenteus_@WP_000673110_1
93061_000013425_1_Saureus_@YP_499901_1
407035_001005905_1_Shalodurans_@WP_046790336_1
1461582_000756715_2_Jsp_@CEA00381_1
458233_000010585_1_Mcaseolyticus_@WP_012656947_1
1034809_000270465_1_Slugdunensis_@WP_002478198_1
1283_000972725_1_Shaemolyticus_@WP_046309904_1
342451_000010125_1_Ssaprophyticu_@WP_011303147_1
246432_001432245_1_Sequorum_@WP_056935625_1
1288_000709415_1_Sxylosus_@WP_029377346_1
396513_000009405_1_Scarnosus_@WP_015900269_1
1295_001188875_1_Sschleiferi_@WP_050332185_1
1284_000816085_1_Shyicus_@WP_039645843_1
985762_001442815_1_Sagnetis_@WP_060551012_1
937773_000185885_1_Spseudinterme_@WP_015729100_1
0.2

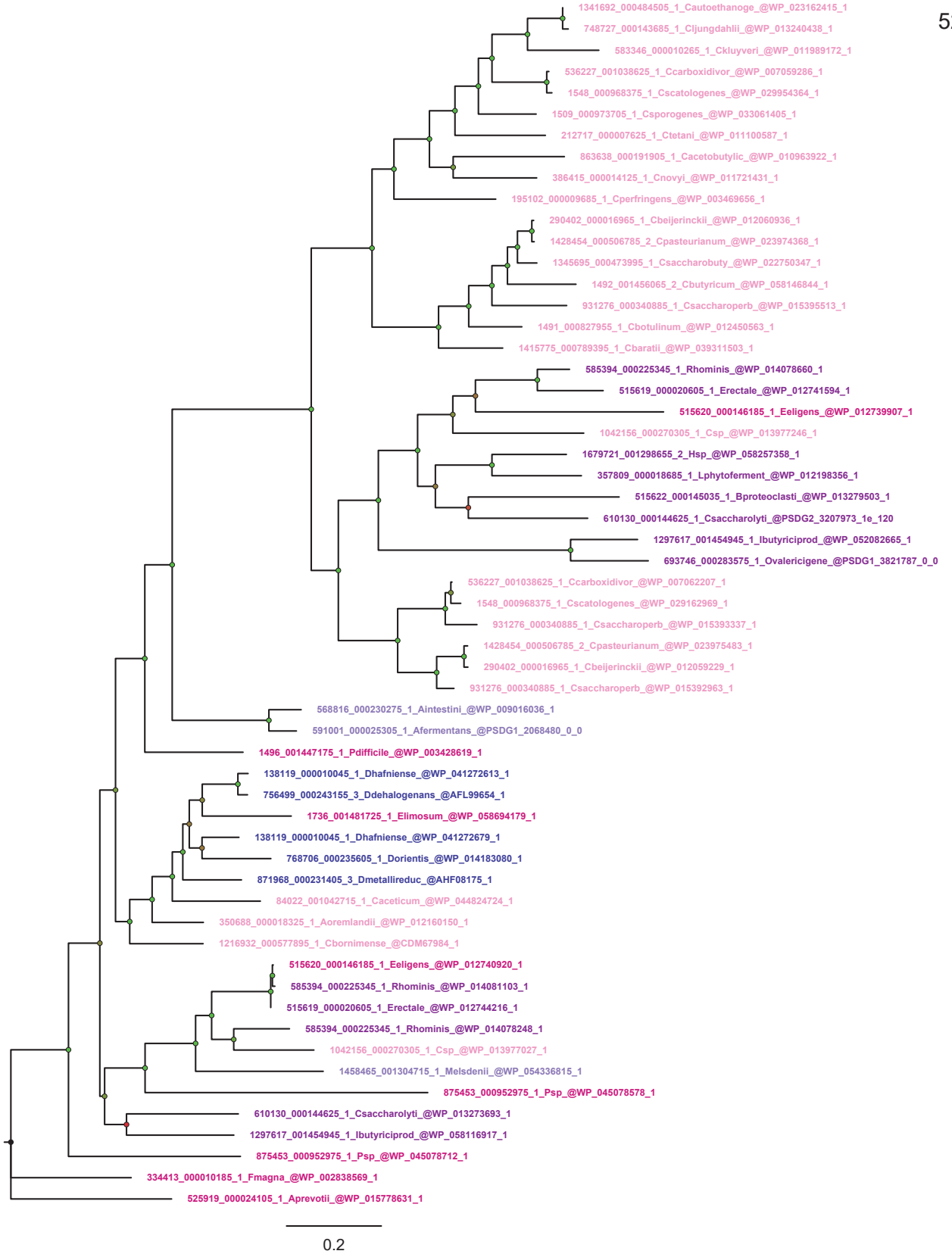

1288_000709415_1_Sxylosus_@WP_042362253_1
342451_000010125_1_Ssaprophyticu_@WP_011302589_1
246432_001432245_1_Sequorum_@WP_002506953_1
93061_000013425_1_Saureus_@YP_500991_1
985002_000236925_1_Sargenteus_@WP_000413872_1
1282_000759555_1_Sepidermidis_@WP_002477644_1
72758_001028645_1_Scapitis_@WP_037549993_1
1276282_000494875_1_Spasteuri_@WP_023373622_1
1194526_000332735_1_Swarneri_@WP_002466510_1
1283_000972725_1_Shaemolyticus_@WP_046309106_1
1034809_000270465_1_Slugdunensis_@WP_002478823_1
396513_000009405_1_Scarnosus_@WP_015900992_1
985762_001442815_1_Sagnetis_@WP_060551611_1
1284_000816085_1_Shyicus_@WP_039644425_1
937773_000185885_1_Spseudinterme_@WP_015729520_1
1295_001188875_1_Sschleiferi_@WP_050330227_1
458233_000010585_1_Mcaseolyticus_@WP_041636268_1
1461582_000756715_2_Jsp_@CEA01462_1
407035_001005905_1_Shalodurans_@WP_046789160_1
866775_000193205_1_Aurinae_@WP_013669557_1
1087448_000299435_1_Eantarcticum_@WP_014970641_1
262543_000019905_1_Esibiricum_@WP_012370699_1
1399115_000496635_1_Esp_@WP_023468672_1
334413_000010185_1_Fmagna_@WP_012290635_1
697284_000511405_1_Plarvae_@WP_042119547_1
1041504_000283555_1_Carthromitus_@WP_014094066_1
195102_000009685_1_Cperfringens_@WP_003451563_1
646529_000255115_3_Dacidiphilus_@AFM40830_1
138119_000010045_1_Dhafniense_@WP_011461306_1
762051_000092505_1_Lkimchii_@WP_013103580_1
979982_000219785_1_Lsp_@WP_013103580_1
762550_000196855_1_Lgelidum_@WP_013231247_1
1229758_000300135_1_Lcarnosum_@WP_014973828_1
349519_000026405_1_Lcitreum_@WP_004905568_1
1107880_000234825_3_Lmesenteroide_@WP_014325100_1
203123_000014385_1_Ooeni_@WP_011677714_1
1045004_000241055_1_Okitaharae_@WP_007747304_1
759620_000750515_1_Wceti_@WP_009496187_1
714313_000225325_1_Lsanfrancisce_@WP_014081380_1
137591_001308145_1_Wcibaria_@WP_043710662_1
1045854_000219805_1_Wkoreensis_@WP_006845901_1
137591_001308145_1_Wcibaria_@WP_043708854_1
525919_000024105_1_Aprevotii_@WP_015777250_1
699246_000025225_2_Mindolicus_@WP_012992702_1
697329_000179635_2_Ralbus_@WP_013499466_1
1160721_000723465_1_Rbicirculans_@WP_041337421_1
663278_000178115_2_Eharbinense_@WP_013486040_1
29343_000953215_1_Ccellulosi_@WP_034840899_1
697281_000213255_1_Maustraliensi_@WP_013781818_1
555079_000144645_1_Toceani_@WP_013275122_1
644966_000184705_1_Tmarianensis_@WP_013494951_1
1209989_000213235_1_Tacetatoxydan_@WP_013779475_1
1089553_000305935_1_Tphaeum_@WP_015051810_1
138119_000010045_1_Dhafniense_@WP_011462256_1
756499_000243155_3_Ddehalogenans_@AFM02416_1
871963_000243135_3_Ddichloroelim_@AGA70852_1
871968_000231405_3_Dmetallireduc_@AHF08296_1
768704_000231385_3_Dmeridiei_@AFQ46388_1
768706_000235605_1_Dorientis_@WP_014187723_1
646529_000255115_3_Dacidiphilus_@AFM43487_1
871738_000512895_1_Drestrictus_@WP_019225010_1
1147129_000305775_1_Dsp_@WP_015044937_1
645991_000190635_1_Sglycolicus_@WP_013626286_1
498761_000019165_1_Hmodesticaldu_@WP_049754224_1
1525_001267405_1_Mthermoacetic_@WP_011393885_1
643648_000092405_1_Slipocalidus_@WP_049765119_1
335541_000014725_1_Swolfei_@WP_041428102_1
0.4


985002_000236925_1_Sargenteus_@WP_000672873_1
93061_000013425_1_Saureus_@YP_499756_1
1283_000972725_1_Shaemolyticus_@WP_011275952_1
1034809_000270465_1_Slugdunensis_@WP_014533319_1
1295_001188875_1_Sschleiferi_@WP_050331672_1
1276282_000494875_1_Spasteuri_@WP_023374480_1
1194526_000332735_1_Swarneri_@WP_002465919_1
246432_001432245_1_Sequorum_@WP_002511588_1
396513_000009405_1_Scarnosus_@PSDG1_1655986_1e_156
407035_001005905_1_Shalodurans_@WP_046790335_1
1461582_000756715_2_Jsp_@CEA00377_1
985762_001442815_1_Sagnetis_@WP_060550841_1
1284_000816085_1_Shyicus_@WP_039646220_1
458233_000010585_1_Mcaseolyticus_@WP_012656039_1
1284_000816085_1_Shyicus_@WP_039643939_1
985762_001442815_1_Sagnetis_@WP_060551856_1
937773_000185885_1_Spseudinterme_@WP_015729788_1
342451_000010125_1_Ssaprophyticu_@WP_011302445_1
93061_000013425_1_Saureus_@YP_501158_1
985002_000236925_1_Sargenteus_@WP_001012242_1
1283_000972725_1_Shaemolyticus_@WP_011274967_1
1034809_000270465_1_Slugdunensis_@WP_002478696_1
1276282_000494875_1_Spasteuri_@WP_023373928_1
1194526_000332735_1_Swarneri_@WP_015364722_1
1282_000759555_1_Sepidermidis_@WP_038398732_1
72758_001028645_1_Scapitis_@WP_047797103_1
0.4

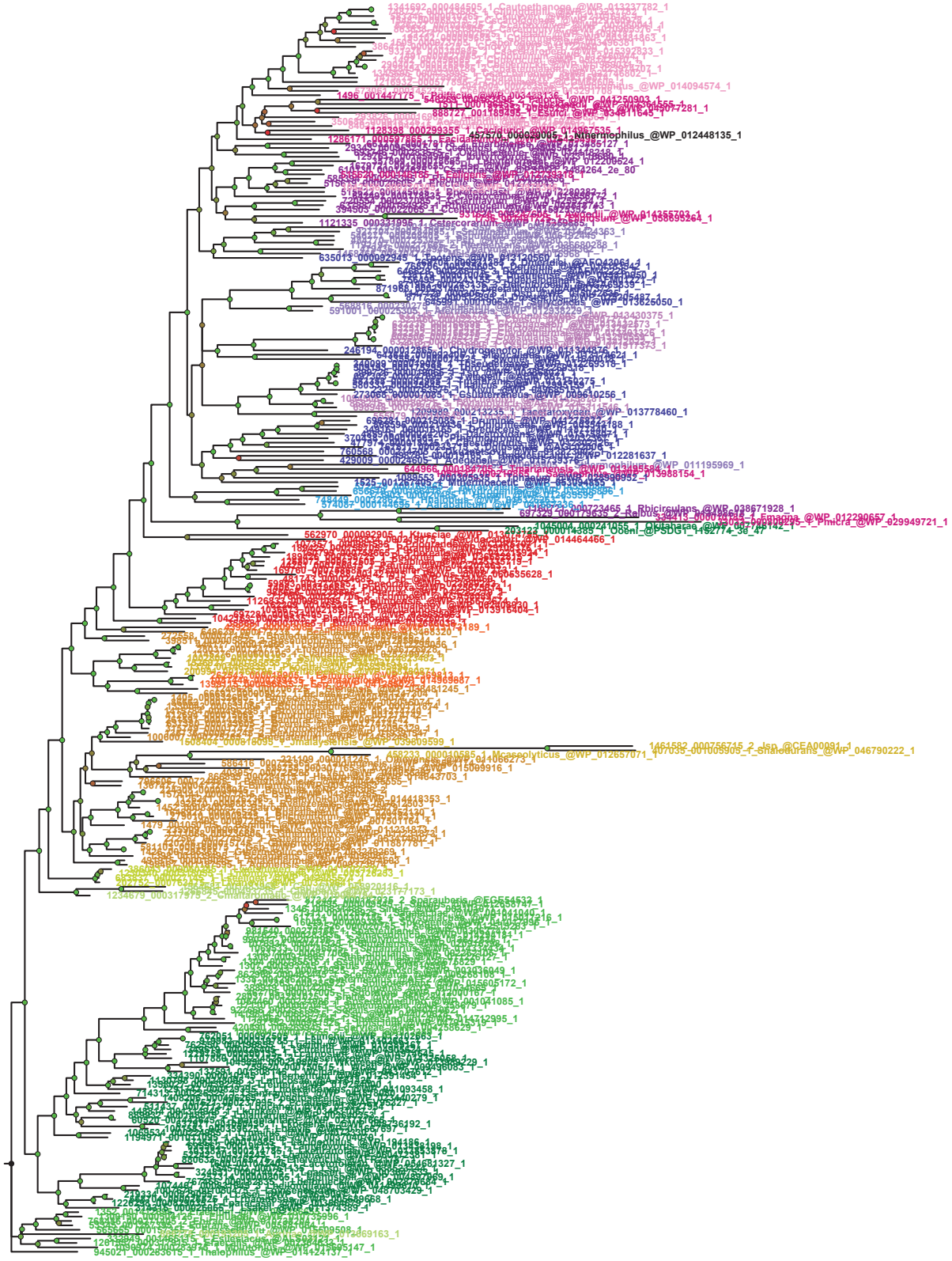

457570_000020005_1_Nthermophilus_@WP_012447765_1
650150_000270085_1_Erhusiopathia_@WP_013853066_1
0.5

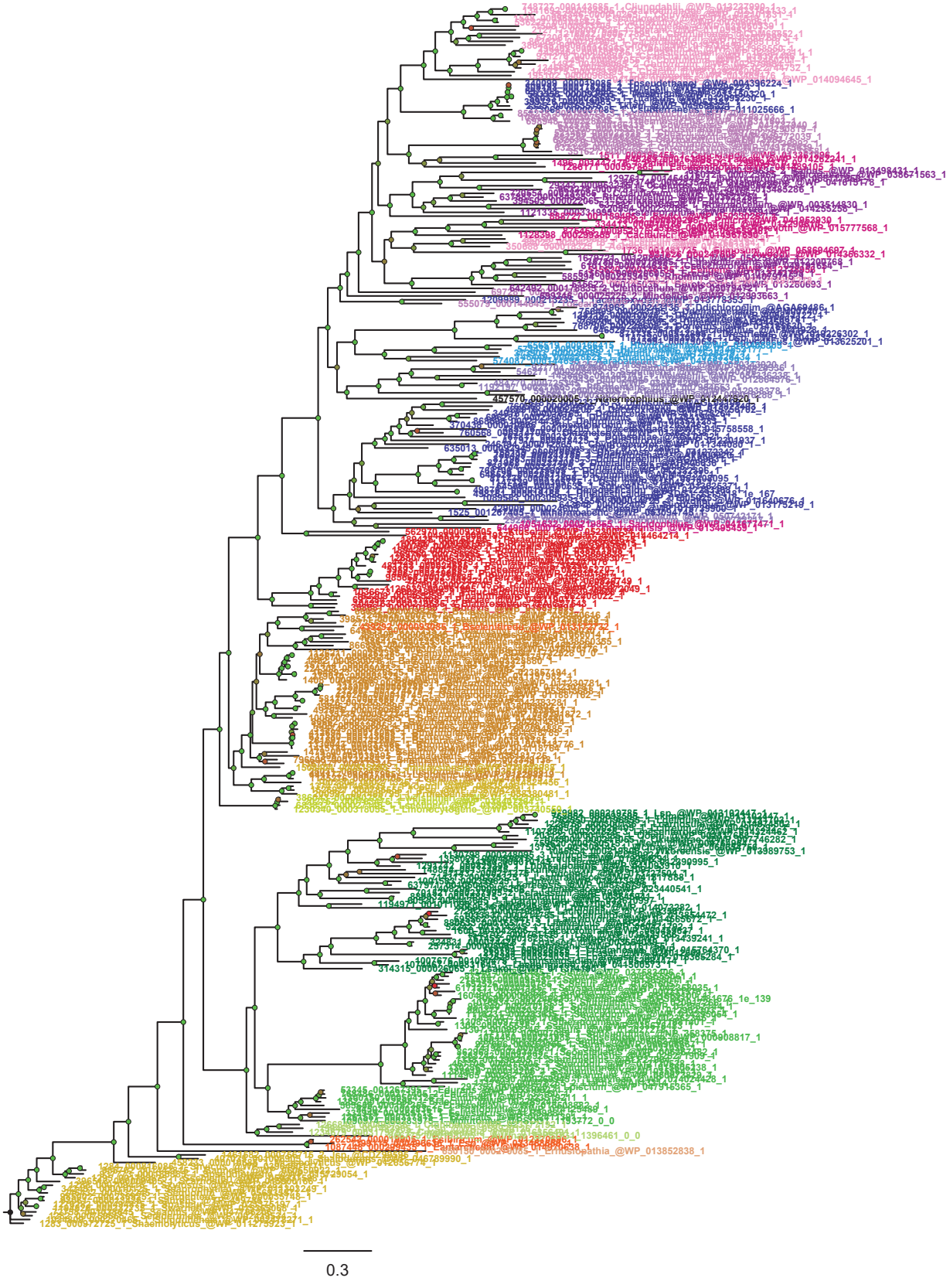

1069534_000224985_1_Lruminis_@WP_014073981_1
0.3

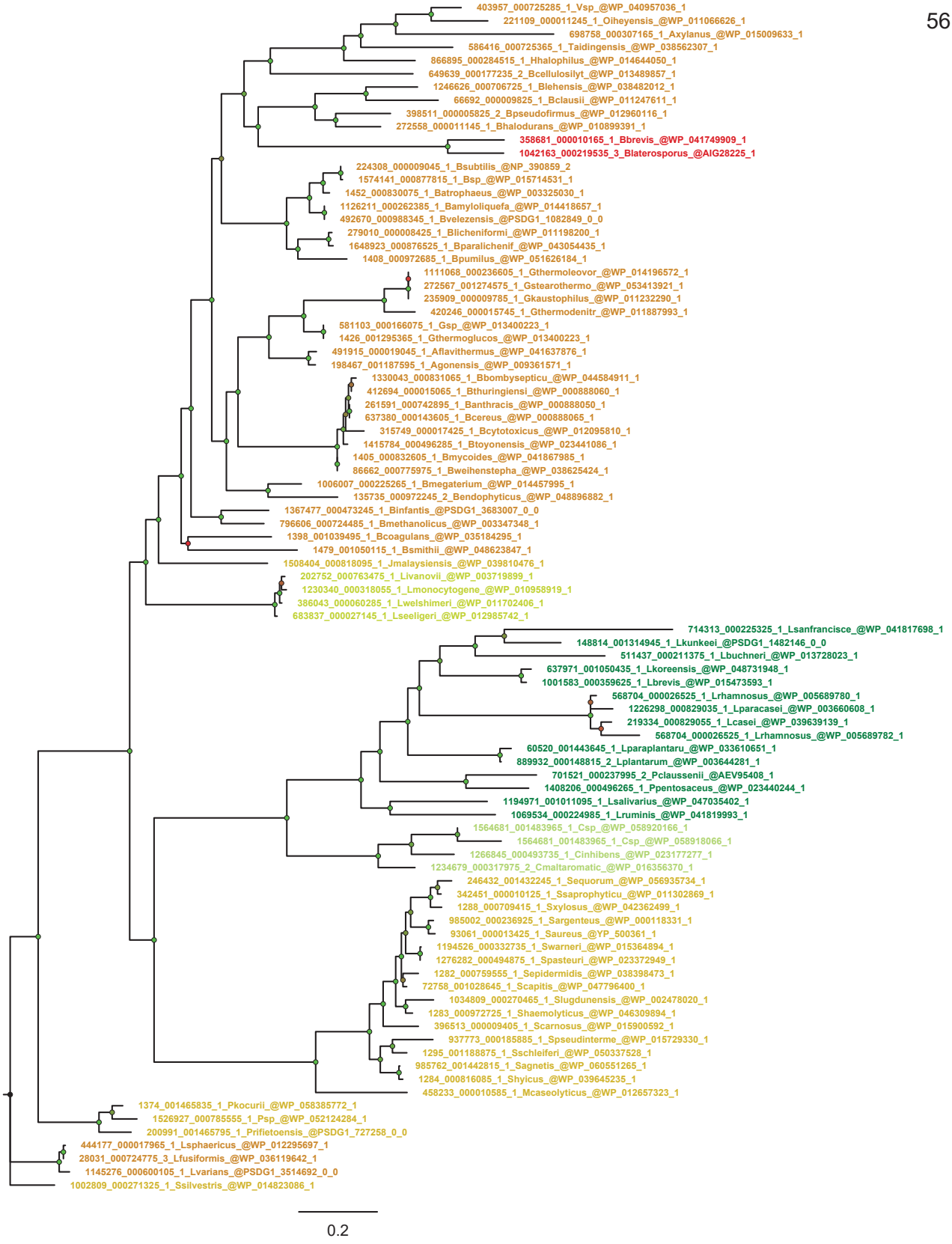

0.3

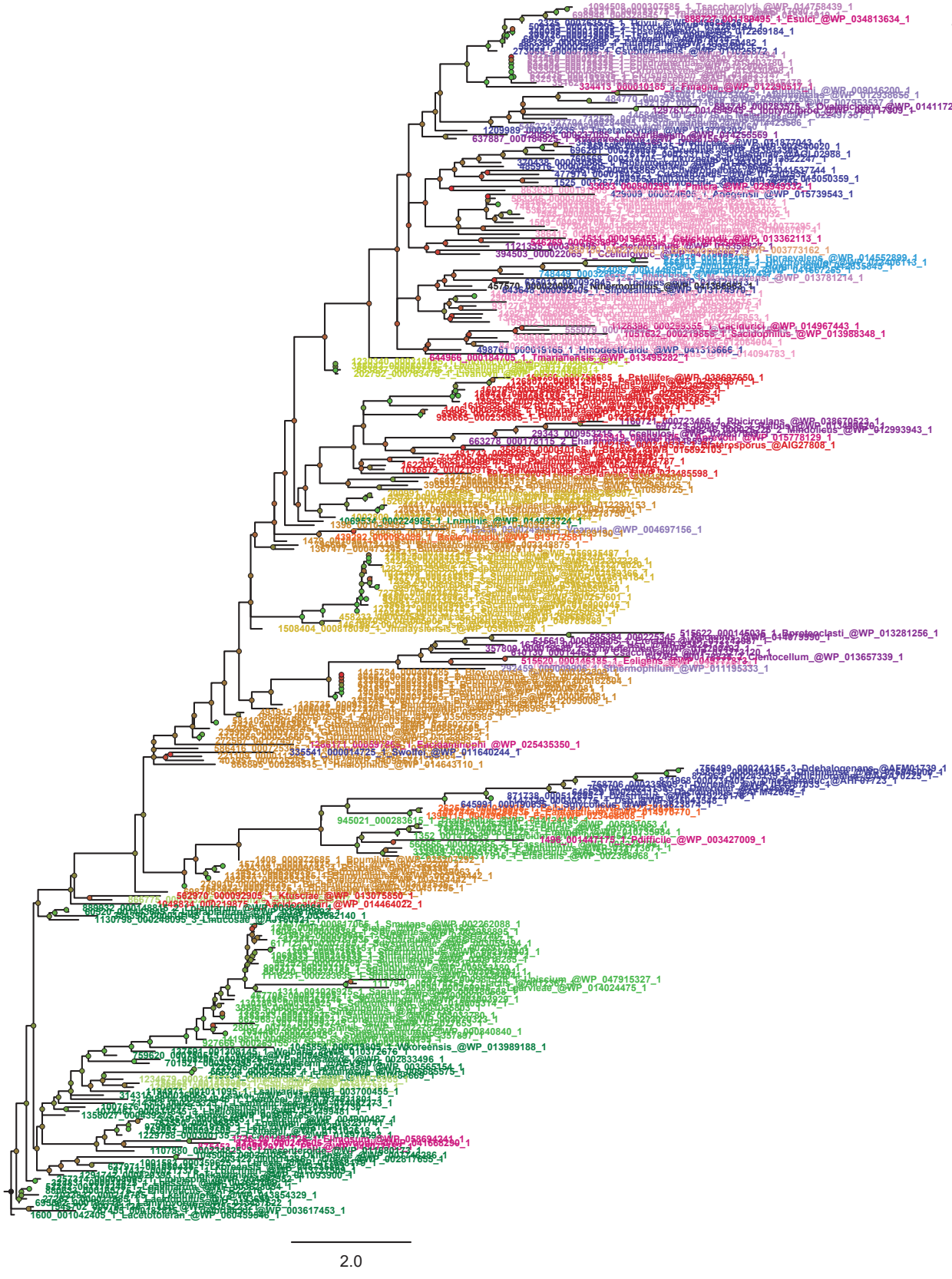

1341692_000484505_1_Cautoethanoge_@WP_023162415_1
748727_000143685_1_Cljungdahlii_@WP_013240438_1
583346_000010265_1_Ckluyveri_@WP_011989172_1
536227_001038625_1_Ccarboxidivor_@WP_007059286_1
1548_000968375_1_Cscatologenes_@WP_029954364_1
1509_000973705_1_Csporogenes_@WP_033061405_1
212717_000007625_1_Ctetani_@WP_011100587_1
863638_000191905_1_Cacetobutylic_@WP_010963922_1
386415_000014125_1_Cnovyi_@WP_011721431_1
195102_000009685_1_Cperfringens_@WP_003469656_1
290402_000016965_1_Cbeijerinckii_@WP_012060936_1
1428454_000506785_2_Cpasteurianum_@WP_023974368_1
1345695_000473995_1_Csaccharobuty_@WP_022750347_1
1492_001456065_2_Cbutyricum_@WP_058146844_1
931276_000340885_1_Csaccharoperb_@WP_015395513_1
1491_000827955_1_Cbotulinum_@WP_012450563_1
1415775_000789395_1_Cbaratii_@WP_039311503_1
585394_000225345_1_Rhominis_@WP_014078660_1
515619_000020605_1_Erectale_@WP_012741594_1
515620_000146185_1_Eeligens_@WP_012739907_1
1042156_000270305_1_Csp_@WP_013977246_1
1679721_001298655_2_Hsp_@WP_058257358_1
357809_000018685_1_Lphytoferment_@WP_012198356_1
515622_000145035_1_Bproteoclasti_@WP_013279503_1
610130_000144625_1_Csaccharolyti_@PSDG2_3207973_1e_120
1297617_001454945_1_Ibutyriciprod_@WP_052082665_1
693746_000283575_1_Ovalericigene_@PSDG1_3821787_0_0
536227_001038625_1_Ccarboxidivor_@WP_007062207_1
1548_000968375_1_Cscatologenes_@WP_029162969_1
931276_000340885_1_Csaccharoperb_@WP_015393337_1
1428454_000506785_2_Cpasteurianum_@WP_023975483_1
290402_000016965_1_Cbeijerinckii_@WP_012059229_1
931276_000340885_1_Csaccharoperb_@WP_015392963_1
568816_000230275_1_Aintestini_@WP_009016036_1
591001_000025305_1_Afermentans_@PSDG1_2068480_0_0
1496_001447175_1_Pdifficile_@WP_003428619_1
138119_000010045_1_Dhafniense_@WP_041272613_1
756499_000243155_3_Ddehalogenans_@AFL99654_1
1736_001481725_1_Elimosum_@WP_058694179_1
138119_000010045_1_Dhafniense_@WP_041272679_1
768706_000235605_1_Dorientis_@WP_014183080_1
871968_000231405_3_Dmetallireduc_@AHF08175_1
84022_001042715_1_Caceticum_@WP_044824724_1
350688_000018325_1_Aoremlandii_@WP_012160150_1
1216932_000577895_1_Cbornimense_@CDM67984_1
515620_000146185_1_Eeligens_@WP_012740920_1
585394_000225345_1_Rhominis_@WP_014081103_1
515619_000020605_1_Erectale_@WP_012744216_1
585394_000225345_1_Rhominis_@WP_014078248_1
1042156_000270305_1_Csp_@WP_013977027_1
1458465_001304715_1_Melsdenii_@WP_054336815_1
875453_000952975_1_Psp_@WP_045078578_1
610130_000144625_1_Csaccharolyti_@WP_013273693_1
1297617_001454945_1_Ibutyriciprod_@WP_058116917_1
875453_000952975_1_Psp_@WP_045078712_1
334413_000010185_1_Fmagna_@WP_002838569_1
525919_000024105_1_Aprevotii_@WP_015778631_1
0.2

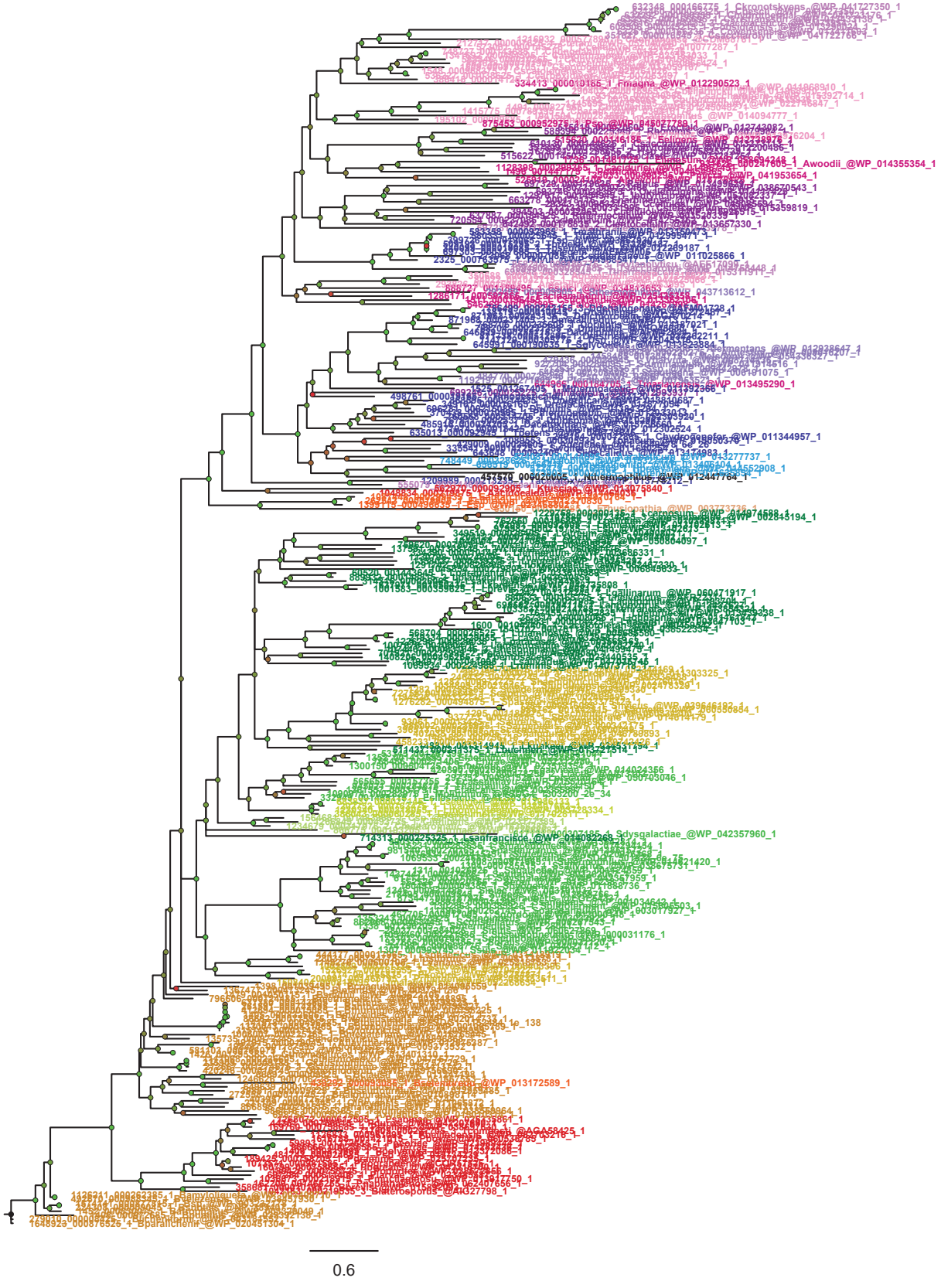
1341692_000484505_1_Cautoethanoge_@WP_023162805_1
748727_000143685_1_Cljungdahlii_@WP_013237293_1
583346_000010265_1_Ckluyveri_@WP_012103063_1
1548_000968375_1_Cscatologenes_@WP_029160597_1
536227_001038625_1_Ccarboxidivor_@WP_007062026_1
1509_000973705_1_Csporogenes_@WP_004461562_1
212717_000007625_1_Ctetani_@WP_011098819_1
573061_000145275_1_Ccellulovoran_@WP_010075703_1
863638_000191905_1_Cacetobutylic_@WP_010964273_1
386415_000014125_1_Cnovyi_@WP_039225449_1
1428454_000506785_2_Cpasteurianum_@WP_012058741_1
290402_000016965_1_Cbeijerinckii_@WP_012058741_1
931276_000340885_1_Csaccharoperb_@WP_015394223_1
1345695_000473995_1_Csaccharobuty_@WP_022746519_1
1216932_000577895_1_Cbornimense_@CDM67425_1
293826_000016985_1_Ametalliredig_@WP_012064275_1
84022_001042715_1_Caceticum_@WP_044823661_1
350688_000018325_1_Aoremlandii_@WP_041719736_1
1128398_000299355_1_Cacidurici_@WP_014967125_1
697281_000213255_1_Maustraliensi_@WP_041644117_1
1679721_001298655_2_Hsp_@WP_058259164_1
357809_000018685_1_Lphytoferment_@WP_012200071_1
642492_000178835_2_Clentocellum_@WP_013656462_1
1511_000196455_1_Csticklandii_@WP_013360440_1
536227_001038625_1_Ccarboxidivor_@WP_007061677_1
509193_000175295_2_Tbrockii_@WP_012269064_1
340099_000019085_1_Tpseudethanol_@WP_012269064_1
399726_000019065_1_Tsp_@WP_009052774_1
697303_000147695_3_Twiegelii_@AEM79363_1
580331_000025645_1_Titalicus_@WP_012995753_1
583358_000092965_1_Tmathranii_@WP_013150693_1
273068_000007085_1_Csubterraneus_@WP_009610946_1
2325_000763575_1_Tkivui_@WP_049685625_1
858215_000189775_3_Txylanolyticu_@AEF16547_1
1094508_000307585_1_Tsaccharolyti_@WP_014758111_1
698948_000328545_1_Tthermosaccha_@WP_015311033_1
350688_000018325_1_Aoremlandii_@WP_041719931_1
293826_000016985_1_Ametalliredig_@WP_012062847_1
84022_001042715_1_Caceticum_@WP_044823364_1
1496_001447175_1_Pdifficile_@WP_003439135_1
1128398_000299355_1_Cacidurici_@WP_014967059_1
1367477_000473245_1_Binfantis_@WP_009793688_1
796606_000724485_1_Bmethanolicus_@WP_004434237_1
509193_000175295_2_Tbrockii_@WP_003867331_1
399726_000019065_1_Tsp_@WP_003867331_1
340099_000019085_1_Tpseudethanol_@WP_003867331_1
583358_000092965_1_Tmathranii_@WP_012994286_1
580331_000025645_1_Titalicus_@WP_012994286_1
2325_000763575_1_Tkivui_@WP_049684325_1
697303_000147695_3_Twiegelii_@AEM77643_1
273068_000007085_1_Csubterraneus_@WP_011024629_1
858215_000189775_3_Txylanolyticu_@AEF16167_1
1094508_000307585_1_Tsaccharolyti_@WP_014757608_1
698948_000328545_1_Tthermosaccha_@WP_015310671_1
760568_000214705_1_Dkuznetsovii_@WP_013823431_1
767817_000233715_3_Dgibsoniae_@AGL01327_1
370438_000010565_1_Pthermopropio_@WP_050748135_1
477974_000018425_1_Cdesulforudis_@WP_012301941_1
429009_000024605_1_Adegensii_@WP_015738651_1
485916_000024205_1_Dacetoxidans_@WP_015758550_1
696281_000215085_1_Druminis_@WP_013841960_1
868595_000214435_1_Dnigrificans_@WP_013810199_1
349161_000016165_1_Dreducens_@WP_011878255_1
635013_000092945_1_Tpotens_@WP_013120310_1
748449_000328625_1_Hhalobius_@WP_015327268_1
574087_000144695_1_Aarabaticum_@WP_013278471_1
373903_000020485_1_Horenii_@WP_012636128_1
457570_000020005_1_Nthermophilus_@WP_012447925_1
555079_000144645_1_Toceani_@WP_013276015_1
1209989_000213235_1_Tacetatoxydan_@WP_013778357_1
484770_000725345_1_Psp_@WP_038672110_1
1192197_000271665_2_Pfermentans_@WP_007952686_1
335541_000014725_1_Swolfei_@WP_011640669_1
643648_000092405_1_Slipocalidus_@WP_013175226_1
1051632_000219855_1_Sacidophilus_@WP_013986577_1
644966_000184705_1_Tmarianensis_@WP_013495478_1
292459_000009905_1_Sthermophilum_@WP_011195767_1
1089553_000305935_1_Tphaeum_@WP_015050258_1
498761_000019165_1_Hmodesticaldu_@WP_012283419_1
1525_001267405_1_Mthermoacetic_@WP_011392591_1
138119_000010045_1_Dhafniense_@WP_005812900_1
756499_000243155_3_Ddehalogenans_@AFM00910_1
871963_000243135_3_Ddichloroelim_@AGA69619_1
871968_000231405_3_Dmetallireduc_@AHF06645_1
768704_000231385_3_Dmeridiei_@AFQ44655_1
768706_000235605_1_Dorientis_@WP_014185782_1
646529_000255115_3_Dacidiphilus_@AFM42116_1
562970_000092905_1_Ktusciae_@WP_013075672_1
1048834_000219875_1_Aacidocaldari_@WP_014464336_1
358681_000010165_1_Bbrevis_@WP_012686196_1
1042163_000219535_3_Blaterosporus_@AIG26284_1
189425_000758705_1_Pgraminis_@WP_025703687_1
1073571_000981585_1_Priograndensi_@CQR56831_1
160799_000758665_1_Pborealis_@WP_042215447_1
189426_000758725_1_Podorifer_@WP_038572410_1
44251_000756615_1_Pdurus_@WP_042207268_1
1268072_000612505_1_Psabinae_@WP_025335537_1
169760_000758685_1_Pstellifer_@WP_038696987_1
481743_000024685_1_Psp_@WP_015734565_1
1616788_001421015_1_Pbovis_@WP_060535521_1
1406_000819665_1_Ppolymyxa_@WP_013371632_1
59893_001272655_1_Ppeoriae_@WP_053325659_1
985665_000235585_1_Pterrae_@WP_014282052_1
717605_000227705_3_Tcomposti_@AGA57939_1
1126833_000961095_1_Pbeijingensis_@WP_045672554_1
697284_000511405_1_Plarvae_@WP_042118838_1
162209_001465255_1_Pnaphthalenov_@WP_062408830_1
1036673_000218915_1_Pmucilaginosu_@WP_013916555_1
0.4

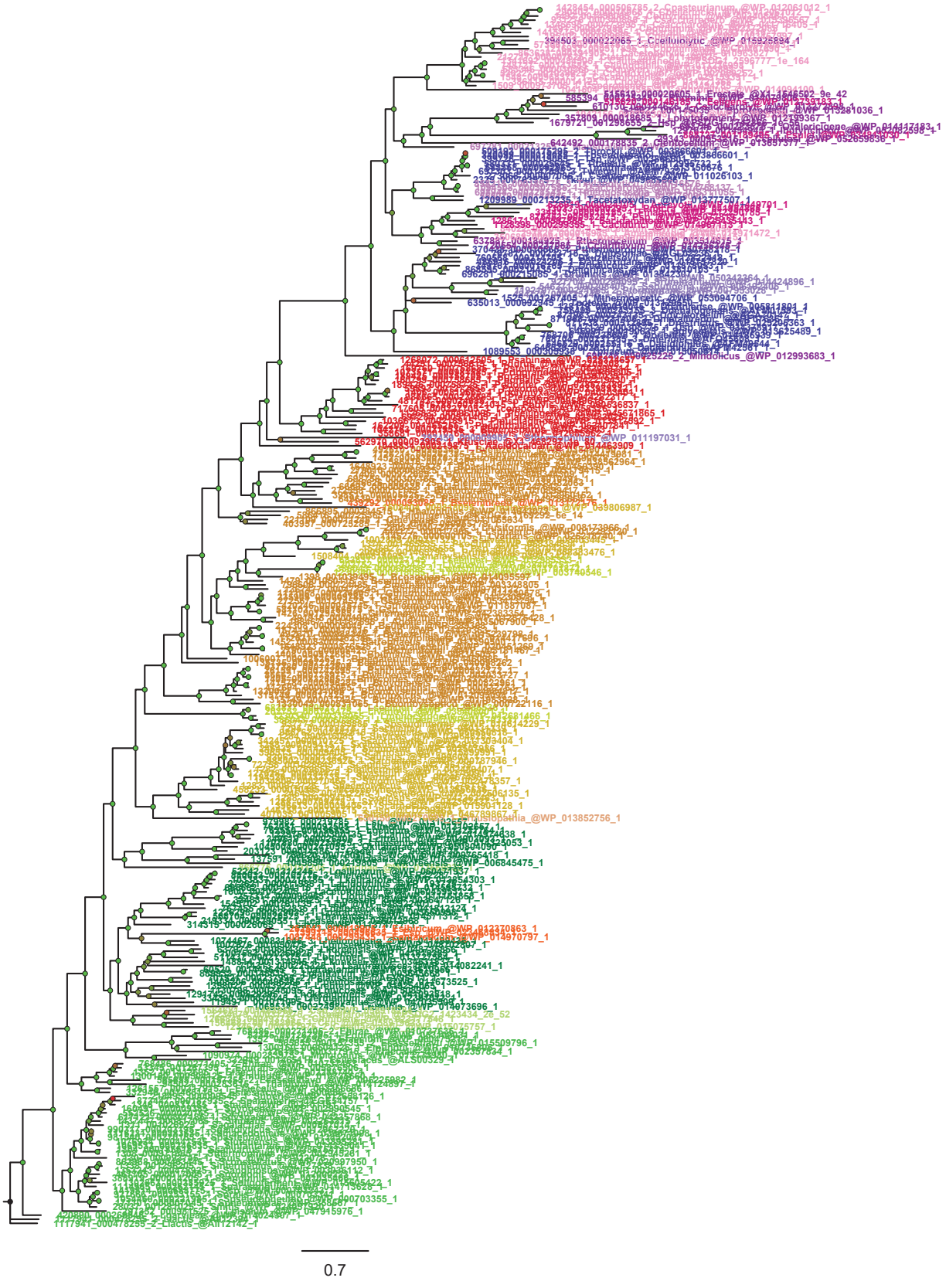

0.4

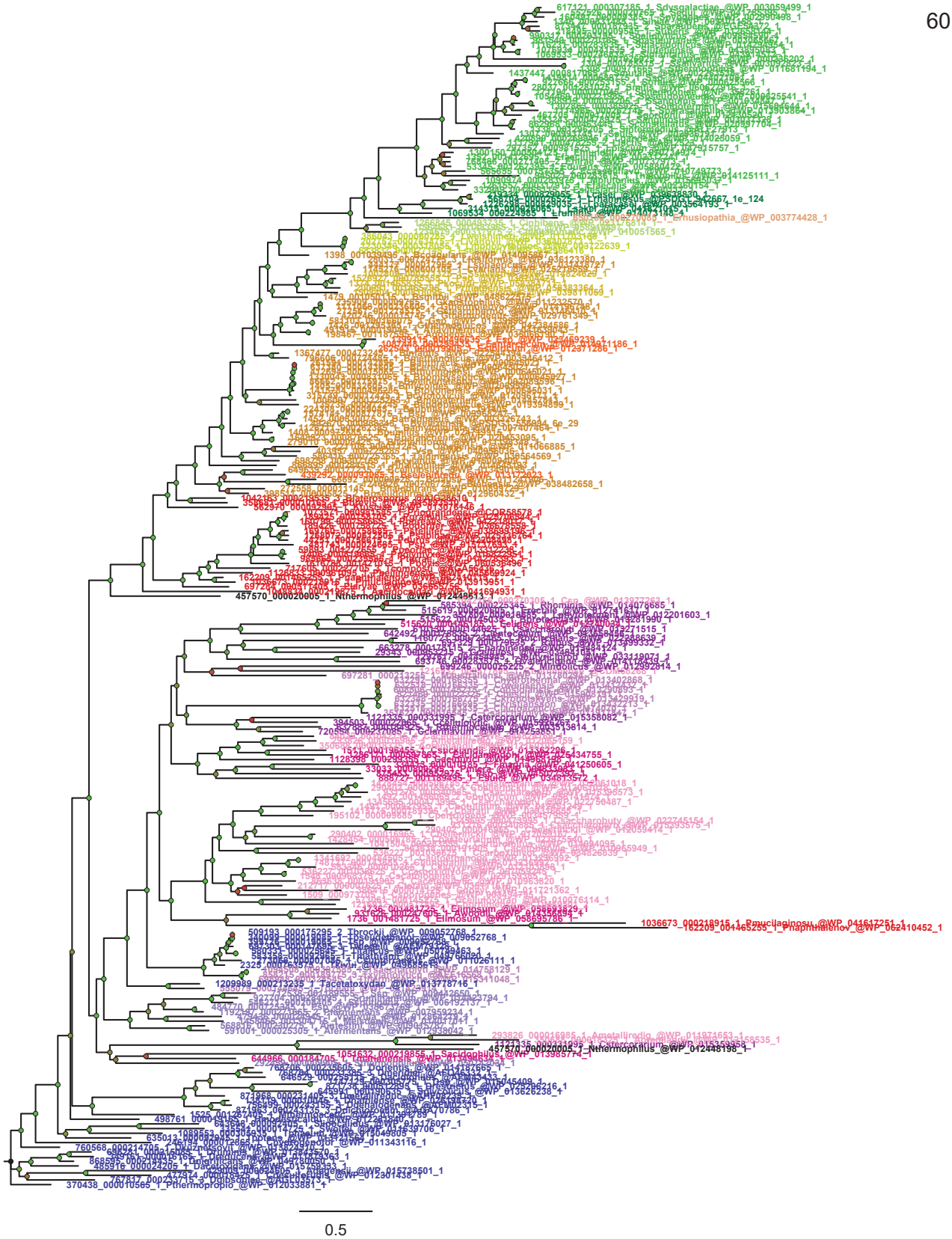

0.3

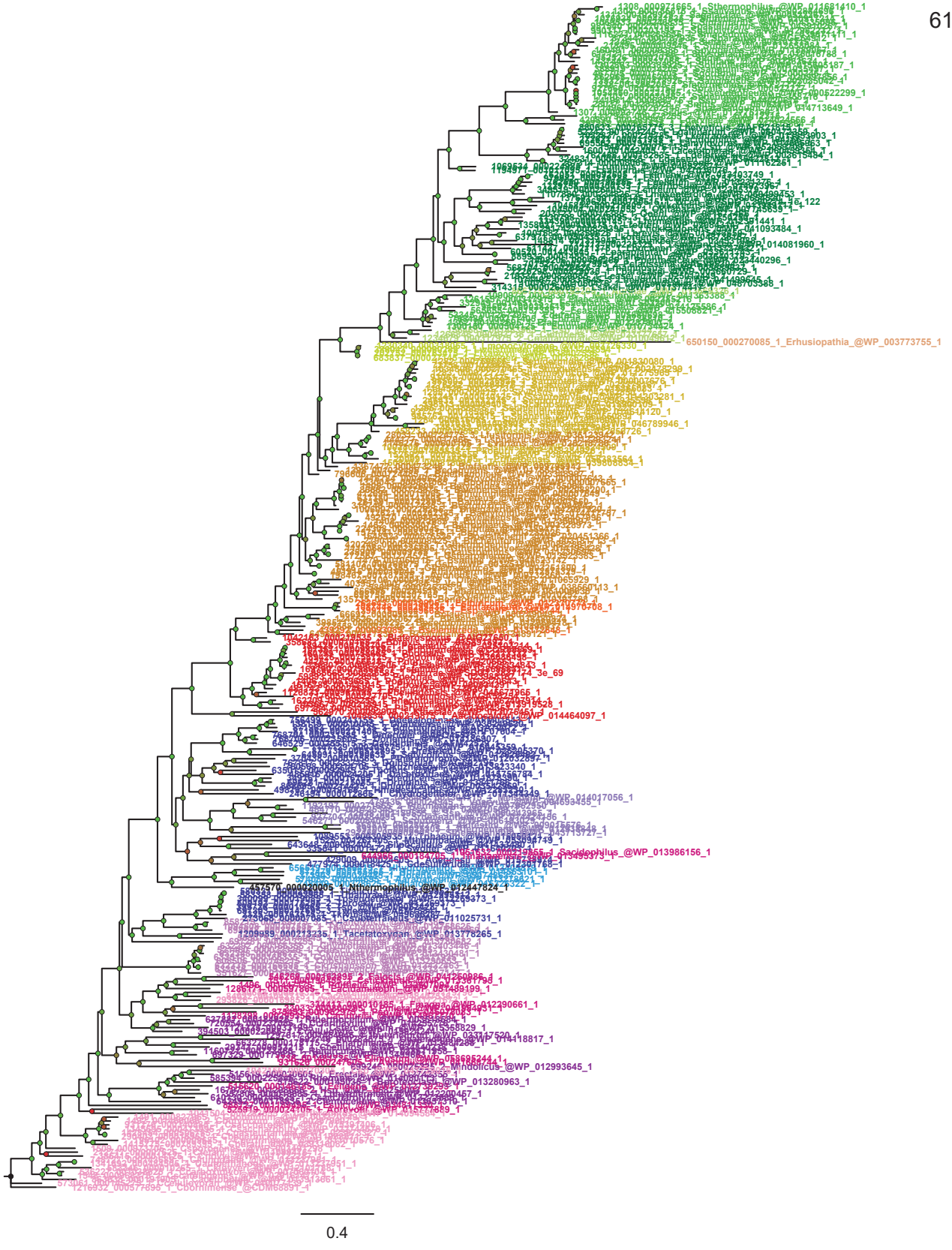

403957_000725285_1_Vsp_@WP_040957036_1
221109_000011245_1_Oiheyensis_@WP_011066626_1
698758_000307165_1_Axylanus_@WP_015009633_1
586416_000725365_1_Taidingensis_@WP_038562307_1
866895_000284515_1_Hhalophilus_@WP_014644050_1
649639_000177235_2_Bcellulosilyt_@WP_013489857_1
1246626_000706725_1_Blehensis_@WP_038482012_1
66692_000009825_1_Bclausii_@WP_011247611_1
398511_000005825_2_Bpseudofirmus_@WP_012960116_1
272558_000011145_1_Bhalodurans_@WP_010899391_1
358681_000010165_1_Bbrevis_@WP_041749909_1
1042163_000219535_3_Blaterosporus_@AIG28225_1
224308_000009045_1_Bsubtilis_@NP_390859_2
1574141_000877815_1_Bsp_@WP_015714531_1
1452_000830075_1_Batrophaeus_@WP_003325030_1
1126211_000262385_1_Bamyloliquefa_@WP_014418657_1
492670_000988345_1_Bvelezensis_@PSDG1_1082849_0_0
279010_000008425_1_Blicheniformi_@WP_011198200_1
1648923_000876525_1_Bparalichenif_@WP_043054435_1
1408_000972685_1_Bpumilus_@WP_051626184_1
1111068_000236605_1_Gthermoleovor_@WP_014196572_1
272567_001274575_1_Gstearothermo_@WP_053413921_1
235909_000009785_1_Gkaustophilus_@WP_011232290_1
420246_000015745_1_Gthermodenitr_@WP_011887993_1
581103_000166075_1_Gsp_@WP_013400223_1
1426_001295365_1_Gthermoglucos_@WP_013400223_1
491915_000019045_1_Aflavithermus_@WP_041637876_1
198467_001187595_1_Agonensis_@WP_009361571_1
1330043_000831065_1_Bbombysepticu_@WP_044584911_1
412694_000015065_1_Bthuringiensi_@WP_000888060_1
261591_000742895_1_Banthracis_@WP_000888050_1
637380_000143605_1_Bcereus_@WP_000888065_1
315749_000017425_1_Bcytotoxicus_@WP_012095810_1
1415784_000496285_1_Btoyonensis_@WP_023441086_1
1405_000832605_1_Bmycoides_@WP_041867985_1
86662_000775975_1_Bweihenstepha_@WP_038625424_1
1006007_000225265_1_Bmegaterium_@WP_014457995_1
135735_000972245_2_Bendophyticus_@WP_048896882_1
1367477_000473245_1_Binfantis_@PSDG1_3683007_0_0
796606_000724485_1_Bmethanolicus_@WP_003347348_1
1398_001039495_1_Bcoagulans_@WP_035184295_1
1479_001050115_1_Bsmithii_@WP_048623847_1
1508404_000818095_1_Jmalaysiensis_@WP_039810476_1
202752_000763475_1_Livanovii_@WP_003719899_1
1230340_000318055_1_Lmonocytogene_@WP_010958919_1
386043_000060285_1_Lwelshimeri_@WP_011702406_1
683837_000027145_1_Lseeligeri_@WP_012985742_1
714313_000225325_1_Lsanfrancisce_@WP_041817698_1
148814_001314945_1_Lkunkeei_@PSDG1_1482146_0_0
511437_000211375_1_Lbuchneri_@WP_013728023_1
637971_001050435_1_Lkoreensis_@WP_048731948_1
1001583_000359625_1_Lbrevis_@WP_015473593_1
568704_000026525_1_Lrhamnosus_@WP_005689780_1
1226298_000829035_1_Lparacasei_@WP_003660608_1
219334_000829055_1_Lcasei_@WP_039639139_1
568704_000026525_1_Lrhamnosus_@WP_005689782_1
60520_001443645_1_Lparaplantaru_@WP_033610651_1
889932_000148815_2_Lplantarum_@WP_003644281_1
701521_000237995_2_Pclaussenii_@AEV95408_1
1408206_000496265_1_Ppentosaceus_@WP_023440244_1
1194971_001011095_1_Lsalivarius_@WP_047035402_1
1069534_000224985_1_Lruminis_@WP_041819993_1
1564681_001483965_1_Csp_@WP_058920166_1
1564681_001483965_1_Csp_@WP_058918066_1
1266845_000493735_1_Cinhibens_@WP_023177277_1
1234679_000317975_2_Cmaltaromatic_@WP_016356370_1
246432_001432245_1_Sequorum_@WP_056935734_1
342451_000010125_1_Ssaprophyticu_@WP_011302869_1
1288_000709415_1_Sxylosus_@WP_042362499_1
985002_000236925_1_Sargenteus_@WP_000118331_1
93061_000013425_1_Saureus_@YP_500361_1
1194526_000332735_1_Swarneri_@WP_015364894_1
1276282_000494875_1_Spasteuri_@WP_023372949_1
1282_000759555_1_Sepidermidis_@WP_038398473_1
72758_001028645_1_Scapitis_@WP_047796400_1
1034809_000270465_1_Slugdunensis_@WP_002478020_1
1283_000972725_1_Shaemolyticus_@WP_046309894_1
396513_000009405_1_Scarnosus_@WP_015900592_1
937773_000185885_1_Spseudinterme_@WP_015729330_1
1295_001188875_1_Sschleiferi_@WP_050337528_1
985762_001442815_1_Sagnetis_@WP_060551265_1
1284_000816085_1_Shyicus_@WP_039645235_1
458233_000010585_1_Mcaseolyticus_@WP_012657323_1
1374_001465835_1_Pkocurii_@WP_058385772_1
1526927_000785555_1_Psp_@WP_052124284_1
200991_001465795_1_Prifietoensis_@PSDG1_727258_0_0
444177_000017965_1_Lsphaericus_@WP_012295697_1
28031_000724775_3_Lfusiformis_@WP_036119642_1
1145276_000600105_1_Lvarians_@PSDG1_3514692_0_0
1002809_000271325_1_Ssilvestris_@WP_014823086_1
0.2

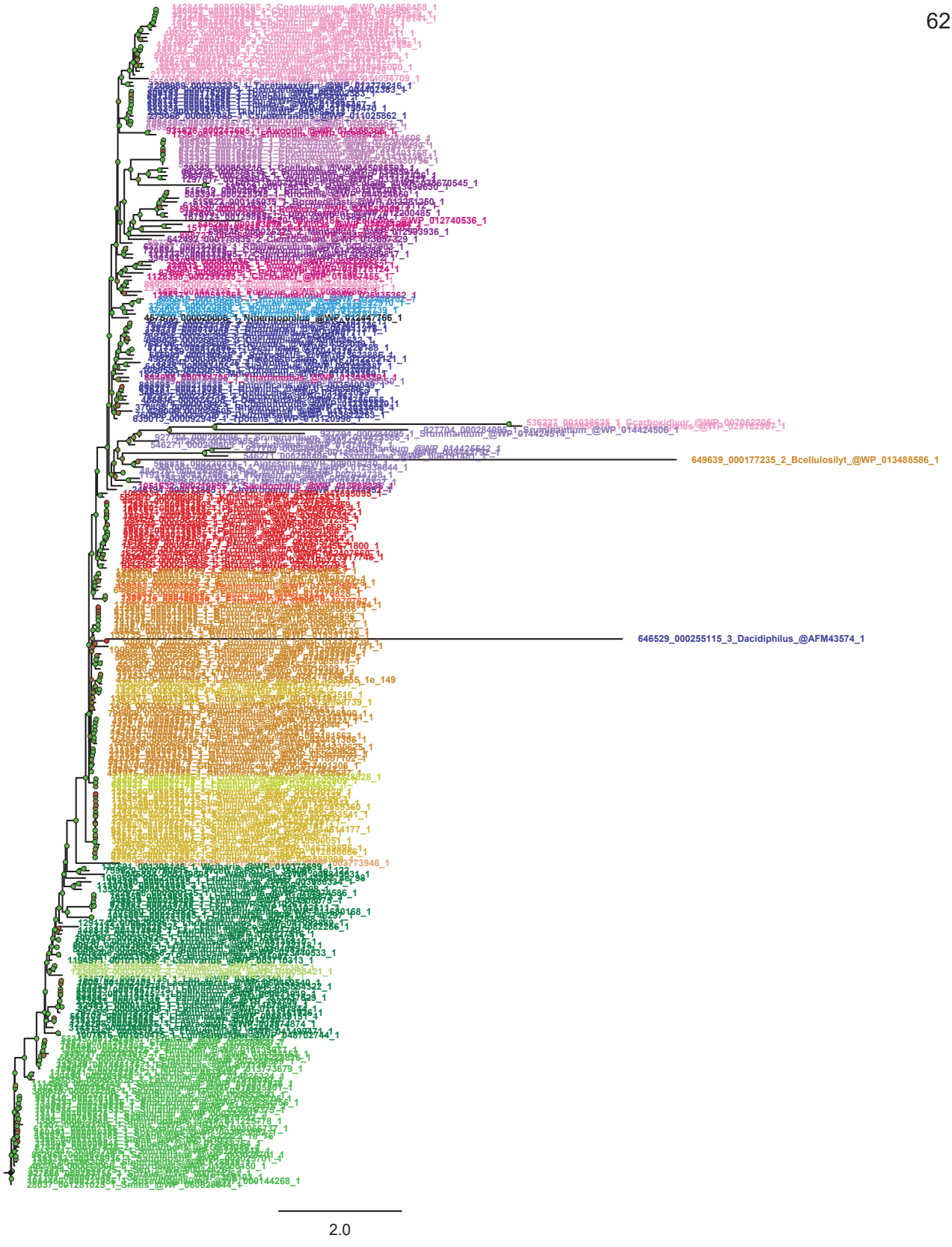

2.0

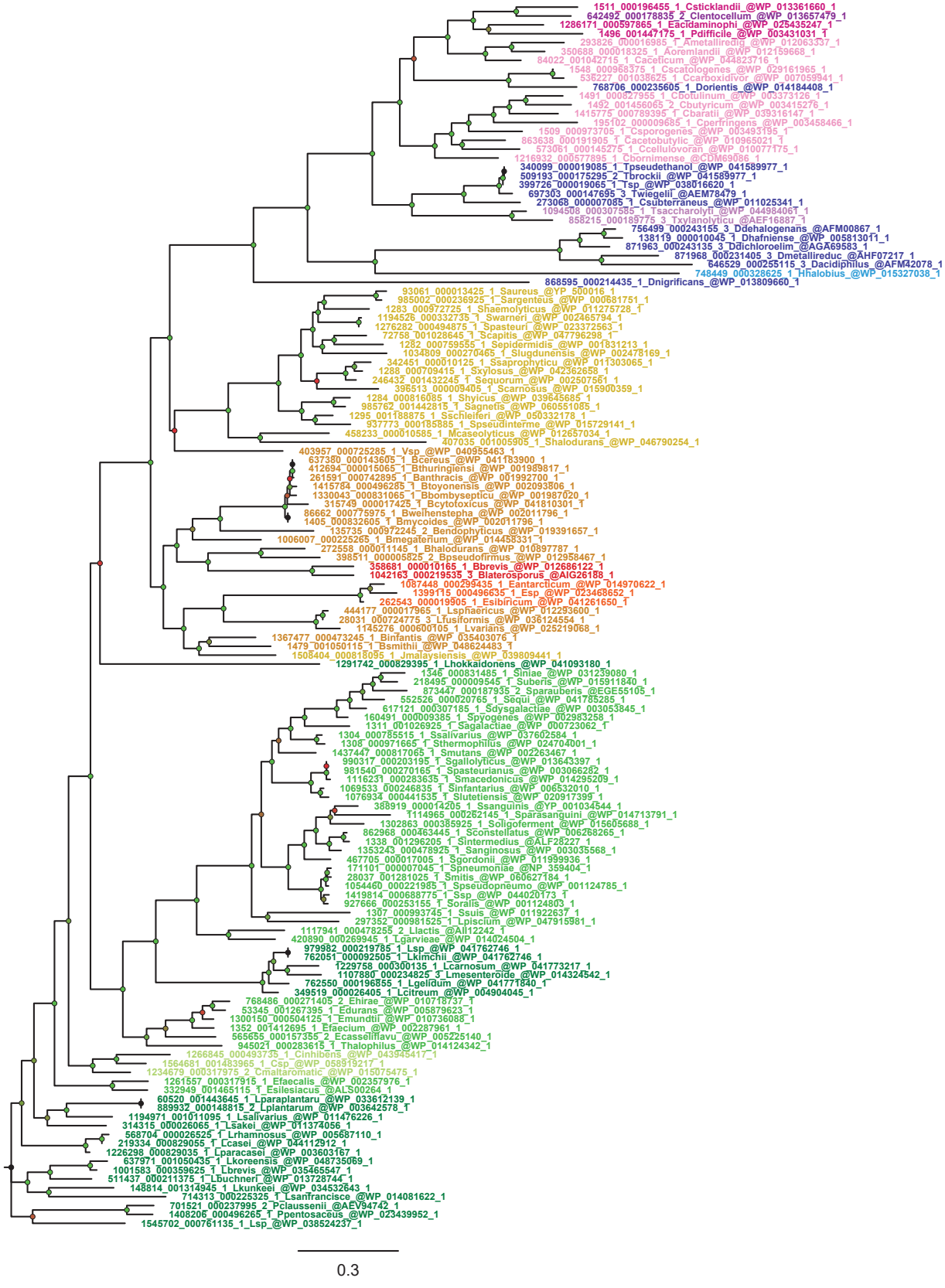

0.6

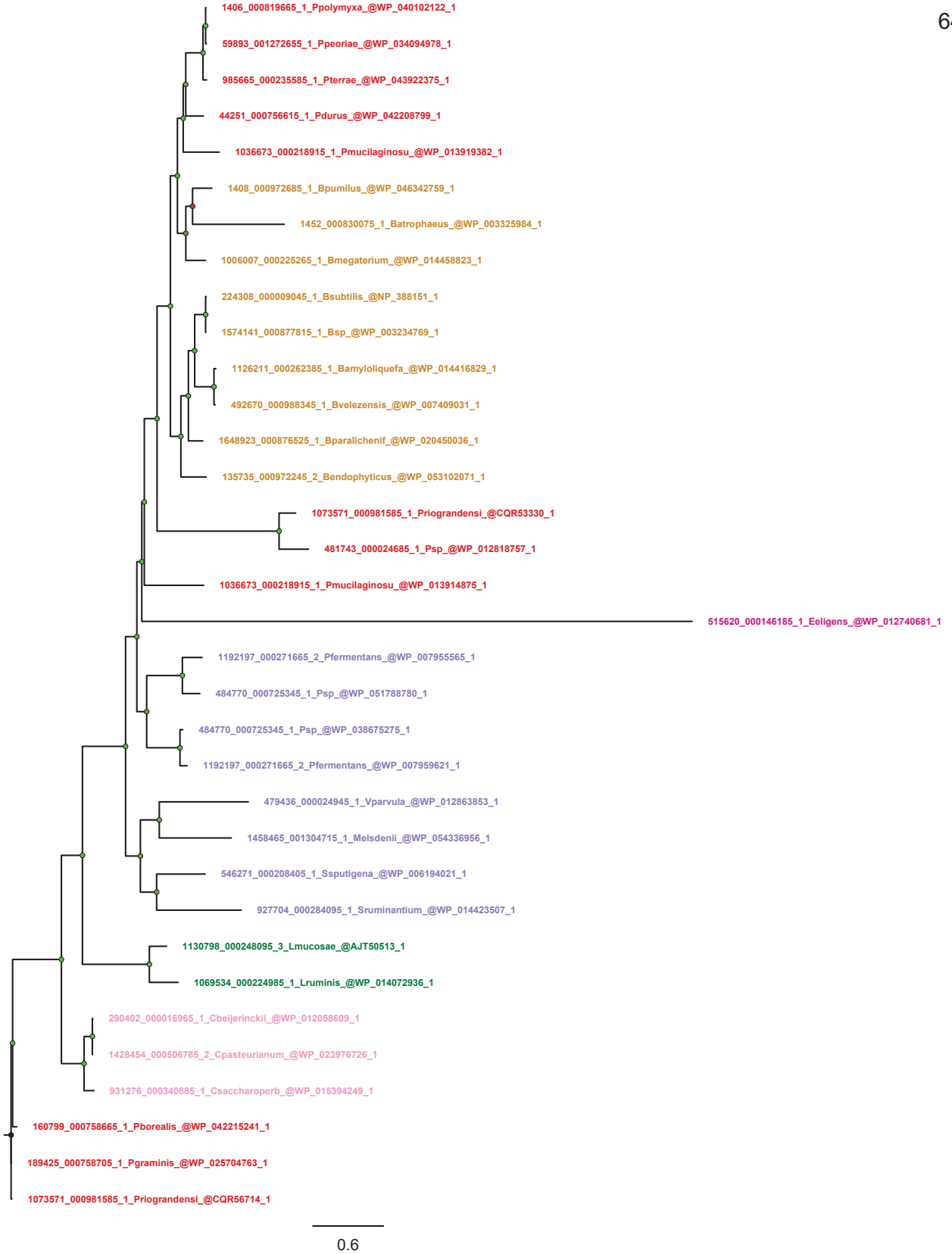

0.7

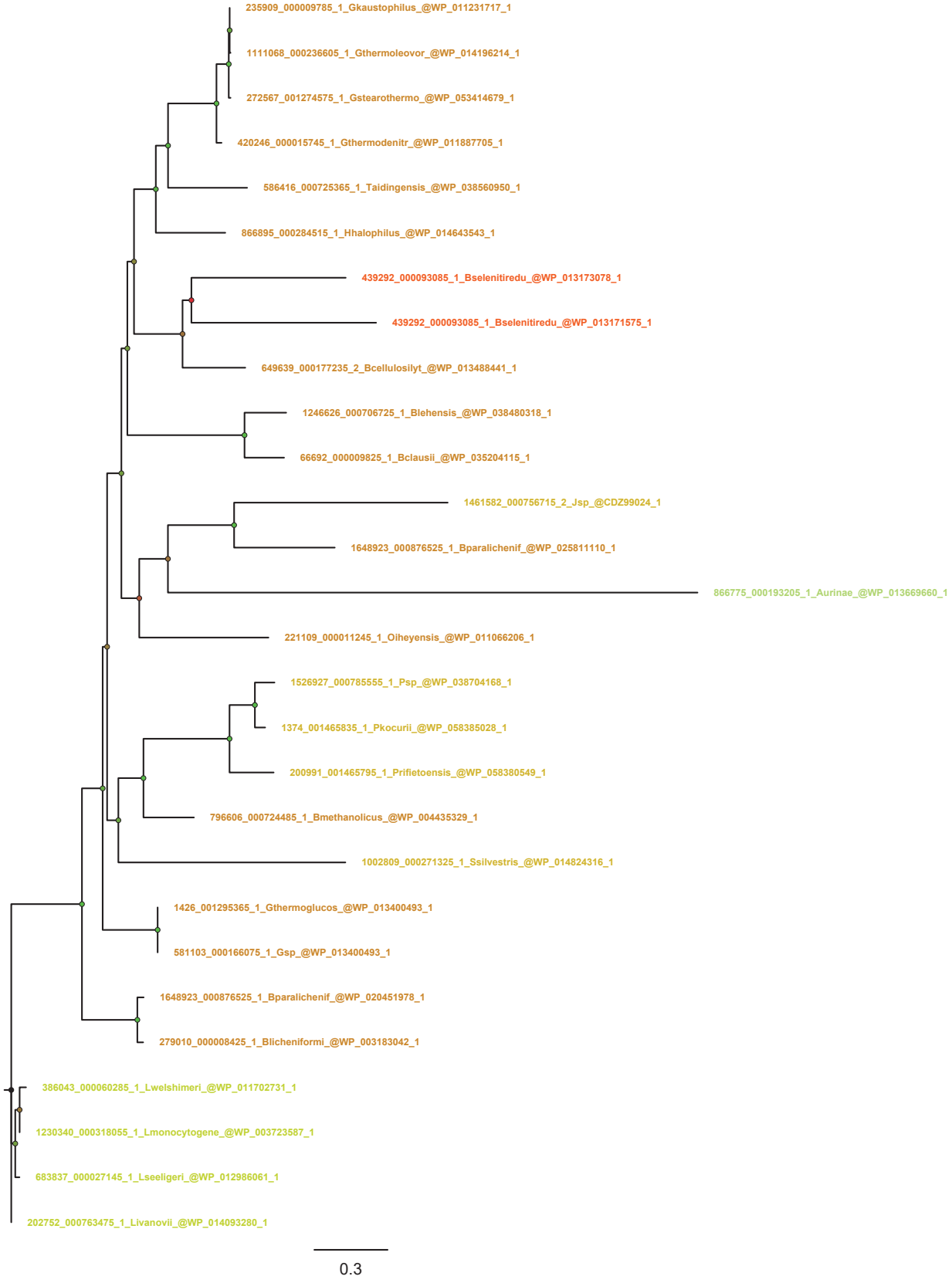

0.5

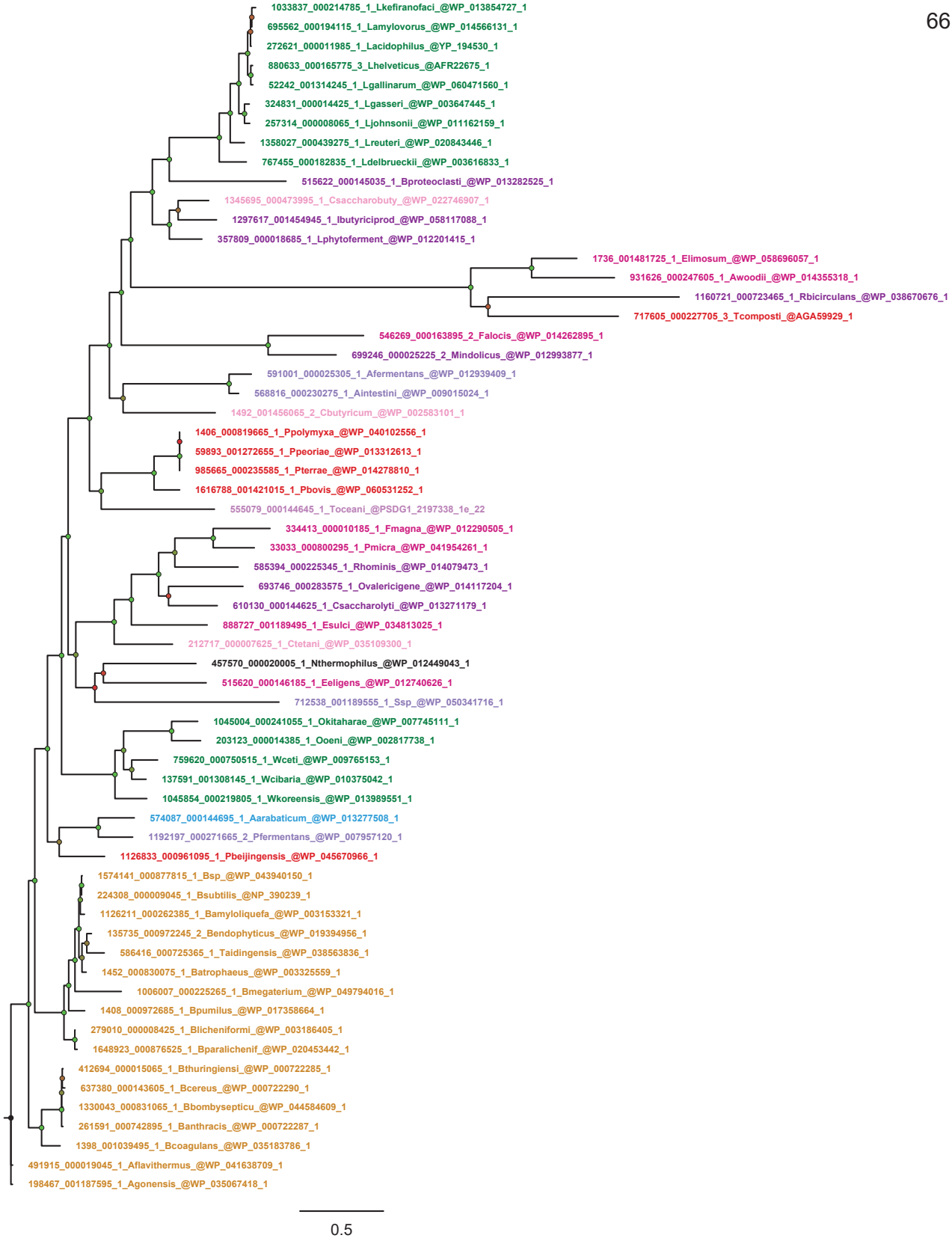

650150_000270085_1_Erhusiopathia_@WP_003773755_1
457570_000020005_1_Nthermophilus_@WP_012447824_1
0.4

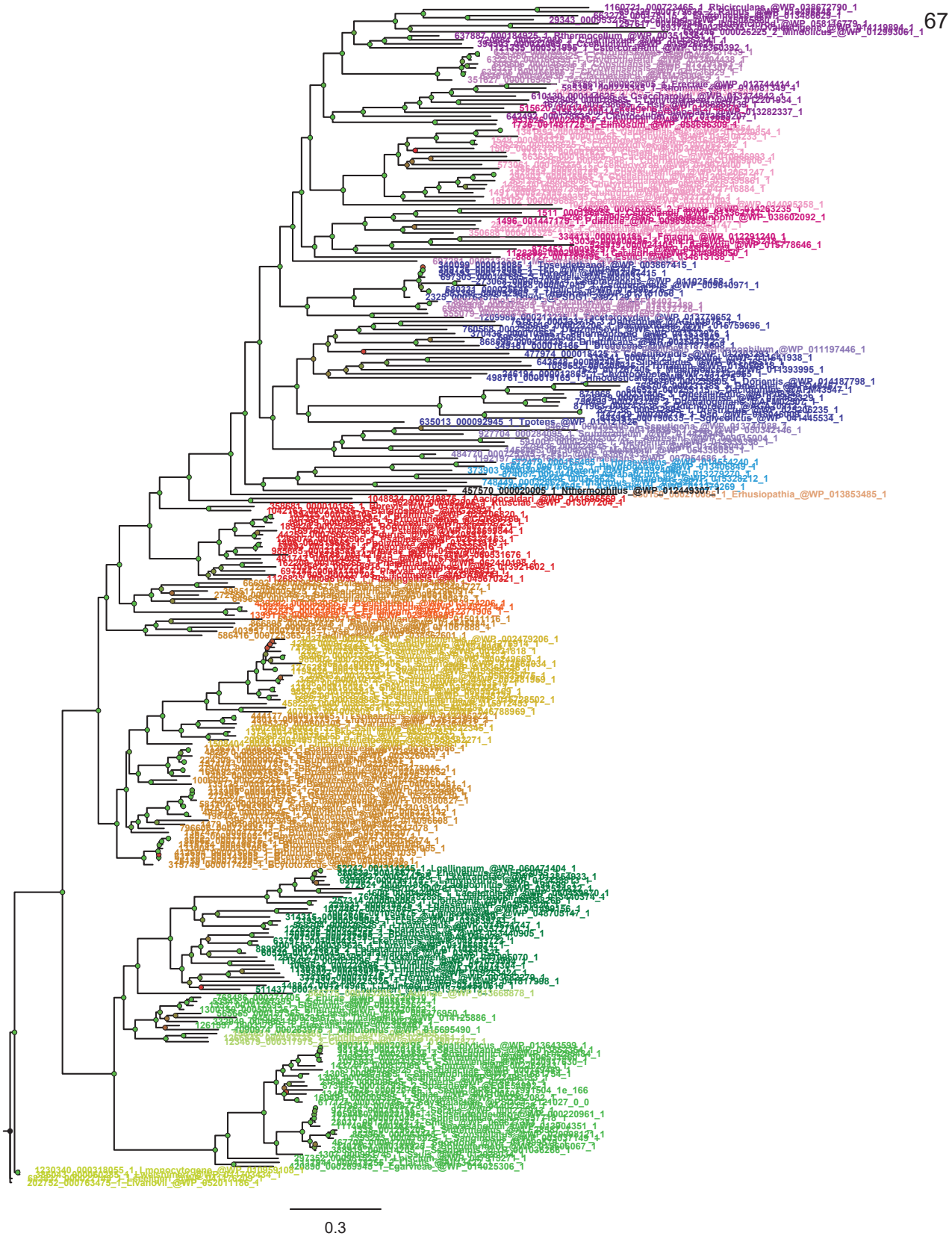

649639_000177235_2_Bcellulosilyt_@WP_013488586_1
646529_000255115_3_Dacidiphilus_@AFM43574_1
2.0

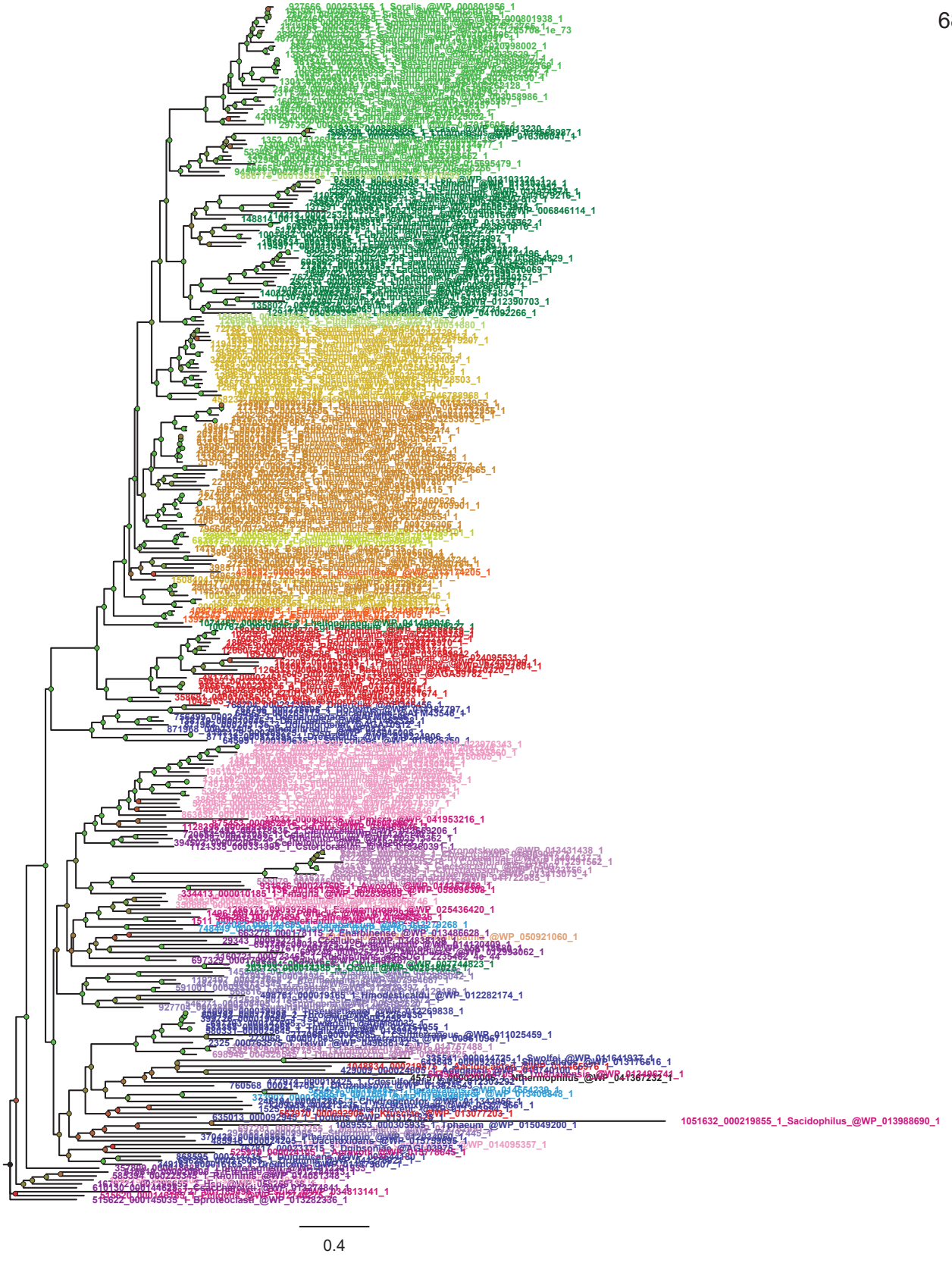

1511_000196455_1_Csticklandii_@WP_013361660_1
642492_000178835_2_Clentocellum_@WP_013657479_1
1286171_000597865_1_Eacidaminophi_@WP_025435247_1
1496_001447175_1_Pdifficile_@WP_003431031_1
293826_000016985_1_Ametalliredig_@WP_012063337_1
350688_000018325_1_Aoremlandii_@WP_012159668_1
84022_001042715_1_Caceticum_@WP_044823716_1
1548_000968375_1_Cscatologenes_@WP_029161965_1
536227_001038625_1_Ccarboxidivor_@WP_007059941_1
768706_000235605_1_Dorientis_@WP_014184408_1
1491_000827955_1_Cbotulinum_@WP_003373126_1
1492_001456065_2_Cbutyricum_@WP_003415276_1
1415775_000789395_1_Cbaratii_@WP_039316147_1
195102_000009685_1_Cperfringens_@WP_003458466_1
1509_000973705_1_Csporogenes_@WP_003493195_1
863638_000191905_1_Cacetobutylic_@WP_010965021_1
573061_000145275_1_Ccellulovoran_@WP_010077175_1
1216932_000577895_1_Cbornimense_@CDM69086_1
340099_000019085_1_Tpseudethanol_@WP_041589977_1
509193_000175295_2_Tbrockii_@WP_041589977_1
399726_000019065_1_Tsp_@WP_038016620_1
697303_000147695_3_Twiegelii_@AEM78479_1
273068_000007085_1_Csubterraneus_@WP_011025341_1
1094508_000307585_1_Tsaccharolyti_@WP_044984061_1
858215_000189775_3_Txylanolyticu_@AEF16887_1
756499_000243155_3_Ddehalogenans_@AFM00867_1
138119_000010045_1_Dhafniense_@WP_005813011_1
871963_000243135_3_Ddichloroelim_@AGA69583_1
871968_000231405_3_Dmetallireduc_@AHF07217_1
646529_000255115_3_Dacidiphilus_@AFM42078_1
748449_000328625_1_Hhalobius_@WP_015327038_1
868595_000214435_1_Dnigrificans_@WP_013809660_1
93061_000013425_1_Saureus_@YP_500016_1
985002_000236925_1_Sargenteus_@WP_000681751_1
1283_000972725_1_Shaemolyticus_@WP_011275728_1
1194526_000332735_1_Swarneri_@WP_002465794_1
1276282_000494875_1_Spasteuri_@WP_023372563_1
72758_001028645_1_Scapitis_@WP_047796298_1
1282_000759555_1_Sepidermidis_@WP_001831213_1
1034809_000270465_1_Slugdunensis_@WP_002478169_1
342451_000010125_1_Ssaprophyticu_@WP_011303065_1
1288_000709415_1_Skylosus_@WP_042362659_1
246432_001432245_1_Sequorum_@WP_002507561_1
396513_000009405_1_Scarnosus_@WP_015900359_1
1284_000816085_1_Shyicus_@WP_039645685_1
985762_001442815_1_Sagnetis_@WP_060551085_1
1295_001188875_1_Sschleiferi_@WP_050332178_1
937773_000185885_1_Spseudintermе_@WP_015729141_1
458233_000010585_1_Mcaseolyticus_@WP_012657034_1
407035_001005905_1_Shalodurans_@WP_046790254_1
403957_000725285_1_Vsp_@WP_040955463_1
637380_000143605_1_Bcereus_@WP_041183900_1
412694_000015065_1_Bthuringiensi_@WP_001089817_1
261591_000742895_1_Banthracis_@WP_001932700_1
1415784_000496285_1_Btoyonensis_@WP_002093806_1
1330043_000831065_1_Bbombysepticu_@WP_001987020_1
315749_000017425_1_Bcytotoxicus_@WP_041810301_1
86662_000775975_1_Bweihenstepha_@WP_002011796_1
1405_000832605_1_Bmycoides_@WP_002011796_1
135735_000972245_2_Bendophyticus_@WP_019391657_1
1006007_000225265_1_Bmegaterium_@WP_014458331_1
272558_000011145_1_Bhalodurans_@WP_010897787_1
398511_000005825_2_Bpseudofirmus_@WP_012958467_1
358681_000010165_1_Bbrevis_@WP_012686122_1
1042163_000219535_3_Blateros porus_@AIG26188_1
1087448_000299435_1_Eantarcticum_@WP_014970622_1
1399115_000496635_1_Esp_@WP_023468652_1
262543_000019905_1_Esibiricum_@WP_041261650_1
444177_000017965_1_Lsphaericus_@WP_012293600_1
28031_000724775_3_Lfusiformis_@WP_036124554_1
1145276_000600105_1_Lvarians_@WP_025219068_1
1367477_000473245_1_Binfantis_@WP_035403076_1
1479_001050115_1_Bsmithii_@WP_048624483_1
1508404_000818095_1_Jmalaysiensis_@WP_039809441_1
1291742_000829395_1_Lhokkaidonens_@WP_041093180_1
1346_000831485_1_Siniae_@WP_031259080_1
218495_000009545_1_Suberis_@WP_015911840_1
873447_000187935_2_Sparauberis_@EGE55105_1
552526_000020765_1_Sequi_@WP_041785285_1
617121_000307185_1_Sdysgalactiae_@WP_003053845_1
160491_000009385_1_Spyogenes_@WP_002983258_1
1311_001026925_1_Sagalactiae_@WP_000723062_1
1304_000785515_1_Ssalivarius_@WP_037602584_1
1308_000971665_1_Sthermophilus_@WP_024704001_1
1437447_000817065_1_Smutans_@WP_002263467_1
990317_000203195_1_Sgallolyticus_@WP_013643397_1
981540_000270165_1_Spasseriani_@WP_003066282_1
1116231_000283635_1_Smacedonicus_@WP_014295209_1
1069533_000246835_1_Sinfantarius_@WP_006532010_1
1076934_000441535_1_Slutetiensis_@WP_020917399_1
388919_000014205_1_Ssanguinis_@YP_001034544_1
T1149665_000262145_1_Sparasanguini_@WP_014713791_1
1302863_000385925_1_Soligofermen t_@WP_015605688_1
862968_000463445_1_Sconstellatus_@WP_006268265_1
1338_001296205_1_Sintermedius_@ALF28227_1
1353243_000478925_1_Sanginosus_@WP_003035568_1
467705_000017005_1_Sgordonii_@WP_011999936_1
171100_000007045_1_Spneumoniae_@NP_359404_1
28037_001281025_1_Smitis_@WP_060627184_1
1054460_000221985_1_Spseudopneumo_@WP_001124785_1
1419814_000688775_1_Ssp_@WP_044020173_1
927666_000253155_1_Soralis_@WP_001124803_1
1307_000993745_1_Ssuis_@WP_011922637_1
297352_000981525_1_Lpisc ium_@WP_047915981_1
1117941_000478255_2_Llactis_@AII12242_1
420890_000269945_1_Lgarvieae_@WP_014024504_1
979982_000219785_1_Lsp_@WP_041762746_1
762051_000092505_1_Lkimchii_@WP_041762746_1
1229758_000300135_1_Lcarnosum_@WP_041773217_1
1107880_000234825_3_Lmesenteroide_@WP_014324542_1
762550_000196855_1_Lgelidum_@WP_041771840_1
349519_000026405_1_Lcitreum_@WP_004904045_1
768486_000271405_2_Ehirae_@WP_010718737_1
53345_001267395_1_Edurans_@WP_005879623_1
1300150_000504125_1_Emundtii_@WP_010736088_1
1352_001412695_1_Efaecium_@WP_002287961_1
565655_000157355_2_Ecasseliflavu_@WP_005225140_1
945021_000283615_1_Thalophilus_@WP_014124342_1
1266845_000493735_1_Cinhibens_@WP_043945417_1
1564691_001483965_1_Csp_@WP_058918217_1
1234679_000317975_2_Cmaltaromatic_@WP_015075475_1
1261557_000317915_1_Efaecalis_@WP_002357976_1
332949_001465115_1_Esilesiacus_@ALS00264_1
60520_001443645_1_Lparaplantaru_@WP_033612139_1
889932_000148815_2_Lplantarum_@WP_003642578_1
1194971_001011095_1_Lsalivarius_@WP_011476226_1
314315_000026065_1_Lsakei_@WP_011374056_1
568704_000026525_1_Lrhamnosus_@WP_005687110_1
219334_000829055_1_Lcasei_@WP_044112912_1
1226298_000829035_1_Lparacasei_@WP_003603167_1
637971_001050435_1_Lkoreensis_@WP_048735069_1
1001583_000359625_1_Lbrevis_@WP_035465547_1
511437_000211375_1_Lbuchneri_@WP_013728744_1
148814_001314945_1_Lkunkeei_@WP_034532643_1
714313_000225325_1_Lsanfrancisce_@WP_014081622_1
701521_000237995_2_Plaussenii_@AEV94742_1
1408206_000496265_1_Ppentosaceus_@WP_023439952_1
1545702_000761135_1_Lsp_@WP_038524237_1
0.3

