Histoire évolutive de Xanthomonas arboricola, espèce bactérienne composée de souches pathogènes et commensales

Déborah Merda

▶ **To cite this version:**

Déborah Merda. Histoire évolutive de Xanthomonas arboricola, espèce bactérienne composée de souches pathogènes et commensales. Sciences agricoles. Université d'Angers, 2016. Français. NNT : 2016ANGE0028 . tel-01460117

HAL Id: tel-01460117
https://theses.hal.science/tel-01460117

Submitted on 7 Feb 2017

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.

UNIVERSITE BRETAGNE LOIRE

Thèse de Doctorat

Déborah MERDA

Mémoire présenté en vue de l'obtention du
grade de Docteur de l'Université d'Angers
sous le sceau de l'Université Bretagne Loire

École doctorale : *Végétal, Environnement, Nutrition, Agroalimentaire, Mer (VENAM)*

Discipline : *Biologie des organismes, des populations, écologie*
Spécialité : *Biologie des organismes*
Unité de recherche : *EmerSys, IRHS (Université d'angers, INRA, Agrocampus Ouest)-UMR1345*

Soutenue le *29 Novembre 2016*

Thèse N° : *135629*

Histoire évolutive de *Xanthomonas arboricola*, espèce bactérienne composée de souches pathogènes et commensales

JURY

Rapporteurs :	**Virginie RAVIGNE**, chargé de recherche, CIRAD **Olivier TENAILLON**, directeur de recherche, Inserm
Examinateurs :	**Didier THARREAU,** directeur de recherche, CIRAD **Philippe SIMONEAU**, professeur des universités, Université d'Angers
Directeur de Thèse :	**Marie-Agnès JACQUES,** directeur de recherche, INRA
Co-encadrant de Thèse :	**Marion FISCHER-LE SAUX,** chargé de recherche, INRA

Remerciements

Entre manips, stress et début de rédaction du rapport du stage de Master 2 à Lille, j'ai pensé que ce serait une bonne idée de poursuivre en thèse. J'ai alors découvert ce projet qui m'a fascinée dès le début. Malgré les difficultés que j'ai pu rencontrer, ma motivation est toujours restée intacte, et cela grâce aux personnes qui m'ont entourée au cours de ces trois années.

Tout d'abord, **Marie-Agnès** et **Marion** je souhaite vous remercier d'avoir accepté que l'on travaille ensemble sur ce projet. Merci à vous deux pour l'agréable ambiance de travail qui a toujours été au rendez-vous. Avec ce duo d'encadrantes que vous formez, je savais que je pouvais avoir confiance, aucun détail ne vous échappait. Marie-Agnès, je te remercie d'avoir toujours fait passer ma thèse en priorité malgré un emploi du temps très chargé. Marion, je te remercie pour tout ce temps que tu as passé à mes côtés, notamment lors de la rédaction des articles puis de la thèse.

Merci à **Martial Briand**, bio-informaticien de l'équipe. Quelle aide précieuse pour moi qui savait à peine ouvrir un terminal en arrivant ! Tu as toujours été là pour régler mes problèmes de scripts qui ne fonctionnaient pas… Martial, je te remercie pour ta disponibilité, ta patience quand tu m'as enseigné ces quelques lignes de code, mais également pour toutes nos discussions et l'ambiance qui régnait dans notre bureau.

Merci à **Christophe Lemaire** de m'avoir fait aimer la génét des pops à travers ses cours. Ce n'était pas vraiment ce à quoi je me destinais en allant en fac de bio, mais je suis très contente d'avoir changé de trajectoire. Merci de m'avoir donné l'opportunité de découvrir le monde de la recherche grâce à mon stage de M1. Merci également d'avoir toujours été là, pendant la thèse, pour me rappeler les bases théoriques lorsqu'elles me semblaient un peu floues. Enfin, Merci pour ton aide dans la rédaction de l'article.

Au cours de cette thèse, j'ai eu la chance de pouvoir collaborer avec deux personnes : **Eran Bosis** et **Adrien Rieux**. Eran, merci pour nos discussions lors du congrès à Banyuls-sur-mer. Adrien, merci également pour nos discussions et pour ta relecture.

Merci aux deux stagiaires qui m'ont bien aidée : **Alexis De Wever** et **Louis Legoff**. Alexis, bonne chance pour ton début de thèse, et Louis, pour ton M2.

Je souhaite remercier les membres du jury qui ont accepté d'évaluer mon travail. Merci à mes deux rapporteurs : **Olivier Tenaillon** et **Virginie Ravigné**, pour leurs commentaires sur le manuscrit. Et merci à mes deux examinateurs : **Didier Tharreau** et **Philippe Simoneau**.

Merci aux trois personnes qui étaient présentes lors de mon comité de thèse, et qui m'ont aiguillée lors de ces réunions : **Laurent Noel**, **Sylvain Brisse** et **Mahendra Mariadassou**. Mahendra, je te remercie également pour ton accueil au sein de ton laboratoire.

Je remercie les deux financeurs de ce projet de thèse : **La région Pays de la Loire** et **l'INRA**.

Je remercie également l'ensemble les membres du réseau **FNX** (French Network on Xanthomonads) pour les réunions annuelles qui se font toujours dans la bonne humeur et pour vos conseils. Merci d'avoir financé une partie de mon voyage à Bogota pour que je puisse y présenter mes travaux. Cela a été une formidable expérience. J'en profite pour remercier particulièrement **Aude Cerutti** pour la bonne ambiance pendant ce voyage. Bonne chance pour ta dernière année ! Et, Merci à **Ralf Koebnik** pour sa relecture du chapitre 4.

Merci aux personnes d'EmerSys qui ont participé activement à ce projet. **Sophie C**, merci pour tous tes commentaires sur ce manuscrit, pour tes conseils au jour le jour et pour ta bonne humeur. **Anne**, merci pour ton aide lors des manips plantes et pour nos discussions le midi. Et, **Sophie B**. merci pour la partie « répertoire » du chapitre 1.

Merci à **Chrystelle** pour les pauses et ta bonne humeur. Merci à **Karine**, pour tes conseils dans les analyses. Merci à **Nico D** pour toutes nos discussions et tes blagues souvent pas drôles…. **Nico C**, merci les soirées, merci à **Mathieu**, pour ta bonne humeur, tes conseils, et ton aide dans le séquencage des génomes. Merci à **Perrine**, pour les souches commandées souvent à la dernière minute. Enfin, merci à **Cécile** et **Géraldine**.

Merci à l'ensemble des personnes qui mangent à la cafet et qui contribuent à la bonne humeur qui y règne : **Marie Noëlle**, **Anne**, **Chrystelle**, **Jérôme**, **Ludo**, **Armelle**, **Tristan**, **Valérian** et **Christophe**.

Merci **Ludo** pour tes conseils avec Mendeley qui m'ont bien simplifié la vie.

Merci aux doctorants : **Mylène**, **Samir**, **Valérian**, et merci aux anciens **Arnaud** et **Céline**.

Merci à **Thibault Leroy** d'avoir toujours été là pour me conseiller, m'écouter et de m'avoir donné le courage de me lancer dans l'aventure du doctorat. Thibault, je te remercie pour ton amitié, ton soutien et pour les différentes relectures que tu as pu faire. Pour te répondre ici aux remerciements de ta thèse, oui mon passage chez EcoFun a été une formidable expérience.

Je pense également à mon passage à Lille, cette année aura été nécessaire pour moi. Un très grand merci à **Céline Poux** et **Xavier Vekemans** pour tout ce que j'ai pu apprendre auprès de vous. Merci à **Vincent Castric** pour tes encouragements. Merci également à toutes les personnes que j'ai pu y rencontrer, particulièrement **Hélène**

Martin et **Raphaël Meheust**. Merci pour votre soutien pendant toute la thèse et également pendant la rédaction. C'était rassurant d'avancer ensemble tous les trois.

Merci aux angevins qui ont contribué à me changer les idées. **Alban**, merci de m'avoir fait relativiser quand j'en avais besoin. **Jeff**, c'est un vrai plaisir de t'avoir rencontré. Merci pour toutes nos discussions toujours très enrichissantes, et nos soirées Quick… Enfin, un grand merci aux 4 bisounours (**Marie**, **Mélanie**, **Amandine** et **Gaëlle**), pour les discussions chat / lapin, les soirées et autres fous rires. Vous avez été un réel soutien !

A Rome, il y a également des personnes qui ont participé à me faire lever le nez de la thèse. Un grand merci à **Alex** pour ta vision de la vie, ta joie de vivre et toutes tes paroles réconfortantes. Merci à **Dave** pour ton accueil lors de mon dernier passage avec ma sœur. Merci à **Veronica**, **Aly** et **Anzu**, sans oublier il bambino **Ibrahima**.

Merci aux copains de fac : **Amandine**, **Pym**, **Gaby**, **David**, **Elodie**, **Hélène** et **Aurélie**.

Merci aux amis de longue date, tellement lourds que j'ai parfois du mal à les supporter.

Jordi, je te remercie pour toutes nos heures passées au téléphone. C'est vrai qu'on n'est pas toujours très sage, mais un peu d'insouciance de temps en temps ça fait vraiment du bien. Merci pour ta compréhension quand j'ai raté le rendez-vous lycée à cause de la rédaction. Et, je te remercie également pour ce super voyage que nous avons fait en Sicile avec la Lilie, un vrai bol d'air pur au milieu de la thèse. Là, où nous avons appris qu'un petit déjeuner se prenait toujours en sautant !

Elise, merci pour ta façon de tout rationnaliser, c'est impressionnant ! Mais, en même temps, c'est nécessaire. Quand la fatigue était trop grande, un petit coup de téléphone à la Lilie qui me faisait prendre conscience que mon esprit divaguait. Elise, je te remercie également pour « toutes les certaines incroyables qu'il nous est raconté » !

Perrine, c'est toujours comme lorsqu'on avait 15 ans quand on se voit, et ça c'est juste un pur bonheur. Donc merci pour ça. La fin de thèse a fait qu'on n'a pas profité assez avant ton prochain grand voyage, mais on se rattrapera.

Florian, l'arnaqueur, quelle patience ! Je t'embête souvent, je le reconnais. Si, malgré tout, un jour tu acceptais de revenir chez moi, je te ferais un meilleur accueil, promis !

Benoit, le dernier arrivé mais qui a su très vite se prendre au jeu, merci pour tous ces bons moments passés en ta compagnie, et d'avoir fait un petit arrêt par Angers pendant la rédaction. Par contre, Goji ne te remercie pas, vu tout ce que tu lui as fait vivre.

Enfin, Merci à **Florence**, **Mathieu**, **Vanessa** et **Wilfried** pour tous ces moments !

Merci à ma famille pour tout votre amour, vous êtes ma force. Merci à **mes parents** qui, peu importe mes choix, m'ont toujours soutenue et fais confiance. Merci à mes frères et sœurs : **Tatien** pour toutes les vidéos loufoques que tu me fais voir, **Gégé**

pour toutes nos choré et ton côté sauveur en cas d'invasion, **Rox**, merci pour ta patience dans le choix du film du soir et **Chouchou** pour avoir été ma petite taupe et pour avoir toujours supporté mes plans un peu foireux. Merci à **Stef** pour toutes nos discussions depuis tant d'années, et aussi d'être venue avec Tatien pendant la rédaction. Merci à **Nico** pour tous nos duos de folie, vive les hippocampes ! Merci à **Clara** d'être venue à Angers avec Chouchou me faire à manger. Merci à **Malik**, promis, on pourra faire plein de jeux tous ensembles maintenant. Enfin, Merci aux trois p'tits loulous pour toute la joie que vous m'apportez.

Merci à **Bastien**, pour ton amour, ta présence et ta patience (il en faut !). Merci pour tous ces weekends où j'ai laissé la thèse de côté grâce à toi. Ce sera tellement plus simple maintenant, on oublie la distance et en route pour la vie à deux ! J'ai hâte de vivre cette nouvelle aventure à tes côtés.

Sommaire

Liste de figures **2**

Liste de tableaux **6**

Synthèse bibliographique **9**

I. Emergence des maladies infectieuses **9**

1. Définition 9
2. Scénarios d'émergence 10
    a. Mécanismes moléculaires conduisant à l'émergence d'un agent pathogène 10
    b. Scénarios d'émergence de maladie 13
3. Facteurs favorisant l'émergence de maladies 15
    a. Les forces évolutives 16
    b. Facteurs environnementaux biotiques et abiotiques 17
4. Rôle des réservoirs environnementaux dans l'émergence de maladie . . . 19
5. Emergence de phytobactérioses dans les agroécosystèmes 21

II. Mise en place de l'interaction hôte-pathogène chez les bactéries phytopathogènes **23**

1. La phase d'attraction 24
    a. Le chimiotactisme et la cascade de transduction du signal activent le flagelle 24
    b. Perception de la bactérie par la plante et stratégie de contournement 26
2. La phase d'installation 27
3. La phase de multiplication 28
    a. Le système de sécrétion de type 3 29
    b. Les effecteurs de type 3, au cœur de l'interaction hôte-pathogène . 30

c. Défenses de la plante pendant la phase de multiplication : Le modèle en zigzag 31

III. Le genre *Xanthomonas* **32**
1. Morphologie, biochimie et physiologie 32
2. Taxonomie et phylogénie du genre 32
3. Caractéristiques génomiques des *Xanthomonas* 36

IV. *Xanthomonas arboricola* et ses hôtes **38**
1. *Xanthomonas arboricola* pv. *pruni* 38
2. *Xanthomonas arboricola* pv. *juglandis* 39
3. *Xanthomonas arboricola* pv. *corylina* 40
4. Les pathovars *arracaciae*, *guizotiae* et *zantedeschiae* 41
5. *Xanthomonas arboricola* pv. *populi* 41
6. *Xanthomonas arboricola* pv. *celebensis* 42
7. *Xanthomonas arboricola* pv. *fragariae* 42
8. Les souches pathogènes n'appartenant pas à un pathovar 42
9. Les souches commensales 43

V. Questions de recherche **44**

Chapitre I **48**

Introduction **48**

Recombination-prone bacterial strains form a reservoir from which epidemic clones emerge in agroecosystems **50**
Abstract 50
I. Introduction 51
II. Results 53
III. Discussion 57
IV. Conclusion 59
V. Experimental procedures 60
Acknowledgments 65

Supplemental materials **66**

Chapitre II **68**

Introduction **68**

Persistence of an asymmetric gene flow from pathogenic strains to commensal ones along divergence of epidemic clones and recombinogenic background. **70**
I. Introduction . . . 70
II. Materials and methods . . . 73
III. Results . . . 76
IV. Discussion . . . 79

Datation phylogénétique de l'émergence des trois clones épidémiques chez *X. arboricola* **83**
I. Introduction . . . 83
II. Matériel et Méthodes . . . 84
III. Résultats et Discussion . . . 85
IV. Perspectives . . . 87

Supplemental materials **88**

Chapitre III **90**

Etude phénotypique des souches commensales et pathogènes appartenant à l'espèce *Xanthomonas arboricola* **90**
I. Introduction . . . 90
II. Matériel et Méthodes . . . 92
III. Résultats . . . 96
IV. Discussion . . . 98

Données supplémentaires **103**

Chapitre IV **105**

Introduction **105**

Evolution of the type three secretion system and type three effector repertoires in *Xanthomonas* **107**
Abstract . . . 107
I. Introduction . . . 108
II. Results . . . 111
III. Discussion . . . 118
IV. Materials and Methods . . . 123

Acknowledgments 125

Supplemental materials **126**

Discussion générale **128**

I. Le scénario évolutif de l'espèce *X. arboricola* 128

1. Les souches commensales dans l'émergence de nouveaux agents pathogènes 128

2. L'émergence des trois pathovars performants 130

3. L'émergence du pathovar *guizotiae* 132

II. L'écologie des souches commensales 133

III. Apport de la génomique 134

IV. Conclusion et Perspectives 136

Perspectives permettant d'améliorer nos connaissances de l'émergence de maladie 136

Perspectives permettant d'améliorer nos connaissances sur l'écologie des souches commensales 137

Références **141**

Liste de figures

Synthèse bibliographique

Figure I-1 : Les différents mécanismes de modification de l'ADN.

Figure I-2 : Croissance des souches de *Salmonella enterica* pré-émergentes et sauvages.

Figure I-3 : Résultats des études de phénotypes améliorés avec la présence du plasmide pESI.

Figure I-4 : Vue schématique des scénarios d'émergences de maladies.

Figure I-5 : Schéma des nombreux facteurs favorisant l'émergence de maladies infectieuses.

Figure I-6 : Représentation des effets des différents types de sélection sur le polymorphisme.

Figure I-7 : Effet d'un goulot d'étranglement et de la dérive génétique sur la diversité allélique.

Figure I-8 : Représentation des principales différences entre les milieux naturels et les agroécosystèmes.

Figure I-9 : Cartographie du plasmide p-PATH de *Pantoea agglomerans*.

Figure I-10 : Représentation du cycle infectieux des bactéries épiphytes pathogènes sur plantes.