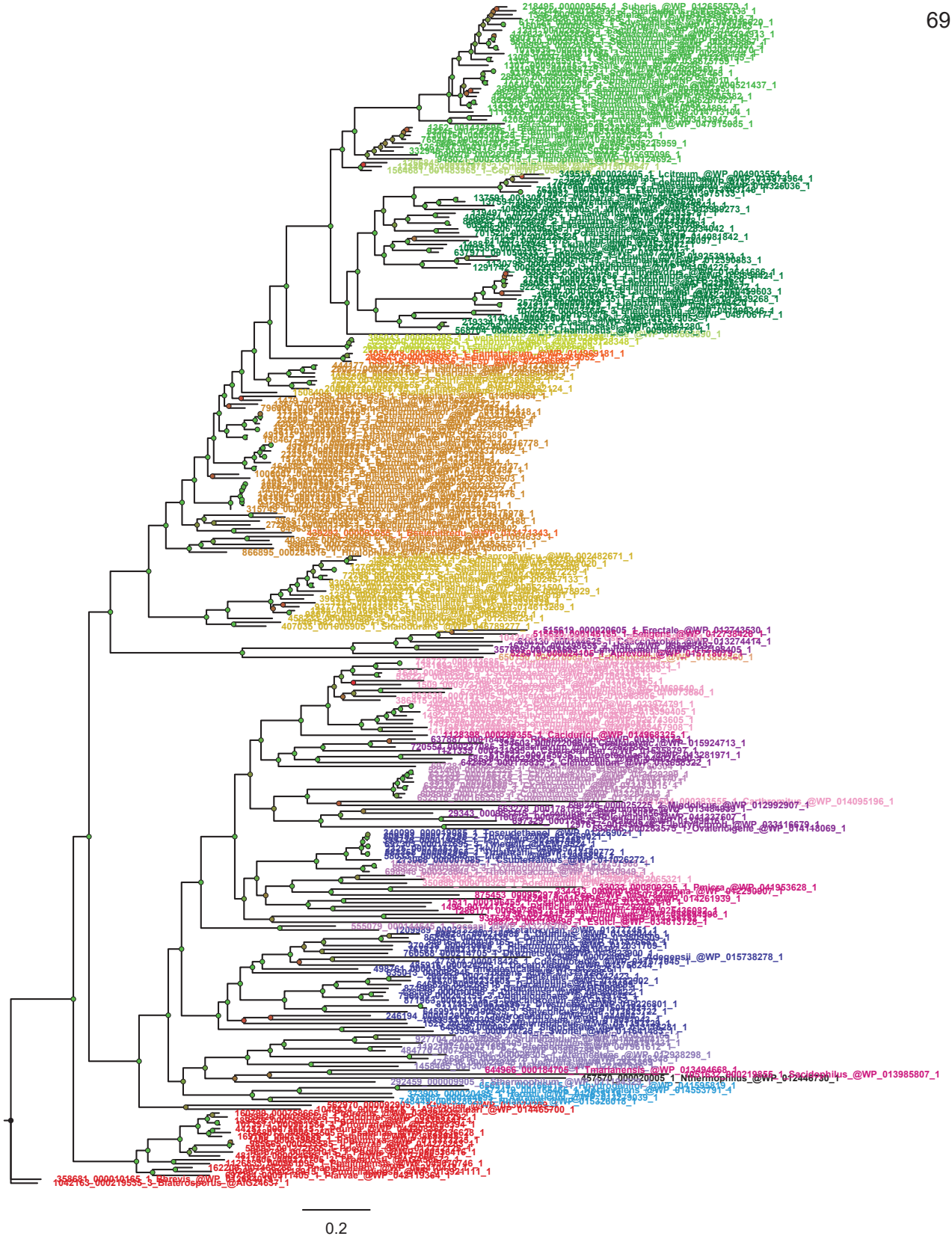

1406_000819665_1_Ppolymyxa_@WP_040102122_1
59893_001272655_1_Ppeoriae_@WP_034094978_1
985665_000235585_1_Pterrae_@WP_043922375_1
44251_000756615_1_Pdurus_@WP_042208799_1
1036673_000218915_1_Pmucilaginosu_@WP_013919382_1
1408_000972685_1_Bpumilus_@WP_046342759_1
1452_000830075_1_Batrophaeus_@WP_003325984_1
1006007_000225265_1_Bmegaterium_@WP_014458823_1
224308_000009045_1_Bsubtilis_@NP_388151_1
1574141_000877815_1_Bsp_@WP_003234769_1
1126211_000262385_1_Bamyloliquefa_@WP_014416829_1
492670_000988345_1_Bvelezensis_@WP_007409031_1
1648923_000876525_1_Bparalichenif_@WP_020450036_1
135735_000972245_2_Bendophyticus_@WP_053102071_1
1073571_000981585_1_Priograndensi_@CQR53330_1
481743_000024685_1_Psp_@WP_012818757_1
1036673_000218915_1_Pmucilaginosu_@WP_013914875_1
515620_000146185_1_Eeligens_@WP_012740681_1
1192197_000271665_2_Pfermentans_@WP_007955565_1
484770_000725345_1_Psp_@WP_051788780_1
484770_000725345_1_Psp_@WP_038675275_1
1192197_000271665_2_Pfermentans_@WP_007959621_1
479436_000024945_1_Vparvula_@WP_012863853_1
1458465_001304715_1_Melsdenii_@WP_054336956_1
546271_000208405_1_Ssputigena_@WP_006194021_1
927704_000284095_1_Sruminantium_@WP_014423507_1
1130798_000248095_3_Lmucosae_@AJT50513_1
1069534_000224985_1_Lruminis_@WP_014072936_1
290402_000016965_1_Cbeijerinckii_@WP_012058609_1
1428454_000506785_2_Cpasteurianum_@WP_023976726_1
931276_000340885_1_Csaccharoperb_@WP_015394249_1
160799_000758665_1_Pborealis_@WP_042215241_1
189425_000758705_1_Pgraminis_@WP_025704763_1
1073571_000981585_1_Priograndensi_@CQR56714_1
0.6

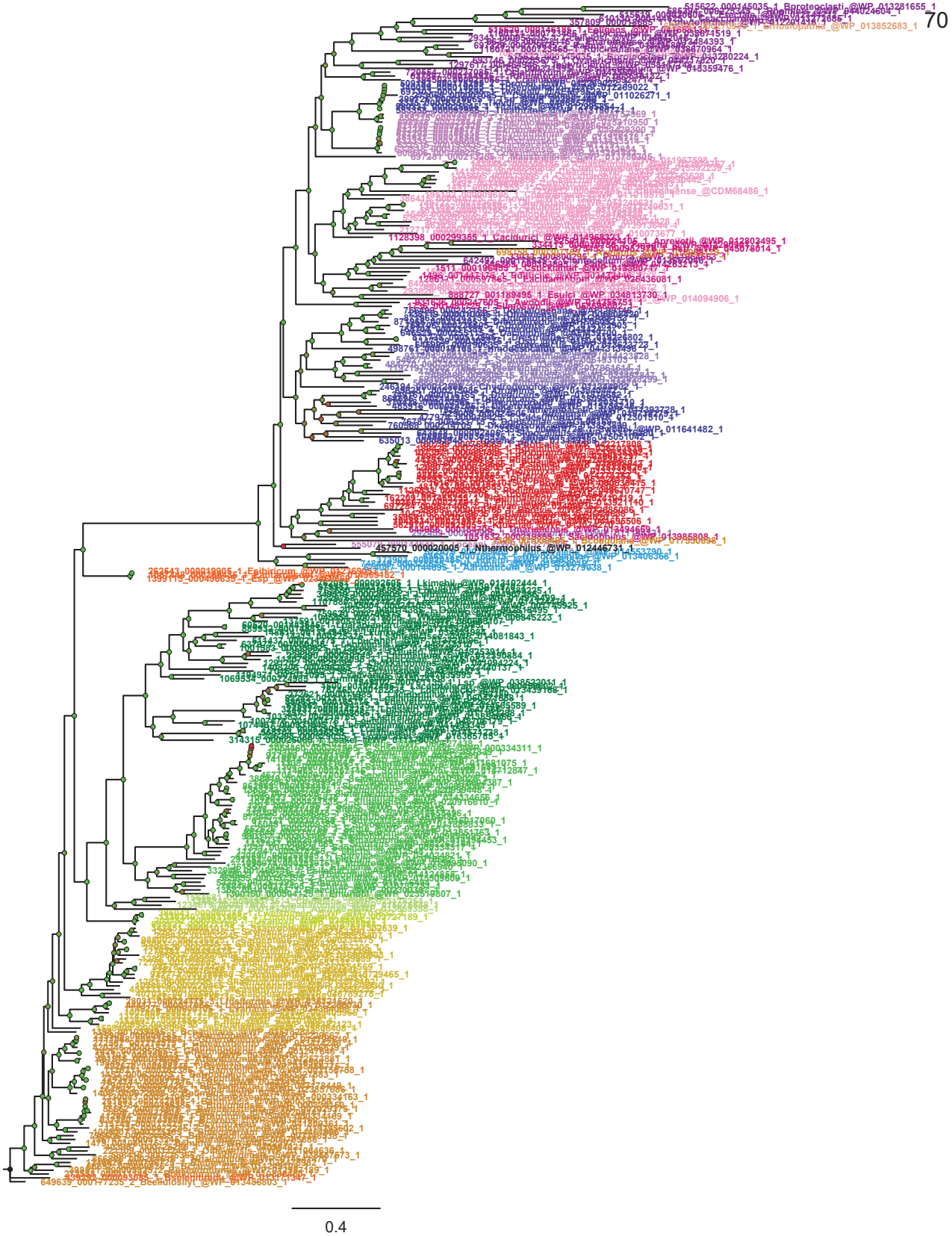

235909_000009785_1_Gkaustophilus_@WP_011231717_1
1111068_000236605_1_Gthermoleovor_@WP_014196214_1
272567_001274575_1_Gstearothermo_@WP_053414679_1
420246_000015745_1_Gthermodenitr_@WP_011887705_1
586416_000725365_1_Taidingensis_@WP_038560950_1
866895_000284515_1_Hhalophilus_@WP_014643543_1
439292_000093085_1_Bselenitiredu_@WP_013173078_1
439292_000093085_1_Bselenitiredu_@WP_013171575_1
649639_000177235_2_Bcellulosilyt_@WP_013488441_1
1246626_000706725_1_Blehensis_@WP_038480318_1
66692_000009825_1_Bclausii_@WP_035204115_1
1461582_000756715_2_Jsp_@CDZ99024_1
1648923_000876525_1_Bparalichenif_@WP_025811110_1
866775_000193205_1_Aurinae_@WP_013669660_1
221109_000011245_1_Oiheyensis_@WP_011066206_1
1526927_000785555_1_Psp_@WP_038704168_1
1374_001465835_1_Pkocurii_@WP_058385028_1
200991_001465795_1_Prifietoensis_@WP_058380549_1
796606_000724485_1_Bmethanolicus_@WP_004435329_1
1002809_000271325_1_Ssilvestris_@WP_014824316_1
1426_001295365_1_Gthermoglucos_@WP_013400493_1
581103_000166075_1_Gsp_@WP_013400493_1
1648923_000876525_1_Bparalichenif_@WP_020451978_1
279010_000008425_1_Blicheniformi_@WP_003183042_1
386043_000060285_1_Lwelshimeri_@WP_011702731_1
1230340_000318055_1_Lmonocytogene_@WP_003723587_1
683837_000027145_1_Lseeligeri_@WP_012986061_1
202752_000763475_1_Livanovii_@WP_014093280_1
0.3

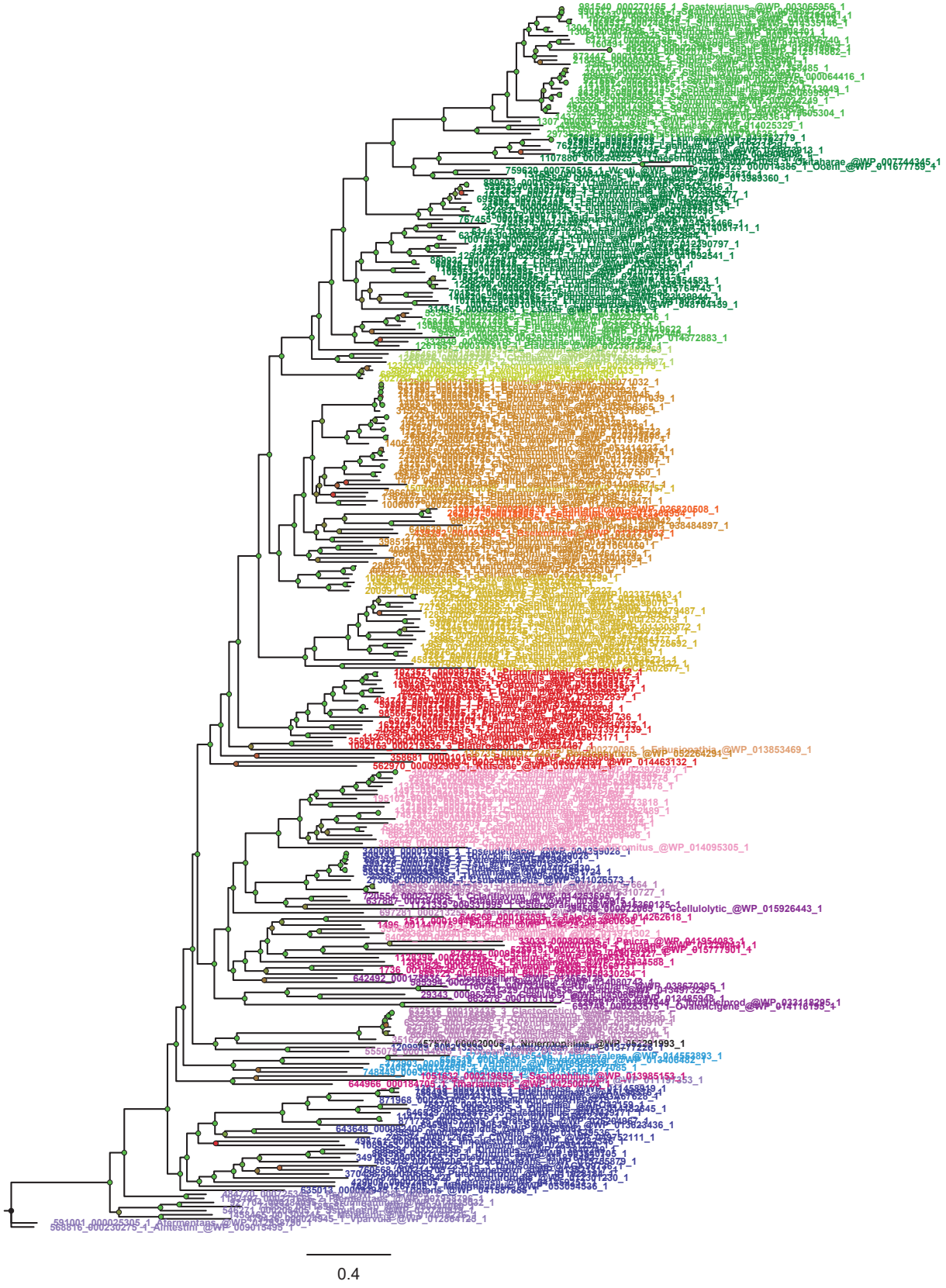

1033837_000214785_1_Lkefiranofaci_@WP_013854727_1
695562_000194115_1_Lamylovorus_@WP_014566131_1
272621_000011985_1_Lacidophilus_@YP_194530_1
880633_000165775_3_Lhelveticus_@AFR22675_1
52242_001314245_1_Lgallinarum_@WP_060471560_1
324831_000014425_1_Lgasseri_@WP_003647445_1
257314_000008065_1_Ljohnsonii_@WP_011162159_1
1358027_000439275_1_Lreuteri_@WP_020843446_1
767455_000182835_1_Ldelbrueckii_@WP_003616833_1
515622_000145035_1_Bproteoclasti_@WP_013282525_1
1345695_000473995_1_Csaccharobuty_@WP_022746907_1
1297617_001454945_1_Ibutyriciprod_@WP_058117088_1
357809_000018685_1_Lphytoferment_@WP_012201415_1
1736_001481725_1_Elimosum_@WP_058696057_1
931626_000247605_1_Awoodii_@WP_014355318_1
1160721_000723465_1_Rbicirculans_@WP_038670676_1
717605_000227705_3_Tcomposti_@AGA59929_1
546269_000163895_2_Falocis_@WP_014262895_1
699246_000025225_2_Mindolicus_@WP_012993877_1
591001_000025305_1_Afermentans_@WP_012939409_1
568816_000230275_1_Aintestini_@WP_009015024_1
1492_001456065_2_Cbutyricum_@WP_002583101_1
1406_000819665_1_Ppolymyxa_@WP_040102556_1
59893_001272655_1_Ppeoriae_@WP_013312613_1
985665_000235585_1_Pterrae_@WP_014278810_1
1616788_001421015_1_Pbovis_@WP_060531252_1
555079_000144645_1_Toceani_@PSDG1_2197338_1e_22
334413_000010185_1_Fmagna_@WP_012290505_1
33033_000800295_1_Pmicra_@WP_041954261_1
585394_000225345_1_Rhominis_@WP_014079473_1
693746_000283575_1_Ovalericigene_@WP_014117204_1
610130_000144625_1_Csaccharolyti_@WP_013271179_1
888727_001189495_1_Esulci_@WP_034813025_1
212717_000007625_1_Ctetani_@WP_035109300_1
457570_000020005_1_Nthermophilus_@WP_012449043_1
515620_000146185_1_Eeligens_@WP_012740626_1
712538_001189555_1_Ssp_@WP_050341716_1
1045004_000241055_1_Okitaharae_@WP_007745111_1
203123_000014385_1_Ooeni_@WP_002817738_1
759620_000750515_1_Wceti_@WP_009765153_1
137591_001308145_1_Wcibaria_@WP_010375042_1
1045854_000219805_1_Wkoreensis_@WP_013989551_1
574087_000144695_1_Aarabaticum_@WP_013277508_1
1192197_000271665_2_Pfermentans_@WP_007957120_1
1126833_000961095_1_Pbeijingensis_@WP_045670966_1
1574141_000877815_1_Bsp_@WP_043940150_1
224308_000009045_1_Bsubtilis_@NP_390239_1
1126211_000262385_1_Bamyloliquefa_@WP_003153321_1
135735_000972245_2_Bendophyticus_@WP_019394956_1
586416_000725365_1_Taidingensis_@WP_038563836_1
1452_000830075_1_Batrophaeus_@WP_003325559_1
1006007_000225265_1_Bmegaterium_@WP_049794016_1
1408_000972685_1_Bpumilus_@WP_017358664_1
279010_000008425_1_Blicheniformi_@WP_003186405_1
1648923_000876525_1_Bparalichenif_@WP_020453442_1
412694_000015065_1_Bthuringiensi_@WP_000722285_1
637380_000143605_1_Bcereus_@WP_000722290_1
1330043_000831065_1_Bbombysepticu_@WP_044584609_1
261591_000742895_1_Banthracis_@WP_000722287_1
1398_001039495_1_Bcoagulans_@WP_035183786_1
491915_000019045_1_Aflavithermus_@WP_041638709_1
198467_001187595_1_Agonensis_@WP_035067418_1
0.5

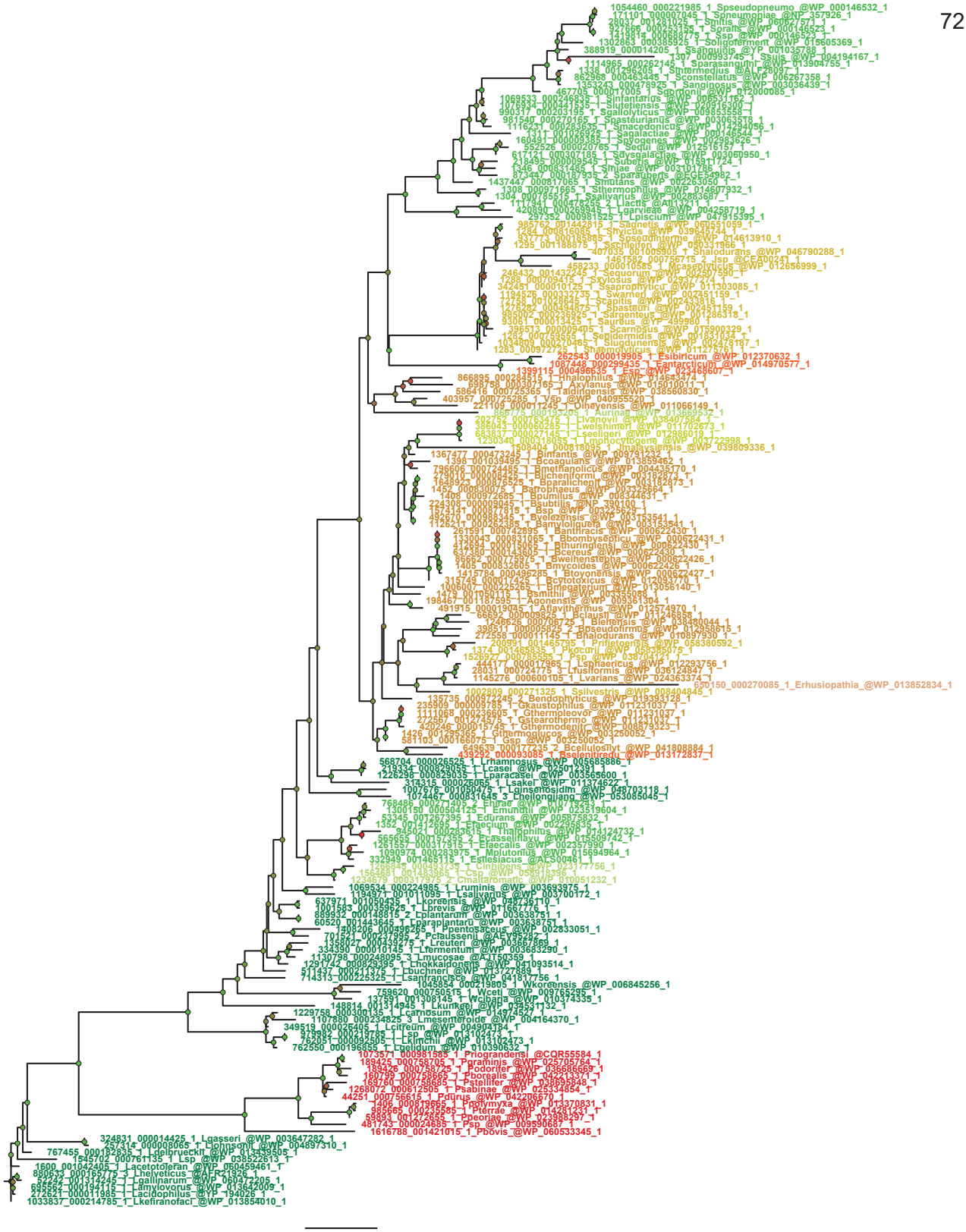

0.3

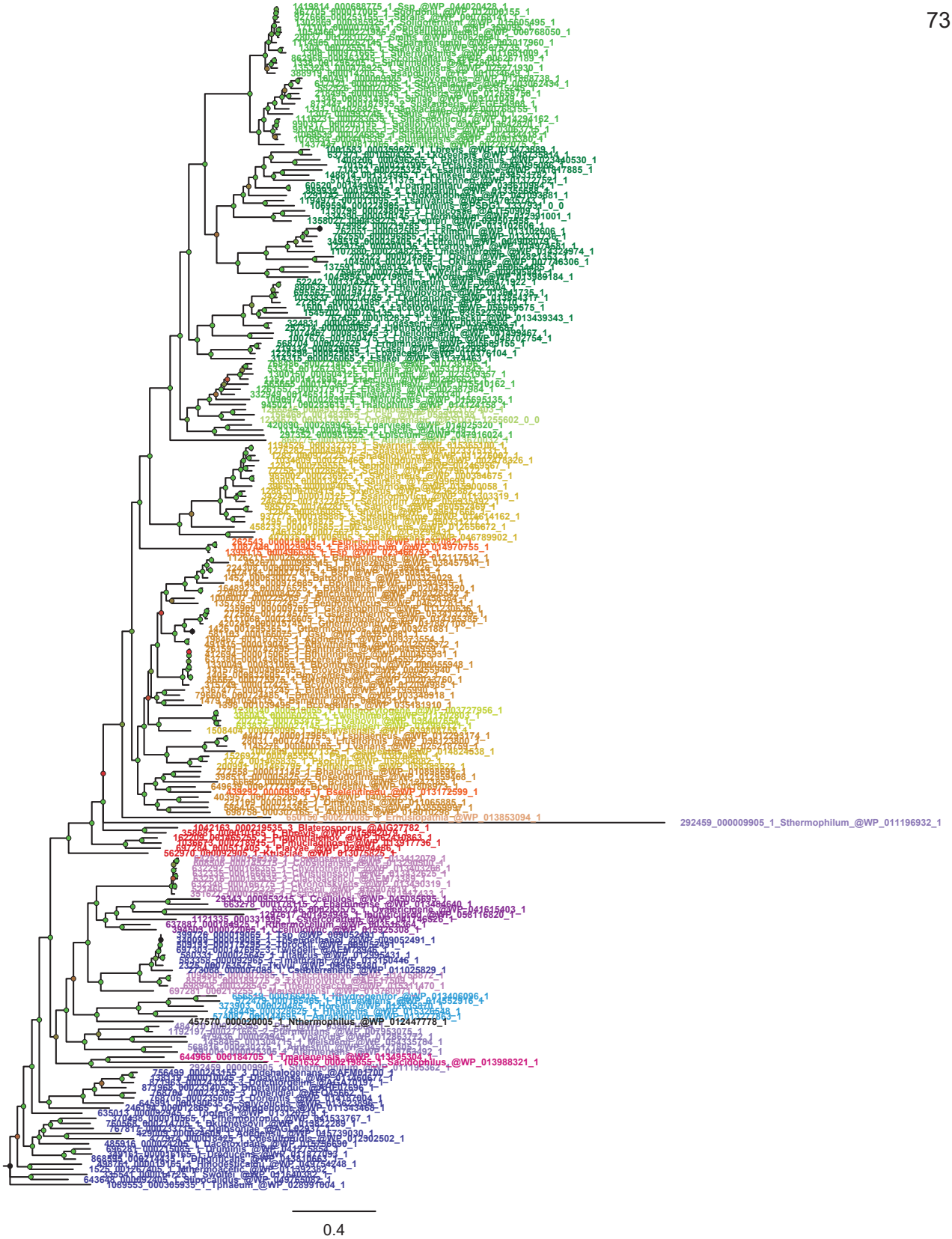

1051632_000219855_1_Sacidophilus_@WP_013988690_1
0.4

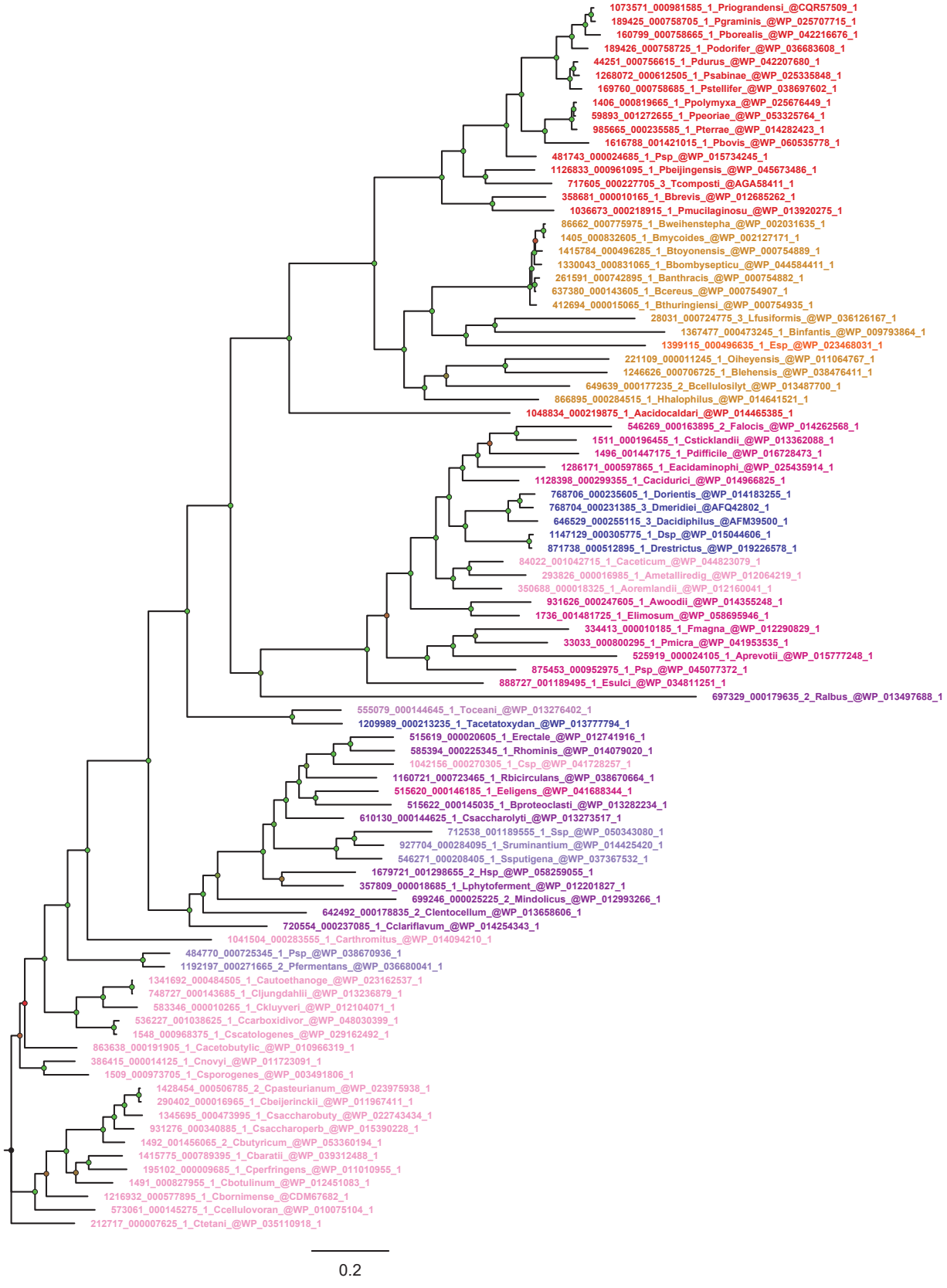

0.2

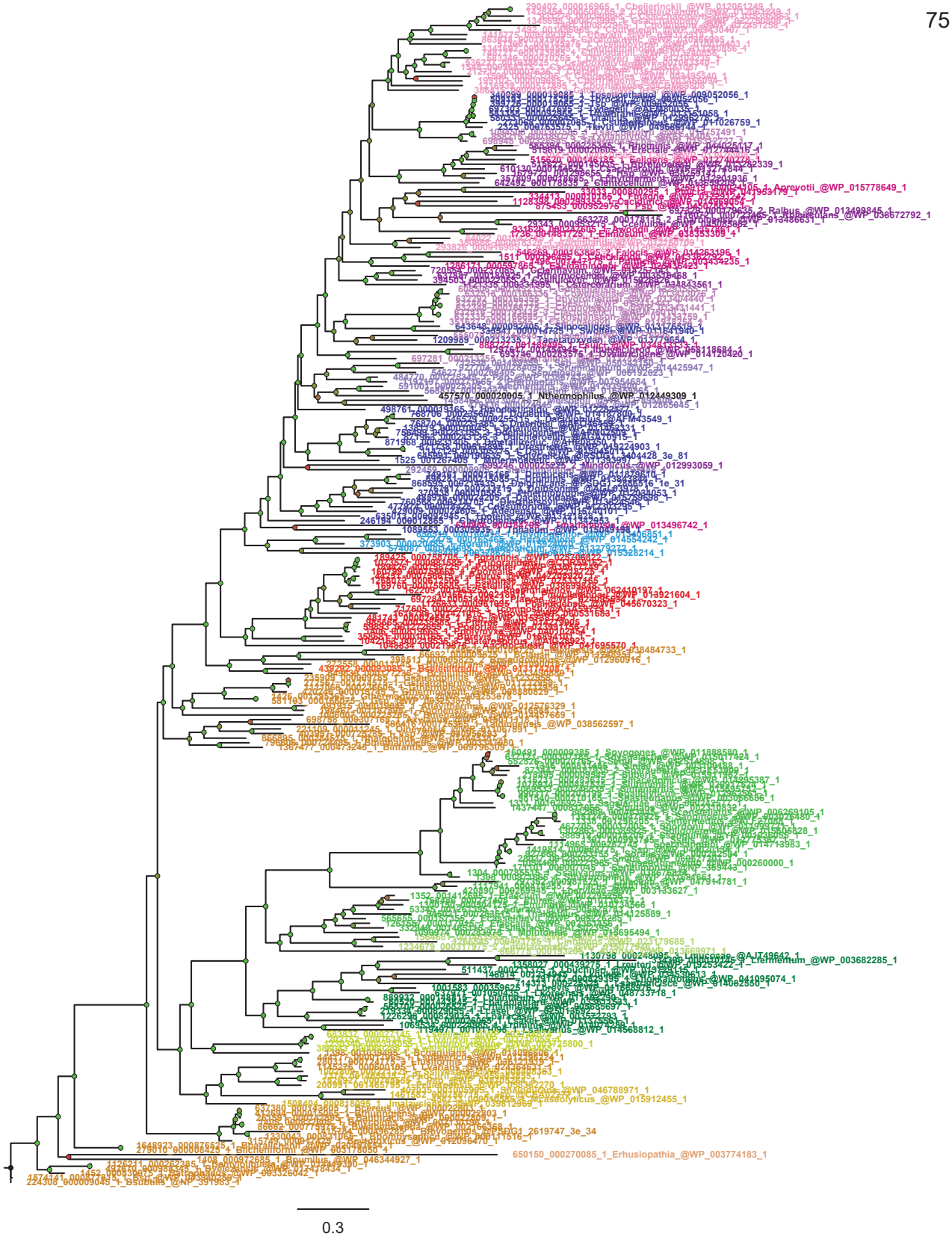

0.4

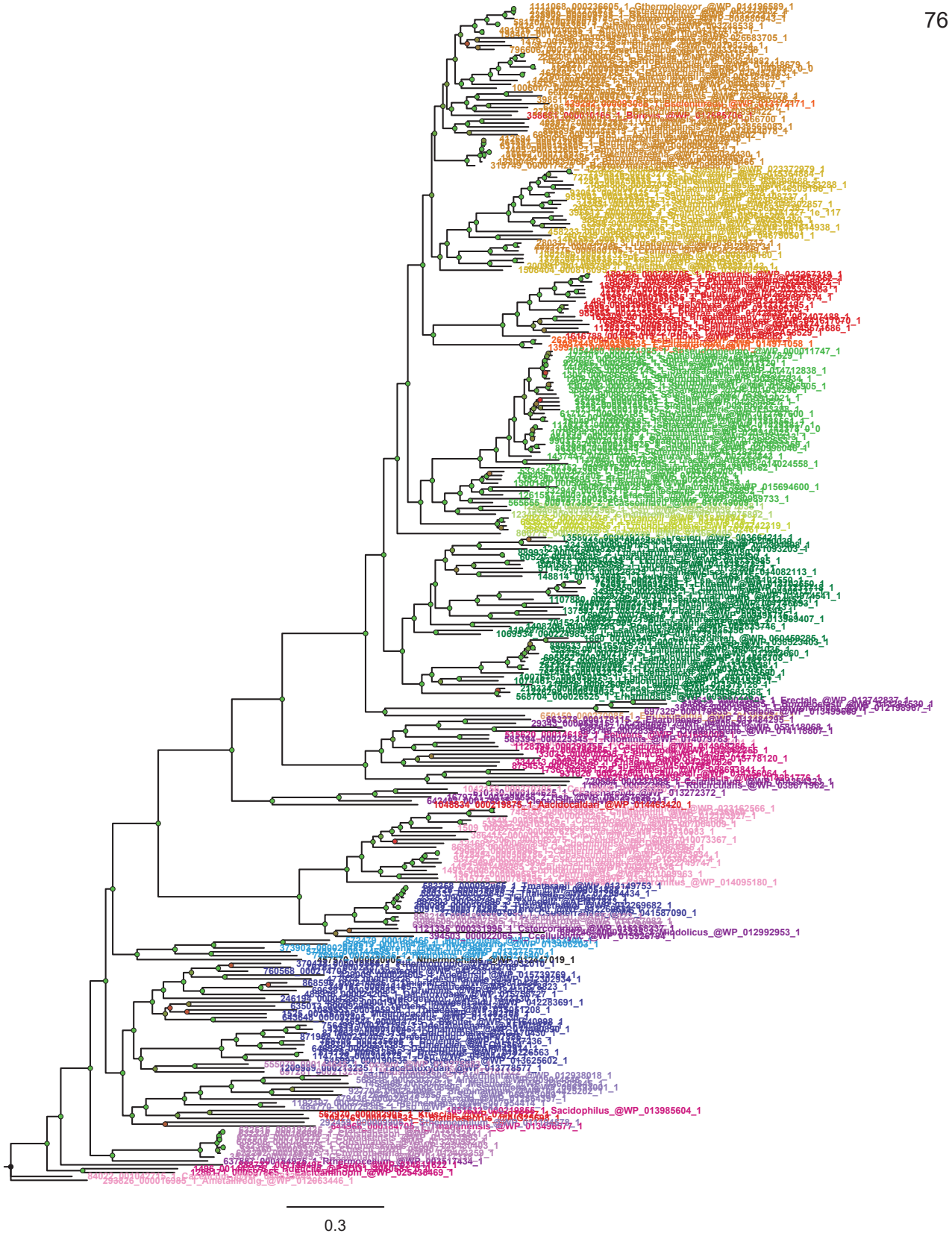

0.4

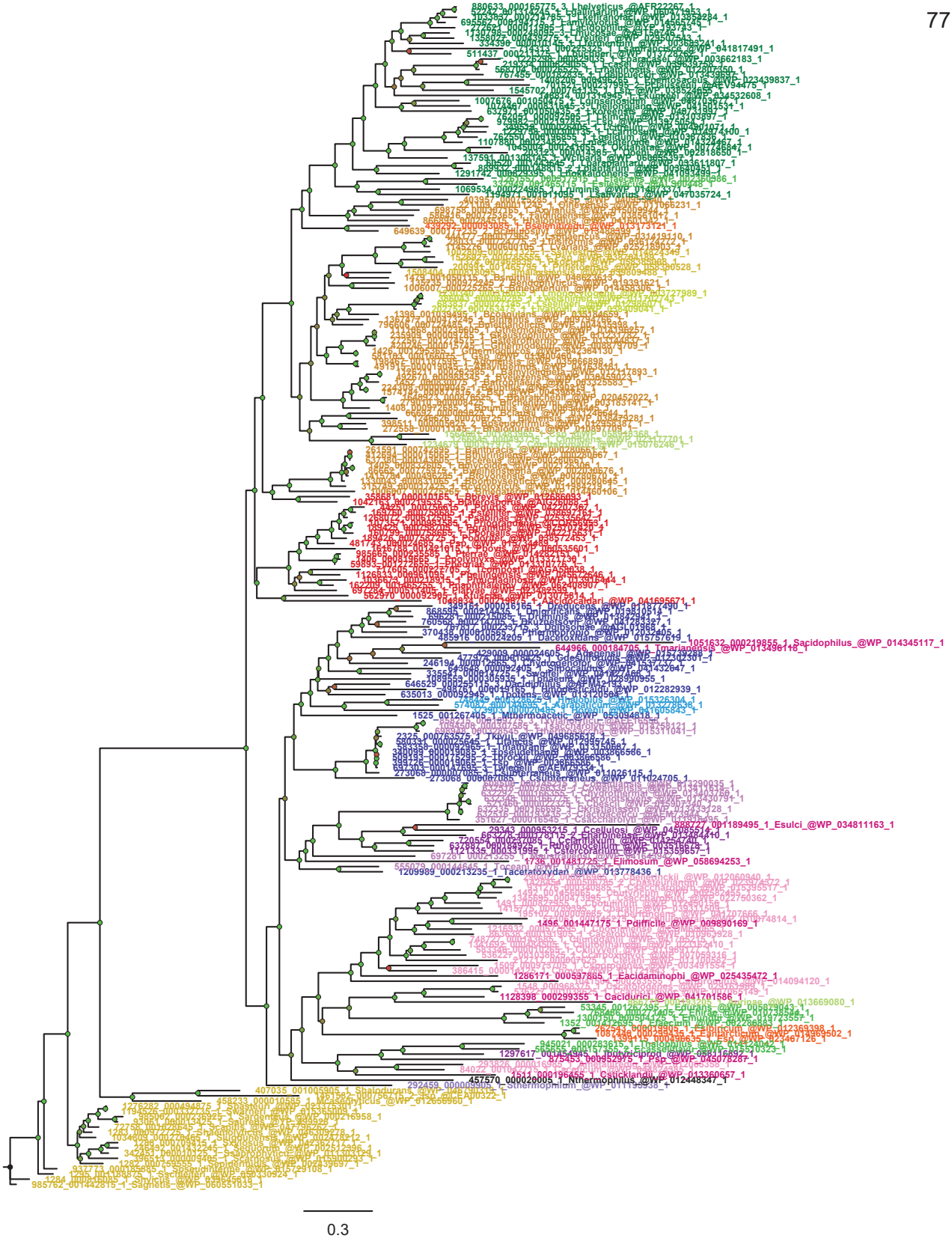

0.4

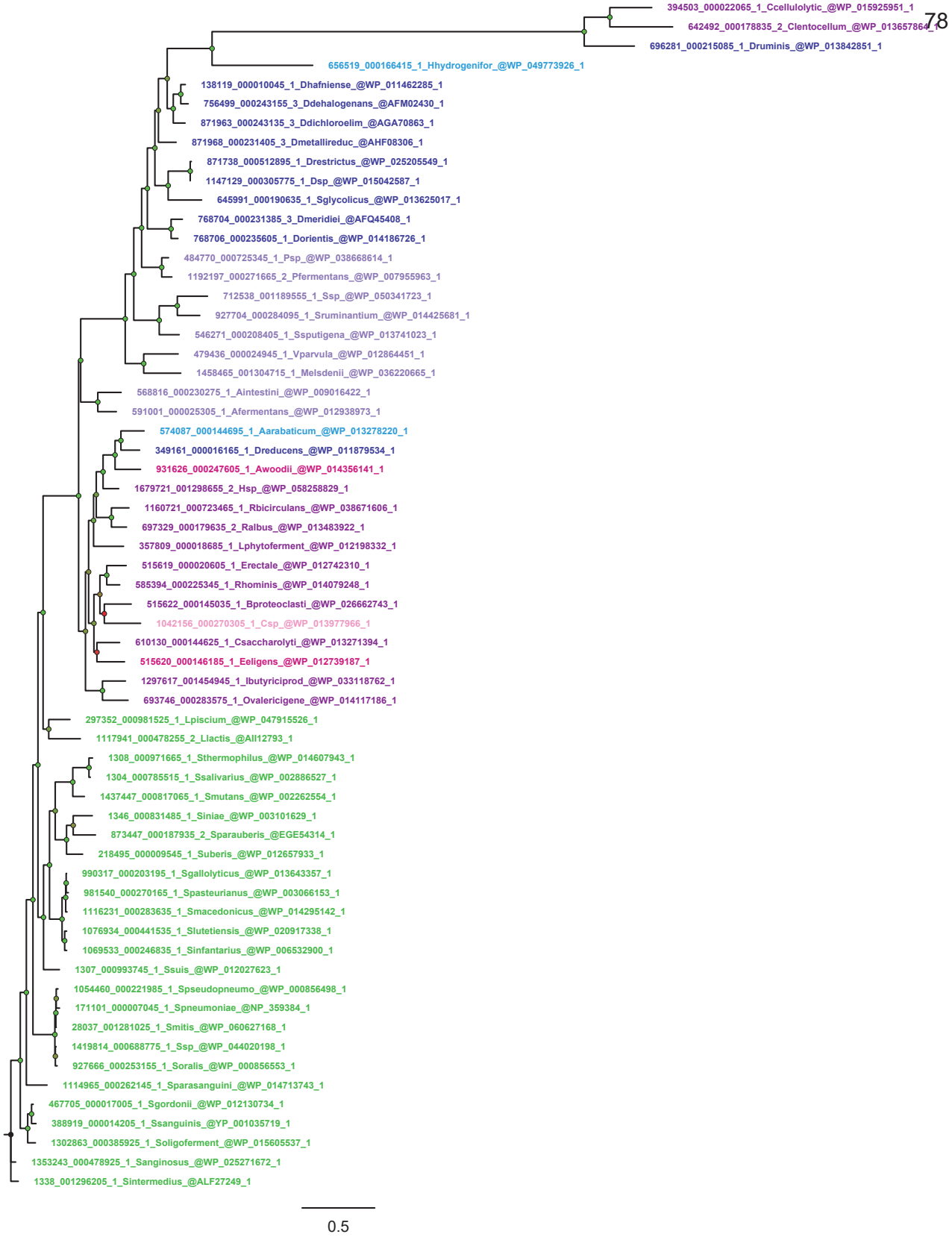

292459_000009905_1_Sthermophilum_@WP_011196932_1
0.4

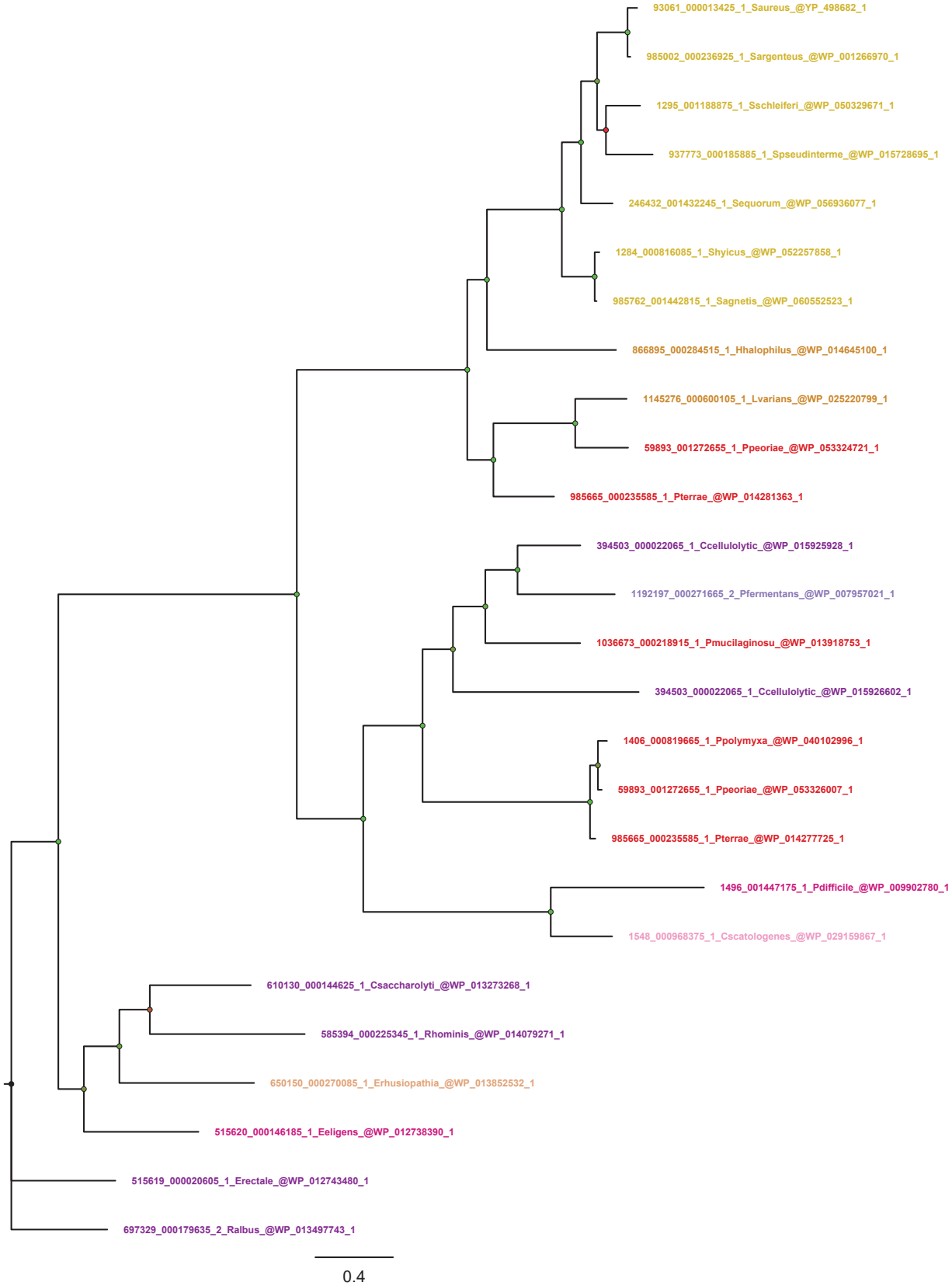

1073571_000981585_1_Priograndensi_@CQR57509_1
189425_000758705_1_Pgraminis_@WP_025707715_1
160799_000758665_1_Pborealis_@WP_042216676_1
189426_000758725_1_Podorifer_@WP_036683608_1
44251_000756615_1_Pdurus_@WP_042207680_1
1268072_000612505_1_Psabinae_@WP_025335848_1
169760_000758685_1_Pstellifer_@WP_038697602_1
1406_000819665_1_Ppolymyxa_@WP_025676449_1
59893_001272655_1_Ppeoriae_@WP_053325764_1
985665_000235585_1_Pterrae_@WP_014282423_1
1616788_001421015_1_Pbovis_@WP_060535778_1
481743_000024685_1_Psp_@WP_015734245_1
1126833_000961095_1_Pbeijingensis_@WP_045673486_1
717605_000227705_3_Tcomposti_@AGA58411_1
358681_000010165_1_Bbrevis_@WP_012685262_1
1036673_000218915_1_Pmucilaginosu_@WP_013920275_1
86662_000775975_1_Bweihenstepha_@WP_002031635_1
1405_000832605_1_Bmycoides_@WP_002127171_1
1415784_000496285_1_Btoyonensis_@WP_000754889_1
1330043_000831065_1_Bbombysepticu_@WP_044584411_1
261591_000742895_1_Banthracis_@WP_000754882_1
637380_000143605_1_Bcereus_@WP_000754907_1
412694_000015065_1_Bthuringiensi_@WP_000754935_1
28031_000724775_3_Lfusiformis_@WP_036126167_1
1367477_000473245_1_Binfantis_@WP_009793864_1
1399115_000496635_1_Esp_@WP_023468031_1
221109_000011245_1_Oiheyensis_@WP_011064767_1
1246626_000706725_1_Blehensis_@WP_038476411_1
649639_000177235_2_Bcellulosilyt_@WP_013487700_1
866895_000284515_1_Hhalophilus_@WP_014641521_1
1048834_000219875_1_Aacidocaldari_@WP_014465385_1
546269_000163895_2_Falocis_@WP_014262568_1
1511_000196455_1_Csticklandii_@WP_013362088_1
1496_001447175_1_Pdifficile_@WP_016728473_1
1286171_000597865_1_Eacidaminophi_@WP_025435914_1
1128398_000299355_1_Cacidurici_@WP_014966825_1
768706_000235605_1_Dorientis_@WP_014183255_1
768704_000231385_3_Dmeridiei_@AFQ42802_1
646529_000255115_3_Dacidiphilus_@AFM39500_1
1147129_000305775_1_Dsp_@WP_015044606_1
871738_000512895_1_Drestrictus_@WP_019226578_1
84022_001042715_1_Caceticum_@WP_044823079_1
293826_000016985_1_Ametalliredig_@WP_012064219_1
350688_000018325_1_Aoremlandii_@WP_012160041_1
931626_000247605_1_Awoodii_@WP_014355248_1
1736_001481725_1_Elimosum_@WP_058695946_1
334413_000010185_1_Fmagna_@WP_012290829_1
33033_000800295_1_Pmicra_@WP_041953535_1
525919_000024105_1_Aprevotii_@WP_015777248_1
875453_000952975_1_Psp_@WP_045077372_1
888727_001189495_1_Esulci_@WP_034811251_1
697329_000179635_2_Ralbus_@WP_013497688_1
555079_000144645_1_Toceani_@WP_013276402_1
1209989_000213235_1_Tacetatoxydan_@WP_013777794_1
515619_000020605_1_Erectale_@WP_012741916_1
585394_000225345_1_Rhominis_@WP_014079020_1
1042156_000270305_1_Csp_@WP_041728257_1
1160721_000723465_1_Rbicirculans_@WP_038670664_1
515620_000146185_1_Eeligens_@WP_041688344_1
515622_000145035_1_Bproteoclasti_@WP_013282234_1
610130_000144625_1_Csaccharolyti_@WP_013273517_1
712538_001189555_1_Ssp_@WP_050343080_1
927704_000284095_1_Sruminantium_@WP_014425420_1
546271_000208405_1_Ssputigena_@WP_037367532_1
1679721_001298655_2_Hsp_@WP_058259055_1
357809_000018685_1_Lphytoferment_@WP_012201827_1
699246_000025225_2_Mindolicus_@WP_012993266_1
642492_000178835_2_Clentocellum_@WP_013658606_1
720554_000237085_1_Cclariflavum_@WP_014254343_1
1041504_000283555_1_Carthromitus_@WP_014094210_1
484770_000725345_1_Psp_@WP_038670936_1
1192197_000271665_2_Pfermentans_@WP_036680041_1
1341692_000484505_1_Cautoethanoge_@WP_023162537_1
748727_000143685_1_Cljungdahlii_@WP_013236879_1
583346_000010265_1_Ckluyveri_@WP_012104071_1
536227_001038625_1_Ccarboxidivor_@WP_048030399_1
1548_000968375_1_Cscatologenes_@WP_029162492_1
863638_000191905_1_Cacetobutylic_@WP_010966319_1
386415_000014125_1_Cnovyi_@WP_011723091_1
1509_000973705_1_Csporogenes_@WP_003491806_1
1428454_000506785_2_Cpasteurianum_@WP_023975938_1
290402_000016965_1_Cbeijerinckii_@WP_011967411_1
1345695_000473995_1_Csaccharobuty_@WP_022743434_1
931276_000340885_1_Csaccharoperb_@WP_015390228_1
1492_001456065_2_Cbutyricum_@WP_053360194_1
1415775_000789395_1_Cbaratii_@WP_039312488_1
195102_000009685_1_Cperfringens_@WP_011010955_1
1491_000827955_1_Cbotulinum_@WP_012451083_1
1216932_000577895_1_Cbornimense_@CDM67682_1
573061_000145275_1_Ccellulovoran_@WP_010075104_1
212717_000007625_1_Ctetani_@WP_035110918_1
0.2

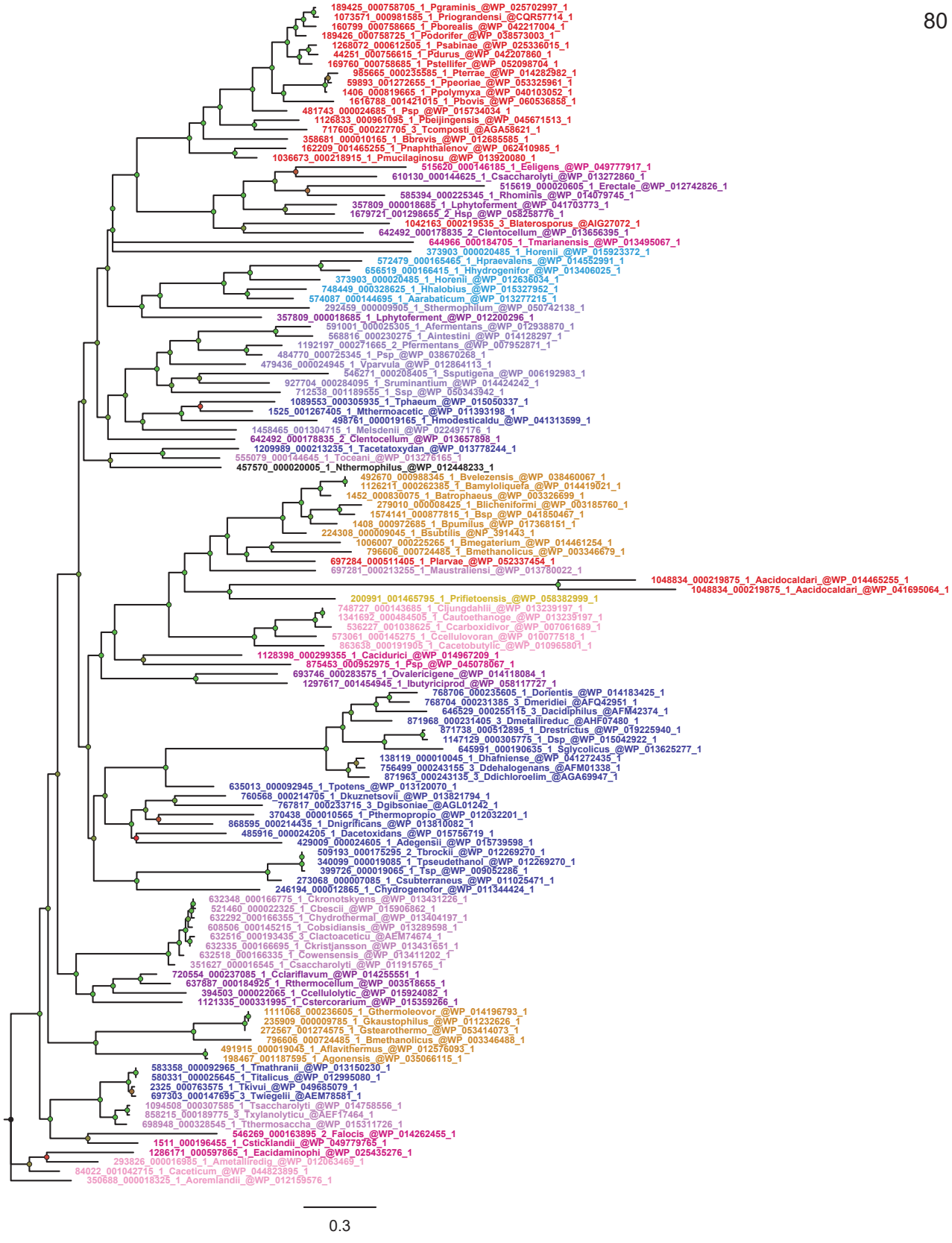

457570_000020005_1_Nthermophilus_@WP_012449309_1
650150_000270085_1_Erhusiopathia_@WP_003774183_1
0.3


0.3

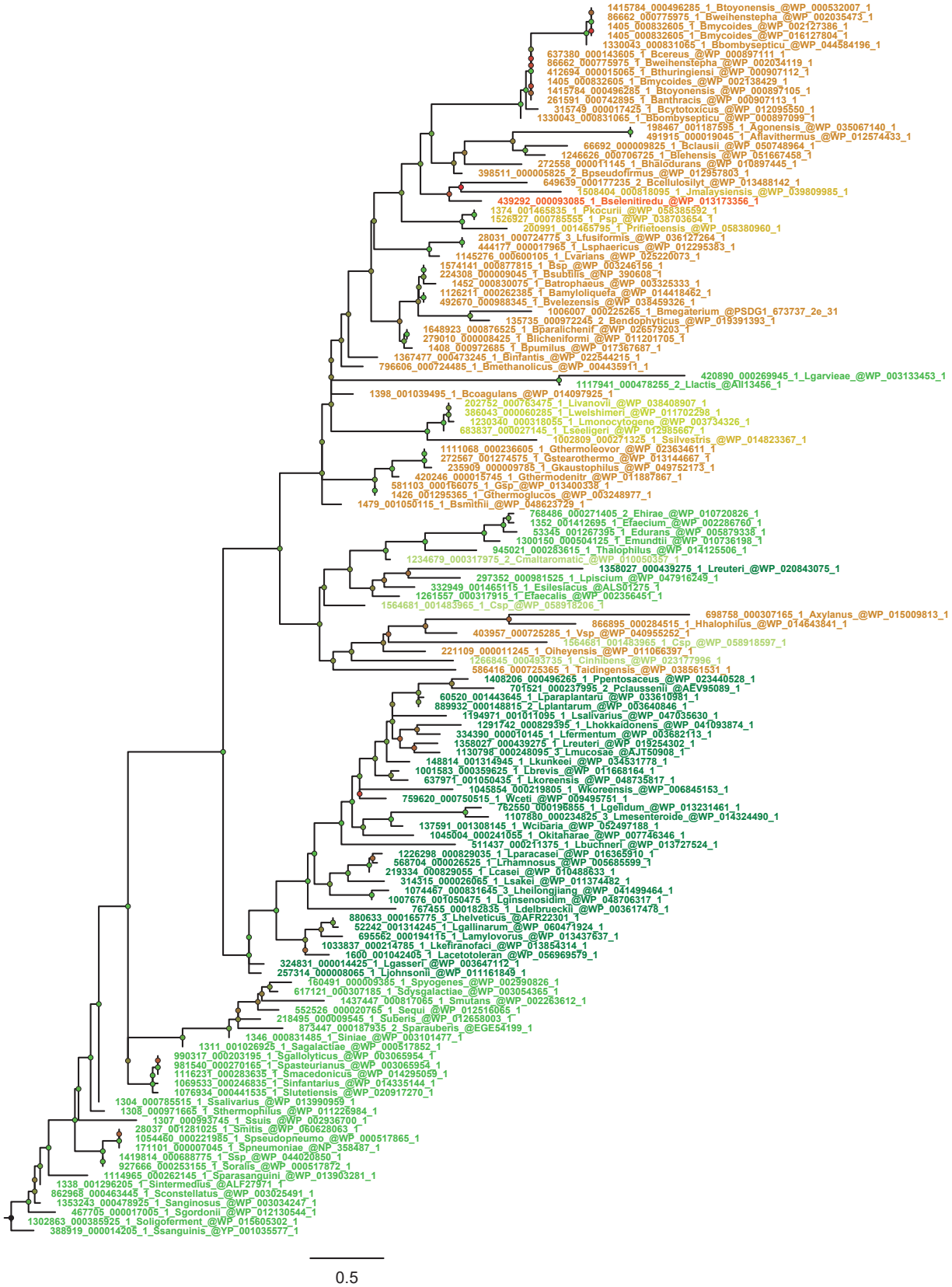

0.3

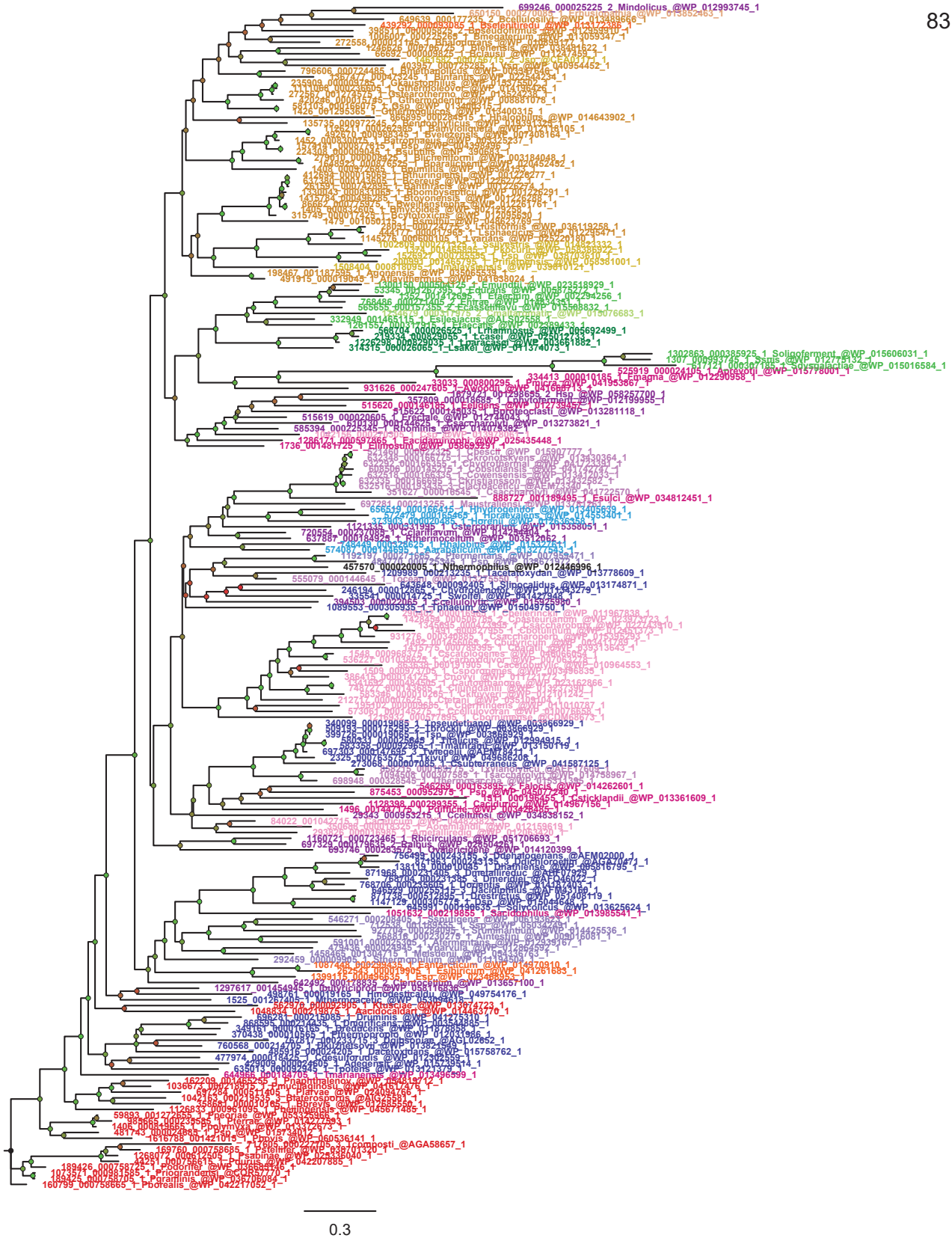

394503_000022065_1_Ccellulolytic_@WP_015925951_1
642492_000178835_2_Clentocellum_@WP_013657864_1
696281_000215085_1_Druminis_@WP_013842851_1
656519_000166415_1_Hhydrogenifor_@WP_049773926_1
138119_000010045_1_Dhafniense_@WP_011462285_1
756499_000243155_3_Ddehalogenans_@AFM02430_1
871963_000243135_3_Ddichloroelim_@AGA70863_1
871968_000231405_3_Dmetallireduc_@AHF08306_1
871738_000512895_1_Drestrictus_@WP_025205549_1
1147129_000305775_1_Dsp_@WP_015042587_1
645991_000190635_1_Sglycolicus_@WP_013625017_1
768704_000231385_3_Dmeridiei_@AFQ45408_1
768706_000235605_1_Dorientis_@WP_014186726_1
484770_000725345_1_Psp_@WP_038668614_1
1192197_000271665_2_Pfermentans_@WP_007955963_1
712538_001189555_1_Ssp_@WP_050341723_1
927704_000284095_1_Sruminantium_@WP_014425681_1
546271_000208405_1_Ssputigena_@WP_013741023_1
479436_000024945_1_Vparvula_@WP_012864451_1
1458465_001304715_1_Melsdenii_@WP_036220665_1
568816_000230275_1_Aintestini_@WP_009016422_1
591001_000025305_1_Afermentans_@WP_012938973_1
574087_000144695_1_Aarabaticum_@WP_013278220_1
349161_000016165_1_Dreducens_@WP_011879534_1
931626_000247605_1_Awoodii_@WP_014356141_1
1679721_001298655_2_Hsp_@WP_058258829_1
1160721_000723465_1_Rbicirculans_@WP_038671606_1
697329_000179635_2_Ralbus_@WP_013483922_1
357809_000018685_1_Lphytoferment_@WP_012198332_1
515619_000020605_1_Erectale_@WP_012742310_1
585394_000225345_1_Rhominis_@WP_014079248_1
515622_000145035_1_Bproteoclasti_@WP_026662743_1
1042156_000270305_1_Csp_@WP_013977966_1
610130_000144625_1_Csaccharolyti_@WP_013271394_1
515620_000146185_1_Eeligens_@WP_012739187_1
1297617_001454945_1_Ibutyriciprod_@WP_033118762_1
693746_000283575_1_Ovalericigene_@WP_014117186_1
297352_000981525_1_Lpiscium_@WP_047915526_1
1117941_000478255_2_Llactis_@AII12793_1
1308_000971665_1_Sthermophilus_@WP_014607943_1
1304_000785515_1_Ssalivarius_@WP_002886527_1
1437447_000817065_1_Smutans_@WP_002262554_1
1346_000831485_1_Siniae_@WP_003101629_1
873447_000187935_2_Sparauberis_@EGE54314_1
218495_000009545_1_Suberis_@WP_012657933_1
990317_000203195_1_Sgallolyticus_@WP_013643357_1
981540_000270165_1_Spasteurianus_@WP_003066153_1
1116231_000283635_1_Smacedonicus_@WP_014295142_1
1076934_000441535_1_Slutetiensis_@WP_020917338_1
1069533_000246835_1_Sinfantarius_@WP_006532900_1
1307_000993745_1_Ssuis_@WP_012027623_1
1054460_000221985_1_Spseudopneumo_@WP_000856498_1
171101_000007045_1_Spneumoniae_@NP_359384_1
28037_001281025_1_Smitis_@WP_060627168_1
1419814_000688775_1_Ssp_@WP_044020198_1
927666_000253155_1_Soralis_@WP_000856553_1
1114965_000262145_1_Sparasanguini_@WP_014713743_1
467705_000017005_1_Sgordonii_@WP_012130734_1
388919_000014205_1_Ssanguinis_@YP_001035719_1
1302863_000385925_1_Soligoferment_@WP_015605537_1
1353243_000478925_1_Sanginosus_@WP_025271672_1
1338_001296205_1_Sintermedius_@ALF27249_1
0.5

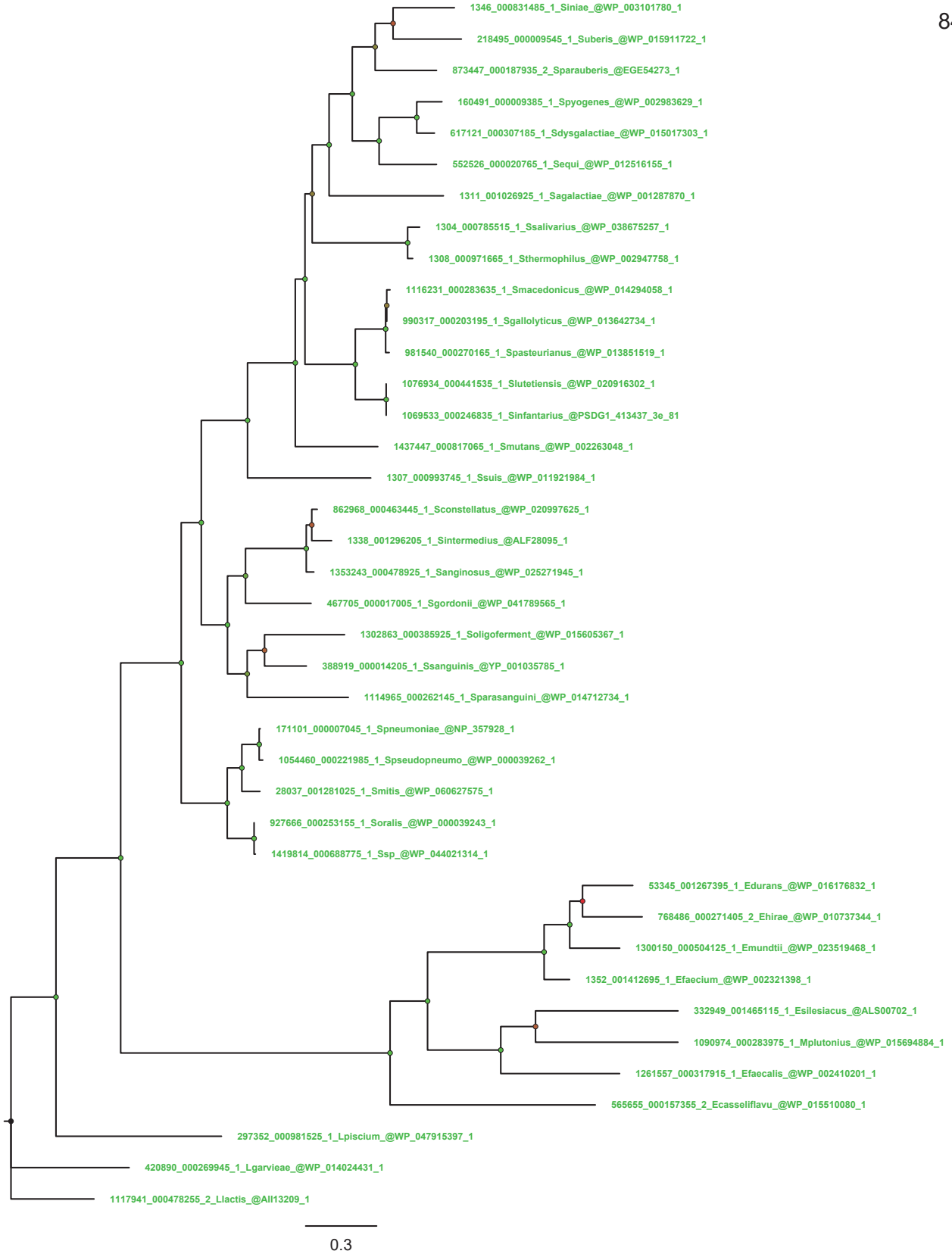

93061_000013425_1_Saureus_@YP_498682_1
985002_000236925_1_Sargenteus_@WP_001266970_1
1295_001188875_1_Sschleiferi_@WP_050329671_1
937773_000185885_1_Spseudinterme_@WP_015728695_1
246432_001432245_1_Sequorum_@WP_056936077_1
1284_000816085_1_Shyicus_@WP_052257858_1
985762_001442815_1_Sagnetis_@WP_060552523_1
866895_000284515_1_Hhalophilus_@WP_014645100_1
1145276_000600105_1_Lvarians_@WP_025220799_1
59893_001272655_1_Ppeoriae_@WP_053324721_1
985665_000235585_1_Pterrae_@WP_014281363_1
394503_000022065_1_Ccellulolytic_@WP_015925928_1
1192197_000271665_2_Pfermentans_@WP_007957021_1
1036673_000218915_1_Pmucilaginosu_@WP_013918753_1
394503_000022065_1_Ccellulolytic_@WP_015926602_1
1406_000819665_1_Ppolymyxa_@WP_040102996_1
59893_001272655_1_Ppeoriae_@WP_053326007_1
985665_000235585_1_Pterrae_@WP_014277725_1
1496_001447175_1_Pdifficile_@WP_009902780_1
1548_000968375_1_Cscatologenes_@WP_029159867_1
610130_000144625_1_Csaccharolyti_@WP_013273268_1
585394_000225345_1_Rhominis_@WP_014079271_1
650150_000270085_1_Erhusiopathia_@WP_013852532_1
515620_000146185_1_Eeligens_@WP_012738390_1
515619_000020605_1_Erectale_@WP_012743480_1
697329_000179635_2_Ralbus_@WP_013497743_1
0.4

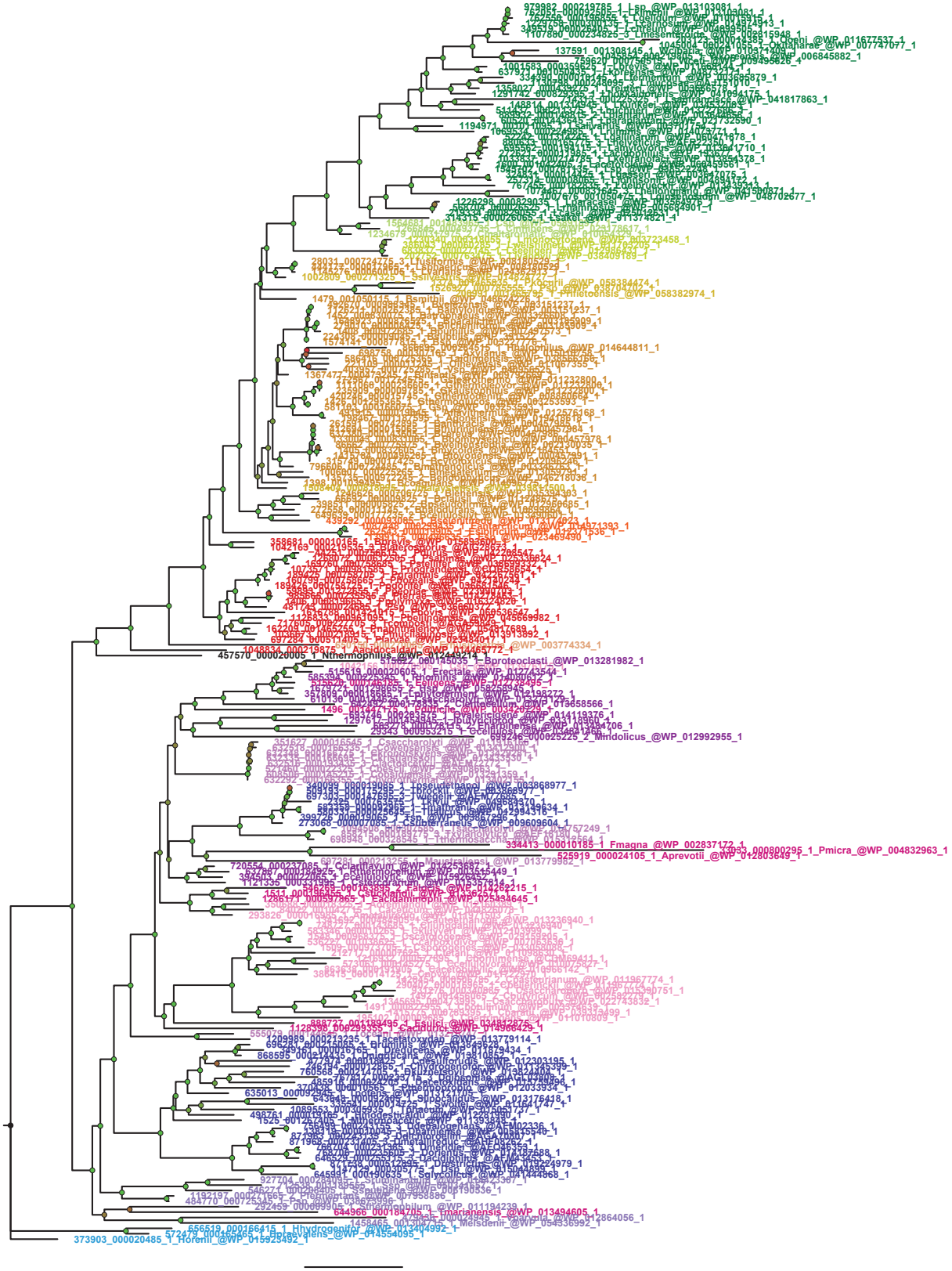

189425_000758705_1_Pgraminis_@WP_025702997_1
1073571_000981585_1_Priograndensi_@CQR57714_1
160799_000758665_1_Pborealis_@WP_042217004_1
189426_000758725_1_Podorifer_@WP_038573003_1
1268072_000612505_1_Psabinae_@WP_025336015_1
44251_000756615_1_Pdurus_@WP_042207860_1
169760_000758685_1_Pstellifer_@WP_052098704_1
985665_000235585_1_Pterrae_@WP_014282982_1
59893_001272655_1_Ppeoriae_@WP_053325961_1
1406_000819665_1_Ppolymyxa_@WP_040103052_1
1616788_001421015_1_Pbovis_@WP_060536858_1
481743_000024685_1_Psp_@WP_015734034_1
1126833_000961095_1_Pbeijingensis_@WP_045671513_1
717605_000227705_3_Tcomposti_@AGA58621_1
358681_000010165_1_Bbrevis_@WP_012685585_1
162209_001465255_1_Pnaphthalenov_@WP_062410985_1
1036673_000218915_1_Pmucilaginosu_@WP_013920080_1
515620_000146185_1_Eeligens_@WP_049777917_1
610130_000144625_1_Csaccharolyti_@WP_013272860_1
515619_000020605_1_Erectale_@WP_012742826_1
585394_000225345_1_Rhominis_@WP_014079745_1
357809_000018685_1_Lphytoferment_@WP_041703773_1
1679721_001298655_2_Hsp_@WP_058258776_1
1042163_000219535_3_Blaterosporus_@AIG27072_1
642492_000178835_2_Clentocellum_@WP_013656395_1
644966_000184705_1_Tmarianensis_@WP_013495067_1
373903_000020485_1_Horenii_@WP_015923372_1
572479_000165465_1_Hpraevalens_@WP_014552991_1
656519_000166415_1_Hhydrogenifor_@WP_013406025_1
373903_000020485_1_Horenii_@WP_012636034_1
748449_000328625_1_Hhalobius_@WP_015327952_1
574087_000144695_1_Aarabaticum_@WP_013277215_1
292459_000009905_1_Sthermophilum_@WP_050742138_1
357809_000018685_1_Lphytoferment_@WP_012200296_1
591001_000025305_1_Afermentans_@WP_012938870_1
568816_000230275_1_Aintestini_@WP_014128297_1
1192197_000271665_2_Pfermentans_@WP_007952871_1
484770_000725345_1_Psp_@WP_038670268_1
479436_000024945_1_Vparvula_@WP_012864113_1
546271_000208405_1_Ssputigena_@WP_006192983_1
927704_000284095_1_Sruminantium_@WP_014424242_1
712538_001189555_1_Ssp_@WP_050343942_1
1089553_000305935_1_Tphaeum_@WP_015050337_1
1525_001267405_1_Mthermoacetic_@WP_011393198_1
488761_000019165_1_Hmodesticaldu_@WP_041313599_1
1458465_001304715_1_Melsdenii_@WP_022497176_1
642492_000178835_2_Clentocellum_@WP_013657898_1
1209989_000213235_1_Tacetatoxydan_@WP_013778244_1
555079_000144645_1_Toceani_@WP_013276165_1
457570_000020005_1_Nthermophilus_@WP_012448233_1
492670_000988345_1_Bvelezensis_@WP_038460067_1
1126211_000262385_1_Bamyloliquefa_@WP_014419021_1
1452_000830075_1_Batrophaeus_@WP_003326699_1
279010_000084825_1_Blicheniformi_@WP_003185760_1
1574141_000877815_1_Bsp_@WP_041850467_1
1408_000972685_1_Bpumilus_@WP_017368151_1
224308_000009045_1_Bsubtilis_@NP_391443_1
1006007_000225265_1_Bmegaterium_@WP_014461254_1
796606_000724485_1_Bmethanolicus_@WP_003346679_1
697284_000511405_1_Plarvae_@WP_052337454_1
697281_000213255_1_Maustraliensi_@WP_013780022_1
1048834_000219875_1_Aacidocaldari_@WP_014465255_1
1048834_000219875_1_Aacidocaldari_@WP_041695064_1
200991_001465795_1_Prifietoensis_@WP_058382999_1
748727_000143685_1_Cljungdahlii_@WP_013239197_1
1341692_000484505_1_Cautoethanoge_@WP_013239197_1
536227_001038625_1_Ccarboxidivor_@WP_007061689_1
573061_000145275_1_Ccellulovoran_@WP_010077518_1
863638_000191905_1_Cacetobutylic_@WP_010965801_1
1128398_000299355_1_Cacidurici_@WP_014967209_1
875453_000952975_1_Psp_@WP_045078067_1
693746_000283575_1_Ovalericigene_@WP_014118084_1
1297617_001454945_1_Ibutyriciprod_@WP_058117727_1
768706_000235605_1_Dorientis_@WP_014183425_1
768704_000231385_3_Dmeridiei_@AFQ42951_1
646529_000255115_3_Dacidiphilus_@AFM42374_1
871968_000231405_3_Dmetallireduc_@AHF07480_1
871738_000512895_1_Drestrictus_@WP_019225940_1
1147129_000305775_1_Dsp_@WP_015042922_1
645991_000190635_1_Sglycolicus_@WP_013625277_1
138119_000010045_1_Dhafniense_@WP_041272435_1
756499_000243155_3_Ddehalogenans_@AFM01338_1
871963_000243135_3_Ddichloroelim_@AGA69947_1
635013_000092945_1_Tpotens_@WP_013120070_1
760568_000214705_1_Dkuznetsovii_@WP_013821794_1
767817_000233715_3_Dgibsoniae_@AGL01242_1
370438_000010565_1_Pthermopropio_@WP_012032201_1
868595_000214435_1_Dnigrificans_@WP_013810082_1
485916_000024205_1_Dacetoxidans_@WP_015756719_1
429009_000024605_1_Adegensii_@WP_015739598_1
509193_000175295_2_Tbrockii_@WP_012269270_1
340099_000019085_1_Tpseudethanol_@WP_012269270_1
399726_000019065_1_Tsp_@WP_009052286_1
273068_000007085_1_Csubterraneus_@WP_011025471_1
246194_000012865_1_Chydrogenofor_@WP_011344424_1
632348_000166775_1_Ckronotskyens_@WP_013431226_1
521460_000022325_1_Cbescii_@WP_015906862_1
632292_000166355_1_Chydrothermal_@WP_013404197_1
608506_000145215_1_Cobsidiansis_@WP_013289598_1
632516_000193435_3_Clactoaceticu_@AEM74674_1
632335_000166695_1_Ckristjansson_@WP_013431651_1
632518_000166335_1_Cowensensis_@WP_013411202_1
351627_000016545_1_Csaccharolyti_@WP_011915765_1
720554_000237085_1_Cclariflavum_@WP_014255551_1
637887_000184925_1_Rthermocellum_@WP_003518655_1
394503_000022065_1_Ccellulolytic_@WP_015924082_1
1121335_000331995_1_Cstercorarium_@WP_015359266_1
1111068_000236605_1_Gthermoleovor_@WP_014196793_1
235909_000009785_1_Gkaustophilus_@WP_011232626_1
272567_001274575_1_Gstearothermo_@WP_053414073_1
796606_000724485_1_Bmethanolicus_@WP_003346488_1
491915_000019045_1_Aflavithermus_@WP_012576093_1
198467_001187595_1_Agonensis_@WP_035066115_1
583358_000092965_1_Tmathranii_@WP_013150230_1
580331_000025645_1_Ttalicus_@WP_012995080_1
2325_000763575_1_Tkivui_@WP_049685079_1
697303_000147695_3_Twiegelii_@AEM78581_1
1094508_000307585_1_Tsaccharolyti_@WP_014758556_1
858215_000189775_3_Txylanolyticu_@AEF17464_1
698948_000328545_1_Tthermosaccha_@WP_015311726_1
546269_000163895_2_Falocis_@WP_014262455_1
1511_000196455_1_Csticklandii_@WP_049779765_1
1286171_000597865_1_Eacidaminophi_@WP_025435276_1
293826_000016935_1_Ametalliredig_@WP_012063469_1
84022_001042715_1_Caceticum_@WP_044823895_1
350688_000018325_1_Aoremlandii_@WP_012159576_1
0.3

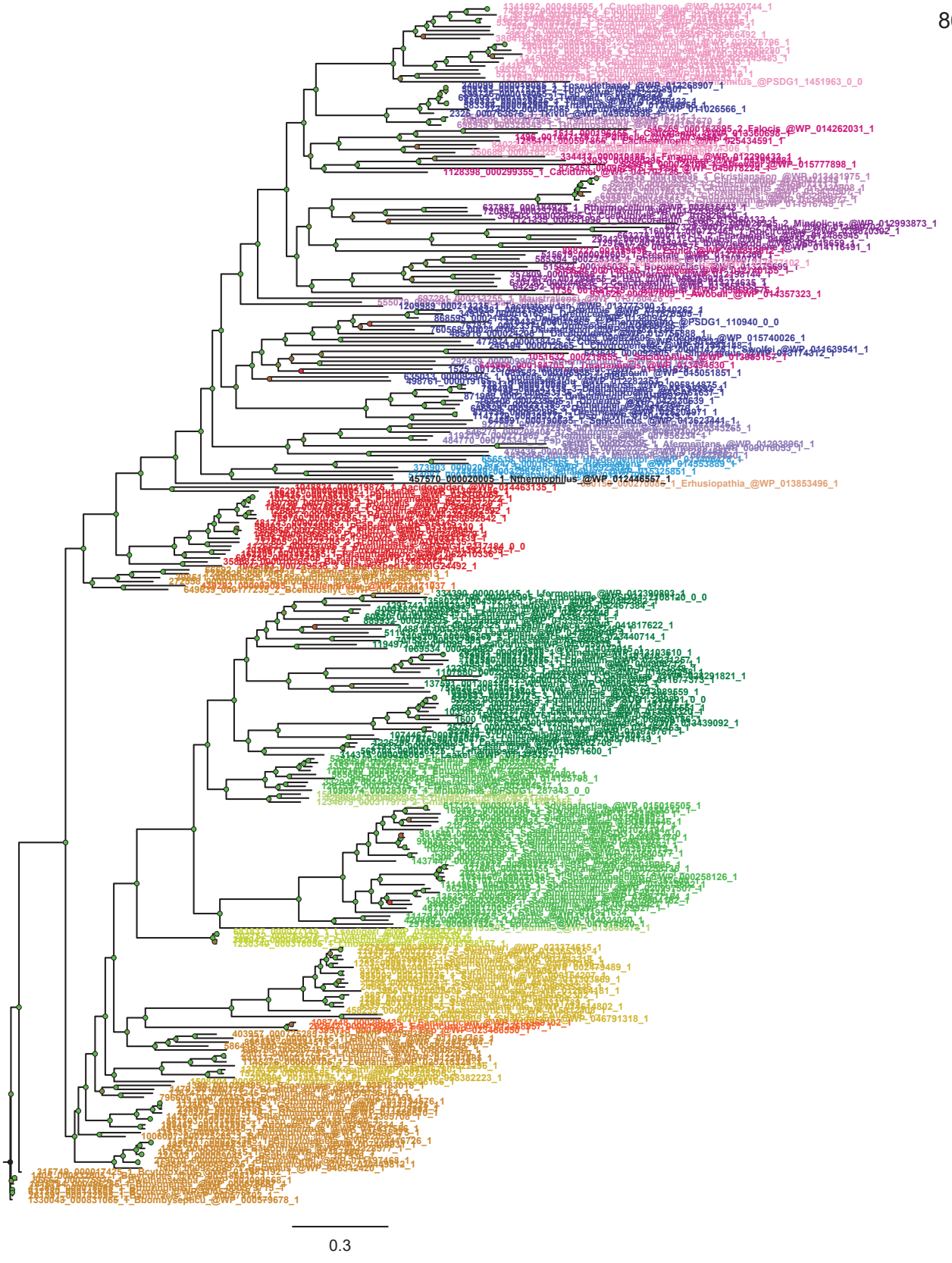
1419814_000688775_1_Ssp_@WP_044020926_1
81
927666_000253155_1_Soralis_@WP_041170860_1
28037_001281025_1_Smitis_@WP_060628463_1
1054460_000221985_1_Spseudopneumo_@WP_000757762_1
171101_000007045_1_Spneumoniae_@NP_358461_1
0.07

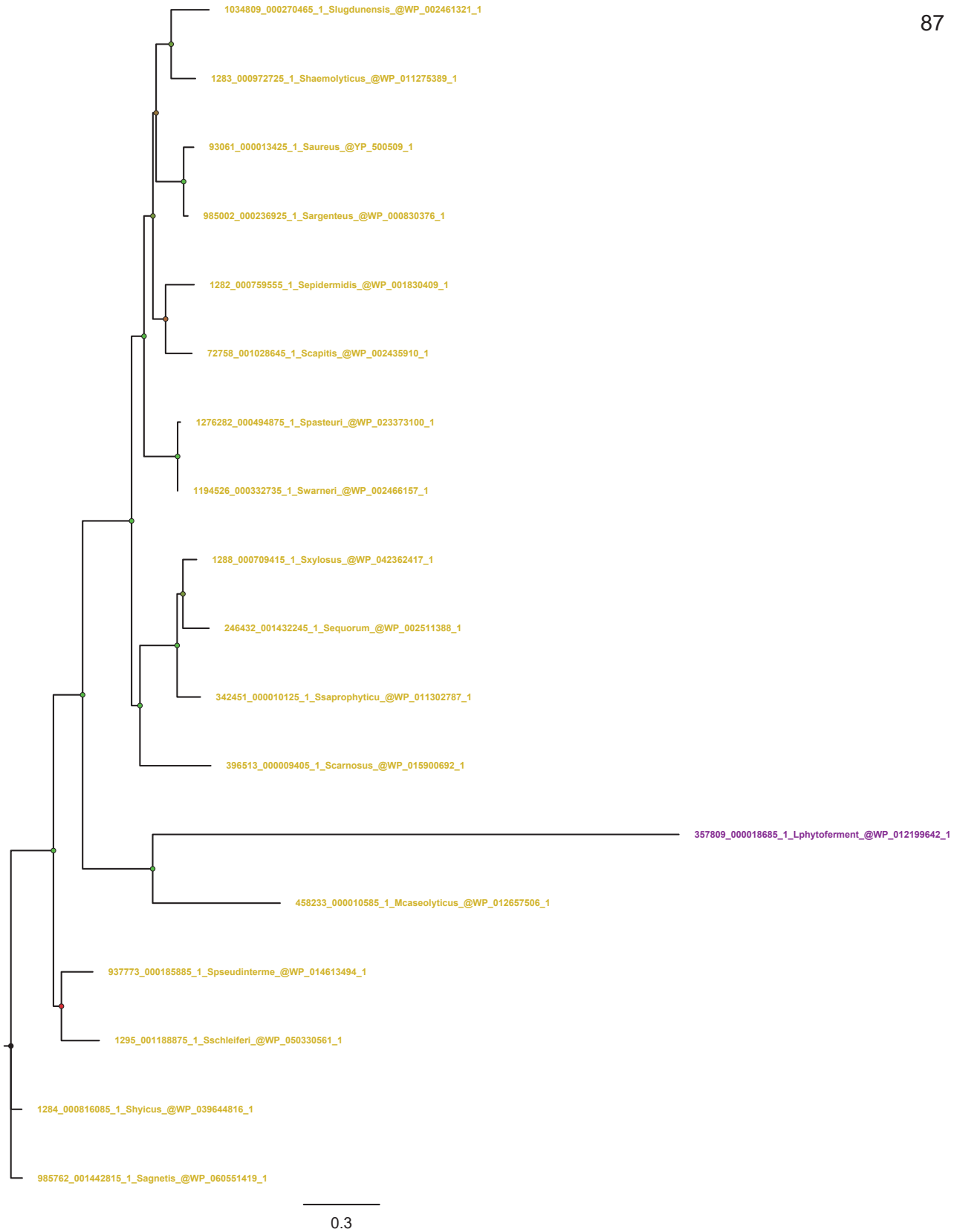

1415784_000496285_1_Btoyonensis_@WP_000532007_1
86662_000775975_1_Bweihenstepha_@WP_002035473_1
1405_000832605_1_Bmycoides_@WP_002127386_1
1405_000832605_1_Bmycoides_@WP_016127804_1
1330043_000831065_1_Bbombysepticu_@WP_044584196_1
637380_000143605_1_Bcereus_@WP_000897111_1
86662_000775975_1_Bweihenstepha_@WP_002034119_1
412694_000015065_1_Bthuringiensi_@WP_000907112_1
1405_000832605_1_Bmycoides_@WP_002138429_1
1415784_000496285_1_Btoyonensis_@WP_000897105_1
261591_000742895_1_Banthracis_@WP_000907113_1
315749_000017425_1_Bcytotoxicus_@WP_012095550_1
1330043_000831065_1_Bbombysepticu_@WP_000897099_1
198467_001187595_1_Agonensis_@WP_035067140_1
491915_000019045_1_Aflavithermus_@WP_012574433_1
66692_000009825_1_Bclausii_@WP_050748964_1
1246626_000706725_1_Blehensis_@WP_051667458_1
272558_000011145_1_Bhalodurans_@WP_010897445_1
398511_000005825_2_Bpseudofirmus_@WP_012957803_1
649639_000177235_2_Bcellulosilyt_@WP_013488142_1
1508404_000818095_1_Jmalaysiensis_@WP_039809985_1
439292_000093085_1_Bselenitiredu_@WP_013173356_1
1374_001465835_1_Pkocurii_@WP_058395592_1
1526927_000785555_1_Psp_@WP_038703654_1
200991_001465795_1_Prifietoensis_@WP_058380960_1
28031_000724775_3_Lfusiformis_@WP_036127264_1
444177_000017965_1_Lsphaericus_@WP_012295383_1
1145276_000600105_1_Lvarians_@WP_025220073_1
1574141_000877815_1_Bsp_@WP_003246156_1
224308_000009045_1_Bsubtilis_@NP_390608_1
1452_000830075_1_Batrophaeus_@WP_003325333_1
1126211_000262385_1_Bamyloliquefa_@WP_014418462_1
492670_000988345_1_Bvelezensis_@WP_038459326_1
1006007_000225265_1_Bmegaterium_@PSDG1_673737_2e_31
135735_000972245_2_Bendophyticus_@WP_019391393_1
1648923_000876525_1_Bparalichenif_@WP_026579203_1
279010_000008425_1_Blicheniformi_@WP_011201705_1
1408_000972685_1_Bpumilus_@WP_017367687_1
1367477_000473245_1_Binfantis_@WP_022544215_1
796606_000724485_1_Bmethanolicus_@WP_004435911_1
420890_000269945_1_Lgarvieae_@WP_003133453_1
1117941_000478255_2_Llactis_@All13456_1
1398_001039495_1_Bcoagulans_@WP_014097925_1
202752_000763475_1_Livanovii_@WP_038408907_1
386043_000060285_1_Lwelshimeri_@WP_011702298_1
1230340_000318055_1_Lmonocytogene_@WP_003734326_1
683837_000027145_1_Lseeligeri_@WP_012985667_1
1002809_000271325_1_Lsilvestris_@WP_014823367_1
1111068_000236605_1_Gthermoleovor_@WP_023634611_1
272567_001274575_1_Gstearothermo_@WP_013144667_1
235909_000009785_1_Gkaustophilus_@WP_049752173_1
420246_000015745_1_Gthermodenitr_@WP_011887867_1
581103_000186075_1_Gsp_@WP_013400338_1
1426_001295365_1_Gthermoglucos_@WP_003248977_1
1479_001050115_1_Bsmithii_@WP_048623729_1
768486_000271405_2_Ehirae_@WP_010720826_1
1352_001412695_1_Efaecium_@WP_002286760_1
53345_001267395_1_Edurans_@WP_005879338_1
1300150_000504125_1_Emundtii_@WP_010736198_1
945021_000283615_1_Thalophilus_@WP_014125506_1
1234679_000317975_2_Cmaltaromatic_@WP_010050057_1
1358027_000439275_1_Lreuteri_@WP_020843075_1
297352_000981525_1_Lpiscium_@WP_047916249_1
332949_001465115_1_Esilesiacus_@ALS01275_1
1261557_000317915_1_Efaecalis_@WP_002356451_1
1564681_001483965_1_Csp_@WP_058918206_1
698758_000307165_1_Axylanus_@WP_015009813_1
866895_000284515_1_Hhalophilus_@WP_014643841_1
403957_000725285_1_Vsp_@WP_040955252_1
1564681_001483965_1_Csp_@WP_058918597_1
221109_000011245_1_Oiheyensis_@WP_011066397_1
1266845_000493735_1_Cinhibens_@WP_023177996_1
586416_000725365_1_Taidingensis_@WP_038561531_1
1408206_000496265_1_Ppentosaceus_@WP_023440528_1
701521_000237995_2_Pclaussenii_@AEV95089_1
60520_001443645_1_Lparaplantaru_@WP_033610981_1
889932_000148815_2_Lplantarum_@WP_003640846_1
1194971_001011095_1_Lsalivarius_@WP_047035630_1
1291742_000829395_1_Lhokkaidonens_@WP_041093874_1
334390_000010145_1_Lfermentum_@WP_003682113_1
1358027_000439275_1_Lreuteri_@WP_019254302_1
1130798_000248095_3_Lmucosae_@AJT50908_1
148814_001314945_1_Lkunkeei_@WP_034531778_1
1001583_000359625_1_Lbrevis_@WP_011668164_1
637971_001050435_1_Lkoreensis_@WP_048735817_1
1045854_000219805_1_Wkoreensis_@WP_006845153_1
759620_000750515_1_Wceti_@WP_009495751_1
762550_000198855_1_Lgelidum_@WP_013231461_1
1107880_000234825_3_Lmesenteroide_@WP_014324490_1
137591_001308145_1_Wcibaria_@WP_052497188_1
1045004_000241055_1_Okitaharae_@WP_007746346_1
511437_000211375_1_Lbuchneri_@WP_013727524_1
1226298_000829035_1_Lparacasei_@WP_016365910_1
568704_000026525_1_Lrhamnosus_@WP_005685599_1
219334_000829055_1_Lcasei_@WP_010488633_1
314315_000026065_1_Lsakei_@WP_011374482_1
1074467_000831645_3_Lheilongjiang_@WP_041499464_1
1007676_001050475_1_Lginsenosidim_@WP_048706317_1
767455_000182835_1_Ldelbrueckii_@WP_003617478_1
880633_000165775_3_Lhelveticus_@AFR22301_1
52242_001314245_1_Lgallinarum_@WP_060471924_1
695562_000194115_1_Lamylovorus_@WP_013437637_1
1033837_000214785_1_Lkefiranofaci_@WP_013854314_1
1600_001042405_1_Lacetotoleran_@WP_056969579_1
324831_000014425_1_Lgasseri_@WP_003647112_1
257314_000008065_1_Ljohnsonii_@WP_011161849_1
160491_000009385_1_Spyogenes_@WP_002990826_1
617121_000307185_1_Sdysgalactiae_@WP_003054365_1
1437447_000817065_1_Smutans_@WP_002263612_1
552526_000020765_1_Sequi_@WP_012516065_1
218495_000009545_1_Suberis_@WP_012658003_1
873447_000187935_2_Sparauberis_@EGE54199_1
1346_000831485_1_Siniae_@WP_003101477_1
1311_001026925_1_Sagalactiae_@WP_000517852_1
990317_000203195_1_Sgallolyticus_@WP_003065954_1
981540_000270165_1_Spasteurianus_@WP_003065954_1
1116231_000283635_1_Smacedonicus_@WP_014295059_1
1069533_000246835_1_Sinfantarius_@WP_014335144_1
1076934_000441535_1_Slutetiensis_@WP_020917270_1
1304_000785515_1_Ssalivarius_@WP_013990959_1
1308_000971665_1_Sthermophilus_@WP_011226984_1
1307_000993745_1_Ssuis_@WP_002936700_1
28037_001281025_1_Smitis_@WP_060628063_1
1054460_000221985_1_Spseudopneumo_@WP_000517865_1
171101_000007045_1_Spneumoniae_@NP_358487_1
1419814_000688775_1_Ssp_@WP_044020650_1
927666_000253155_1_Soralis_@WP_000517872_1
1114965_000262145_1_Sparasanguini_@WP_013903281_1
1338_001296205_1_Sintermedius_@ALF27971_1
862968_000463445_1_Sconstellatus_@WP_003025491_1
1353243_000478925_1_Sanginosus_@WP_003034247_1
467705_000017005_1_Sgordonii_@WP_012130544_1
1302863_000385925_1_Soligoferment_@WP_015605302_1
388919_000014205_1_Ssanguinis_@YP_001035577_1
0.5

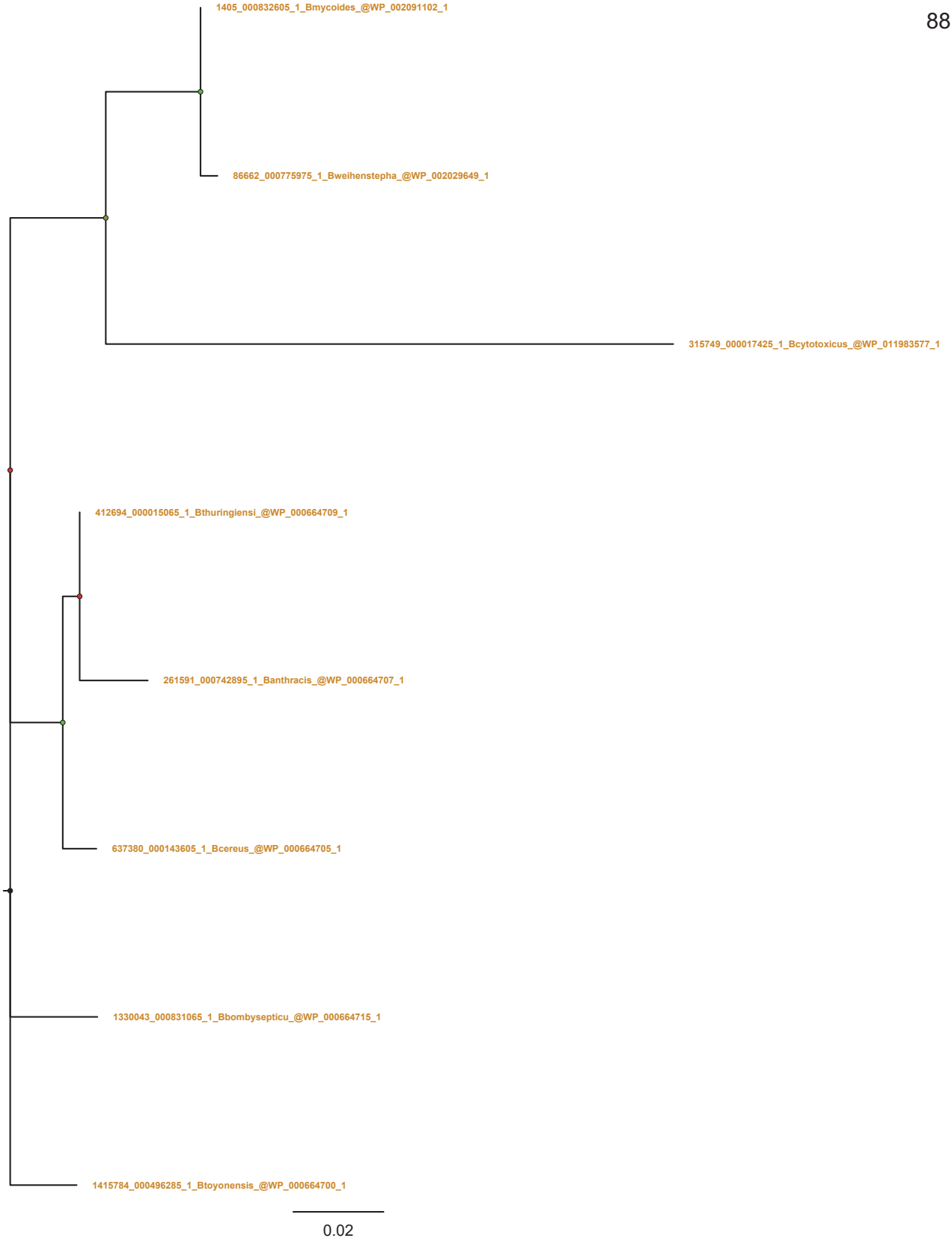

0.3

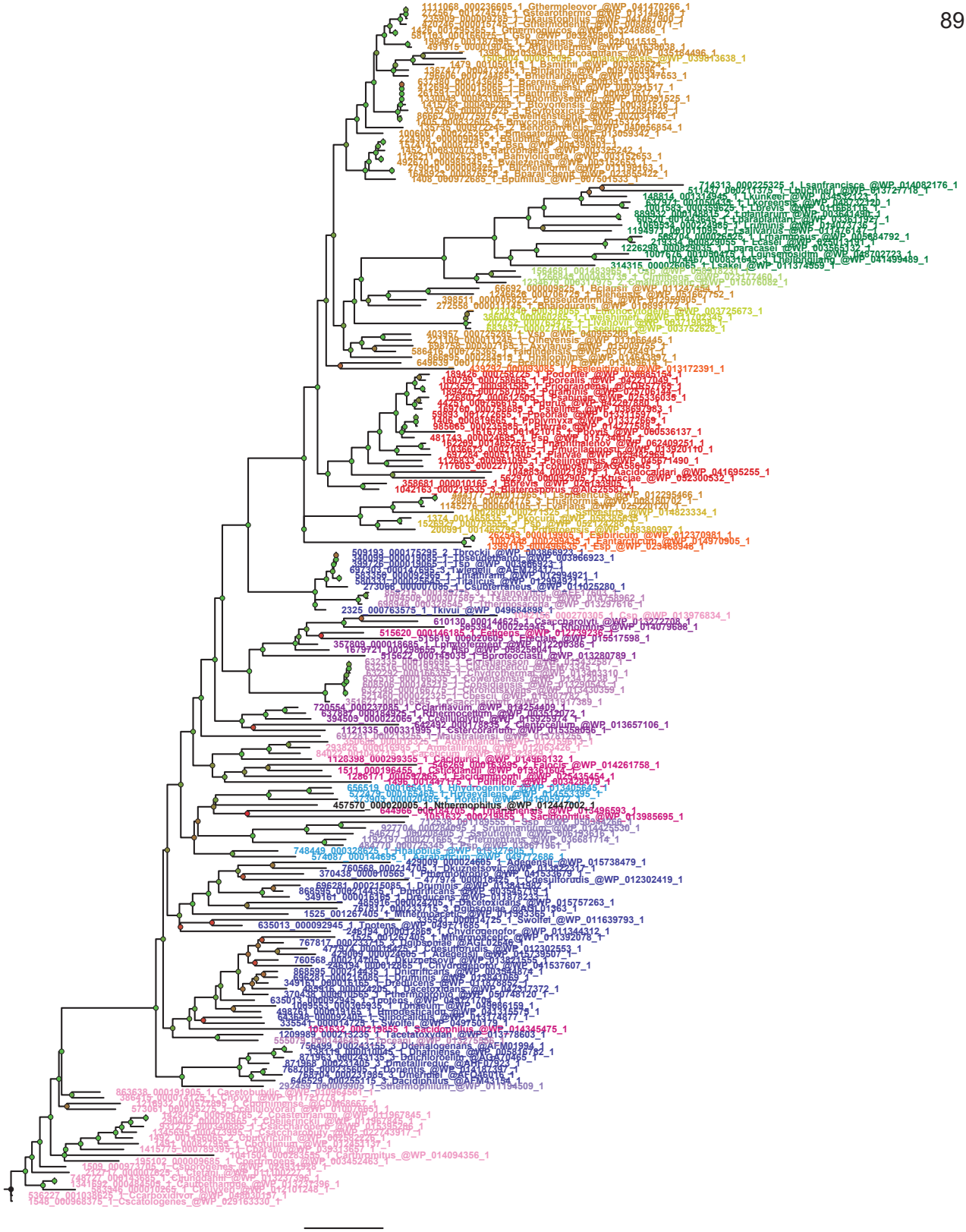

1346_000831485_1_Siniae_@WP_003101780_1
218495_000009545_1_Suberis_@WP_015911722_1
873447_000187935_2_Sparauberis_@EGE54273_1
160491_000009385_1_Spyogenes_@WP_002983629_1
617121_000307185_1_Sdysgalactiae_@WP_015017303_1
552526_000020765_1_Sequi_@WP_012516155_1
1311_001026925_1_Sagalactiae_@WP_001287870_1
1304_000785515_1_Ssalivarius_@WP_038675257_1
1308_000971665_1_Sthermophilus_@WP_002947758_1
1116231_000283635_1_Smacedonicus_@WP_014294058_1
990317_000203195_1_Sgallolyticus_@WP_013642734_1
981540_000270165_1_Spasteurianus_@WP_013851519_1
1076934_000441535_1_Slutetiensis_@WP_020916302_1
1069533_000246835_1_Sinfantarius_@PSDG1_413437_3e_81
1437447_000817065_1_Smutans_@WP_002263048_1
1307_000993745_1_Ssuis_@WP_011921984_1
862968_000463445_1_Sconstellatus_@WP_020997625_1
1338_001296205_1_Sintermedius_@ALF28095_1
1353243_000478925_1_Sanginosus_@WP_025271945_1
467705_000017005_1_Sgordonii_@WP_041789565_1
1302863_000385925_1_Soligoferment_@WP_015605367_1
388919_000014205_1_Ssanguinis_@YP_001035785_1
1114965_000262145_1_Sparasanguini_@WP_014712734_1
171101_000007045_1_Spneumoniae_@NP_357928_1
1054460_000221985_1_Spseudopneumo_@WP_000039262_1
28037_001281025_1_Smitis_@WP_060627575_1
927666_000253155_1_Soralis_@WP_000039243_1
1419814_000688775_1_Ssp_@WP_044021314_1
53345_001267395_1_Edurans_@WP_016176832_1
768486_000271405_2_Ehirae_@WP_010737344_1
1300150_000504125_1_Emundtii_@WP_023519468_1
1352_001412695_1_Efaecium_@WP_002321398_1
332949_001465115_1_Esilesiacus_@ALS00702_1
1090974_000283975_1_Mplutonius_@WP_015694884_1
1261557_000317915_1_Efaecalis_@WP_002410201_1
565655_000157355_2_Ecasseliflavu_@WP_015510080_1
297352_000981525_1_Lpiscium_@WP_047915397_1
420890_000269945_1_Lgarvieae_@WP_014024431_1
1117941_000478255_2_Llactis_@AII13209_1
0.3

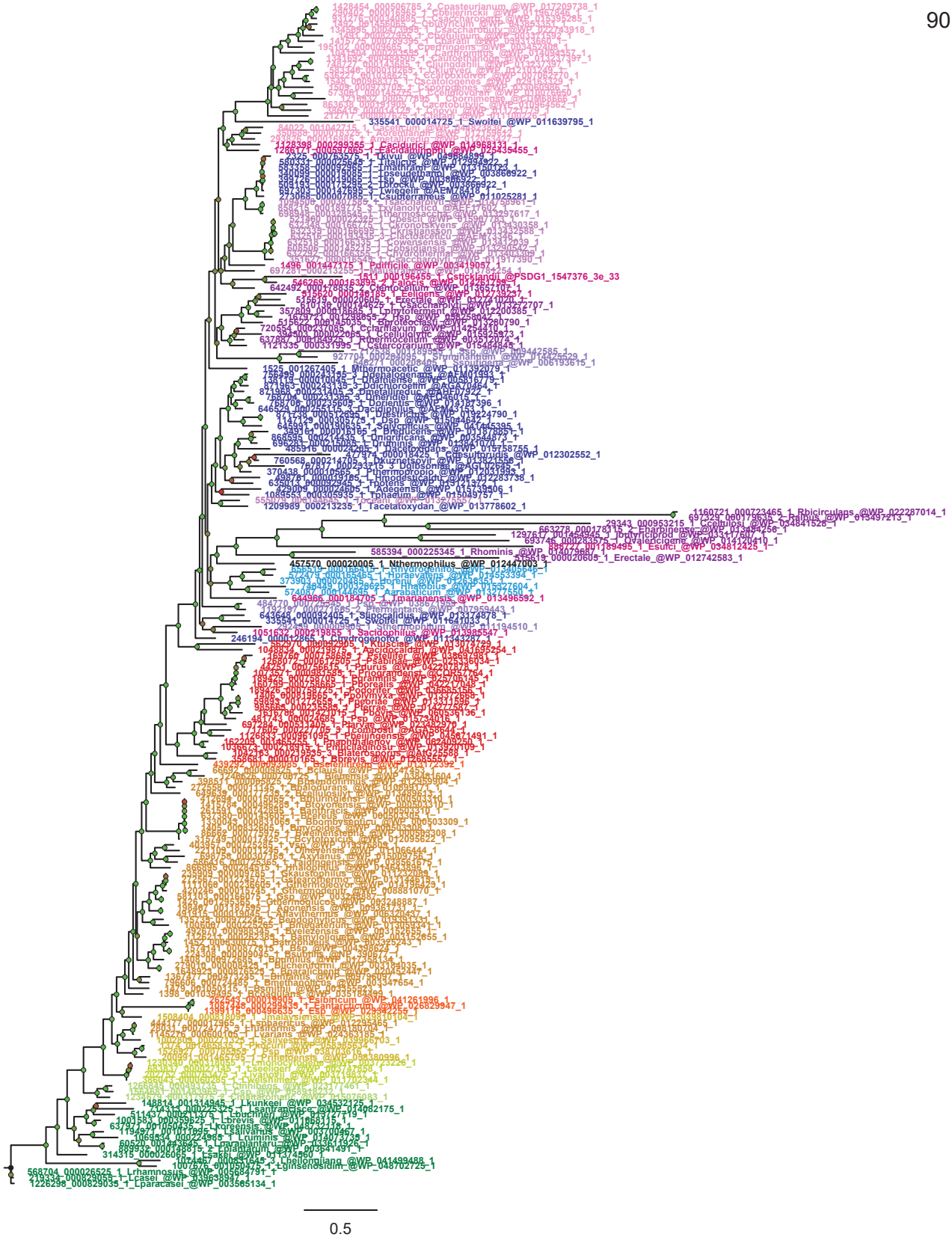

0.3

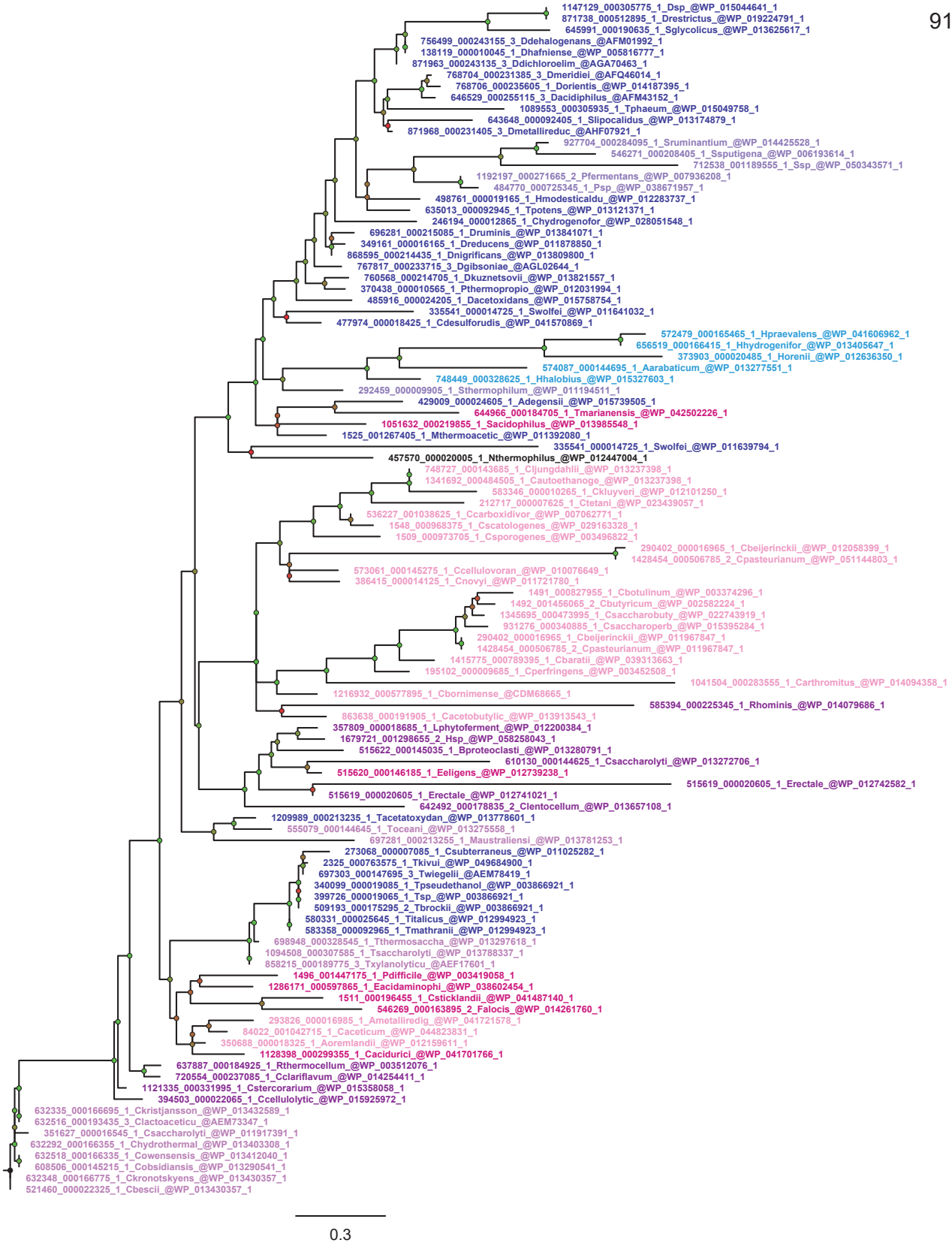

0.3

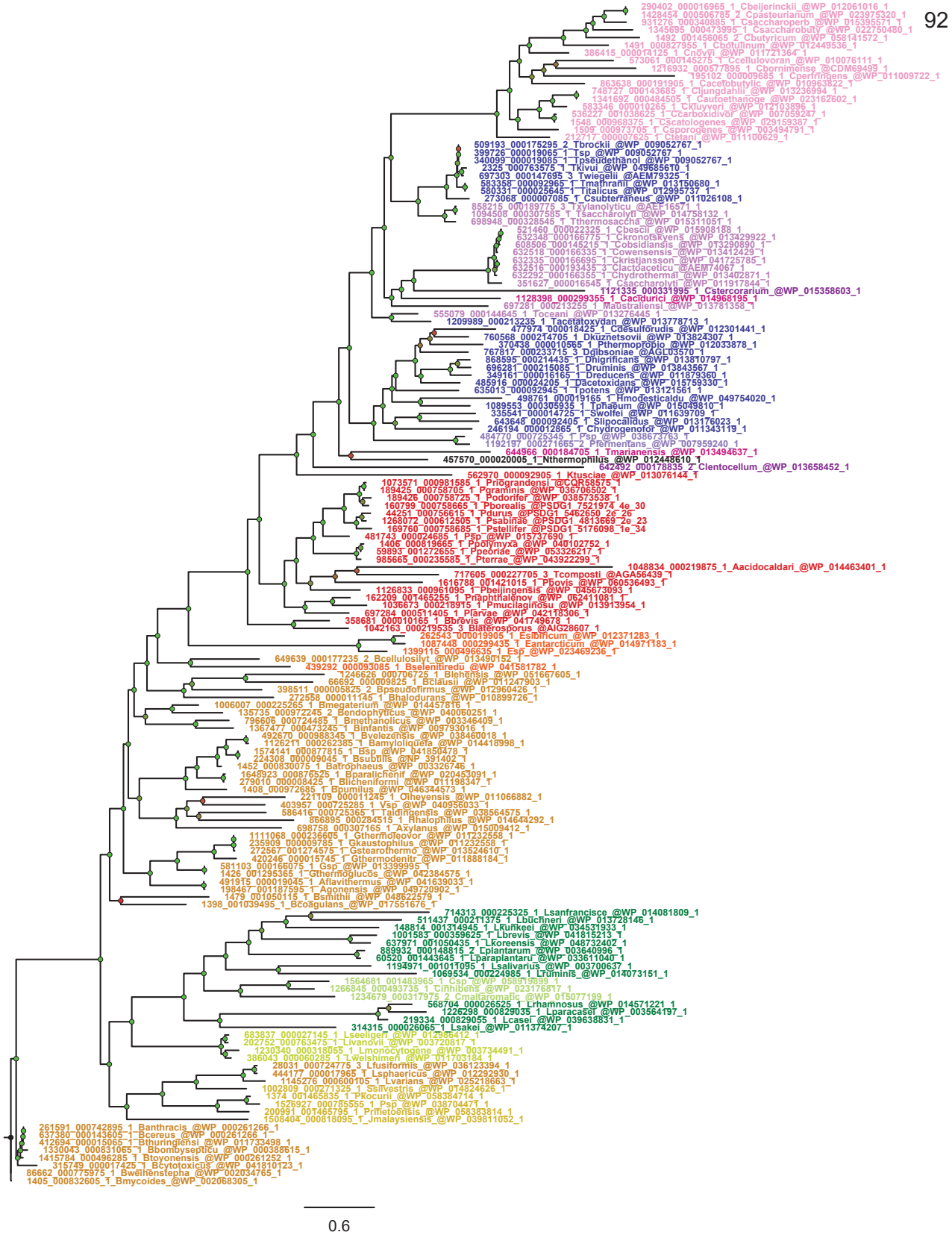

1034809_000270465_1_Slugdunensis_@WP_002461321_1
1283_000972725_1_Shaemolyticus_@WP_011275389_1
93061_000013425_1_Saureus_@YP_500509_1
985002_000236925_1_Sargenteus_@WP_000830376_1
1282_000759555_1_Sepidermidis_@WP_001830409_1
72758_001028645_1_Scapitis_@WP_002435910_1
1276282_000494875_1_Spasteuri_@WP_023373100_1
1194526_000332735_1_Swarneri_@WP_002466157_1
1288_000709415_1_Sxylosus_@WP_042362417_1
246432_001432245_1_Sequorum_@WP_002511388_1
342451_000010125_1_Ssaprophyticu_@WP_011302787_1
396513_000009405_1_Scarnosus_@WP_015900692_1
357809_000018685_1_Lphytoferment_@WP_012199642_1
458233_000010585_1_Mcaseolyticus_@WP_012657506_1
937773_000185885_1_Spseudinterme_@WP_014613494_1
1295_001188875_1_Sschleiferi_@WP_050330561_1
1284_000816085_1_Shyicus_@WP_039644816_1
985762_001442815_1_Sagnetis_@WP_060551419_1
0.3

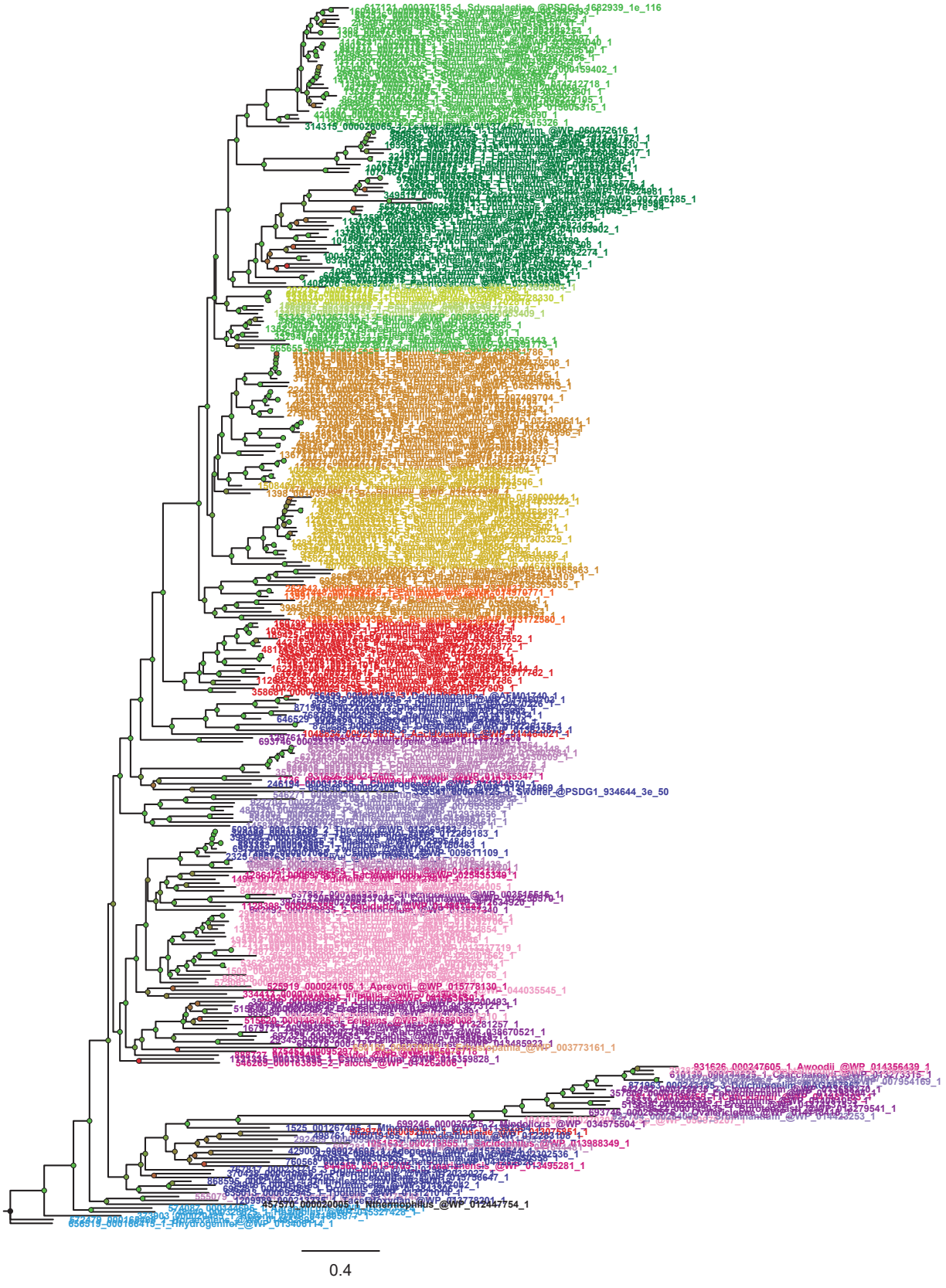

1405_000832605_1_Bmycoides_@WP_002091102_1
86662_000775975_1_Bweihenstepha_@WP_002029649_1
315749_000017425_1_Bcytotoxicus_@WP_011983577_1
412694_000015065_1_Bthuringiensi_@WP_000664709_1
261591_000742895_1_Banthracis_@WP_000664707_1
637380_000143605_1_Bcereus_@WP_000664705_1
1330043_000831065_1_Bbombysepticu_@WP_000664715_1
1415784_000496285_1_Btoyonensis_@WP_000664700_1
0.02

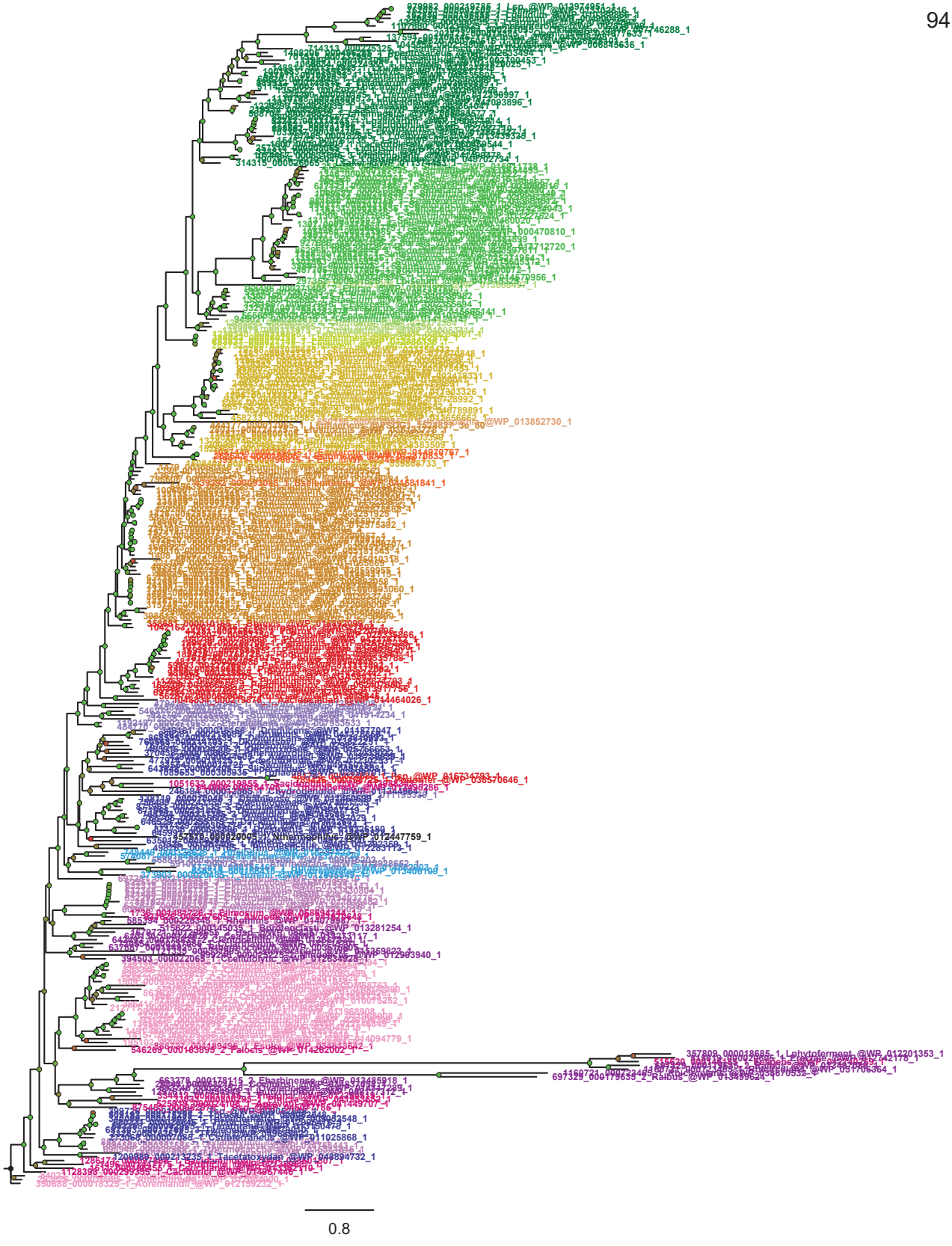

0.5

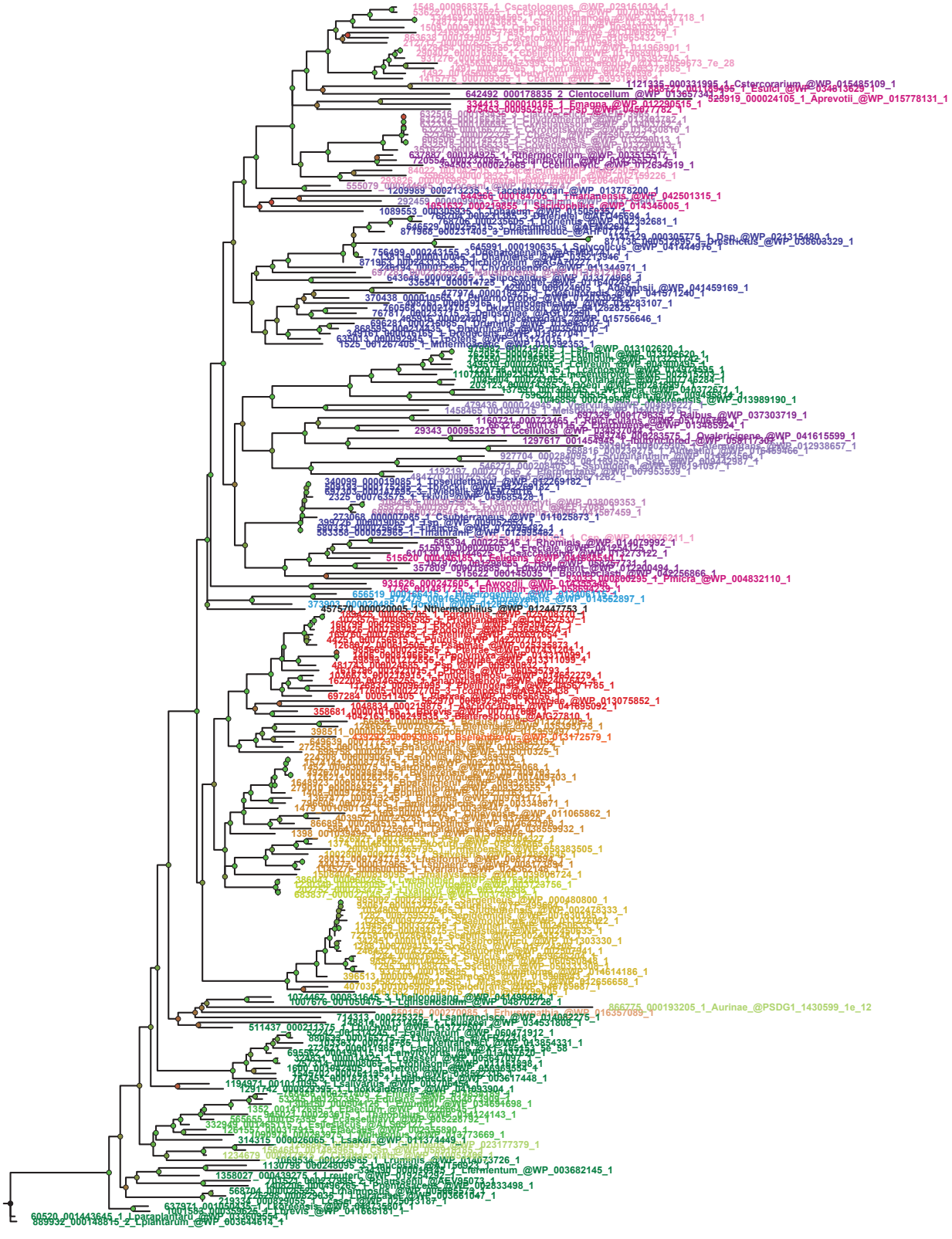

0.5

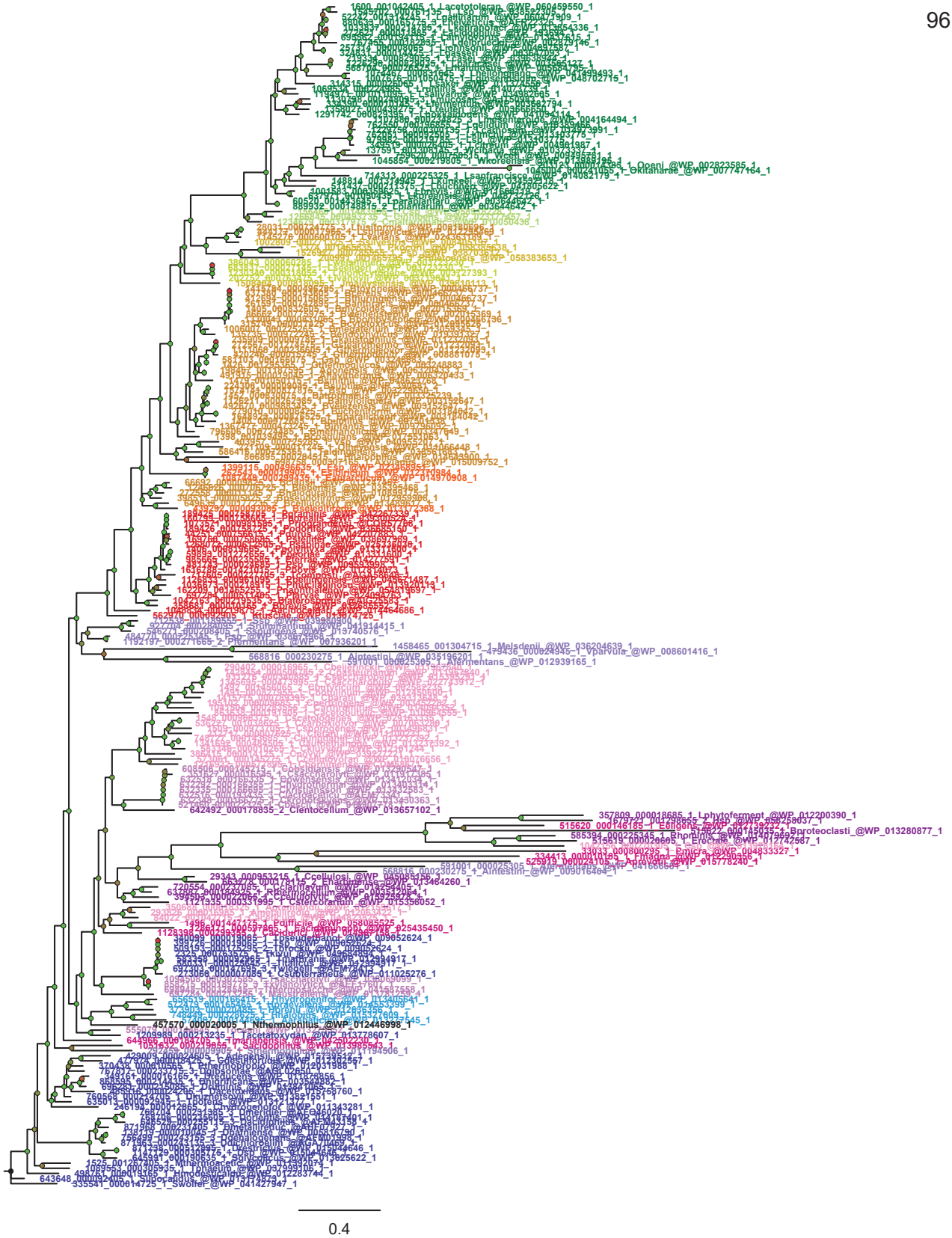

1147129_000305775_1_Dsp_@WP_015044641_1
871738_000512895_1_Drestrictus_@WP_019224791_1
645991_000190635_1_Sglycolicus_@WP_013625617_1
756499_000243155_3_Ddehalogenans_@AFM01992_1
138119_000010045_1_Dhafniense_@WP_005816777_1
871963_000243135_3_Ddichloroelim_@AGA70463_1
768704_000231385_3_Dmeridiei_@AFQ46014_1
768706_000235605_1_Dorientis_@WP_014187395_1
646529_000255115_3_Dacidiphilus_@AFM43152_1
1089553_000305935_1_Tphaeum_@WP_015049758_1
643648_000092405_1_Slipocalidus_@WP_013174879_1
871968_000231405_3_Dmetallireduc_@AHF07921_1
927704_000284095_1_Sruminantium_@WP_014425528_1
546271_000208405_1_Ssputigena_@WP_006193614_1
712538_001189555_1_Ssp_@WP_050343571_1
1192197_000271665_2_Pfermentans_@WP_007936208_1
484770_000725345_1_Psp_@WP_038671957_1
498761_000019165_1_Hmodesticaldu_@WP_012283737_1
635013_000092945_1_Tpotens_@WP_013121371_1
246194_000012865_1_Chydrogenofor_@WP_028051548_1
696281_000215085_1_Druminis_@WP_013841071_1
349161_000016165_1_Dreducens_@WP_011878850_1
868595_000214435_1_Dnigrificans_@WP_013809800_1
767817_000233715_3_Dgibsoniae_@AGL02644_1
760568_000214705_1_Dkuznetsovii_@WP_013821557_1
370438_000010565_1_Pthermopropio_@WP_012031994_1
485916_000024205_1_Dacetoxidans_@WP_015758754_1
335541_000014725_1_Swolfei_@WP_011641032_1
477974_000018425_1_Cdesulforudis_@WP_041570869_1
572479_000165465_1_Hpraevalens_@WP_041606962_1
656519_000166415_1_Hhydrogenifor_@WP_013405647_1
373903_000020485_1_Horenii_@WP_012636350_1
574087_000144695_1_Aarabaticum_@WP_013277551_1
748449_000328625_1_Hhalobius_@WP_015327603_1
292459_000009905_1_Sthermophilum_@WP_011194511_1
429009_000024605_1_Adegensii_@WP_015739505_1
644966_000184705_1_Tmarianensis_@WP_042502226_1
1051632_000219855_1_Sacidophilus_@WP_013985548_1
1525_001267405_1_Mthermoacetic_@WP_011392080_1
335541_000014725_1_Swolfei_@WP_011639794_1
457570_000020005_1_Nthermophilus_@WP_012447004_1
748727_000143685_1_Cljungdahlii_@WP_013237398_1
1341692_000484505_1_Cautoethanoge_@WP_013237398_1
583346_000010265_1_Ckluyveri_@WP_012101250_1
212717_000007625_1_Ctetani_@WP_023439057_1
536227_001038625_1_Ccarboxidivor_@WP_007062771_1
1548_000968375_1_Cscatologenes_@WP_029163328_1
1509_000973705_1_Csporogenes_@WP_003496822_1
290402_000016965_1_Cbeijerinckii_@WP_012058399_1
1428454_000506785_2_Cpasteurianum_@WP_051144803_1
573061_000145275_1_Ccellulovoran_@WP_010076649_1
386415_000014125_1_Cnovyi_@WP_011721780_1
1491_000827955_1_Cbotulinum_@WP_003374296_1
1492_001456065_2_Cbutyricum_@WP_002582224_1
1345695_000473995_1_Csaccharobuty_@WP_022743919_1
931276_000340885_1_Csaccharoperb_@WP_015395284_1
290402_000016965_1_Cbeijerinckii_@WP_011967847_1
1428454_000506785_2_Cpasteurianum_@WP_011967847_1
1415775_000789395_1_Cbaratii_@WP_039313663_1
195102_000009685_1_Cperfringens_@WP_003452508_1
1041504_000283555_1_Carthromitus_@WP_014094358_1
1216932_000577895_1_Cbornimense_@CDM68665_1
585394_000225345_1_Rhominis_@WP_014079686_1
863638_000191905_1_Cacetobutylic_@WP_013913543_1
357809_000018685_1_Lphytoferment_@WP_012200384_1
1679721_001298655_2_Hsp_@WP_058258043_1
515622_000145035_1_Bproteoclasti_@WP_013280791_1
610130_000144625_1_Csaccharolyti_@WP_013272706_1
515620_000146185_1_Eeligens_@WP_012739238_1
515619_000020605_1_Erectale_@WP_012742582_1
515619_000020605_1_Erectale_@WP_012741021_1
642492_000178835_2_Clentocellum_@WP_013657108_1
1209989_000213235_1_Tacetatoxydan_@WP_013778601_1
555079_000144645_1_Toceani_@WP_013275558_1
697281_000213255_1_Maustraliensi_@WP_013781253_1
273068_000007085_1_Csubterraneus_@WP_011025282_1
2325_000763575_1_Tkivui_@WP_049684900_1
697303_000147695_3_Twiegelii_@AEM78419_1
340099_000019085_1_Tpseudethanol_@WP_003866921_1
399726_000019065_1_Tsp_@WP_003866921_1
509193_000175295_2_Tbrockii_@WP_003866921_1
580331_000025645_1_Titalicus_@WP_012994923_1
583358_000092965_1_Tmathranii_@WP_012994923_1
698948_000328545_1_Tthermosaccha_@WP_013297618_1
1094508_000307585_1_Tsaccharolyti_@WP_013788337_1
858215_000189775_3_Txylanolyticu_@AEF17601_1
1496_001447175_1_Pdifficile_@WP_003419058_1
1286171_000597865_1_Eacidaminophi_@WP_038602454_1
1511_000196455_1_Csticklandii_@WP_041487140_1
546269_000163895_2_Falocis_@WP_014261760_1
293826_000016985_1_Ametalliredig_@WP_041721578_1
84022_001042715_1_Caceticum_@WP_044823831_1
350688_000018325_1_Aoremlandii_@WP_012159611_1
1128398_000299355_1_Caciduric_@WP_041701766_1
637887_000184925_1_Rthermocellum_@WP_003512076_1
720554_000237085_1_Cclariflavum_@WP_014254411_1
1121335_000331995_1_Cstercorarium_@WP_015358058_1
394503_000022065_1_Ccellulolytic_@WP_015925972_1
632335_000166695_1_Ckristjansson_@WP_013432589_1
632516_000193435_3_Clactoaceticu_@AEM73347_1
351627_000016545_1_Csaccharolyti_@WP_011917391_1
632292_000166355_1_Chydrothermal_@WP_013403308_1
632518_000166335_1_Cowensensis_@WP_013412040_1
608506_000145215_1_Cobsidiansis_@WP_013290541_1
632348_000166775_1_Ckronotskyens_@WP_013430357_1
521460_000022325_1_Cbescii_@WP_013430357_1
0.3

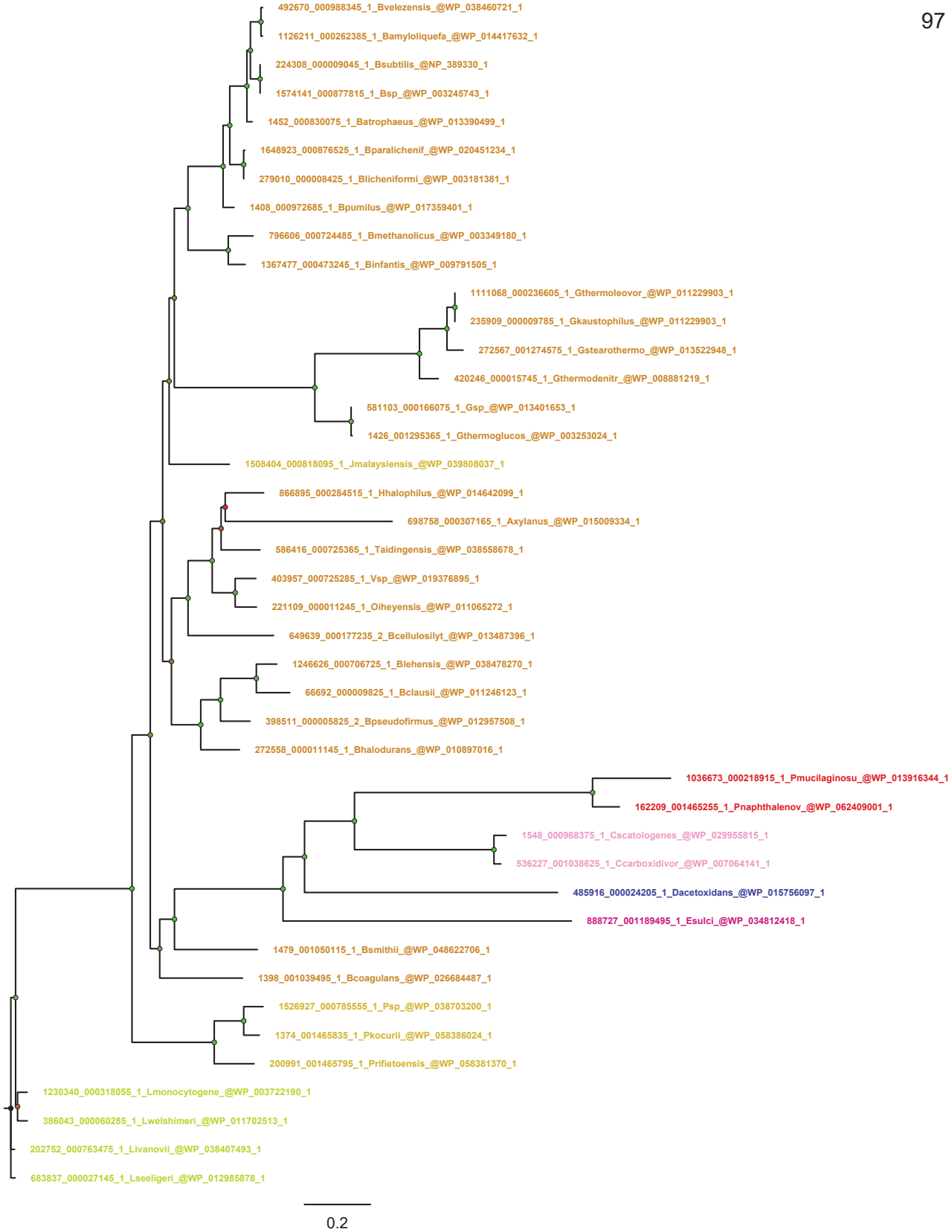

290402_000016965_1_Cbeijerinckii_@WP_012061016_1
1428454_000506785_2_Cpasteurianum_@WP_023975320_1
931276_000340885_1_Csaccharoperb_@WP_015395571_1
1345695_000473995_1_Csaccharobuty_@WP_022750480_1
1492_001456065_2_Cbutyricum_@WP_058141572_1
1491_000827955_1_Cbotulinum_@WP_012449536_1
386415_000014125_1_Cnovyi_@WP_011721364_1
573061_000145275_1_Ccellulovoran_@WP_010076111_1
1216932_000577895_1_Cbornimense_@CDM69499_1
795102_000009685_1_Cperfringens_@WP_011009722_1
863638_000191905_1_Cacetobutylic_@WP_010963922_1
748727_000143685_1_Cljungdahlii_@WP_013236994_1
1341692_000484505_1_Cautoethanoge_@WP_023162602_1
583346_000010265_1_Ckluyveri_@WP_012103896_1
536227_001038625_1_Ccarboxidivor_@WP_007059247_1
1548_000968375_1_Cscatologenes_@WP_029159387_1
1509_000973705_1_Csporogenes_@WP_003494791_1
212717_000007625_1_Ctetani_@WP_011100629_1
509193_000175295_2_Tbrockii_@WP_009052767_1
399726_000019065_1_Tsp_@WP_009052767_1
340099_000019085_1_Tpseudethanol_@WP_009052767_1
2325_000763575_1_Tkivui_@WP_045686570_1
697303_000147695_3_Twiegelii_@AEM79325_1
583358_000092965_1_Tmathranii_@WP_013150680_1
580331_000025645_1_Ttalicus_@WP_012995737_1
273068_000007085_1_Csubterraneus_@WP_011026108_1
858215_000189775_3_Txylanolyticu_@AEF16571_1
1094508_000307585_1_Tsaccharolyti_@WP_014758132_1
698948_000328545_1_Tthermosaccha_@WP_015311051_1
521460_000022325_1_Cbescii_@WP_015908188_1
632348_000166775_1_Ckronotskyens_@WP_013429922_1
608506_000145215_1_Cobsidiansis_@WP_013290890_1
632518_000166335_1_Cowensensis_@WP_013412429_1
632335_000166695_1_Ckristjansson_@WP_041725785_1
632516_000193435_3_Clactoaceticu_@AEM74067_1
632292_000166355_1_Chydrothermal_@WP_013402871_1
351627_000016545_1_Csaccharolyti_@WP_011917844_1
1121335_000331995_1_Cstercorarium_@WP_015358603_1
1128398_000299355_1_Cacidurici_@WP_014968195_1
697281_000213255_1_Maustraliensi_@WP_013781358_1
555079_000144645_1_Tocean_@WP_013276445_1
1209989_000213235_1_Tacetatoxydan_@WP_013778713_1
477974_000018425_1_Cdesulfurudis_@WP_012301441_1
760568_000214705_1_Dkuznetsovii_@WP_013823307_1
370438_000010585_1_Pthermotropio_@WP_012033878_1
767817_000233715_3_Dgibsoniae_@AGL03570_1
868595_000214435_1_Dhafniferans_@WP_013810797_1
696281_000215085_1_Druminis_@WP_013843567_1
349161_000016165_1_Dreducens_@WP_011879360_1
485916_000024205_1_Dacetoxidans_@WP_015759330_1
635013_000092945_1_Tpotens_@WP_013121561_1
498761_000019165_1_Hmodesticaldu_@WP_049754020_1
1089553_000305935_1_Tphaeum_@WP_015049810_1
335541_000014725_1_Swolfei_@WP_011639709_1
643648_000092405_1_Slipocaldius_@WP_013176023_1
246194_000012865_1_Chydrogenofor_@WP_011343119_1
484770_000725345_1_Psp_@WP_038673763_1
1192197_000271665_2_Pthermtans_@WP_007959240_1
644966_000184705_1_Tmarianensis_@WP_013494637_1
457570_000020005_1_Nthermophilus_@WP_012448610_1
642492_000178835_2_Clentocellum_@WP_013658452_1
562970_000092905_1_Ktusciae_@WP_013076144_1
1073571_000981585_1_Phiograndensi_@CQR58575_1
189425_000758705_1_Pgramlinis_@WP_036706502_1
189426_000758725_1_Podorifer_@WP_038573538_1
160799_000758665_1_Pborealis_@PSDG1_7521974_4e_30
44251_000756615_1_Pdurus_@PSDG1_5462650_2e_26
1268072_000612505_1_Psabinae_@PSDG1_4813669_2e_23
169760_000758685_1_Pstelifer_@PSDG1_5176098_1e_34
481743_000024685_1_Psp_@WP_015737690_1
1406_000819665_1_Ppolymyxa_@WP_040102752_1
59893_001272655_1_Ppeoriae_@WP_053326217_1
985665_000235585_1_Pterrae_@WP_043922299_1
1048834_000219875_1_Aacidocaldari_@WP_014463401_1
717605_000227705_3_Tcomposti_@AGA56439_1
1616788_001421015_1_Pbovis_@WP_060536493_1
1126833_000961095_1_Pbeijingensis_@WP_045673093_1
162209_001465255_1_Prhizosphaerae_@WP_052411081_1
1036673_000218915_1_Pmucilaginosu_@WP_013913954_1
697284_000511405_1_Plarvae_@WP_042118306_1
358681_000010165_1_Bbrevis_@WP_041749678_1
1042163_000219535_3_Blateroporus_@AIG28607_1
262543_000019905_1_Esibiricum_@WP_012371283_1
1087448_000299435_1_Eantarcticum_@WP_014971183_1
1398115_000496635_1_Esp_@WP_023469236_1
649639_000177235_2_Bcelluliosilyt_@WP_013490152_1
439292_000093085_1_Bselenitiredu_@WP_041581782_1
1246626_000706725_1_Blehensis_@WP_051667605_1
66692_000009925_1_Bclausii_@WP_011247903_1
398511_000005825_2_Bpseudofirmus_@WP_012960426_1
272558_000011145_1_Bhalodurans_@WP_010899726_1
1006007_000225265_1_Bmegaterium_@WP_014457816_1
135735_000972245_2_Bendophyticus_@WP_040080251_1
796606_000724485_1_Bmethanolicus_@WP_003346409_1
1367477_000473245_1_Binfantis_@WP_009793016_1
492670_000983545_1_Bvelezensis_@WP_038460018_1
1126211_000262385_1_Bamyloliquefa_@WP_014418998_1
1574141_000877815_1_Bsp_@WP_041850478_1
224308_000009045_1_Bsubtilis_@NP_391402_1
1452_000830075_1_Batrophaeus_@WP_003326746_1
1648923_000876525_1_Bparalichenif_@WP_020453091_1
279010_000008425_1_Blicheniformi_@WP_011198347_1
1408_000972685_1_Bpumilus_@WP_046344573_1
221109_000011245_1_Olheyensis_@WP_011066882_1
403957_000725285_1_Vsp_@WP_040056033_1
586416_000725365_1_Taldingensis_@WP_038564575_1
866895_000284515_1_Rhalophilus_@WP_014644292_1
698758_000307165_1_Axylanus_@WP_015094412_1
1111068_000236605_1_Gthermoleovor_@WP_011232558_1
235909_000009785_1_Gkaustophilus_@WP_011232558_1
272567_001274575_1_Gstearothermo_@WP_013524610_1
420246_000015745_1_Gthermodenitr_@WP_011888184_1
581103_000166075_1_Gsp_@WP_013399995_1
1426_001295365_1_Gthermoglucos_@WP_042384575_1
491975_000019045_1_Aflavithermus_@WP_041639033_1
198467_001187595_1_Agonensis_@WP_049720902_1
1479_001050115_1_Bsmithii_@WP_048522579_1
1398_001039495_1_Bcoagulans_@WP_017551676_1
714313_000225325_1_Lsanfrancisce_@WP_014081809_1
511437_000211375_1_Lbuchneri_@WP_013728146_1
148814_001314945_1_Lkunkeei_@WP_034531933_1
1001583_000359625_1_Lbrevis_@WP_041815213_1
637971_001050435_1_Lcoreensis_@WP_048732402_1
889932_000148815_2_Lplantarum_@WP_003640996_1
60520_001443645_1_Lparaplantaru_@WP_033611040_1
1194971_001011095_1_Lsalivarius_@WP_003700637_1
1069534_000224985_1_Lruminis_@WP_014073151_1
1564681_001483965_1_Csp_@WP_058519599_1
1266845_000493735_1_Cinhibens_@WP_023176817_1
1234679_000317975_2_Gmaltaromatic_@WP_015077199_1
568704_000026525_1_Lrhamnosus_@WP_014571221_1
1226298_000829035_1_Lparacasei_@WP_003564197_1
219334_000829055_1_Lcasei_@WP_039638831_1
314315_000026065_1_Lsakei_@WP_011374207_1
683837_000027145_1_Lseeligeri_@WP_012566412_1
202752_000763475_1_Livanovii_@WP_003720817_1
1230340_000318055_1_Lmonocytogene_@WP_003734491_1
386043_000060285_1_Lwelshimeri_@WP_011703184_1
28031_000724775_3_Lfusiformis_@WP_036123394_1
444177_000017965_1_Lsphaericus_@WP_012292930_1
1145276_000600105_1_Lvarians_@WP_025218663_1
1002809_000271325_1_Ssilvestris_@WP_014824626_1
1374_001465835_1_Pkocurii_@WP_058384714_1
1526927_000785555_1_Psp_@WP_038704477_1
200391_001465795_1_Prifletoensis_@WP_058383814_1
1508404_000618095_1_Jmalaysiensis_@WP_039811052_1
261591_000742895_1_Banthracis_@WP_000261266_1
637380_000143605_1_Bcereus_@WP_000261266_1
412694_000015065_1_Bthuringiensi_@WP_011733498_1
1330043_000831065_1_Bbombysepticu_@WP_000388615_1
1415784_000496285_1_Btoyonensis_@WP_000261252_1
315749_000017425_1_Bcytotoxicus_@WP_041810123_1
86662_000775975_1_Bweihenstepha_@WP_002034765_1
1405_000832605_1_Bmycoides_@WP_002068305_1
0.6

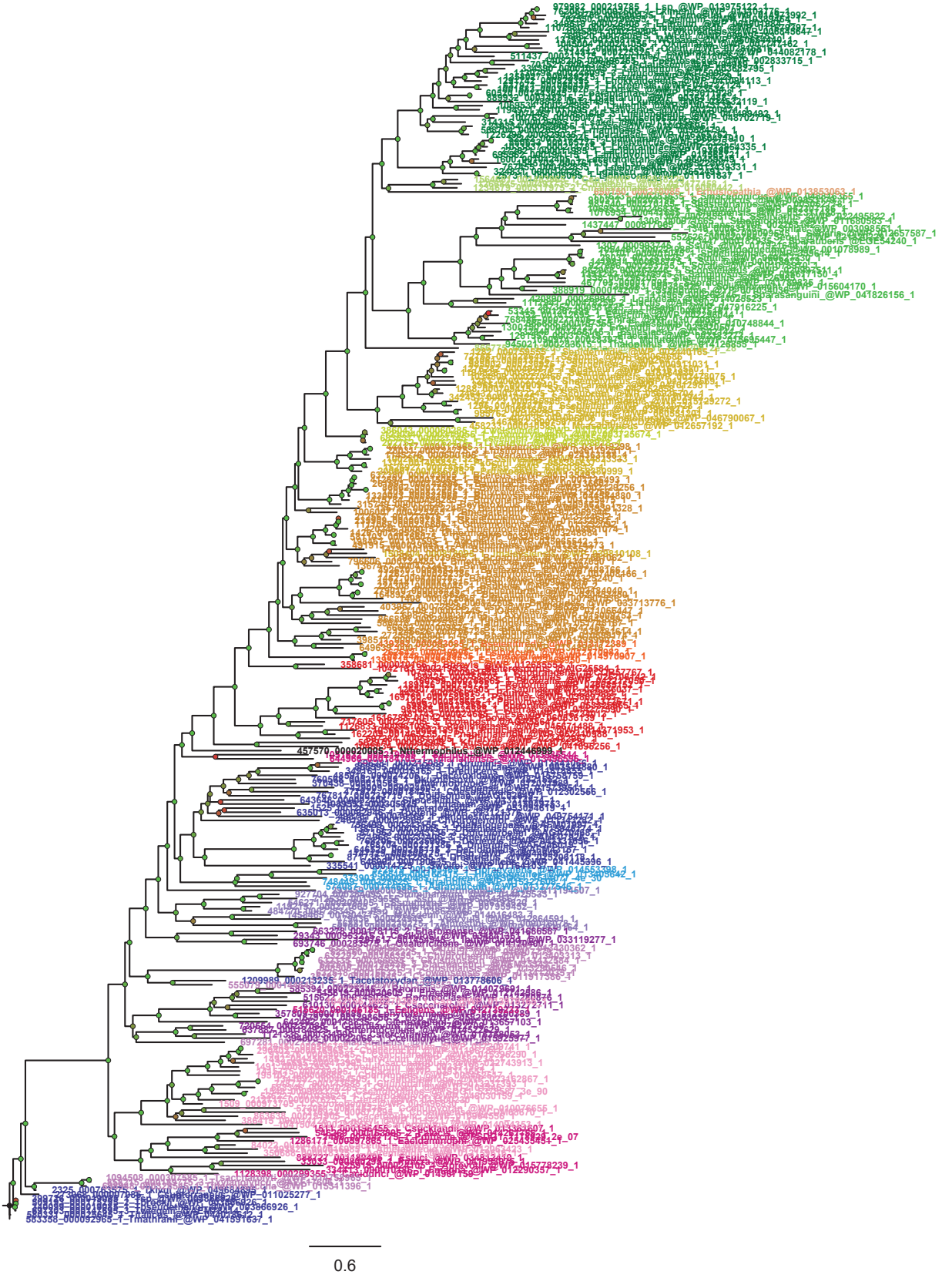

0.4

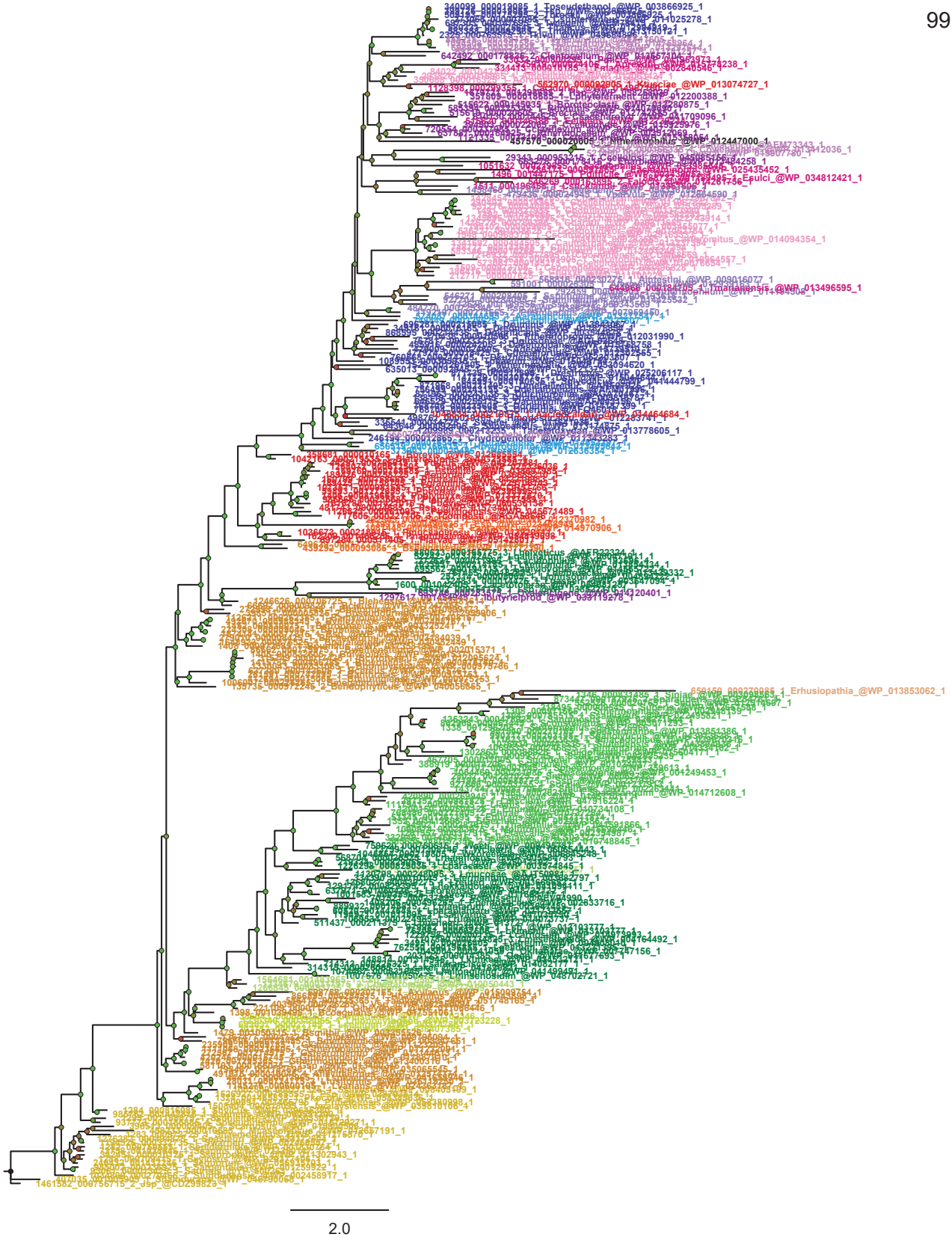

0.8

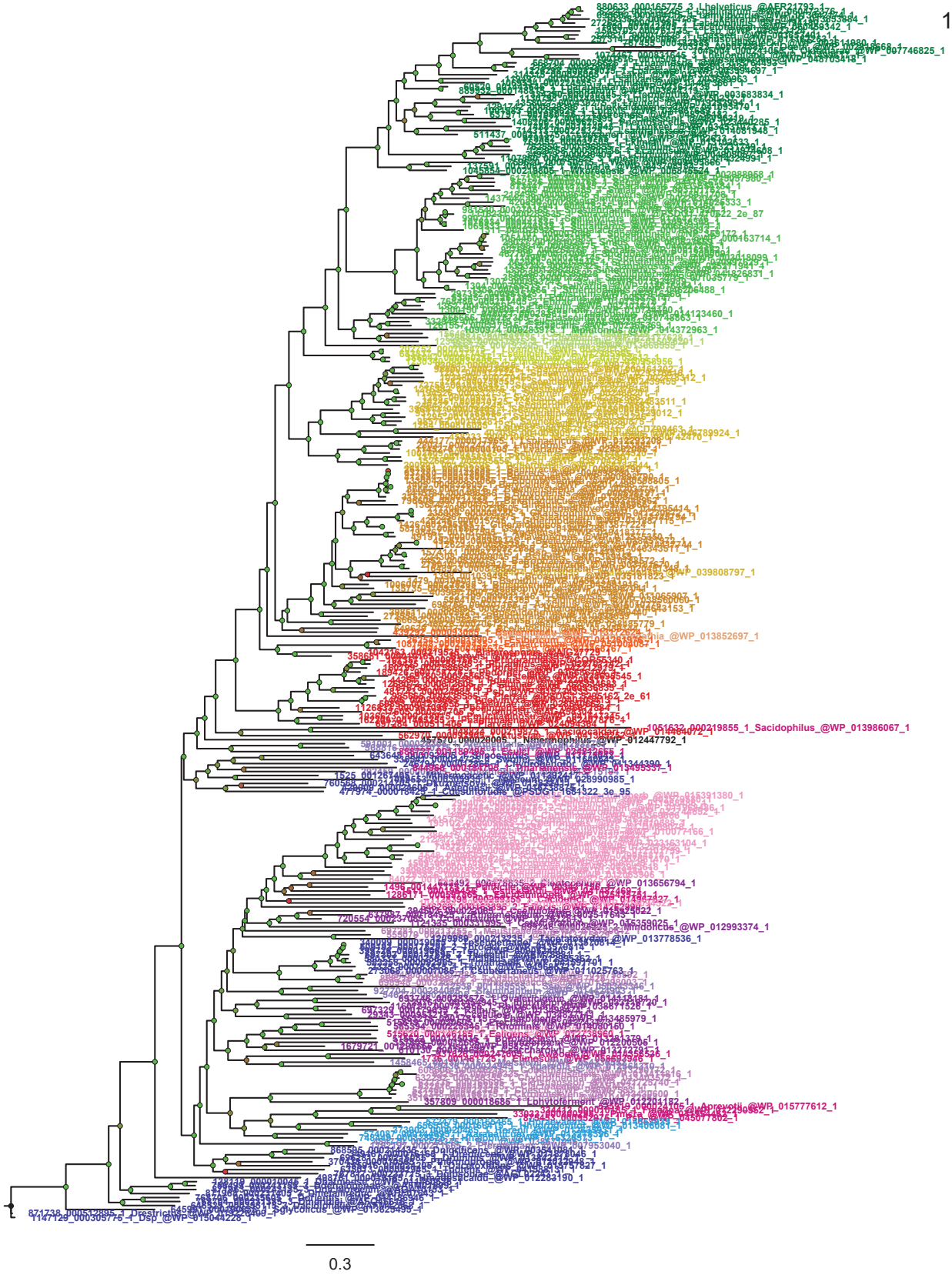

0.3

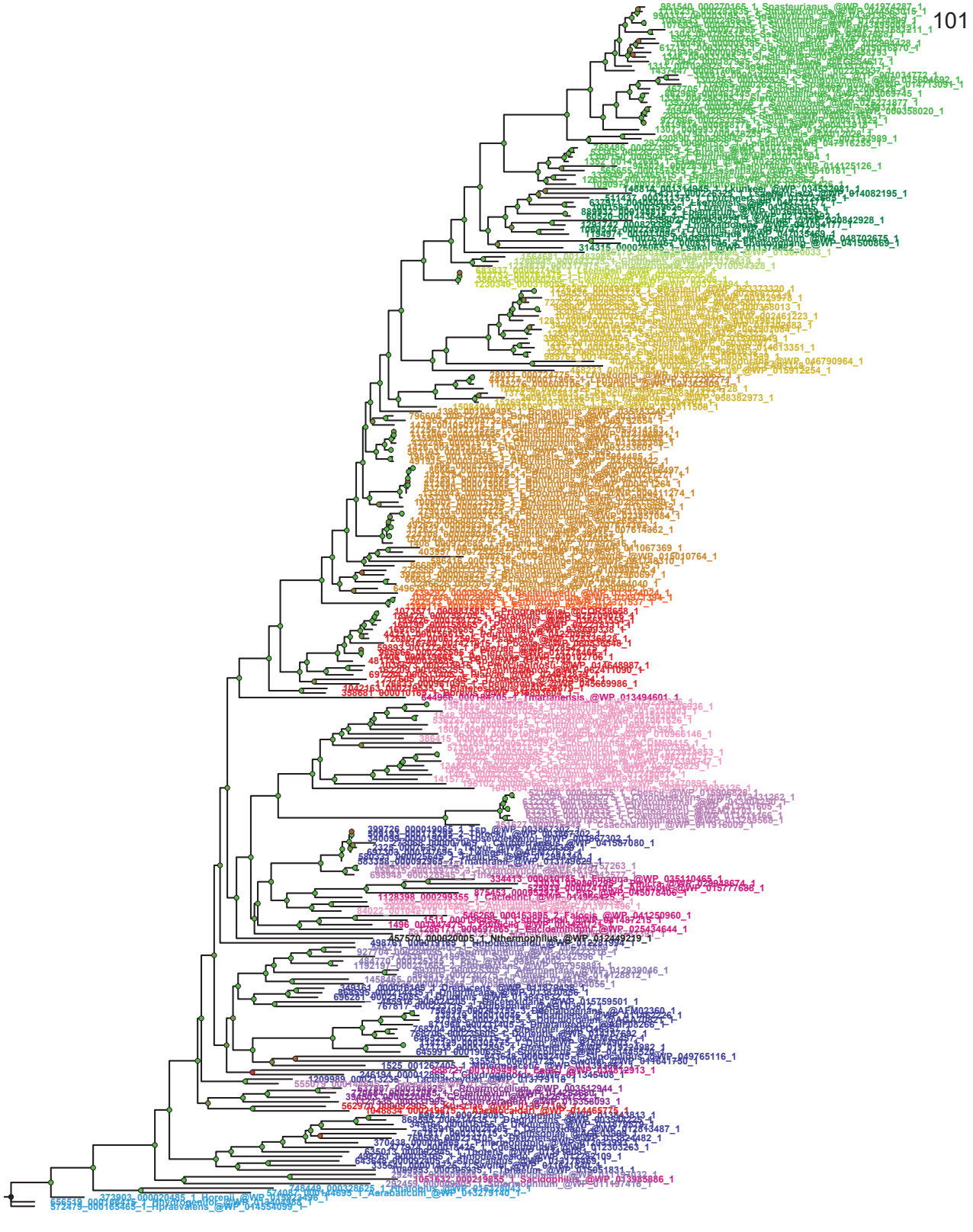

0.4

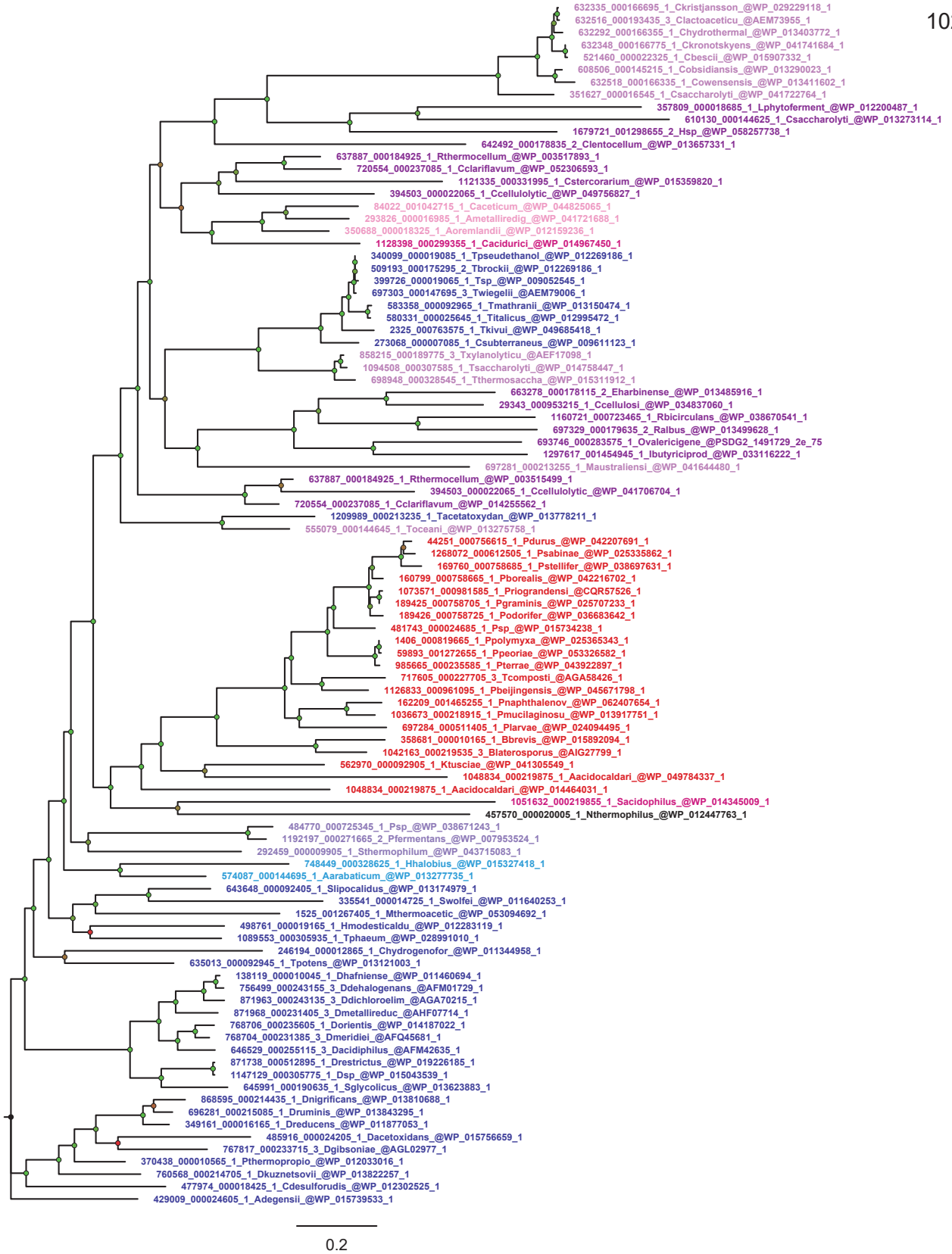

492670_000988345_1_Bvelezensis_@WP_038460721_1
1126211_000262385_1_Bamyloliquefa_@WP_014417632_1
224308_000009045_1_Bsubtilis_@NP_389330_1
1574141_000877815_1_Bsp_@WP_003245743_1
1452_000830075_1_Batrophaeus_@WP_013390499_1
1648923_000876525_1_Bparalichenif_@WP_020451234_1
279010_000008425_1_Blicheniformi_@WP_003181381_1
1408_000972685_1_Bpumilus_@WP_017359401_1
796606_000724485_1_Bmethanolicus_@WP_003349180_1
1367477_000473245_1_Binfantis_@WP_009791505_1
1111068_000236605_1_Gthermoleovor_@WP_011229903_1
235909_000009785_1_Gkaustophilus_@WP_011229903_1
272567_001274575_1_Gstearothermo_@WP_013522948_1
420246_000015745_1_Gthermodenitr_@WP_008881219_1
581103_000166075_1_Gsp_@WP_013401653_1
1426_001295365_1_Gthermoglucos_@WP_003253024_1
1508404_000818095_1_Jmalaysiensis_@WP_039808037_1
866895_000284515_1_Hhalophilus_@WP_014642099_1
698758_000307165_1_Axylanus_@WP_015009334_1
586416_000725365_1_Taidingensis_@WP_038558678_1
403957_000725285_1_Vsp_@WP_019376895_1
221109_000011245_1_Oiheyensis_@WP_011065272_1
649639_000177235_2_Bcellulosilyt_@WP_013487396_1
1246626_000706725_1_Blehensis_@WP_038478270_1
66692_000009825_1_Bclausii_@WP_011246123_1
398511_000005825_2_Bpseudofirmus_@WP_012957508_1
272558_000011145_1_Bhalodurans_@WP_010897016_1
1036673_000218915_1_Pmucilaginosu_@WP_013916344_1
162209_001465255_1_Pnaphthalenov_@WP_062409001_1
1548_000968375_1_Cscatologenes_@WP_029955815_1
536227_001038625_1_Ccarboxidivor_@WP_007064141_1
485916_000024205_1_Dacetoxidans_@WP_015756097_1
888727_001189495_1_Esulci_@WP_034812418_1
1479_001050115_1_Bsmithii_@WP_048622706_1
1398_001039495_1_Bcoagulans_@WP_026684487_1
1526927_000785555_1_Psp_@WP_038703200_1
1374_001465835_1_Pkocurii_@WP_058386024_1
200991_001465795_1_Prifietoensis_@WP_058381370_1
1230340_000318055_1_Lmonocytogene_@WP_003722190_1
386043_000060285_1_Lwelshimeri_@WP_011702513_1
202752_000763475_1_Livanovii_@WP_038407493_1
683837_000027145_1_Lseeligeri_@WP_012985878_1
0.2