Figure I-11 : Cascades de phosphorylation et de déphosphorylation des protéines kinases du chimiotactisme chez *E. coli*.

Figure I-12 : Réponse de la souche K60 appartenant à l'espèce *Ralstonia solanacearum*, aux composés issus de la tomate et du riz.

Figure I-13 : Représentation des séquences protéiques du domaine cytoplasmique conservé entre les 7 familles de MCP.

Figure I-14 : Représentation de la distribution de chaque protéine du flagelle.

Figure I-15 : Description de la PTI.

Figure I-16 : Représentation de la stratégie mise en place par l'agent pathogène *Pseudomonas aeruginosa* pour échapper à la reconnaissance via la flagelline.

Figure I-17 : Comparaison des structures du système de sécrétion de type 3 et du flagelle.

Figure I-18 : Représentation des deux hypothèses des relations évolutives entre le flagelle et le système de sécrétion de type 3.

Figure I-19 : Modèle d'évolution et de diversification du système de sécrétion de type 3 proposé par Abby & Rocha, 2012.

Figure I-20 : Description du mécanisme de réarrangement terminal, à l'origine de nouveaux effecteurs de type 3.

Figure I-21 : Le modèle gène pour gène de Flor 1971.

Figure I-22 : Modèle en zigzag de l'interaction plante-agent pathogène.

Figure I-23 : Reclassification des espèces au sein du genre *Xanthomonas*.

Figure I-24 : Représentations des relations phylogénétiques entre les différentes espèces du genre *Xanthomonas*.

Figure I-25: Schéma représentant la délétion des gènes dans le cluster flagellaire.

Figure I-26 : Comparaison du cluster de gènes du système de sécrétion de type 3.

Figure I-27 : Inoculation de souches appartenant à l'espèce *X. arboricola* sur un milieu succinate-quinate.

Figure I-28 : Photos de symptômes du chancre vertical suintant sur le noyer.

Figure I-29 : Photos de symptômes de la bactériose du noisetier.

Figure I-30 : Arbre phylogénétique des souches non pathogènes isolées de noyer.

Chapitre I

Figure II-1 : Genealogies, results of Bayesian clustering , and depiction of recombination events for the 187 strains in collection I.

Figure II-2 : Splitgraph of the concatenated dataset (*atpD*, *dnaK*, *efP*, *fuyA*, *glnA*, *gyrB*, and *rpoD*) for the 187 *X. arboricola* strains studied.

Figure II-3 : Minimum spanning tree created using the allelic profiles of the 187 strains.

Figure II-4 : Inference of evolutionary gains and losses of LHT-associated genes.

Figure II-5 : Phylogeny of the 187 strains in collection I generated using a concatenated dataset of partial sequences for seven housekeeping genes.

Figure II-6 : Maximum-likelihood phylogenies generated using the concatenated sequences of seven housekeeping genes and sequences for *xopR*.

Figure II-7 : Genomic environments of *xopR*.

Figure II-8 : Phylogeny of nonpathogenic strains used in this study and *Xanthomonas* type strains generated using a concatenated dataset of partial sequences for seven housekeeping genes.

Chapitre II

Figure III-1 : The seven tested model in δaδi.

Figure III-2 : Multidimensional scaling of ANI based distance matrix.

Figure III-3 : Principal component analysis based on core genome of 36 strains representing *X. arboricola* species.

Figure III-4 : Maximum likelihood phylogeny of 36 strains representing *X. arboricola* based on non-recombinant SNPs of the core genome.

Figure III-5 : Representation of gene flow between strains of groups A and B.

Figure III-6 : Distribution of derived alleles in group A and B of the best model IM2M.

Figure III-7 : ρ value for each SNP belonging to group A core genome.

Figure III-8 : Histogram representing the origins of recombinant sequences.

Figure III-9 : Representation of MGE repertoires (phages, plasmids and IS) in the 36 strains belonging to *X. arboricola*.

Figure III-S1 : Results obtained with δaδi for the 6 other models.

Résultats complémentaires

Figure III-10 : Régression linéaire des distances de chaque échantillon à la racine en fonction de la date d'isolement, pour le jeu de données complet.

Figure III-11 : Régression linéaire des distances de chaque échantillon à la racine, en fonction de la date d'isolement pour les 4 souches du pathovar *pruni*.

Figure III-12 : Phylogénie basée sur les SNP non recombinants des 4 souches du pathovar *pruni*.

Figure III-13 : Taux de mutations estimés pour les agents pathogènes humains.

Figure III-14 : Comparaison des taux de substitutions.

Figure III-15 : Phylogénie basée sur les SNP non recombinants des 10 souches appartenant aux trois pathovars *pruni*, *juglandis* et *corylina*.

Chapitre III

Figure IV-1 : Représentation des phénotypes des souches en fonction de leur positionnement phylogénétique.

Figure IV-2 : Inférence de gain et de perte de MCP, TBDT et adhésines au cours de l'histoire évolutive de l'espèce *X. arboricola*.

Figure IV-3 : Résultats des analyses d'antibiose suite à la confrontation d'une souche pathogène et une souche commensale.

Figure IV-4 : Dynamiques des tailles de populations sur haricot de souches commensales et pathogènes isolées de haricot.

Chapitre IV

Figure V-1 : Maximum likelihood phylogeny based on the concatenated sequences of the core proteome

Figure V-2 : Comparison of organism and T3SS phylogenies and schematic representation of T3SS genomic environments.

Figure V-3 : Representation of recombination events affecting 16 *hrc/hrp* genes of the T3SS cluster in *Xanthomonas* genus.

Figure V-4 : Representation of ratios of recombination rate vs mutation rate (ρ/Θ) along the T3SS cluster using 61 genomes representing the whole genus diversity.

Figure V-5 : Representation of T3E repertoires in *X. arboricola* strains.

Figure V-6 : Graphical circular representation of the draft genomes of *X. arboricola* strains belonging to the three successful pathovars.

Figure V-S1 : Genomic environments of T3SS clusters

Figure V-S2 : Genomic environments of a large region of T3SS cluster

Figure V-S3 : individual phylogenies of T3SS coding genes

Figure V-S4 : Genomic environments of *avrBs2*

Figure V-S5 : Genomic environments of 10 T3Es acquired in *X. arboricola* group A strains

Discussion générale

Figure VI-1 : Modèle évolutif retraçant l'histoire évolutive de l'espèce *X. arboricola*.

Figure VI-2 : Résultats des tests de pouvoir pathogène sur piment et tomate.

Liste de tableaux

Chapitre I

Table II-1 : Estimates of recombination and mutation.

Table II-S1 : Strains used in this study and their characteristics.

Table II-S2 : MLST allelic profiles.

Table II-S3 : Estimates of diversity and neutrality for each group and each locus.

Table II-S4 : PCR results for the presence (in black) and absence (in white) of T3SS and T3E genes.

Table II-S5 : PCR results for the presence (in black) and absence (in white) of MCP, TBDT, and adhesin genes.

Table II-S6 : PCR primers used to amplify the T3SS and T3E genes.

Table II-S7 : PCR primers used to amplify the MCP, TBDT, and adhesin genes.

Table II-S8 : Primers used to sequence *xopR*.

Chapitre II

Table III-1 : likelihood obtained for the seven tested models.

Table III-2 : Mean number and variance of MGE (phages, plasmids and IS) found in 36 strains of *X. arboricola*.

Table III-S1 : Collection used for chapter 3.

Table III-S2 : ANI values for each pairwise genome comparisons of 44 genomes.

Chapitre III

Table IV-S1 : Collection de souches

Chapitre IV

Table V-S1 : Collection of genomes

Table V-S2 : T3SS coding genes used as query in Blast searches

Table V-S3 : Results of SH tests comparing phylogenies of T3SS coding genes in *Xanthomonas*

Table V-S4 : T3E Repertoires in *X. arboricola*

Synthèse bibliographique

I. Emergence des maladies infectieuses

1. Définition

Les émergences de nouvelles maladies infectieuses concernent tous les environnements et une grande majorité d'espèces vivantes. Elles posent des problèmes en santé humaine, dans le domaine vétérinaire, en conservation de la biodiversité et au sein des agroécosystèmes. L'importance d'étudier ces émergences semble donc cruciale, mais pour cela, il est nécessaire de bien définir ce que l'on entend par maladie émergente. Plusieurs définitions ont été proposées ; la définition la plus répandue (Morens et al. 2004 ; Bartoli et al. 2016) correspond à celle proposée par Morse en 1995, selon laquelle une maladie est émergente si elle apparait dans une nouvelle population, ou si elle existait déjà mais que l'on observe une augmentation de son incidence ou de son aire géographique. Cependant, en 2001, Woolhouse distingue l'augmentation de l'incidence qui résulte de changements profonds de l'agent pathogène et qui entraine des conséquences mondiales de celle ayant un effet local et à court terme. L'auteur reconnait que cette notion est très subjective. Morens et al. 2004 distingue, quant à lui, trois types d'émergence de maladie : les maladies nouvellement émergentes, les réémergences ou encore les émergences délibérées (actes malveillants). Anderson en 2004, fait apparaitre une dernière notion, qui correspond à la découverte de la maladie émergente ou le fait de reconnaitre une maladie comme telle. En effet, cette notion est importante puisqu'elle prend en compte le fait que certaines émergences peuvent n'être découvertes que fortuitement ou tardivement si elles ne posent pas de problème sanitaire, si l'espèce touchée par la maladie ne présente pas un intérêt particulier, que ce soit économique ou en biologie de la conservation, ou bien si les moyens techniques ne permettent pas de révéler le caractère infectieux de la maladie (Schrag & Wiener 1995). On peut, par exemple, citer le cas de l'identification par Warren en 1983 de l'agent pathogène responsable de l'ulcère gastrique, *Helicobacter pylori*. Cette maladie existait déjà mais n'était pas connue pour être causée par un agent pathogène bactérien. L'identification de cet agent pathogène dans la muqueuse gastrique a permis de mettre en évidence que l'ulcère gastrique était en réalité une maladie infectieuse. Cette découverte est également intéressante puisque l'analyse de génétique des populations de cet agent pathogène a été faite et utilisée pour tracer les migrations de son hôte, l'homme, dans différentes aires géographiques (Falush et al., 2003 ; Yamaoka, 2009).

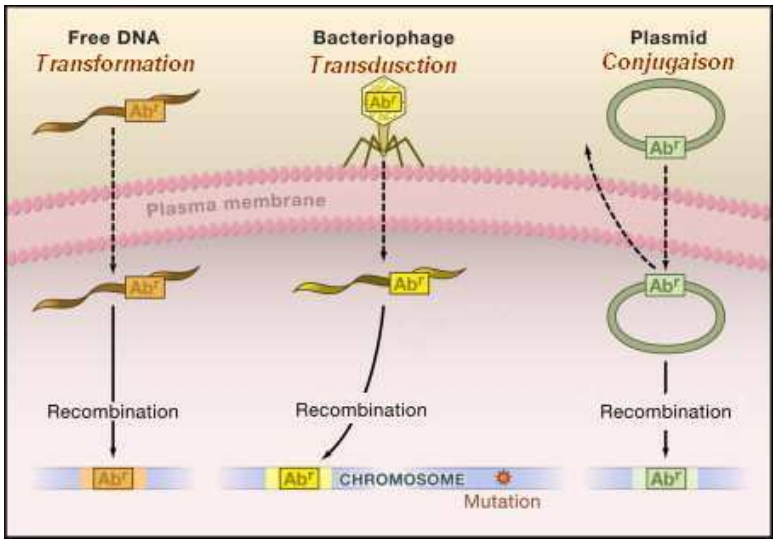


Figure I-1 : Les différents mécanismes de modification de l'ADN. En partie haute de la figure sont représentées les différentes voies d'acquisition d'ADN exogène par une bactérie. De gauche à droite : la transformation correspond à l'acquisition de fragments d'ADN libre dans le milieu par des bactéries compétentes, la transduction correspond à l'acquisition d'ADN par l'intermédiaire d'un bactériophage lytique ou tempéré et la conjugaison fait intervenir un plasmide conjugatif. Un fragment d'ADN étranger ayant ainsi pénétré dans la cellule bactérienne peut être incorporé dans le génome par recombinaison. Enfin, la mutation d'une ou plusieurs bases sur le chromosome est la dernière voie de modification de l'ADN (d'après Anlekshun & Levy, 2007).

2. Scénarios d'émergence

a. Mécanismes moléculaires conduisant à l'émergence d'un agent pathogène

L'émergence d'une maladie implique tout d'abord l'émergence d'un nouvel agent pathogène, et ensuite que cet agent pathogène se disperse et rencontre les conditions favorables à sa multiplication. L'émergence d'un nouvel agent pathogène est consécutive à trois types de mécanismes moléculaires : la recombinaison homologue, le transfert horizontal de gènes (THG), et la mutation (Bartoli et al. 2016). Ces trois mécanismes peuvent avoir différentes conséquences au sein des génomes bactériens. En effet, ils peuvent conduire à de nouveaux allèles (recombinaison et mutation), à l'acquisition de nouveaux gènes (THG), à des réarrangements génomiques (recombinaison) qui peuvent entrainer la perte ou la modification de l'expression d'un gène, ou encore, à l'inactivation d'un ou de plusieurs gènes (mutation, recombinaison).

La recombinaison homologue

La recombinaison homologue peut conduire à un changement allélique. Elle peut s'apparenter à l'effet de la multiplication sexuée des eucaryotes quand elle a lieu entre deux cellules bactériennes (Vos & Didelot, 2009). En effet, elle peut avoir lieu suite à l'intégration de séquences d'ADN exogènes par transformation, par transduction et par conjugaison (fig. I-1). Une fois que l'ADN est incorporé dans la cellule bactérienne, il peut recombiner avec l'ADN génomique en autorisant un certain taux de mésappariement. Ainsi, une nouvelle séquence allélique peut remplacer celle déjà présente, on parle alors de recombinaison efficace si cette recombinaison a lieu entre deux séquences différentes (Didelot & Maiden, 2010). Si cet allèle confère un avantage dans la pathogénie, ou s'il remplace un allèle qui était non fonctionnel, cela peut entrainer l'apparition d'un nouvel agent pathogène. Le cas de *Neisseria meningitidis* est particulièrement intéressant pour étudier l'émergence d'agents pathogènes causée par la recombinaison homologue. En effet, cette bactérie pathogène de l'homme, responsable de la méningite, est connue pour avoir une structure de population épidémique, telle que définie par Maynard-Smith et al. (1993), dans laquelle la recombinaison est la force évolutive la plus importante. Cette espèce est subdivisée en 5 sérogroupes nommés A, B, C, W et Y. Dans les années 2000, un clone très agressif[1] a émergé et a causé des épidémies dans de nombreux pays, il s'agit

[1]Dans ce manuscrit, le terme agressivité est utilisé au sens phytopathologique proposé par Vanderplanck (1963). C'est donc un caractère quantitatif, contrairement à la virulence qui lui est un caractère qualitatif.

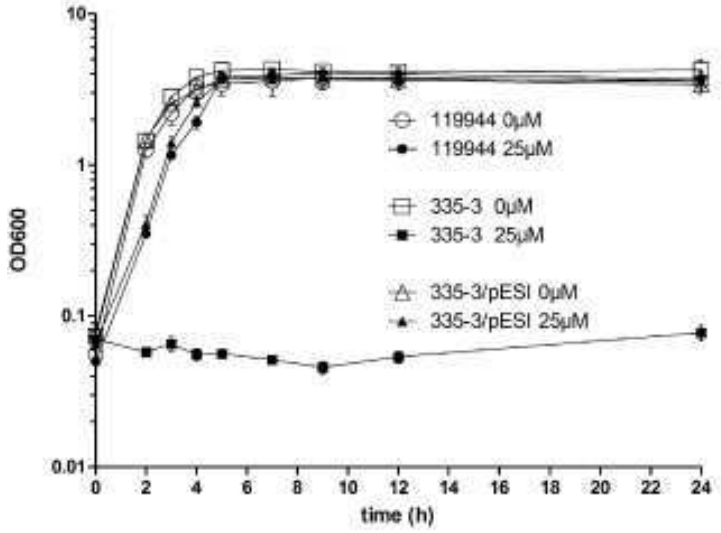


Figure I-2 : Croissance des souches de *Salmonella enterica* pré-émergentes (qui diffèrent des souches émergentes par l'absence du plasmide) sauvage (335-3), transformée avec le plasmide pESI résistante à l'azide (335-3/pESI) et émergente (119944) sur milieu minimum en présence ou absence de 25 µM de $HgCl_2$. Chaque point représente la moyenne de densité optique (OD600) mesurée à différentes heures après inoculation (abscisse) (Aviv et al. 2014).