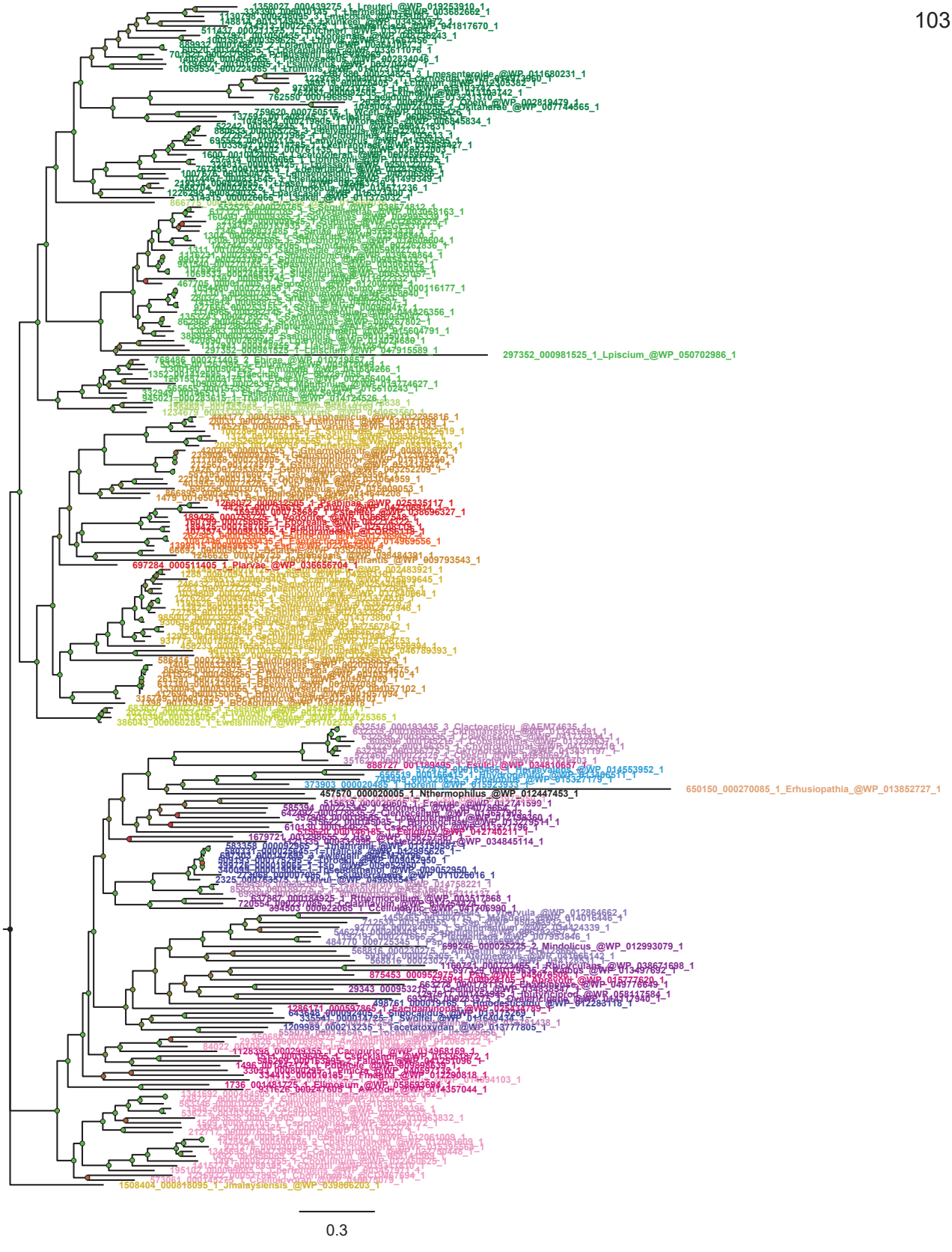

0.6

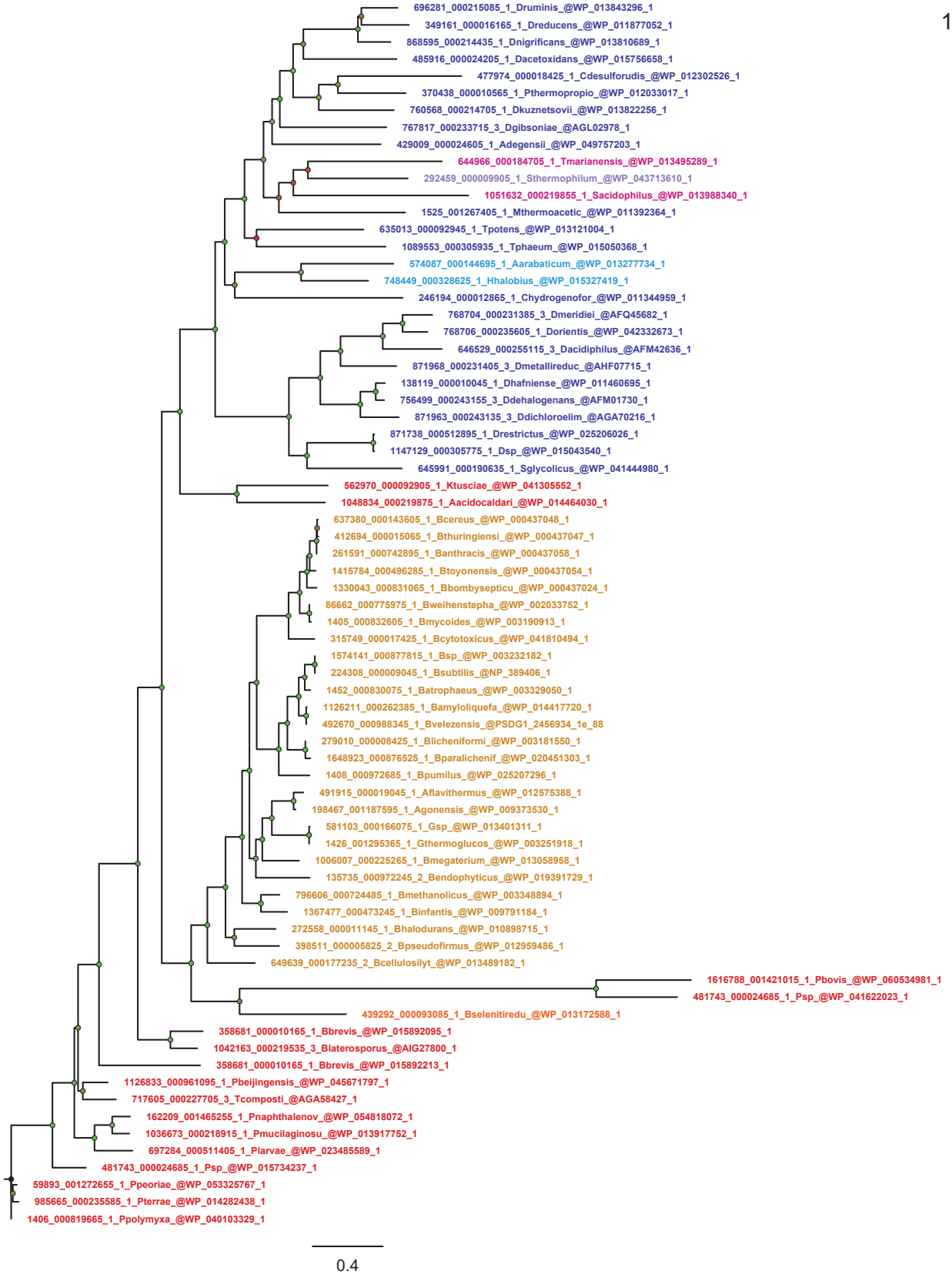

2.0

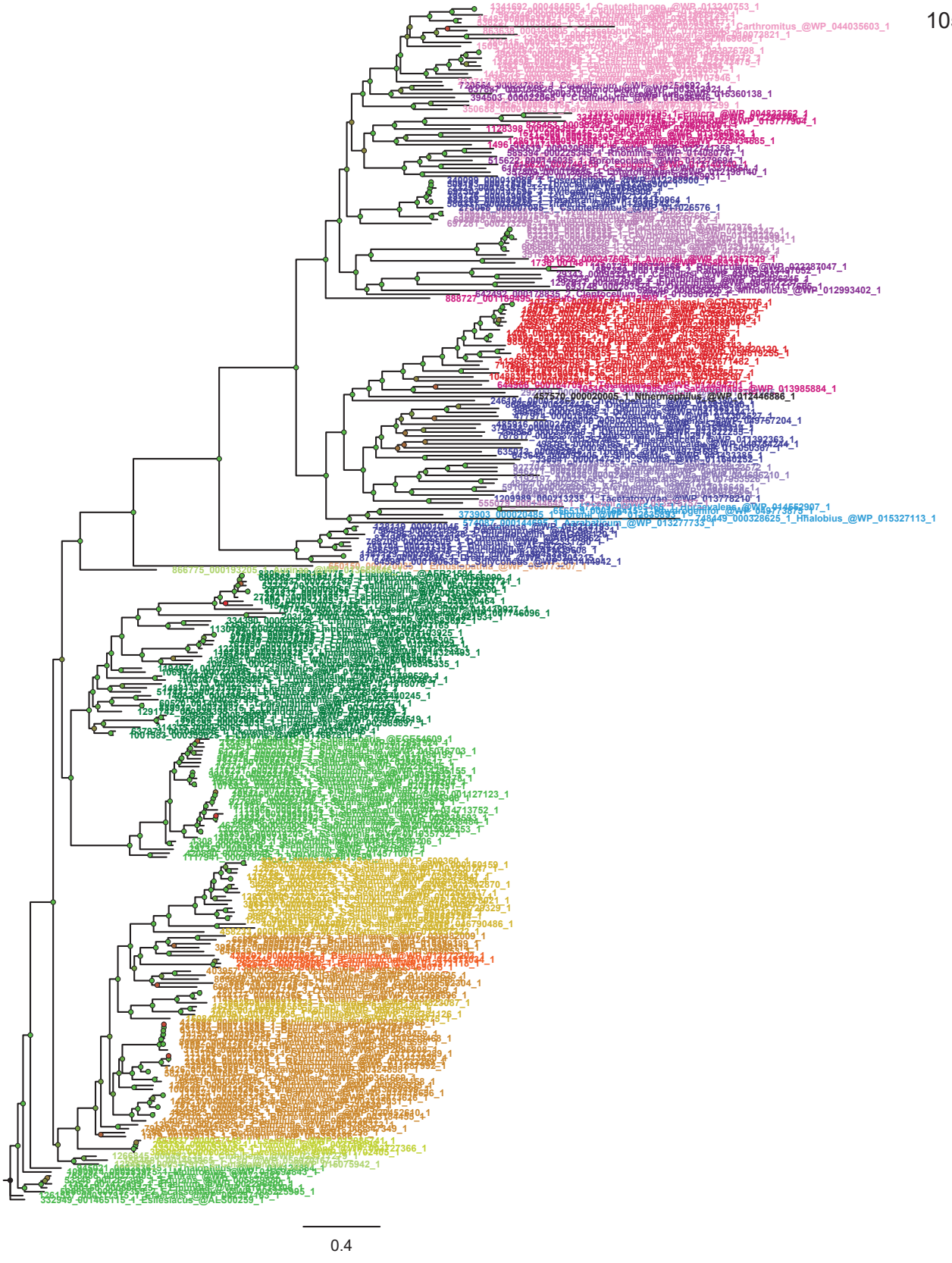

0.3

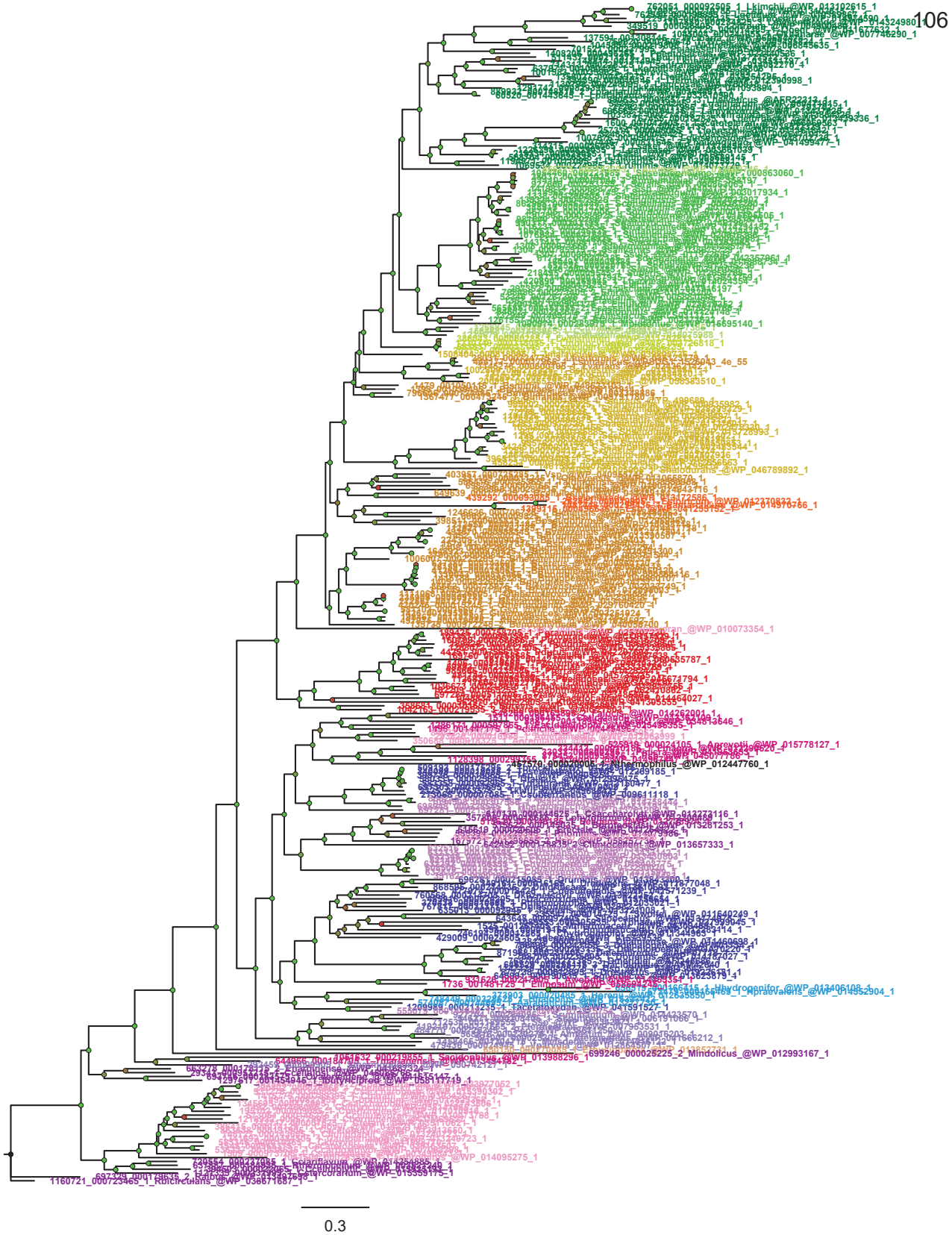

101
0.3

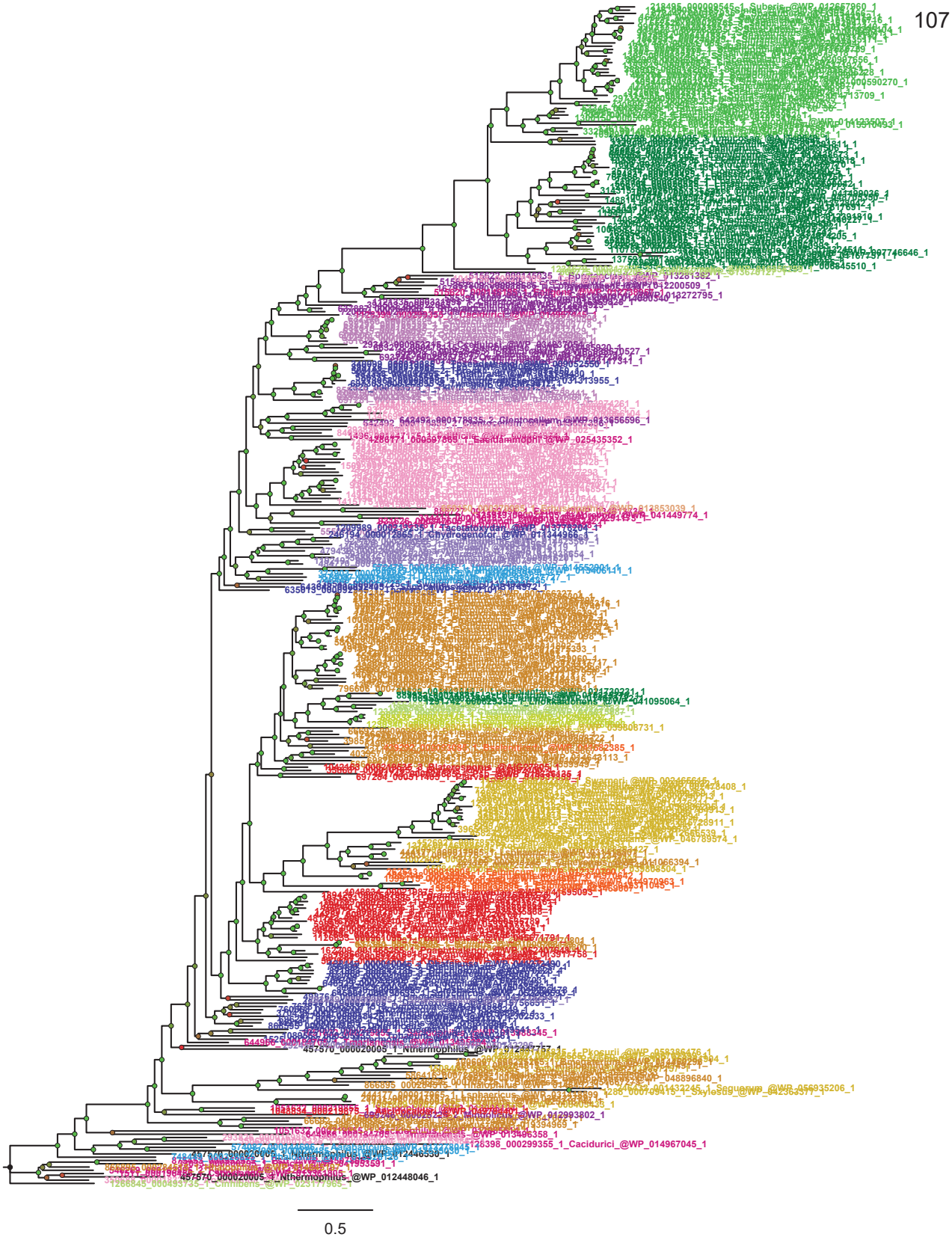

632335_000166695_1_Ckristjansson_@WP_029229118_1
632516_000193435_3_Clactoaceticu_@AEM73955_1
632292_000166355_1_Chydrothermal_@WP_013403772_1
632348_000166775_1_Ckronotskyens_@WP_041741684_1
521460_000022325_1_Cbescii_@WP_015907332_1
608506_000145215_1_Cobsidiansis_@WP_013290023_1
632518_000166335_1_Cowensensis_@WP_013411602_1
351627_000016545_1_Csaccharolyti_@WP_041722764_1
357809_000018685_1_Lphytoferment_@WP_012200487_1
610130_000144625_1_Csaccharolyti_@WP_013273114_1
1679721_001298655_2_Hsp_@WP_058257738_1
642492_000178835_2_Clentocellum_@WP_013657331_1
637887_000184925_1_Rthermocellum_@WP_003517893_1
720554_000237085_1_Cclariflavum_@WP_052306593_1
1121335_000331995_1_Cstercorarium_@WP_015359820_1
394503_000022065_1_Ccellulolytic_@WP_049756827_1
84022_001042715_1_Caceticum_@WP_044825065_1
293826_000016985_1_Ametalliredig_@WP_041721688_1
350688_000018325_1_Aoremlandii_@WP_012159236_1
1128398_000299355_1_Cacidurici_@WP_014967450_1
340099_000019085_1_Tpseudethanol_@WP_012269186_1
509193_000175295_2_Tbrockii_@WP_012269186_1
399726_000019065_1_Tsp_@WP_009052545_1
697303_000147695_3_Twiegelii_@AEM79006_1
583358_000092965_1_Tmathranii_@WP_013150474_1
580331_000025645_1_Titalicus_@WP_012995472_1
2325_000763575_1_Tkivui_@WP_049685418_1
273068_000007085_1_Csubterraneus_@WP_009611123_1
858215_000189775_3_Txylanolyticu_@AEF17098_1
1094508_000307585_1_Tsaccharolyti_@WP_014758447_1
698948_000328545_1_Tthermosaccha_@WP_015311912_1
663278_000178115_2_Eharbinense_@WP_013485916_1
29343_000953215_1_Ccellulosi_@WP_034837060_1
1160721_000723465_1_Rbicirculans_@WP_038670541_1
697329_000179635_2_Ralbus_@WP_013499628_1
693746_000283575_1_Ovalericigene_@PSDG2_1491729_2e-75
1297617_001454945_1_Ibutyriciprod_@WP_033116222_1
697281_000213255_1_Maustraliensi_@WP_041644480_1
637887_000184925_1_Rthermocellum_@WP_003515499_1
394503_000022065_1_Ccellulolytic_@WP_041706704_1
720554_000237085_1_Cclariflavum_@WP_014255562_1
1209989_000213235_1_Tacetatoxydan_@WP_013778211_1
555079_000144645_1_Toceani_@WP_013275758_1
44251_000756615_1_Pdurus_@WP_042207691_1
1268072_000612505_1_Psabinae_@WP_025335862_1
169760_000758685_1_Pstellifer_@WP_038697631_1
160799_000758665_1_Pborealis_@WP_042216702_1
1073571_000981585_1_Priograndensi_@CQR57526_1
189425_000758705_1_Pgraminis_@WP_025707233_1
189426_000758725_1_Podorifer_@WP_036683642_1
481743_000024685_1_Psp_@WP_015734238_1
1406_000819665_1_Ppolymyxa_@WP_025365343_1
59893_001272655_1_Ppeoriae_@WP_053326582_1
985665_000235585_1_Pterrae_@WP_043922897_1
717605_000227705_3_Tcomposti_@AGA58426_1
1126833_000961095_1_Pbeijingensis_@WP_045671798_1
162209_001465255_1_Pnaphthalenov_@WP_062407654_1
1036673_000218915_1_Pmucilaginosu_@WP_013917751_1
697284_000511405_1_Plarvae_@WP_024094495_1
358681_000010165_1_Bbrevis_@WP_015892094_1
1042163_000219535_3_Blaterosporus_@AIG27799_1
562970_000092905_1_Ktusciae_@WP_041305549_1
1048834_000219875_1_Aacidocaldari_@WP_049784337_1
1048834_000219875_1_Aacidocaldari_@WP_014464031_1
1051632_000219855_1_Sacidophilus_@WP_014345009_1
457570_000020005_1_Nthermophilus_@WP_012447763_1
484770_000725345_1_Psp_@WP_038671243_1
1192197_000271665_2_Pfermentans_@WP_007953524_1
292459_000009905_1_Sthermophilum_@WP_043715083_1
748449_000328625_1_Hhalobius_@WP_015327418_1
574087_000144695_1_Aarabaticum_@WP_013277735_1
643648_000092405_1_Slipocalidus_@WP_013174979_1
335541_000014725_1_Swolfei_@WP_011640253_1
1525_001267405_1_Mthermoacetic_@WP_053094692_1
498761_000019165_1_Hmodesticaldu_@WP_012283119_1
1089553_000305935_1_Tphaeum_@WP_028991010_1
246194_000012865_1_Chydrogenofor_@WP_011344958_1
635013_000092945_1_Tpotens_@WP_013121003_1
138119_000010045_1_Dhafniense_@WP_011460694_1
756499_000243155_3_Ddehalogenans_@AFM01729_1
871963_000243135_3_Ddichloroelim_@AGA70215_1
871968_000231405_3_Dmetallireduc_@AHF07714_1
768706_000235605_1_Dorientis_@WP_014187022_1
768704_000231385_3_Dmeridiei_@AFQ45681_1
646529_000255115_3_Dacidiphilus_@AFM42635_1
871738_000512895_1_Drestrictus_@WP_019226185_1
1147129_000305775_1_Dsp_@WP_015043539_1
645991_000190635_1_Sglycolicus_@WP_013623883_1
868595_000214435_1_Dnigrificans_@WP_013810688_1
696281_000215085_1_Druminis_@WP_013843295_1
349161_000016165_1_Dreducens_@WP_011877053_1
485916_000024205_1_Dacetoxidans_@WP_015756659_1
767817_000233715_3_Dgibsoniae_@AGL02977_1
370438_000010565_1_Pthermopropio_@WP_012033016_1
760568_000214705_1_Dkuznetsovii_@WP_013822257_1
477974_000018425_1_Cdesulforudis_@WP_012302525_1
429009_000024605_1_Adegensii_@WP_015739533_1
0.2

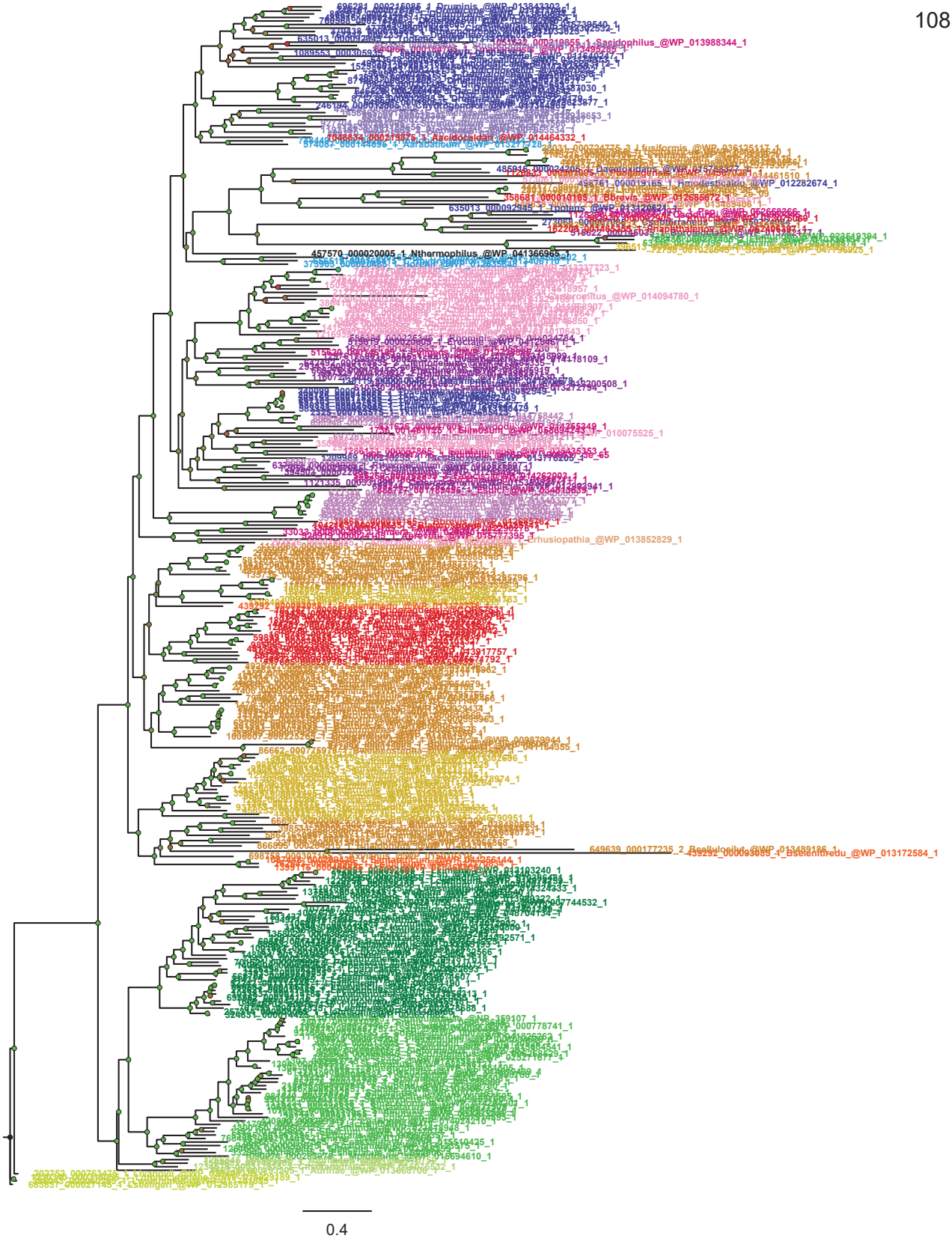

297352_000981525_1_Lpiscium_@WP_050702986_1
650150_000270085_1_Erhusiopathia_@WP_013852727_1
457570_000020005_1_Nthermophilus_@WP_012447453_1
0.3

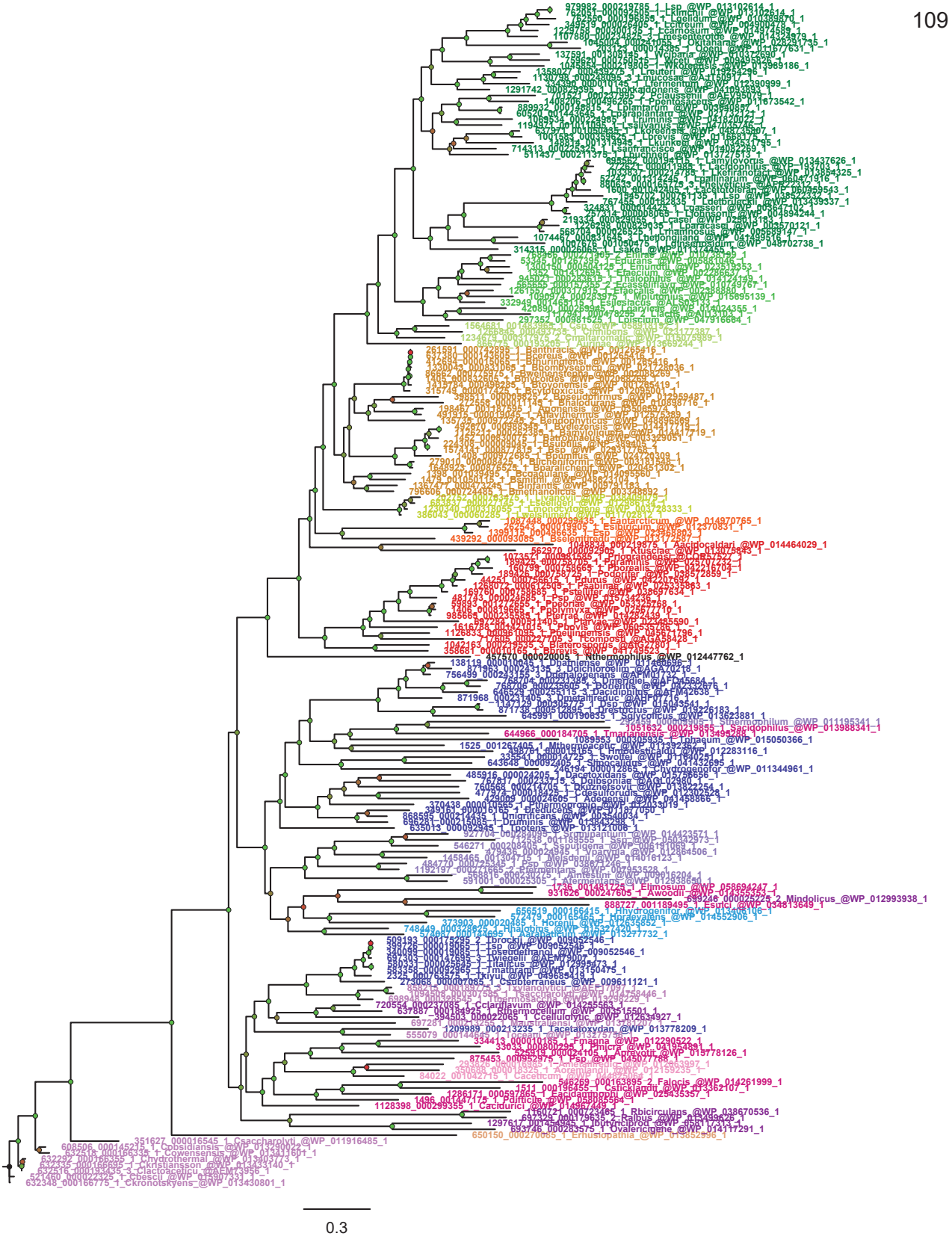

696281_000215085_1_Druminis_@WP_013843296_1
349161_000016165_1_Dreducens_@WP_011877052_1
868595_000214435_1_Dnigrificans_@WP_013810689_1
485916_000024205_1_Dacetoxidans_@WP_015756658_1
477974_000018425_1_Cdesulforudis_@WP_012302526_1
370438_000010565_1_Pthermopropio_@WP_012033017_1
760568_000214705_1_Dkuznetsovii_@WP_013822256_1
767817_000233715_3_Dgibsoniae_@AGL02978_1
429009_000024605_1_Adegensii_@WP_049757203_1
644966_000184705_1_Tmarianensis_@WP_013495289_1
292459_000009905_1_Sthermophilum_@WP_043713610_1
1051632_000219855_1_Sacidophilus_@WP_013988340_1
1525_001267405_1_Mthermoacetic_@WP_011392364_1
635013_000092945_1_Tpotens_@WP_013121004_1
1089553_000305935_1_Tphaeum_@WP_015050368_1
574087_000144695_1_Aarabaticum_@WP_013277734_1
748449_000328625_1_Hhalobius_@WP_015327419_1
246194_000012865_1_Chydrogenofor_@WP_011344959_1
768704_000231385_3_Dmeridiei_@AFQ45682_1
768706_000235605_1_Dorientis_@WP_042332673_1
646529_000255115_3_Dacidiphilus_@AFM42636_1
871968_000231405_3_Dmetallireduc_@AHF07715_1
138119_000010045_1_Dhafniense_@WP_011460695_1
756499_000243155_3_Ddehalogenans_@AFM01730_1
871963_000243135_3_Ddichloroelim_@AGA70216_1
871738_000512895_1_Drestrictus_@WP_025206026_1
1147129_000305775_1_Dsp_@WP_015043540_1
645991_000190635_1_Sglycolicus_@WP_041444980_1
562970_000092905_1_Ktusciae_@WP_041305552_1
1048834_000219875_1_Aacidocaldari_@WP_014464030_1
637380_000143605_1_Bcereus_@WP_000437048_1
412694_000015065_1_Bthuringiensi_@WP_000437047_1
261591_000742895_1_Banthracis_@WP_000437058_1
1415784_000496285_1_Btoyonensis_@WP_000437054_1
1330043_000831065_1_Bbombysepticu_@WP_000437024_1
86662_000775975_1_Bweihenstepha_@WP_002033752_1
1405_000832605_1_Bmycoides_@WP_003190913_1
315749_000017425_1_Bcytotoxicus_@WP_041810494_1
1574141_000877815_1_Bsp_@WP_003232182_1
224308_000009045_1_Bsubtilis_@NP_389406_1
1452_000830075_1_Batrophaeus_@WP_003329050_1
1126211_000262385_1_Bamyloliquefa_@WP_014417720_1
492670_000988345_1_Bvelezensis_@PSDG1_2456934_1e_88
279010_000008425_1_Blicheniformi_@WP_003181550_1
1648923_000876525_1_Bparalichenif_@WP_020451303_1
1408_000972685_1_Bpumilus_@WP_025207296_1
491915_000019045_1_Aflavithermus_@WP_012575388_1
198467_001187595_1_Agonensis_@WP_009373530_1
581103_000166075_1_Gsp_@WP_013401311_1
1426_001295365_1_Gthermoglucos_@WP_003251918_1
1006007_000225265_1_Bmegaterium_@WP_013058958_1
135735_000972245_2_Bendophyticus_@WP_019391729_1
796606_000724485_1_Bmethanolicus_@WP_003348894_1
1367477_000473245_1_Binfantis_@WP_009791184_1
272558_000011145_1_Bhalodurans_@WP_010898715_1
398511_000005825_2_Bpseudofirmus_@WP_012959486_1
649639_000177235_2_Bcellulosilyt_@WP_013489182_1
1616788_001421015_1_Pbovis_@WP_060534981_1
481743_000024685_1_Psp_@WP_041622023_1
439292_000093085_1_Bselenitiredu_@WP_013172588_1
358681_000010165_1_Bbrevis_@WP_015892095_1
1042163_000219535_3_Blaterosporus_@AIG27800_1
358681_000010165_1_Bbrevis_@WP_015892213_1
1126833_000961095_1_Pbeijingensis_@WP_045671797_1
717605_000227705_3_Tcomposti_@AGA58427_1
162209_001465255_1_Pnaphthalenov_@WP_054818072_1
1036673_000218915_1_Pmucilaginosu_@WP_013917752_1
697284_000511405_1_Plarvae_@WP_023485589_1
481743_000024685_1_Psp_@WP_015734237_1
59893_001272655_1_Ppeoriae_@WP_053325767_1
985665_000235585_1_Pterrae_@WP_014282438_1
1406_000819665_1_Ppolymyxa_@WP_040103329_1
0.4

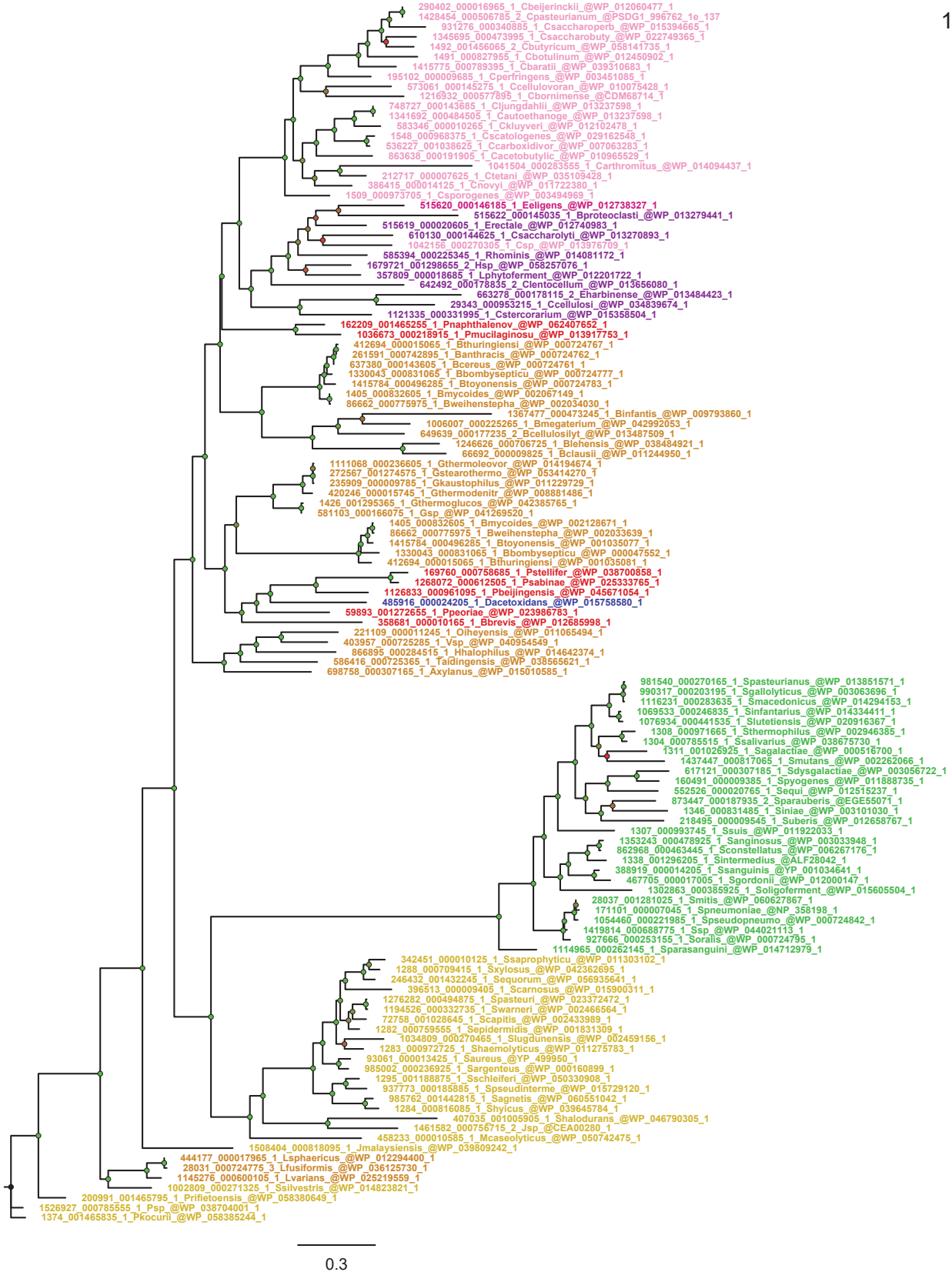

0.4

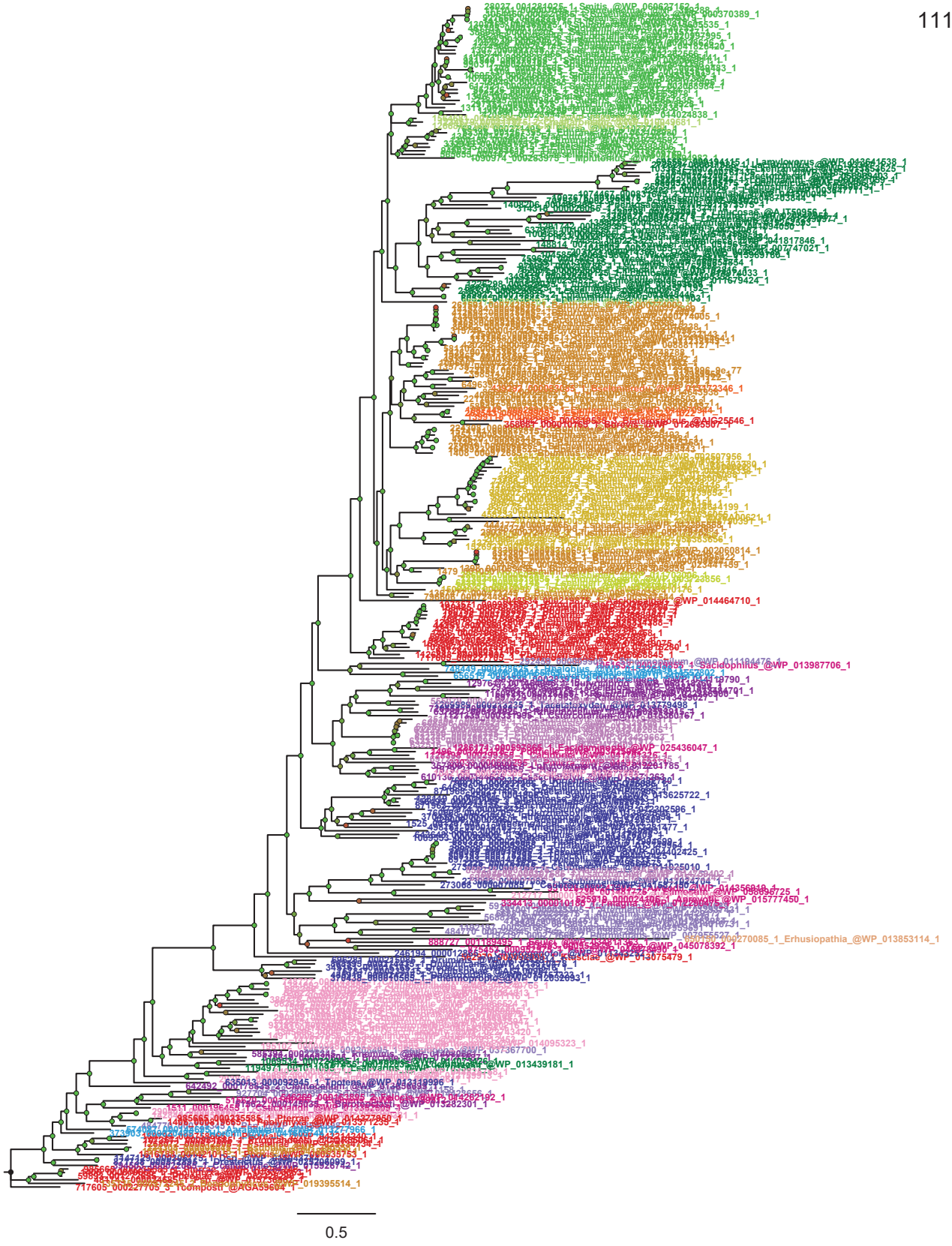

0.3

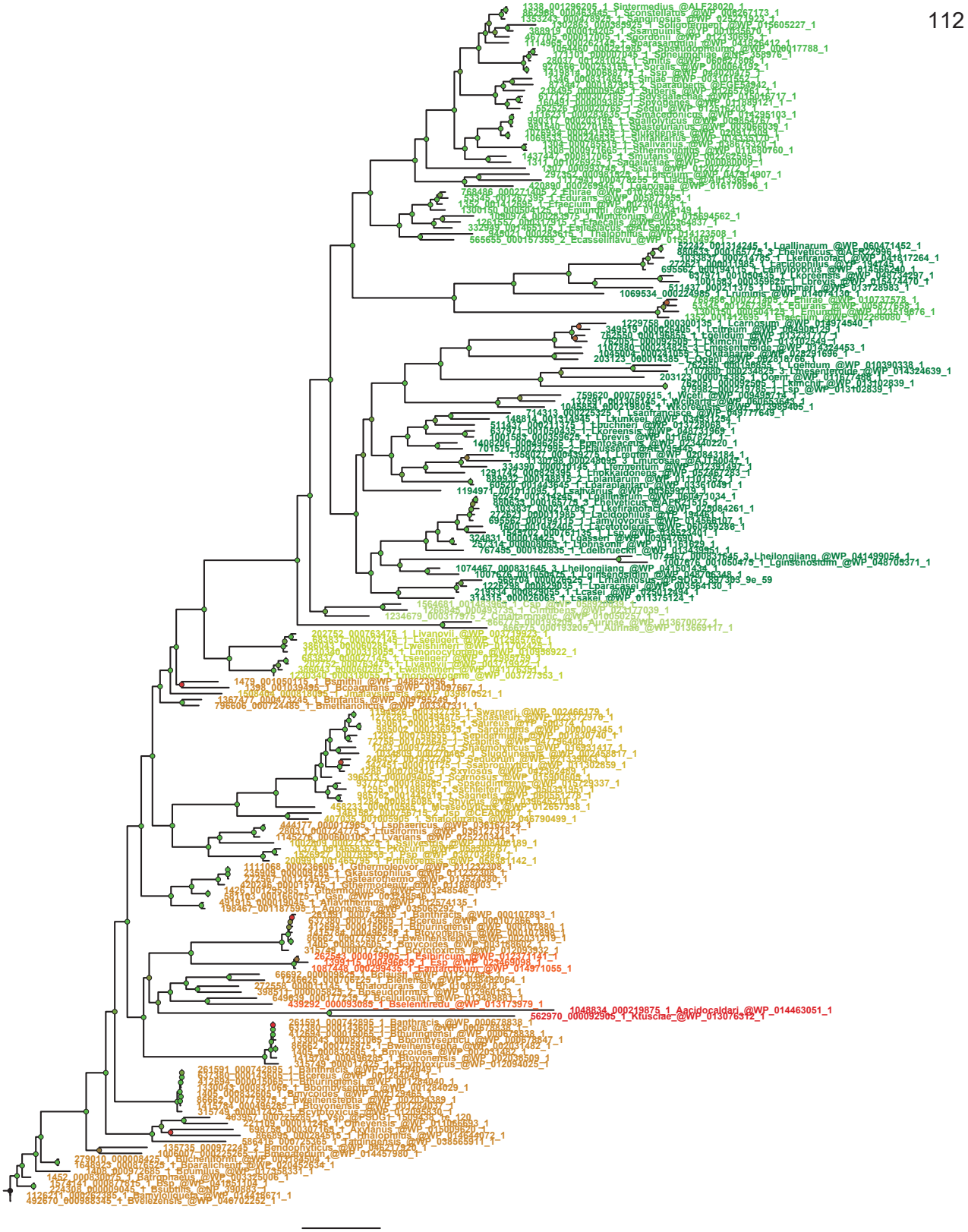

457570_000020005_1_Nthermophilus_@WP_012448046_1
0.5

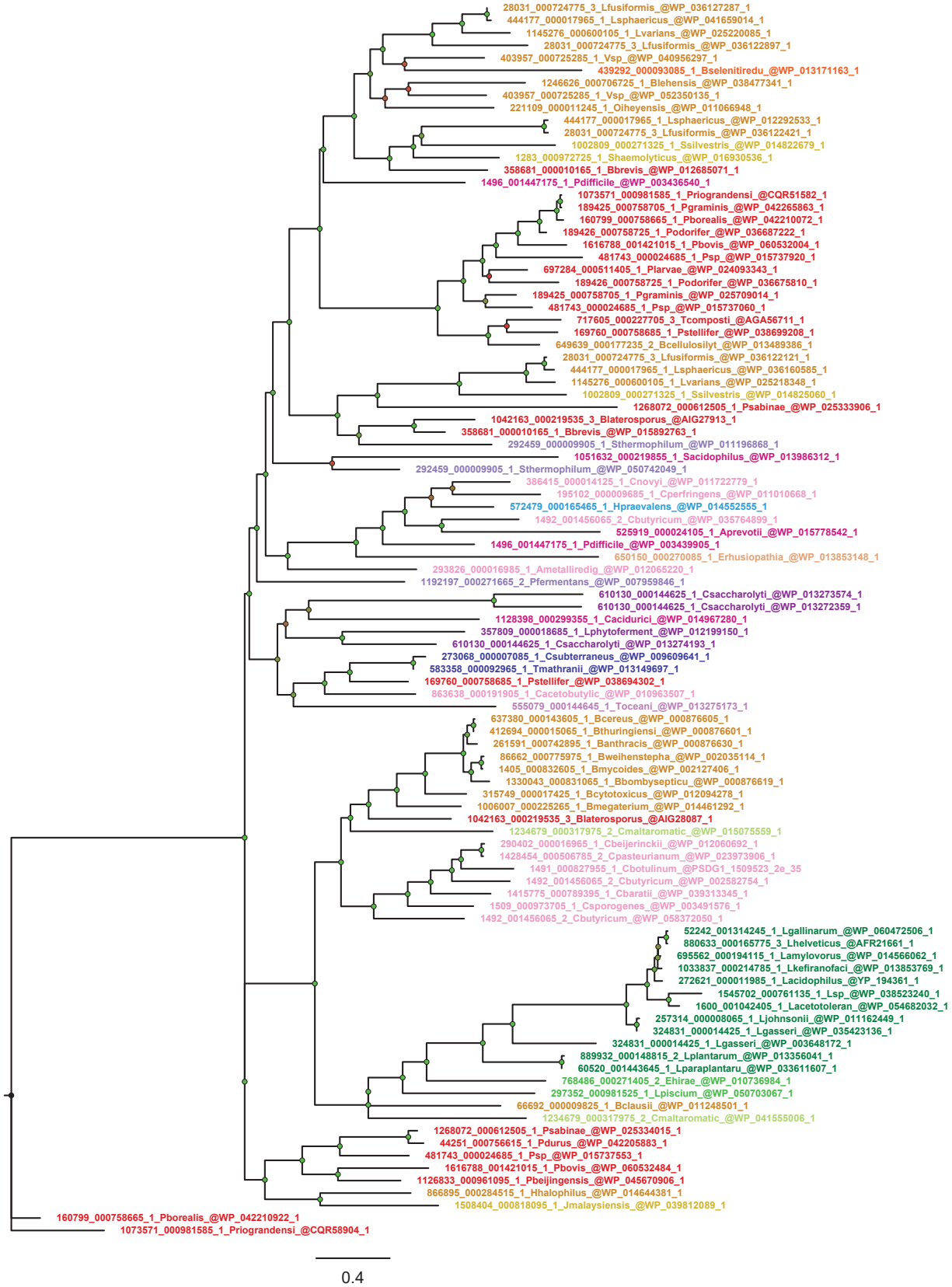

649639_000177235_2_Bcellulosilyt_@WP_013489186_1
439292_000093985_4_Bselenitireedu_@WP_013172584_1
457570_000020005_1_Nthermophilus_@WP_041366955_1
635013_000092945_1_Tpotens_@WP_013489406_1
485916_000024205_1_Paerolubricans_@WP_013172584_1
358681_000010165_1_Bbrevis_@WP_012685872_1
0.4

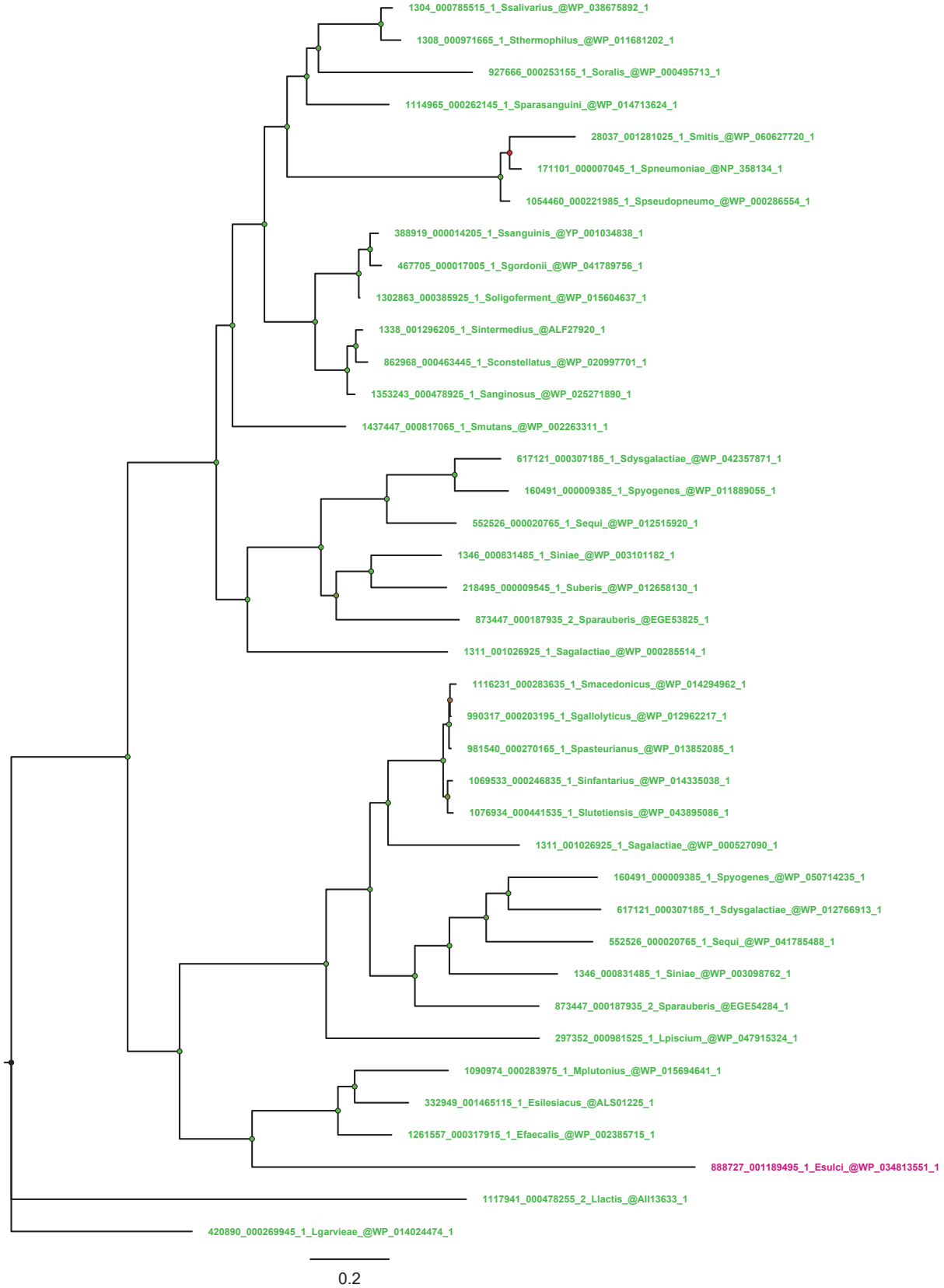

0.3

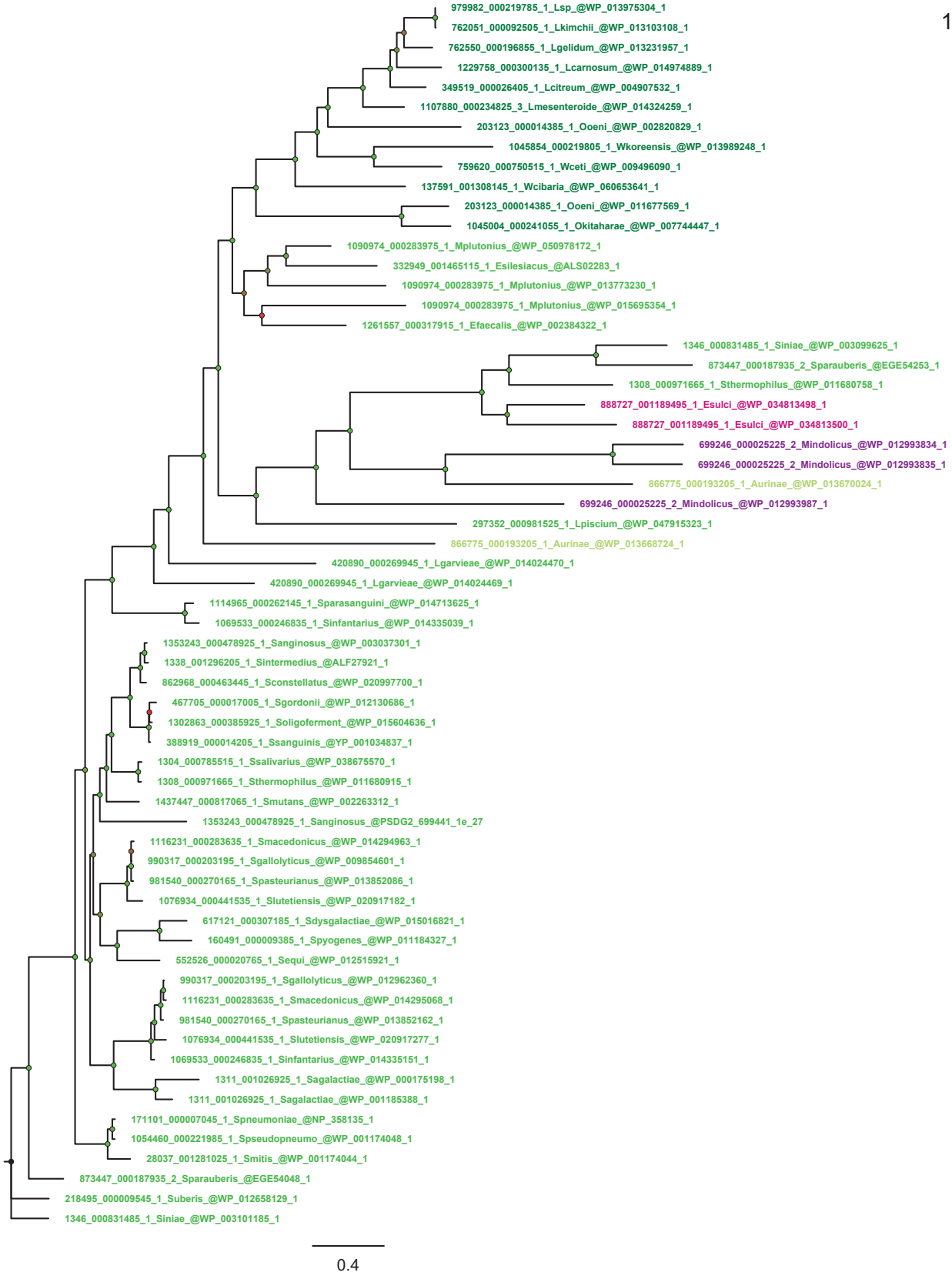

290402_000016965_1_Cbeijerinckii_@WP_012060477_1
1428454_000506785_2_Cpasteurianum_@PSDG1_996762_1e_137
931276_000340885_1_Csaccharoperb_@WP_015394665_1
1345695_000473995_1_Csaccharobuty_@WP_022749365_1
1492_001456065_2_Cbutyricum_@WP_058141735_1
1491_000827955_1_Cbotulinum_@WP_012450902_1
1415775_000789395_1_Cbaratii_@WP_039310683_1
195102_000009685_1_Cperfringens_@WP_003451085_1
573061_000145275_1_Ccellulovoran_@WP_010075428_1
1216932_000577895_1_Cbornimense_@CDM68714_1
748727_000143685_1_Cljungdahlii_@WP_013237598_1
1341692_000484505_1_Cautoethanoge_@WP_013237598_1
583346_000010265_1_Ckluyveri_@WP_012102478_1
1548_000968375_1_Cscatologenes_@WP_029162548_1
536227_001038625_1_Ccarboxidivor_@WP_007063283_1
863638_000191905_1_Cacetobutylic_@WP_010965529_1
1041504_000283555_1_Carthromitus_@WP_014094437_1
212717_000007625_1_Ctetani_@WP_035109428_1
386415_000014125_1_Cnovyi_@WP_011722380_1
1509_000973705_1_Csporogenes_@WP_003494969_1
515620_000146185_1_Eeligens_@WP_012738327_1
515622_000145035_1_Bproteoclasti_@WP_013279441_1
515619_000020605_1_Erectale_@WP_012740983_1
610130_000144625_1_Csaccharolyti_@WP_013270893_1
1042156_000270305_1_Csp_@WP_013976709_1
585394_000225345_1_Rhominis_@WP_014081172_1
1679721_001298655_2_Hsp_@WP_058257076_1
357809_000018685_1_Lphytoferment_@WP_012201722_1
642492_000178835_2_Clentocellum_@WP_013656080_1
663278_000178115_2_Eharbinense_@WP_013484423_1
29343_000953215_1_Ccelluliosi_@WP_034839674_1
1121335_000331995_1_Cstercorarium_@WP_015358504_1
162209_001465255_1_Pnaphthalenov_@WP_062407652_1
1036673_000218915_1_Pmucilaginosu_@WP_013917753_1
412694_000015065_1_Bthuringiensi_@WP_000724767_1
261591_000742895_1_Banthracis_@WP_000724762_1
637380_000143605_1_Bcereus_@WP_000724761_1
1330043_000831065_1_Bbombysepticu_@WP_000724777_1
1415784_000496285_1_Btoyonensis_@WP_000724783_1
1405_000832605_1_Bmycoides_@WP_002067149_1
86662_000775975_1_Bweihenstepha_@WP_002034030_1
1367477_000473245_1_Binfantis_@WP_009793860_1
1006007_000225265_1_Bmegaterium_@WP_042992053_1
649639_000177235_2_Bcellulosilyt_@WP_013487509_1
1246626_000706725_1_Blehensis_@WP_038484921_1
66692_000009825_1_Bclausii_@WP_011244950_1
1111068_000236605_1_Gthermoleovor_@WP_014194674_1
272567_001274575_1_Gstearothermo_@WP_053414270_1
235909_000009785_1_Gkaustophilus_@WP_011229729_1
420246_000015745_1_Gthermodenitr_@WP_008881486_1
1426_001295365_1_Gthermoglucos_@WP_042385765_1
581103_000166075_1_Gsp_@WP_041269520_1
1405_000832605_1_Bmycoides_@WP_002128671_1
86662_000775975_1_Bweihenstepha_@WP_002033639_1
1415784_000496285_1_Btoyonensis_@WP_001035077_1
1330043_000831065_1_Bbombysepticu_@WP_000047552_1
412694_000015065_1_Bthuringiensi_@WP_001035081_1
169760_000758685_1_Pstellifer_@WP_038700858_1
1268072_000612505_1_Psabinae_@WP_025333765_1
1126833_000961095_1_Pbeijingensis_@WP_045671054_1
485916_000024205_1_Dacetoxidans_@WP_015758580_1
59893_001272655_1_Ppeoriae_@WP_023986783_1
358681_000010165_1_Bbrevis_@WP_012685998_1
221109_000011245_1_Oiheyensis_@WP_011065494_1
403957_000725285_1_Vsp_@WP_040954549_1
866895_000284515_1_Hhalophilus_@WP_014642374_1
586416_000725365_1_Taidingensis_@WP_038565621_1
698758_000307165_1_Axylanus_@WP_015010585_1
981540_000270165_1_Spasteurianus_@WP_013851571_1
990317_000203195_1_Sgallolyticus_@WP_003063696_1
1116231_000283635_1_Smacedonicus_@WP_014294153_1
1069533_000246835_1_Sinfantarius_@WP_014334411_1
1076934_000441535_1_Slutetiensis_@WP_020916367_1
1308_000971665_1_Sthermophilus_@WP_002946385_1
1304_000785515_1_Ssalivarius_@WP_038675730_1
1311_001026925_1_Sagalactiae_@WP_000516700_1
1437447_000817065_1_Smutans_@WP_002262066_1
617121_000307185_1_Sdysgalactiae_@WP_003056722_1
160491_000009385_1_Spyogenes_@WP_011888735_1
552526_000020765_1_Sequi_@WP_012515237_1
873447_000187935_2_Sparauberis_@EGE55071_1
1346_000831485_1_Siniae_@WP_003101030_1
218495_000009545_1_Suberis_@WP_012658767_1
1307_000993745_1_Ssuis_@WP_011922033_1
1353243_000478925_1_Sanginosus_@WP_003033948_1
862968_000463445_1_Sconstellatus_@WP_006267176_1
1338_001296205_1_Sintermedius_@ALF28042_1
388919_000014205_1_Ssanguinis_@YP_001034641_1
467705_000017005_1_Sgordonii_@WP_012000147_1
1302863_000385925_1_Soligoferment_@WP_015605504_1
28037_001281025_1_Smitis_@WP_060627867_1
171101_000007045_1_Spneumoniae_@NP_358198_1
1054460_000221985_1_Spseudopneumo_@WP_000724842_1
1419814_000688775_1_Ssp_@WP_044021113_1
927666_000253155_1_Soralis_@WP_000724795_1
1114965_000262145_1_Sparasanguini_@WP_014712979_1
342451_000010125_1_Ssaprophyticu_@WP_011303102_1
1288_000709415_1_Sxylosus_@WP_042362695_1
246432_001432245_1_Sequorum_@WP_056935641_1
396513_000009405_1_Scarnosus_@WP_015900311_1
1276282_000494875_1_Spasteuri_@WP_023372472_1
1194526_000332735_1_Swarneri_@WP_002466564_1
72758_001028645_1_Scapitis_@WP_002433989_1
1282_000759555_1_Sepidermidis_@WP_001831309_1
1034809_000270465_1_Slugdunensis_@WP_002459156_1
1283_000972725_1_Shaemolyticus_@WP_011275783_1
93061_000013425_1_Saureus_@YP_499950_1
985002_000236925_1_Sargenteus_@WP_000160899_1
1295_001188875_1_Sschleiferi_@WP_050330908_1
937773_000185885_1_Spseudinterme_@WP_015729120_1
985762_001442815_1_Sagnetis_@WP_060551042_1
1284_000816085_1_Shyicus_@WP_039645784_1
407035_001005905_1_Shalodurans_@WP_046790305_1
1461582_000756715_2_Jsp_@CEA00280_1
458233_000010585_1_Mcaseolyticus_@WP_050742475_1
1508404_000818095_1_Jmalaysiensis_@WP_039809242_1
444177_000017965_1_Lsphaericus_@WP_012294400_1
28031_000724775_3_Lfusiformis_@WP_036125730_1
1145276_000600105_1_Lvarians_@WP_025219559_1
1002809_0002271325_1_Ssilvestris_@WP_014823821_1
200991_001465795_1_Prifietoensis_@WP_058380649_1
1526927_000785555_1_Psp_@WP_038704001_1
1374_001465835_1_Pkocurii_@WP_058385244_1
0.3

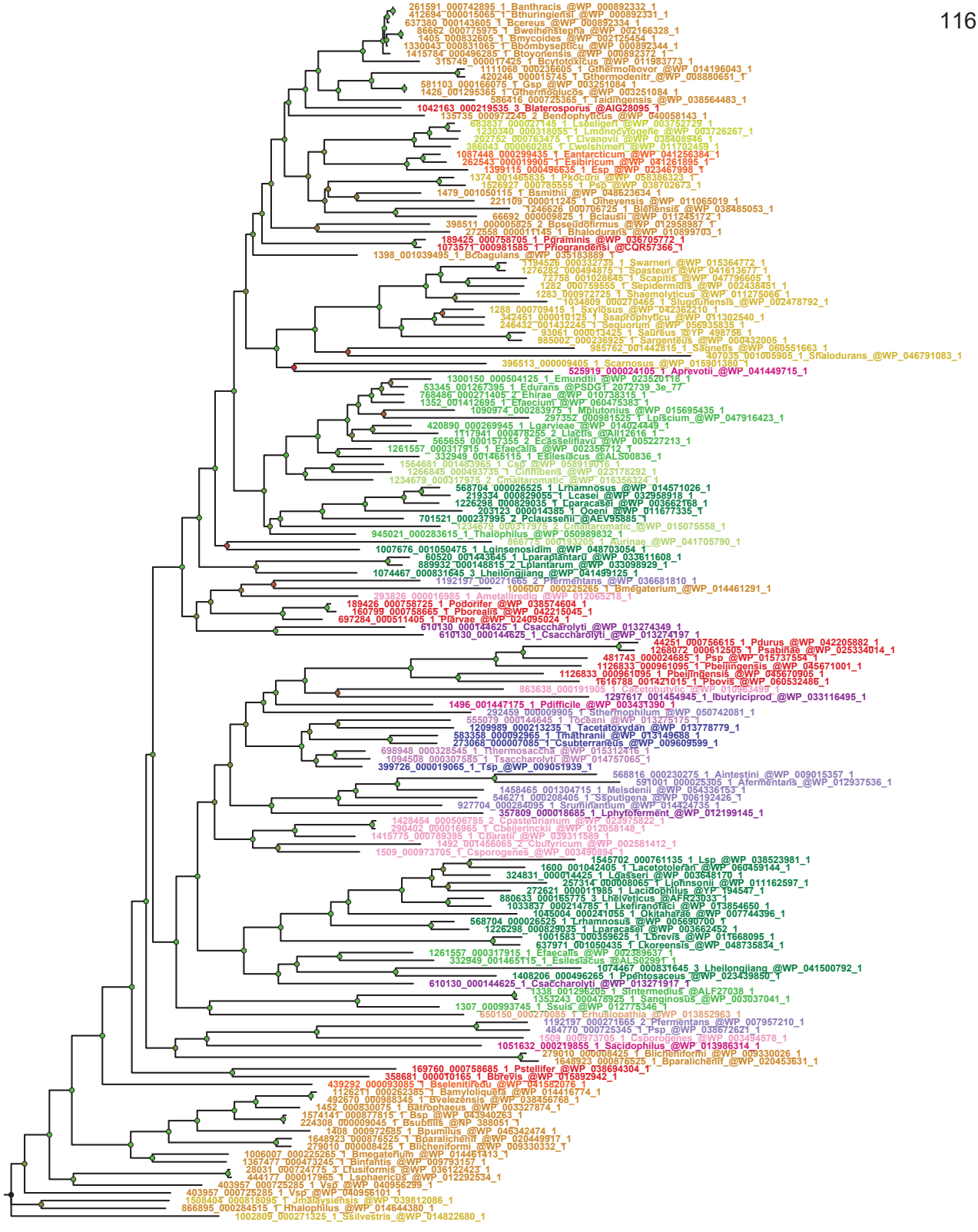

0.5

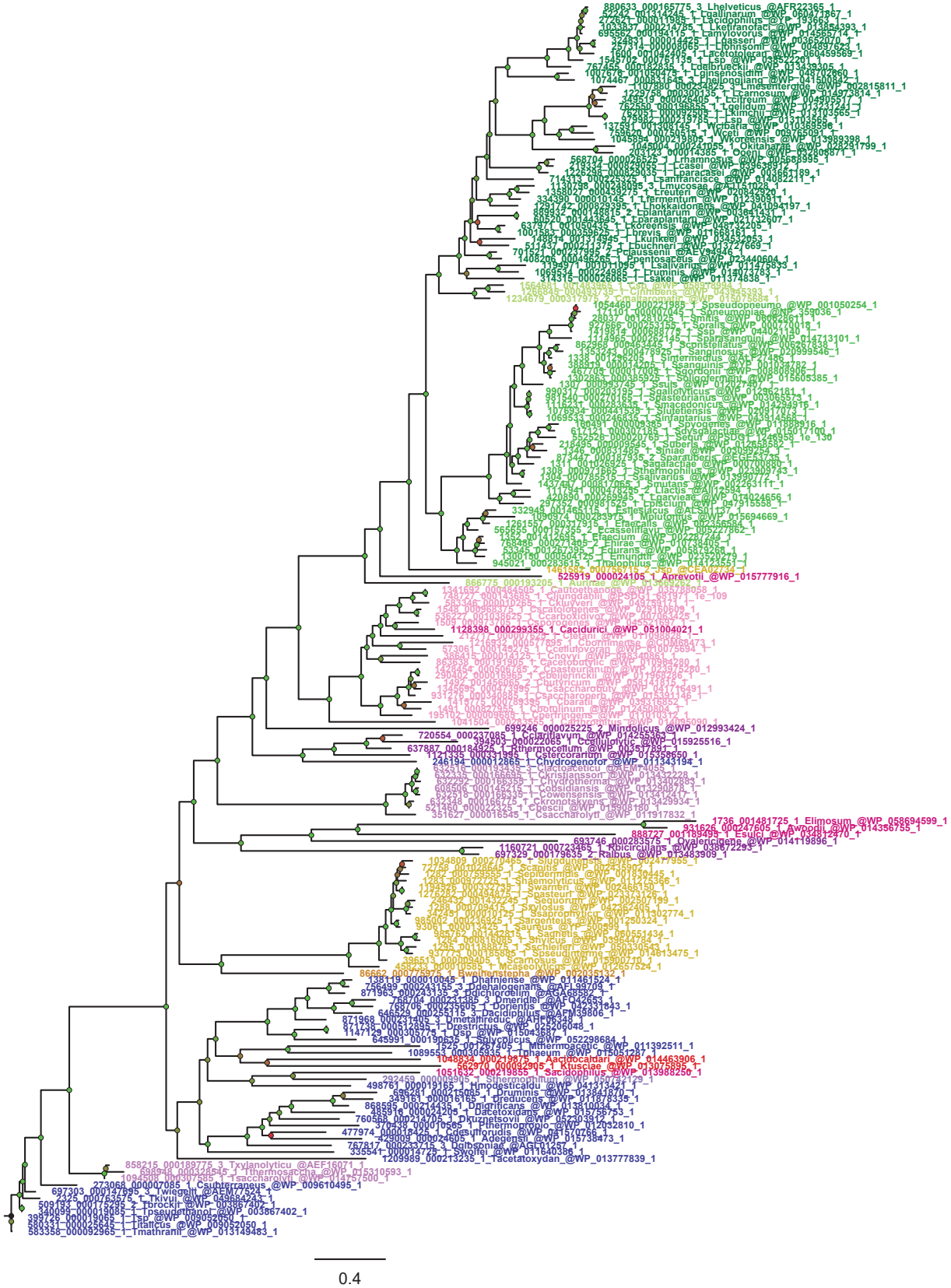

0.7

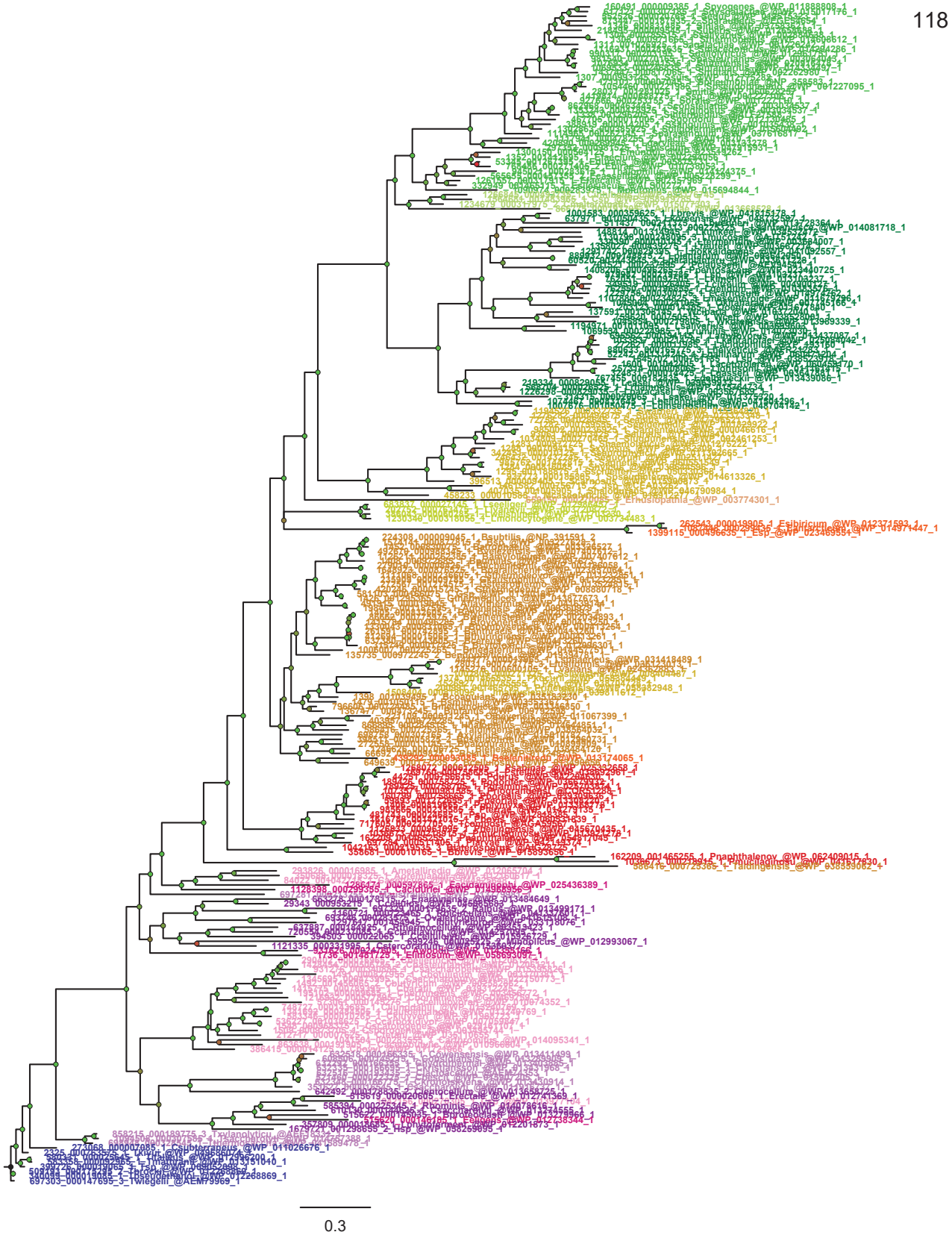

28031_000724775_3_Lfusiformis_@WP_036127287_1
444177_000017965_1_Lsphaericus_@WP_041659014_1
1145276_000600105_1_Lvarians_@WP_025220085_1
28031_000724775_3_Lfusiformis_@WP_036122897_1
403957_000725285_1_Vsp_@WP_040956297_1
439292_000093085_1_Bselenitiredu_@WP_013171163_1
1246626_000706725_1_Blehensis_@WP_038477341_1
403957_000725285_1_Vsp_@WP_052350135_1
221109_000011245_1_Oiheyensis_@WP_011066948_1
444177_000017965_1_Lsphaericus_@WP_012292533_1
28031_000724775_3_Lfusiformis_@WP_036122421_1
1002809_000271325_1_Ssilvestris_@WP_014822679_1
1283_000972725_1_Shaemolyticus_@WP_016930536_1
358681_000010165_1_Bbrevis_@WP_012685071_1
1496_001447175_1_Pdifficile_@WP_003436540_1
1073571_000981585_1_Priograndensi_@CQR51582_1
189425_000758705_1_Pgraminis_@WP_042265863_1
160799_000758665_1_Pborealis_@WP_042210072_1
189426_000758725_1_Podorifer_@WP_036687222_1
1616788_001421015_1_Pbovis_@WP_060532004_1
481743_000024685_1_Psp_@WP_015737920_1
697284_000511405_1_Plarvae_@WP_024093343_1
189426_000758725_1_Podorifer_@WP_036675810_1
189425_000758705_1_Pgraminis_@WP_025709014_1
481743_000024685_1_Psp_@WP_015737060_1
717605_000227705_3_Tcomposti_@AGA56711_1
169760_000758685_1_Pstellifer_@WP_038699208_1
649639_000177235_2_Bcellulosilyt_@WP_013489386_1
28031_000724775_3_Lfusiformis_@WP_036122121_1
444177_000017965_1_Lsphaericus_@WP_036160585_1
1145276_000600105_1_Lvarians_@WP_025218348_1
1002809_000271325_1_Ssilvestris_@WP_014825060_1
1268072_000612505_1_Psabinae_@WP_025333906_1
1042163_000219535_3_Blaterosporus_@AIG27913_1
358681_000010165_1_Bbrevis_@WP_015892763_1
292459_000009905_1_Sthermophilum_@WP_011196868_1
1051632_000219855_1_Sacidophilus_@WP_013986312_1
292459_000009905_1_Sthermophilum_@WP_050742049_1
386415_000014125_1_Cnovyi_@WP_011722779_1
195102_000009685_1_Cperfringens_@WP_011010668_1
572479_000165465_1_Hpraevalens_@WP_014552555_1
1492_001456065_2_Cbutyricum_@WP_035764899_1
525919_000024105_1_Aprevotii_@WP_015778542_1
1496_001447175_1_Pdifficile_@WP_003439905_1
650150_000270085_1_Erhusiopathia_@WP_013853148_1
293826_000016985_1_Ametalliredig_@WP_012065220_1
1192197_000271665_2_Pfermentans_@WP_007959846_1
610130_000144625_1_Csaccharolyti_@WP_013273574_1
610130_000144625_1_Csaccharolyti_@WP_013272359_1
1128398_000299355_1_Cacidurici_@WP_014967280_1
357809_000018685_1_Lphytoferment_@WP_012199150_1
610130_000144625_1_Csaccharolyti_@WP_013274193_1
273068_000007085_1_Csubterraneus_@WP_009609641_1
583358_000092965_1_Tmathranii_@WP_013149697_1
169760_000758685_1_Pstellifer_@WP_038694302_1
863638_000191905_1_Cacetobutylic_@WP_010963507_1
555079_000144645_1_Toceani_@WP_013275173_1
637380_000143605_1_Bcereus_@WP_000876605_1
412694_000015065_1_Bthuringiensi_@WP_000876601_1
261591_000742895_1_Banthracis_@WP_000876630_1
86662_000775975_1_Bweihenstepha_@WP_002035114_1
1405_000832605_1_Bmycoides_@WP_002127406_1
1330043_000831065_1_Bbombysepticu_@WP_000876619_1
315749_000017425_1_Bcytotoxicus_@WP_012094278_1
1006007_000225265_1_Bmegaterium_@WP_014461292_1
1042163_000219535_3_Blaterosporus_@AIG28087_1
1234679_000317975_2_Cmaltaromatic_@WP_015075559_1
290402_000016965_1_Cbeijerinckii_@WP_012060692_1
1428454_000506785_2_Cpasteurianum_@WP_023973906_1
1491_000827955_1_Cbotulinum_@PSDG1_1509523_2e_35
1492_001456065_2_Cbutyricum_@WP_002582754_1
1415775_000789395_1_Cbaratii_@WP_039313345_1
1509_000973705_1_Csporogenes_@WP_003491576_1
1492_001456065_2_Cbutyricum_@WP_058372050_1
52242_001314245_1_Lgallinarum_@WP_060472506_1
880633_000165775_3_Lhelveticus_@AFR21661_1
695562_000194115_1_Lamylovorus_@WP_014566062_1
1033837_000214785_1_Lkefiranofaci_@WP_013853769_1
272621_000011985_1_Lacidophilus_@YP_194361_1
1545702_000761135_1_Lsp_@WP_038523240_1
1600_001042405_1_Lacetotoleran_@WP_054682032_1
257314_000008065_1_Ljohnsonii_@WP_011162449_1
324831_000014425_1_Lgasseri_@WP_035423136_1
324831_000014425_1_Lgasseri_@WP_003648172_1
889932_000148815_2_Lplantarum_@WP_013356041_1
60520_001443645_1_Lparaplantaru_@WP_033611607_1
768486_000271405_2_Ehirae_@WP_010736984_1
297352_000981525_1_Lpiscium_@WP_050703067_1
66692_000009825_1_Bclausii_@WP_011248501_1
1234679_000317975_2_Cmaltaromatic_@WP_041555006_1
1268072_000612505_1_Psabinae_@WP_025334015_1
44251_000756615_1_Pdurus_@WP_042205883_1
481743_000024685_1_Psp_@WP_015737553_1
1616788_001421015_1_Pbovis_@WP_060532484_1
1126833_000961095_1_Pbeijingensis_@WP_045670906_1
866895_000284515_1_Hhalophilus_@WP_014644381_1
1508404_000818095_1_Jmalaysiensis_@WP_039812089_1
160799_000758665_1_Pborealis_@WP_042210922_1
1073571_000981585_1_Priograndensi_@CQR58904_1
0.4

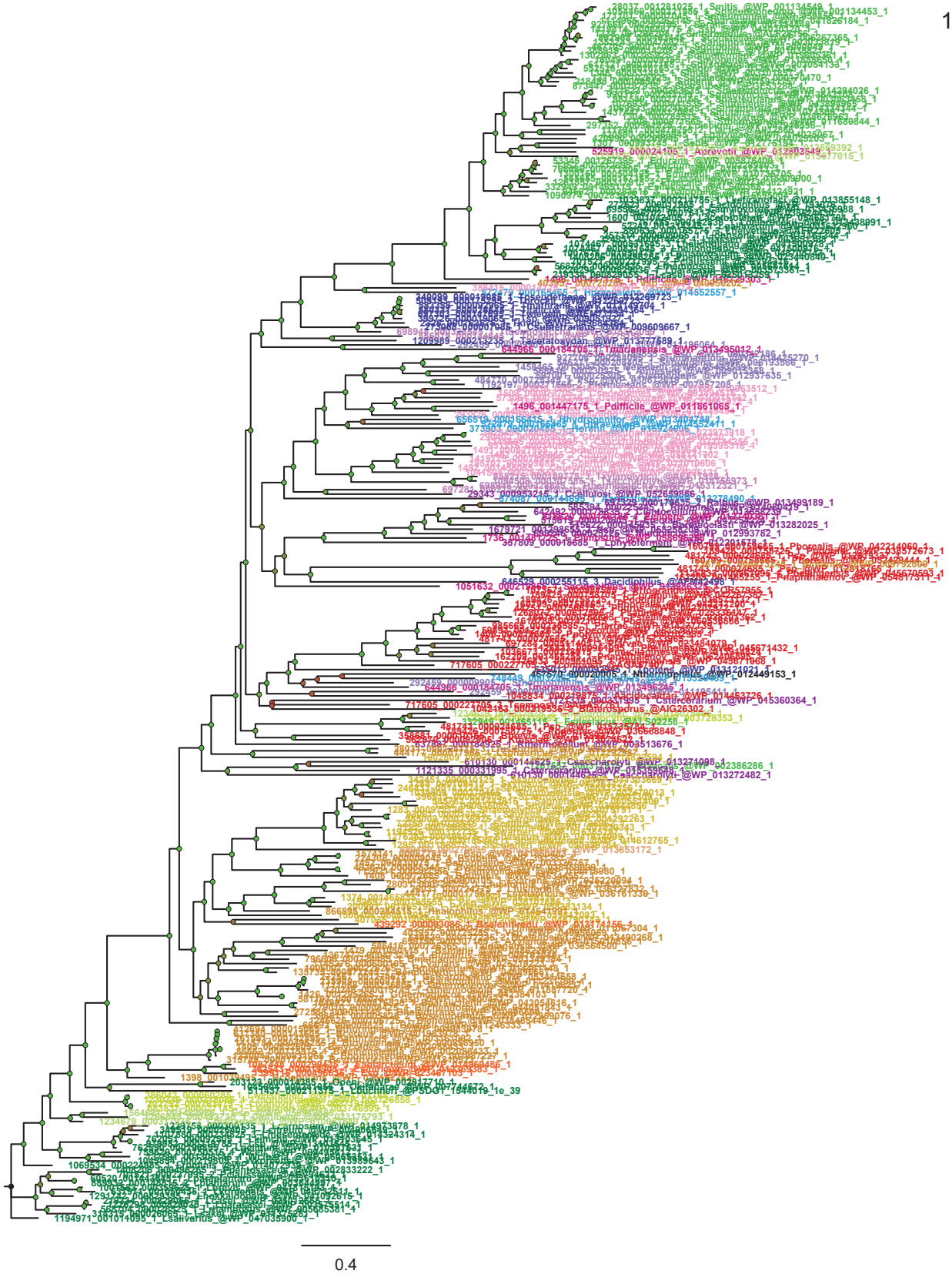

1304_000785515_1_Ssalivarius_@WP_038675892_1
1308_000971665_1_Sthermophilus_@WP_011681202_1
927666_000253155_1_Soralis_@WP_000495713_1
1114965_000262145_1_Sparasanguini_@WP_014713624_1
28037_001281025_1_Smitis_@WP_060627720_1
171101_000007045_1_Spneumoniae_@NP_358134_1
1054460_000221985_1_Spseudopneumo_@WP_000286554_1
388919_000014205_1_Ssanguinis_@YP_001034838_1
467705_000017005_1_Sgordonii_@WP_041789756_1
1302863_000385925_1_Soligoferment_@WP_015604637_1
1338_001296205_1_Sintermedius_@ALF27920_1
862968_000463445_1_Sconstellatus_@WP_020997701_1
1353243_000478925_1_Sanginosus_@WP_025271890_1
1437447_000817065_1_Smutans_@WP_002263311_1
617121_000307185_1_Sdysgalactiae_@WP_042357871_1
160491_000009385_1_Spyogenes_@WP_011889055_1
552526_000020765_1_Sequi_@WP_012515920_1
1346_000831485_1_Siniae_@WP_003101182_1
218495_000009545_1_Suberis_@WP_012658130_1
873447_000187935_2_Sparauberis_@EGE53825_1
1311_001026925_1_Sagalactiae_@WP_000285514_1
1116231_000283635_1_Smacedonicus_@WP_014294962_1
990317_000203195_1_Sgallolyticus_@WP_012962217_1
981540_000270165_1_Spasteurianus_@WP_013852085_1
1069533_000246835_1_Sinfantarius_@WP_014335038_1
1076934_000441535_1_Slutetiensis_@WP_043895086_1
1311_001026925_1_Sagalactiae_@WP_000527090_1
160491_000009385_1_Spyogenes_@WP_050714235_1
617121_000307185_1_Sdysgalactiae_@WP_012766913_1
552526_000020765_1_Sequi_@WP_041785488_1
1346_000831485_1_Siniae_@WP_003098762_1
873447_000187935_2_Sparauberis_@EGE54284_1
297352_000981525_1_Lpiscium_@WP_047915324_1
1090974_000283975_1_Mplutonius_@WP_015694641_1
332949_001465115_1_Esilesiacus_@ALS01225_1
1261557_000317915_1_Efaecalis_@WP_002385715_1
888727_001189495_1_Esulci_@WP_034813551_1
1117941_000478255_2_Llactis_@All13633_1
420890_000269945_1_Lgarvieae_@WP_014024474_1
0.2

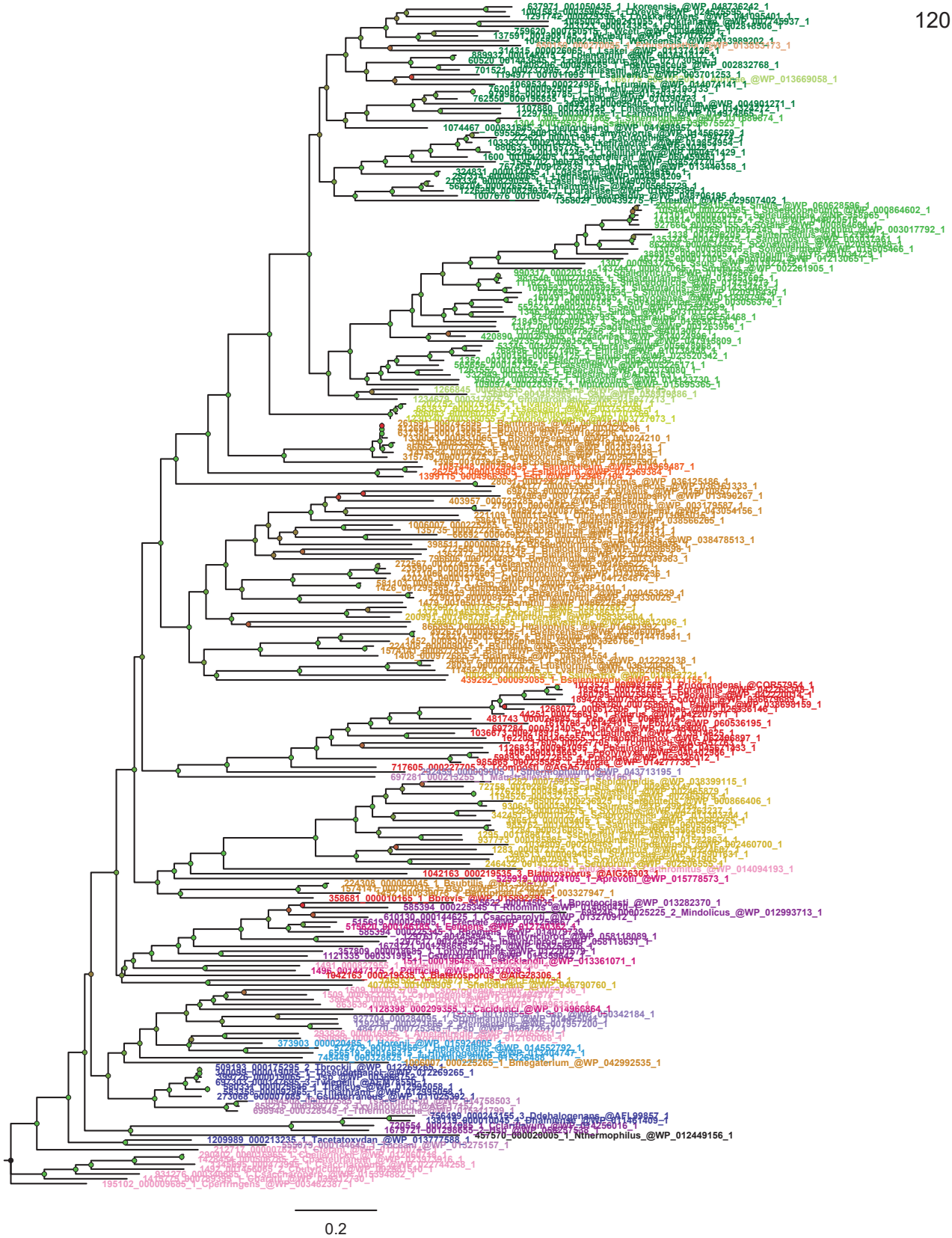

979982_000219785_1_Lsp_@WP_013975304_1
762051_000092505_1_Lkimchii_@WP_013103108_1
762550_000196855_1_Lgelidum_@WP_013231957_1
1229758_000300135_1_Lcarnosum_@WP_014974889_1
349519_000026405_1_Lcitreum_@WP_004907532_1
1107880_000234825_3_Lmesenteroide_@WP_014324259_1
203123_000014385_1_Ooeni_@WP_002820829_1
1045854_000219805_1_Wkoreensis_@WP_013989248_1
759620_000750515_1_Wceti_@WP_009496090_1
137591_001308145_1_Wcibaria_@WP_060653641_1
203123_000014385_1_Ooeni_@WP_011677569_1
1045004_000241055_1_Okitaharae_@WP_007744447_1
1090974_000283975_1_Mplutonius_@WP_050978172_1
332949_001465115_1_Esilesiacus_@ALS02283_1
1090974_000283975_1_Mplutonius_@WP_013773230_1
1090974_000283975_1_Mplutonius_@WP_015695354_1
1261557_000317915_1_Efaecalis_@WP_002384322_1
1346_000831485_1_Siniae_@WP_003099625_1
873447_000187935_2_Sparauberis_@EGE54253_1
1308_000971665_1_Sthermophilus_@WP_011680758_1
888727_001189495_1_Esulci_@WP_034813498_1
888727_001189495_1_Esulci_@WP_034813500_1
699246_000025225_2_Mindolicus_@WP_012993834_1
699246_000025225_2_Mindolicus_@WP_012993835_1
866775_000193205_1_Aurinae_@WP_013670024_1
699246_000025225_2_Mindolicus_@WP_012993987_1
297352_000981525_1_Lpiscium_@WP_047915323_1
866775_000193205_1_Aurinae_@WP_013668724_1
420890_000269945_1_Lgarvieae_@WP_014024470_1
420890_000269945_1_Lgarvieae_@WP_014024469_1
1114965_000262145_1_Sparasanguini_@WP_014713625_1
1069533_000246835_1_Sinfantarius_@WP_014335039_1
1353243_000478925_1_Sanginosus_@WP_003037301_1
1338_001296205_1_Sintermedius_@ALF27921_1
862968_000463445_1_Sconstellatus_@WP_020997700_1
467705_000017005_1_Sgordonii_@WP_012130686_1
1302863_000385925_1_Soligoferment_@WP_015604636_1
388919_000014205_1_Ssanguinis_@YP_001034837_1
1304_000785515_1_Ssalivarius_@WP_038675570_1
1308_000971665_1_Sthermophilus_@WP_011680915_1
1437447_000817065_1_Smutans_@WP_002263312_1
1353243_000478925_1_Sanginosus_@PSDG2_699441_1e_27
1116231_000283635_1_Smacedonicus_@WP_014294963_1
990317_000203195_1_Sgallolyticus_@WP_009854601_1
981540_000270165_1_Spasteurianus_@WP_013852086_1
1076934_000441535_1_Slutetiensis_@WP_020917182_1
617121_000307185_1_Sdysgalactiae_@WP_015016821_1
160491_000009385_1_Spyogenes_@WP_011184327_1
552526_000020765_1_Sequi_@WP_012515921_1
990317_000203195_1_Sgallolyticus_@WP_012962360_1
1116231_000283635_1_Smacedonicus_@WP_014295068_1
981540_000270165_1_Spasteurianus_@WP_013852162_1
1076934_000441535_1_Slutetiensis_@WP_020917277_1
1069533_000246835_1_Sinfantarius_@WP_014335151_1
1311_001026925_1_Sagalactiae_@WP_000175198_1
1311_001026925_1_Sagalactiae_@WP_001185388_1
171101_000007045_1_Spneumoniae_@NP_358135_1
1054460_000221985_1_Spseudopneumo_@WP_001174048_1
28037_001281025_1_Smitis_@WP_001174044_1
873447_000187935_2_Sparauberis_@EGE54048_1
218495_000009545_1_Suberis_@WP_012658129_1
1346_000831485_1_Siniae_@WP_003101185_1
0.4

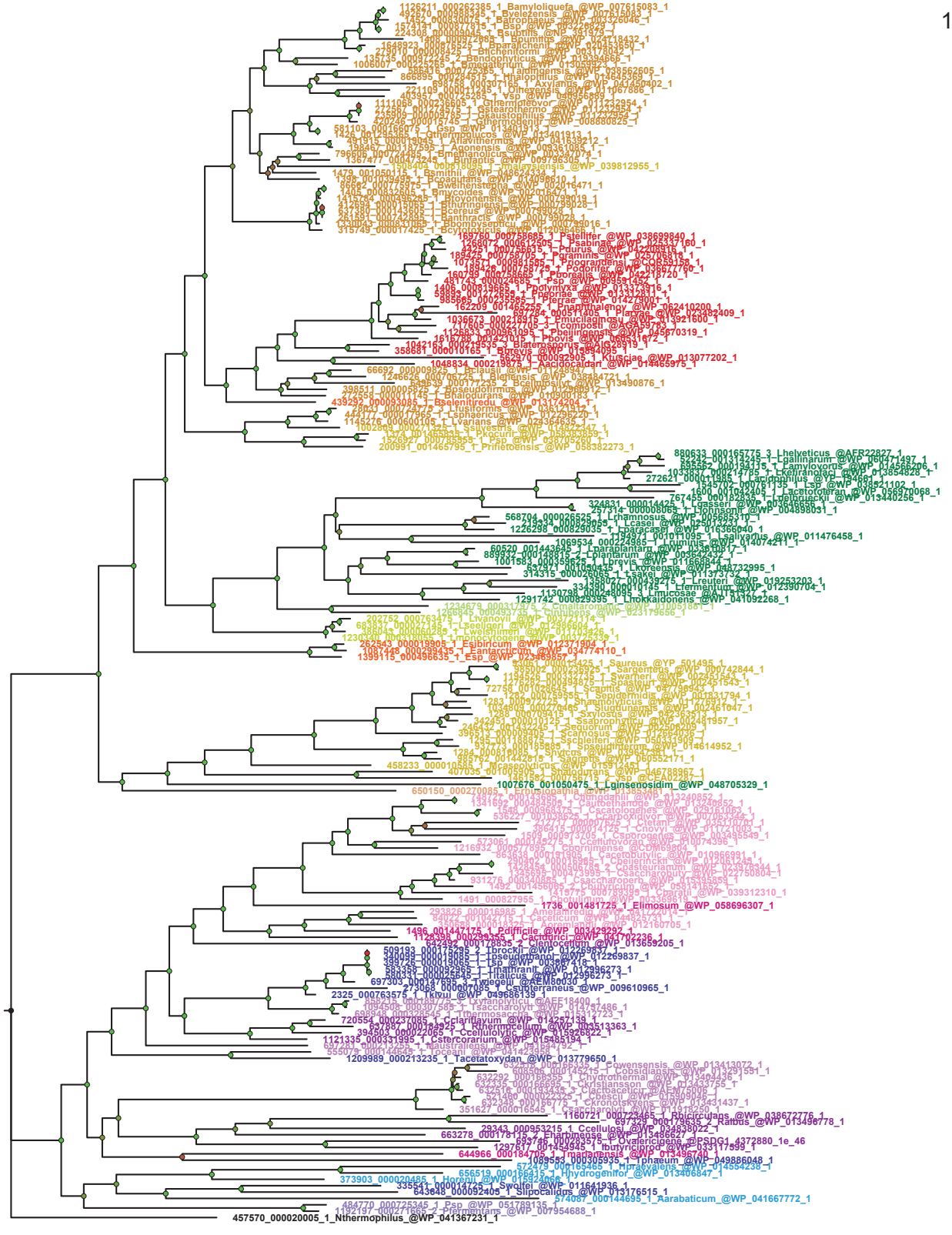

0.2

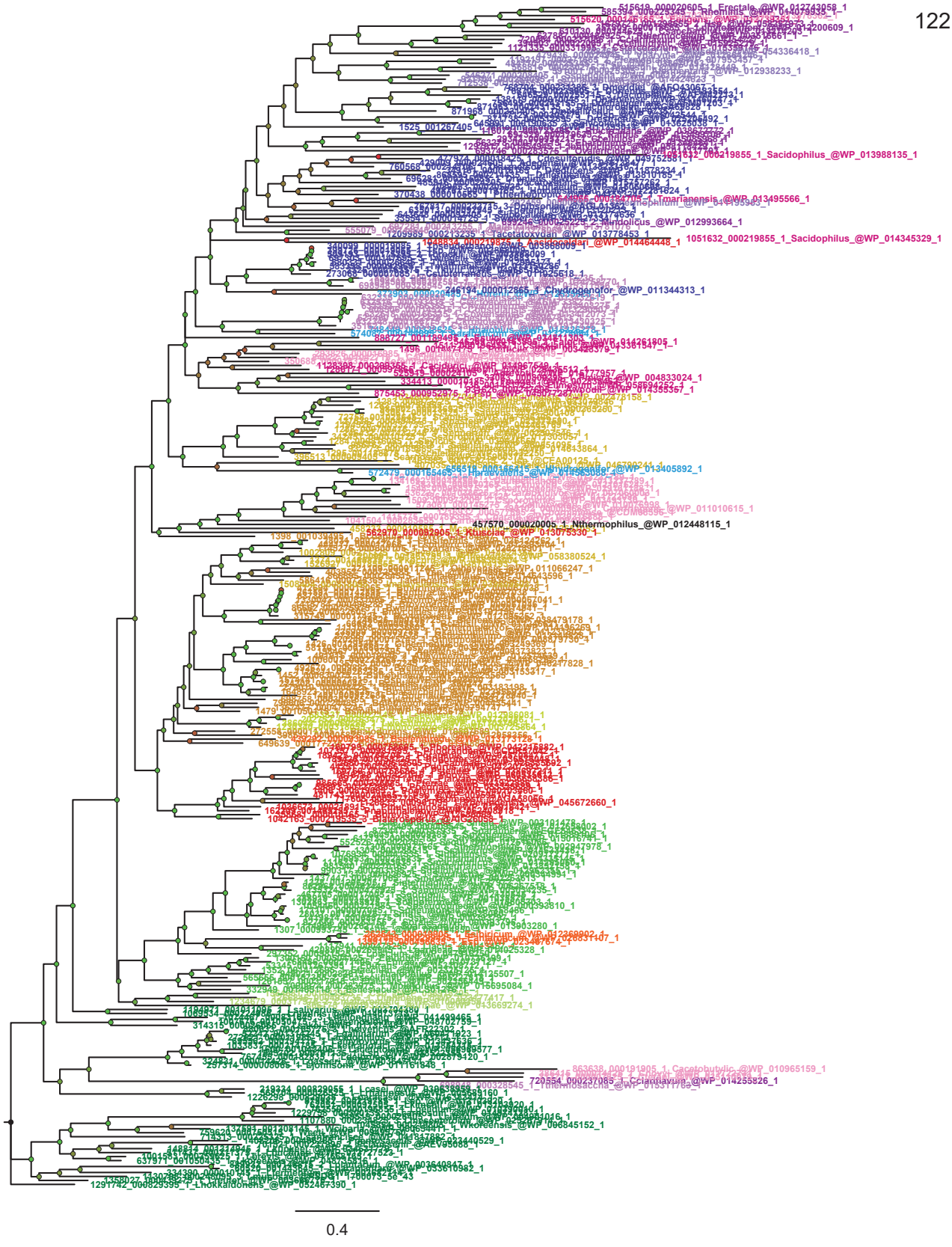

0.4

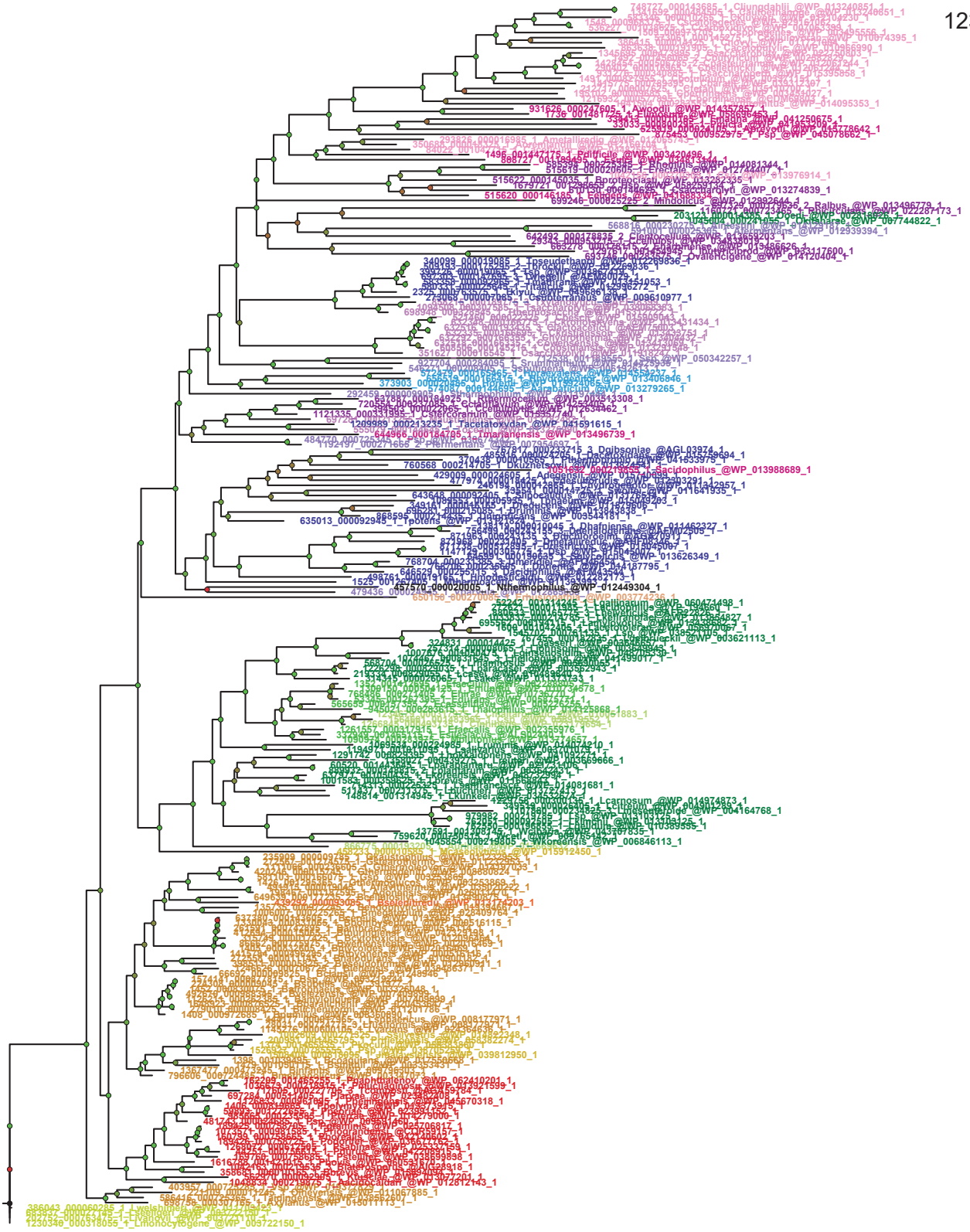

0.3

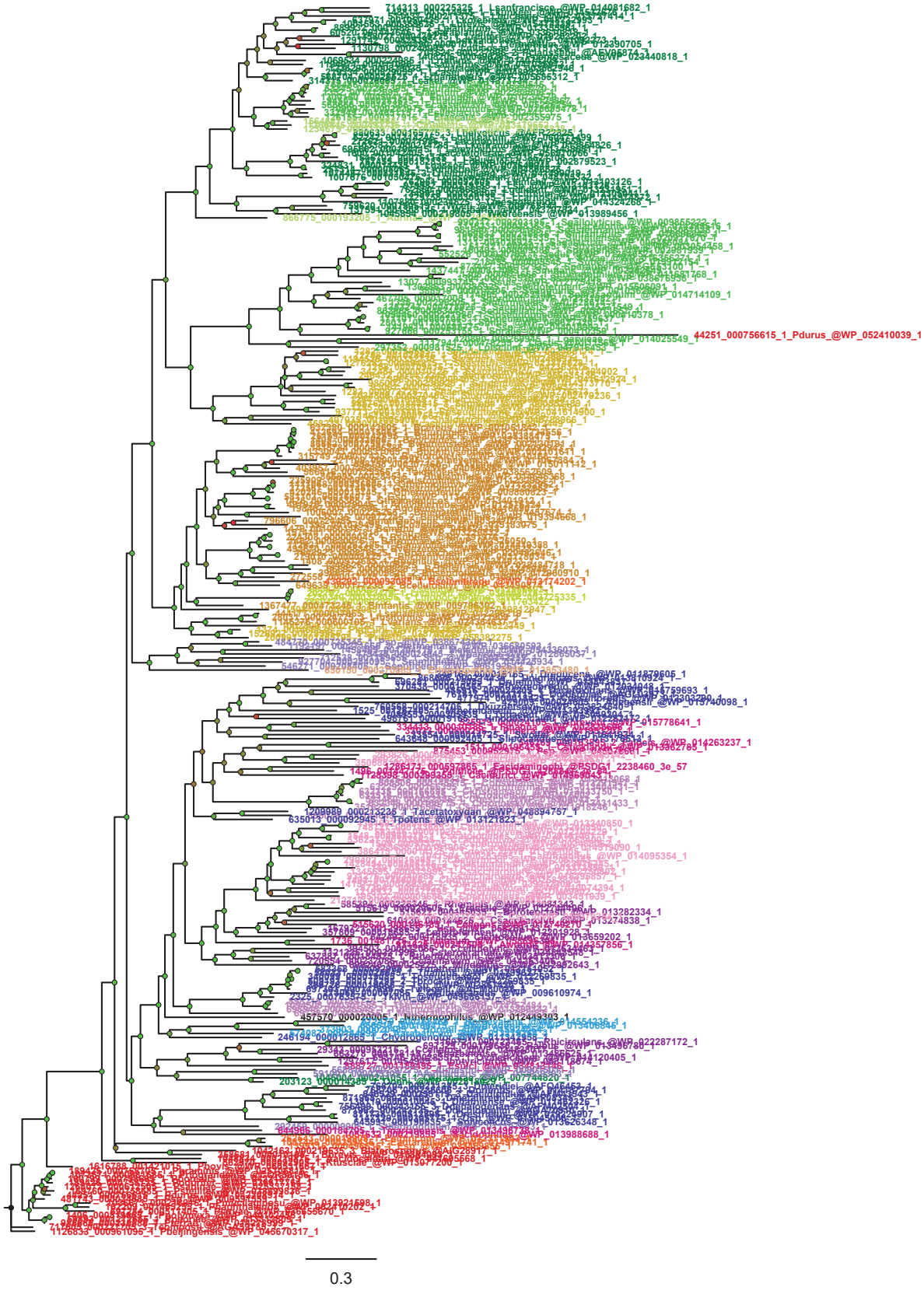

0.4

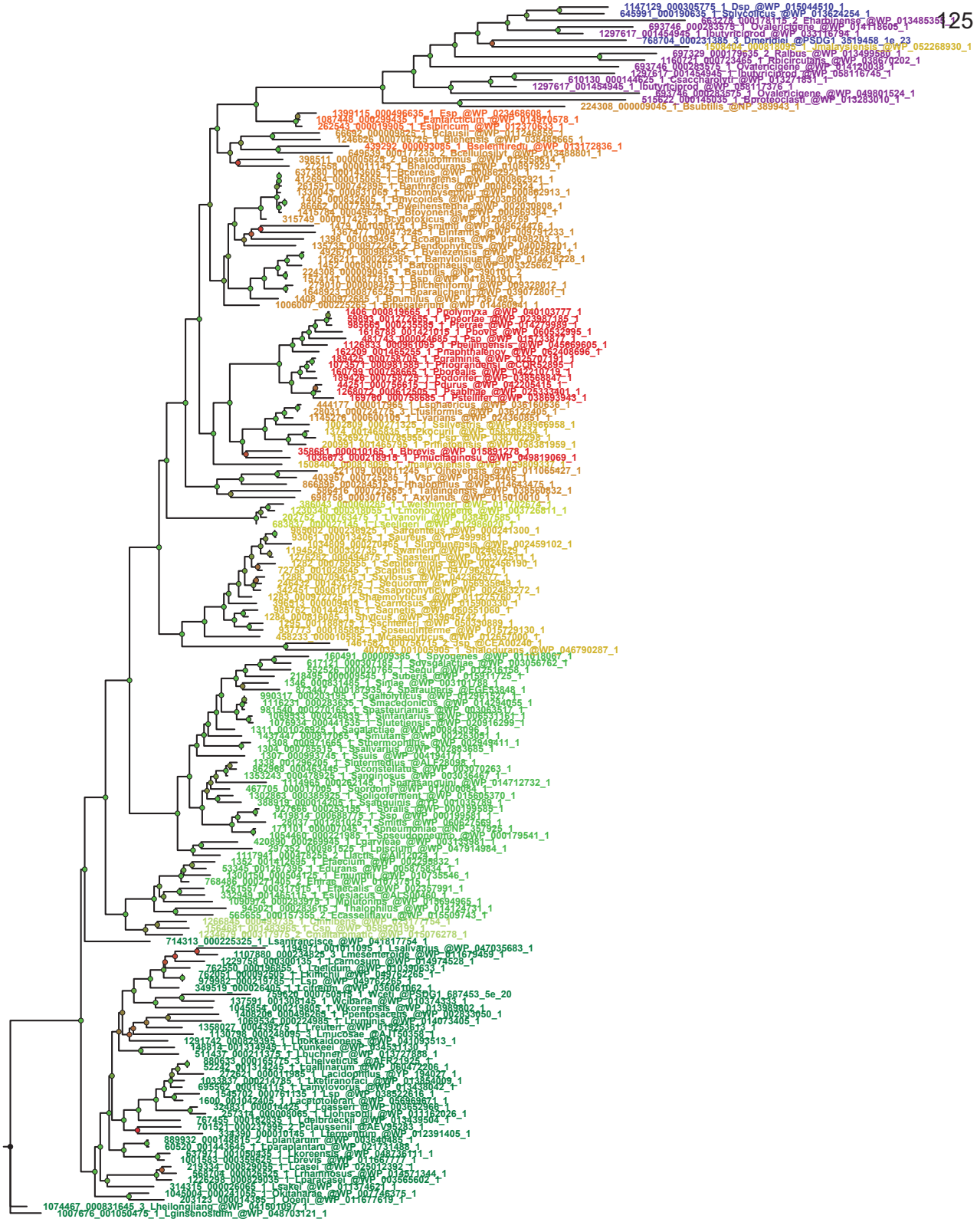

637971_001050435_1_Lkoreensis_@WP_048736242_1
1209989_000213235_1_Tacetatoxydan_@WP_013775588_1
457570_000020005_1_Nthermophilus_@WP_012449156_1
0.2

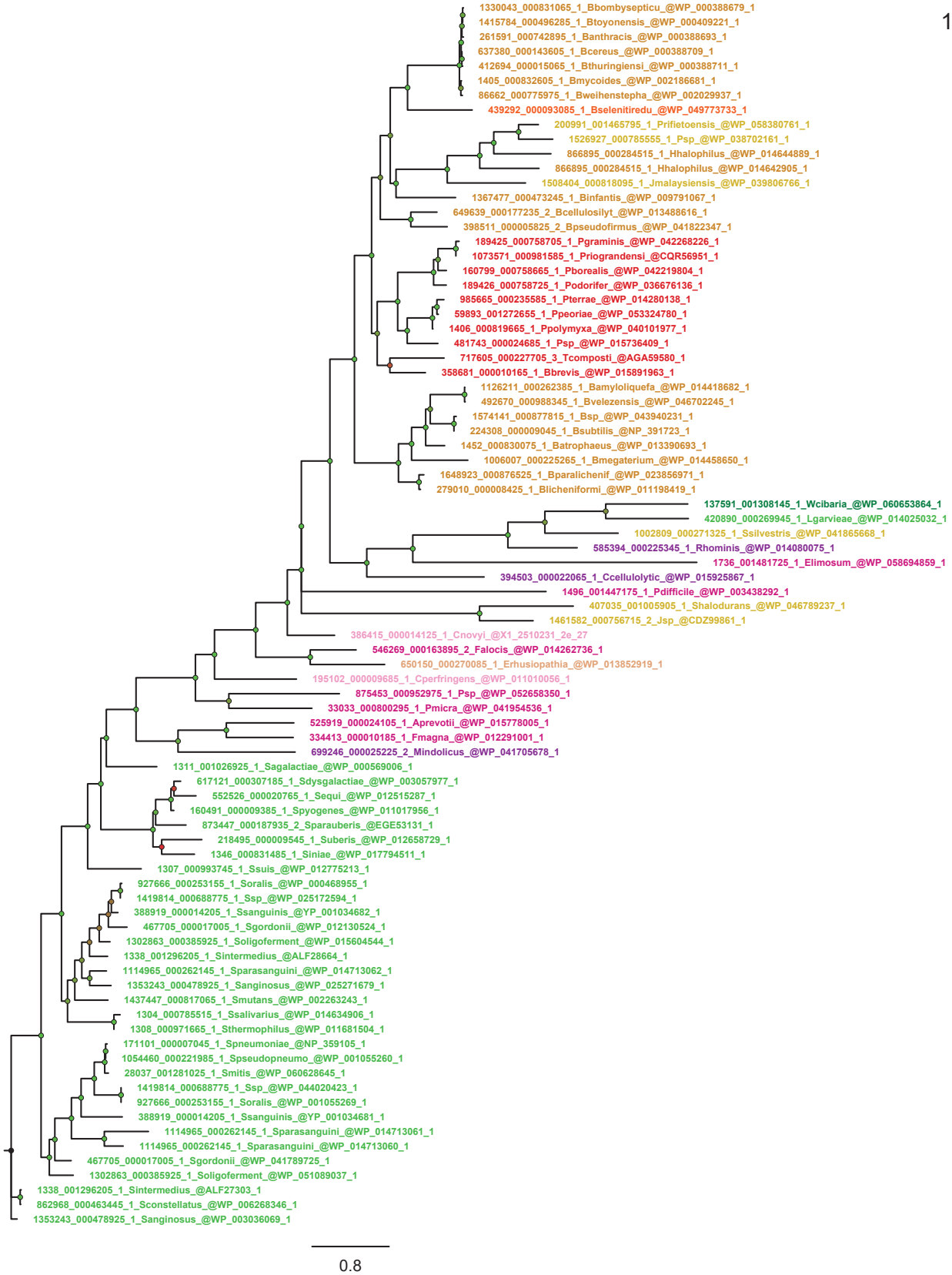

0.2

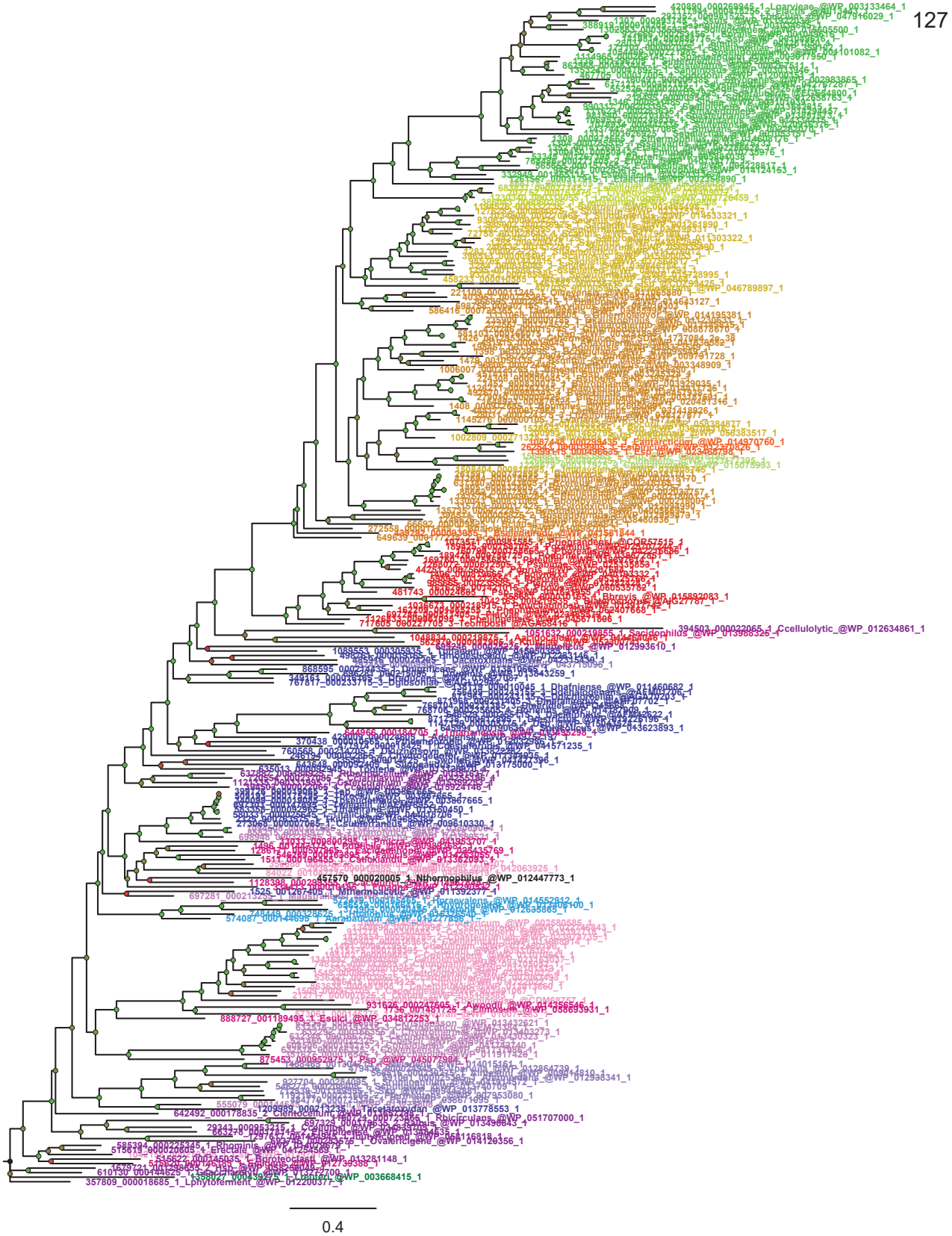

457570_000020005_1_Nthermophilus_@WP_012448115_1
1051632_000219855_1_Sacidophilus_@WP_014345329_1
0.4

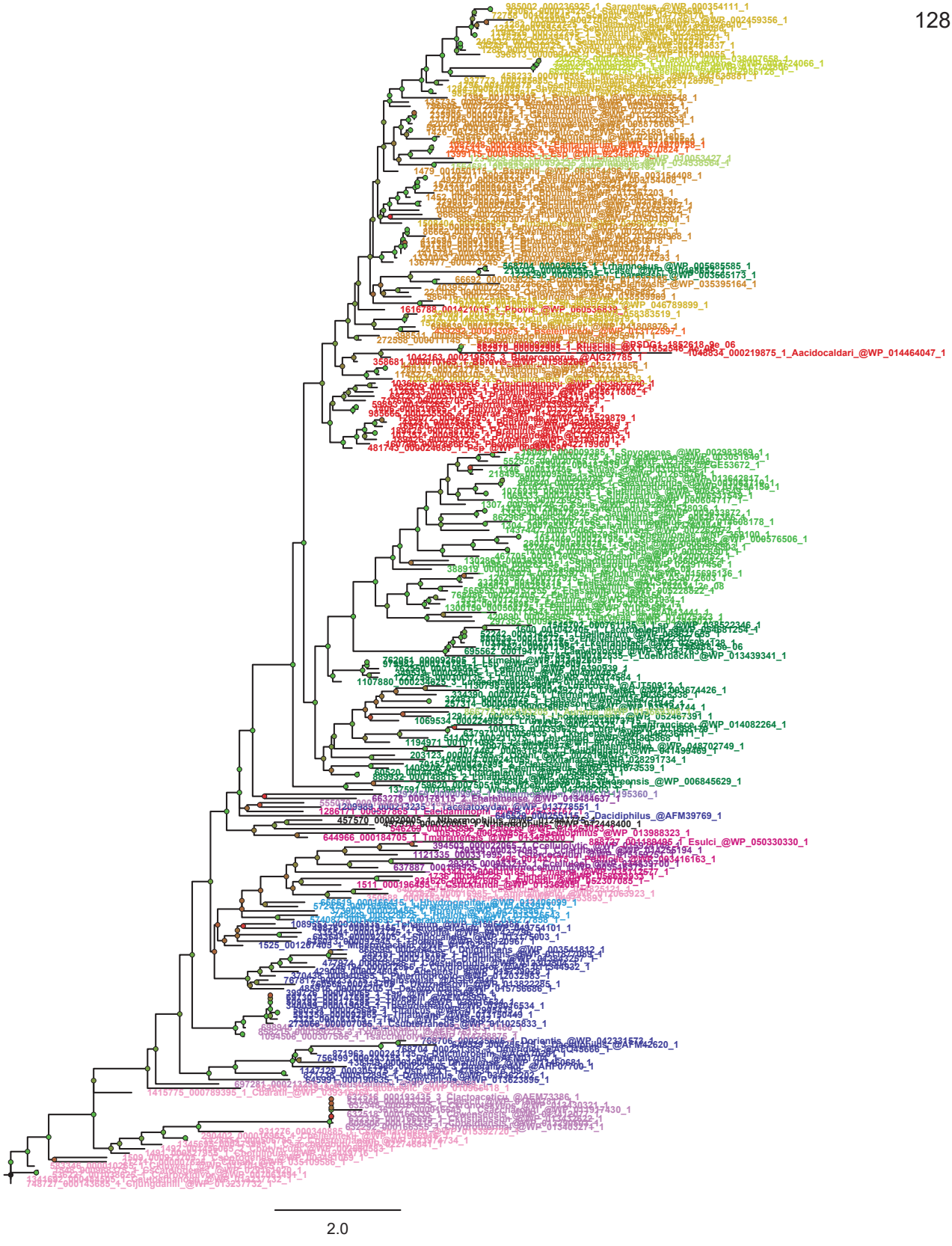

0.3

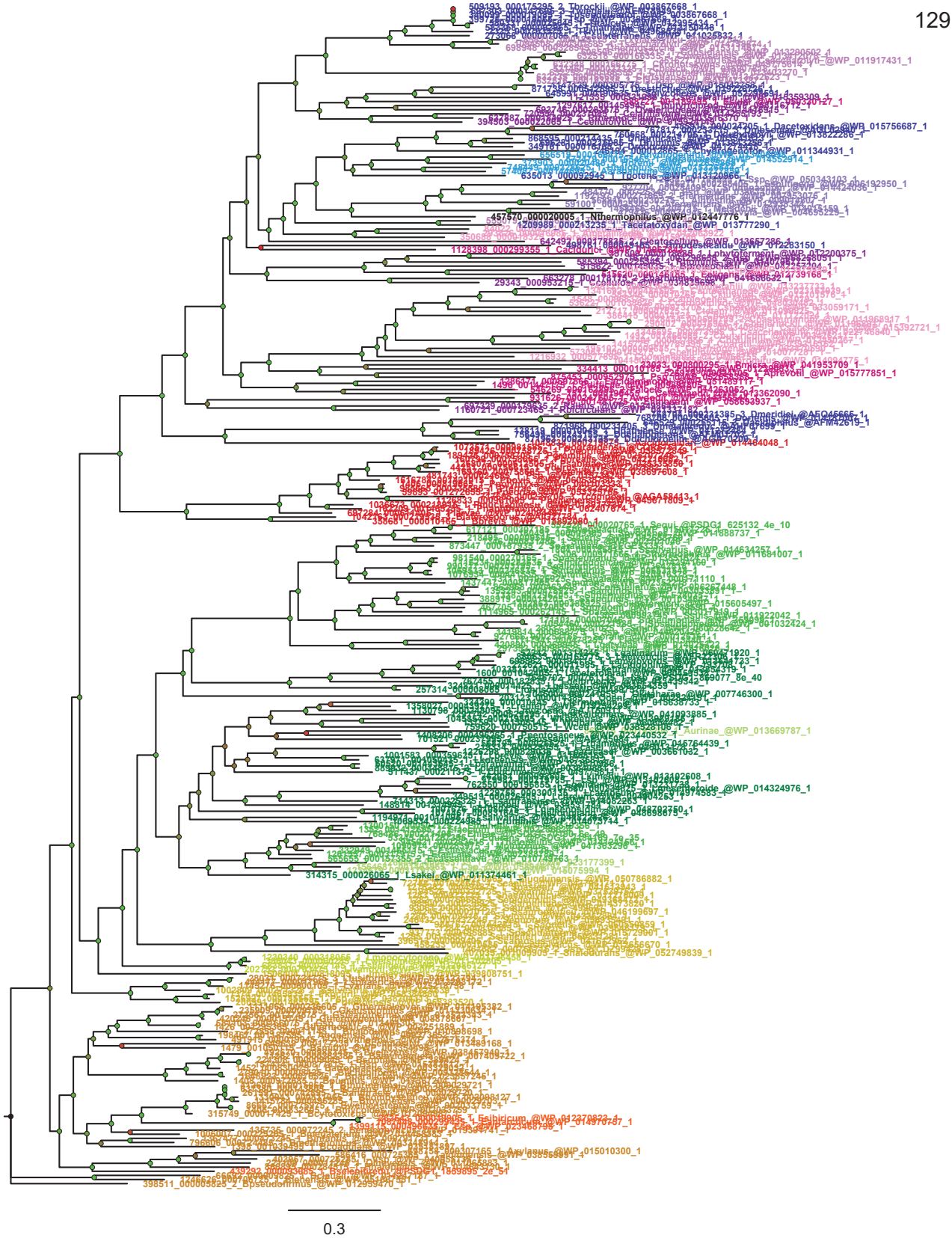

44251_000756615_1_Pdurus_@WP_052410039_1
0.3

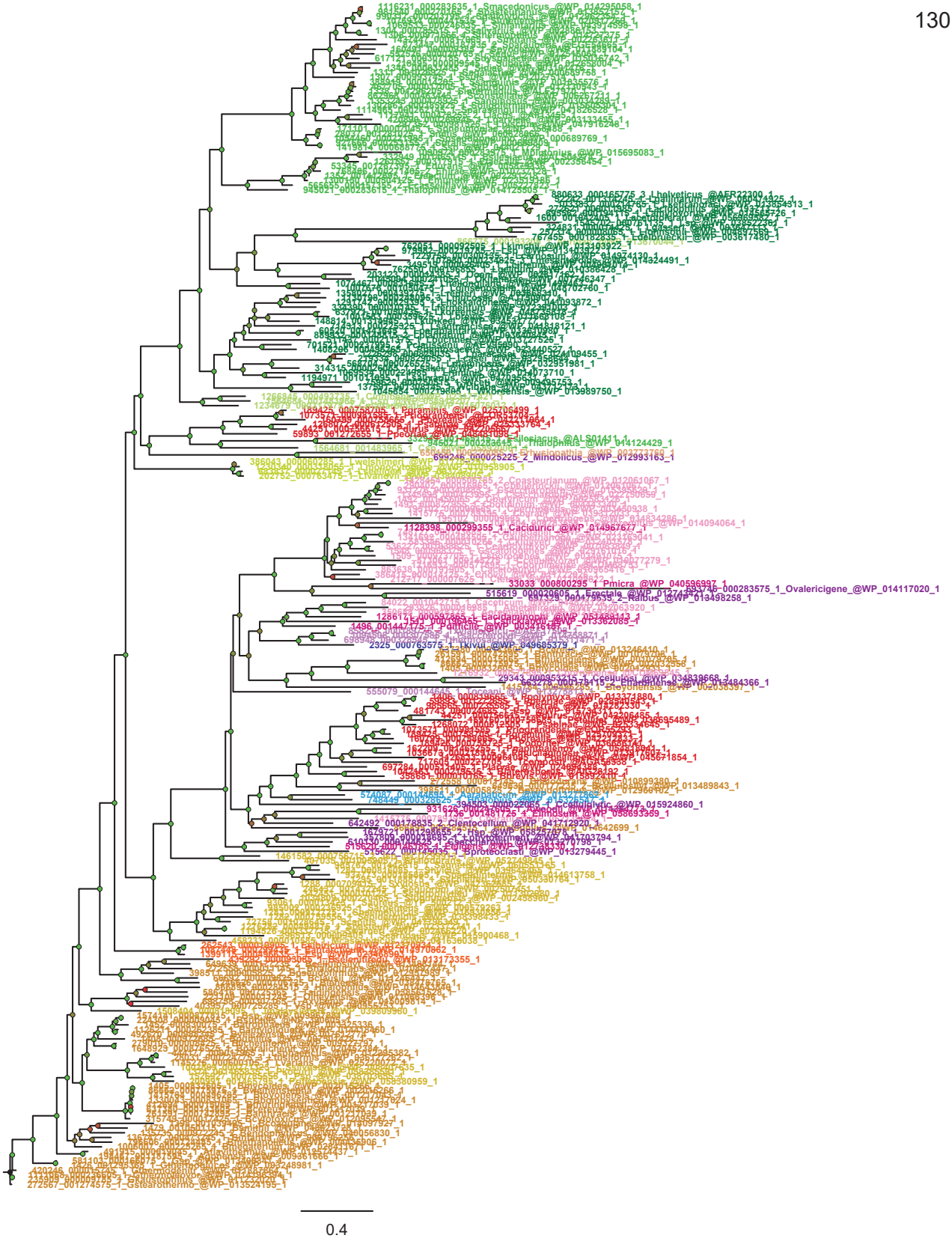

0.6

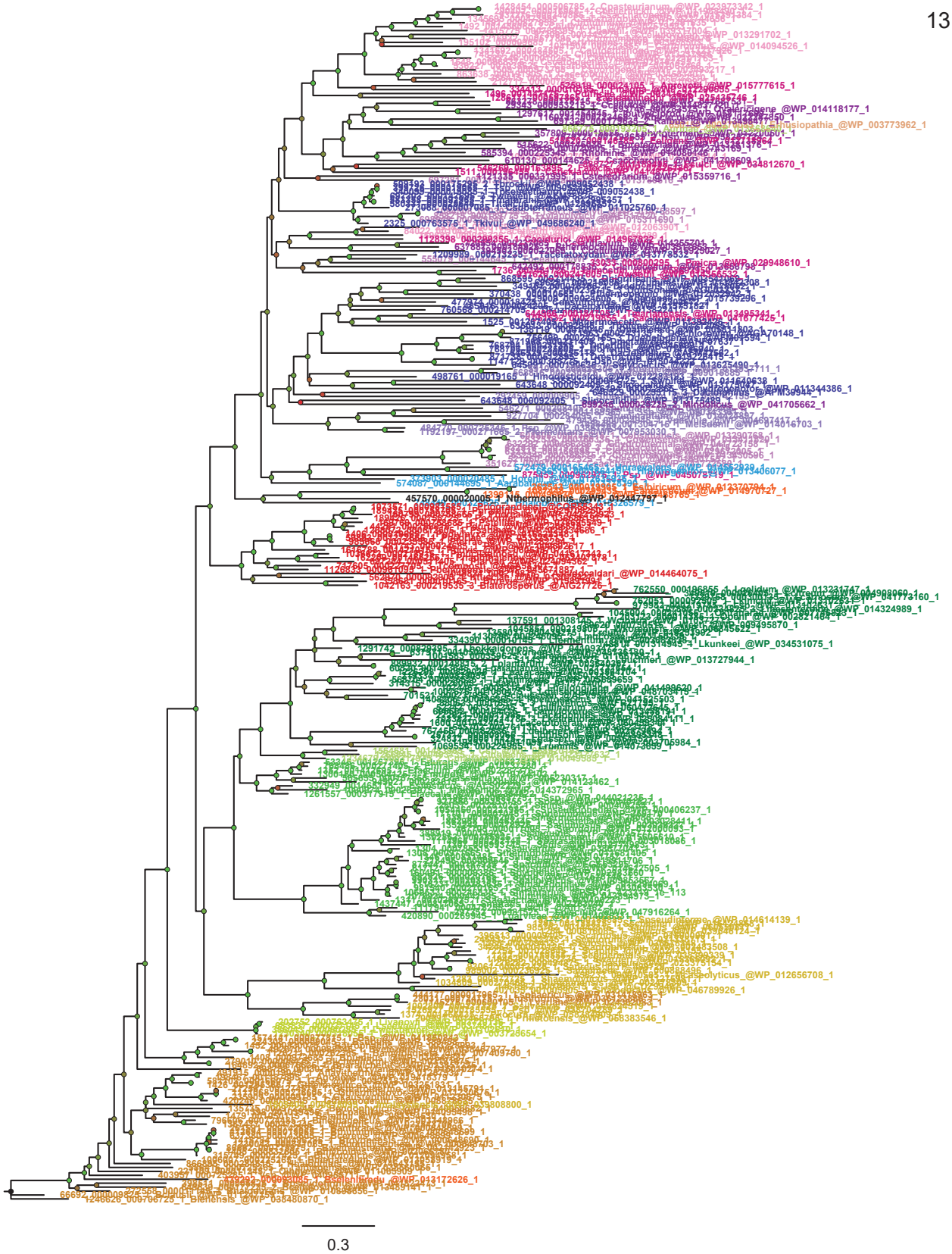

1330043_000831065_1_Bbombysepticu_@WP_000388679_1
1415784_000496285_1_Btoyonensis_@WP_000409221_1
261591_000742895_1_Banthracis_@WP_000388693_1
637380_000143605_1_Bcereus_@WP_000388709_1
412694_000015065_1_Bthuringiensi_@WP_000388711_1
1405_000832605_1_Bmycoides_@WP_002186681_1
86662_000775975_1_Bweihenstepha_@WP_002029937_1
439292_000093085_1_Bselenitiredu_@WP_049773733_1
200991_001465795_1_Prifietoensis_@WP_058380761_1
1526927_000785555_1_Psp_@WP_038702161_1
866895_000284515_1_Hhalophilus_@WP_014644889_1
866895_000284515_1_Hhalophilus_@WP_014642905_1
1508404_000818095_1_Jmalaysiensis_@WP_039806766_1
1367477_000473245_1_Binfantis_@WP_009791067_1
649639_000177235_2_Bcellulosilyt_@WP_013488616_1
398511_000005825_2_Bpseudofirmus_@WP_041822347_1
189425_000758705_1_Pgraminis_@WP_042268226_1
1073571_000981585_1_Priograndensi_@CQR56951_1
160799_000758665_1_Pborealis_@WP_042219804_1
189426_000758725_1_Podorifer_@WP_036676136_1
985665_000235585_1_Pterrae_@WP_014280138_1
59893_001272655_1_Ppeoriae_@WP_053324780_1
1406_000819665_1_Ppolymyxa_@WP_040101977_1
481743_000024685_1_Psp_@WP_015736409_1
717605_000227705_3_Tcomposti_@AGA59580_1
358681_000010165_1_Bbrevis_@WP_015891963_1
1126211_000262385_1_Bamyloliquefa_@WP_014418682_1
492670_000988345_1_Bvelezensis_@WP_046702245_1
1574141_000877815_1_Bsp_@WP_043940231_1
224308_000009045_1_Bsubtilis_@NP_391723_1
1452_000830075_1_Batrophaeus_@WP_013390693_1
1006007_000225265_1_Bmegaterium_@WP_014458650_1
1648923_000876525_1_Bparalichenif_@WP_023856971_1
279010_000008425_1_Blicheniformi_@WP_011198419_1
137591_001308145_1_Wcibaria_@WP_060653864_1
420890_000269945_1_Lgarvieae_@WP_014025032_1
1002809_000271325_1_Ssilvestris_@WP_041865668_1
585394_000225345_1_Rhominis_@WP_014080075_1
1736_001481725_1_Elimosum_@WP_058694859_1
394503_000022065_1_Ccellulolytic_@WP_015925867_1
1496_001447175_1_Pdifficile_@WP_003438292_1
407035_001005905_1_Shalodurans_@WP_046789237_1
1461582_000756715_2_Jsp_@CDZ99861_1
386415_000014125_1_Cnovyi_@X1_2510231_2e_27
546269_000163895_2_Falocis_@WP_014262736_1
650150_000270085_1_Erhusiopathia_@WP_013852919_1
195102_000009685_1_Cperfringens_@WP_011010056_1
875453_000952975_1_Psp_@WP_052658350_1
33033_000800295_1_Pmicra_@WP_041954536_1
525919_000024105_1_Aprevotii_@WP_015778005_1
334413_000010185_1_Fmagna_@WP_012291001_1
699246_000025225_2_Mindolicus_@WP_041705678_1
1311_001026925_1_Sagalactiae_@WP_000569006_1
617121_000307185_1_Sdysgalactiae_@WP_003057977_1
552526_000020765_1_Sequi_@WP_012515287_1
160491_000009385_1_Spyogenes_@WP_011017956_1
873447_000187935_2_Sparauberis_@EGE53131_1
218495_000009545_1_Suberis_@WP_012658729_1
1346_000831485_1_Siniae_@WP_017794511_1
1307_000993745_1_Ssuis_@WP_012775213_1
927666_000253155_1_Soralis_@WP_000468955_1
1419814_000688775_1_Ssp_@WP_025172594_1
388919_000014205_1_Ssanguinis_@YP_001034682_1
467705_000017005_1_Sgordonii_@WP_012130524_1
1302863_000385925_1_Soligoferment_@WP_015604544_1
1338_001296205_1_Sintermedius_@ALF28664_1
1114965_000262145_1_Sparasanguini_@WP_014713062_1
1353243_000478925_1_Sanginosus_@WP_025271679_1
1437447_000817065_1_Smutans_@WP_002263243_1
1304_000785515_1_Ssalivarius_@WP_014634906_1
1308_000971665_1_Sthermophilus_@WP_011681504_1
171101_000007045_1_Spneumoniae_@NP_359105_1
1054460_000221985_1_Spseudopneumo_@WP_001055260_1
28037_001281025_1_Smitis_@WP_060628645_1
1419814_000688775_1_Ssp_@WP_044020423_1
927666_000253155_1_Soralis_@WP_001055269_1
388919_000014205_1_Ssanguinis_@YP_001034681_1
1114965_000262145_1_Sparasanguini_@WP_014713061_1
1114965_000262145_1_Sparasanguini_@WP_014713060_1
467705_000017005_1_Sgordonii_@WP_041789725_1
1302863_000385925_1_Soligoferment_@WP_051089037_1
1338_001296205_1_Sintermedius_@ALF27303_1
862968_000463445_1_Sconstellatus_@WP_006268346_1
1353243_000478925_1_Sanginosus_@WP_003036069_1
0.8

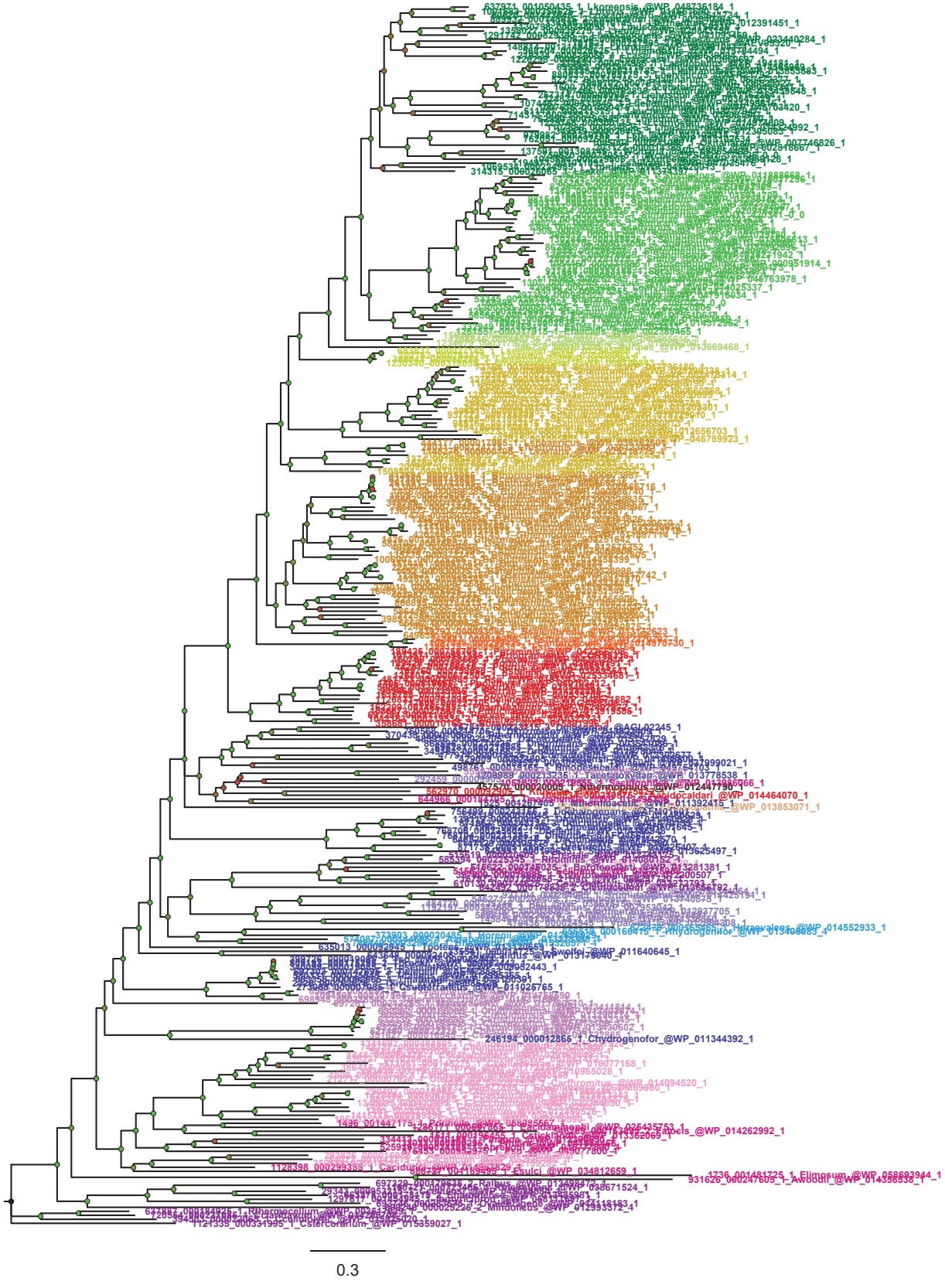

457570_000020005_1_Nthermophilus_@WP_012447773_1
0.4

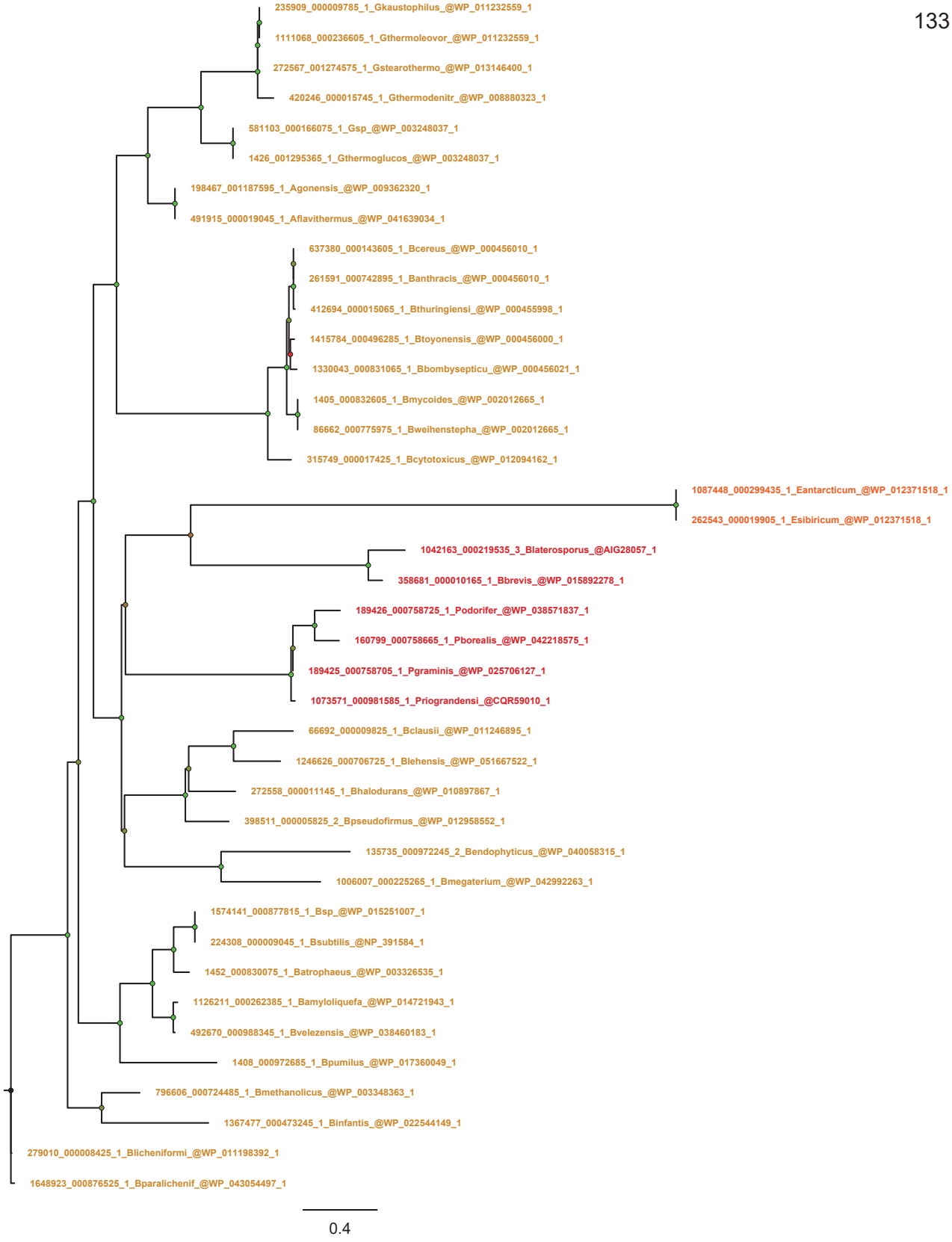

2.0

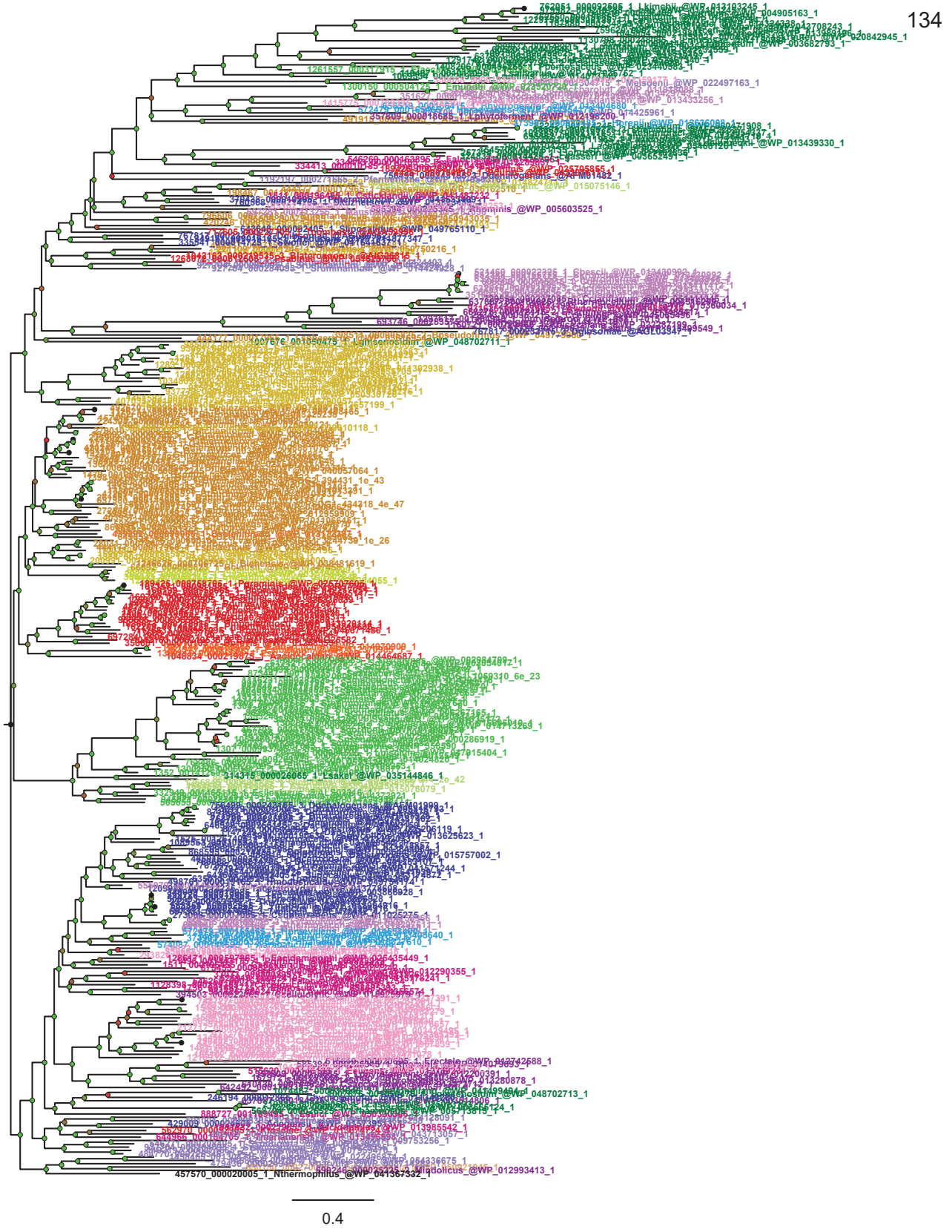

0.3

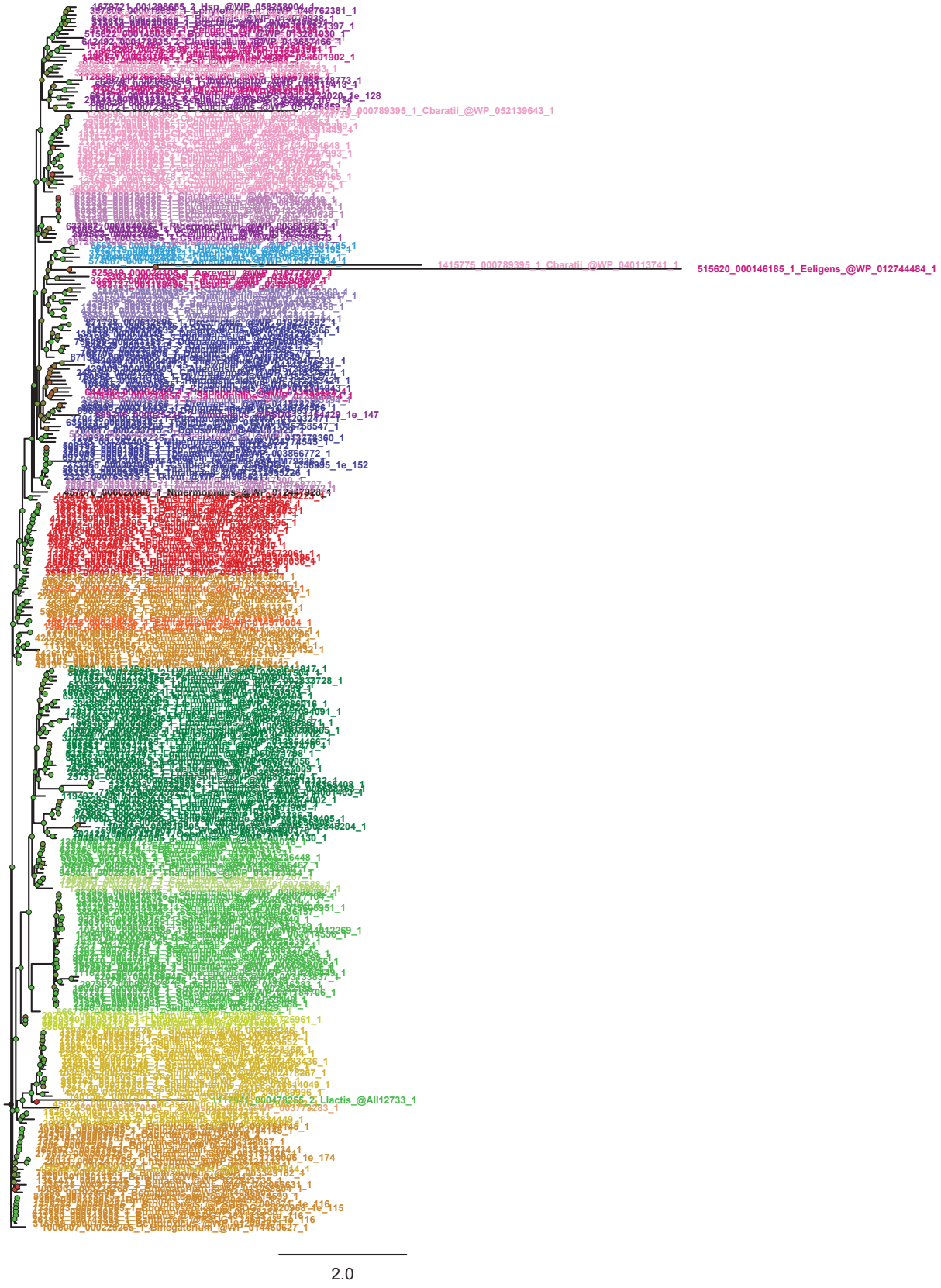

0.4

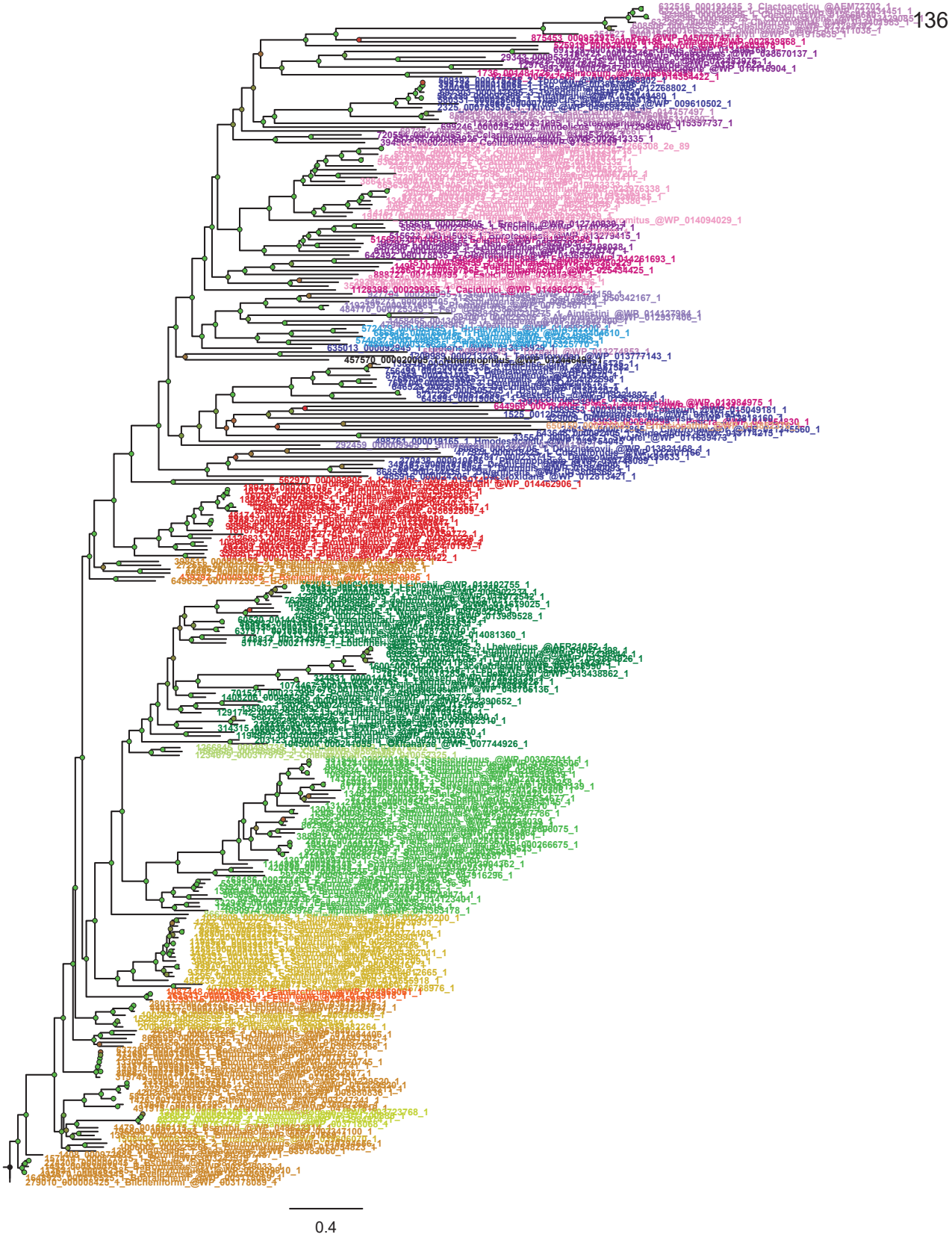

0.3

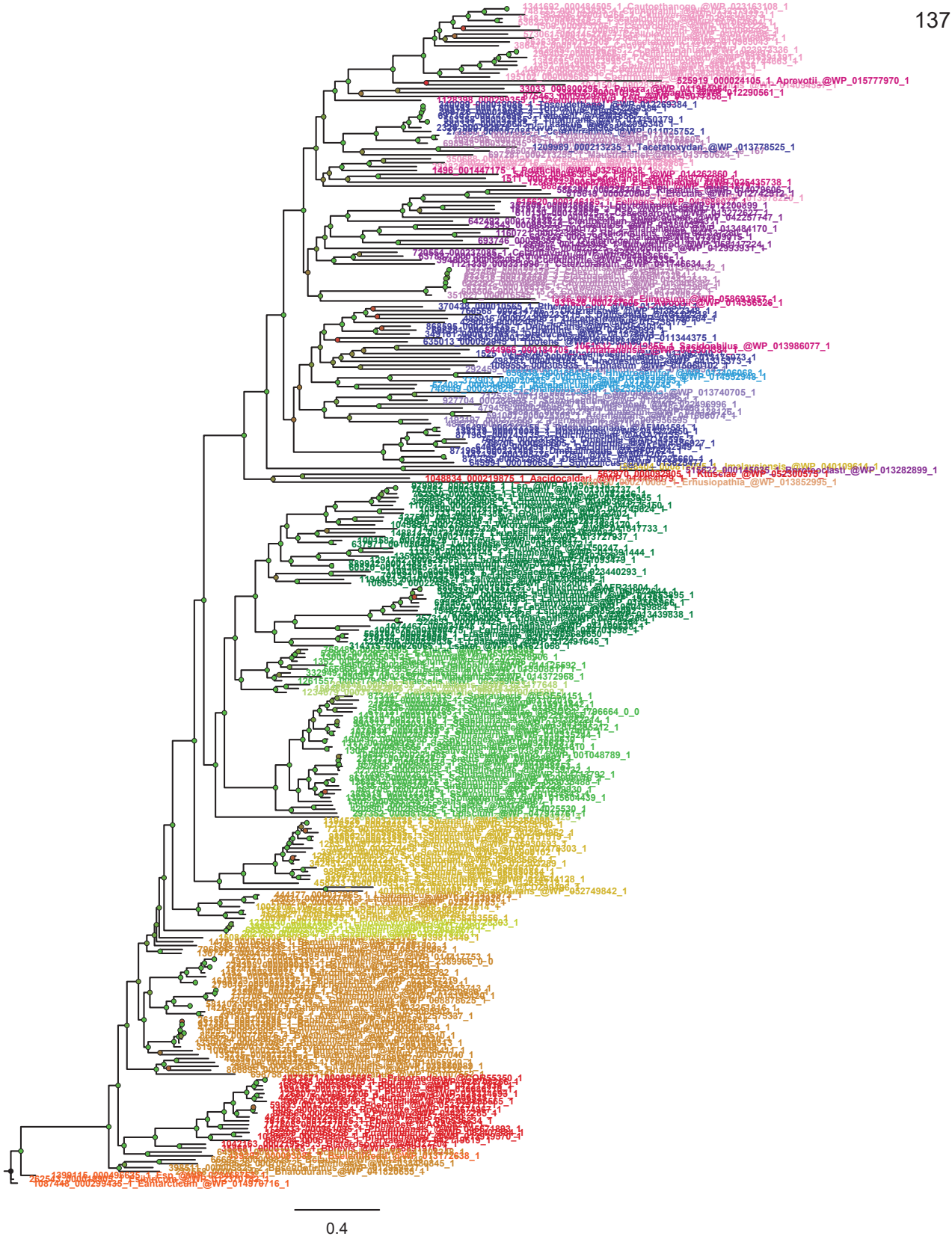

246194_000012865_1_Chydrogenofor_@WP_011344392_1
0.3

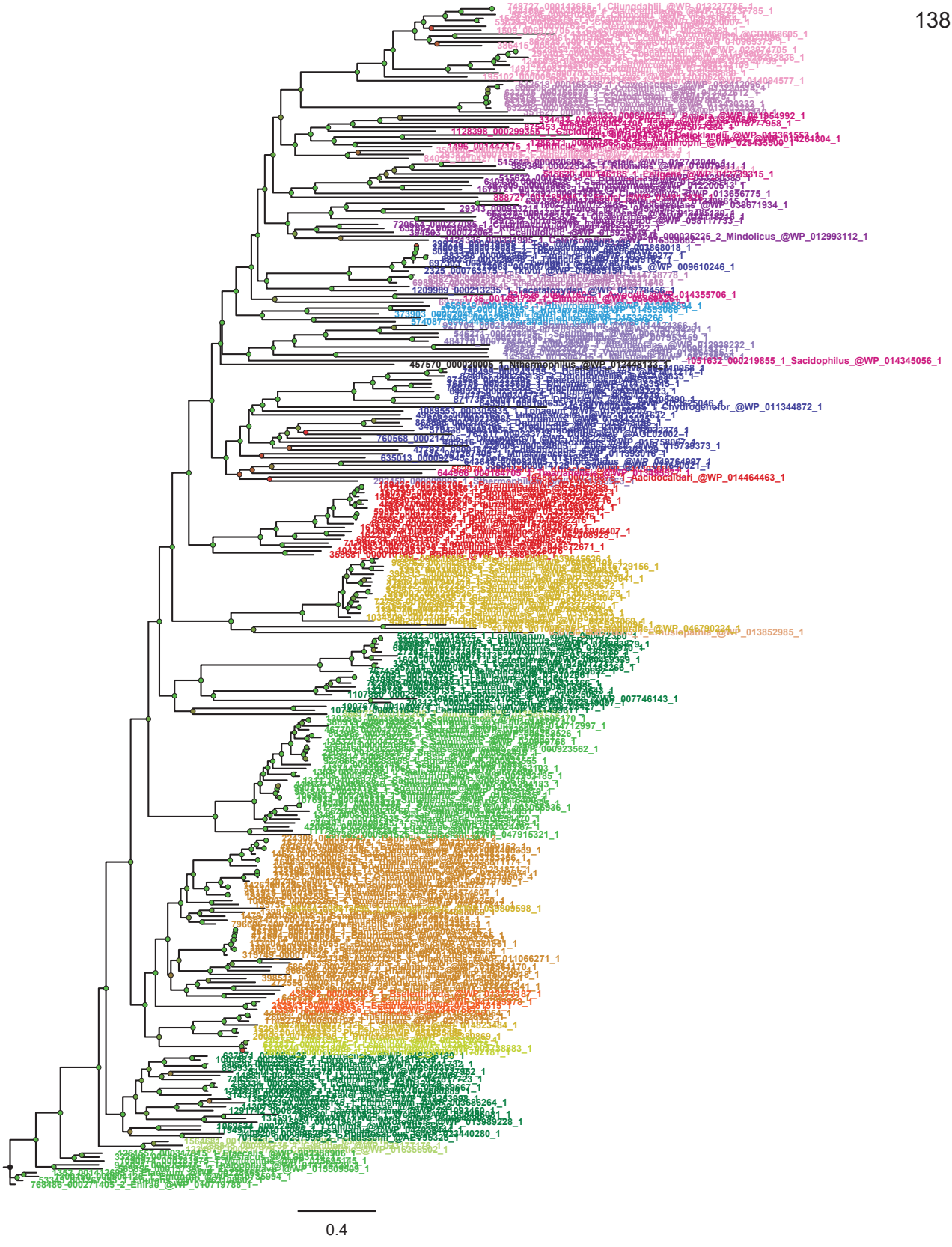

235909_000009785_1_Gkaustophilus_@WP_011232559_1
1111068_000236605_1_Gthermoleovor_@WP_011232559_1
272567_001274575_1_Gstearothermo_@WP_013146400_1
420246_000015745_1_Gthermodenitr_@WP_008880323_1
581103_000166075_1_Gsp_@WP_003248037_1
1426_001295365_1_Gthermoglucos_@WP_003248037_1
198467_001187595_1_Agonensis_@WP_009362320_1
491915_000019045_1_Aflavithermus_@WP_041639034_1
637380_000143605_1_Bcereus_@WP_000456010_1
261591_000742895_1_Banthracis_@WP_000456010_1
412694_000015065_1_Bthuringiensi_@WP_000455998_1
1415784_000496285_1_Btoyonensis_@WP_000456000_1
1330043_000831065_1_Bbombysepticu_@WP_000456021_1
1405_000832605_1_Bmycoides_@WP_002012665_1
86662_000775975_1_Bweihenstepha_@WP_002012665_1
315749_000017425_1_Bcytotoxicus_@WP_012094162_1
1087448_000299435_1_Eantarcticum_@WP_012371518_1
262543_000019905_1_Esibiricum_@WP_012371518_1
1042163_000219535_3_Blaterosporus_@AIG28057_1
358681_000010165_1_Bbrevis_@WP_015892278_1
189426_000758725_1_Podorifer_@WP_038571837_1
160799_000758665_1_Pborealis_@WP_042218575_1
189425_000758705_1_Pgraminis_@WP_025706127_1
1073571_000981585_1_Priograndensi_@CQR59010_1
66692_000009825_1_Bclausii_@WP_011246895_1
1246626_000706725_1_Blehensis_@WP_051667522_1
272558_000011145_1_Bhalodurans_@WP_010897867_1
398511_000005825_2_Bpseudofirmus_@WP_012958552_1
135735_000972245_2_Bendophyticus_@WP_040058315_1
1006007_000225265_1_Bmegaterium_@WP_042992263_1
1574141_000877815_1_Bsp_@WP_015251007_1
224308_000009045_1_Bsubtilis_@NP_391584_1
1452_000830075_1_Batrophaeus_@WP_003326535_1
1126211_000262385_1_Bamyloliquefa_@WP_014721943_1
492670_000988345_1_Bvelezensis_@WP_038460183_1
1408_000972685_1_Bpumilus_@WP_017360049_1
796606_000724485_1_Bmethanolicus_@WP_003348363_1
1367477_000473245_1_Binfantis_@WP_022544149_1
279010_000008425_1_Blicheniformi_@WP_011198392_1
1648923_000876525_1_Bparalichenif_@WP_043054497_1
0.4