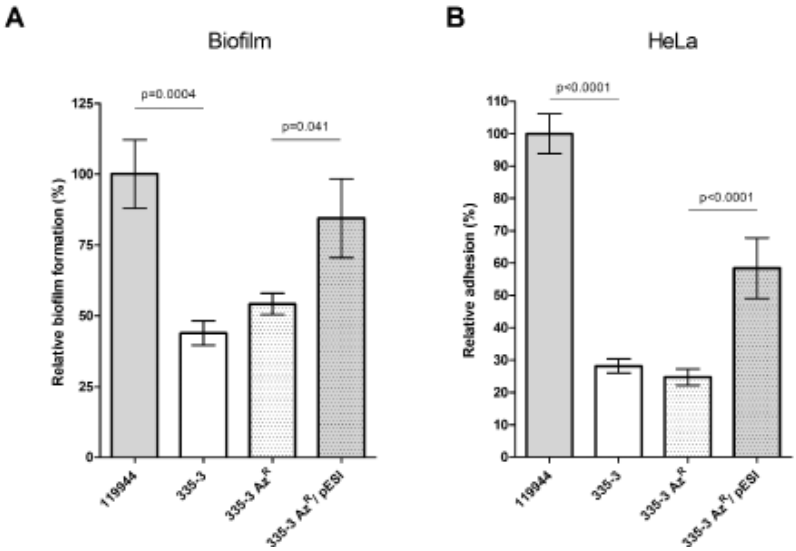


Figure I-3 : Résultats des études de phénotypes améliorés avec la présence du plasmide pESI. La souche émergente de *S. enterica* est 119944, 335-3 correspond à la souche pré-émergente sauvage, 335-3 Az^R correspond à la souche résistante à l'azide et la souche 335-3 Az^R/pESI correspond à la souche transformée avec le plasmide pESI.

A : Formation de biofilm par les souches après 48h de croissance dans des tubes recouverts de cholestérol

B : Adhésion (HeLa) des souches sur cellules épithéliales humaines

du clone Hajj qui appartient au sérogroupe W. Une étude a permis d'identifier que ce clone Hajj présente des allèles recombinants aux locus impliqués dans l'agressivité qui ne sont pas retrouvés dans le sérogroupe d'origine (Mustapha et al. 2015).

La recombinaison homologue peut également entrainer des réarrangements génomiques, puisqu'elle peut avoir lieu au sein du génome d'une même cellule bactérienne. Cela peut alors modifier l'expression de certains gènes s'ils sont situés à proximité d'un promoteur plus ou moins fort. Cette modification peut également affecter un ou plusieurs gènes impliqués dans la virulence et conduire à un nouvel agent pathogène. Par exemple, un réarrangement chromosomique a été mis en évidence chez *Streptococcus pyogenes* qui affectait l'expression du gène *slo* impliqué dans la virulence de cet agent pathogène (Savic & Ferretti, 2003). Les séquences d'insertion (IS) peuvent jouer un rôle dans les réarrangements génomiques. En effet, lors de la recombinaison homologue entre deux IS, celles-ci peuvent alors emporter avec elles des gènes qui sont proches et être responsables de réarrangements génomiques (Bennett et al. 2004).

Le transfert horizontal de matériel génétique

Le transfert horizontal de gènes (ou recombinaison hétérologue) peut permettre l'acquisition de nouveaux gènes. En effet, ce mécanisme, bien plus fréquent chez les procaryotes que chez les eucaryotes, permet directement l'échange de gènes entre deux cellules, sans transmission verticale. Ainsi, une cellule peut acquérir de nouveaux traits au cours de sa vie. Les transferts horizontaux peuvent avoir lieu par conjugaison, transformation ou transduction (fig. I-1). Ils peuvent faire intervenir des éléments génétiques mobiles, tels que les plasmides, les bactériophages ou encore les séquences d'insertion. Des gènes codants pour des facteurs de virulence peuvent être associés à ces éléments génétiques mobiles et être acquis par la cellule receveuse (Friesen et al. 2006 ; Bartoli et al. 2016). Le changement de phénotype est donc immédiat, et il est possible qu'une souche non-pathogène devienne pathogène. Par exemple, la comparaison des génomes et des phénotypes associés aux souches émergentes et pré-émergentes de *Salmonella enterica* serovar *infantis* a montré que cette émergence était probablement due à l'acquisition d'un mégaplasmide, pESI. Ce plasmide confère la résistance à 3 antibiotiques et contient un opéron de résistance au mercure (fig. I-2) (Aviv et al. 2013). Les auteurs ont, par ailleurs, montré que l'incorporation de ce plasmide dans une souche pré-émergente, isolée avant l'épidémie et qui diffère des souches émergentes par l'absence du plasmide, améliorait la formation de biofilm, l'adhésion (fig. I-3) et l'invasion de l'hôte.

Le transfert horizontal de matériel génétique peut entrainer l'inactivation d'un gène. En effet, l'intégration des IS dans le génome peut conduire à la pseudo-

genéisation, si le fragment acquis s'intègre au sein d'un gène. Une étude a été réalisée chez *Escherichia coli*, afin de tester la capacité d'adaptation des souches commensales face au système immunitaire humain. Dans cette étude, *in vitro*, il a été montré que le même phénotype apparait dans plusieurs populations, lorsque les bactéries se retrouvent en présence de la pression de sélection exercée par le système immunitaire (Miskinyte et al. 2013). Ce phénotype est dû à l'insertion d'une transposase dans le promoteur du gène *yrfF*, sa fonction n'est pas connue chez *E. coli*, mais son homologue chez *Shigella* régule la synthèse de la capsule.

La mutation

Les mutations modifient la séquence allélique d'un gène et peuvent affecter des gènes directement impliqués dans l'expression du pouvoir pathogène. Elles peuvent également affecter d'autres gènes, dont les gènes d'antivirulence[2], ce qui peut conduire à une perte de fonction réprimant le pouvoir pathogène de la cellule bactérienne (Maurelli, 2007). Ces mutations peuvent être de différents types : insertion, délétion ou changement d'un ou de plusieurs nucléotides. Ainsi, le rôle de la mutation adaptative a été démontré dans le cas des émergences de la résistance aux antibiotiques, chez les agents pathogènes humains (Woodford & Ellington, 2007). L'impact des mutations comme mécanisme de pathoadaptation a été initialement défini par Sokurenko et al. (1998). En effet, il a été montré chez *E. coli* que des événements de mutation du gène *fimH*, qui code une adhésine, pouvait modifier l'écologie des souches commensales, qui deviennent des agents pathogènes (Sokurenko et al. 1998).

Régulation de ces mécanismes

La recombinaison homologue et la mutation sont régulées par un système moléculaire, appelé " réparation des mésappariements " ou MMR (pour mismatch repair) (Shofield & Hsieh, 2003). Le rôle de ce système est de conserver l'intégrité du génome propre à chaque individu en corrigeant les mésappariements, et en bloquant la recombinaison homologue avec des séquences divergentes. Ainsi, seules les séquences peu divergentes pourront être incorporées au génome de la cellule bactérienne lors de la recombinaison homologue. C'est pourquoi, la recombinaison homologue se réalise de manière plus efficace lorsque les individus sont proches génétiquement, c'est-à-dire entre individus appartenant à la même espèce ou à des espèces voisines (Boto, 2010).

Cependant, il a été montré que certaines souches appelées " mutators " pou-

[2]Un gène d'antivirulence correspond à un gène dont l'expression dans l'agent pathogène est incompatible avec la virulence de l'agent pathogène (Maurelli et al. 2007)

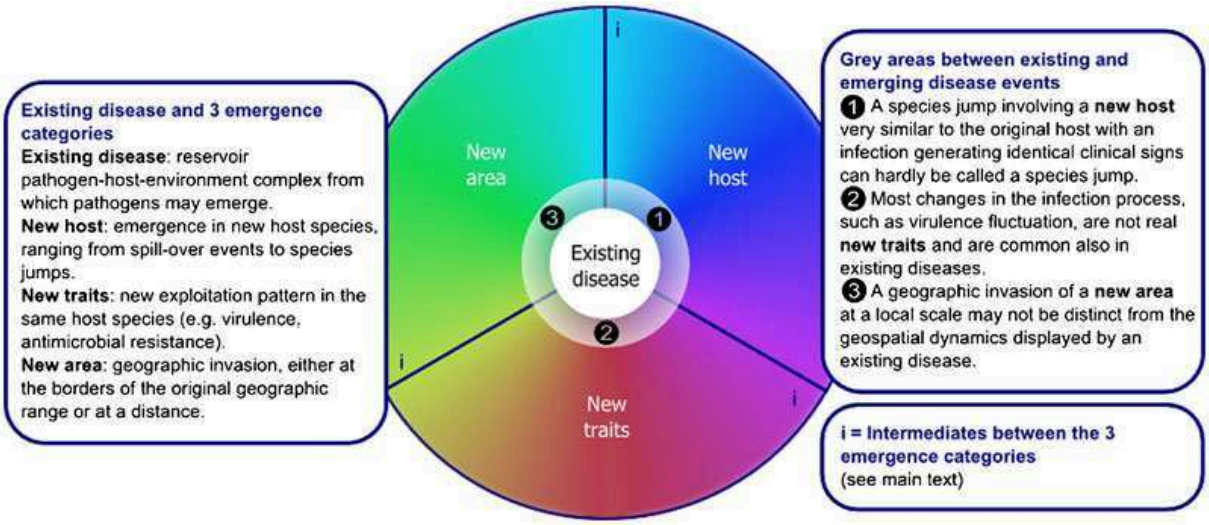


Figure I-4 : Vue schématique des scénarios d'émergences de maladies à partir d'agents pathogènes existants. Ces scénarios peuvent être liés les uns aux autres, cela est représenté par l'ensemble des couleurs de l'arc-en-ciel qui représentent toutes les possibilités. Trois catégories de scénarios sont proposées : émergences dans une nouvelle aire, sur un nouvel hôte ou avec un nouveau trait (Engering et al. 2013).

vaient avoir un taux de mutation plus élevé qu'attendu ou pouvaient incorporer des séquences d'ADN très divergentes par recombinaison homologue (Denamur & Matic, 2006). En effet, un stress peut induire la " réponse SOS " chez la bactérie ce qui va entrainer un relâchement du système de réparation des mésappariements et donc, un état " mutator " transitoire (McKenzie et al. 2001). Parmi ces stress favorisant l'apparition des mutations, on peut retrouver des variations de températures (faibles ou élevées), le manque de ressource nutritive, l'oxydation ou l'irradiation aux UV (Hengge-Aronis, 2002 ; Tenaillon et al. 2004). Il a été montré que chez *E. coli*, cette réponse pouvait avoir un rôle dans l'acquisition de la résistance aux antibiotiques (Handel et al. 2016). La capacité d'adaptation des bactéries dépend donc principalement de leur capacité à déclencher cette réponse SOS.

b. Scénarios d'émergence de maladie

Une fois que l'agent pathogène a émergé, son évolution, ainsi que celle de son environnement peuvent conduire à l'émergence d'une nouvelle maladie, telle que définie précédemment. Plusieurs conditions sont nécessaires pour qu'une maladie se déclenche. Tout d'abord, une condition qui peut sembler triviale est que l'agent pathogène et son hôte doivent être en sympatrie, c'est-à-dire qu'ils doivent pouvoir entrer en contact, et ensuite que l'agent pathogène doit être adapté à son hôte, c'est-à-dire capable de se multiplier, de provoquer des symptômes et de se disséminer (Woolhouse et al. 2005). Différents scénarios ont donc été proposés pour expliquer ces émergences. Le premier correspond à la colonisation d'une nouvelle aire géographique, le second à la colonisation d'un nouvel hôte et le dernier à l'acquisition d'un nouveau trait (fig. I-4) (Engering et al. 2013).

La colonisation d'une nouvelle aire géographique suggère que l'agent pathogène n'a pas subi de changement autre que son aire de répartition. Cette colonisation peut résulter de grands événements de migration de l'hôte. En effet, lors d'événements de migration ou d'introduction de nouvelles espèces dans un milieu, l'agent pathogène peut suivre son hôte. C'est le cas d'*H. pylori*, agent d'ulcères gastriques chez l'homme mentionné précédemment. Par ailleurs, la comparaison de la phylogéographie de *Mycobacterium tuberculosis*, agent pathogène de la tuberculose et la phylogéographie de l'homme a permis de montrer que cet agent pathogène infectait déjà les premières populations humaines il y a plusieurs dizaines de milliers d'années, en Afrique de l'Est. Cet agent pathogène s'est ensuite dispersé avec son hôte en Afrique et hors d'Afrique. De plus, une expansion majeure de la tuberculose coïncide avec l'augmentation de la taille de population humaine et la densité humaine au néolithique (Comas et al. 2013). Plus récemment, une étude phylogénétique a permis de déterminer l'événement fondateur de l'épidémie de VIH

(virus de l'immunodéficience humaine) au Groenland, qui aurait émergée en 1986, suite à la migration d'un homme contaminé provenant du Danemark (Bruhn et al. 2014). Les scénarios de colonisation d'une nouvelle aire géographique suggèrent que les agents pathogènes doivent se disséminer. Cela est donc principalement dépendant de la capacité de dispersion de l'agent pathogène. Outre la dispersion via leur hôte, certains agents pathogènes peuvent également être véhiculés par des vecteurs. Par exemple les agents phytopathogènes des genres *Liberibacter*, *Phytoplasma*, *Spiroplasma* et *Xylella*, ont pour vecteurs des hémiptères (Perilla-Henao & Casteel, 2016). Les agents pathogènes peuvent également être dispersés par les éléments naturels tels que l'eau, le vent et les courants marins. Pour de nombreux agents phytopathogènes, la dissémination à longue distance et donc la colonisation d'une nouvelle aire résulte avant tout de facteurs anthropiques tels que l'introduction d'un hôte avec son cortège microbien dans une nouvelle aire géographique. Ce point sera détaillé plus loin.

La colonisation d'un nouvel hôte dépend de l'écologie de l'agent pathogène, de sa spécialisation à son hôte (plus il est spécialisé, plus il est difficile de coloniser un nouvel hôte), du degré d'association écologique entre l'agent pathogène et le nouvel hôte, et enfin de la distance génétique entre l'hôte d'origine et le nouvel hôte (Parker & Gilbert, 2004). Cette distance génétique permet de distinguer deux scénarios : la colonisation d'un hôte phylogénétiquement proche de l'hôte initial, on parle alors d' " host shift " et la colonisation d'un hôte phylogénétiquement très éloigné, on parle d' " host jump " (Stuckenbrock & McDonald, 2008 ; Kirzinger & Stavrinides, 2012). L' " host shift " peut se réaliser dans le cas d'un agent pathogène pré-adapté à son nouvel hôte (Parker et al. 2004). En effet, un agent pathogène peut posséder la capacité à coloniser un hôte et à exprimer une maladie sur celui-ci, sans pour autant avoir été en contact avec. Et, si l'hôte et l'agent pathogène entrent en contact, alors ce dernier aura la capacité de provoquer une maladie. Cela peut être illustré, par exemple, par l'histoire évolutive de l'agent pathogène du pommier, *Venturia inaequalis*. Deux types de populations existent chez cet agent pathogène, une virulente sur les pommiers domestiques (*Malus x domestica*) porteur du gène de résistance *vf* et une autre, avirulente sur ces pommiers. L'introgression du gène de résistance dans les pommiers domestiques a été faite depuis le pommier ornemental, *M. floribunda* et l'inférence de scénarios évolutifs a montré que la population virulente de *V. inaequalis* correspondait à celle existant déjà sur le pommier ornemental (Lemaire et al. 2016). Ainsi, l'agent pathogène a migré sur le pommier domestique suite à la modification de ce dernier. Le " host jump ", quant à lui, peut être illustré par l'histoire évolutive de *Staphylococcus aureus*. En effet, au cours de l'histoire évolutive de cet agent pathogène, il y a eu, au moins, 5 événements de colonisation d'un nouvel hôte de type " host jump ", de l'homme vers les bovins (Weinert et al. 2012). Ces

évènements auraient eu lieu lors de la domestication des bovins, suggérant un lien étroit entre l'homme et ces animaux qui est considéré par les auteurs comme le principal facteur dirigeant ces sauts d'hôte. En effet, le premier évènement aurait eu lieu il y a 5512 ans, correspondant à l'expansion de l'agriculture au néolithique. Les autres événements auraient eu lieu il y a 3100 ans, 1800 ans, 1130 ans et 83 ans. Le " host jump " peut être associé à des changements profonds de la spécificité d'hôte induits par des mécanismes moléculaires, tels que la recombinaison, la mutation ou le transfert de gènes (Kirzinger & Stavrinides, 2012). Ainsi, il est associé au scénario de colonisation d'un nouvel hôte, mais également à l'acquisition d'un nouveau trait, cela illustre le fait que ces scénarios ne sont pas toujours indépendants les uns des autres (fig. I-4) (Engering et al. 2013).

L'acquisition d'un nouveau trait par l'agent pathogène peut entrainer la colonisation d'un nouvel hôte, mais il peut également permettre à un agent pathogène de coloniser de nouvelles niches écologiques sur son hôte initial ou d'améliorer sa fitness sur ce dernier. En effet, il peut permettre à l'agent pathogène de contourner les défenses de l'hôte, d'avoir accès à de nouveaux nutriments, ou d'améliorer sa compétitivité vis-à-vis des autres micro-organismes de l'environnement. Par exemple, l'acquisition de la résistance aux antibiotiques peut jouer un rôle en milieu naturel dans la compétition avec les autres micro-organismes (Flemming et al. 2016) ou, dans le cas d'un hôte traité aux antibiotiques. Cette acquisition de résistance chez les agents pathogènes de l'homme conduisant à l'émergence d'un clone résistant a largement été démontrée ces dernières années (Wellington et al. 2013 ; Chang et al. 2015). Par exemple, dans les années 1990, le taux d'éradication des infections à *Helicobacter pylori* était d'environ 80 %, ce taux a diminué avec l'acquisition de résistances aux antibiotiques, passant à 60 % dans certains pays (Thung et al. 2016).

3. Facteurs favorisant l'émergence de maladies

Bien évidemment, l'émergence d'une maladie est le plus souvent le résultat de la succession de plusieurs scénarios et plusieurs mécanismes. Par exemple, l'émergence de *Mycobacterium tuberculosis* à partir de l'ancêtre commun partagé avec *M. canetti* serait le résultat de la perte de fonction de gènes, de l'accumulation de facteurs de virulence par transfert horizontal de gènes, de la recombinaison homologue entre les différentes souches et enfin, de la fixation d'allèles soumis à la sélection (Supply et al. 2013). Il a également été montré chez *Staphylococcus aureus* qu'une mutation localisée dans le gène *dltB*, impliqué dans la régulation de la formation de biofilm, était responsable d'un changement d'hôte, de l'homme vers le lapin (Viana et al. 2015). Il existe différents facteurs (génétiques ou environnementaux) qui peuvent favoriser ces

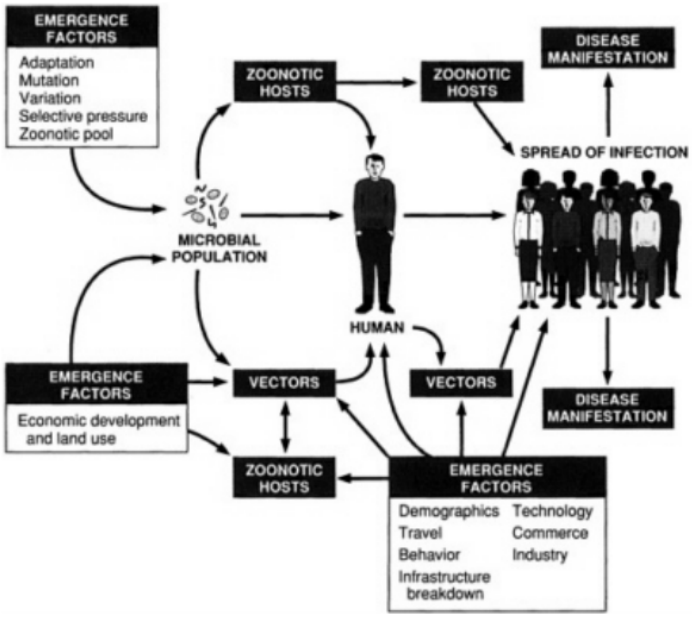


Figure I-5 : Schéma des nombreux facteurs favorisant l'émergence de maladies infectieuses chez l'homme (Lederberg & Shope, 2008).

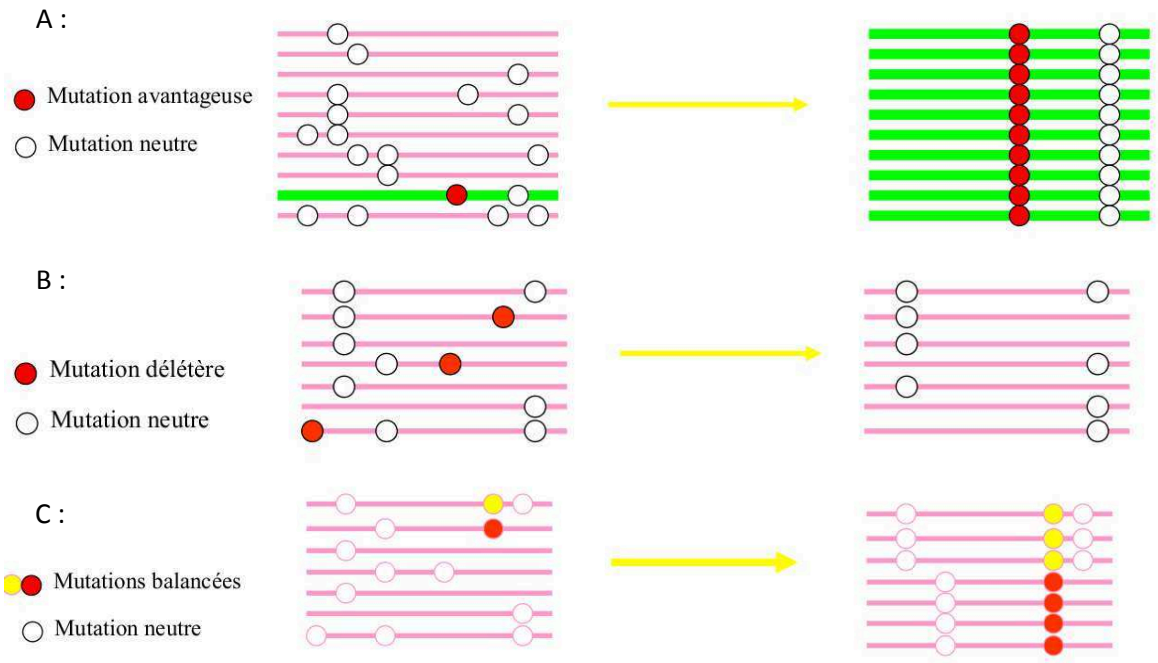


Figure I-6 : Représentation des effets des différents types de sélection sur le polymorphisme. A : cas d'une mutation avantageuse sous sélection positive. Après plusieurs générations, la mutation se fixe dans la population, et on observe un balayage sélectif des locus avoisinants. B : cas d'une mutation délétère sous sélection purifiante, cette mutation est éliminée au cours des générations. C : cas de mutations sous sélection balancée, le polymorphisme est maintenu dans la population au cours des générations (d'après Pauwells, comm. pers.).

émergences (fig. I-5). La probabilité qu'une émergence se produise dépend à la fois du caractère invasif de l'agent pathogène et de la structure de l'environnement qui va accueillir cette émergence (Engering et al. 2013). Ainsi, en connaissant l'intensité des différents facteurs qui s'exercent sur les agents pathogènes et sur leurs conditions environnementales, il est possible de retracer le scénario ayant conduit à leur émergence, ou d'évaluer les risques que peuvent représenter certaines populations de microorganismes pour de potentielles émergences (Lindeberg et al. 2008).

a. Les forces évolutives

Comme nous l'avons vu précédemment, l'adaptation à l'hôte fait partie des pré-requis nécessaires pour l'expression de la maladie. En effet, avec l'apparition de polymorphisme dans une population différentes forces évolutives s'exercent, et selon l'intensité de ces différentes forces, le maintien de ce polymorphisme peut varier dans les populations. Ainsi, soit le nouveau variant se fixe dans la population c'est–à-dire que tous les individus possèdent le même allèle, soit il disparait. Ces forces évolutives correspondent à la sélection et à la dérive génétique. Le mécanisme de pathoadaptation laisse supposer que l'hôte exerce une pression de sélection sur l'agent pathogène, conduisant ce dernier à s'adapter aux variations du milieu imposées par l'hôte. Cette force évolutive correspond à la sélection naturelle telle que définie par Darwin en 1859 qu'il distingue de la sélection imposée par l'homme.

La sélection

Différents types de sélection existent et ont un effet différent sur la fréquence des allèles au sein d'une population (fig. I-6). En effet, sous l'effet de la sélection positive, le polymorphisme au sein d'une population disparait rapidement et l'allèle avantageux pour les individus se fixe par balayage sélectif. Par exemple, une analyse de génomique des populations chez *Pseudomonas aeruginosa* a montré que l'un des clades appartenant à cette espèce, le clade B, avait émergé suite à un balayage sélectif, qui correspond à la propagation d'un nouveau clone (Caballero et al. 2015). La sélection purifiante, quant à elle va contre-sélectionner les allèles désavantageux dans une population. Le polymorphisme sera donc restreint, mais contrairement à la sélection positive, ce sera le résultat de la disparition des nouvelles mutations et non pas de leur fixation. Par exemple, chez *Helicobacter pylori*, deux gènes codant pour des adhésines (*flaA* et *flaB*) ainsi que *vacA*, qui code pour une cytotoxine seraient soumis à une forte sélection purifiante. Cela a, comme conséquence, de diminuer le polymorphisme de ces gènes comparativement à *E. coli* (Suerbaum et al. 1998). Enfin, le dernier type de sélection qui peut s'exercer est la sélection diversifiante

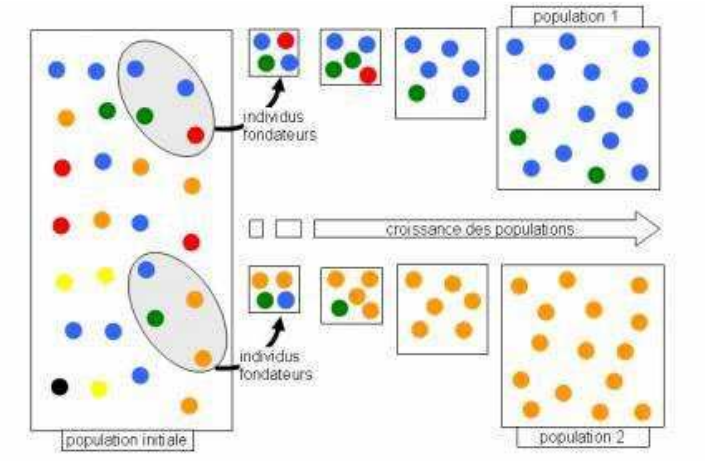


Figure I-7 : Effet d'un goulot d'étranglement et de la dérive génétique sur la diversité allélique d'une population. Le goulot d'étranglement diminue la diversité allélique au sein des populations, et la dérive génétique fixe ou élimine de manière aléatoire les allèles au cours des générations

(ou balancée). Cette sélection permet un maintien du polymorphisme dans une population. Ce maintien de polymorphisme dans la population pourrait conférer un avantage adaptatif face aux défenses de l'hôte. Par exemple, cette sélection peut agir sur certains effecteurs de type III de *Pseudomonas syringae*, protéines impliquées directement dans le dialogue moléculaire plante-agent pathogène (voir ci-dessous) (Rohmer et al. 2004).

La dérive génétique

La dérive génétique est une force évolutive dont les effets dépendent des événements démographiques que subit la population (fig. I-7). Plus la taille de la population est faible, plus l'effet de la dérive génétique est important. Elle peut, comme la sélection, fixer un variant dans une population ou le faire disparaitre, mais cela de manière aléatoire et en fonction de la fréquence initiale du variant dans la population. La population de l'agent pathogène peut subir un goulot d'étranglement lors de sa transmission à un nouvel hôte, ce qui réduit sa taille de population efficace et accentue l'effet de la dérive génétique (Didelot et al. 2016), cela peut conduire à la perte d'allèles favorables (Toft & Andersson, 2010). La taille de population du microorganisme pathogène peut aussi varier au cours de l'infection d'un individu, comme cela a été montré chez l'agent pathogène *Staphylococcus aureus* (Golubchik et al. 2013).

b. Facteurs environnementaux biotiques et abiotiques

Intensification des échanges commerciaux et domestication des hôtes

La principale cause de dispersion des agents pathogènes reste les activités anthropiques. En effet, l'homme ne se déplace pas seul, il déplace également les espèces végétales ou animales, qui peuvent être porteuses d'agents pathogènes. En particulier, les échanges de matériel végétal se sont largement intensifiés, et l'introduction de nouvelles espèces dans un nouvel environnement a favorisé, soit la colonisation d'une nouvelle aire géographique, soit la colonisation d'un nouvel hôte qui était jusqu'alors inconnu. Par exemple, en Avril 2012, des plants de caféiers provenant d'Amérique latine et cultivés en serre en France, ont été déclarés infectés par l'agent pathogène, *Xylella fastidiosa* (Jacques et al. 2016a). Cependant, cette importation de *Xylella fastidiosa* n'a pas entrainé d'émergence en France, puisque la bactérie a été détectée dans une zone confinée et les plants ont été détruits. Par contre, l'émergence de *X. arboricola* pv. *corylina* en France pourrait être dû à l'introduction de plants contaminés en provenance des Etats-Unis (Fischer-Le Saux et al. 2015).

En effet, les génotypes des souches retrouvées en France dans les années 1970 sont identiques à ceux retrouvés dans l'Oregon, là où a été identifiée la maladie pour la première fois (Barss, 1915 ; Miller et al. 1940). De plus, les plants utilisés pour implanter le verger dans lequel a été observée la maladie en France étaient originaires de l'Oregon (Luisetti et al. 1975).

La domestication peut également jouer un rôle dans l'émergence de nouvelles maladies. En effet, lors de la domestication, l'homme sélectionne une espèce et peut également sélectionner l'agent pathogène associé. On parle alors d' " host tracking ". Dans ce cas, le centre d'origine de l'agent pathogène est le même que celui de son hôte (Stukenbrock & McDonald 2008). Ce phénomène a été décrit chez plusieurs agents fongiques, comme *Phytophtora infestans* (Gomez-Alpizar et al. 2007) et *Venturia inaequalis* (Gladieux et al. 2008), pathogènes respectivement sur pomme de terre et sur pommier.

Les changements de l'environnement abiotique

Les conditions environnementales constituent un facteur important pour l'expression des maladies (Engering et al. 2013). En effet, les variations des conditions environnementales peuvent entrainer une modification d'expression d'un caractère déjà existant (Mekalanos, 1992). Ceci suggère qu'un agent pathogène ne peut exprimer une maladie uniquement lorsque les conditions sont favorables pour cela. Ainsi, le changement global observé actuellement pourrait favoriser l'émergence de maladies (Garrett et al. 2006). En effet, il peut affecter les précipitations et augmenter les variations de température. Par exemple, le phénomène océanique à grande échelle affectant périodiquement l'océan Pacifique, appelé El-Nino, entraine un courant côtier d'eau chaude vers l'Amérique du Sud. Ce courant pourrait être responsable de la dispersion de l'Asie vers l'Amérique du Sud des deux agents pathogènes *Vibrio cholerae* (ayant causée l'épidémie de choléra en 1991) et de *Vibrio parahaemolyticus* (responsable d'une épidémie en 1997) (Martinez-Urtaza et al. 2008). Ainsi, le changement global, et par conséquent, des courants marins pourraient avoir un rôle dans la dispersion de ces agents pathogènes (Martinez-Urtaza et al. 2016). Il peut également conduire à des conditions favorables pour l'émergence d'une maladie dont l'agent pathogène était pré-existant dans l'environnement, ou encore ces changements peuvent fournir des conditions favorables pour l'introduction d'un nouvel agent pathogène dans le milieu (Anderson et al. 2004). Ce changement environnemental peut également affecter la transmission des agents pathogènes. En effet, la perte de biodiversité causée par ce changement (Cardinale et al. 2012) peut favoriser la transmission des agents pathogènes suite à l'uniformisation des milieux (Altizer et al. 2013). Enfin, le changement climatique peut étendre la répartition

des insectes vecteurs. Il a été montré que l'augmentation des températures et de l'humidité relative jouait un rôle sur la dissémination du virus de la dengue dont le vecteur appartient au genre *Aedes* (Naish et al. 2014), ces conditions favorisant le développement du vecteur.

4. Rôle des réservoirs environnementaux dans l'émergence de maladie

Les réservoirs environnementaux désignent les environnements (en dehors de l'hôte) qui peuvent héberger l'agent pathogène et être sources d'inoculum (Vezzulli et al. 2010). Ainsi, l'agent pathogène peut migrer depuis cet environnement sur son hôte. En agriculture, ces réservoirs ont été principalement considérés dans l'émergence de maladies associées aux espèces animales, appelées zoonoses. On parle alors de spill-back[3] lorsqu'une maladie présente dans les milieux naturels émerge chez les espèces domestiquées (Daszak et al. 2000). Il est également connu que les plantes ornementales constituent un réservoir pour l'émergence de maladies associées aux plantes. En effet, dans les années 70, il a été montré que l'aubépine pouvait représenter un réservoir d'inoculum pour la bactérie *Erwinia amylovora* responsable du feu bactérien du pommier (Eden-green & Billing, 1972). Par ailleurs, les agro-ecosystèmes peuvent être des réservoirs pour des agents pathogènes de l'homme. Par exemple, en 1974, les plantes agricoles ont été reconnues comme étant des réservoirs pour *Pseudomonas aeruginosa*, agent pathogène de l'homme (Green et al. 1975).