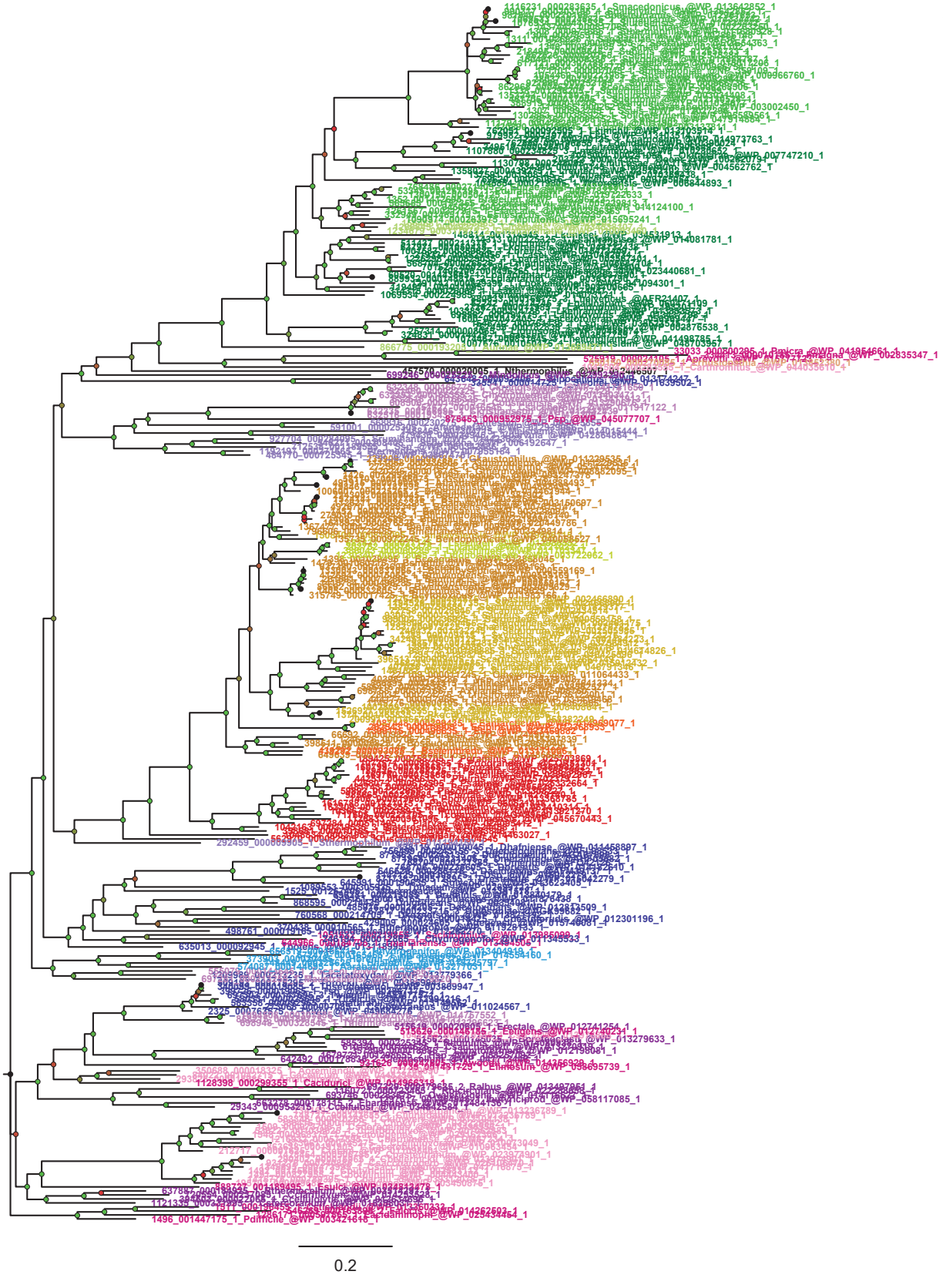

457570_000020005_1_Nthermophilus_@WP_041367332_1
0.4

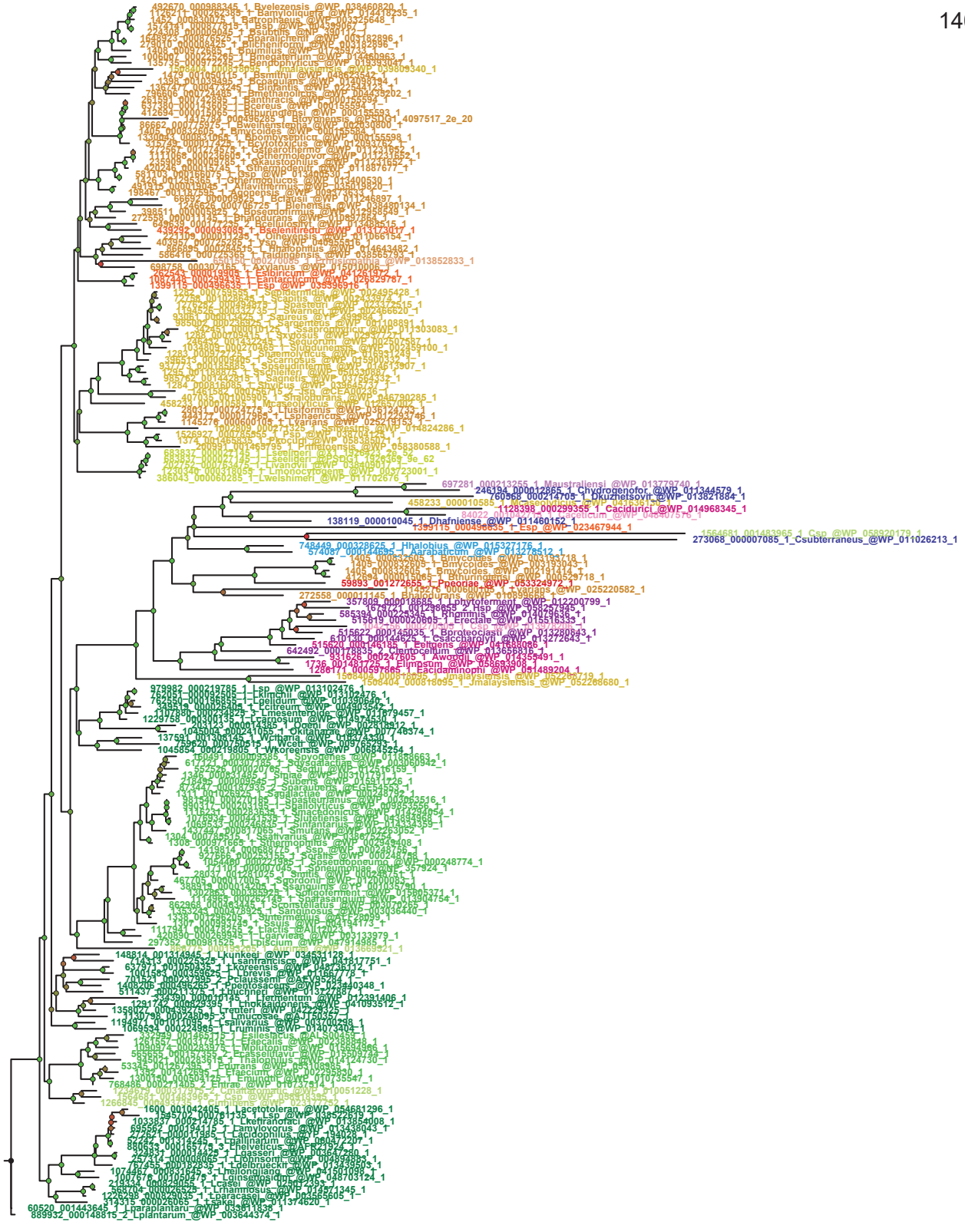

1415775_000789395_1_Cbaratii_@WP_040113741_1
515620_000146185_1_Eeligens_@WP_012744484_1
2.0

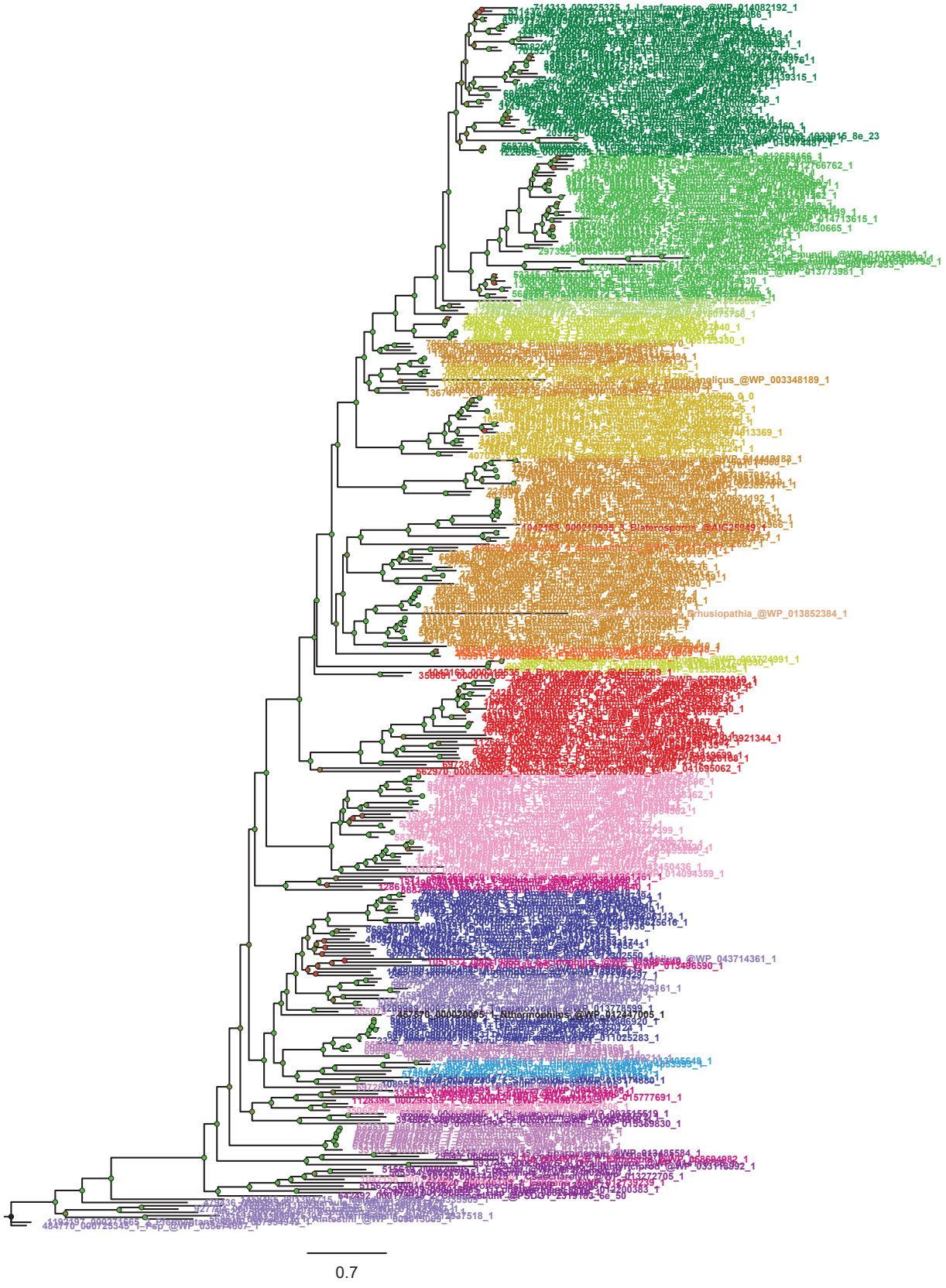

0.4

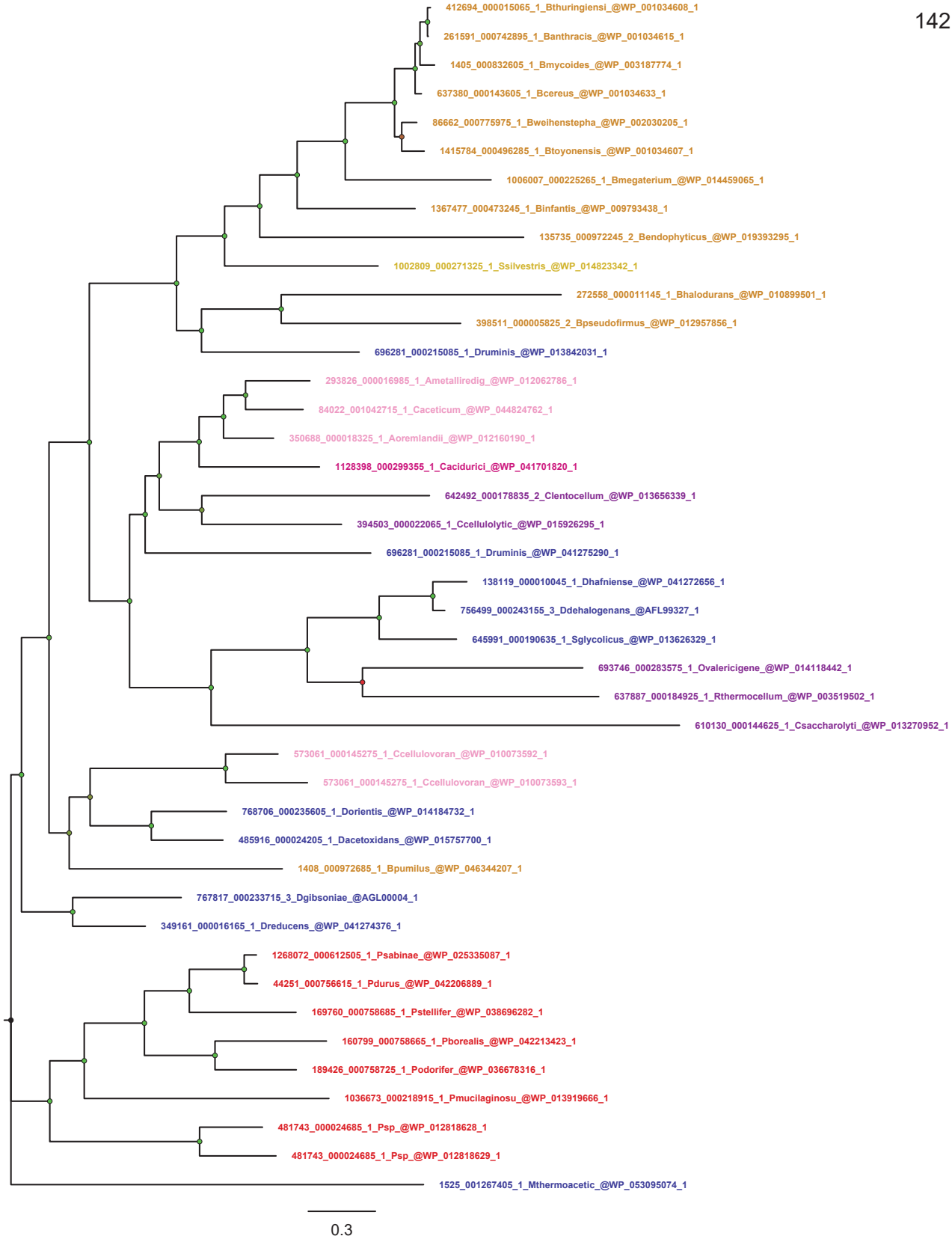

0.4

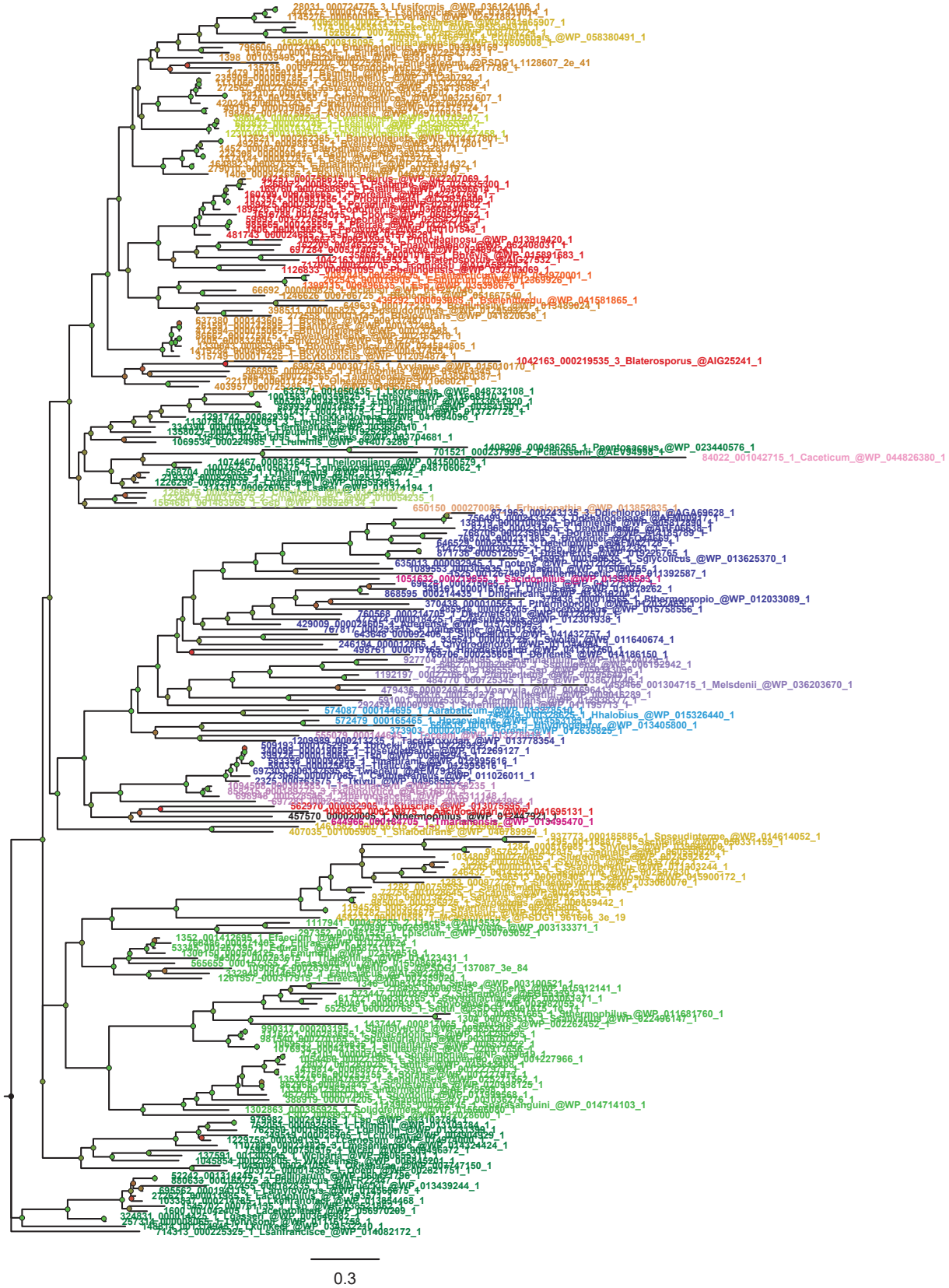

0.4

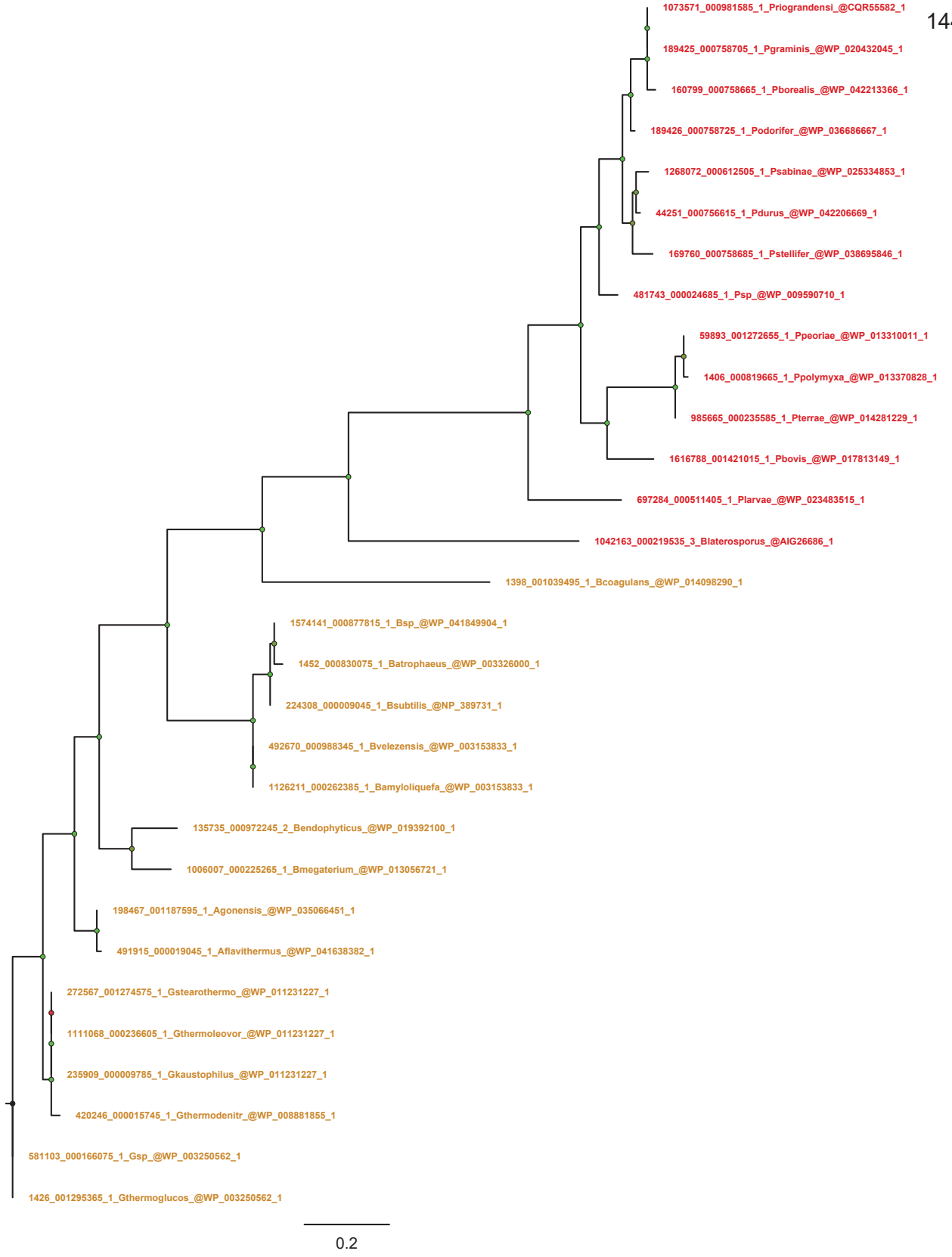

0.2

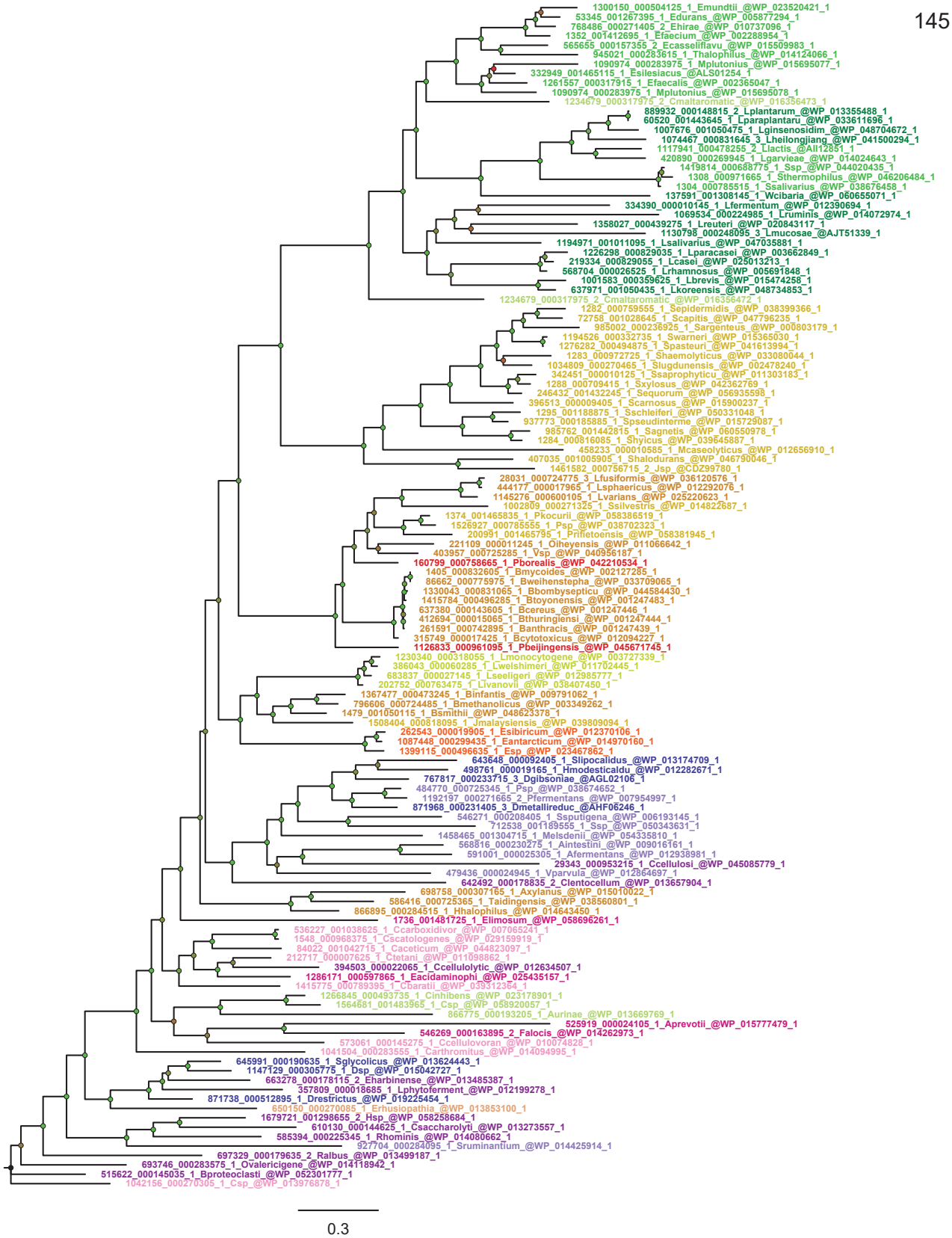

0.6

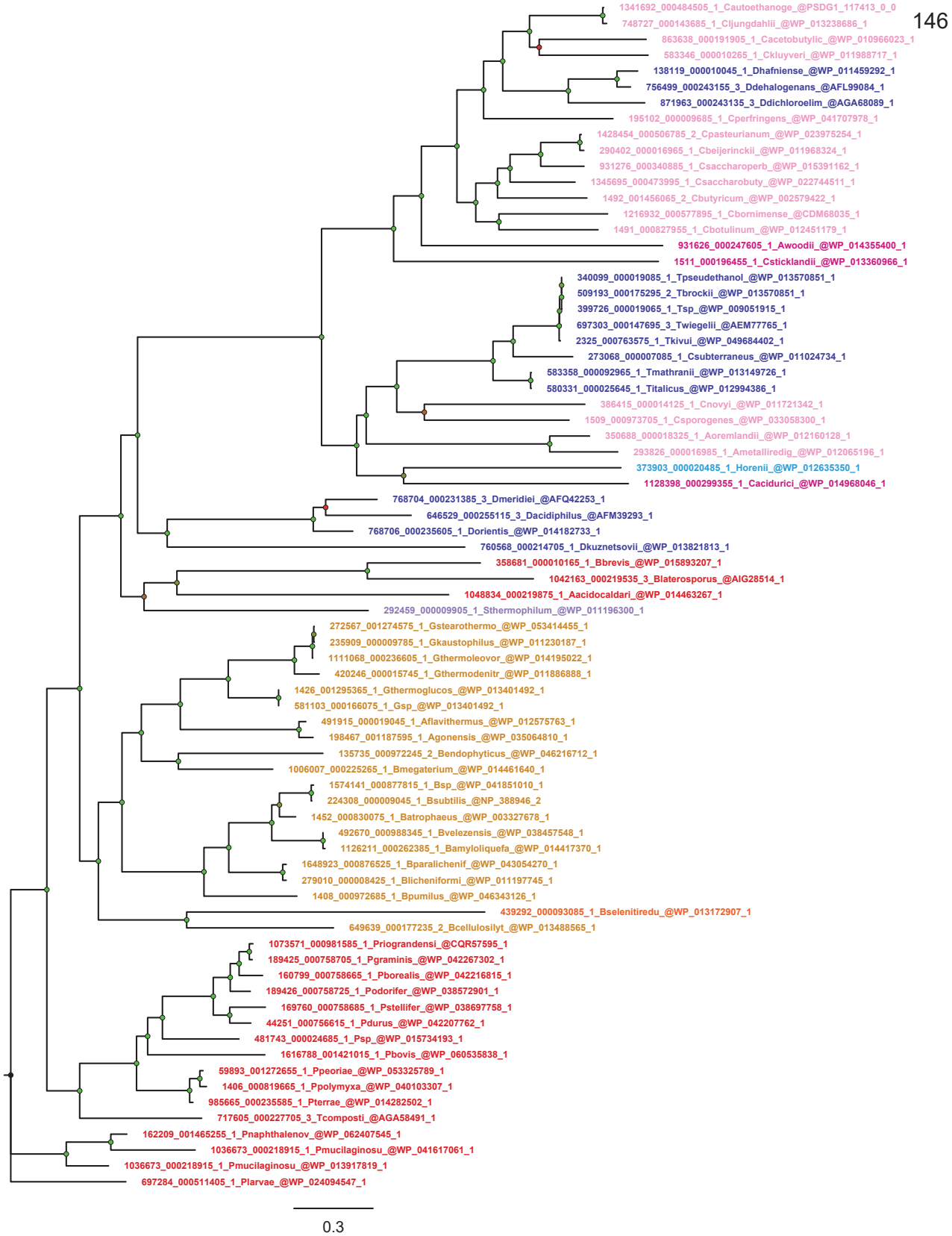

0.7

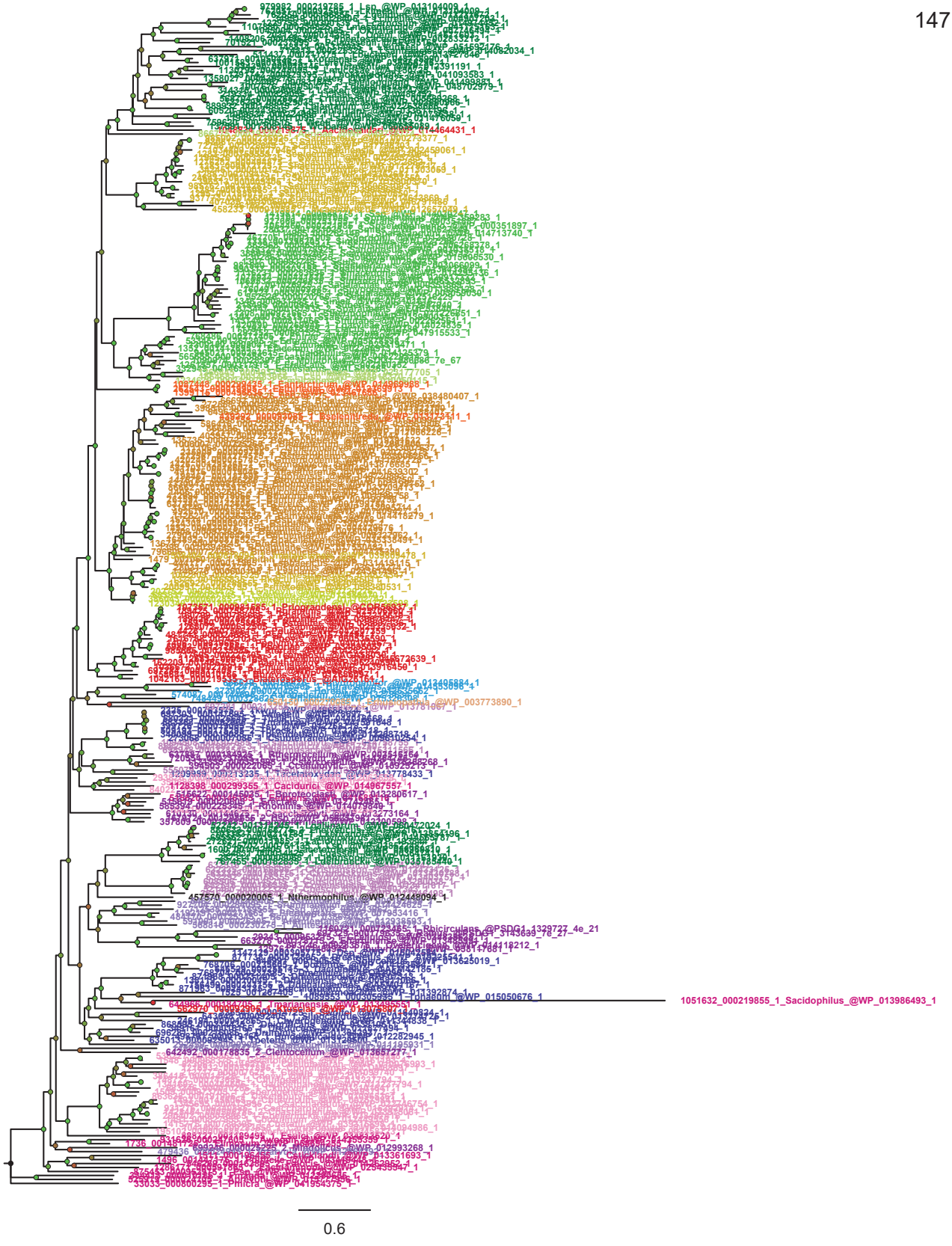

412694_000015065_1_Bthuringiensi_@WP_001034608_1
261591_000742895_1_Banthracis_@WP_001034615_1
1405_000832605_1_Bmycoides_@WP_003187774_1
637380_000143605_1_Bcereus_@WP_001034633_1
86662_000775975_1_Bweihenstepha_@WP_002030205_1
1415784_000496285_1_Btoyonensis_@WP_001034607_1
1006007_000225265_1_Bmegaterium_@WP_014459065_1
1367477_000473245_1_Binfantis_@WP_009793438_1
135735_000972245_2_Bendophyticus_@WP_019393295_1
1002809_000271325_1_Ssilvestris_@WP_014823342_1
272558_000011145_1_Bhalodurans_@WP_010899501_1
398511_000005825_2_Bpseudofirmus_@WP_012957856_1
696281_000215085_1_Druminis_@WP_013842031_1
293826_000016985_1_Ametalliredig_@WP_012062786_1
84022_001042715_1_Caceticum_@WP_044824762_1
350688_000018325_1_Aoremlandii_@WP_012160190_1
1128398_000299355_1_Cacidurici_@WP_041701820_1
642492_000178835_2_Clentocellum_@WP_013656339_1
394503_000022065_1_Ccellulolytic_@WP_015926295_1
696281_000215085_1_Druminis_@WP_041275290_1
138119_000010045_1_Dhafniense_@WP_041272656_1
756499_000243155_3_Ddehalogenans_@AFL99327_1
645991_000190635_1_Sglycolicus_@WP_013626329_1
693746_000283575_1_Ovalericigene_@WP_014118442_1
637887_000184925_1_Rthermocellum_@WP_003519502_1
610130_000144625_1_Csaccharolyti_@WP_013270952_1
573061_000145275_1_Ccellulovoran_@WP_010073592_1
573061_000145275_1_Ccellulovoran_@WP_010073593_1
768706_000235605_1_Dorientis_@WP_014184732_1
485916_000024205_1_Dacetoxidans_@WP_015757700_1
1408_000972685_1_Bpumilus_@WP_046344207_1
767817_000233715_3_Dgibsoniae_@AGL00004_1
349161_000016165_1_Dreducens_@WP_041274376_1
1268072_000612505_1_Psabinae_@WP_025335087_1
44251_000756615_1_Pdurus_@WP_042206889_1
169760_000758685_1_Pstellifer_@WP_038696282_1
160799_000758665_1_Pborealis_@WP_042213423_1
189426_000758725_1_Podorifer_@WP_036678316_1
1036673_000218915_1_Pmucilaginosu_@WP_013919666_1
481743_000024685_1_Psp_@WP_012818628_1
481743_000024685_1_Psp_@WP_012818629_1
1525_001267405_1_Mthermoacetic_@WP_053095074_1
0.3

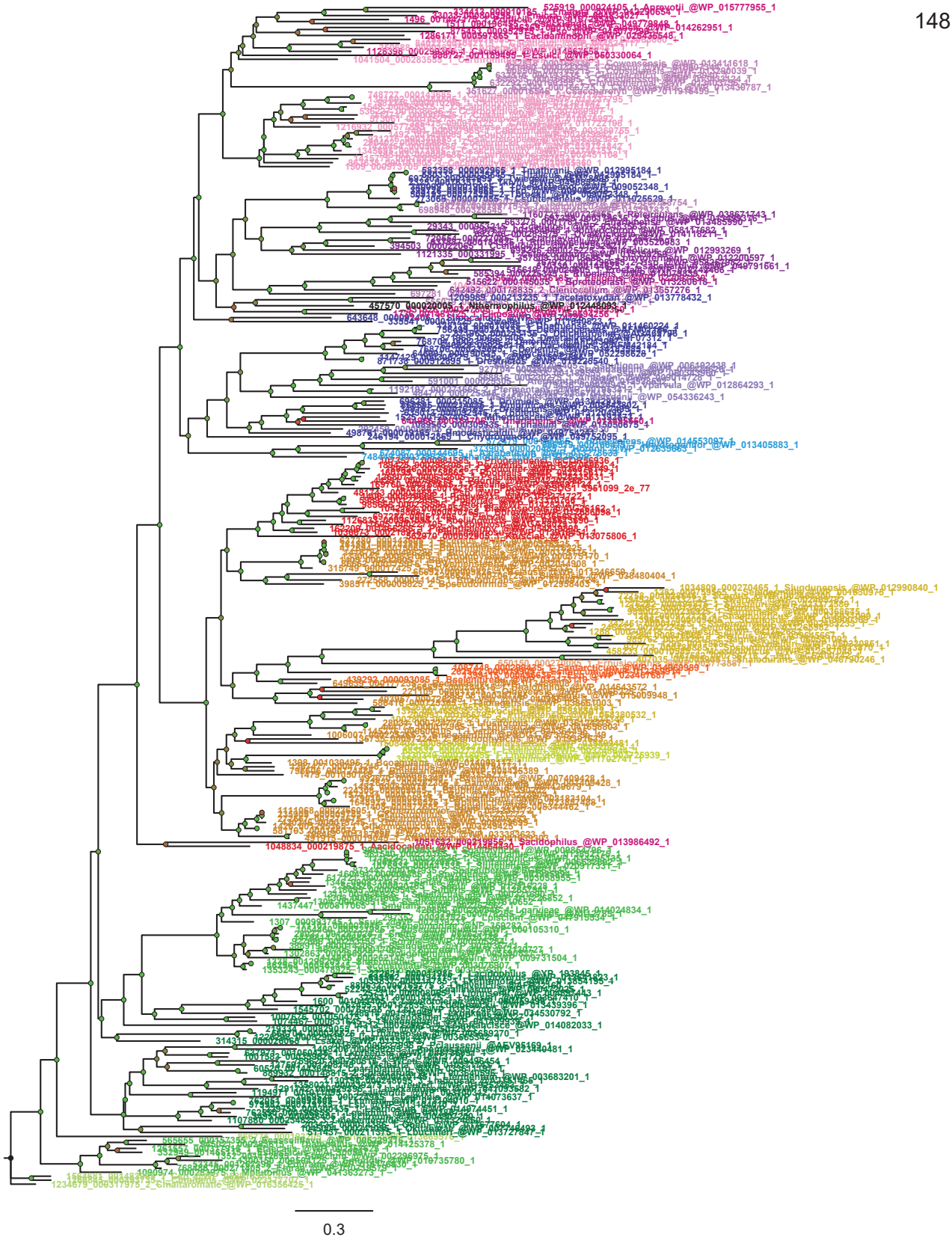

1042163_000219535_3_Blaterosporus_@AIG25241_1
84022_001042715_1_Caceticum_@WP_044826380_1
650150_000270085_1_Ethnolinealia_@WP_013550235_1
1042163_000219535_3_Blaterosporus_@AIG25241_1
1042163_000219535_3_Blaterosporus_@AIG25241_1
0.3

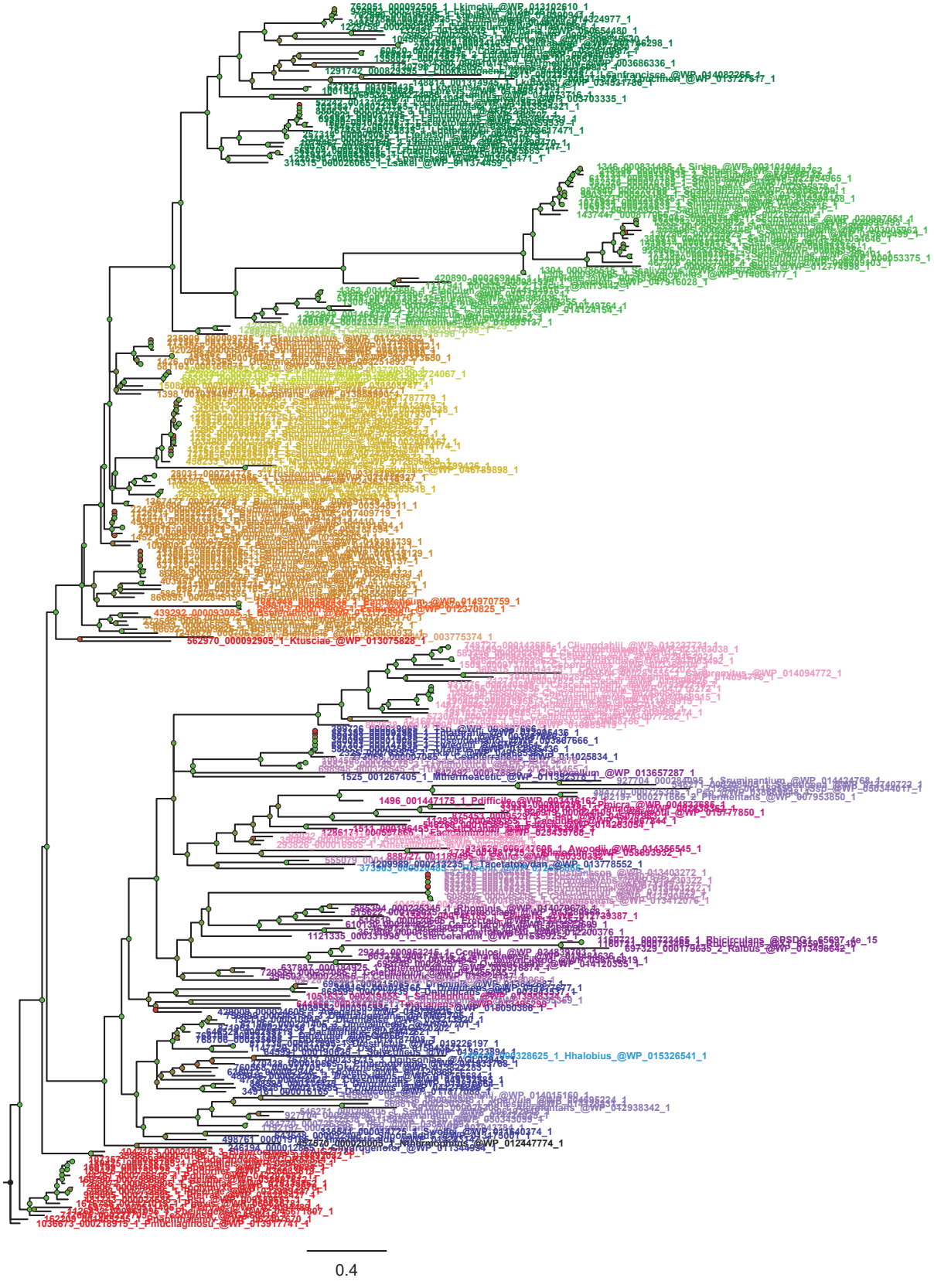

1073571_000981585_1_Priograndensi_@CQR55582_1
189425_000758705_1_Pgraminis_@WP_020432045_1
160799_000758665_1_Pborealis_@WP_042213366_1
189426_000758725_1_Podorifer_@WP_036686667_1
1268072_000612505_1_Psabinae_@WP_025334853_1
44251_000756615_1_Pdurus_@WP_042206669_1
169760_000758685_1_Pstellifer_@WP_038695846_1
481743_000024685_1_Psp_@WP_009590710_1
59893_001272655_1_Ppeoriae_@WP_013310011_1
1406_000819665_1_Ppolymyxa_@WP_013370828_1
985665_000235585_1_Pterrae_@WP_014281229_1
1616788_001421015_1_Pbovis_@WP_017813149_1
697284_000511405_1_Plarvae_@WP_023483515_1
1042163_000219535_3_Blaterosporus_@AIG26686_1
1398_001039495_1_Bcoagulans_@WP_014098290_1
1574141_000877815_1_Bsp_@WP_041849904_1
1452_000830075_1_Batrophaeus_@WP_003326000_1
224308_000009045_1_Bsubtilis_@NP_389731_1
492670_000988345_1_Bvelezensis_@WP_003153833_1
1126211_000262385_1_Bamyloliquefa_@WP_003153833_1
135735_000972245_2_Bendophyticus_@WP_019392100_1
1006007_000225265_1_Bmegaterium_@WP_013056721_1
198467_001187595_1_Agonensis_@WP_035066451_1
491915_000019045_1_Aflavithermus_@WP_041638382_1
272567_001274575_1_Gstearothermo_@WP_011231227_1
1111068_000236605_1_Gthermoleovor_@WP_011231227_1
235909_000009785_1_Gkaustophilus_@WP_011231227_1
420246_000015745_1_Gthermodenitr_@WP_008881855_1
581103_000166075_1_Gsp_@WP_003250562_1
1426_001295365_1_Gthermoglucos_@WP_003250562_1
0.2

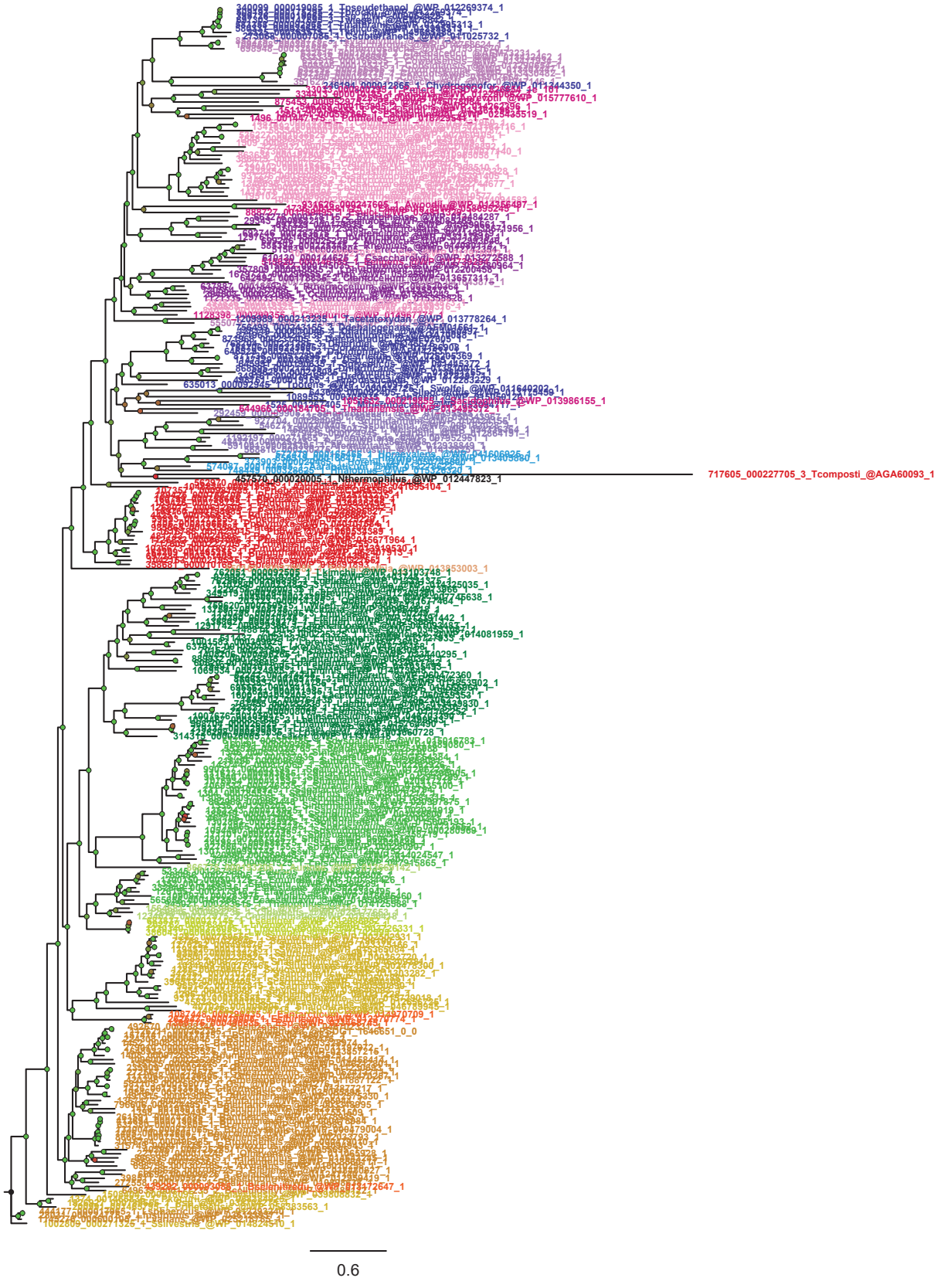

1300150_000504125_1_Emundtii_@WP_023520421_1
53345_001267395_1_Edurans_@WP_005877294_1
768486_000271405_2_Ehirae_@WP_010737096_1
1352_001412695_1_Efaecium_@WP_002288954_1
565655_000157355_2_Ecasseliflavu_@WP_015509983_1
945021_000283615_1_Thalophilus_@WP_014124066_1
1090974_000283975_1_Mplutonius_@WP_015695077_1
332949_001465115_1_Esilesiacus_@ALS01254_1
1261557_000317915_1_Efaecalis_@WP_002365047_1
1090974_000283975_1_Mplutonius_@WP_015695078_1
1234679_000317975_2_Cmaltaromatic_@WP_016356473_1
889932_000148815_2_Lplantarum_@WP_013355488_1
60520_001443645_1_Lparaplantaru_@WP_033611696_1
1007676_001050475_1_Lginsenosidim_@WP_048704672_1
1074467_000831645_3_Lheilongjiang_@WP_041500294_1
1117941_000478255_2_Llactis_@AII12851_1
420890_000269945_1_Lgarvieae_@WP_014024643_1
1419814_000688775_1_Ssp_@WP_044020435_1
1308_000971665_1_Sthermophilus_@WP_046206484_1
1304_000785515_1_Ssalivarius_@WP_038676458_1
137591_001308145_1_Wcibaria_@WP_060655071_1
334390_000010145_1_Lfermentum_@WP_012390694_1
1069534_000224985_1_Lruminis_@WP_014072974_1
1358027_000439275_1_Lreuteri_@WP_020843117_1
1130798_000248095_3_Lmucosae_@AJT51339_1
1194971_001011095_1_Lsalivarius_@WP_047035881_1
1226298_000829035_1_Lparacasei_@WP_003662849_1
219334_000829055_1_Lcasei_@WP_025013213_1
568704_000026525_1_Lrhamnosus_@WP_005691848_1
1001583_000359625_1_Lbrevis_@WP_015474258_1
637971_001050435_1_Lkoreensis_@WP_048734853_1
1234679_000317975_2_Cmaltaromatic_@WP_016356472_1
1282_000759555_1_Sepidermidis_@WP_038399366_1
72758_001028645_1_Scapitis_@WP_047796235_1
985002_000236925_1_Sargenteus_@WP_000803179_1
1194526_000332735_1_Swarneri_@WP_015365030_1
1276282_000494875_1_Spasteuri_@WP_041613994_1
1283_000972725_1_Shaemolyticus_@WP_033080044_1
1034809_000270465_1_Slugdunensis_@WP_002478240_1
342451_000010125_1_Ssaprophyticu_@WP_011303183_1
1288_000709415_1_Sxylosus_@WP_042362769_1
246432_001432245_1_Sequorum_@WP_056935598_1
396513_000009405_1_Scarnosus_@WP_015900237_1
1295_001188875_1_Sschleiferi_@WP_050331048_1
937773_000185885_1_Spseudinterme_@WP_015729087_1
985762_001442815_1_Sagnetis_@WP_060550978_1
1284_000816085_1_Shyicus_@WP_039645887_1
458233_000010585_1_Mcaseolyticus_@WP_012656910_1
407035_001005905_1_Shalodurans_@WP_046790046_1
1461582_000756715_2_Jsp_@CDZ99780_1
28031_000724775_3_Lfusiformis_@WP_036120576_1
444177_000017965_1_Lsphaericus_@WP_012292076_1
1145276_000600105_1_Lvarians_@WP_025220623_1
1002809_000271325_1_Ssilvestris_@WP_014822687_1
1374_001465835_1_Pkocurii_@WP_058386519_1
1526927_000785555_1_Psp_@WP_038702323_1
200991_001465795_1_Prifletoensis_@WP_058381945_1
221109_000011245_1_Oiheyensis_@WP_011066642_1
403957_000725285_1_Vsp_@WP_040956187_1
160799_000758665_1_Pborealis_@WP_042210534_1
1405_000832605_1_Bmycoides_@WP_002127285_1
86662_000775975_1_Bweihenstepha_@WP_033709065_1
1330043_000831065_1_Bbombysepticu_@WP_044584430_1
1415784_000496285_1_Btoyonensis_@WP_001247483_1
637380_000143605_1_Bcereus_@WP_001247446_1
412694_000015065_1_Bthuringiensi_@WP_001247444_1
261591_000742895_1_Banthracis_@WP_001247439_1
315749_000017425_1_Bcytotoxicus_@WP_012094227_1
1126833_000961095_1_Pbeijingensis_@WP_045671745_1
1230340_000318055_1_Lmonocytogene_@WP_003727339_1
386043_000060285_1_Lwelshimeri_@WP_011702445_1
683837_000027145_1_Lseeligeri_@WP_012985777_1
202752_000763475_1_Livanovii_@WP_038407450_1
1367477_000473245_1_Binfantis_@WP_009791062_1
796606_000724485_1_Bmethanolicus_@WP_003349262_1
1479_001050115_1_Bsmithii_@WP_048623378_1
1508404_000818095_1_Jmalaysiensis_@WP_039809094_1
262543_000019905_1_Esibiricum_@WP_012370106_1
1087448_000299435_1_Eantarcticum_@WP_014970160_1
1399115_000496635_1_Esp_@WP_023467862_1
643648_000092405_1_Slipocalidus_@WP_013174709_1
498761_000019165_1_Hmodesticaldu_@WP_012282671_1
767817_000233715_3_Dgibsoniae_@AGL02106_1
484770_000725345_1_Psp_@WP_038674652_1
1192197_000271665_2_Pfermentans_@WP_007954997_1
871968_000231405_3_Dmetallireduc_@AHF06246_1
546271_000208405_1_Ssputigena_@WP_006193145_1
712538_001189555_1_Ssp_@WP_050343631_1
1458465_001304715_1_Melsdenii_@WP_054335810_1
568816_000230275_1_Aintestini_@WP_009016761_1
591001_000025305_1_Afermentans_@WP_012938981_1
29343_000953215_1_Ccellulosi_@WP_045085779_1
479436_000024945_1_Vparvula_@WP_012864697_1
642492_000178835_2_Clentocellum_@WP_013657904_1
698758_000307165_1_Axylanus_@WP_015010022_1
586416_000725365_1_Taidingensis_@WP_038560801_1
866895_000284515_1_Hhalophilus_@WP_014643450_1
1736_001481725_1_Elimosum_@WP_058696261_1
536227_001038625_1_Ccarboxidivor_@WP_007065241_1
1548_000968375_1_Cscatologenes_@WP_029159919_1
84022_001042715_1_Caceticum_@WP_044823097_1
212717_000007625_1_Ctetani_@WP_011098862_1
394503_000022065_1_Ccellulolytic_@WP_012634507_1
1286171_000597865_1_Eacidaminophi_@WP_025435157_1
141575_000789395_1_Cbaratii_@WP_039312364_1
1266845_000493735_1_Cinhibens_@WP_023178901_1
1564681_001483565_1_Csp_@WP_058920057_1
866775_000193205_1_Aurinae_@WP_013669769_1
525919_000024105_1_Aprevotii_@WP_015777479_1
546269_000163895_2_Falocis_@WP_014262973_1
573061_000145275_1_Ccellulovoran_@WP_010074828_1
1041504_000283555_1_Carthromitus_@WP_014094995_1
645991_000190635_1_Sglycolicus_@WP_013624443_1
1147129_000305775_1_Dsp_@WP_015042727_1
663278_000178115_2_Eharbinense_@WP_013485387_1
357809_000018685_1_Lphytoferment_@WP_012199278_1
871738_000512895_1_Drestrictus_@WP_019225454_1
650150_000270085_1_Erhusiopathia_@WP_013853100_1
1679721_001298655_2_Hsp_@WP_058258684_1
610130_000144625_1_Csaccharolyti_@WP_013273557_1
585394_000225345_1_Rhominis_@WP_014080662_1
927704_000284095_1_Sruminantium_@WP_014425914_1
697329_000179635_2_Ralbus_@WP_013499187_1
693746_000283575_1_Ovalericigene_@WP_014118942_1
515622_000145035_1_Bproteoclasti_@WP_052301777_1
1042156_000270305_1_Csp_@WP_013976878_1
0.3

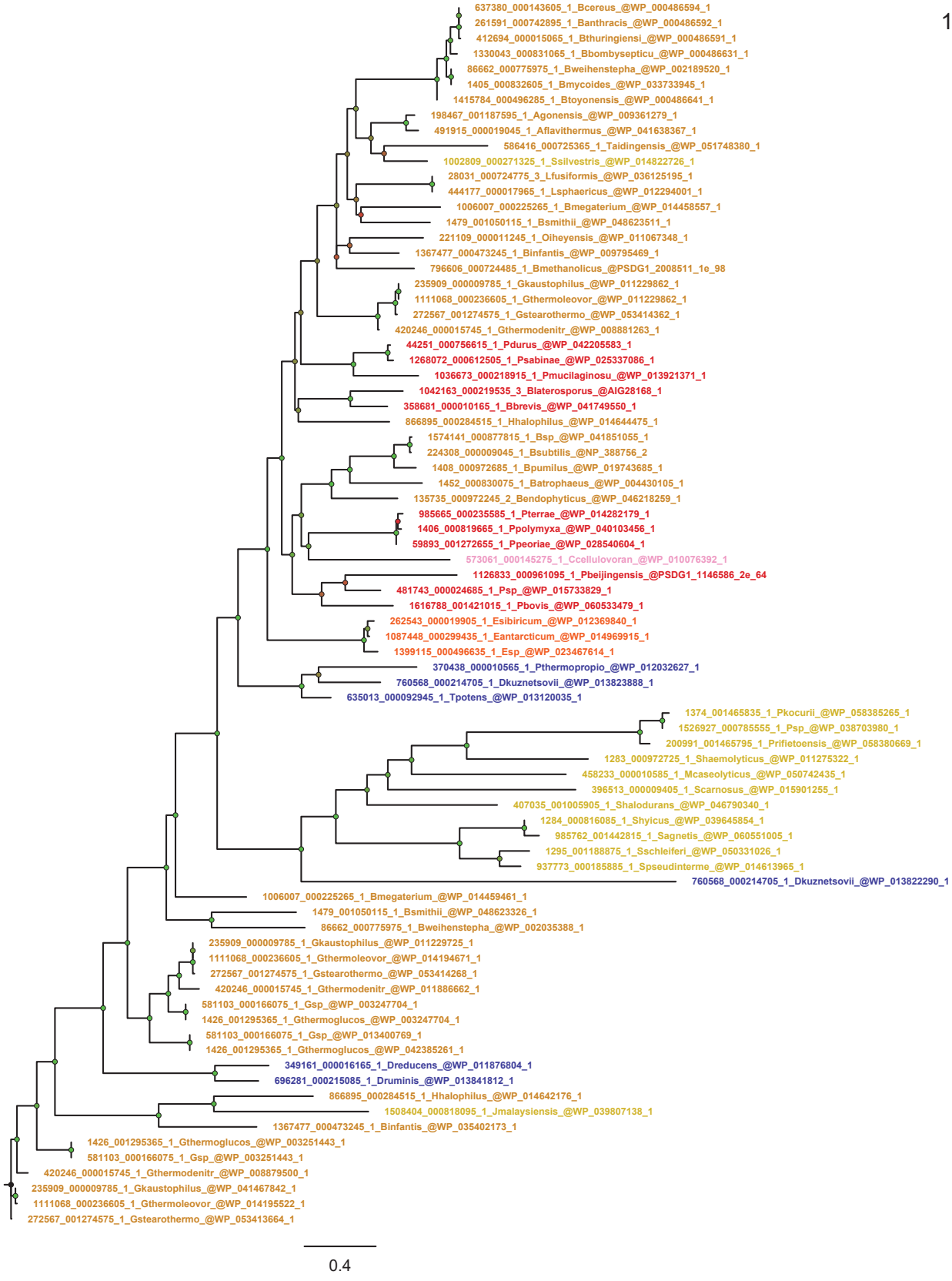

1341692_000484505_1_Cautoethanoge_@PSDG1_117413_0_0
748727_000143685_1_Cljungdahlii_@WP_013238686_1
863638_000191905_1_Cacetobutylic_@WP_010966023_1
583346_000010265_1_Ckluyveri_@WP_011988717_1
138119_000010045_1_Dhafniense_@WP_011459292_1
756499_000243155_3_Ddehalogenans_@AFL99084_1
871963_000243135_3_Ddichloroelim_@AGA68089_1
195102_000009685_1_Cperfringens_@WP_041707978_1
1428454_000506785_2_Cpasteurianum_@WP_023975254_1
290402_000016965_1_Cbeijerinckii_@WP_011968324_1
931276_000340885_1_Csaccharoperb_@WP_015391162_1
1345695_000473995_1_Csaccharobuty_@WP_022744511_1
1492_001456065_2_Cbutyricum_@WP_002579422_1
1216932_000577895_1_Cbornimense_@CDM68035_1
1491_000827955_1_Cbotulinum_@WP_012451179_1
931626_000247605_1_Awoodii_@WP_014355400_1
1511_000196455_1_Csticklandii_@WP_013360966_1
340099_000019085_1_Tpseudethanol_@WP_013570851_1
509193_000175295_2_Tbrockii_@WP_013570851_1
399726_000019065_1_Tsp_@WP_009051915_1
697303_000147695_3_Twiegelii_@AEM77765_1
2325_000763575_1_Tkivui_@WP_049684402_1
273068_000007085_1_Csubterraneus_@WP_011024734_1
583358_000092965_1_Tmathranii_@WP_013149726_1
580331_000025645_1_Titalicus_@WP_012994386_1
386415_000014125_1_Cnovyi_@WP_011721342_1
1509_000973705_1_Csporogenes_@WP_033058300_1
350688_000018325_1_Aoremlandii_@WP_012160128_1
293826_000016985_1_Ametalliredig_@WP_012065196_1
373903_000020485_1_Horenii_@WP_012635350_1
1128398_000299355_1_Cacidurici_@WP_014968046_1
768704_000231385_3_Dmeridiei_@AFQ42253_1
646529_000255115_3_Dacidiphilus_@AFM39293_1
768706_000235605_1_Dorientis_@WP_014182733_1
760568_000214705_1_Dkuznetsovii_@WP_013821813_1
358681_000010165_1_Bbrevis_@WP_015893207_1
1042163_000219535_3_Blaterosporus_@AIG28514_1
1048834_000219875_1_Aacidocaldari_@WP_014463267_1
292459_000009905_1_Sthermophilum_@WP_011196300_1
272567_001274575_1_Gstearothermo_@WP_053414455_1
235909_000009785_1_Gkaustophilus_@WP_011230187_1
1111068_000236605_1_Gthermoleovor_@WP_014195022_1
420246_000015745_1_Gthermodenitr_@WP_011886888_1
1426_001295365_1_Gthermoglucos_@WP_013401492_1
581103_000166075_1_Gsp_@WP_013401492_1
491915_000019045_1_Aflavithermus_@WP_012575763_1
198467_001187595_1_Agonensis_@WP_035064810_1
135735_000972245_2_Bendophyticus_@WP_046216712_1
1006007_000225265_1_Bmegaterium_@WP_014461640_1
1574141_000877815_1_Bsp_@WP_041851010_1
224308_000009045_1_Bsubtilis_@NP_388946_2
1452_000830075_1_Batrophaeus_@WP_003327678_1
492670_000988345_1_Bvelezensis_@WP_038457548_1
1126211_000262385_1_Bamyloliquefa_@WP_014417370_1
1648923_000876525_1_Bparalichenif_@WP_043054270_1
279010_000008425_1_Blicheniformi_@WP_011197745_1
1408_000972685_1_Bpumilus_@WP_046343126_1
439292_000093085_1_Bselenitiredu_@WP_013172907_1
649639_000177235_2_Bcellulosilyt_@WP_013488565_1
1073571_000981585_1_Priograndensi_@CQR57595_1
189425_000758705_1_Pgraminis_@WP_042267302_1
160799_000758665_1_Pborealis_@WP_042216815_1
189426_000758725_1_Podorifer_@WP_038572901_1
169760_000758685_1_Pstellifer_@WP_038697758_1
44251_000756615_1_Pdurus_@WP_042207762_1
481743_000024685_1_Psp_@WP_015734193_1
1616788_001421015_1_Pbovis_@WP_060535838_1
59893_001272655_1_Ppeoriae_@WP_053325789_1
1406_000819665_1_Ppolymyxa_@WP_040103307_1
985665_000235585_1_Pterrae_@WP_014282502_1
717605_000227705_3_Tcomposti_@AGA58491_1
162209_001465255_1_Pnaphthalenov_@WP_062407545_1
1036673_000218915_1_Pmucilaginosu_@WP_041617061_1
1036673_000218915_1_Pmucilaginosu_@WP_013917819_1
697284_000511405_1_Plarvae_@WP_024094547_1
0.3

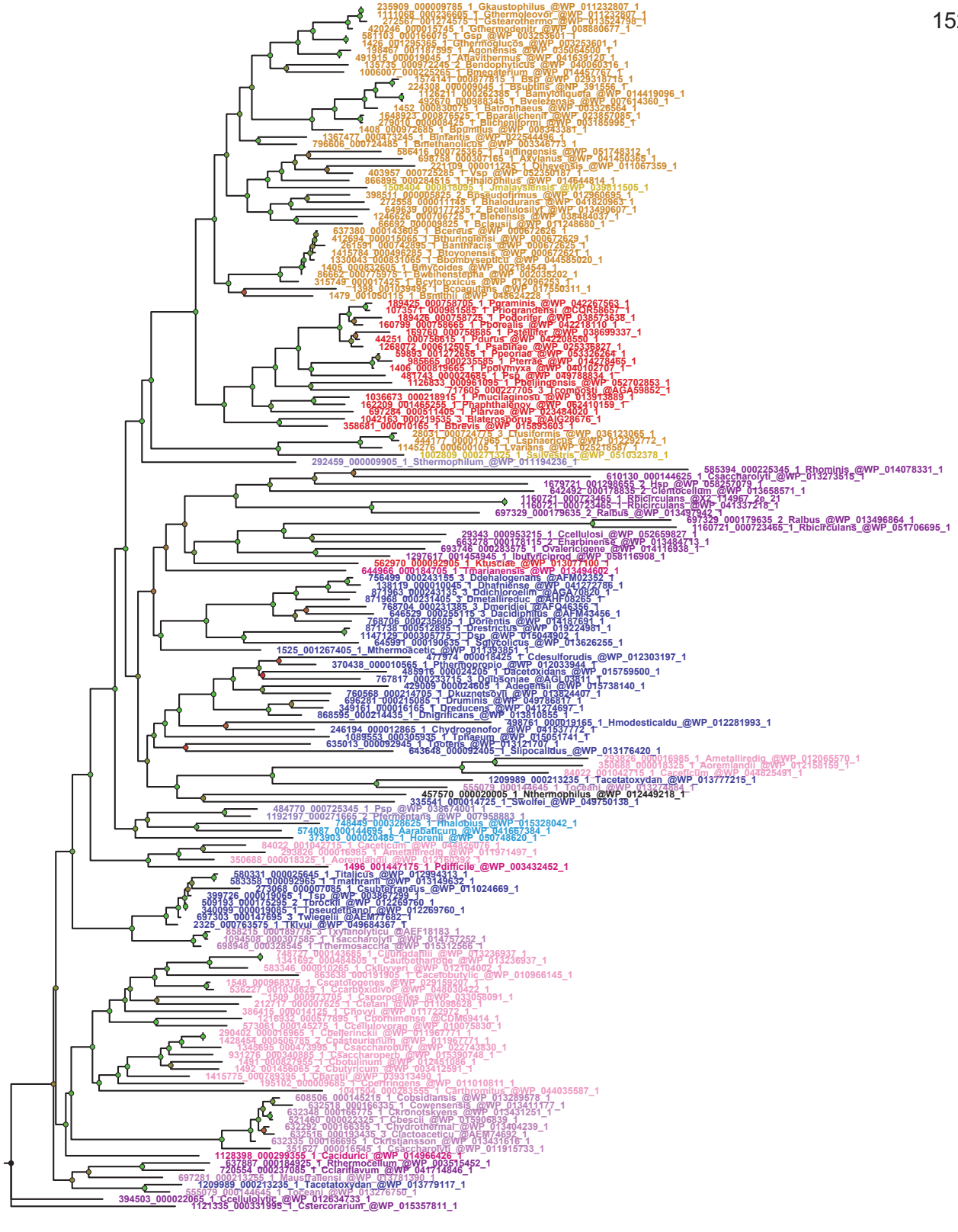

457570_000020005_1_Nthermophilus_@WP_012448094_1
1051632_000219855_1_Sacidophilus_@WP_013986493_1
0.6

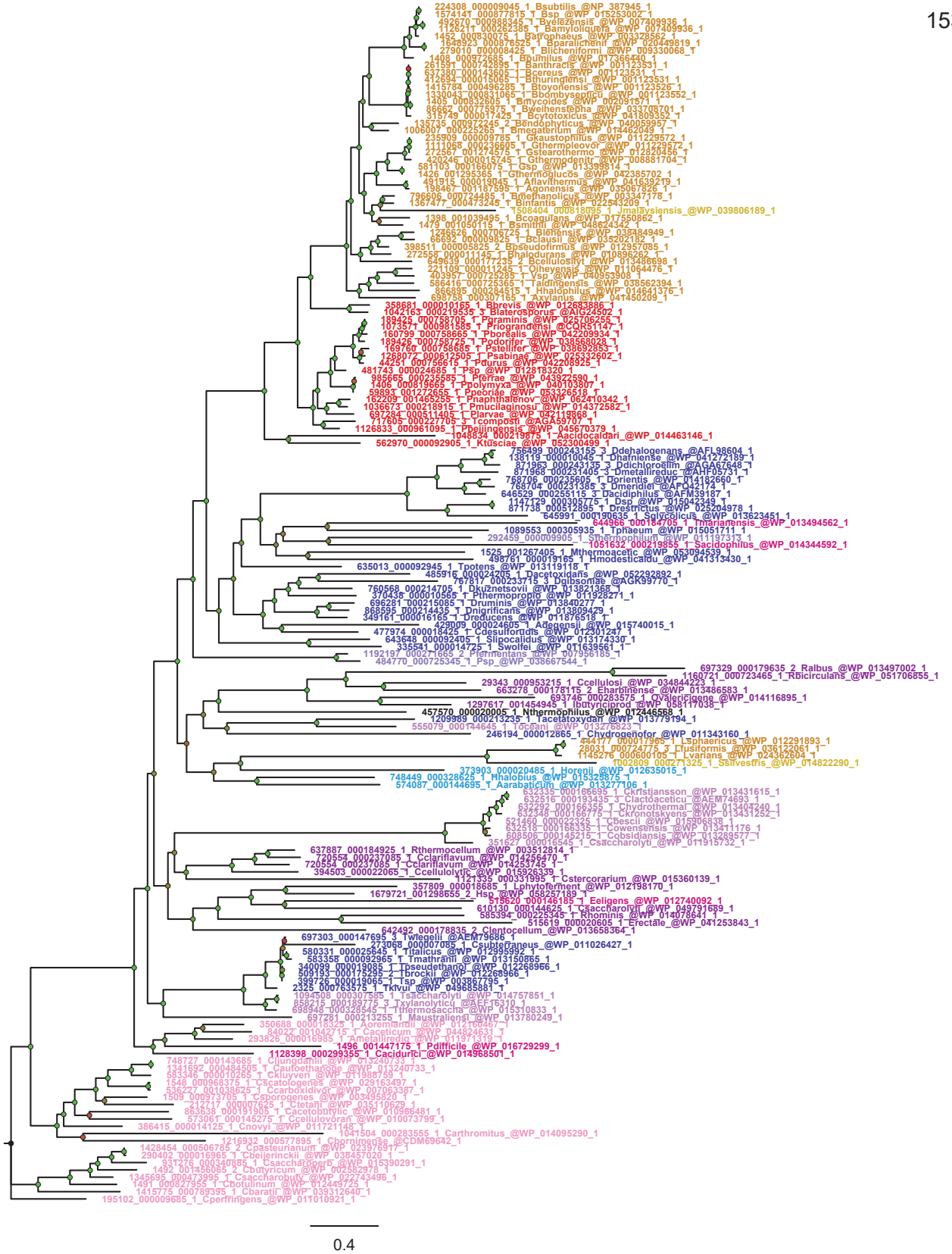

0.3

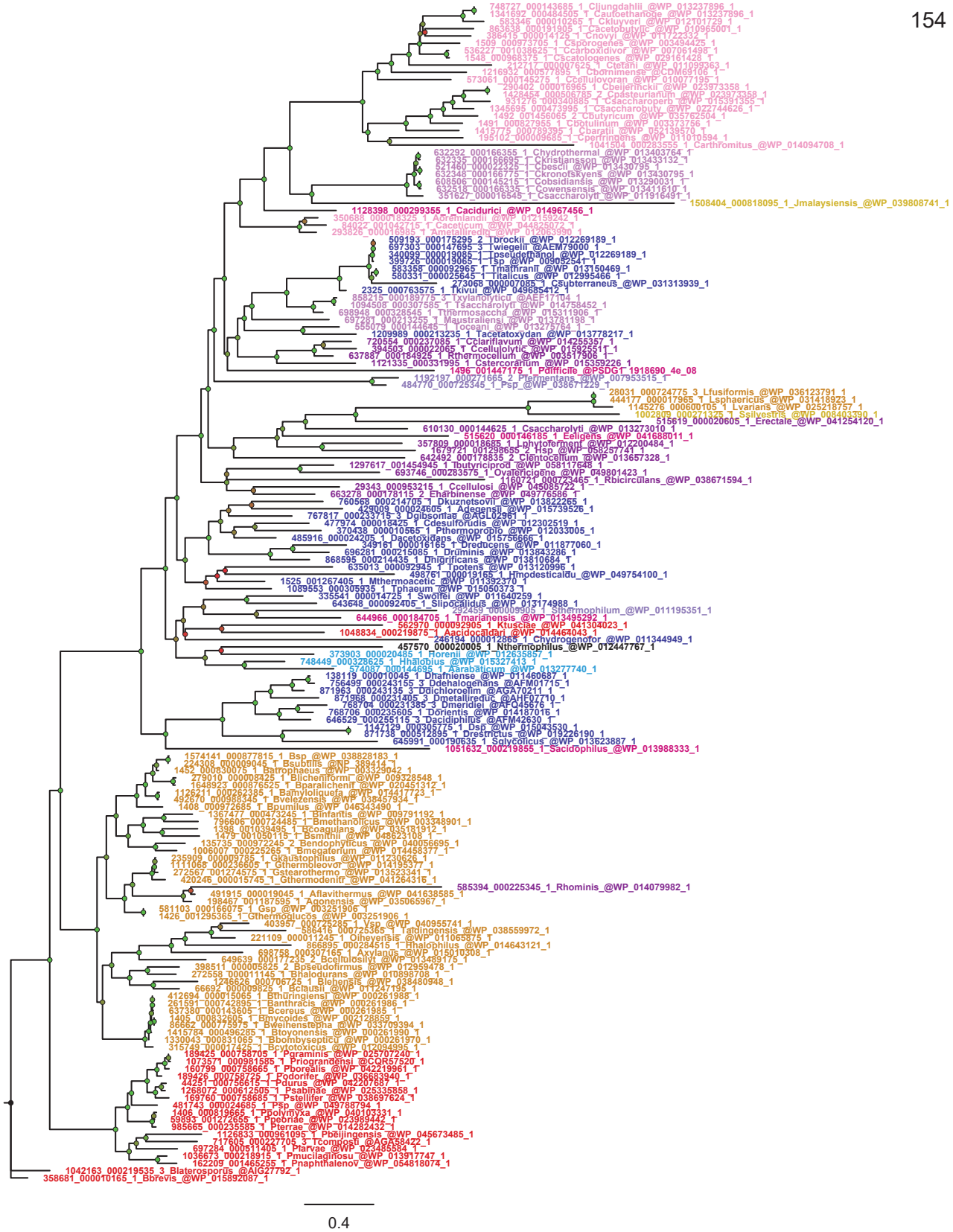

0.4

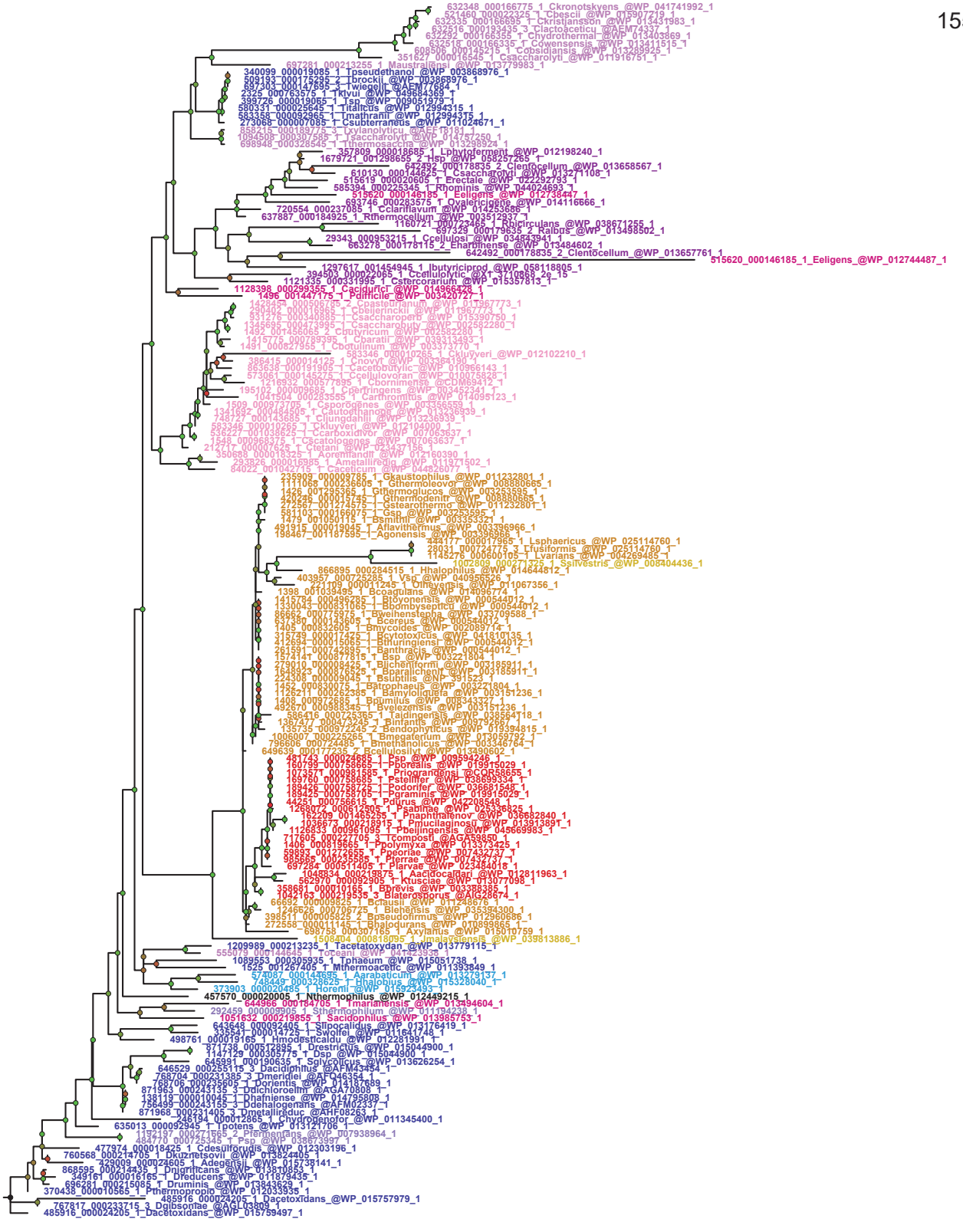

717605_000227705_3_Tcomposti_@AGA60093_1
0.6

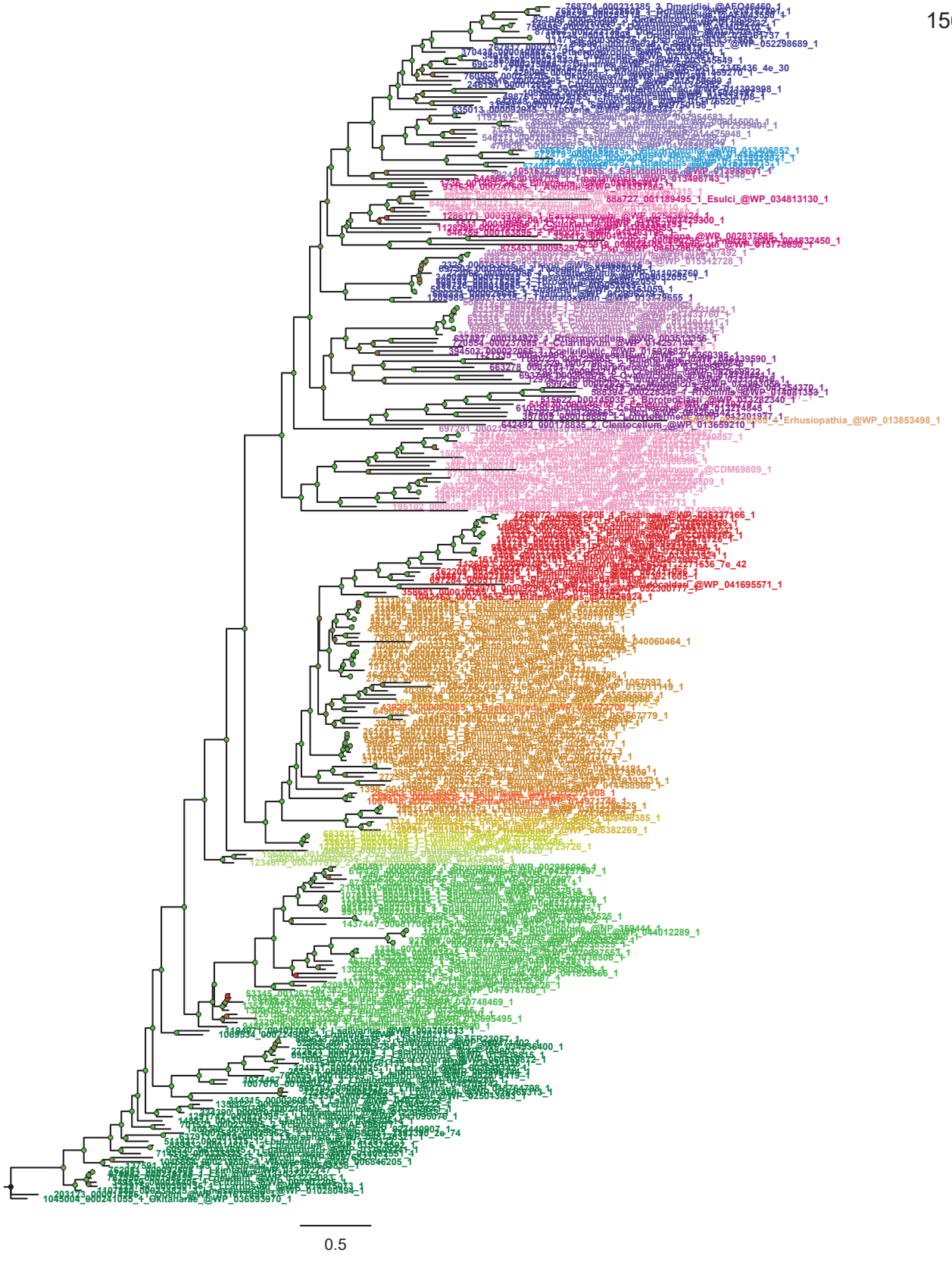

637380_000143605_1_Bcereus_@WP_000486594_1
261591_000742895_1_Banthracis_@WP_000486592_1
412694_000015065_1_Bthuringiensi_@WP_000486591_1
1330043_000831065_1_Bbombysepticu_@WP_000486631_1
86662_000775975_1_Bweihenstepha_@WP_002189520_1
1405_000832605_1_Bmycoides_@WP_033733945_1
1415784_000496285_1_Btoyonensis_@WP_000486641_1
198467_001187595_1_Agonensis_@WP_009361279_1
491915_000019045_1_Aflavithermus_@WP_041638367_1
586416_000725365_1_Taidingensis_@WP_051748380_1
1002809_000271325_1_Ssilvestris_@WP_014822726_1
28031_000724775_3_Lfusiformis_@WP_036125195_1
444177_000017965_1_Lsphaericus_@WP_012294001_1
1006007_000225265_1_Bmegaterium_@WP_014458557_1
1479_001050115_1_Bsmithii_@WP_048623511_1
221109_000011245_1_Oiheyensis_@WP_011067348_1
1367477_000473245_1_Binfantis_@WP_009795469_1
796606_000724485_1_Bmethanolicus_@PSDG1_2008511_1e_98
235909_000009785_1_Gkaustophilus_@WP_011229862_1
1111068_000236605_1_Gthermoleovor_@WP_011229862_1
272567_001274575_1_Gstearothermo_@WP_053414362_1
420246_000015745_1_Gthermodenitr_@WP_008881263_1
44251_000756615_1_Pdurus_@WP_042205583_1
1268072_000612505_1_Psabinae_@WP_025337086_1
1036673_000218915_1_Pmucilaginosu_@WP_013921371_1
1042163_000219535_3_Blaterosporus_@AIG28168_1
358681_000010165_1_Bbrevis_@WP_041749550_1
866895_000284515_1_Hhalophilus_@WP_014644475_1
1574141_000877815_1_Bsp_@WP_041851055_1
224308_000009045_1_Bsubtilis_@NP_388756_2
1408_000972685_1_Bpumilus_@WP_019743685_1
1452_000830075_1_Batrophaeus_@WP_004430105_1
135735_000972245_2_Bendophyticus_@WP_046218259_1
985665_000235585_1_Pterrae_@WP_014282179_1
1406_000819665_1_Ppolymyxa_@WP_040103456_1
59893_001272655_1_Ppeoriae_@WP_028540604_1
573061_000145275_1_Ccellulovoran_@WP_010076392_1
1126833_000961095_1_Pbeijingensis_@PSDG1_1146586_2e_64
481743_000024685_1_Psp_@WP_015733829_1
1616788_001421015_1_Pbovis_@WP_060533479_1
262543_000019905_1_Esibiricum_@WP_012369840_1
1087448_000299435_1_Eantarcticum_@WP_014969915_1
1399115_000496635_1_Esp_@WP_023467614_1
370438_000010565_1_Pthermopropio_@WP_012032627_1
760568_000214705_1_Dkuznetsovii_@WP_013823888_1
635013_000092945_1_Tpotens_@WP_013120035_1
1374_001465835_1_Pkocurii_@WP_058385265_1
1526927_000785555_1_Psp_@WP_038703980_1
200991_001465795_1_Prifietoensis_@WP_058380669_1
1283_000972725_1_Shaemolyticus_@WP_011275322_1
458233_000010585_1_Mcaseolyticus_@WP_050742435_1
396513_000009405_1_Scarnosus_@WP_015901255_1
407035_001005905_1_Shalodurans_@WP_046790340_1
1284_000816085_1_Shyicus_@WP_039645854_1
985762_001442815_1_Sagnetis_@WP_060551005_1
1295_001188875_1_Sschleiferi_@WP_050331026_1
937773_000185885_1_Spseudinterme_@WP_014613965_1
760568_000214705_1_Dkuznetsovii_@WP_013822290_1
1006007_000225265_1_Bmegaterium_@WP_014459461_1
1479_001050115_1_Bsmithii_@WP_048623326_1
86662_000775975_1_Bweihenstepha_@WP_002035388_1
235909_000009785_1_Gkaustophilus_@WP_011229725_1
1111068_000236605_1_Gthermoleovor_@WP_014194671_1
272567_001274575_1_Gstearothermo_@WP_053414268_1
420246_000015745_1_Gthermodenitr_@WP_011886662_1
581103_000166075_1_Gsp_@WP_003247704_1
1426_001295365_1_Gthermoglucos_@WP_003247704_1
581103_000166075_1_Gsp_@WP_013400769_1
1426_001295365_1_Gthermoglucos_@WP_042385261_1
349161_000016165_1_Dreducens_@WP_011876804_1
696281_000215085_1_Druminis_@WP_013841812_1
866895_000284515_1_Hhalophilus_@WP_014642176_1
1508404_000818095_1_Jmalaysiensis_@WP_039807138_1
1367477_000473245_1_Binfantis_@WP_035402173_1
1426_001295365_1_Gthermoglucos_@WP_003251443_1
581103_000166075_1_Gsp_@WP_003251443_1
420246_000015745_1_Gthermodenitr_@WP_008879500_1
235909_000009785_1_Gkaustophilus_@WP_041467842_1
1111068_000236605_1_Gthermoleovor_@WP_014195522_1
272567_001274575_1_Gstearothermo_@WP_053413664_1
0.4