L'ensemble de la communauté microbienne dans laquelle se développe un agent pathogène peut constituer le réservoir environnemental à partir duquel pourront émerger de nouveaux agents pathogènes par acquisition de nouveaux traits. L'importance de ces réservoirs dans l'émergence d'agents pathogènes de l'homme ou d'animaux n'est plus à démontrer. En effet, la mise en évidence de la dissémination de la résistance aux antibiotiques à partir de ces réservoirs a bouleversé la vision du monde médical dans l'utilisation des antibiotiques pour le traitement des maladies infectieuses. Pourtant, le monde de la phytopathologie considère encore trop souvent uniquement les agents pathogènes sans prendre en compte l'environnement biotique et en particulier les individus commensaux comme acteurs potentiels dans l'émergence de maladie. Monteil et al. (2013) proposent que les populations de *Pseudomonas syringae* retrouvées dans les précipitations et les eaux de surface auraient joué un rôle important dans l'histoire évolutive de l'agent pathogène, *Pseudomonas syringae* pv. *tomato*, et met l'accent sur l'importance de la prise en compte des en-

[3]Le spill-back correspond à l'inverse du spill-over, qui correspond à l'émergence d'une zoonose du milieu domestique vers le milieu sauvage (Daszak et al. 2000).

Encadré I : Systèmes de sécrétion chez les bactéries

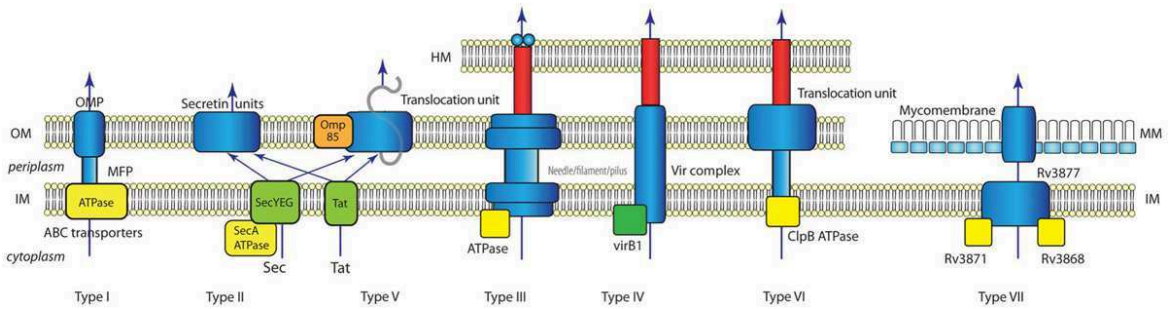


Figure représentant l'ensemble des 7 systèmes de sécrétions des bactéries à Gram-négatif (Tseng et al. 2009)

Le système de sécrétion de type I permet la sécrétion de protéines dans le domaine extracellulaire en une seule étape (Tseng et al, 2004). Il est composé de 3 protéines (ATP-binding cassette transporteur, les facteurs de membranes externes (OMP) et les protéines de fusion membranaires). Ce système peut sécréter des protéines de taille variée (10KDa à 520 KDA) (Boyd et al. 2014).

Le système de sécrétion de type II est impliqué dans la biogenèse des pilus de type IV, qui ont un rôle majeur dans la formation de biofilm et dans la virulence. De plus, il peut sécréter des protéines importantes pour la survie sur l'hôte et des toxines, comme la toxine cholérique chez *Vibrio cholerae* (Nivaskumar & Francetic, 2014 ; Costa et al. 2015). Il sécrète des protéines dans le périplasme puis dans l'environnement extracellulaire.

Le système de sécrétion de type IV peut sécréter de l'ADN et des protéines, à la fois dans des cellules eucaryotes et procaryotes. Ce système a un rôle dans la conjugaison puisqu'il permet de transférer des plasmides (Christie et al. 2005). De plus, il peut jouer un rôle dans la virulence, par exemple chez *Helicobacter pylori*, l'antigène CagA est sécrété par ce système directement dans les cellules gastriques de l'hôte (Odenbreit et al. 2000).

Le système de sécrétion de type V sécrète essentiellement des facteurs de virulence mais il peut avoir un rôle dans la formation de biofilm. C'est un système autotransporteur à deux partenaires (Desvaux et al. 2004).

Le système de sécrétion de type VI sécrète des protéines effectrices dans les cellules bactériennes qui peuvent être impliquées dans la compétition bactérienne. Sa structure est similaire à celle d'un bactériophage (Record, 2011). Il joue également un rôle dans la virulence, notamment chez *Burkholderia mallei* (Schell et al. 2007).

Le système de sécrétion de type VII est retrouvé chez les mycobactéries où il joue un rôle majeur dans la virulence (Houben et al. 2012).

vironnements non agricoles pour étudier l'émergence des bactéries phytopathogènes. Cependant, ces émergences peuvent également provenir des populations colonisant les environnements cultivés, puisque les agents pathogènes sont en interaction avec le microbiote associé aux différents organes des plantes hôtes. En effet, au sein de ces communautés, il est possible que des individus génétiquement proches des agents pathogènes coexistent avec ceux-ci. C'est ce qui est observé par exemple, dans le microbiote instestinal qui est en partie colonisé par des souches commensales d'*Escherichia coli* qui ainsi peuvent se trouver en sympatrie avec des agents pathogènes . Il a été montré que ces bactéries commensales pouvaient acquérir des facteurs de virulence (Wirth et al. 2006). Chez les agents phytopathogènes, il existe également des espèces composées à la fois de souches pathogènes et non pathogènes, c'est le cas, par exemple de l'espèce *Clavibacter michiganensis*, pour laquelle des individus pathogènes et commensaux sont observés chez la même espèce hôte, la tomate (Jacques et al. 2012). La comparaison génomique de souches pathogènes et commensales appartenant à cette espèce a montré que la principale différence correspondait à leur contenu en facteurs de virulence, puisque les souches non pathogènes ne possèdent pas les deux principaux facteurs de virulence de cette espèce, les gènes *celA* et *part−1* (Jacques et al. 2012 ; Zaluga et al. 2014). Bien que la divergence entre ces souches non pathogènes et pathogènes soit élevée (du niveau de la sous-espèce), Zaluga et al. 2014 supposent qu'une émergence d'un nouvel agent pathogène, à partir de ces souches non pathogènes, pourrait se réaliser après l'acquisition de ces deux facteurs de virulence et d'autres facteurs impliqués dans la colonisation et l'adaptation à l'hôte.

La génomique comparative est une méthode en plein développement, ainsi les comparaisons des séquences génomiques entre les souches pathogènes et commensales permettent de mieux comprendre les mécanismes d'émergence à partir de ces réservoirs environnementaux. Chez *Gardnerella vaginalis*, la comparaison génomique et phénotypique d'une souche commensale et d'une souche pathogène a montré une grande divergence entre les deux et que la souche pathogène était plus adhérente et formait des biofilms plus importants (Harwich et al. 2010). Ce type d'analyse a été réalisé, par exemple, chez *Escherichia coli* où il a montré que le génome de la souche commensale correspondait à une mosaïque des génomes des agents pathogènes de la même espèce (Rasko et al. 2008). Les auteurs supposent qu'avec cette structure génomique mosaïque, les souches commensales pourraient représenter un réservoir de facteurs de virulence.

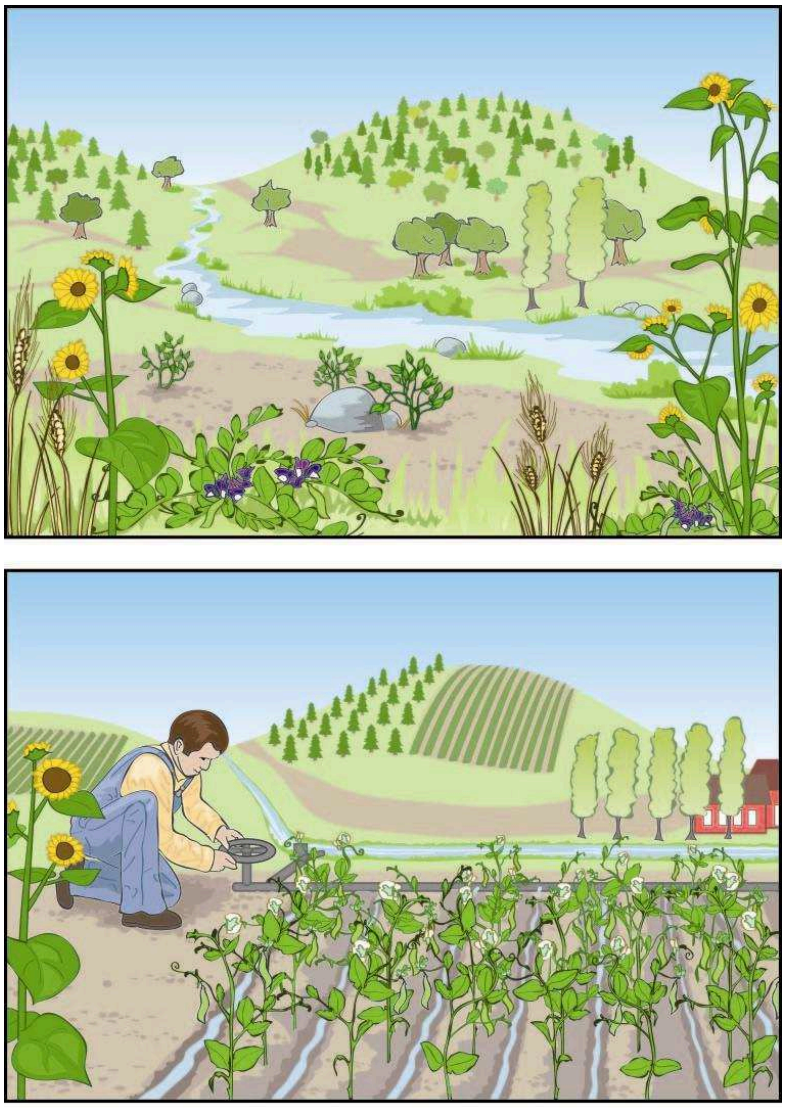


Figure I-8 : Représentation des principales différences entre les milieux naturels (en haut) et les agroécosystèmes (en bas). Les environnements naturels sont hétérogènes et montrent une diversité spécifique importante. Les agroécosystèmes quant à eux sont plus homogènes, avec une faible diversité spécifique et une densité en hôte importante, ce qui facilite la transmission des agents pathogènes d'un individu à un autre (Stuckenbrock & McDonald, 2008).

5. Emergence de phytobactérioses dans les agroécosystèmes

Le cadre dans lequel on se place pour étudier les émergences doit être défini puisque finalement, une maladie va être considérée comme émergente selon ce cadre et les conséquences qu'elle va entrainer (Guiserix, 2009). Pour ce travail de thèse, je me suis donc placée dans le cadre des émergences de nouvelles maladies bactériennes ou de réémergences sur plantes au sein des agroécosystèmes.

Les agroécosystèmes sont souvent sujets à des émergences de maladies bactériennes. Ces environnements présentent une diversité spécifique très faible pouvant même aller jusqu'à des parcelles monospécifiques où la densité en individus potentiellement hôtes pour les bactéries phytopathogènes est très importante (fig. I-8). Ces milieux, étant contrôlés par l'homme ils sont moins variables que les milieux naturels. Ils sont donc plus sensibles que les milieux naturels à de nouvelles émergences, étant donné que l'équilibre hôte-pathogène est en faveur de l'agent pathogène. En effet, les milieux naturels sont des environnements hypervariables ce qui les rend moins favorables au maintien et à la dissémination de souches fortement agressives et très spécialisées sur leur hôte (Stuckenbrock et McDonald, 2008). Dans les agroécosystèmes, on peut donc voir la mise en place de l'émergence de complexes clonaux, dont les souches sont fortement agressives. Ces dernières années, plusieurs épidémies ont été dramatiques pour ces agroécosystèmes. Par exemple, on peut noter l'épidémie causée par *Xylella fastidiosa*, qui a émergé en Italie sur olivier en 2013 (Saponari et al. 2013) et celle causée par *Pseudomonas syringae* pv. *actinidiae* qui a émergé en France dans les vergers de kiwi en 2010 (Vanneste et al. 2013).

L'agent pathogène *Xylella fastidiosa* est une bactérie limitée au xylème transmise par des insectes. Elle est composée de 6 sous-espèces, *fastidiosa*, *sandyi*, *multiplex*, *morus*, *tashke* et *pauca* (Jacques et al. 2016). Chaque sous-espèce présente une gamme d'hôtes plus ou moins large et une aire d'origine spécifique. Il a été montré que cette bactérie était naturellement compétente, et donc qu'elle était capable de recevoir de l'ADN extérieur par transformation (Kung et al. 2011). La recombinaison homologue est une force évolutive importante dans l'émergence de ces différentes sous-espèces et serait à l'origine de l'émergence de la sous-espèce *morus* pour laquelle, un changement d'hôte aurait été consécutif aux évènements de recombinaison entre les sous-espèces *fastidiosa* et *multiplex* (Nunney et al. 2014). Alors, que cet agent pathogène était connu pour être présent uniquement en Amérique, il a été isolé sur olivier dans la région des Pouilles, au sud de l'Italie en 2013, provoquant la mort de milliers d'oliviers (Saponari et al. 2013). La souche responsable de cette épidémie est une souche de la sous-espèce *pauca*, appelée souche CoDiRO, et aurait

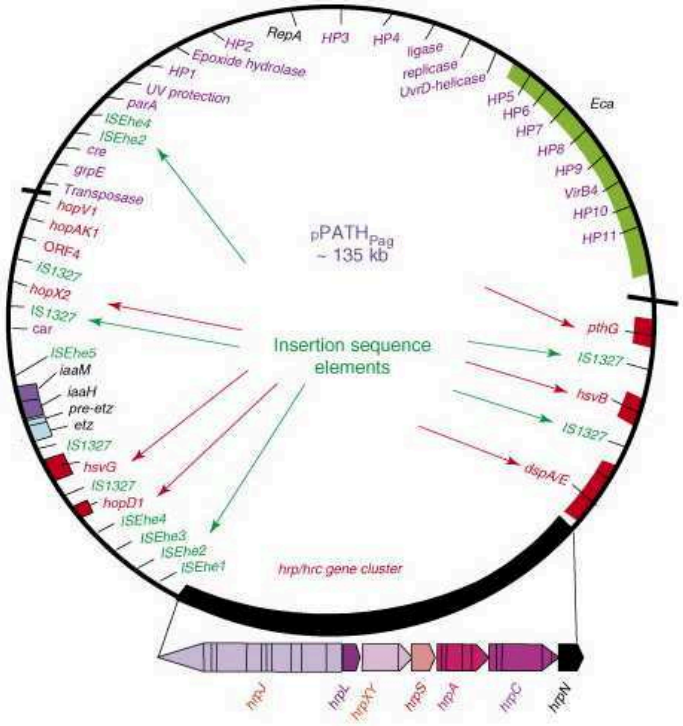


Figure I-9 : Cartographie du plasmide p-PATH de *Pantoea agglomerans*. L'ilot de pathogénie est représenté par le bandeau noir, au sein duquel sont retrouvés les gènes du cluster *hrp/hrc*. Les flèches rouges indiquent 6 effecteurs de type 3 fonctionnels, et les flèches vertes indiquent les séquences d'insertions (Barash & Manulis-Sasson, 2007).

pour origine le Costa Rica (Loconsole et al. 2016). Il semblerait que cette émergence dans la région des Pouilles soit due à l'introduction de plantes ornementales contaminées provenant du Costa Rica (Martelli et al. 2016). L'histoire évolutive de cette espèce est donc marquée à la fois par l'émergence d'un nouvel agent pathogène suite à la recombinaison homologue, et par son introduction dans une nouvelle aire géographique. Les outils de génétique des populations ont permis de retracer cette histoire évolutive et de mettre en évidence les mécanismes moléculaires impliqués dans cette émergence.

L'agent pathogène *Pseudomonas syringae* pv. *actinidiae*, quant à lui, est responsable de la bactériose du kiwi. Ce pathovar[4] (pv.) est divisé en 4 biovars (Vanneste et al. 2013), qui regroupent les souches ayant des caractéristiques biochimiques ou physiologiques communes (Lepage et al. 1992). L'un de ces 4 biovars, le biovar 3 est responsable d'une épidémie à l'échelle mondiale, causant des pertes économiques importantes (Scortichini et al. 2012). Il a été observé pour la première fois dans le Latium en 2008, une région située au centre de l'Italie. Ensuite, ce biovar a été détecté en Nouvelle-Zélande, le principal pays producteur de Kiwi, puis en Espagne, au Portugal et en France. L'analyse de l'origine de ce biovar en France a montré qu'il avait la même origine que l'émergence observée en Italie (Cunty et al. 2015). En effet, plusieurs haplotypes identiques ont été retrouvés dans ces deux pays. L'origine de l'émergence de cet agent pathogène est donc due à l'introduction dans une nouvelle aire géographique d'un agent pathogène sans doute pré-existant, mais non encore découvert depuis l'aire d'origine du kiwi qui se trouve en Chine (Mazzaglia et al. 2012). De plus, son étude épidémiologique a permis de mettre en évidence que cette épidémie était causée par un complexe clonal et donc de développer des outils de détection spécifique à cet agent pathogène.

Les agents phytopathogènes appartenant à l'espèce *Pantoea agglomerans* auraient émergé à partir d'une souche commensale grâce à l'acquisition d'un plasmide, p-PATH, porteur d'un ilot de pathogénie (Manulis & Barash, 2003 ; Barash & Manulis-Sasson, 2007). Cette espèce a un mode de vie épiphyte et commensal et est associée à un grand nombre d'espèces végétales. Cependant, les agents pathogènes ayant émergé à partir de souches commensales présentent une forte spécificité d'hôte. En effet, deux pathovars existent au sein de cette espèce, le pathovar *gypsophilae*, responsable de la galle sur gypsophile, une herbacée, et le pathovar *betae* capable d'infecter la betterave, mais également le gypsophile (Manulis & Barash, 2003). De plus, il a été montré que la pathogénie de ces deux pathovars était due au plasmide p-PATH qui porte un cluster Hrp (fig. I-9) (Nizan et al. 1997), codant le complexe protéique du système de sécrétion de type III, principal facteur de viru-

[4]Un pathovar correspond à un rang taxonomique infra-subspécifique qui rassemble l'ensemble des souches responsables des mêmes symptômes sur une même gamme d'hôtes (Dye et al., 1980).

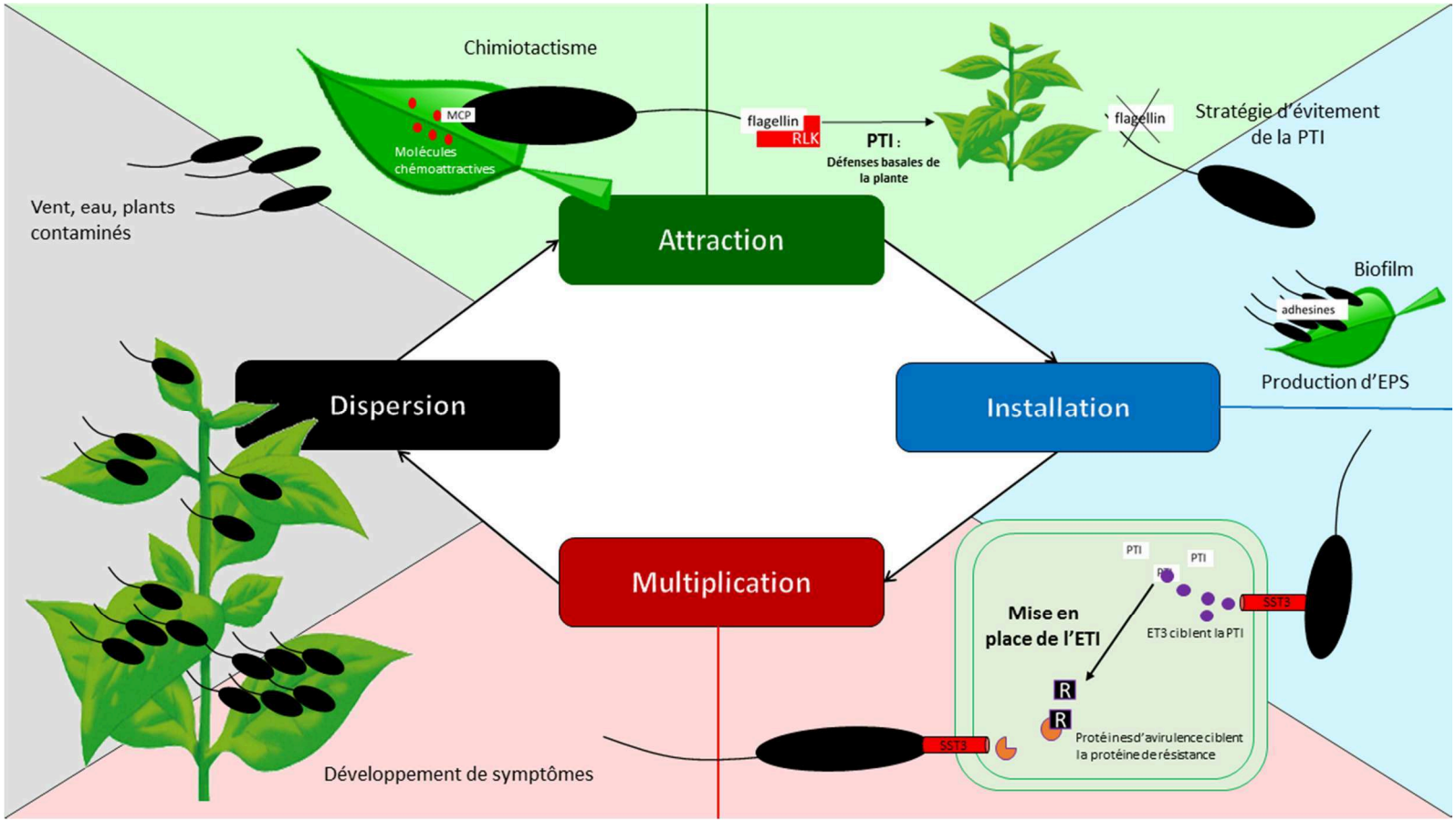


Figure I-10 : Représentation du cycle infectieux des bactéries épiphytes pathogènes sur plantes. La première étape correspond à l'attraction et fait intervenir le chimiotactisme, qui permet la détection de sites favorables grâce aux MCP (Methyl-accepting chemotaxis protein). La plante reconnait des PAMP via ses protéines RLK tels que la flagelline et déclenche son premier niveau de défense, la PTI. L'agent pathogène peut échapper à ces défenses et entrer dans la phase d'installation, qui va se traduire par la formation de biofilms à la surface des tissus et dans les sites du mésophylle qu'elle a pu atteindre. Ensuite lors de son installation, la bactérie contourne le deuxième niveau de défenses de la plante (ETI) grâce aux effecteurs de type 3 (ET3), sécrétés via le système de sécrétion de type 3 (SST3) et grâce à ses autres facteurs de virulence va entrainer le développement de symptômes de la maladie. Lors de cette phase la bactérie se multiplie activement et se disperse dans la plante. Enfin, la dernière phase du cycle correspond à la dispersion à plus grande distance qui se fait grâce au vent, à l'eau ou aux semences et plants contaminés.

lence des bactéries pathogènes à Gram-négatif. Il semblerait que les deux pathovars aient divergé à partir d'un ancêtre capable d'induire une maladie sur la betterave et sur le gypsophile (Barash & Manulis-Sasson, 2009). Dans cet exemple, les agents pathogènes ont donc émergé suite à un transfert horizontal de gènes. Cet exemple permet également de montrer l'importance du rôle que peuvent avoir les souches commensales dans l'émergence de nouvelles maladies associées aux plantes.

II. Mise en place de l'interaction hôte-pathogène chez les bactéries phytopathogènes

Dans cette partie, je présenterai le cycle infectieux des bactéries à Gram-négatif colonisant les parties aériennes des plantes. En effet, il correspond notamment au cycle infectieux des *Xanthomonas* qui sont le modèle d'étude de cette thèse (voir ci-dessous).

Au cours de l'interaction hôte-pathogène, différents facteurs de virulence sont secrétés par l'agent pathogène, ils vont être reconnus par la plante et vont déclencher une cascade de réactions immunitaires. Contrairement aux animaux, qui possèdent un système immunitaire inné et acquis, les plantes ne possèdent qu'un système inné (Numberger et al. 2004). Cependant, elles ne possèdent pas de cellules spécialisées dans la détection des agents pathogènes, mais au contraire, chaque cellule est capable de reconnaitre un agent pathogène et d'induire les défenses (Zipfel, 2008). Les plantes possèdent une première barrière de défenses constitutives, qui peuvent être physiques, notamment grâce à la paroi cellulaire ou la cuticule, ou biochimiques grâce aux métabolites secondaires, par exemple. Ces défenses ne sont pas spécifiques des attaques de micro-organismes, puisqu'elles peuvent également empêcher les attaques des herbivores (Wittstock & Gershenzon, 2002). Ensuite, si l'agent pathogène réussit à franchir ces premières barrières, il sera confronté au système immunitaire à proprement parler. Ces différents niveaux du système de défense de la plante induisent une pression de sélection sur l'agent pathogène qui développe des stratégies au cours de son cycle infectieux lui permettant de contourner ces différentes défenses (Jones & Dangl, 2006). Chaque étape du cycle infectieux de la bactérie sur la plante fait donc intervenir des facteurs spécifiques (fig. I-10). Le cycle infectieux d'une bactérie épiphyte sur la plante est divisé en trois étapes : la phase d'attraction, la phase d'installation, la phase de multiplication. Les phases d'attraction et d'installation sont regroupées dans la phase dite

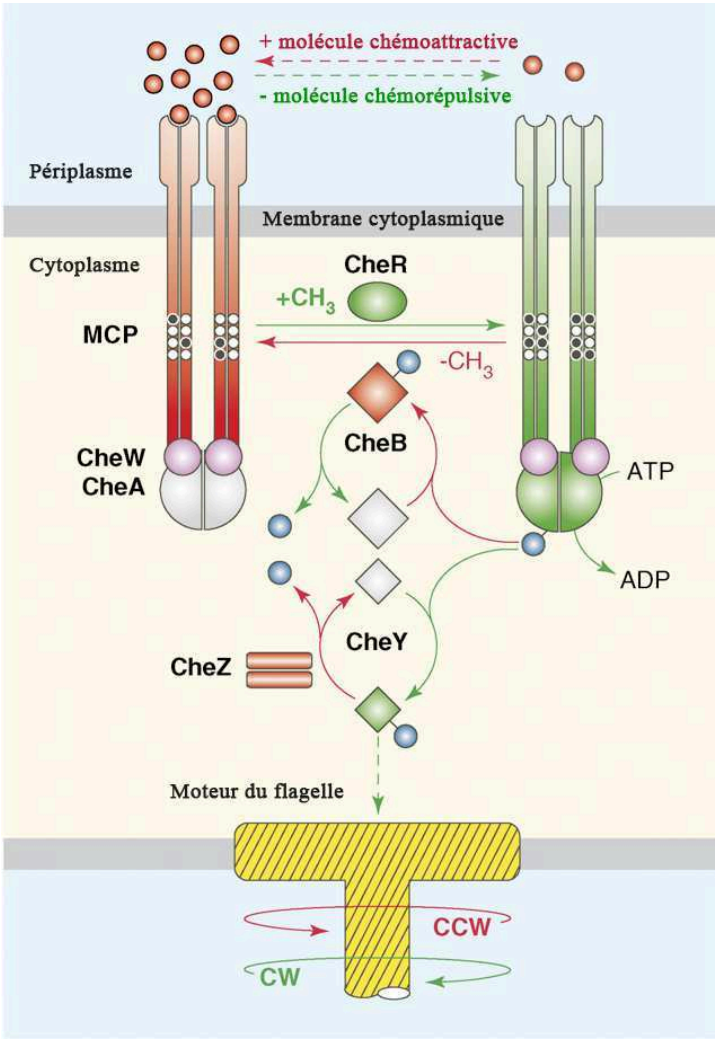


Figure I-11 : Cascades de phosphorylation et de déphosphorylation des protéines kinases du chimiotactisme chez *E. coli* (d'après Hazelbauer et al, 2007 ; Medbhi-Hajri, 2010). En rouge est représentée la cascade induite par une molécule attractive et en vert celle induite par une molécule répulsive. Cette cascade induit la mobilité du flagelle, dont le moteur va tourner soit dans le sens horaire (CW) ou dans le sens anti-horaire (CCW).

asymptomatique, tandis que, généralement, la phase de multiplication correspond au moins partiellement à la phase d'expression des symptômes (Alfano & Collmer, 1996). Ensuite, il y a une phase de dissémination qui peut se faire par l'eau, le vent ou les plantes contaminées par les agents pathogènes (via les semences par exemple).

1. La phase d'attraction

Une fois que la bactérie arrive sur la plante hôte, elle peut coloniser les sites favorables à sa survie au sein de cet hôte. En effet, la phyllosphère, qui correspond aux parties aériennes de la plante, constitue un environnement stressant pour les bactéries qui arrivent à la surface foliaire (Lindow & Brandl, 2003). Ces bactéries vont être soumises à différents stress abiotiques, notamment les UV, les variations de températures et de l'humidité relative (Hirano & Upper, 2000 ; Lindow & Brandl, 2003 ; Turner et al. 2013). La surface foliaire ne correspond pas à une surface plane et homogène, mais au contraire à une diversité d'habitats pour les micro-organismes (Lindow & Brandl, 2003). Les bactéries colonisent donc les microhabitats les plus favorables à leur installation, comme la base des trichomes par exemple. La surface foliaire présente également des voies d'entrées pour la bactérie, qui peut alors s'internaliser dans les tissus de la plante. Ces voies d'entrées correspondent aux stomates, aux hydathodes ou aux blessures (Perez-Bueno et al. 2015). Lors de cette étape d'attraction, la bactérie va détecter les sites favorables. Cette détection est possible grâce au chimiotactisme, ce qui va ensuite entrainer l'activation de la mobilité de la bactérie.

a. Le chimiotactisme et la cascade de transduction du signal activent le flagelle

Le chimiotactisme correspond aux mouvements de la cellule dans un gradient de paramètres physico-chimiques variés (Alexandre, 2010). Son rôle va être de transduire un signal jusqu'au flagelle ce qui induira la direction de rotation de ce dernier, en cas de signal attractif ou répulsif (Berg, 2003 ; Wadhams & Armitage, 2004). Il existe donc deux cascades de transduction, respectivement spécifiques de l'attraction vers une source ou de la répulsion. La première cascade liée au déplacement de la bactérie vers une source attractive fait intervenir la déphosphorylation des protéines kinases du chimiotactisme, CheA et CheY (fig. I-11). Au contraire, la phosphorylation de ces deux protéines entraine un mouvement désordonné du flagelle, permettant à la bactérie de ne pas s'approcher d'une source répulsive (Bren & Eisenbach, 2000). Par exemple, une analyse du chimiotactisme, chez *Ralstonia solanacearum*, a montré que la bactérie était attirée vers

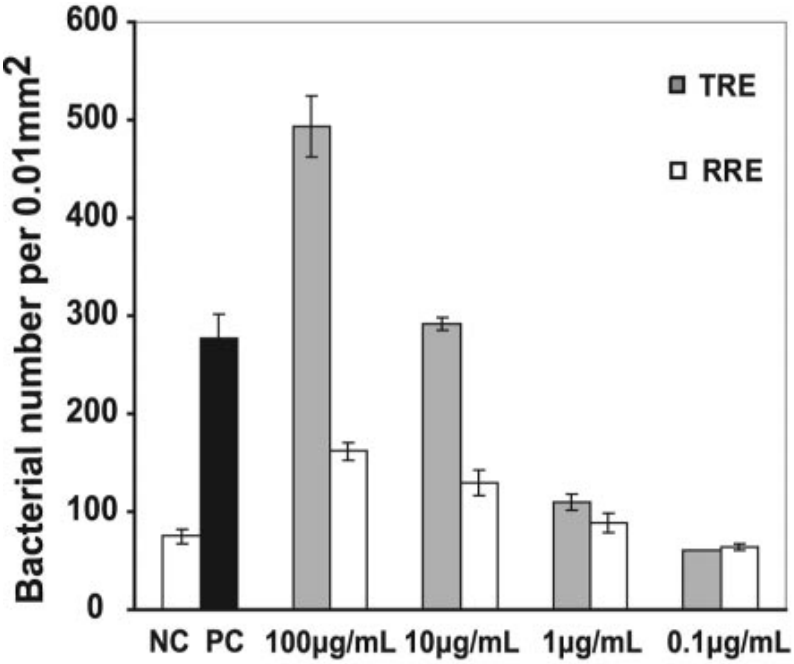


Figure I-12 : Réponse de la souche K60 appartenant à l'espèce *Ralstonia solanacearum*, aux composés issus de la tomate (plante hôte) et du riz (plante non-hôte). Ces histogrammes représentent le nombre de cellules bactériennes dans 0.01 mm² autour d'un bouchon d'agarose contenant les différents composés en fonction des différentes concentrations protéiques des composés utilisées. NC : contrôle négatif, PC : contrôle positif (0.1% d'extrait de levure), TRE : exudats racinaires de tomate, RRE : exudats racinaires de riz. (Yao & Allen, 2006)

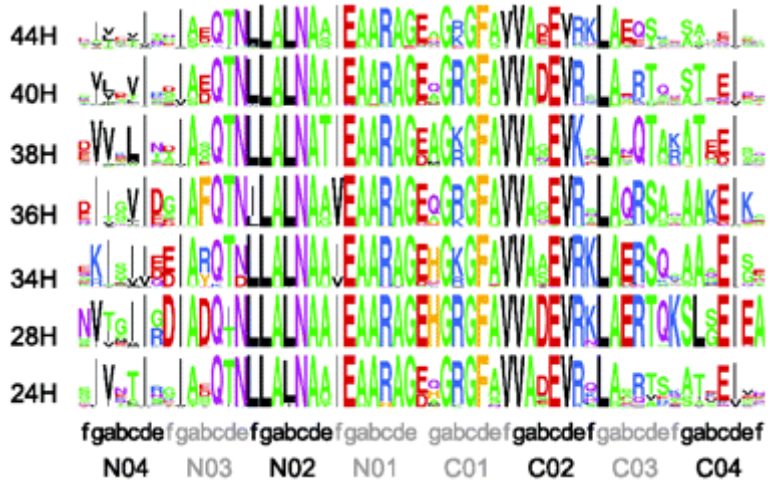


Figure I-13 : Représentation des séquences protéiques du domaine cytoplasmique conservé entre les 7 familles de MCP (Alexander & Zhulin, 2007). Chaque ligne horizontale correspond à une séquence représentative des familles de MCP.

des composés moléculaires secrétés au niveau des racines de la tomate, une plante hôte, mais au contraire, elle ne l'est pas lorsque ces composés proviennent d'une plante non hôte, par exemple le riz (fig. I-12) (Yao & Allen, 2006). De plus, les auteurs ont également montré l'importance des protéines CheA et CheW dans le chimiotactisme puisque les souches mutées aux gènes codants ces deux protéines ne sont plus attirées par les composés provenant de leur plante hôte. Les protéines CheW se lient directement aux chémorécepteurs (Gegner et al. 1992).