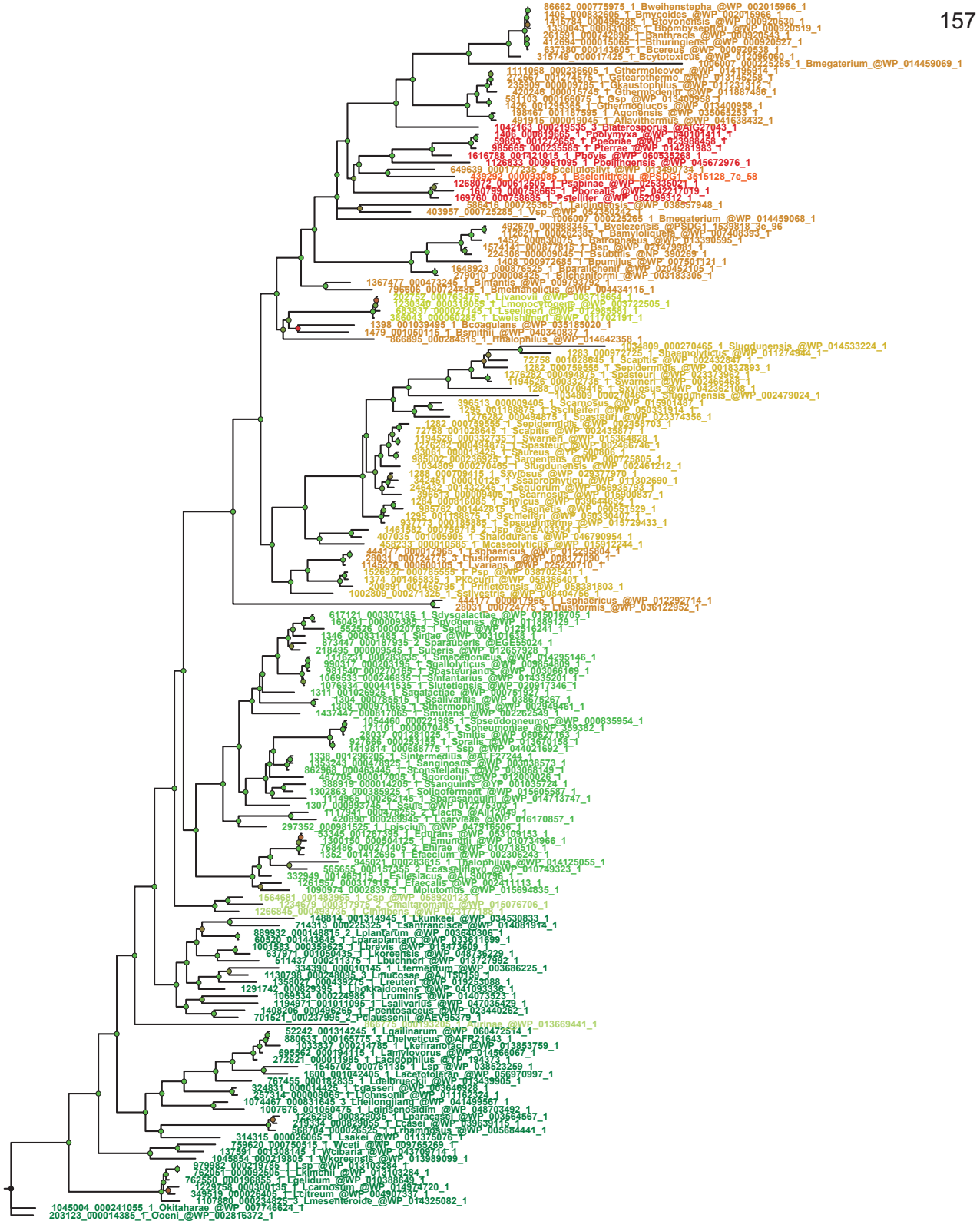

0.3

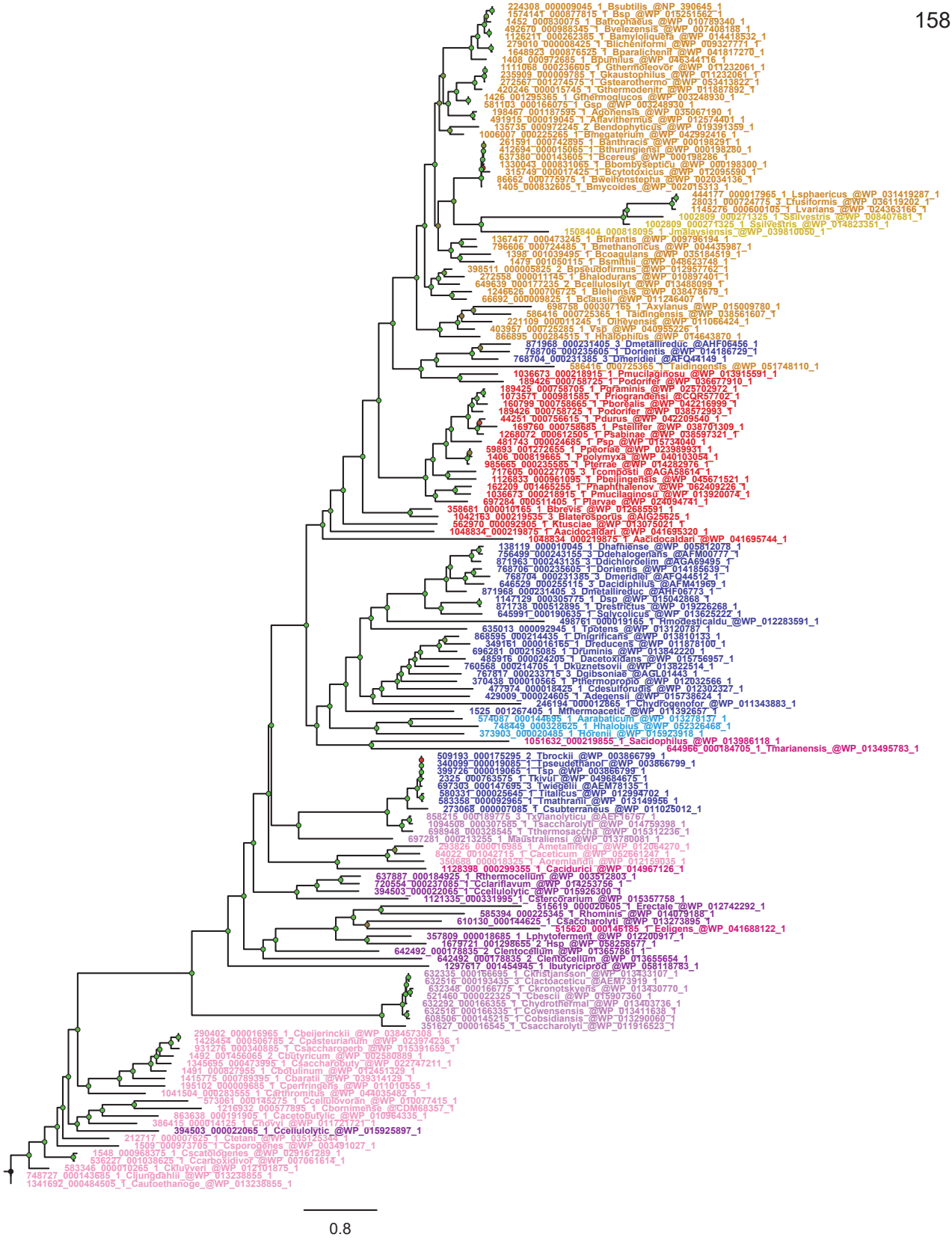

0.4

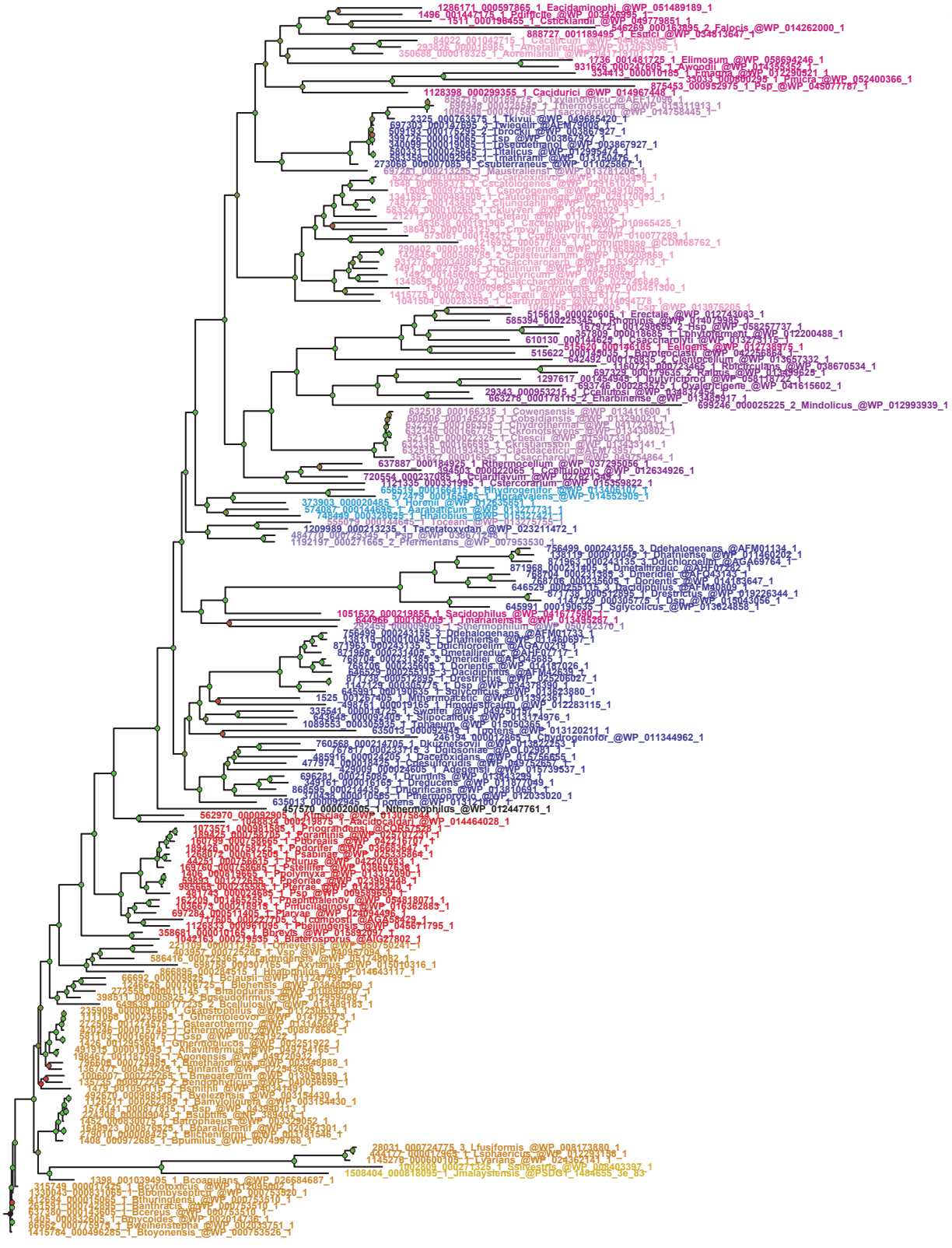

748727_000143685_1_Cljungdahlii_@WP_013237896_1
1341692_000484505_1_Caufoethanogе_@WP_013237896_1
583345_000010265_1_Ckluyveri_@WP_012101729_1
863638_000191905_1_Cacetobutylic_@WP_010965001_1
386415_000014125_1_Cnovyi_@WP_011722332_1
1509_000973705_1_Csporogenes_@WP_003439425_1
536227_001038625_1_Ccarboxidivor_@WP_007061498_1
1548_000968375_1_Cscatologenes_@WP_029161428_1
212717_000007625_1_Ctetani_@WP_011099363_1
1216932_000577895_1_Cbornimense_@CDM69106_1
573061_000145275_1_Ccellulovoran_@WP_010077195_1
290402_000016965_1_Cbeijerinckii_@WP_023973358_1
1428458_000506785_2_Cpasteurianum_@WP_023973358_1
931276_000340885_1_Csaccharoperb_@WP_015391355_1
1345695_000473995_1_Csaccharobuty_@WP_022744626_1
1492_001456065_2_Cbutyricum_@WP_035762504_1
1491_000827955_1_Cbotulinum_@WP_003373756_1
1415775_000789395_1_Cbaratii_@WP_052139570_1
195102_000009685_1_Cperfringens_@WP_011010594_1
1041504_000283555_1_Carthromitus_@WP_014094708_1
632292_000166355_1_Chydrothermal_@WP_013403764_1
632335_000166695_1_Ckristjanssoн_@WP_013433132_1
521460_000022225_1_Cbescii_@WP_013430795_1
632348_000166775_1_Ckronotskyens_@WP_013430795_1
608506_000145215_1_Cobsidiansis_@WP_013290031_1
632518_000166335_1_Cowensensis_@WP_013411610_1
351627_000016545_1_Csaccharolyti_@WP_011916491_1
1508404_000818095_1_Jmalaysiensis_@WP_039808741_1
1128398_000299355_1_Cacidurici_@WP_014967456_1
350688_000018225_1_Adremiandi_@WP_012159242_1
84022_001042715_1_Caceticum_@WP_044825072_1
293826_000018885_1_Jmetalliredis_@WP_015863690_1
509193_000175295_2_Tbrockii_@WP_012269189_1
697303_000147695_3_Twiegelii_@AEM79000_1
340099_000019085_1_Tpseudoethanol_@WP_012269189_1
339726_000019065_1_Tsp_@WP_009082541_1
583358_000092965_1_Thaimranii_@WP_013150469_1
580331_000025645_1_Titalicus_@WP_012995466_1
273068_000007085_1_Csubterraneus_@WP_031313939_1
2325_000763575_1_Tkivui_@WP_049885412_1
858215_000189775_3_Txylanolyticu_@AEF17104_1
1094508_000307585_1_Tsaccharolyti_@WP_014755452_1
698948_000328545_1_Thermosaccha_@WP_015311906_1
697281_000213255_1_Maustraliensi_@WP_013781198_1
555075_000144645_1_Toceani_@WP_013275784_1
1209989_000213235_1_Tacetatoxydan_@WP_013778217_1
720554_000237085_1_Cclariflavum_@WP_014255357_1
394503_000022065_1_Ccellulolytic_@WP_015925511_1
637887_000184925_1_Rthermocellum_@WP_003517906_1
1121335_000331995_1_Cstercorarium_@WP_015359226_1
1496_001447175_1_Pdifficile_@PSDG1_1918690_4e_08
1192197_000276665_2_Pfermentans_@WP_007953515_1
484770_000725345_1_Psp_@WP_038671229_1
28031_000724775_3_Lfusiformis_@WP_036123791_1
444177_000017965_1_Lsphaericus_@WP_031418923_1
1145276_000600705_1_Lvarians_@WP_025218757_1
1002809_000271325_1_Ssilvestris_@WP_008403390_1
515619_000020605_1_Erectale_@WP_041254120_1
610130_000144625_1_Csaccharolyti_@WP_013273010_1
515620_000146185_1_Eeligens_@WP_041688011_1
357809_000018685_1_Lphytofermen_@WP_012200484_1
1679721_001298695_2_Hsp_@WP_058257741_1
642492_000178835_2_Clentocellum_@WP_013657328_1
1297617_001454945_1_Tbutyricprod_@WP_058117648_1
693746_000283575_1_Ovalericigene_@WP_049801423_1
1160721_000723465_1_Rbicirculans_@WP_038671594_1
29343_000953215_1_Ccellulosi_@WP_045085722_1
663278_000178115_2_Ehabrinense_@WP_049776586_1
760568_000214705_1_Dkuznetsovii_@WP_013822265_1
429009_000024605_1_Adegensii_@WP_015739526_1
767817_000233715_3_Dgibsoniae_@AGL02961_1
477974_000018425_1_Cdesulfuricis_@WP_012302519_1
370438_000010565_1_Pthermopropio_@WP_012033005_1
485916_000024205_1_Dacetoxidans_@WP_015756666_1
349181_000818185_1_Dredfcens_@WP_011877060_1
696281_000215085_1_Druminis_@WP_013843286_1
868595_000214435_1_Dhafnifricans_@WP_013810684_1
635013_000092945_1_Tpotens_@WP_013120996_1
498761_000019165_1_Hmpdesticaldu_@WP_049754100_1
1525_001267405_1_Mthermoacetic_@WP_011392370_1
1089553_000305935_1_Tphaeum_@WP_015050373_1
335541_000074725_1_Swolfei_@WP_011640259_1
643648_000092405_1_Slipocalidus_@WP_013174988_1
292459_000009905_1_Sthermophilum_@WP_011195351_1
644966_000184705_1_Tmaritahensis_@WP_013495282_1
562970_000092905_1_Ktuscilae_@WP_041304023_1
1048834_000219875_1_Aacidocaldari_@WP_014464043_1
246194_000012865_1_Chydrogenofor_@WP_011344949_1
457570_000020005_1_Nthermophilus_@WP_012447767_1
373903_000020485_1_Rorenii_@WP_012635857_1
748449_000328625_1_Hhalobius_@WP_015327413_1
574087_000144695_1_Aarabaticum_@WP_013277740_1
138119_000010045_1_Dhafniense_@WP_011460687_1
756499_000243155_3_Ddehalogenans_@AFM01715_1
871963_000243135_3_Ddichloroelim_@AGA70211_1
871968_000231405_3_Dmetallireduc_@AHE07710_1
768704_000231385_3_Dmeridiei_@AFO45676_1
768706_000235605_1_Dorientis_@WP_014187016_1
646529_000255115_3_Dacidiphilus_@AFM42630_1
1147129_000305775_1_Dsp_@WP_015043530_1
871738_000512895_1_Drestrictus_@WP_019226190_1
645991_000190635_1_Solycicus_@WP_013623887_1
1051632_000219855_1_Sacidophilus_@WP_013988333_1
1574141_000877815_1_Bsp_@WP_038828183_1
224308_000009045_1_Bsubtilis_@WP_009414_1
1452_000830075_1_Batrophaeus_@WP_003329042_1
279010_000008425_1_Blicheniformi_@WP_009328548_1
1648923_000876525_1_Bparalicheni_@WP_020451312_1
1126211_000262385_1_Bamyloliquefa_@WP_014417723_1
492670_000988345_1_Bvelezensis_@WP_038457934_1
1408_000972685_1_Bpumilus_@WP_046343490_1
1367477_000473245_1_Binfantis_@WP_009791192_1
796606_000724485_1_Bmethanolicus_@WP_003348901_1
1398_001039495_1_Bcoagulans_@WP_035181912_1
1479_001050115_1_Bsmithii_@WP_048823108_1
135735_000972245_2_Bendophyticus_@WP_040056695_1
1006007_000225265_1_Bmegaterium_@WP_014458377_1
235909_000009735_1_Gkaustophilus_@WP_011230626_1
1111065_000236605_1_Gthermoleovor_@WP_014195977_1
272567_001274575_1_Gstearothermo_@WP_013523341_1
420246_000015745_1_Gthermodenitr_@WP_041264316_1
585394_000225345_1_Rhominis_@WP_014079982_1
491915_000019045_1_Aflavithermus_@WP_041638585_1
198467_001187595_1_Agonensis_@WP_035065967_1
581103_000166075_1_Gsp_@WP_003251906_1
1426_001295365_1_Gthermoglucos_@WP_003251906_1
403957_000725285_1_Vsp_@WP_040955741_1
586418_000225365_1_Tadhingensis_@WP_038559972_1
221109_000011245_1_Oiheyensis_@WP_011065875_1
866895_000284515_1_Rhalophilus_@WP_014643121_1
698758_000307165_1_Axylanus_@WP_015010308_1
649639_000177235_2_Bcellulosilyt_@WP_013489175_1
398511_000005825_2_Bpseudofirmus_@WP_012959478_1
272558_000011145_1_Bhalodurans_@WP_010898708_1
1246626_000706725_1_Blehensis_@WP_038480948_1
66692_000009825_1_Bclausii_@WP_011247195_1
412694_000015065_1_Bthuringiens_@WP_000261988_1
261591_000742895_1_Banthracis_@WP_000261966_1
637380_000143605_1_Bcereus_@WP_000261985_1
1405_000832605_1_Bmycoides_@WP_002128859_1
66662_000775978_1_Bweihenstepha_@WP_033709394_1
1415784_000496285_1_Btoyonensis_@WP_000261990_1
1330043_000831065_1_Bbombysepticu_@WP_000261970_1
315749_000017425_1_Bcytotoxicus_@WP_012094995_1
189425_000758705_1_Pparamicis_@WP_025707240_1
1073571_000981585_1_Priograndensi_@CQR57520_1
160799_000758665_1_Pborealis_@WP_042219961_1
189426_000758725_1_Pdoronifer_@WP_036683940_1
44251_000756615_1_Pdurus_@WP_042207687_1
1268072_000612505_1_Psabinae_@WP_025335858_1
169760_000758685_1_Pstelliferi_@WP_038697624_1
481743_000024685_1_Psp_@WP_049788794_1
1406_000819665_1_Ppolymyxa_@WP_040103331_1
59893_001272655_1_Ppeoriae_@WP_023989442_1
985665_000235585_1_Pterrae_@WP_014282432_1
1126833_000961095_1_Pbaiingensis_@WP_045673485_1
717605_000227705_3_Tcommunti_@AGA58422_1
697284_000511405_1_Ptarvae_@WP_034285884_1
1036673_000218915_1_Pmucilaginosu_@WP_013917747_1
162209_001465255_1_Pnaphthalenov_@WP_054818074_1
1042163_000219535_3_Blaterosporus_@AIG27792_1
358681_000010165_1_Bbrevis_@WP_015892087_1
0.4

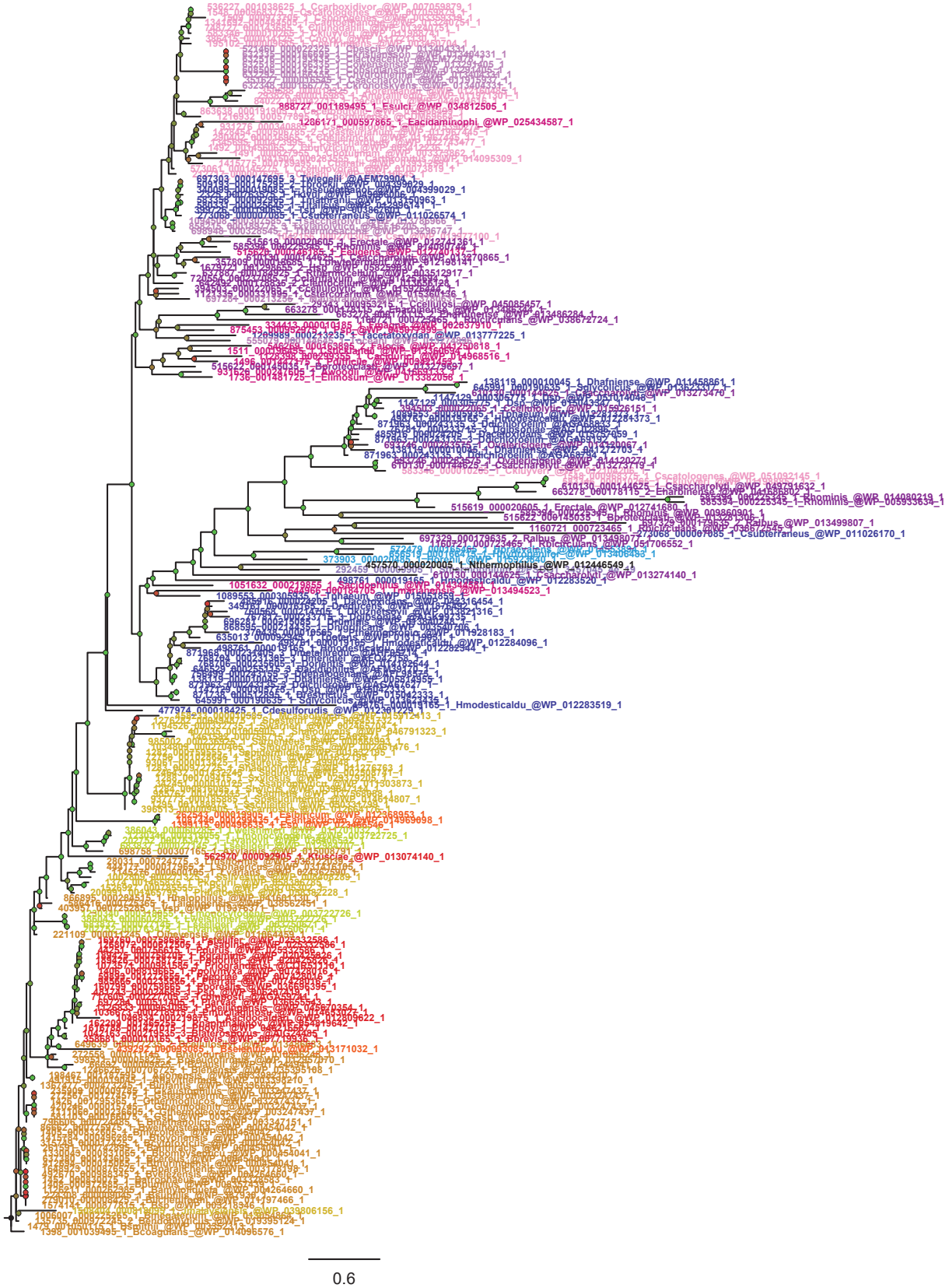

632348_000166775_1_Ckronotskyens_@WP_041741992_1
521460_000022325_1_Cbescii_@WP_015907719_1
632335_000166695_1_Ckristjanssonii_@WP_013431983_1
632516_000193435_3_Clactoaceticu_@AEM74337_1
632292_000166355_1_Chydrothermal_@WP_013403869_1
632518_000166335_1_Cowensensis_@WP_013411515_1
698508_000145215_1_Cobsidiansis_@WP_013289925_1
351627_000016545_1_Csaccharolyti_@WP_011916751_1
697281_000213255_1_Maustralianus_@WP_013779983_1
340099_000019085_1_Tpseudethanol_@WP_003868976_1
509183_000175235_2_Tbrockii_@WP_003868976_1
697303_000147695_3_Twiegelii_@AEM77694_1
2325_000763575_1_Tkivui_@WP_049884369_1
339726_000019065_1_Tsp_@WP_009051973_1
580331_000025645_1_Titalicus_@WP_012994315_1
583358_000092965_1_Tmathranii_@WP_012994315_1
273068_000007085_1_Csubterraneus_@WP_011024671_1
665415_000189775_3_Txylanolyticu_@AEF75161_1
1094508_000307585_1_Tsaccharolytr_@WP_014757250_1
698948_000328545_1_Thermosaccha_@WP_013289924_1
357809_000018685_1_Lbhytoferment_@WP_012198240_1
1679721_001298655_2_Hsp_@WP_058257265_1
642492_000178835_2_Clentocellum_@WP_013658567_1
610130_000144625_1_Csaccharolyti_@WP_013271106_1
515619_000020605_1_Erectale_@WP_022292793_1
585394_000225345_1_Rhominis_@WP_044024693_1
515620_000146185_1_Eeligens_@WP_012738447_1
693746_000283575_1_Ovalericigene_@WP_014116666_1
720554_000237085_1_Cclariflavum_@WP_014253685_1
637887_000184925_1_Rthermocellum_@WP_003512937_1
1160721_000723485_1_Rhiciruclans_@WP_038671255_1
697329_000179635_2_Ralbus_@WP_013498502_1
29343_000953215_1_Ccellulosi_@WP_034643941_1
663278_000178115_2_Ehartbinense_@WP_013484602_1
642492_000178835_2_Clentocellum_@WP_013657761_1
515620_000146185_1_Eeligens_@WP_012744487_1
1297617_001454945_1_Ibutyriciprod_@WP_058118805_1
394503_000922065_1_Ccellulolytic_@XP_3710869_2e-15
1213335_000331985_1_Cstercorarium_@WP_013357813_1
1128398_000299355_1_Cacidurtcl_@WP_014866428_1
1496_001447175_1_Pdifficile_@WP_003420727_1
1491_000627955_1_Cbotulinum_@WP_003373770_1
583346_000010265_1_Ckluyveri_@WP_012102210_1
386415_000014125_1_Cnovyi_@WP_003264198_1
663638_000191965_1_Cacetobutylic_@WP_010966143_1
573061_000145225_1_Ccellulovoran_@WP_010075828_1
1216932_000577895_1_Cbornimensa_@CDM69412_1
195102_000003685_1_Cpertfringens_@WP_003452341_1
1041504_000283555_1_Carthromitus_@WP_014095123_1
1509_000973705_1_Csporogenes_@WP_003366559_1
1341692_000448805_1_Cautoethanogs_@WP_013236939_1
748727_000143685_1_Cljungdahlii_@WP_013236939_1
583346_000010265_1_Ckluyveri_@WP_012104000_1
536227_001038625_1_Ccarboxidivor_@WP_007063637_1
1548_000968375_1_Cscatologenes_@WP_007063637_1
212477_000007635_1_Cfetani_@WP_023437156_1
350688_000018325_1_Aoremfandii_@WP_012160390_1
293826_000016985_1_Ametalliredg_@WP_011971502_1
84022_001042715_1_Cacetobuli_@WP_044826077_1
235909_000009765_1_Gkaustophilus_@WP_011232801_1
1111068_000236605_1_Gthermoleovor_@WP_008880665_1
1426_001295365_1_Gthermoglucos_@WP_003253595_1
420246_000019745_1_Gthermodenitr_@WP_008880665_1
272567_001274575_1_Gstearothermo_@WP_011232801_1
581103_000166075_1_Gsp_@WP_003253595_1
1479_001050115_1_Bsmithii_@WP_003353321_1
491915_000019045_1_Afulgidus_@WP_003396966_1
198467_001187595_1_Agonensis_@WP_003396966_1
444177_000017965_1_Lsphaericus_@WP_025114760_1
28031_000724775_3_Lfusiformis_@WP_025114760_1
1145278_000500105_1_Lvarians_@WP_004269483_1
1002809_000271725_1_Spilvestris_@WP_008404436_1
866895_000284515_1_Hhalophilus_@WP_014644812_1
403957_000725285_1_Vsp_@WP_040956526_1
221109_000011245_1_Olheyangis_@WP_011067356_1
1398_001039495_1_Bcoagulans_@WP_014096774_1
1415784_000496295_1_Blovonensis_@WP_000544012_1
1330043_000831065_1_Bbombysepticu_@WP_000544012_1
69662_000775975_1_Bweihenstepha_@WP_033709588_1
651280_000133605_1_Bcereus_@WP_000544012_1
1405_000832605_1_Bmycoides_@WP_002089714_1
315739_000174275_1_Bcytotoxicus_@WP_041810135_1
412694_000015065_1_Bthuringiensi_@WP_000544012_1
261591_000742995_1_Banthracis_@WP_000544012_1
574141_000877815_1_Bsp_@WP_033221804_1
279019_000008425_1_Blichenformi_@WP_003185911_1
1649923_000076525_1_Bparalichenif_@WP_003185911_1
574308_000009045_1_Bsubtilis_@NP_391523_1
1452_000830075_1_Batrophaeus_@WP_003271804_1
1126211_000262385_1_Bamyloliquefa_@WP_003151236_1
1409_000972685_1_Bpumilus_@WP_003243727_1
492670_000988345_1_Bvelezensis_@WP_003151236_1
586416_000725365_1_Taidingensis_@WP_038564718_1
1367477_000479245_1_Bintfantis_@WP_009792693_1
135735_000972245_2_Bendophyticus_@WP_019394815_1
1006007_000225265_1_Bmegaterium_@WP_013059792_1
796606_000724485_1_Bmethanolicus_@WP_003346764_1
649639_000177235_2_Bcellulosilyt_@WP_013490602_1
481743_000024685_1_Psp_@WP_009594246_1
160799_000758685_1_Pborealis_@WP_019915029_1
1073571_000981585_1_Priograndensi_@CQR58655_1
169760_000758685_1_Pstellifer_@WP_038699334_1
169426_000758725_1_Podorifer_@WP_036681545_1
169425_000758705_1_Pgraminis_@WP_019915029_1
44251_000756615_1_Pdurus_@WP_042208548_1
1268072_000813805_1_Psabinae_@WP_026238625_1
162209_001485255_1_Pnaphthalenov_@WP_036682840_1
1036673_000218915_1_Pmucilaginosu_@WP_013913891_1
1126833_000961095_1_Pbeijingensis_@WP_045669983_1
717605_000227705_3_Tcomposti_@AGA58850_1
1406_000819665_1_Ppolymyxa_@WP_013373425_1
58933_001274555_1_Ppeoriae_@WP_007432737_1
58965_000235585_1_Pterrae_@WP_007432737_1
697284_000511405_1_Plarvae_@WP_023484018_1
1048834_000219875_1_Aacidocaldari_@WP_012811963_1
562970_000092905_1_Ktusciae_@WP_013077098_1
358681_000010165_1_Bprevis_@WP_003388385_1
1042163_000219535_3_Blaterosporus_@AIG28674_1
66692_000098625_1_Bclausii_@WP_011246576_1
1246626_000706725_1_Blehensis_@WP_035394300_1
398511_000005825_2_Bpseudofirmus_@WP_012960686_1
272564_000011145_1_Shaloduran s_@WP_010889665_1
698758_000307165_1_Axylanus_@WP_015010759_1
1508404_000818095_1_Ymalaysiensis_@WP_039813886_1
1209989_000213235_1_Tacetatoxydan_@WP_013779115_1
555079_000144845_1_Toceani_@WP_041623938_1
1089553_000305935_1_Tphaeum_@WP_015051738_1
1529_001267405_1_Mthermoacetic_@WP_011393849_1
574087_000144695_1_Aarabaticum_@WP_013279137_1
748449_000328625_1_Rhalobilus_@WP_015328040_1
373903_000020485_1_Horenii_@WP_015923493_1
457570_000020005_1_Nthermophilus_@WP_012449215_1
644966_000184705_1_Tmaritimensis_@WP_013494604_1
292459_000005905_1_Sthermophilum_@WP_011134238_1
1051632_000219855_1_Sacidophilus_@WP_013985753_1
643648_000092405_1_Slipocalidus_@WP_013176419_1
335541_000014725_1_Svolfei_@WP_011641748_1
498761_000019165_1_Hmodesticaldu_@WP_012281991_1
971738_000512895_1_Drestrictus_@WP_015044900_1
1147129_000385775_1_Dsp_@WP_015044900_1
645991_000190635_1_Sglycolicus_@WP_013626254_1
646529_000255115_3_Dacidiphilus_@AFM43454_1
769704_000231395_3_Dmeridiei_@AFM43564_1
788706_000235505_1_Dorientis_@WP_014187689_1
871963_000245135_3_Dalichloroelhi_@AGA70808_1
138119_000010045_1_Dhafniense_@WP_014795808_1
756499_000243155_3_Ddehalogenans_@AFM02337_1
871968_000231405_3_Dmetallireduc_@AHF08263_1
245194_000012865_1_Ehydrogenofor_@WP_011345400_1
635013_000092945_1_Tpotens_@WP_013121706_1
1134197_000271565_2_Ptremernans_@WP_007938964_1
484770_000725345_1_Psp_@WP_038873997_1
477974_000018425_1_Cdesulfuricus_@WP_012303196_1
760568_000214705_1_Dkuznetsovii_@WP_013824405_1
329009_000024605_1_Adegensii_@WP_015738141_1
868595_000214435_1_Dnitrificans_@WP_013610888_1
349161_000016165_1_Dreducens_@WP_011879435_1
686281_000215085_1_Druminis_@WP_013843633_1
370438_000010565_1_Pthermopropio_@WP_012033935_1
485916_000024205_1_Dacetoxidans_@WP_015759979_1
767817_000233715_3_Dgibsoniae_@AGJ03803_1
485916_000024205_1_Dacetoxidans_@WP_015759497_1
0.5

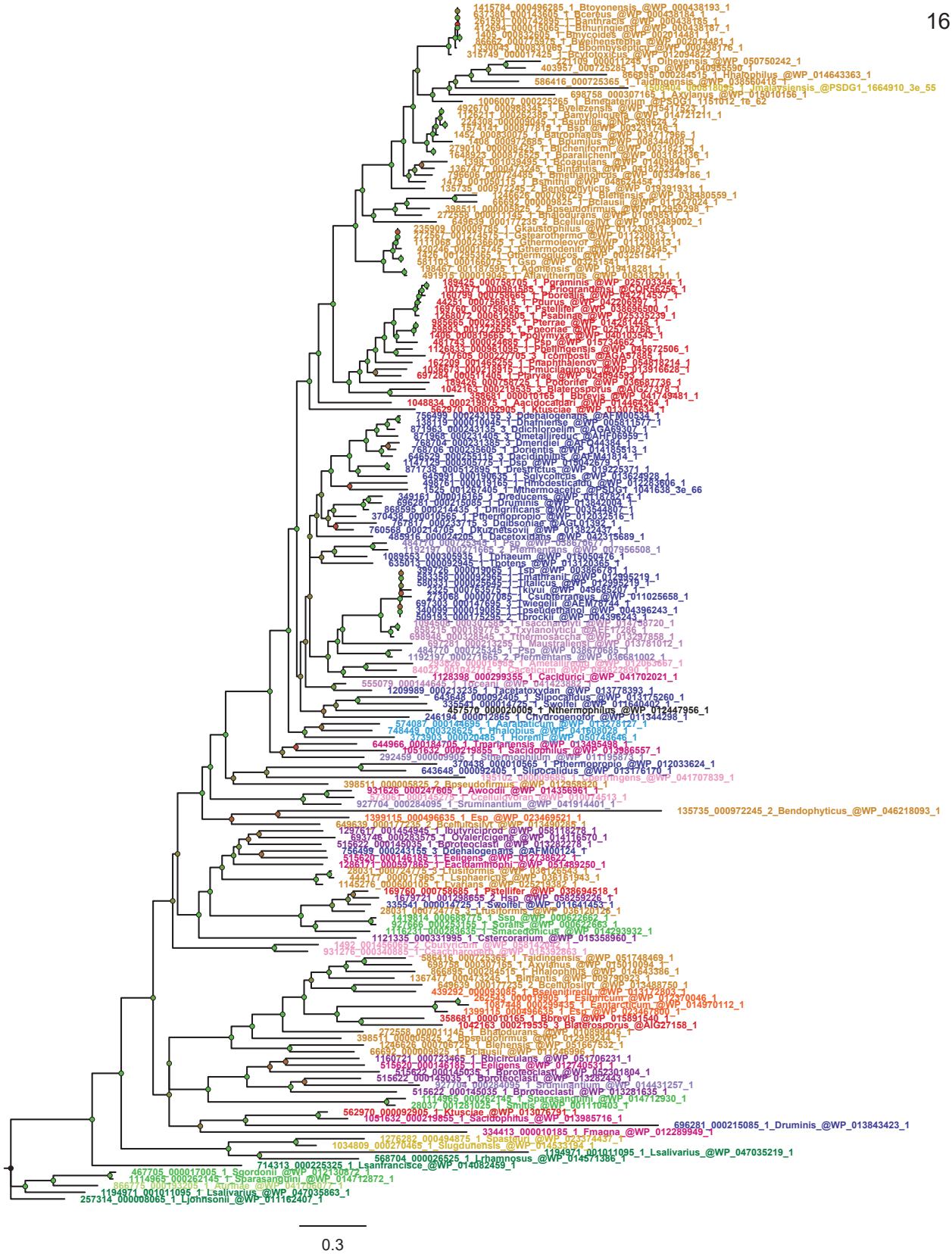

0.5

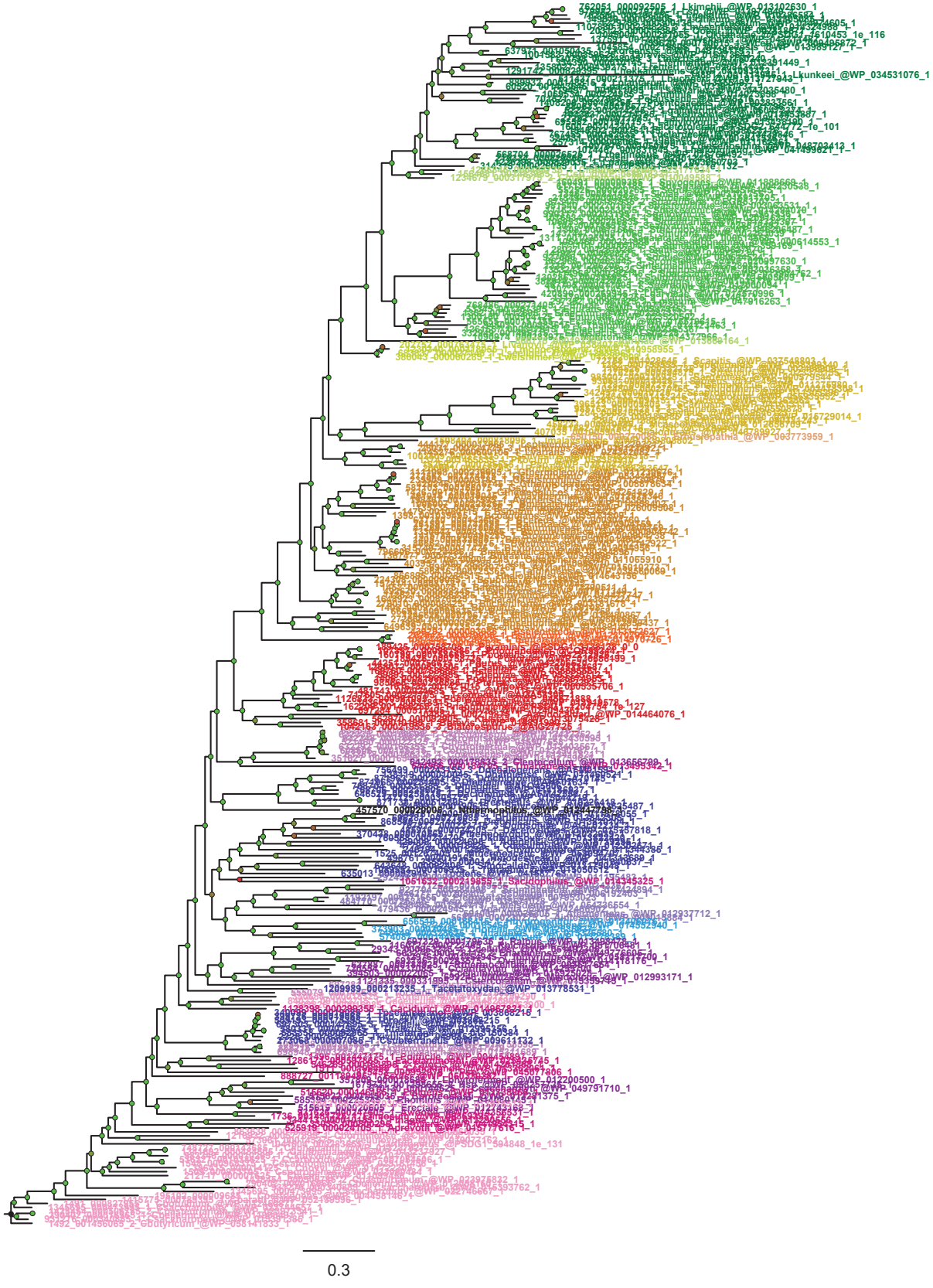

0.4

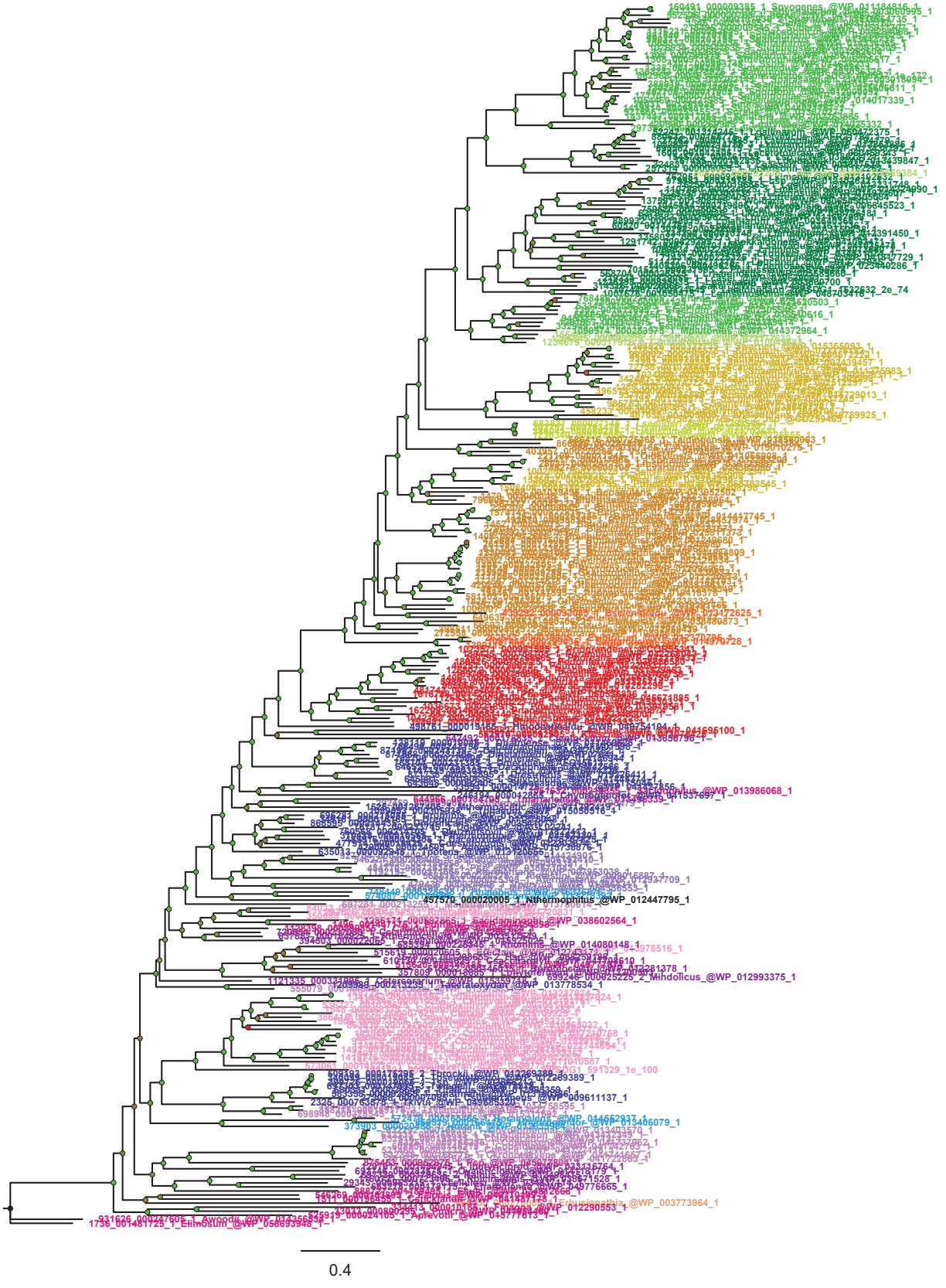

224308_000009045_1_Bsubtilis_@NP_390645_1
157414T_000877815_1_Bsp_@WP_015251562_1
1452_000830075_1_Batrophaeus_@WP_010789340_1
492670_000988345_1_Bvelezensis_@WP_007408188_1
1126211_000262385_1_Bamyloliquefa_@WP_014418532_1
279010_000008425_1_Blichenifolmi_@WP_003927771_1
1648923_000876525_1_Bparalichenif_@WP_041817270_1
1408_000972685_1_Bpumilus_@WP_046344716_1
1111068_000236605_1_Gthermoleovor_@WP_011232061_1
235909_000009785_1_Gkaustophilus_@WP_011232061_1
272567_001274575_1_Gstearothermo_@WP_053413822_1
420246_000015745_1_Gthermodenitr_@WP_011887892_1
1426_001295365_1_Gthermoglucos_@WP_003248930_1
581103_000166075_1_Gsp_@WP_003248930_1
198467_001187595_1_Agoriensis_@WP_035067190_1
491915_000019045_1_Alfavithermus_@WP_012574401_1
135735_000972245_2_Bendophyticus_@WP_019391359_1
1006007_000225265_1_Bmegaterium_@WP_042992416_1
261591_000742895_1_Banthracis_@WP_000198291_1
412694_000015665_1_Bthuringiens_@WP_000198280_1
657380_000143605_1_Bcereus_@WP_000198266_1
1330043_000831065_1_Bbombysepticu_@WP_000198300_1
315749_000017425_1_Bcytotoxicus_@WP_012095590_1
86662_000775975_1_Bweihenstepha_@WP_002034136_1
1405_000832605_1_Bmycoides_@WP_002015313_1
444177_000017965_1_Lsphaericus_@WP_031419287_1
28031_000724775_3_Efusiformis_@WP_036119202_1
145276_000600105_1_Lvarians_@WP_024363166_1
1002809_000271325_1_Ssilvestris_@WP_008407681_1
1002809_000271325_1_Ssilvestris_@WP_014823351_1
1508404_000818095_1_Jmiaiyasiensis_@WP_039810050_1
1367477_000473245_1_Binfantis_@WP_009796194_1
796606_000724485_1_Bmethanolicus_@WP_004435987_1
1396_001039495_1_Bcoagulans_@WP_035184519_1
1479_001050115_1_Bsmithii_@WP_048623748_1
398511_000005825_2_Bpseudofirmus_@WP_012957762_1
272558_000011145_1_Bhalodurans_@WP_010897401_1
649639_000177235_2_Bcellulosilyt_@WP_013468099_1
1246626_000706725_1_Blehensis_@WP_038478679_1
66692_000009825_1_Bclausii_@WP_011246407_1
698758_000307165_1_Axylanus_@WP_015009780_1
586416_000725365_1_Taidingensis_@WP_051748110_1
221109_000011245_1_Olheyensis_@WP_011066424_1
403957_000725285_1_Vsp_@WP_040955225_1
866895_000284515_1_Hhalophilus_@WP_014643870_1
871968_000231405_3_Dmetallireduc_@AHF06456_1
768706_000236605_1_Dorientis_@WP_014186729_1
768704_000231385_3_Dmeridiei_@AFQ44149_1
1036673_000218915_1_Pmucilaginosu_@WP_013915591_1
189426_000758725_1_Podorifer_@WP_036677910_1
189425_000758705_1_Pparaminis_@WP_025702972_1
1073571_000981585_1_Priograndensi_@CQR57702_1
160799_000758665_1_Pborealis_@WP_042216999_1
189426_000759725_1_Podorifer_@WP_035572993_1
44251_000756615_1_Pdurus_@WP_042209540_1
169760_000758685_1_Pstellifer_@WP_038701309_1
1268072_000612505_1_Psabinae_@WP_038597321_1
481743_000024685_1_Psp_@WP_015734040_1
59893_001272655_1_Pporiae_@WP_023989931_1
1406_000819665_1_Ppolymyxa_@WP_040103054_1
985665_000235585_1_Pterrae_@WP_014282976_1
717605_000227705_3_Tcomposti_@AGA58614_1
1126833_000961095_1_Pbeijingensis_@WP_045671521_1
162209_001465255_1_Pnaphthalenov_@WP_062409226_1
1036673_000218915_1_Pmucilaginosu_@WP_013920074_1
697284_000511405_1_Plarvae_@WP_024094741_1
358681_000010165_1_Bbrevis_@WP_012685591_1
1042163_000219535_3_Blateroporus_@AIG25525_1
562970_000092905_1_Ktusciae_@WP_013075021_1
1048834_000219875_1_Aacidocaldari_@WP_041695320_1
1048834_000219875_1_Aacidocaldari_@WP_041695744_1
138119_000010045_1_Dhafniense_@WP_005812078_1
756499_000243155_3_Ddehalogenans_@AFM00777_1
871963_000243135_3_Ddichloroelim_@AGA69495_1
768706_000235605_1_Dorientis_@WP_014185639_1
768704_000231385_3_Dmeridiei_@AFQ44512_1
646529_000255115_3_Dacidiphilus_@AFM41969_1
871968_000231405_3_Dmetallireduc_@AHF06773_1
1147129_000305775_1_Dsp_@WP_015042868_1
871738_000512895_1_Drestrictus_@WP_019226268_1
645991_000190635_1_Solycolicus_@WP_013625222_1
498761_000019165_1_Rmodesticaldu_@WP_012283591_1
635013_000092945_1_Tpotens_@WP_013120787_1
868595_000214435_1_Dnigrificans_@WP_013810133_1
349161_000016165_1_Dreducens_@WP_011878100_1
696281_000215085_1_Dfluminis_@WP_013842220_1
485916_000024205_1_Dacetoxidans_@WP_015756957_1
760568_000214705_1_Dkuznetsovii_@WP_013822514_1
767817_000233715_3_Dgibsoniae_@AGL01443_1
370438_000010565_1_Pthermopropio_@WP_012032566_1
477974_000018425_1_Cdesulfurudis_@WP_012302327_1
429009_000024605_1_Adegensii_@WP_015738624_1
246194_000012865_1_Chydrogenofor_@WP_011343883_1
1525_001267405_1_Mthermoacetic_@WP_011392657_1
574087_000144595_1_Aarabaticum_@WP_013278137_1
748449_000328625_1_Hhalobius_@WP_052326468_1
373903_000020485_1_Hcenii_@WP_015923918_1
1051632_000219855_1_Sacidophilus_@WP_013986118_1
644966_000184705_1_Tmarianensis_@WP_013495783_1
509193_000175295_2_Tbrockii_@WP_003866799_1
340099_000019085_1_Tpseudethanol_@WP_003866799_1
399726_000019065_1_Tsp_@WP_003866799_1
2325_000763575_1_Tkivui_@WP_049684675_1
697303_000147695_3_Twiegelii_@AEM76135_1
580331_000025645_1_Ttalicus_@WP_012994702_1
583358_000092965_1_Tmathranii_@WP_013149956_1
273068_000007085_1_Csubterraneus_@WP_011025012_1
858215_000169775_3_Txylanolyticu_@AEM16757_1
1094508_000307585_1_Tsaccharolyti_@WP_014759398_1
698948_000328545_1_Tthermosaccha_@WP_015312236_1
697281_000213255_1_Maustraliensi_@WP_013780081_1
Z93826_000016985_1_Ametallireduc_@WP_012064270_1
84022_001042715_1_Caceticum_@WP_052661247_1
350585_000018325_1_Avromlaffii_@WP_041159735_1
1128398_000299355_1_Caciduric_@WP_014967126_1
637887_000184925_1_Rthermocellum_@WP_003512803_1
720554_000237085_1_Cclariflavum_@WP_014253756_1
394503_000022065_1_Ccellulolytic_@WP_015926300_1
1121335_000331995_1_Cstercorarium_@WP_015357758_1
515619_000020605_1_Erectale_@WP_012742292_1
585394_000225345_1_Rhominis_@WP_014078198_1
610130_000144625_1_Csaccharolyti_@WP_013273895_1
515620_000146185_1_Eeligens_@WP_041688122_1
357809_000018685_1_Lphytoferment_@WP_012200917_1
1679721_001298555_2_Hsp_@WP_052258577_1
642492_000178835_2_Clentocellum_@WP_013657861_1
642492_000178835_2_Clentocellum_@WP_013655654_1
1297617_001454945_1_Ibutyriciprod_@WP_058118783_1
632335_000166695_1_Ckristjansson_@WP_013433107_1
632516_000193435_3_Cactoaceticu_@AEM73919_1
632345_000166775_1_Ckronotskyens_@WP_013430770_1
521460_000022325_1_Cbescii_@WP_015907360_1
632292_000166355_1_Chydrothermal_@WP_013403736_1
632518_000166335_1_Cowensensis_@WP_013411638_1
608506_000145215_1_Cobsidiansis_@WP_013290060_1
351627_000016545_1_Csaccharolyti_@WP_011916523_1
290402_000016965_1_Cbeijerinckii_@WP_038457308_1
1428454_000506785_2_Cpasteurianum_@WP_023974236_1
931276_000340885_1_Csaccharoperb_@WP_015391659_1
1492_001456065_2_Cbutyricum_@WP_002580889_1
1345695_000473995_1_Csaccharobuty_@WP_022747211_1
1491_000227955_1_Cbotulinum_@WP_012451329_1
1415775_000759395_1_Cbaratii_@WP_033914129_1
195102_000009685_1_Cperfringens_@WP_011010555_1
1041504_000283555_1_Carthromitus_@WP_044035482_1
573061_000145275_1_Ccellulovoran_@WP_010077415_1
1216932_000577895_1_Cbornimense_@CDM68357_1
863638_000191905_1_Cacetobutylic_@WP_010964335_1
386414_000014125_1_Cnovyi_@WP_011721721_1
394503_000022065_1_Cceilulolytic_@WP_015925897_1
212717_000007625_1_Ctetani_@WP_035125345_1
1509_000973705_1_Csporogenes_@WP_003491027_1
1548_000968375_1_Cscatologenes_@WP_029161289_1
536227_001038625_1_Ccarboxidivor_@WP_007061614_1
583346_000010265_1_Cklyveri_@WP_012101875_1
748727_000143655_1_Cljungdahlii_@WP_013238855_1
1341692_000484505_1_Cautoethanoge_@WP_013238855_1
0.8

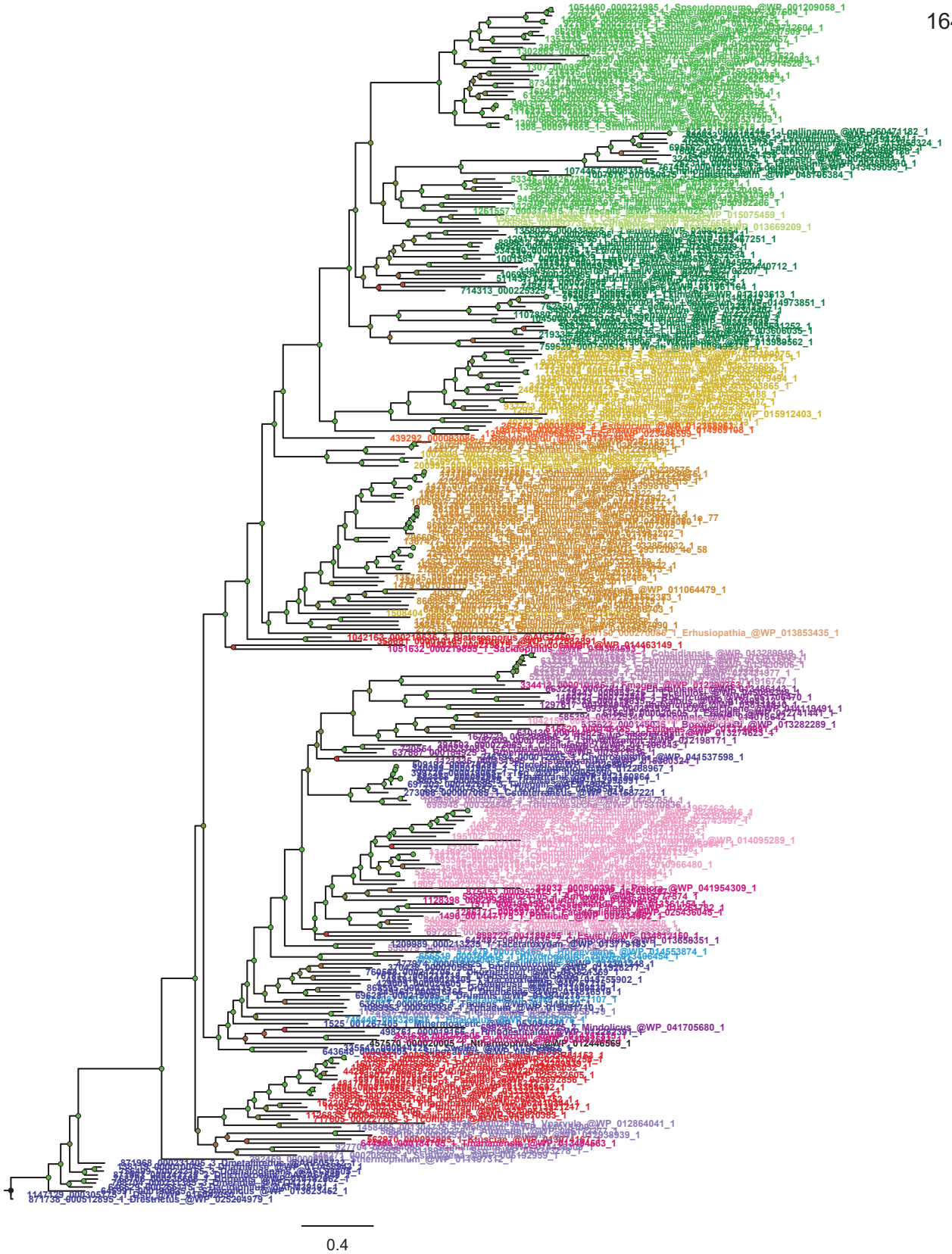

0.4

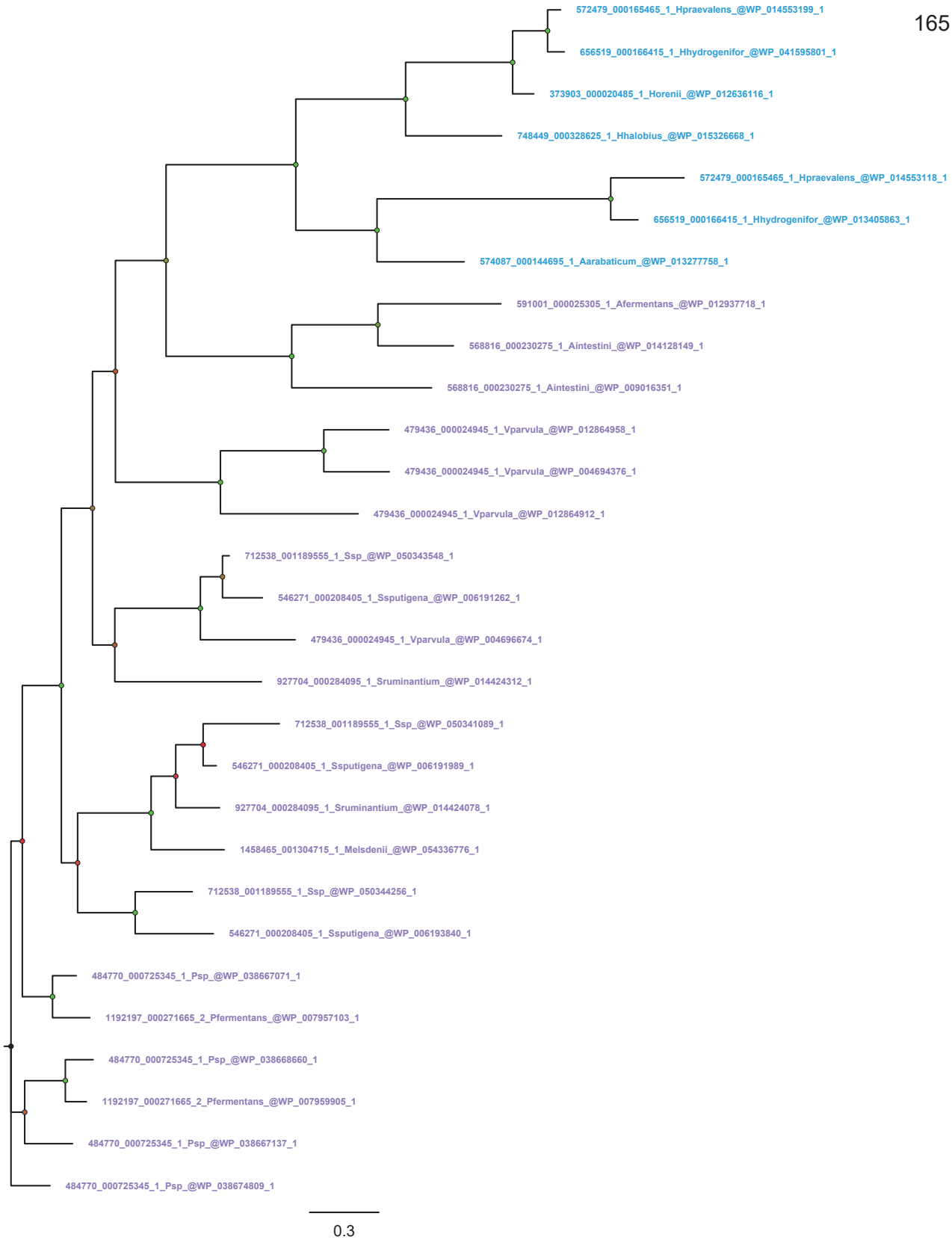

0.6

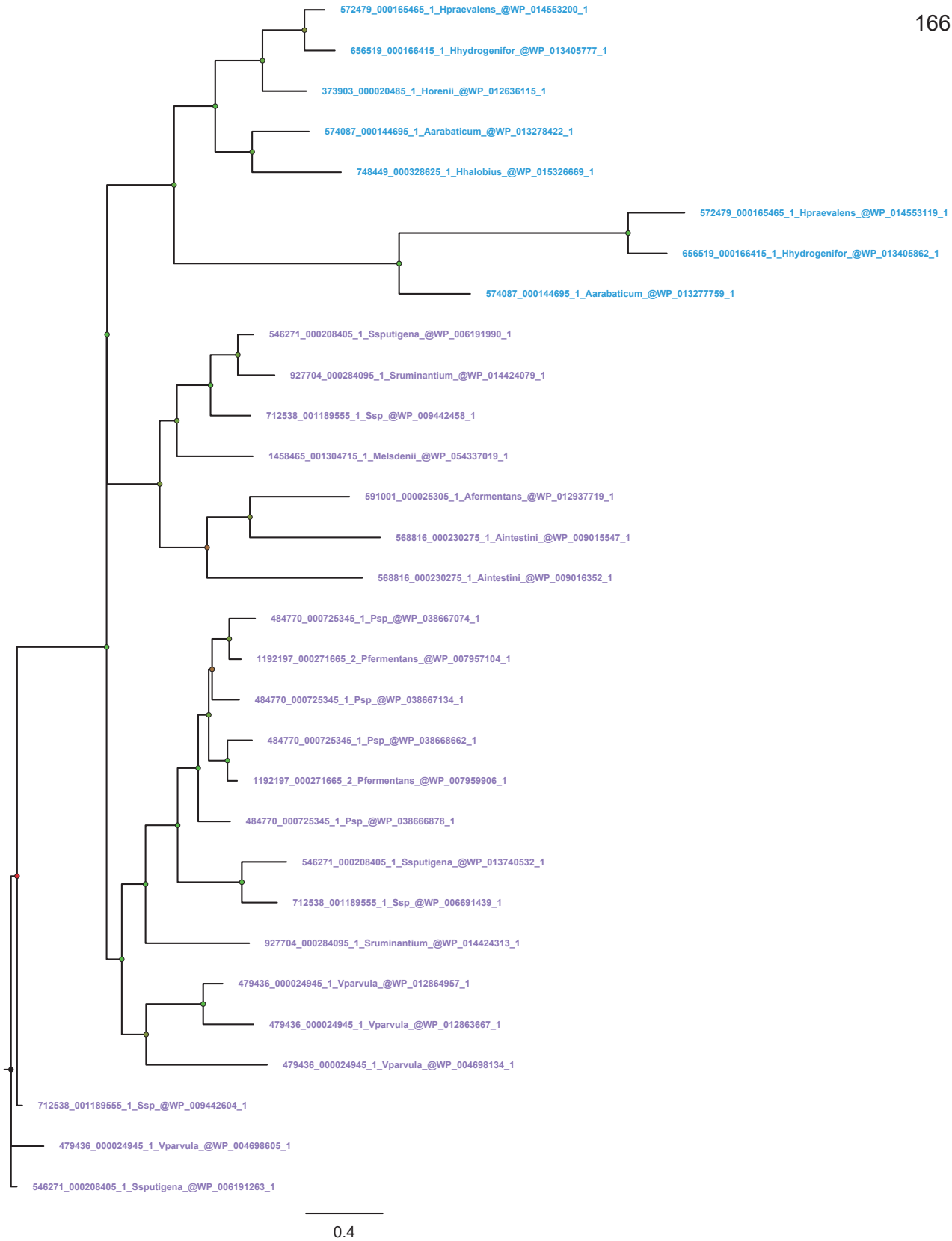

0.3

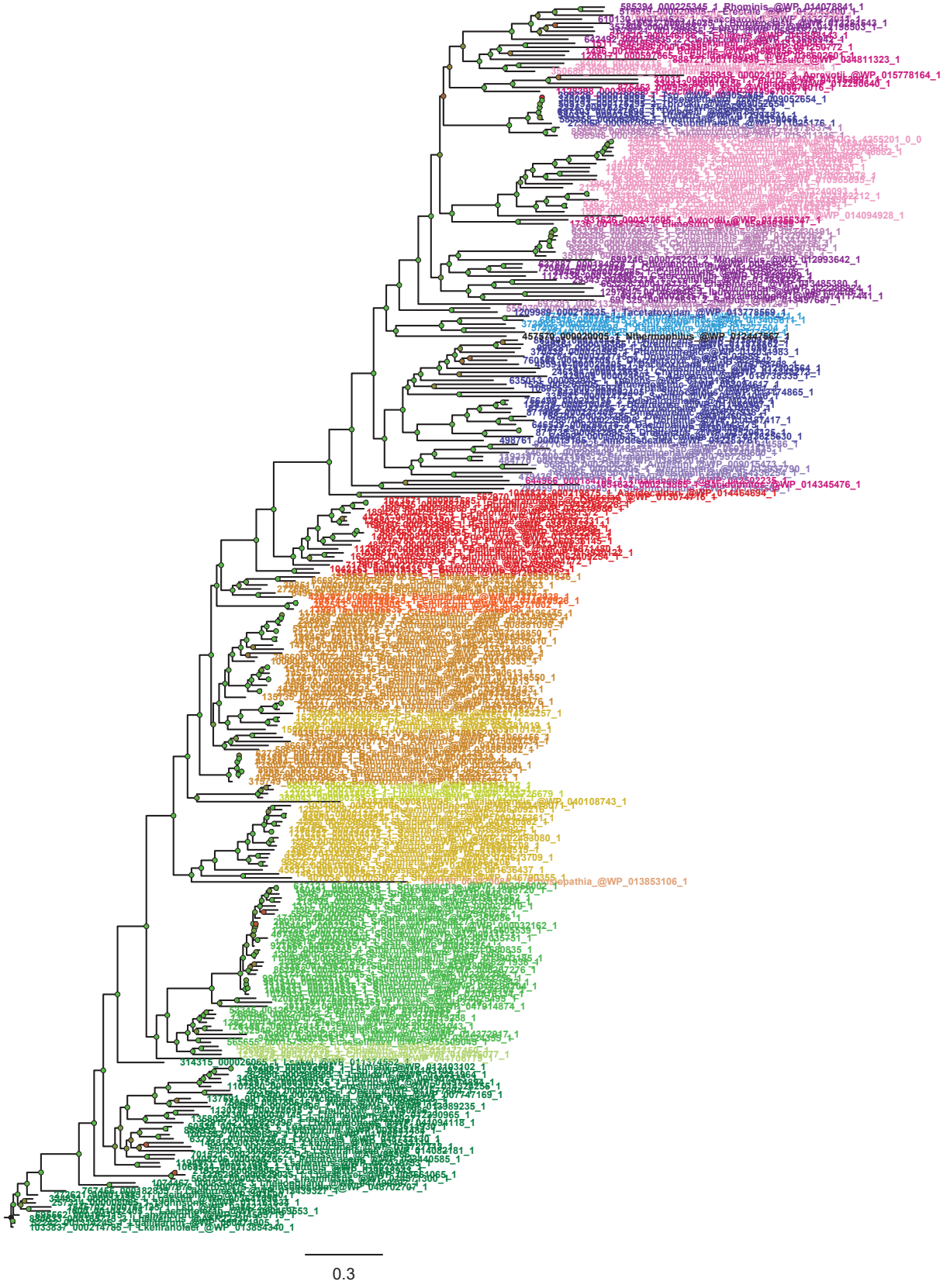

0.3

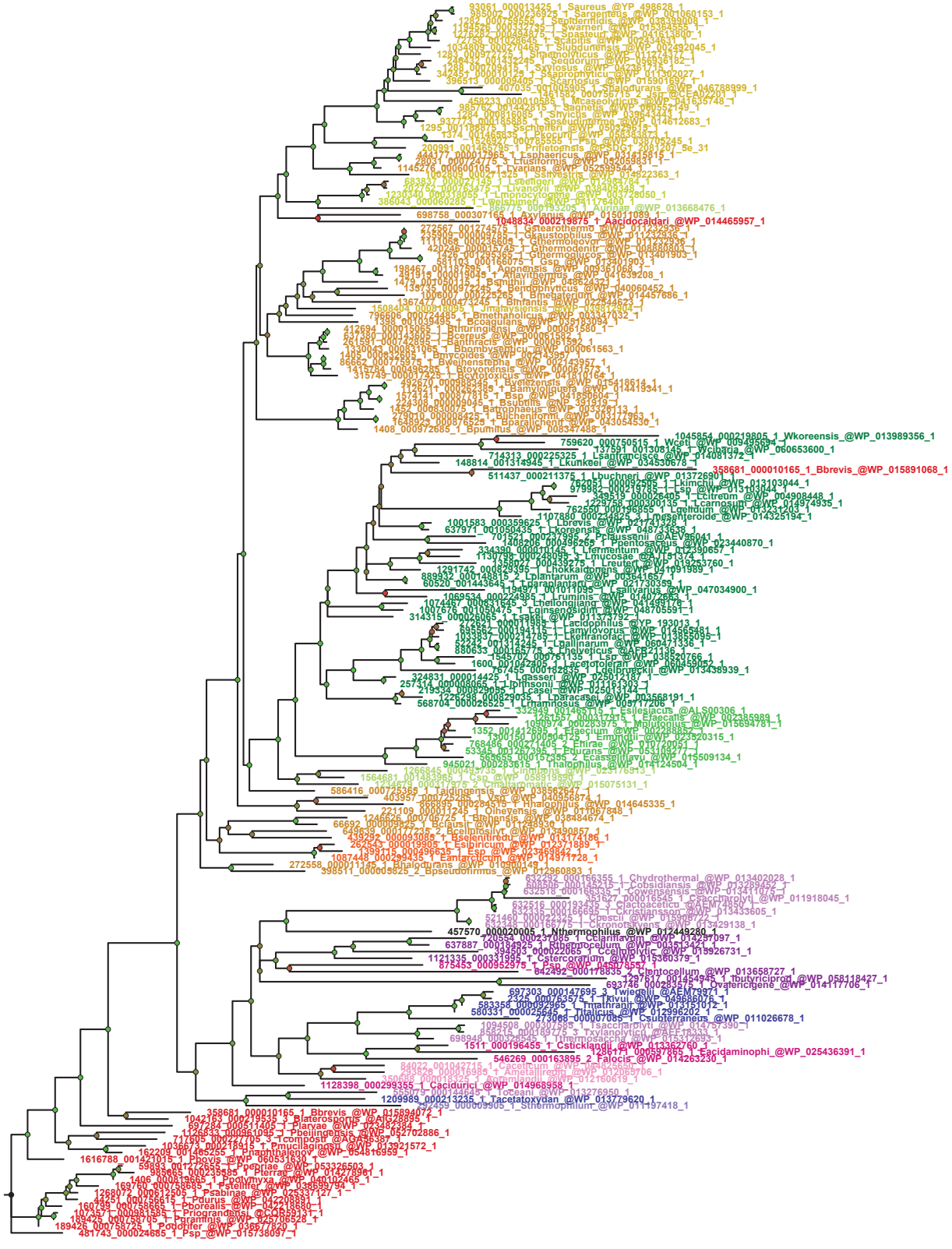

457579_000020005_1_Nthermophilus_@WP_012447795_1
0.4

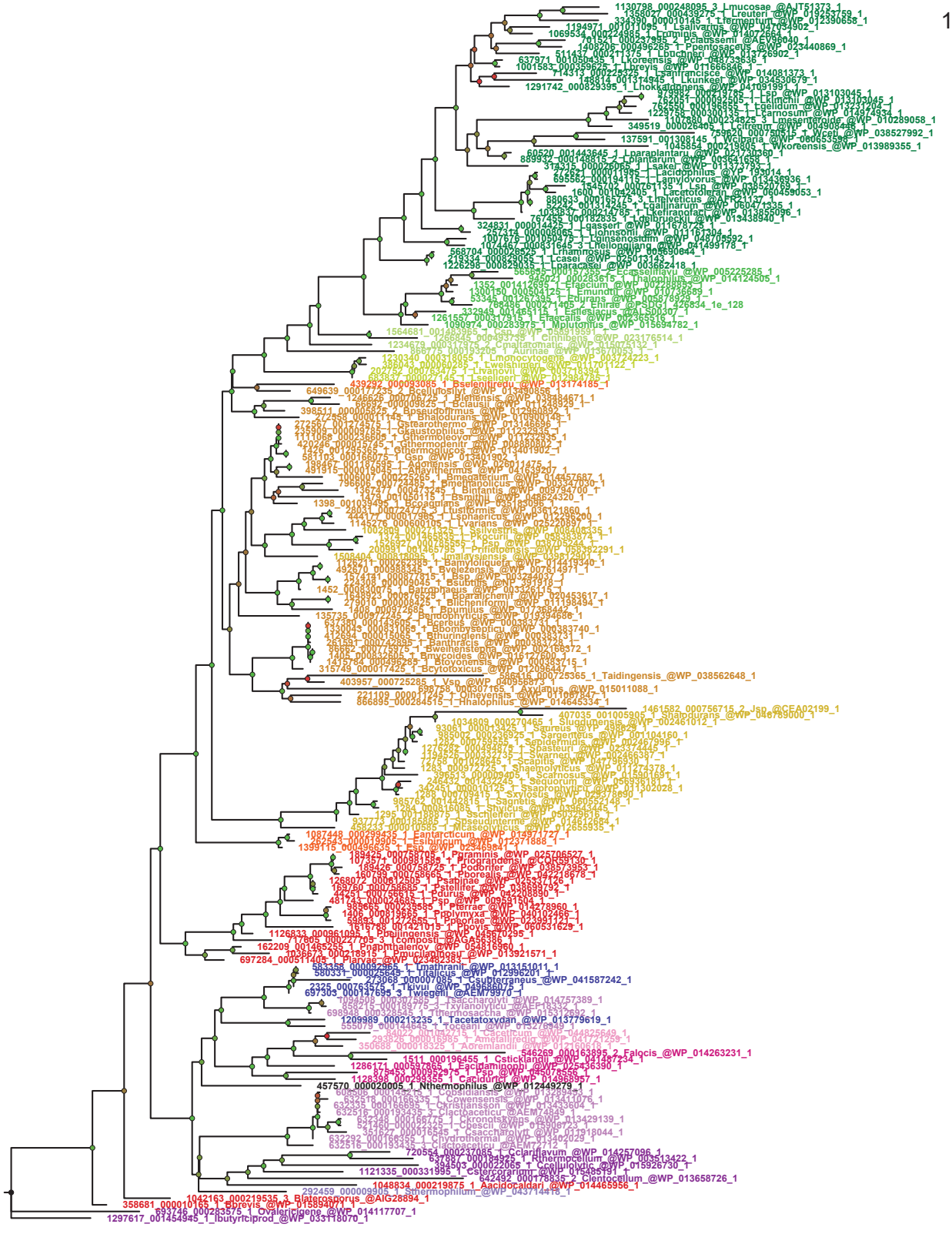

0.4

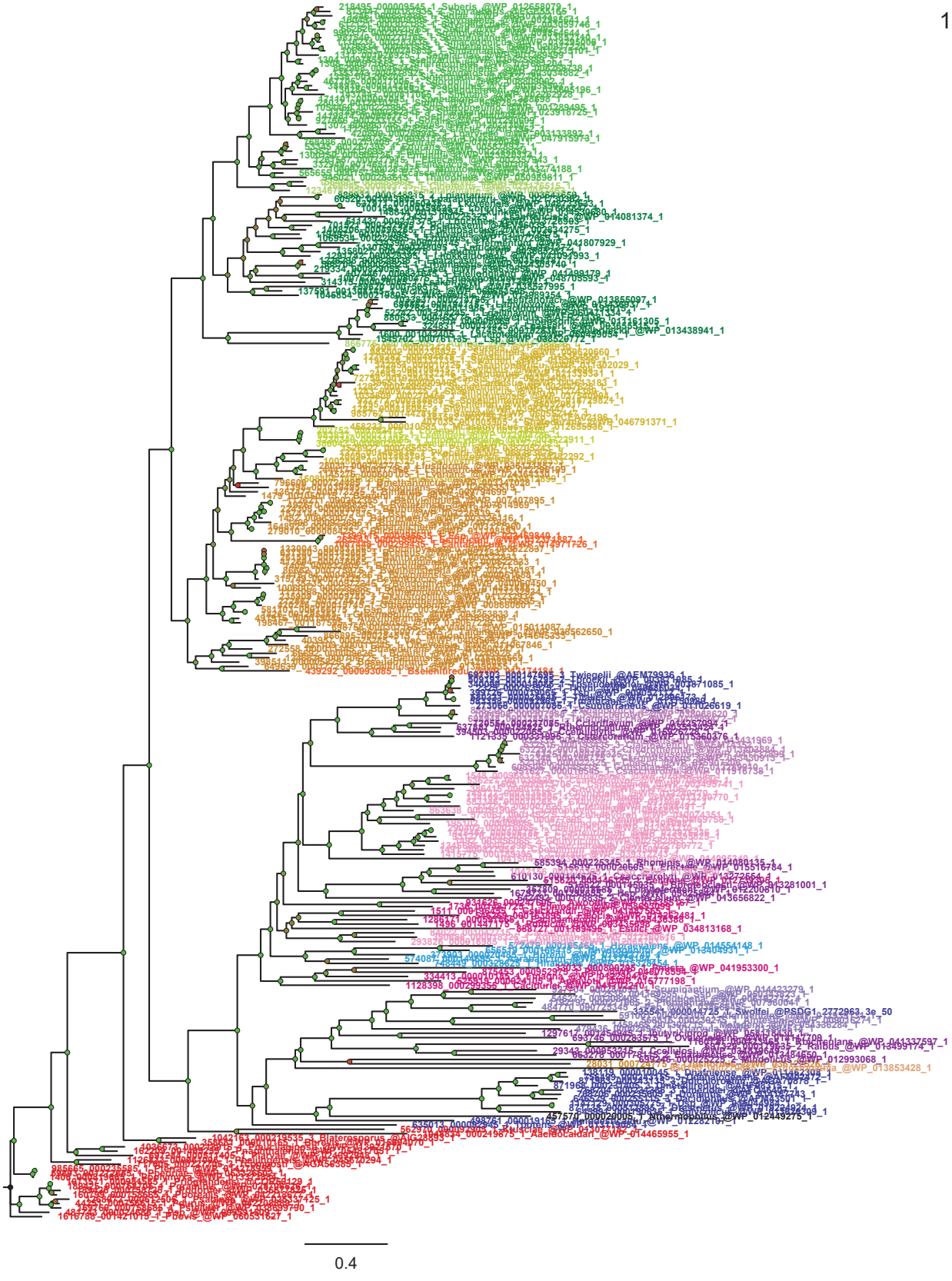

572479_000165465_1_Hpraevalens_@WP_014553199_1
656519_000166415_1_Hhydrogenifor_@WP_041595801_1
373903_000020485_1_Horenii_@WP_012636116_1
748449_000328625_1_Hhalobius_@WP_015326668_1
572479_000165465_1_Hpraevalens_@WP_014553118_1
656519_000166415_1_Hhydrogenifor_@WP_013405863_1
574087_000144695_1_Aarabaticum_@WP_013277758_1
591001_000025305_1_Afermentans_@WP_012937718_1
568816_000230275_1_Aintestini_@WP_014128149_1
568816_000230275_1_Aintestini_@WP_009016351_1
479436_000024945_1_Vparvula_@WP_012864958_1
479436_000024945_1_Vparvula_@WP_004694376_1
479436_000024945_1_Vparvula_@WP_012864912_1
712538_001189555_1_Ssp_@WP_050343548_1
546271_000208405_1_Ssputigena_@WP_006191262_1
479436_000024945_1_Vparvula_@WP_004696674_1
927704_000284095_1_Sruminantium_@WP_014424312_1
712538_001189555_1_Ssp_@WP_050341089_1
546271_000208405_1_Ssputigena_@WP_006191989_1
927704_000284095_1_Sruminantium_@WP_014424078_1
1458465_001304715_1_Melsdenii_@WP_054336776_1
712538_001189555_1_Ssp_@WP_050344256_1
546271_000208405_1_Ssputigena_@WP_006193840_1
484770_000725345_1_Psp_@WP_038667071_1
1192197_000271665_2_Pfermentans_@WP_007957103_1
484770_000725345_1_Psp_@WP_038668660_1
1192197_000271665_2_Pfermentans_@WP_007959905_1
484770_000725345_1_Psp_@WP_038667137_1
484770_000725345_1_Psp_@WP_038674809_1
0.3