Les principaux acteurs moléculaires impliqués dans le chimiotactisme et qui vont permettre de déclencher cette cascade de transduction sont les chémorécepteurs appartenant à la famille des Methyl-accepting Chemotaxis Proteins (MCP). Le rôle de ces MCP dans le chimiotactisme a été mis en évidence en 1975 par Kort et collaborateurs, chez *Escherichia coli* et *Salmonella typhimurium*. Ces protéines sont ancrées dans la membrane plasmique de la bactérie (fig. I-11). Elles ont pour rôle de se lier directement ou indirectement aux molécules présentes dans l'environnement, telles que les sucres, les acides aminés ou les peptides, grâce à leur domaine senseur localisé dans le périplasme (Alexandre, 2010). En effet, les chémorécepteurs sont composés de différents domaines : un domaine de fixation extracellulaire (le domaine senseur), un domaine de signalisation cytoplasmique et le domaine HAMP (Wadhams & Armitage, 2004 ; Hulko et al. 2006). Le domaine HAMP a pour rôle de contrôler la phosphorylation du domaine cytoplasmique, qui est composé de deux régions ainsi que le site de fixation aux protéines CheA et CheW (Alexander & Zhulin, 2007). Si le domaine cytoplasmique est déphosphorylé, cela induit la cascade de déphosphorylation des protéines Che, c'est le cas en présence d'une source attractive. Au contraire, s'il est phosphorylé, cela induit la cascade de transduction concernant une source répulsive. Le domaine cytoplasmique sert donc pour la transduction du signal, il interagit directement avec les protéines kinases du chimiotactisme qui sont localisées dans le cytoplasme de la cellule bactérienne.

Les MCP sont donc essentiels pour la perception de l'environnement par la bactérie et ont un rôle important dans la spécificité d'hôte. En effet, leur répertoire varie en fonction des souches et sont corrélés à leur gamme d'hôte (Mhedbi-Hajri et al. 2011). Une étude réalisée sur des souches appartenant au genre *Xanthomonas* a permis de montrer que cette spécificité d'hôte était exprimée dès les premières phases de la colonisation de l'hôte par les répertoires en gènes codants pour des MCP notamment, et que ces gènes étaient soumis à la sélection positive (Mhedbi-Hajri et al. 2011). Il semblerait cependant, que les pressions de sélection sur ces gènes ne soient pas les mêmes selon les domaines qui les composent. En effet, une analyse des séquences de MCP provenant de 312 génomes procaryotes a montré que le domaine cytoplasmique de ces protéines était très conservé (fig. I-13). Cette étude a également permis de classer les MCP en 7 familles en fonction des in-

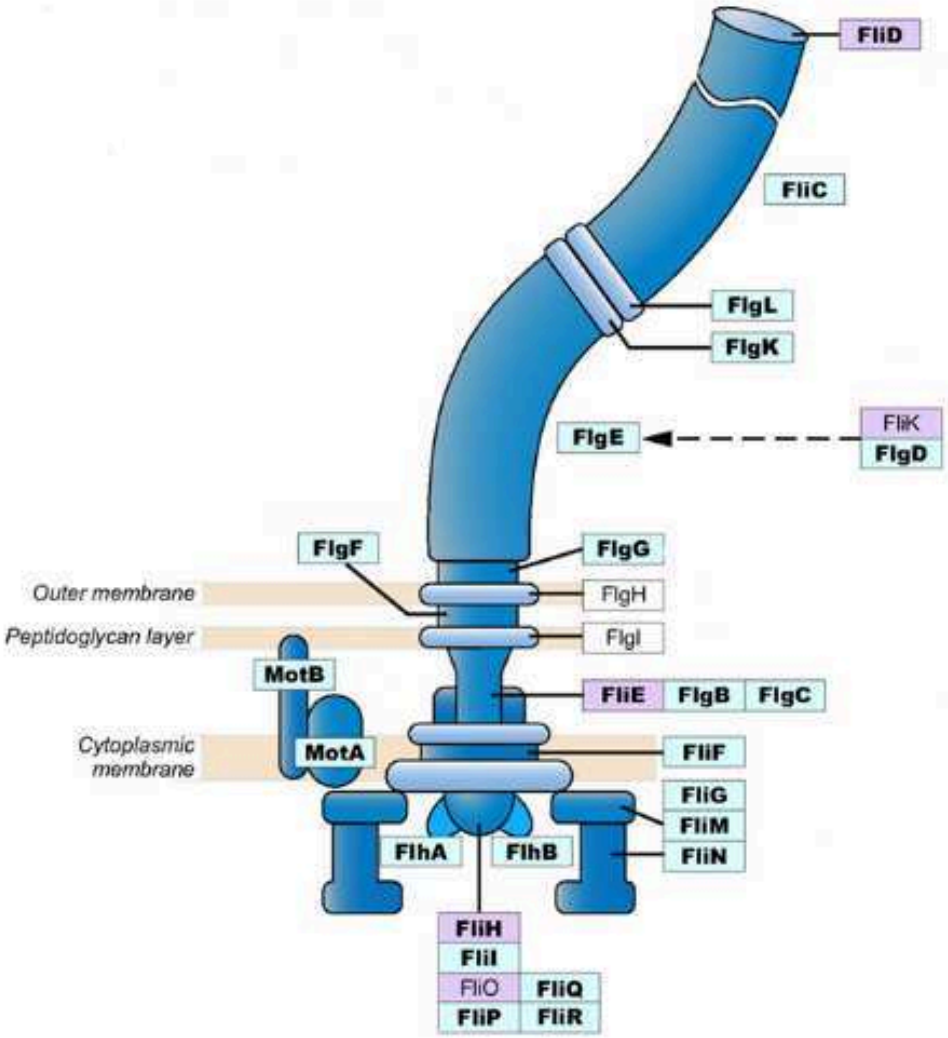


Figure I-14 : Représentation de la distribution de chaque protéine du flagelle (d'après Liu & Ochman, 2007). En bleu clair sont représentées les protéines présentes chez tous les taxons, en violet, les protéines ayant une distribution plus réduite.

sertions ou délétions observées dans ce domaine (Alexander & Zhulin, 2007). Au contraire, le domaine de fixation extracellulaire évolue plus rapidement, et est très variable (Wuichet & Zhulin, 2003). Cela peut s'expliquer par le fait que le domaine cytoplasmique permet les interactions avec les protéines du chimiotactisme, tandis que le domaine extracellulaire est en interaction directe avec les molécules de l'environnement, et est capable d'en détecter un grand nombre (Wuichet et al. 2007 ; Krell et al. 2010). Par exemple, un MCP, Hsb1, spécifique de *X. campestris* pv. *campestris*, agent pathogène des Brassicacées, serait impliqué dans la spécificité d'hôte et permettrait à l'agent pathogène de détecter un glucosinolate et un acide aminé (Indiana, 2014).

Une fois que la bactérie a localisé les sites favorables ou défavorables, elle se déplace grâce aux mouvements rotatoires du flagelle. Cette structure est un complexe protéique d'environ 30 protéines situées dans la membrane plasmique de la bactérie (fig. I-14) (Liu & Ochman, 2007). Il est composé d'un segment extracellulaire, le flagelle, d'un corps basal ancré dans la membrane et de composants cytoplasmiques. Chez *E. coli* et *Salmonella enterica* le moteur du flagelle est actionné grâce à un potentiel électrochimique de protons, appelé la force motrice protonique (Berg, 2003). Deux sens de rotation du moteur sont possibles en fonction des signaux perçus, une rotation horaire ou anti-horaire. Chez *Salmonella*, le sens de rotation anti-horaire des flagelles permet à la bactérie d'avancer, tandis que si l'un des flagelles tourne dans le sens horaire, la bactérie change de direction (Macnab & Ornston, 1977). Le flagelle permet donc la mobilité par nage de la bactérie dans un milieu liquide mais il peut également avoir un rôle dans l'adhésion, la formation de biofilm ou l'expression de la virulence (ces étapes seront décrites par la suite) (Kirov, 2003). En effet, le flagelle possède également une fonction sécrétrice, comme cela a été montré chez *Yersinia enterocolitica*, puisqu'un mutant de la flagelline présente à la fois une perte de biosynthèse du flagelle, mais également une perte de plusieurs protéines sécrétées (Young et al. 1999).

b. Perception de la bactérie par la plante et stratégie de contournement

Lors de cette étape de colonisation de sites favorables, la bactérie doit faire face aux défenses basales de la plante qui sont nommées PTI (PAMP-triggered immunity). Elles ont pour rôle de détecter et de neutraliser les agents pathogènes ou plus généralement les microbes grâce à la reconnaissance de motifs initialement appelés Pathogen-Associated Molecular Patterns (PAMP) puis plus globalement Microbe-Associated Molecular Patterns (MAMP). La plante reconnait ces motifs grâce à des récepteurs qui sont situés à la surface de la cellule, ces récepteurs sont

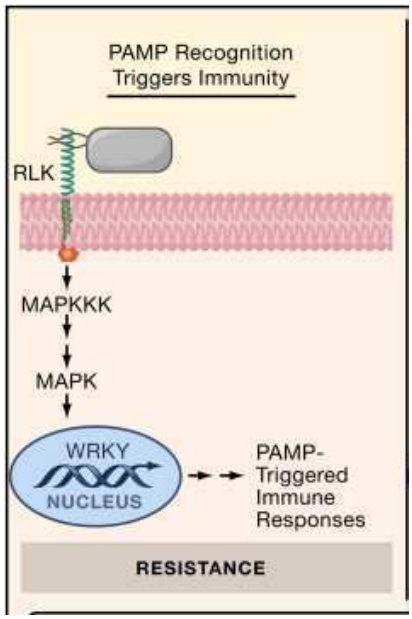


Figure I-15 : Description de la PTI (d'après Chisholm et al. 2006). Reconnaissance d'un PAMP, tel que la flagelline présente dans le flagelle des agents pathogènes bactériens par un récepteur membranaire de la cellule végétale de type RLK. Cette reconnaissance active les protéines MAPK, qui à leur tour activent les facteurs de transcription WRKY.

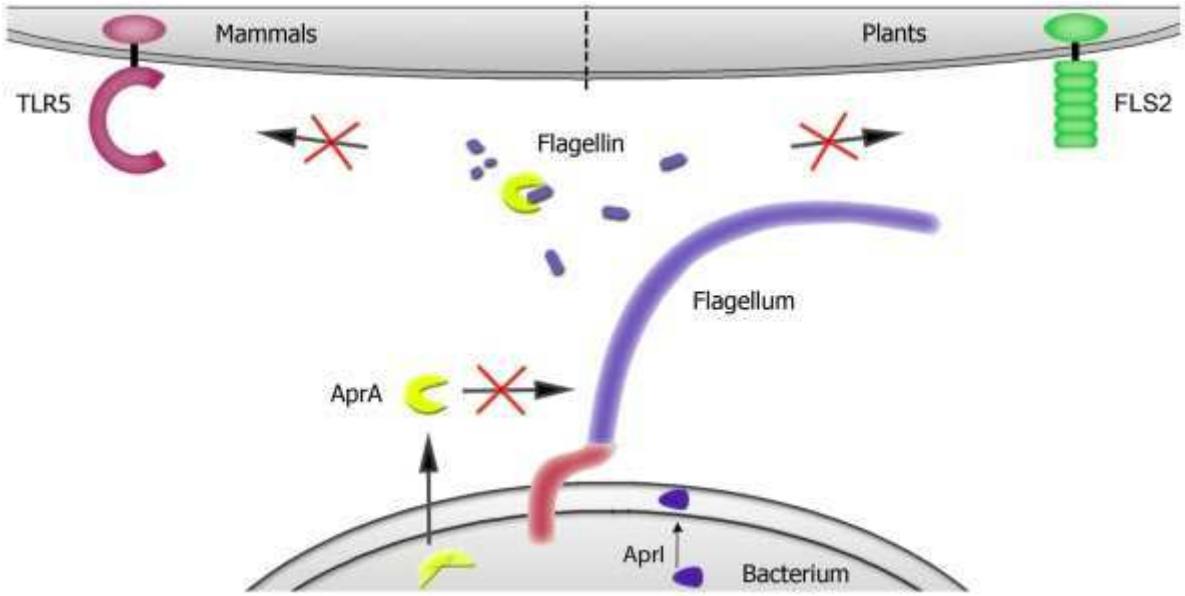


Figure I-16 : Représentation de la stratégie mise en place par l'agent pathogène *Pseudomonas aeruginosa* pour échapper à la reconnaissance via la flagelline (Bardoel et al. 2011). La bactérie sécrète la protéine AprA dans le milieu extracellulaire, cette protéine dégrade la flagelline sans affecter la fonction du flagelle. Ainsi, la bactérie échappe au système immunitaire des mammifères et des plantes.

appelés PRR pour Pattern Recognition Receptors. Ces PRR sont spécifiques d'un MAMP particulier. Les MAMP ont la particularité d'être situés dans des domaines très conservés de l'agent pathogène, et la perte de celui-ci pourrait entrainer une perte de virulence ou de fitness (Boller & He, 2009). Lorsque le motif est reconnu par la plante grâce à une protéine kinase, cela va activer les protéines MAPK qui sont transloquées dans le noyau où elles activent les facteurs de transcription de type WRKY (fig. I-15) (Numberger et al. 2004). Cette défense a été mise en évidence notamment grâce à un MAMP situé dans le flagelle des bactéries à Gram-négatif, qu'il s'agisse d'individus pathogènes ou commensaux (Felix et al. 1999). Chez *Arabidopsis* et chez la tomate, il a été montré que la plante va reconnaitre un motif de 22 acides aminés, appelé flg22, présent dans la partie N-terminale de la flagelline grâce à la protéine FLS2 et induire la cascade précédemment décrite (Chinchilla et al. 2006).

Stratégies d'échappement à la reconnaissance des MAMP

La reconnaissance des MAMP intervient dès la première phase de colonisation de l'hôte. Afin de contourner cette première défense, les bactéries peuvent perdre ou modifier ces motifs reconnus par la plante. En effet, des souches de *Pseudomonas aeruginosa* peuvent sécréter une protéine AprA, via le système de sécrétion de type I, protéine qui clive la flagelline (fig. I-16). Ainsi, la fonction de mobilité de la bactérie n'est pas altérée, mais les défenses de la plante ne sont pas activées (Bardoel et al. 2011). Par ailleurs, il a été montré qu'au sein d'une population d'agents pathogènes, certains individus pouvaient avoir perdu leur flagelle. Dans le cas de *Xanthomonas fuscans* subsp. *fuscans* il a été avancé qu'il pourrait s'agir d'une stratégie mise en place par l'agent pathogène pour que sa présence sur la plante ne déclenche pas des réactions de défenses proportionnées à la taille réelle de sa population (Darrasse et al. 2013). D'autres stratégies d'évitement de la reconnaissance des MAMP sont décrites par Rossez et al. (2015), comme, par exemple, des mutations dans le motif flg22 du flagelle.