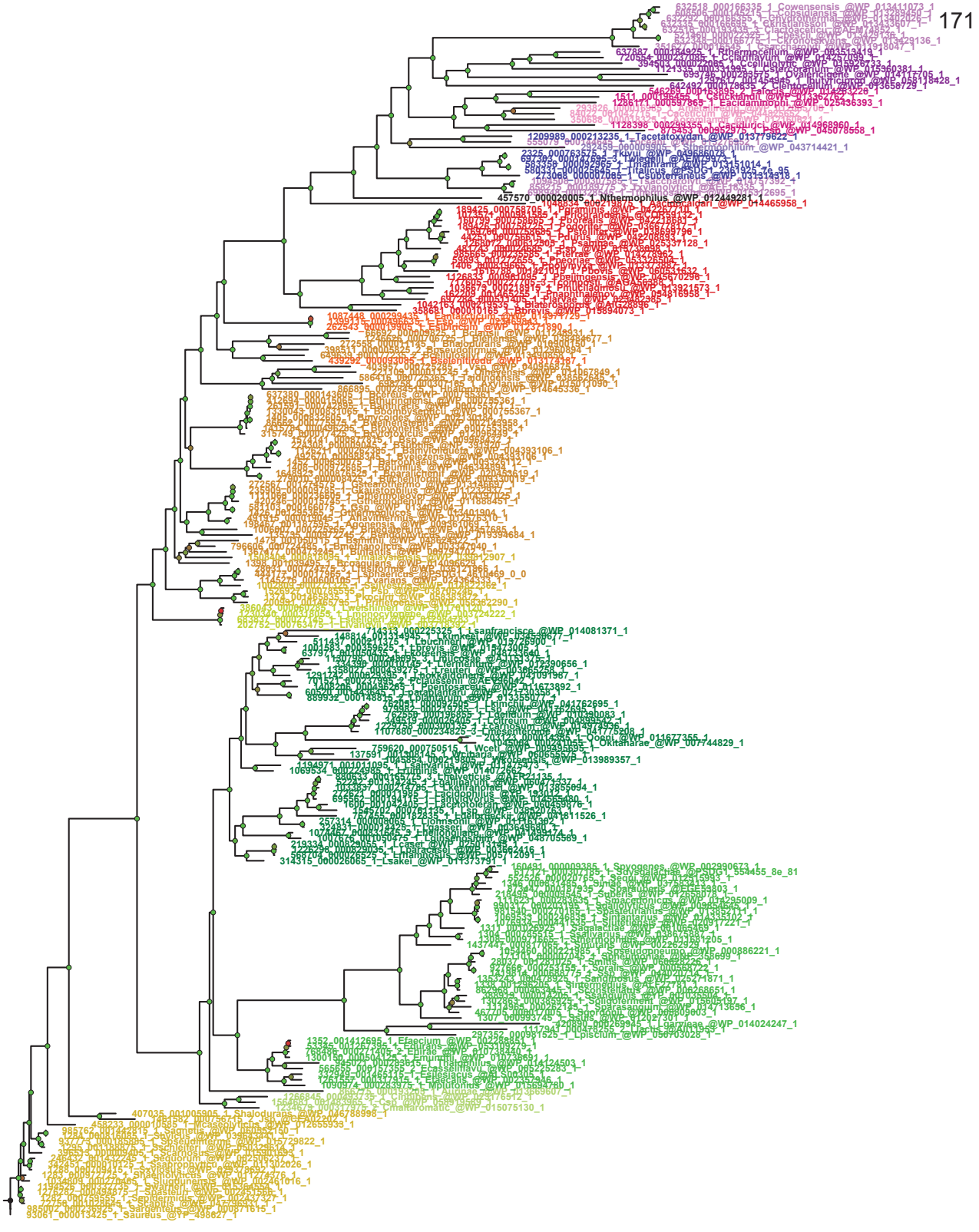

572479_000165465_1_Hpraevalens_@WP_014553200_1
656519_000166415_1_Hhydrogenifor_@WP_013405777_1
373903_000020485_1_Horenii_@WP_012636115_1
574087_000144695_1_Aarabaticum_@WP_013278422_1
748449_000328625_1_Hhalobius_@WP_015326669_1
572479_000165465_1_Hpraevalens_@WP_014553119_1
656519_000166415_1_Hhydrogenifor_@WP_013405862_1
574087_000144695_1_Aarabaticum_@WP_013277759_1
546271_000208405_1_Ssputigena_@WP_006191990_1
927704_000284095_1_Sruminantium_@WP_014424079_1
712538_001189555_1_Ssp_@WP_009442458_1
1458465_001304715_1_Melsdenii_@WP_054337019_1
591001_000025305_1_Afermentans_@WP_012937719_1
568816_000230275_1_Aintestini_@WP_009015547_1
568816_000230275_1_Aintestini_@WP_009016352_1
484770_000725345_1_Psp_@WP_038667074_1
1192197_000271665_2_Pfermentans_@WP_007957104_1
484770_000725345_1_Psp_@WP_038667134_1
484770_000725345_1_Psp_@WP_038668662_1
1192197_000271665_2_Pfermentans_@WP_007959906_1
484770_000725345_1_Psp_@WP_038666878_1
546271_000208405_1_Ssputigena_@WP_013740532_1
712538_001189555_1_Ssp_@WP_006691439_1
927704_000284095_1_Sruminantium_@WP_014424313_1
479436_000024945_1_Vparvula_@WP_012864957_1
479436_000024945_1_Vparvula_@WP_012863667_1
479436_000024945_1_Vparvula_@WP_004698134_1
712538_001189555_1_Ssp_@WP_009442604_1
479436_000024945_1_Vparvula_@WP_004698605_1
546271_000208405_1_Ssputigena_@WP_006191263_1
0.4

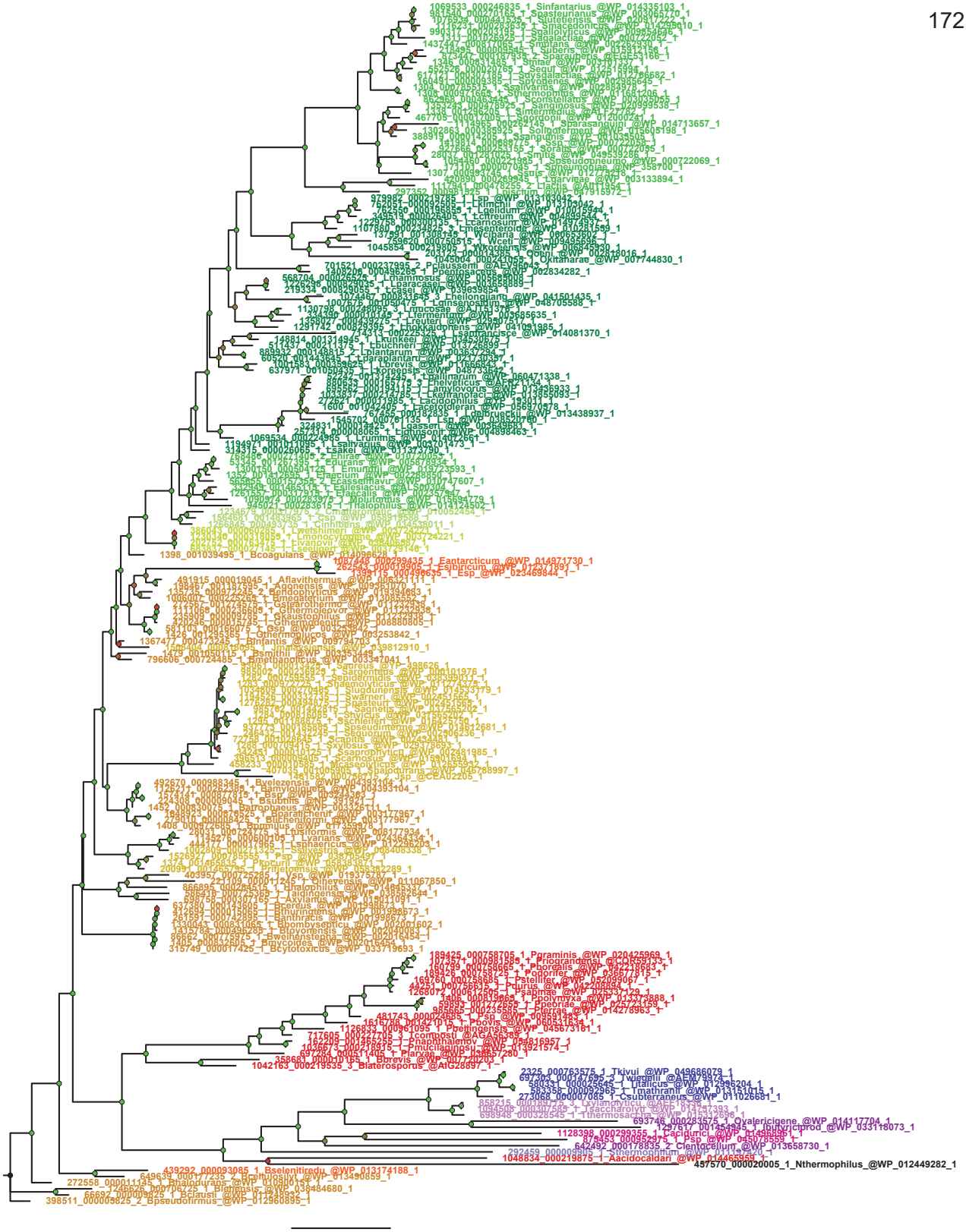

0.3

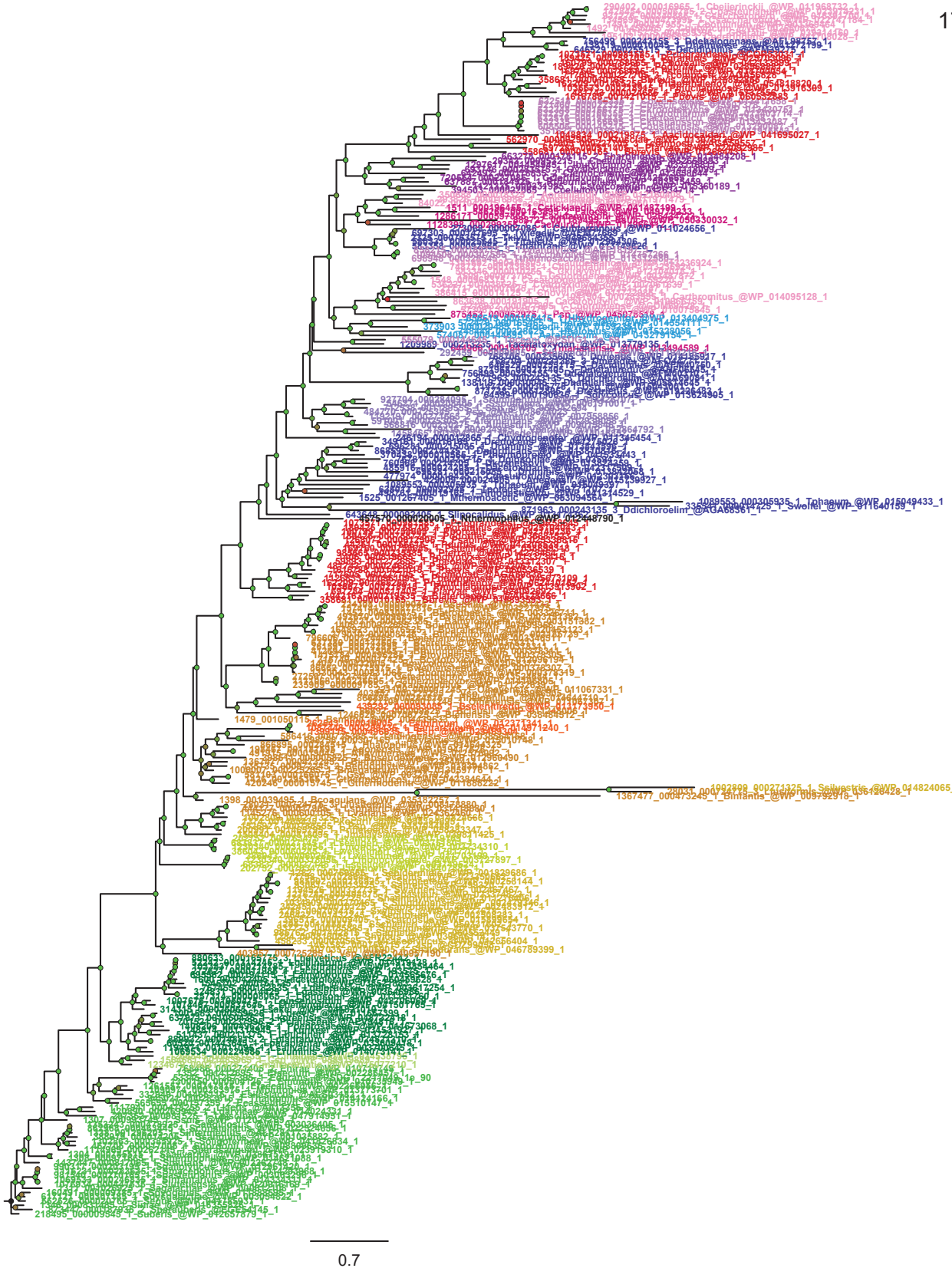

0.5

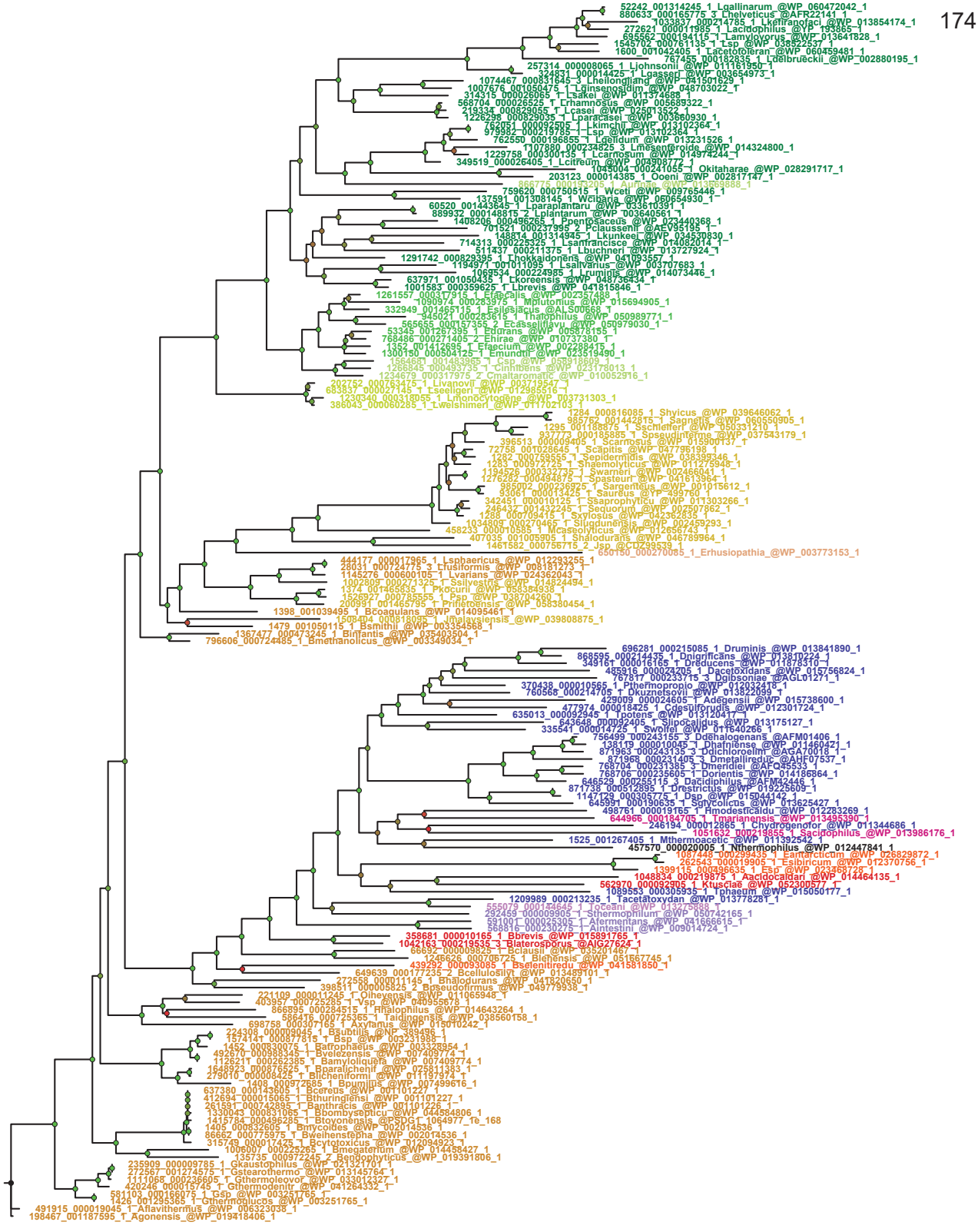

0.5

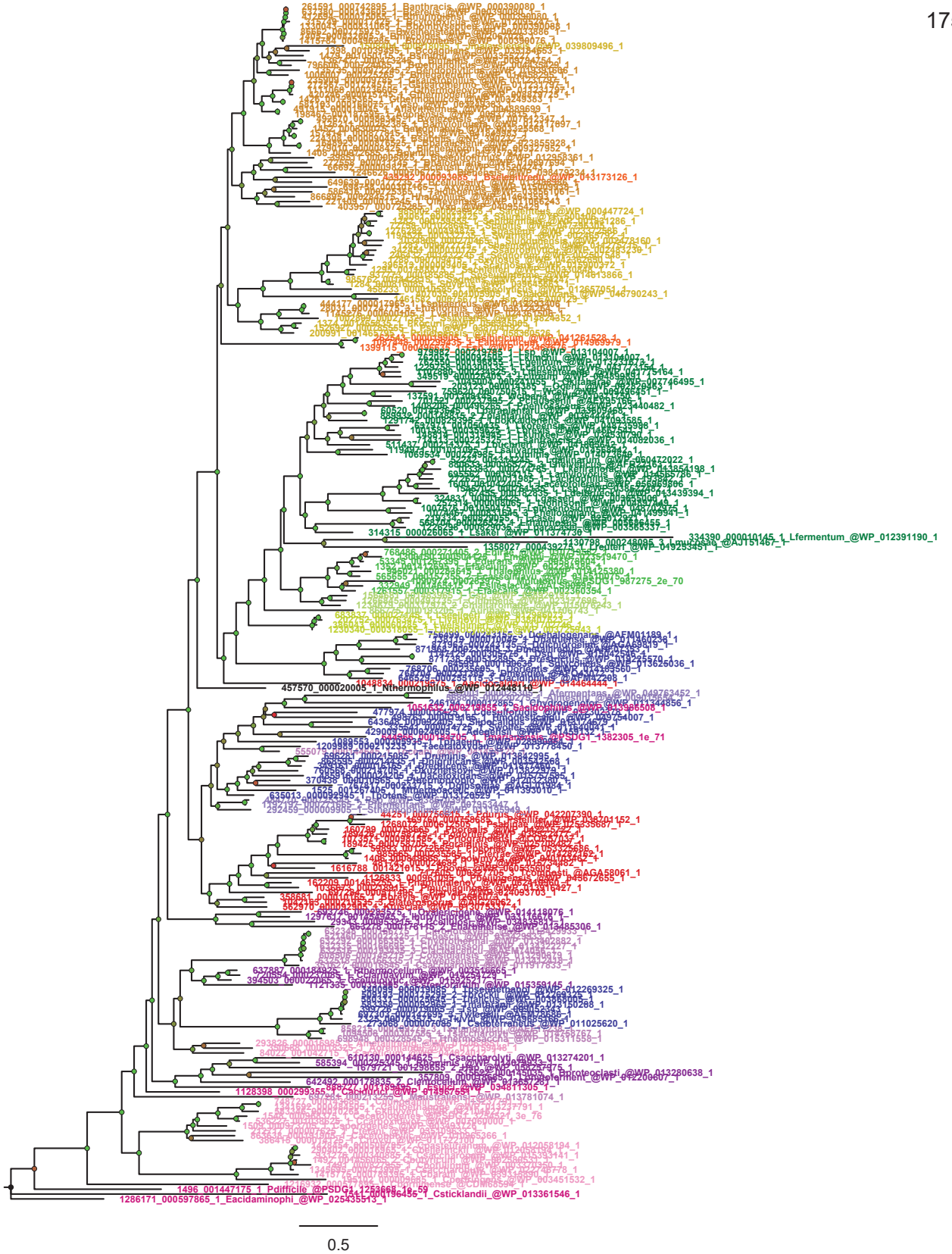

0.4


0.4

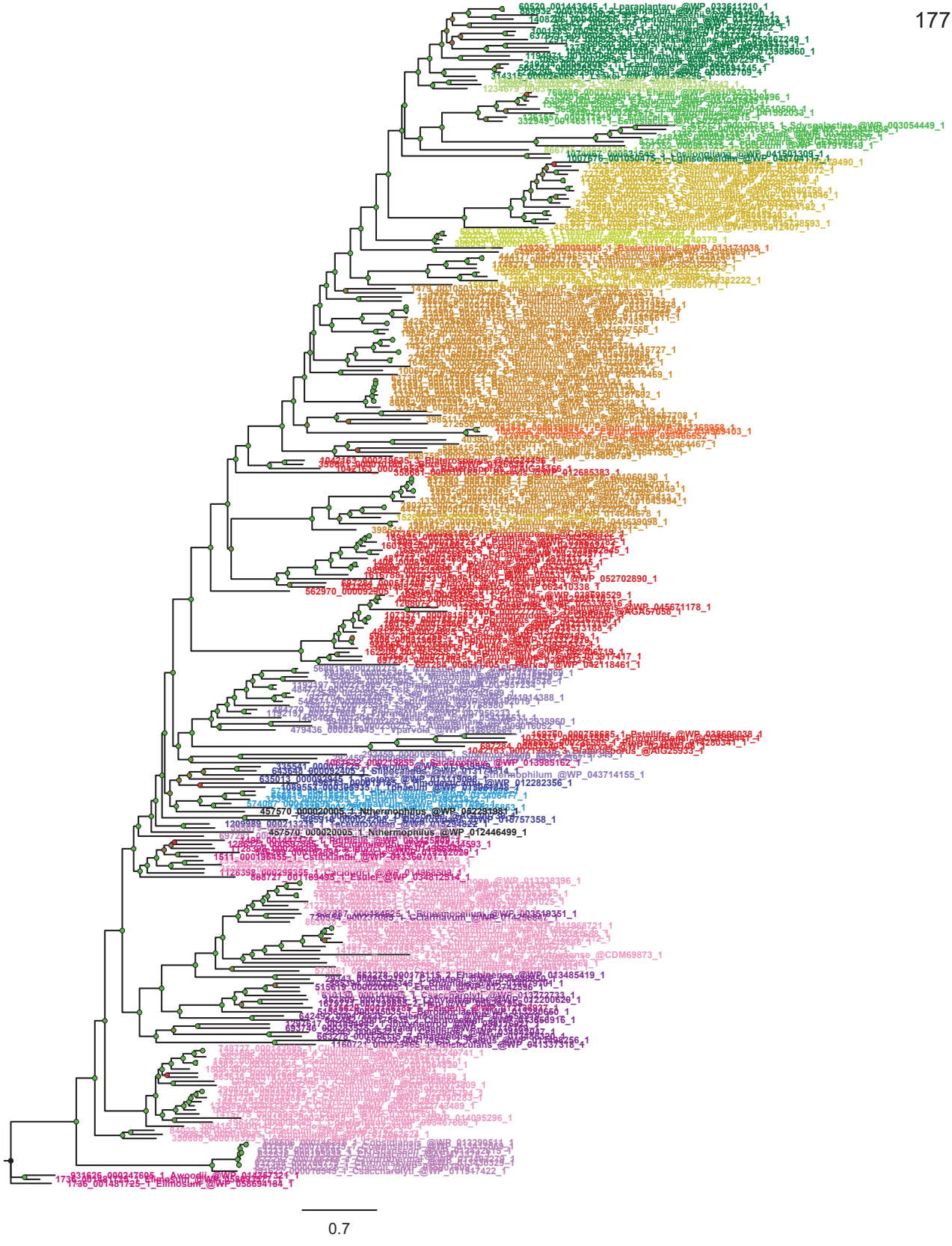

0.3

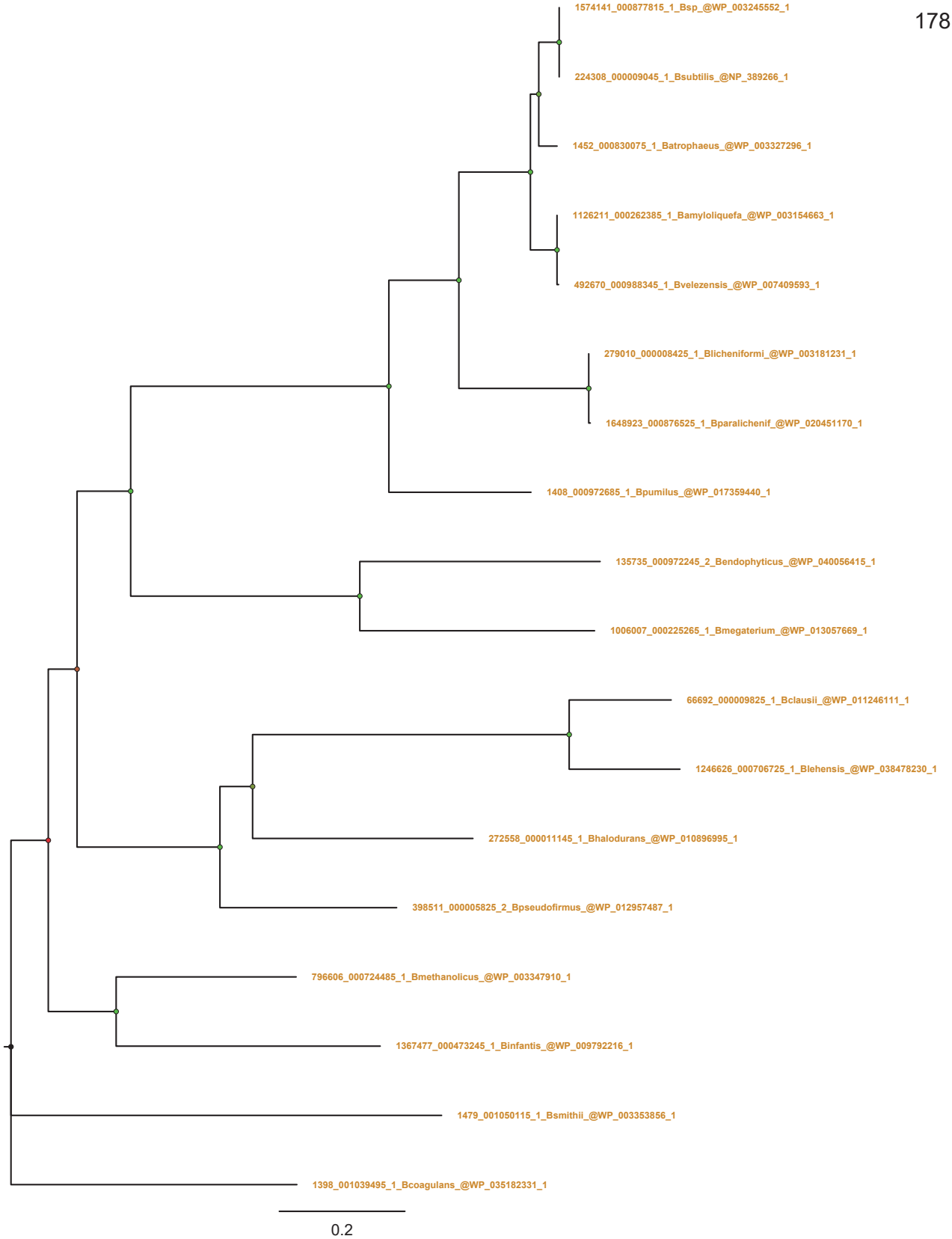

0.7

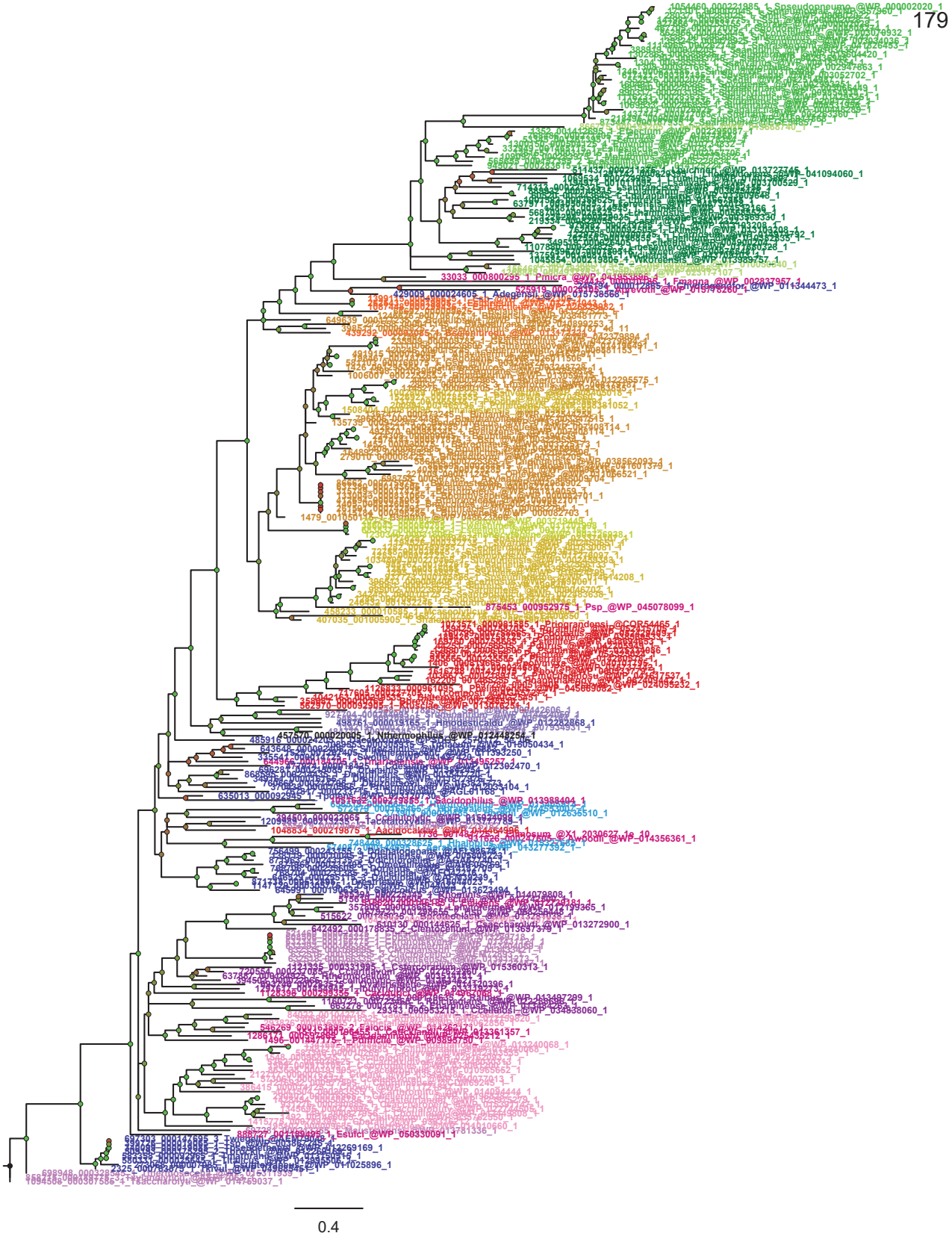

0.4


0.5

1419814_000688775_1_Ssp_@WP_044020774_1
171101_000007045_1_Spneumoniae_@NP_358640_1
28037_001281025_1_Smitis_@WP_060628381_1
927666_000253155_1_Soralis_@WP_000817865_1
1054460_000221985_1_Spseudopneumo_@WP_000817848_1
1114965_000262145_1_Sparasanguini_@WP_041826313_1
862968_000463445_1_Sconstellatus_@WP_006267924_1
1338_001296205_1_Sintermedius_@ALF27630_1
1353243_000478925_1_Sanginosus_@WP_003034462_1
388919_000014205_1_Ssanguinis_@YP_001035127_1
467705_000017005_1_Sgordonii_@WP_012000536_1
1302863_000385925_1_Soligoferment_@WP_015604956_1
981540_000270165_1_Spasteurianus_@WP_013851816_1
990317_000203195_1_Sgallolyticus_@WP_003064732_1
1116231_000283635_1_Smacedonicus_@WP_014294536_1
1069533_000246835_1_Sinfantarius_@WP_014334723_1
1076934_000441535_1_Slutetiensis_@WP_020916721_1
1311_001026925_1_Sagalactiae_@WP_000817933_1
1307_000993745_1_Ssuis_@WP_012775101_1
1437447_000817065_1_Smutans_@WP_002262304_1
160491_000009385_1_Spyogenes_@WP_011888858_1
617121_000307185_1_Sdysgalactiae_@WP_012766932_1
552526_000020765_1_Sequi_@WP_012515628_1
873447_000187935_2_Sparauberis_@EGE54410_1
1346_000831485_1_Siniae_@WP_003099620_1
218495_000009545_1_Suberis_@WP_012658412_1
1304_000785515_1_Ssalivarius_@WP_038676019_1
1308_000971665_1_Sthermophilus_@WP_014608300_1
1117941_000478255_2_Llactis_@All12726_1
297352_000981525_1_Lpiscium_@WP_047915541_1
420890_000269945_1_Lgarvieae_@WP_014024775_1
1117941_000478255_2_Llactis_@All12728_1
1033837_000214785_1_Lkefiranofaci_@WP_013854023_1
272621_000011985_1_Lacidophilus_@YP_194015_1
695562_000194115_1_Lamylovorus_@WP_013641999_1
880633_000165775_3_Lhelveticus_@AFR21940_1
52242_001314245_1_Lgallinarum_@WP_060472195_1
1600_001042405_1_Lacetotoleran_@WP_056970683_1
1545702_000761135_1_Lsp_@WP_038522576_1
324831_000014425_1_Lgasseri_@WP_003656212_1
257314_000008065_1_Ljohnsonii_@WP_011162011_1
767455_000182835_1_Ldelbrueckii_@WP_002878488_1
334390_000010145_1_Lfermentum_@WP_012391233_1
1358027_000439275_1_Lreuteri_@WP_029507410_1
1130798_000248095_3_Lmucosae_@AJT50514_1
1226298_000829035_1_Lparacasei_@WP_003660919_1
568704_000026525_1_Lrhamnosus_@WP_014571331_1
219334_000829055_1_Lcasei_@WP_025012355_1
1033837_000214785_1_Lkefiranofaci_@WP_013851262_1
1001583_000359625_1_Lbrevis_@WP_041815336_1
637971_001050435_1_Lkoreensis_@WP_048736026_1
511437_000211375_1_Lbuchneri_@WP_013727912_1
314315_000026065_1_Lsakei_@WP_011374669_1
1194971_001011095_1_Lsalivarius_@WP_003703247_1
53345_001267395_1_Edurans_@WP_016177643_1
1526927_000785555_1_Psp_@WP_038703832_1
444177_000017965_1_Lsphaericus_@WP_012291724_1
261591_000742895_1_Banthracis_@WP_000070475_1
637380_000143605_1_Bcereus_@WP_000070475_1
1405_000832605_1_Bmycoides_@WP_003193291_1
1330043_000831065_1_Bbombysepticu_@WP_044585296_1
86662_000775975_1_Bweihenstepha_@WP_002035448_1
398511_000005825_2_Bpseudofirmus_@WP_012961259_1
398511_000005825_2_Bpseudofirmus_@WP_012961021_1
293826_000016985_1_Ametalliredig_@WP_012063662_1
350688_000018325_1_Aoremlandii_@WP_041719126_1
84022_001042715_1_Caceticum_@WP_044822899_1
1496_001447175_1_Pdifficile_@PSDG1_2742701_0_0
1511_000196455_1_Csticklandii_@WP_013361450_1
1286171_000597865_1_Eacidaminophi_@WP_025435566_1
546269_000163895_2_Falocis_@WP_014261920_1
888727_001189495_1_Esulci_@WP_034811654_1
1286171_000597865_1_Eacidaminophi_@WP_025436831_1
768704_000231385_3_Dmeridiei_@AFQ44196_1
768706_000235605_1_Dorientis_@WP_014185208_1
646529_000255115_3_Dacidiphilus_@AFM41551_1
138119_000010045_1_Dhafniense_@WP_005811557_1
756499_000243155_3_Ddehalogenans_@AFM00523_1
871968_000231405_3_Dmetallireduc_@AHF06970_1
585394_000225345_1_Rhominis_@WP_014079670_1
515619_000020605_1_Erectale_@WP_012742546_1
1042156_000270305_1_Csp_@WP_013978178_1
515622_000145035_1_Bproteoclasti_@WP_013280822_1
1679721_001298655_2_Hsp_@WP_058258064_1
610130_000144625_1_Csaccharolyti_@WP_013272678_1
515620_000146185_1_Eeligens_@WP_041688082_1
357809_000018685_1_Lphytoferment_@WP_012200215_1
642492_000178835_2_Clentocellum_@PSDG1_2455092_4e_26
1508404_000818095_1_Jmalaysiensis_@WP_040109793_1
515622_000145035_1_Bproteoclasti_@WP_013282478_1
515622_000145035_1_Bproteoclasti_@WP_013282543_1
515622_000145035_1_Bproteoclasti_@WP_013282596_1
398511_000005825_2_Bpseudofirmus_@WP_003321456_1
1073571_000981585_1_Priograndensi_@CQR55940_1
189425_000758705_1_Pgraminis_@WP_025703258_1
985665_000235585_1_Pterrae_@WP_014281650_1
59893_001272655_1_Ppeoriae_@WP_023988768_1
398511_000005825_2_Bpseudofirmus_@WP_012960956_1
358681_000010165_1_Bbrevis_@WP_041749404_1
1042163_000219535_3_Blaterosporus_@AIG26753_1
697284_000511405_1_Plarvae_@WP_023484608_1
1036673_000218915_1_Pmucilaginosu_@WP_013919641_1
162209_001465255_1_Pnaphthalenov_@WP_054819846_1
717605_000227705_3_Tcomposti_@AGA57753_1
1126833_000961095_1_Pbeijingensis_@WP_045672454_1
481743_000024685_1_Psp_@WP_015734723_1
189425_000758705_1_Pgraminis_@WP_025708080_1
1073571_000981585_1_Priograndensi_@CQR56179_1
189426_000758725_1_Podorifer_@WP_036687613_1
160799_000758665_1_Pborealis_@WP_042219653_1
169760_000758685_1_Pstellifer_@WP_038696402_1
44251_000756615_1_Pdurus_@WP_042206940_1
1268072_000612505_1_Psabinae_@WP_025335177_1
1616788_001421015_1_Pbovis_@WP_060535288_1
59893_001272655_1_Ppeoriae_@WP_023988943_1
1406_000819665_1_Ppolymyxa_@WP_040103621_1
985665_000235585_1_Pterrae_@WP_014281529_1
0.5


0.7


1574141_000877815_1_Bsp_@WP_003245552_1
224308_000009045_1_Bsubtilis_@NP_389266_1
1452_000830075_1_Batrophaeus_@WP_003327296_1
1126211_000262385_1_Bamyloliquefa_@WP_003154663_1
492670_000988345_1_Bvelezensis_@WP_007409593_1
279010_000008425_1_Blicheniformi_@WP_003181231_1
1648923_000876525_1_Bparalichenif_@WP_020451170_1
1408_000972685_1_Bpumilus_@WP_017359440_1
135735_000972245_2_Bendophyticus_@WP_040056415_1
1006007_000225265_1_Bmegaterium_@WP_013057669_1
66692_000009825_1_Bclausii_@WP_011246111_1
1246626_000706725_1_Blehensis_@WP_038478230_1
272558_000011145_1_Bhalodurans_@WP_010896995_1
398511_000005825_2_Bpseudofirmus_@WP_012957487_1
796606_000724485_1_Bmethanolicus_@WP_003347910_1
1367477_000473245_1_Binfantis_@WP_009792216_1
1479_001050115_1_Bsmithii_@WP_003353856_1
1398_001039495_1_Bcoagulans_@WP_035182331_1
0.2


0.4

.13 Apparitions des familles de gènes pour les trois points chauds d'apparitions.

Point chaud	Profils phylogénétiques	Réconciliation	Les deux méthodes
Racine des *Firmicutes*	B1, XerC, SbcCD2, WalH, FemX, Noc, NagA, FtsW, Asr, RecN, RecA, RecR, Jag, StkP, MreB, Maf, MreD, NagB, B4, B6, GidA, GidB, Smc, DnaN, ZapA, YabM, LysA1, MurK, MurC, MurD, SpoVK, RodZ, CpsB, Mfd, IleS1, FtsE, MurG1, MurQ, ValS, LytBBS, ParB, RadC, DacB, TilS, ClpX, SpoIIID, MraZ, MraY, MraW, WalI, WalK, FtsH1, FtsH3, Mtf, SpoIIE, PhpP, FtsQ, FtsZ, NudF, GlmM, PCDP6, MinD, MinE, LeuS, Alr, RecF, RecG, DnaB1, GlmS, ScpB, ScpA, MreC, SunL, B5, GlmU, Ddl1, WecA, DnaA, AlaS, DnaG, SpoIIIJ1, CozE4, MurI, Alr_like, MurE, MurF, SpoVB, SpoVG, SpoVE, ParA, DivIVA, WalR, XerD, SpoIID, MurA2, RodA, SpoIIGA, SepF, FtsA, FtsL, FtsK, FtsJ, MurA1, FtsY, FtsX, PCDP7, PriA, DacF, A4, A6, DivIC, MinJ, MinC, MurT, MurB2, Mbl, DnaD	RecF, RecG, DnaB1, GlmS, ScpB, ScpA, MreC, SunL, B5, GlmU, Ddl1, WecA, DnaA, AlaS, DnaG, SpoIIIJ1, CozE4, MurI, Alr_like, MurE, MurF, SpoVB, SpoVG, SpoVE, ParA, DivIVA, WalR, XerD, SpoIID, MurA2, RodA, SpoIIGA, SepF, FtsA, FtsL, FtsK, FtsJ, MurA1, FtsY, FtsX, PCDP7, PriA, DacF, A4, A6, DivIC, MinJ, MinC, MurT, MurB2, Mbl, DnaD	RecF, RecG, DnaB1, GlmS, ScpB, ScpA, MreC, SunL, B5, GlmU, Ddl1, WecA, DnaA, AlaS, DnaG, SpoIIIJ1, CozE4, MurI, Alr_like, MurE, MurF, SpoVB, SpoVG, SpoVE, ParA, DivIVA, WalR, XerD, SpoIID, MurA2, RodA, SpoIIGA, SepF, FtsA, FtsL, FtsK, FtsJ, MurA1, FtsY, FtsX, PCDP7, PriA, DacF, A4, A6, DivIC, MinJ, MinC, MurT, MurB2, Mbl, DnaD,
Émergence des *Clostridia* hors groupe 0	RodA2, SbcCD1, Pfs, AX1, LysA2, GatD4, CpsD, CpsC, WalJ, PCDP8, IleS2, MurB1, BX1	CpsD, CpsC, WalJ, PCDP8, IleS2, MurB1, BX1, RecN, RecA, RecR, Jag, StkP, MreB, Maf, MreD, NagB, B4, B6, GidA, GidB, Smc, DnaN, ZapA, YabM, LysA1, MurC, MurD, SpoVK, RodZ, CpsB, Mfd, IleS1, FtsE, MurG1, MurQ, ValS, LytBBS, ParB, RadC, DacB, TilS, ClpX, SpoIIID, MraZ, MraY, MraW, WalI, WalK, FtsH1, FtsH3, Mtf, SpoIIE, PhpP, FtsQ, FtsZ, NudF, GlmM, PCDP6, MinD, MinE, LeuS, Alr	CpsD, CpsC, WalJ, PCDP8, IleS2, MurB1, BX1
Émergence des *Bacilli*	MacP, A3, CDP1, GatD3, A5, DacC, EzrA, FtsK2, RacA, CozE2, Rtp, CozE3, GpsB, MreBH, PCDP10, SpoIIIJ2, AnmK, MurG2, DacA, DnaBBS, XerS, DnaI, PCDP4, CDP4, B3, MurJ	RecU, GatD3, A5, EzrA, MreBH, PCDP10, GatD1, FtsK2, RacA, SpoIIIJ2, XerC, AnmK, MurG2, DacA, DnaBBS, XerS, DnaI, PCDP4, CDP4, Spo0M, Pfs, SbcCD2, WalH, B3, MurJ, AX1	MreBH, PCDP10, SpoIIIJ2, AnmK, MurG2, DacA, DnaBBS, XerS, DnaI, PCDP4, CDP4, B3, MurJ, EzrA, FtsK2, RacA, GatD3, A5

.14 Nombre de FT proies par FCC.

FCC appât	nombre de FT proies	FCC dans FT	Clan de profil phylogénétique
DnaI	55	oui	Bacilli
DnaBBS	55		
A3	46	oui	
EzrA	44		
MurJ	42	oui	
GpsB	42	oui	
DacA	33		
SpoIIIJ2	27		
RecU	14	oui	
MinE	14	oui	
CDP4	6		
PCDP10	6	oui	
CDP1	5	oui	
MacP	4		
DacF	31		Spo
SpoIIID	31	oui	
SpoIID	31	oui	
SpoIIE	31		
SpoIIGA	31		
SpoVB	23		
DacB	23		
SpoVE	23		
B3	24		
MapZ	5		
WalR	10		
SpoVG	5	oui	
MinD	3	oui	
MurG1	2	oui	
MurG2	2	oui	
IleS1	2	oui	
MurT	2	oui	
BX1	2	oui	
GatD1	1	oui	
GatD4	1	oui	
ParA	1	oui	
NagA	1	oui	
WecA	1	oui	
MurA2	1		
Jag	1		
CozE3	1	oui	
SbcC1	1		
PCDP6	1	oui	
Pfs	1	oui	
LysA2	1	oui	
MreB	1	oui	
LysA1	1	oui	
Maf	1	oui	
Rtp	1	oui	
MraZ	1	oui	
MurM	1	oui	
CpsD	1	oui	
MinC	1		
IleS2	1	oui	
NagB	1	oui	
FtsH3	1		
SbcC2	1		
MurQ	1	oui	
WalK	1	oui	
XerD	1	oui	
Mbl	1	oui	
AnmK	1	oui	
B1	1	oui	

.15 FT proies présentant une synténie avec des FCC.

FT proies présentant une synténie avec des FCC. Un seuil de 0,12 de coefficient de Jaccard a été utilisé. Les FT surlignées en gris correspondent à des fonctions générales ou inconnues.

FT proie	FCC/Clan FCC appât	FCC synténique	Coefficient de Jaccard	Fonction prédite de la FT
FAM003396	SbcC1	SbcC1	0,92	exonuclease sbcCD subunit D
FAM000552	SbcC2	SbcC2	0,882	exonuclease sbcCD subunit D
FAM003953	MurT	MurT	0,806	glutamine amidotransferase
FAM002649	Spo	DacF	0,625	anti-sigma F factor antagonist
FAM002215	Spo	SpoIIGA	0,622	sporulation sigma factor SigE
FAM002883	Spo	DacF	0,606	anti-sigma F factor
FAM007382	Bacilli	MurA1	0,464	membrane protein
FAM002114	Spo	DivIC	0,459	sporulation protein YabP
FAM002355	Spo	SpoIIGA	0,451	hypothetical protein
FAM002658	Spo	RecN	0,438	sporulation transcription factor Spo0A
FAM004209	Bacilli	FtsK2	0,432	tRNA-binding protein
FAM002110	Spo	Mfd	0,424	stage V sporulation protein T
FAM002871	Spo	RecN	0,42	SpoIVB peptidase
FAM001900	Spo	SpoIIGA	0,41	sporulation sigma factor SigG
FAM002600	Spo	FtsK	0,384	translocation-enhancing protein TepA
FAM002110	Spo	YabM	0,373	stage V sporulation protein T
FAM002563	Spo	DacB	0,358	spore maturation protein
FAM002562	Spo	DacB	0,358	spore maturation protein
FAM001359	Bacilli	FtsX	0,347	peptidase S41
FAM007203	B3	SpoVB	0,347	hypothetical protein
FAM002635	Spo	DacB	0,312	sporulation protein YtfJ
FAM002862	Bacilli	DnaG	0,305	deoxyguanosinetriphosphate triphosphohydrolase
FAM001900	SpoVG	DacF	0,295	sporulation sigma factor SigG
FAM001221	Spo	ClpX	0,287	endopeptidase La
FAM002215	Spo	FtsZ	0,279	sporulation sigma factor SigE
FAM002635	Spo	ScpB	0,271	sporulation protein YtfJ
FAM001359	Bacilli	MinJ	0,255	peptidase S41
FAM006999	B3	YabM	0,254	hypothetical protein
FAM002894	Spo	ParB	0,239	spore protease YyaC
FAM002635	Spo	ScpA	0,224	sporulation protein YtfJ
FAM002894	Spo	ParA	0,194	spore protease YyaC
FAM006999	B3	Mfd	0,191	hypothetical protein
FAM001900	SpoVG	FtsZ	0,187	sporulation sigma factor SigG
FAM007225	B3	XerD	0,186	stage II sporulation protein M
FAM002355	Spo	PCDP6	0,182	hypothetical protein
FAM001059	Bacilli	RecA	0,181	ribosomal protein S12 methylthiotransferase RimO
FAM002384	Spo	SpoVB	0,18	membrane protein
FAM002649	Spo	SpoIIGA	0,178	anti-sigma F factor antagonist
FAM007058	B3	Alr	0,178	sporulation protein
FAM002883	Spo	SpoIIGA	0,177	anti-sigma F factor
FAM002114	Spo	YabM	0,174	sporulation protein YabP
FAM002355	Spo	FtsZ	0,172	hypothetical protein
FAM007225	B3	PCDP1	0,17	stage II sporulation protein M
FAM001059	Bacilli	FtsH3	0,169	ribosomal protein S12 methylthiotransferase RimO
FAM001059	Bacilli	FtsK	0,164	ribosomal protein S12 methylthiotransferase RimO
FAM002215	Spo	DacF	0,163	sporulation sigma factor SigE
FAM002950	Bacilli	XerD	0,161	S1 RNA-binding protein
FAM004069	Spo	MurJ	0,159	carbohydrate kinase
FAM002355	Spo	SepF	0,155	hypothetical protein
FAM001900	SpoVG	PCDP6	0,152	sporulation sigma factor SigG
FAM002114	Spo	SpoIIE	0,152	sporulation protein YabP
FAM007232	Bacilli	CozE3	0,151	hypothetical protein
FAM002355	Spo	DacF	0,149	hypothetical protein
FAM007229	B3	GatD3	0,143	germination protein YpeB
FAM004835	Bacilli	MurI	0,142	CBS domain-containing protein
FAM002215	Spo	PCDP6	0,138	sporulation sigma factor SigE
FAM007382	Bacilli	Mbl	0,137	membrane protein
FAM002110	Spo	DivIC	0,135	stage V sporulation protein T
FAM002950	Bacilli	ScpA	0,133	S1 RNA-binding protein
FAM002114	Spo	Mfd	0,128	sporulation protein YabP
FAM007192	B3	MurI	0,128	sporulation protein
FAM001900	SpoVG	SepF	0,126	sporulation sigma factor SigG
FAM002215	Spo	B6	0,122	sporulation sigma factor SigE
FAM004043	Bacilli	CpsB	0,122	LytR family transcriptional regulator
FAM006999	B3	DivIC	0,12	hypothetical protein

.16 FT corrélant avec des FCC du clan Spo.

FT corrélant avec des FCC du clan Spo. La correspondance entre les familles identifiées par [468] et celles identifiées dans cette étude est représentée. Les FT correspondant à des FCC sont indiquées en bleu clair.

FT	annotation fonctionnelle	Famille identifiée dans Traag et al., 2013	Synténie avec FCC
FAM002091	AbrB family transcriptional regulator		
FAM002110	stage V sporulation protein T		Mfd, YabM, DivIC
FAM002114	sporulation protein YabP		YabM, SpoIIE, Mfd
FAM002202	spore cortex-lytic enzyme		
FAM002215	sporulation sigma factor SigE		SpoIIGA, FtsZ, DacF, PCDP6, B6
FAM002330	stage V sporulation protein D		
FAM002377	spore protein		
FAM002415	stage IV sporulation protein A		
FAM002562	spore maturation protein		DacB
FAM002563	spore maturation protein		DacB
FAM002564	stage V sporulation protein S		
FAM002635	sporulation protein YtfJ		DacB, ScpA, ScpB
FAM002647	stage V sporulation protein AC		
FAM002648	stage V sporulation protein AD		
FAM002649	anti-sigma F factor antagonist		DacF, SpoIIGA
FAM002658	sporulation transcription factor Spo0A		RecN
FAM002665	stage III sporulation protein AD		
FAM002666	stage III sporulation protein AC		
FAM002668	stage III sporulation protein AA		
FAM002758	germination protein KA		
FAM002805	sporulation transcriptional regulator SpoIIID		
FAM002851	stage II sporulation protein D		
FAM002871	SpoIVB peptidase		RecN
FAM002883	anti-sigma F factor		SpoIIGA
FAM003273	N-acetylmuramoyl-L-alanine amidase		
FAM002894	spore protease YyaC		ParA, ParB
FAM002606	polysaccharide deacetylase family sporulation protein PdaB		
FAM001221	endopeptidase La	-	ClpX
FAM002384	membrane protein	-	SpoVB
FAM004069	carbohydrate kinase	-	MurJ
FAM002127	hypothetical protein	-	
FAM002280	GPR endopeptidase	-	
FAM002692	asparagine synthetase B	-	
FAM002355	hypothetical protein	ymxH	SpoIIGA, PCDP6, FtsZ, SepF, DacF
FAM002557	hypothetical protein	ylzA	
FAM002600	translocation-enhancing protein TepA	ymfB	FtsK
-	DNA-binding transcriptional regulator	bkdR	
-	Butyrate kinase	buk	
-	Hypothetical protein	ylmC	
-	Polysaccharide deacetylase	ylxY	
-	DksA-like regulator	yteA	
-	DksA-like regulator	ylyA	

.17 FT corrélant avec des FCC du clan *Bacilli*.

FT corrélant avec des FCC du clan *Bacilli*. Les FT correspondant à des FCC sont indiquées en bleu clair. Les FT bien construites sont annotées en vert, celles mal reconstruites en rouge et celles dont la qualité de reconstruction est difficile à établie avec un « ? ».

FT	annotation fonctionnelle	FCC associée	Synténie avec FCC	Vérification
FAM003972	primosomal protein DnaI	DnaI		Bacilli
FAM004011	polysaccharide biosynthesis protein	MurJ		Bacilli
FAM004085	Holliday junction resolvase RecU	RecU		Bacilli
FAM004088	cell division protein GpsB	GpsB		Bacilli
FAM007236	penicillin-binding protein	A3		Bacilli
FAM004051	phosphotransferase	CorZ		Bacilli
FAM002270	cell division topological specificity factor MinE	MinE		Clos/Neg/Tiss
FAM003899	AI-2E family transporter	CozE1/CozE3		Bacilli
FAM007237	hypothetical protein	PCDP10		Bacilli
FAM004136	RNA-binding protein S4	PCDP8		Mal clusterisé
FAM007197	rod shape-determining protein MreC	MreC		Mal clusterisé
FAM007214	DNA primase	DnaG		Mal clusterisé
FAM002318	DNA primase	DnaG		Mal clusterisé
FAM004176	penicillin-binding protein	B4		Mal clusterisé
FAM004231	penicillin-binding protein	B5		Mal clusterisé
FAM007382	membrane protein		MurA1, Mbl	Bacilli
FAM007232	hypothetical protein		CozE3	Bacilli
FAM004043	LytR family transcriptional regulator		CpsB	Bacilli
FAM003916	competence protein ComG			Bacilli
FAM003922	competence protein ComF			Bacilli
FAM004006	DEAD/DEAH box helicase			Bacilli
FAM004042	ferredoxin--NADP(+) reductase			Bacilli
FAM004079	transcriptional regulator Spx			Bacilli
FAM004097	hypothetical protein			Bacilli
FAM004126	hypothetical protein			Bacilli
FAM004137	hypothetical protein			Bacilli
FAM004140	1-acyl-sn-glycerol-3-phosphate acyltransferase			Bacilli
FAM004173	orotate phosphoribosyltransferase			Bacilli
FAM004208	hypothetical protein			Bacilli
FAM004217	hypothetical protein			Bacilli
FAM004218	ribosome biogenesis GTPase YqeH			Bacilli
FAM004225	competence protein ComGC			Bacilli
FAM004237	oligoribonuclease			Bacilli
FAM007182	hypothetical protein			Bacilli
FAM007328	haloacid dehalogenase			Bacilli
FAM007740	hypothetical protein			Bacilli
FAM007321	adenylate cyclase			Bacilli
FAM000027	haloacid dehalogenase			Bacilli
FAM003918	16S rRNA methyltransferase			Bacilli
FAM004166	SprT family protein			Bacilli
FAM007160	5'-nucleotidase			Bacilli
FAM007481	hypothetical protein			Bacilli
FAM000331	nitroreductase			Clos/Neg/Tiss
FAM001097	B12-binding domain-containing radical SAM protein			Clos/Neg/Tiss
FAM000761	3-isopropylmalate dehydratase small subunit			Clos/Neg/Tiss
FAM001166	NADH dehydrogenase			Clos/Neg/Tiss
FAM001167	NADH dehydrogenase			Clos/Neg/Tiss
FAM001218	aminopeptidase			Clos/Neg/Tiss
FAM002862	deoxyguanosinetriphosphate triphosphohydrolase			Clos/Neg/Tiss/Paeni
FAM001059	ribosomal protein S12 methylthiotransferase RimO		RecA, FtsH3, FtsK	Clos/Neg/Tiss/Paeni
FAM000239	rubrerythrin			Clos/Neg/Tiss
FAM000701	thioether cross-link-forming SCIFF peptide maturase			Clos/Neg/Tiss
FAM000717	Fis family transcriptional regulator			Clos/Neg/Tiss/Paeni
FAM000962	hydroxylamine reductase			Clos/Neg/Tiss/Paeni/Cereus
FAM001143	hypothetical protein			Clos/Neg/Tiss/Paeni
FAM001977				Clos/Neg/Tiss/Paeni
FAM002609	exopolyphosphatase			Clos/Neg/Tiss/Paeni
FAM002857	lytic transglycosylase			Clos/Neg/Tiss/Paeni
FAM000436	ferredoxin			?
FAM001359	peptidase S41		MinJ	?
FAM007217	glycerol phosphate lipoteichoic acid synthase			?
FAM004835	CBS domain-containing protein		MurI	?
FAM004209	tRNA-binding protein		FtsK, MurC, SpoIIIE	Mal clusterisé
FAM005647	diacylglycerol kinase			Mal clusterisé
FAM007159	Fe-S cluster assembly protein SufD			Mal clusterisé
FAM007206	single-stranded DNA exonuclease			Mal clusterisé
FAM000739	fumarate hydratase			Mal clusterisé
FAM000740	fumarate hydratase			Mal clusterisé
FAM003104	manganese-dependent inorganic pyrophosphatase			Mal clusterisé
FAM002394	hypothetical protein			Mal clusterisé
FAM004184	HD domain-containing protein			Mal clusterisé
FAM004191	DNA processing protein DprA			Mal clusterisé
FAM004104	NAD(+) kinase			Mal clusterisé
FAM001199	preprotein translocase subunit SecF			Mal clusterisé
FAM001200	preprotein translocase subunit SecD			Mal clusterisé
FAM003952	ribonuclease Z			Mal clusterisé
FAM002627	DNA processing protein DprA			Mal clusterisé
FAM002950	S1 RNA-binding protein		XerD, ScpA	Mal clusterisé

.18 FT corrélant avec d'autres FCC.

FT corrélant avec d'autres FCC. Les FT correspondant à des FCC sont indiquées en bleu clair.

FCC	FT	annotation fonctionnelle	FCC associée	Synténie avec FCC
B3	FAM007202	stage V sporulation protein B	SpoVB	
B3	FAM007474	stage II sporulation protein E	SpoIIE	
B3	FAM004840	AI-2E family transporter	CozE4	
B3	FAM002594	ATP-dependent helicase		
B3	FAM004246	helix-turn-helix transcriptional regulator		
B3	FAM004413	hypothetical protein		
B3	FAM004428	peptidase M23		
B3	FAM004437	spore coat protein		
B3	FAM004459	adapter protein MecA		
B3	FAM004666	stage V sporulation protein M		
B3	FAM004681	transcriptional regulator		
B3	FAM006999	hypothetical protein		YabM, Mfd, DivIC
B3	FAM007058	sporulation protein		Alr
B3	FAM007192	sporulation protein		MurI
B3	FAM007203	hypothetical protein		SpoVB
B3	FAM007221	stage III sporulation protein AG		
B3	FAM007225	stage II sporulation protein M		XerD, NudF
B3	FAM007229	germination protein YpeB		GatD3
B3	FAM007343	hypothetical protein		
B3	FAM007384	stage II sporulation protein R		
B3	FAM007389	cell wall hydrolase		
B3	FAM007390	spore coat protein GerQ		
B3	FAM007893	late competence protein ComER		
B3	FAM008444	hypothetical protein		
WalR	FAM004136	RNA-binding protein S4	PCDP8	
WalR	FAM007237	hypothetical protein	PCDP10	
WalR	FAM004166	SprT family protein		
WalR	FAM003952	ribonuclease Z		
WalR	FAM004042	ferredoxin--NADP(+) reductase		
WalR	FAM004173	orotate phosphoribosyltransferase		
WalR	FAM004184	HD domain-containing protein		
WalR	FAM004218	ribosome biogenesis GTPase YqeH		
WalR	FAM005647	diacylglycerol kinase		
WalR	FAM004104	NAD(+) kinase		
MapZ	FAM033469	transposase		
MapZ	FAM033481	membrane protein		
MapZ	FAM033491	superoxide dismutase		
MapZ	FAM033510	hypothetical protein		
MapZ	FAM033564	copper ABC transporter permease		
SpoVG	FAM001936	septation protein SpoVG	SpoVG	
SpoVG	FAM001900	sporulation sigma factor SigG		SpoIIGA, DacF, FtsZ, PCDP6, SepF
SpoVG	FAM000537	Rrf2 family transcriptional regulator		
SpoVG	FAM001321	2,3-bisphosphoglycerate-independent phosphoglycerate mutase		
SpoVG	FAM003883	general stress protein		
MurT	FAM003953	glutamine amidotransferase		MurT
MurT	FAM004039	UDP-N-acetylmuramyl peptide synthase	MurT	
SbcC1	FAM003396	exonuclease sbcCD subunit D		SbcC1
SbcC2	FAM000552	exonuclease sbcCD subunit D		SbcC2
MurA2	FAM004589	hypothetical protein		

.19 Comparatif des différents outils de visualisation de contextes génomiques.

Comparatif des différents outils de visualisation de contextes génomiques.

	GeConT	JContextExporer	MGcvV	Easyfig	SyntTax	GeneSpy
date	2004	2013	2013	2011	2013	2017
URL	http://operons.ibt.unam.mx/gctNG/	https://omictools.com/jcontextexplorer-tool	http://mgcv.cmbi.ru.nl/	http://mjsull.github.io/Easyfig/	http://archaea.u-psud.fr/synttax/	https://lbbe.univ-lyon1.fr/GeneSpy/
Référence	Martinez-Guerrero et al., 2008	Seitzer et al., 2013	Overmars et al., 2013	Sullivan et al., 2011	Oberto, 2013	Garcia et al., 2018
Entrée	Mots clés Séquence	Mots clés	GI Locus tag Position génomique	GBFF	Séquence	Numéro d'accession + Numéro d'assemblage Fiche GenBank Numéro d'accession
Recherche des homologues	BLASTp Annotation	COG	-	-	tBLASTn	Mots clés BLASTp (local et NCBI)
Coloration des gènes	COG, Pfam, KEGG Pfam KEGG	COG Annotation	COG Pfam Localisation cellulaire %CG	BLASTp	COG	Annotation Customisé
Rapidité de lancement	Lent	Rapide	Rapide	Lent	Très lent	Modéré
Type de base de données	En ligne : RefSeq	Locale : GFF, GBFF	En ligne : RefSeq	-	En ligne : GenBank	Locale : GFF (GenBank et RefSeq)
Composition de la base données	1062 génomes, complets uniquement	-	2773 génomes (2017), complets uniquement	-	8025 chromosomes (2017), complets uniquement	-
Rendu graphique	Faible	Moyen	Moyen	Bon	Bon	Bon
Export	?	PNG, JPG, EPS	PNG, PDF	BMP, SVG	PDF	PNG, JPEG, PDF, SVG, EPS, TIF, iTOL,
Plus	Affichage des promoteurs/terminateurs Taxonomie complète	Arbres NJ des contextes	Sélection des gènes sur la figure	Relations d'homologies entre les gènes affichée	Taxonomie complète Grand nombre de génomes	Mapping des contextes sur une phylogénie Sélection des gènes sur la figure Navigation le long des génomes Construction et gestion simples de la base de données Nombreuses options de customisation des figures
Moins	Très peu de génomes Rendu graphique médiocre Export de figure ne fonctionne pas	Import des bases de données long et à refaire à chaque lancement Pas adapté au nouveau format de GFF Base de données lourde Lent si > 150 contextes	Peu de génomes	GBFF d'une région génomique difficile à obtenir automatiquement Gestion des couleurs difficile Le contexte des gènes en sens inverse est impossible à retourner Limite du nombre de génome basse	Peu d'options pour customiser les figures	Fuite de mémoire Rafraîchissement lent sur Windows et MacOS

.20 Différents formats utilisés par GeneSpy.

1. Format des fichiers GFF (Generic Feature File)
2. Format des fichiers GFM (GFf Minimal content)
3. Format des listes de qsouches
4. Format des fichiers d'entrée
5. Format de fichiers de coloration

Format générique

<Contig><source><region/gene/CDS><start><end><score><strand><phase><attributes>

attributes: Name=<Accession number> ; product=<Biochemical function> ; locus_tag=<locus tag> ; genome/type=<genomic/plasmidic ; gene=<name of gene> ; ID=<Accession number> (RAST/PROKKA)

Exemple

```
##gff-version 3
#!gff-spec-version 1.21
#!processor NCBI annotwriter
#!genome-build ASM1800v1
#!genome-build-accession NCBI_Assembly:GCA_000018005.1
##sequence-region CP000820.1 1 8982042
##species https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?id=298653
CP000820.1  Genbank region  1  8982042  .  +  .  ID=id0;Dbxref=taxon:298653;Is_circular=true;Name=ANONYMOUS;gbkey=Src;genome=chromosome;mol_type=genomic DNA;strain=EAN1.pe
CP000820.1  Genbank gene  71  1654  .  +  .  ID=gene0;Name=Franean1_0001;gbkey=Gene;gene_biotype=protein_coding;locus_tag=Franean1_0001
CP000820.1  Genbank CDS 71  1654  .  +  0
ID=cds0;Parent=gene0;Dbxref=InterPro:IPR001957,InterPro:IPR003593,InterPro:IPR013159,InterPro:IPR013317,NCBI_GP:ABW09470.1;Name=ABW09470.1;Note=KEGG: fal:FRAAL0002 chromosoma
replication initiator protein~TIGRFAM: chromosomal replication initiator protein DnaA~PFAM: Chromosomal replication initiator DnaA domain3B Chromosomal replication initiator
DnaA~SMART: AAA ATPase;gbkey=CDS;product=chromosomal replication initiator protein DnaA;protein_id=ABW09470.1;transl_table=11
CP000820.1  Genbank gene  2937  3959  .  +  .  ID=gene1;Name=Franean1_0002;gbkey=Gene;gene_biotype=protein_coding;locus_tag=Franean1_0002
CP000820.1  Genbank CDS 2937  3959  .  +  0  ID=cds1;Parent=gene1;Dbxref=NCBI_GP:ABW09471.1;Name=ABW09471.1;Note=KEGG: hsa:23524 serine/arginine repetitive matrix
2;gbkey=CDS;product=hypothetical protein;protein_id=ABW09471.1;transl_table=11
CP000820.1  Genbank gene  3956  5116  .  +  .  ID=gene2;Name=Franean1_0003;gbkey=Gene;gene_biotype=protein_coding;locus_tag=Franean1_0003
CP000820.1  Genbank CDS 3956  5116  .  +  0  ID=cds2;Parent=gene2;Dbxref=InterPro:IPR001001,NCBI_GP:ABW09472.1;Name=ABW09472.1;Note=KEGG: fal:FRAAL0004 DNA polymerase
III2C beta chain~TIGRFAM: DNA polymerase III2C beta subunit~PFAM: DNA polymerase III beta chain;gbkey=CDS;product=DNA polymerase III2C beta
subunit;protein_id=ABW09472.1;transl_table=11
CP000820.1  Genbank gene  5118  6491  .  +  .  ID=gene3;Name=Franean1_0004;gbkey=Gene;gene_biotype=protein_coding;locus_tag=Franean1_0004
CP000820.1  Genbank CDS 5118  6491  .  +  0  ID=cds3;Parent=gene3;Dbxref=InterPro:IPR001238,InterPro:IPR003395,NCBI_GP:ABW09473.1;Name=ABW09473.1;Note=TIGRFAM: DNA
replication and repair protein RecF~PFAM: SMC domain protein~KEGG: sen:SACE_0005 DNA replication and repair protein;gbkey=CDS;product=DNA replication and repair protein
RecF;protein_id=ABW09473.1;transl_table=11
CP000820.1  Genbank gene  6712  7245  .  +  .  ID=gene4;Name=Franean1_0005;gbkey=Gene;gene_biotype=protein_coding;locus_tag=Franean1_0005
CP000820.1  Genbank CDS 6712  7245  .  +  0  ID=cds4;Parent=gene4;Dbxref=InterPro:IPR007922,NCBI_GP:ABW09474.1;Name=ABW09474.1;Note=PFAM: protein of unknown function
DUF721~KEGG: fal:FRAAL0006 hypothetical protein;gbkey=CDS;product=protein of unknown function DUF721;protein_id=ABW09474.1;transl_table=11
CP000820.1  Genbank gene  7555  9513  .  +  .  ID=gene5;Name=Franean1_0006;gbkey=Gene;gene_biotype=protein_coding;locus_tag=Franean1_0006
CP000820.1  Genbank CDS 7555  9513  .  +  0
ID=cds5;Parent=gene5;Dbxref=InterPro:IPR000565,InterPro:IPR001241,InterPro:IPR002288,InterPro:IPR003594,InterPro:IPR006171,InterPro:IPR011557,InterPro:IPR011558,InterPro:IPR0
fal:FRAAL0008 DNA gyrase subunit B~TIGRFAM: DNA gyrase2C B subunit~PFAM: DNA gyrase subunit B domain protein3B ATP-binding region ATPase domain protein3B TOPRIM domain protei
DNA topoisomerase type IIA subunit B region 2 domain protein~SMART: DNA topoisomerase II;gbkey=CDS;product=DNA gyrase2C B subunit;protein_id=ABW09475.1;transl_table=11
CP000820.1  Genbank gene  9711  12224  .  +  .  ID=gene6;Name=Franean1_0007;gbkey=Gene;gene_biotype=protein_coding;locus_tag=Franean1_0007
CP000820.1  Genbank CDS 9711  12224  .  +  0
ID=cds6;Parent=gene6;Dbxref=InterPro:IPR002205,InterPro:IPR005743,InterPro:IPR006691,NCBI_GP:ABW09476.1;Name=ABW09476.1;Note=KEGG: fal:FRAAL0009 DNA gyrase2C subunit A2C type
topoisomerase~TIGRFAM: DNA gyrase2C A subunit~PFAM: DNA gyrase/topoisomerase IV subunit A3B DNA gyrase repeat beta-propeller;gbkey=CDS;product=DNA gyrase2C A
subunit;protein_id=ABW09476.1;transl_table=11
CP000820.1  Genbank gene  12333  14270  .  +  .  ID=gene7;Name=Franean1_0008;gbkey=Gene;gene_biotype=protein_coding;locus_tag=Franean1_0008
CP000820.1  Genbank CDS 12333  14270  .  +  0  ID=cds7;Parent=gene7;Dbxref=NCBI_GP:ABW09477.1;Name=ABW09477.1;Note=KEGG: fal:FRAAL0011 hypothetical
protein;gbkey=CDS;product=hypothetical protein;protein_id=ABW09477.1;transl_table=11
```

ormat générique

nf>\t<sup>\t<orientation>\t<accession>\t<locus tag>\t<annotation>\t[<name>]

xemple

```
#CP000820.1 (chromosome)
71  1654    +   ABW09470.1  Franean1_0001   chromosomal replication initiator protein DnaA
2937    3959    +   ABW09471.1  Franean1_0002   hypothetical protein
3956    5116    +   ABW09472.1  Franean1_0003   DNA polymerase III%2C beta subunit
5118    6491    +   ABW09473.1  Franean1_0004   DNA replication and repair protein RecF
6712    7245    +   ABW09474.1  Franean1_0005   protein of unknown function DUF721
7555    9513    +   ABW09475.1  Franean1_0006   DNA gyrase%2C B subunit
9711    12224   +   ABW09476.1  Franean1_0007   DNA gyrase%2C A subunit
12333   14270   +   ABW09477.1  Franean1_0008   hypothetical protein
14343   14419   +   NA  Franean1_R0001  tRNA-Ile    tRNA
15245   15862   +   ABW09478.1  Franean1_0009   conserved hypothetical protein
15821   17065   -   ABW09479.1  Franean1_0010   conserved hypothetical protein
17052   17921   -   ABW09480.1  Franean1_0011   Methyltransferase type 11
17915   19291   -   ABW09481.1  Franean1_0012   glycosyl transferase group 1
19531   19959   -   ABW09482.1  Franean1_0013   OsmC family protein
20169   20873   +   ABW09483.1  Franean1_0014   transcriptional regulator%2C TetR family
20881   21498   -   ABW09484.1  Franean1_0015   Endopeptidase Clp
21745   22848   -   ABW09485.1  Franean1_0016   secretory lipase
23241   24332   +   ABW09486.1  Franean1_0017   SNARE associated Golgi protein
24683   25021   -   ABW09487.1  Franean1_0018   nitrogen regulatory protein P-II
25136   25678   +   ABW09488.1  Franean1_0019   PEBP family protein
25853   26017   +   ABW09489.1  Franean1_0020   hypothetical protein
26159   26716   -   ABW09490.1  Franean1_0021   binding-protein-dependent transport systems inner membrane component
26953   28206   -   ABW09491.1  Franean1_0022   extracellular solute-binding protein family 1
28203   29546   -   ABW09492.1  Franean1_0023   spermidine/putrescine ABC transporter ATPase subunit
29898   30662   +   ABW09493.1  Franean1_0024   conserved hypothetical protein
30889   31950   +   ABW09494.1  Franean1_0025   Cobyrinic acid ac-diamide synthase
33852   33924   +   NA  Franean1_R0002  tRNA-Ala    tRNA
35609   36013   -   ABW09495.1  Franean1_0027   hypothetical protein
36044   36967   -   ABW09496.1  Franean1_0028   hypothetical protein
37290   37865   +   ABW09497.1  Franean1_0029   hypothetical protein
37862   38839   +   ABW09498.1  Franean1_0030   hypothetical protein
39584   40471   -   ABW09499.1  Franean1_0031   hypothetical protein
40475   41677   -   ABW09500.1  Franean1_0032   Protein-L-isoaspartate(D-aspartate) O-methyltransferase
42120   42485   -   ABW09501.1  Franean1_0033   hypothetical protein
42461   43171   -   ABW09502.1  Franean1_0034   hypothetical protein
43221   43496   -   ABW09503.1  Franean1_0035   hypothetical protein
43699   44568   +   ABW09504.1  Franean1_0036   helix-turn-helix domain protein
44565   44774   +   ABW09505.1  Franean1_0037   protein of unknown function DUF397
45693   46814   +   ABW09506.1  Franean1_0038   integrase family protein
46811   47149   +   ABW09507.1  Franean1_0039   putative transcriptional regulator%2C XRE family
47146   48660   +   ABW09508.1  Franean1_0040   hypothetical protein
48701   49600   -   ABW09509.1  Franean1_0041   aminoglycoside phosphotransferase
49995   50705   +   ABW09510.1  Franean1_0042   Resolvase domain
```

Format générique

<Assembly>\t<Name of strain>\t<Other name (displayed in iTOL formats, useful to fit with phylogenies)>

Exemple

```
GCF_900187245.1 Streptococcus acidominimus NCTC11291    Streptococcus_acidominimus_NCTC11291
GCF_001729925.1 Streptococcus agalactiae WC1535 Streptococcus_agalactiae_WC1535
GCF_000478925.1 Streptococcus anginosus subsp. whileyi MAS624   Streptococcus_anginosus_subsp__whileyi_MAS624
GCF_000463445.1 Streptococcus constellatus subsp. pharyngis C818    Streptococcus_constellatus_subsp__pharyngis_C818
GCF_000385925.1 Streptococcus cristatus AS 1.3089   Streptococcus_cristatus_AS_1_3089
GCF_000307185.1 Streptococcus dysgalactiae subsp. equisimilis RE378 Streptococcus_dysgalactiae_subsp__equisimilis_RE378
GCF_000020765.1 Streptococcus equi subsp. zooepidemicus MGCS10565   Streptococcus_equi_subsp__zooepidemicus_MGCS10565
GCF_002000985.1 Streptococcus gallolyticus subsp. gallolyticus DSM 16831    Streptococcus_gallolyticus_subsp__gallolyticus_DSM_16831
GCF_000017005.1 Streptococcus gordonii str. Challis substr. CH1 Streptococcus_gordonii_str__Challis_substr__CH1
GCF_001598035.1 Streptococcus halotolerans HTS9 Streptococcus_halotolerans_HTS9
GCF_001708305.1 Streptococcus himalayensis HTS2 Streptococcus_himalayensis_HTS2
GCF_000246835.1 Streptococcus infantarius subsp. infantarius CJ18   Streptococcus_infantarius_subsp__infantarius_CJ18
GCF_000831485.1 Streptococcus iniae YSFST01-82  Streptococcus_iniae_YSFST01_82
GCF_002356055.1 Streptococcus intermedius TYG1620   Streptococcus_intermedius_TYG1620
GCF_000441535.1 Streptococcus lutetiensis 033   Streptococcus_lutetiensis_033
GCF_000283635.1 Streptococcus macedonicus ACA-DC 198    Streptococcus_macedonicus_ACA_DC_198
GCF_001623565.1 Streptococcus marmotae HTS5 Streptococcus_marmotae_HTS5
GCF_900187085.1 Streptococcus merionis NCTC13788    Streptococcus_merionis_NCTC13788
GCF_001560895.1 Streptococcus mitis SVGS_061    Streptococcus_mitis_SVGS_061
GCF_000817065.1 Streptococcus mutans UA159-FR   Streptococcus_mutans_UA159_FR
GCF_002356415.1 Streptococcus oralis subsp. tigurinus   Streptococcus_oralis_subsp__tigurinus
GCF_001642085.1 Streptococcus pantholopis TA 26 Streptococcus_pantholopis_TA_26
GCF_000262145.1 Streptococcus parasanguinis FW213   Streptococcus_parasanguinis_FW213
GCF_000213825.1 Streptococcus parauberis KCTC 11537 Streptococcus_parauberis_KCTC_11537
GCF_000270165.1 Streptococcus pasteurianus ATCC 43144   Streptococcus_pasteurianus_ATCC_43144
GCF_002953735.1 Streptococcus pluranimalium TH11417 Streptococcus_pluranimalium_TH11417
GCF_000299015.1 Streptococcus pneumoniae gamPNI0373 Streptococcus_pneumoniae_gamPNI0373
GCF_000221985.1 Streptococcus pseudopneumoniae IS7493   Streptococcus_pseudopneumoniae_IS7493
GCF_000009385.1 Streptococcus pyogenes str. Manfredo    Streptococcus_pyogenes_str__Manfredo
GCF_000785515.1 Streptococcus salivarius NCTC 8618  Streptococcus_salivarius_NCTC_8618
GCF_000014205.1 Streptococcus sanguinis SK36    Streptococcus_sanguinis_SK36
GCF_001553685.1 Streptococcus sp. oral taxon 431    Streptococcus_sp__oral_taxon_431
GCF_000993745.1 Streptococcus suis ZY05719  Streptococcus_suis_ZY05719
GCF_002286255.1 Streptococcus thermophilus ST3  Streptococcus_thermophilus_ST3
GCF_002814135.1 Streptococcus uberis NZ01   Streptococcus_uberis_NZ01
```

ormat générique

<Assembly>\t<Accession>

xemple

```
 1 GCF_001708305.1 WP_068991451.1
 2 GCF_001983955.1 WP_000028988.1
 3 GCF_000253155.1 WP_000028988.1
 4 GCF_002356415.1 WP_000028988.1
 5 GCF_002355895.1 WP_000028988.1
 6 GCF_000180515.1 WP_078373048.1
 7 GCF_002073835.2 WP_037600939.1
 8 GCF_000091645.1 WP_002277624.1
 9 GCF_000772245.1 WP_002989099.1
10 GCF_001021955.1 WP_002989099.1
11 GCF_000307535.1 WP_002989099.1
12 GCF_000993765.1 WP_002989099.1
13 GCF_001620285.1 WP_002989099.1
14 GCF_000743015.1 WP_002989099.1
15 GCF_000230295.1 WP_002989099.1
16 GCF_001014305.1 WP_002989099.1
17 GCF_001559175.2 WP_002989099.1
18 GCF_000772185.1 WP_002989099.1
19 GCF_001051095.1 WP_002989099.1
20 GCF_000422045.1 WP_002989099.1
21 GCF_002557755.1 WP_002989099.1
22 GCF_001014285.1 WP_002989099.1
23 GCF_001547715.1 WP_002989099.1
24 GCF_002844355.1 WP_002989099.1
25 GCF_000349925.2 WP_002989099.1
26 GCF_000756485.1 WP_002989099.1
27 GCF_001020185.2 WP_002989099.1
28 GCF_001535565.1 WP_002989099.1
29 GCF_001535505.1 WP_002989099.1
30 GCF_000011665.1 WP_002989099.1
31 GCF_000013545.1 WP_002989099.1
32 GCF_000013525.1 WP_002989099.1
```

Format générique

><color>

<Annotation (accession/locus tag/function/name)>

Exemple

```
1 >red
2 ftsa
3 >blue
4 ftsz
5 >green
6 murC
7 >gray
8 sensor histidine kinase
```

Format générique

><color>

>Assembly>\t<Accession>

Exemple

```
1 >red
2 GCF_000008565.1 NP_294353.1 #cell division protein FtsA
3 GCF_000020685.1 WP_012693534.1  #cell division protein FtsA
4 GCF_000091545.1 YP_144354.1 #cell division protein FtsA
5 GCF_001399775.1 WP_003048578.1  #cell division protein FtsA
6 >blue
7 GCF_001399775.1 WP_003048580.1  #cell division protein FtsZ
8 GCF_000091545.1 YP_144355.1 #cell division protein FtsZ
9 GCF_000020685.1 WP_041227514.1  #cell division protein FtsZ
10 GCF_000008565.1 NP_294354.1 #cell division protein FtsZ
11 >green
12 GCF_001399775.1 WP_003048572.1  #UDP-N-acetylmuramate--L-alanine ligase
13 GCF_000091545.1 YP_144351.1 #UDP-N-acetylmuramate--alanine ligase
14 GCF_000020685.1 WP_012693531.1  #UDP-N-acetylmuramate--L-alanine ligase
15 GCF_000008565.1 NP_294350.1 #UDP-N-acetylmuramate--alanine ligase
16 >gray
17 GCF_001399775.1 WP_003048585.1  #sensor histidine kinase
```

.21 Publications scientifiques publiées au cours de ma thèse.

Publications acceptées

Garcia PS, Simorre JP, Brochier-armanet C, Grangeasse C. Cell division of Streptococcus pneumoniae : think positive !. Curr Opin Microbiol. 2016 ;34 :18-23.

Zucchini L, Mercy C, Garcia PS, et al. PASTA repeats of the protein kinase StkP interconnect cell constriction and separation of Streptococcus pneumoniae. Nat Microbiol. 2018 ;3(2) :197-209.

Fenton AK, Manuse S, Flores-kim J, et al. Phosphorylation-dependent activation of the cell wall synthase PBP2a in by MacP. Proc Natl Acad Sci USA. 2018 ;115(11) :2812-2817.

Garcia PS, Jauffrit F, Grangeasse C, Brochier-armanet C. GeneSpy, a user-friendly and flexible genomic context visualizer. Bioinformatics. 2018 ;

Carriel D and Garcia PS, Castelli F, et al. A novel subfamily of bacterial AAT-fold basic amino acid decarboxylases and functional characterization of its first representative : Pseudomonas aeruginosa LdcA. Genome Biol Evol. 2018 ;

Publications en révisions

Mercy C, Lavergne JP,Slager J, Ducret A, Garcia PS, Noirot-Gros MF, Dubarry N, Nourikyan J, Veening JW, Grangeasse C. RocS drives chromosome segregation and nucleoid occlusion in Streptococcus pneumoniae. BioRxiv (https://www.biorxiv.org/content/early/2018/07/03/359943). 2018.

Kandiah E, Carriel-Lopez D, Garcia PS, Banzhaf M, Félix J, Bacia M, Kritikos G, Typas A, Brochier-Armanet C, Elsen S and Gutsche I. Structural, functional, and evolutionary insights

into the lysine decarboxylase LdcA, a novel player in Pseudomonas aeruginosa polyamine metabolism and antibiotics resistance.

Publications en cours d'écriture

Garcia PS, Duchemin W, Brochier-Armanet C, Grangeasse C. Etude phylogénomique des protéines du cycle cellulaire chez les *Firmicutes*.

Cohen D, Garcia PS, Brochier-Armanet C, Chemin I. Impact de l'aflatoxine sur les mutations chez HBV.