2. La phase d'installation

Au cours de cette phase d'installation, la bactérie va former un biofilm, composé de microorganismes agrégés dans une matrice d'exopolysaccharides et attachés à une surface (Costerton et al. 1995). Un biofilm peut être composé d'une seule espèce ou de plusieurs espèces microbiennes et constitue un " hot spot " d'échange de matériel génétique (Danhorn & Fuqua, 2007). Cette structure permet également aux bactéries qui les composent de faire face aux stress de l'environnement, et joue donc

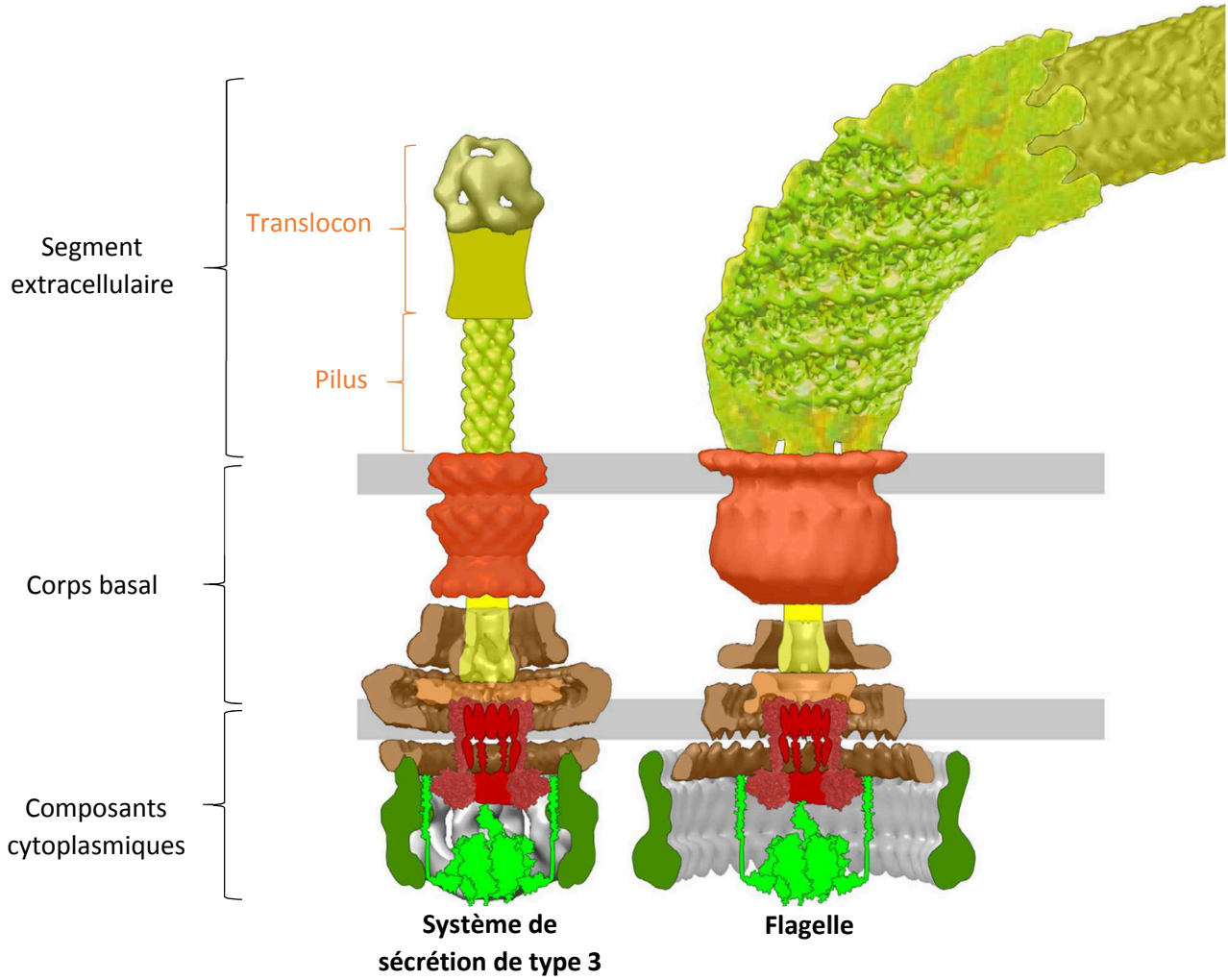


Figure I-17 : Comparaison des structures du système de sécrétion de type 3 (à gauche) et du flagelle (à droite) (d'après Portaliou et al. 2016). Ces images ont été obtenues en microscopie électronique. Les 3 segments : cytoplasmique, le corps basal et les composants extracellulaires sont représentés par les accolades noires à gauche. Les accolades oranges représentent les deux structures du segment extracellulaire du système de sécrétion de type 3. Les structures homologues entre les deux complexes protéiques sont représentées de la même couleur.

un rôle important pour les bactéries inféodées à la phyllosphère. En effet, comme indiqué précédemment, cet environnement peut être particulièrement hostile pour les bactéries qui le colonisent (Jacques et al., 2005 ; Morris & Monier, 2003). La formation de biofilms peut avoir un rôle dans le contournement des défenses de la plante par les agents pathogènes. En effet, une étude réalisée sur la formation de biofilm chez les souches de *Pseudomonas aeruginosa* pathogènes du basilic montre que le biofilm confère à cet agent pathogène une résistance contre les composés antimicrobiens sécrétés par la plante hôte (Walker et al. 2004). Cependant, la formation de biofilm est un processus pratiquement universel, initialement décrit pour les microorganismes aquatiques puis observé chez les microorganismes dans des environnements industriels, alimentaires, mais également hospitaliers ou vétérinaires (Kong et al. 2006 ; Otto, 2013). Un aspect initialement considéré comme important dans la formation de biofilms était la saturation de l'environnement en eau. Cependant des milieux en partie hydraté seulement comme la rhizosphère ou la phyllosphère hébergent des biofilms (Morris et al. 1997).

La formation de biofilm dépend d'adhésines et de polysaccharides (EPS pour exopolysaccharides et LPS pour lipopolysaccharides). Il existe deux types d'adhésines : fibrillaires et non fibrillaires. Ces deux types de protéines n'interviennent pas au même moment lors de la formation du biofilm. En effet, les adhésines non fibrillaires servent à l'attachement initial des cellules bactériennes sur une surface abiotique ou biotique (mais non bactérienne), tandis que les adhésines fibrillaires ont un rôle dans l'agrégation des bactéries entre elles (Danhorn & Fuqua, 2007). Parmi les adhésines fibrillaires, on retrouve les pili de type IV qui interviennent également dans la mobilité des bactéries. Cependant, cette mobilité nécessite une surface solide sur laquelle se fixe le pilus, puis avec des mouvements de rétraction de ce dernier, la bactérie peut bouger, ce processus est appelé " twitching " (Mattick, 2002).

3. La phase de multiplication

Une fois installé au niveau des sites favorables à sa survie, l'agent pathogène va se multiplier, pour atteindre des tailles de population très importantes. Les facteurs de virulence sont exprimés au cours de la phase de multiplication de la bactérie. Comme nous l'avons vu, une fois que la plante utilise ses défenses basales contre l'agent pathogène, celui-ci peut contourner ces défenses en utilisant ses facteurs de virulence. Le principal facteur de virulence des bactéries à Gram-négatif est le système de sécrétion de type 3 et ses effecteurs, dont certains ont pour cible les protéines induites lors des défenses basales, et d'autres ont pour rôle d'inhiber l'immunité déclenchée par les effecteurs (ETI pour Effector-triggered immunity).

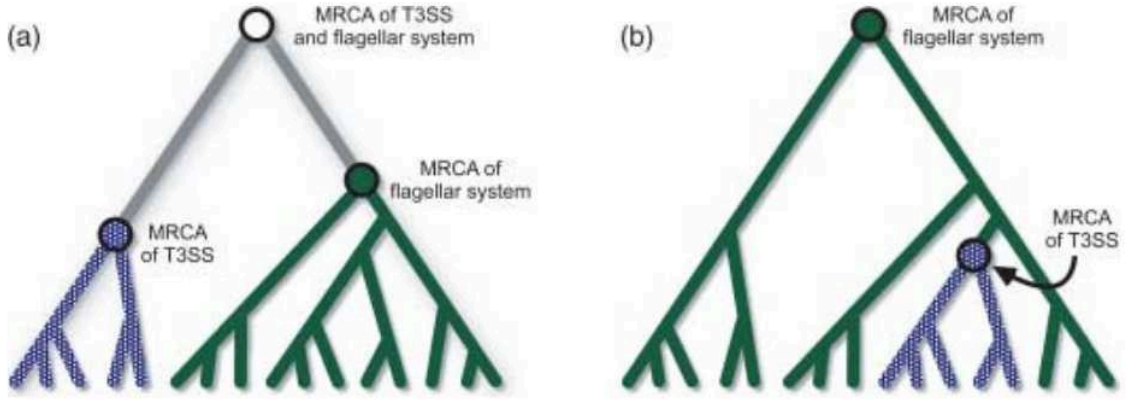


Figure I-18 : Représentation des deux hypothèses des relations évolutives entre le flagelle et le système de sécrétion de type 3 (McCann & Guttman 2008). a : Le flagelle et le système de sécrétion de type 3 ont évolué indépendamment l'un de l'autre, mais possèdent un ancêtre commun représenté par le cercle blanc. b : Le système de sécrétion de type 3 est un descendant direct du système flagellaire, ainsi il correspond à une lignée du système flagellaire.

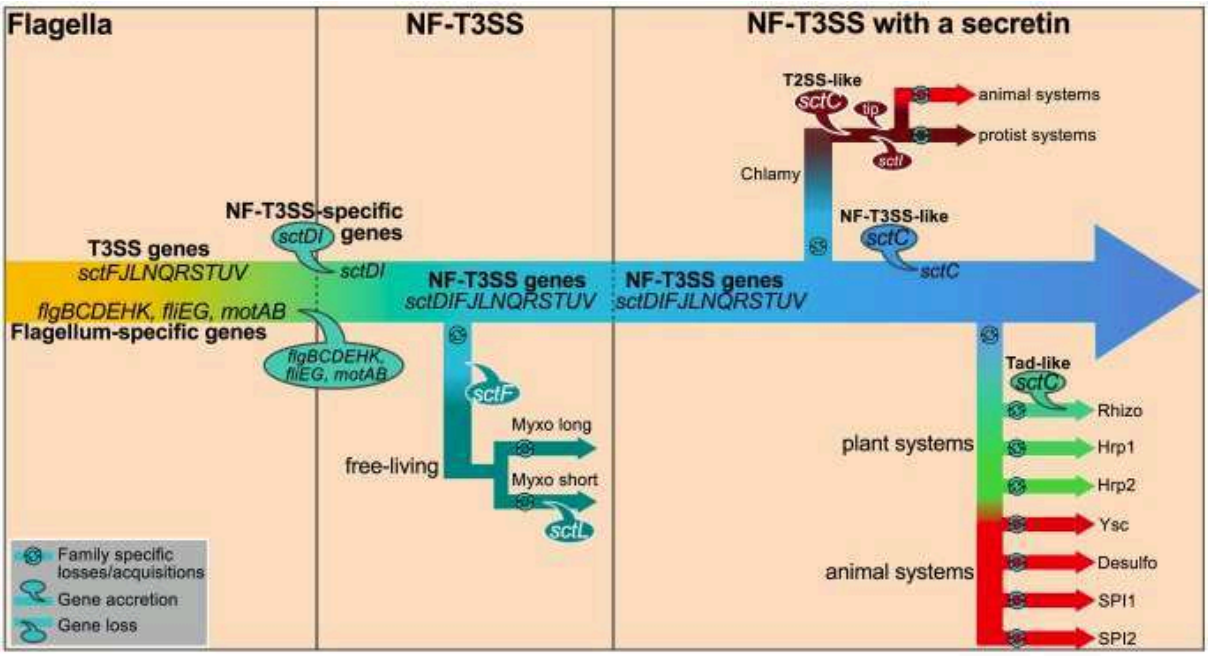


Figure I-19 : Modèle d'évolution et de diversification du système de sécrétion de type 3 proposé par Abby & Rocha, 2012. Le système de sécrétion de type 3 correspond à une lignée du système flagellaire (hypothèse b de la fig. 16). La première étape a été de perdre les gènes impliqués dans la mobilité. Ensuite, au cours de la divergence du SST3, il a perdu puis acquis des gènes qui ont permis sa diversification en 8 familles. La famille Myxo a divergé précocement suite à la perte de *sctF*. Ensuite, on observe l'acquisition de la sécrétine dans l'ancêtre commun des autres familles, suivie d'une diversification dirigée par l'adaptation aux différentes cellules hôtes.

Cependant, il existe d'autres facteurs de virulence tels que les enzymes pectinolytiques, les toxines ou encore les effecteurs des autres systèmes de sécrétion (voir encadré I).

a. Le système de sécrétion de type 3

Le système de sécrétion de type 3 (SST3) est un complexe protéique intégré dans la membrane de la cellule bactérienne. Il forme une seringue qui va permettre de libérer les effecteurs de type 3 (ET3) dans la cellule de l'hôte. Le SST3 est composé de trois structures (fig. I-17) : un corps basal, inclus dans la membrane interne et externe de la bactérie, un pilus situé dans la partie extracellulaire, et enfin des protéines, qui forment le translocon, situées dans la membrane de la cellule hôte (Izoré et al. 2011). Grâce à ce complexe fixé dans la membrane cellulaire de la cellule hôte, l'agent pathogène va injecter des protéines effectrices directement dans le cytoplasme de celle-ci (Galan & Collmer, 1999 ; Izoré et al. 2011 ; Portaliou et al. 2016). Les gènes codants cette machinerie cellulaire sont organisés en un cluster d'environ 23 gènes selon les familles de SST3 et sont régulés par un ou plusieurs régulateurs. Chez les bactéries pathogènes des animaux, le SST3 est connu pour être porté par un plasmide tandis qu'il est le plus souvent intégré dans le chromosome des bactéries phytopathogènes (Groisman & Ochman, 1996 ; Hacker et al. 1997 ; Oh et al. 2005 ; Toth et al. 2006).

Les gènes et la structure du SST3 sont semblables à ceux du flagelle. En effet, sur la fig. I-17 qui compare les structures homologues entre le flagelle et le SST3, on peut constater que la structure de ces deux complexes protéiques est conservée. L'origine exacte du SST3 est toutefois, assez débattue (McCann & Guttman, 2008). Deux hypothèses sont opposées (fig. I-18), soit le SST3 est dérivé du flagelle, soit ces deux structures possèdent un ancêtre commun qui avait potentiellement un rôle dans la mobilité et dans la sécrétion. Récemment, un modèle d'évolution du cluster du SST3 a été proposé par Abby & Rocha (2012) sur la base de l'analyse comparative de séquences nucléotidiques. Ces auteurs ont montré que le SST3 dérivait du cluster flagellaire, avec des pertes et acquisitions de gènes au cours de son histoire évolutive (fig. I-19) qui ont, en partie, provoqué la diversification en plusieurs familles du SST3. Le SST3 aurait divergé du cluster flagellaire suite à la perte de gènes ayant une fonction dans la mobilité. Une famille a divergé très tôt suite à la perte du gène *SctF*, il s'agit de la famille Myxo. Ensuite, le SST3 a acquis la sécrétine et le SST3 s'est diversifié en 8 familles : Clamy, Rhizo, Hrp1, Hrp2, Ysc, Desulfo, SPI1 et SPI2.

Ces différentes familles de SST3 sont spécifiques de l'écologie des bactéries. En effet, Hrp1, Hrp2 et Rhizo ne sont trouvés que chez des bactéries associées aux plantes, tandis que les 5 autres familles sont principalement identifiées chez

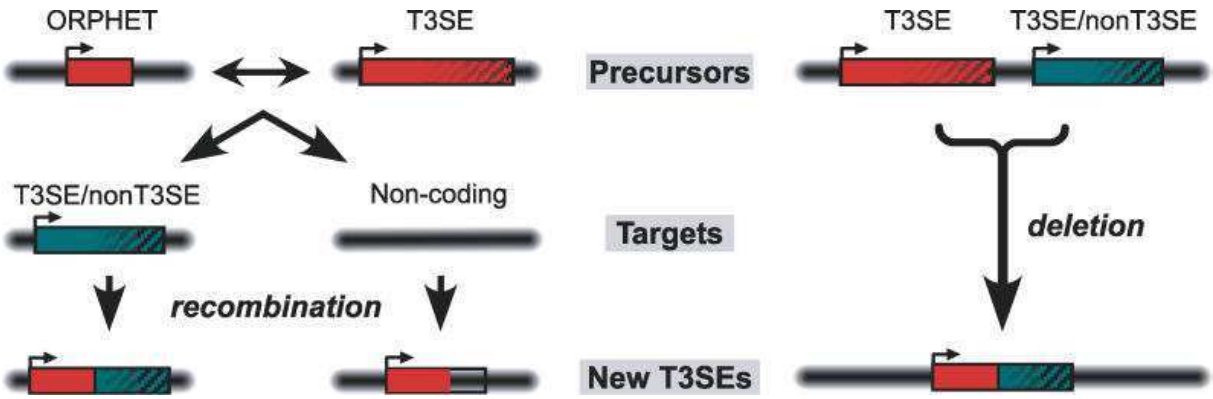


Figure I-20 : Description du mécanisme de réarrangement terminal, à l'origine de nouveaux effecteurs de type 3 (Stavrinides et al. 2006). A gauche, est représenté le cas où l'ORPHET (qui représente la séquence terminale d'un effecteur orphelin) qui contient le signal de sécrétion (représenté par la flèche) recombine avec soit des effecteurs de type 3, soit des protéines qui ne sont pas des effecteurs, soit des régions non codantes. A droite, est représenté le cas où deux effecteurs de type 3 ou un effecteur de type 3 et une protéine fusionnent ce qui génère un nouvel effecteur.

des microorganismes associés aux animaux. Bien que la transmission du SST3 soit connue pour être verticale, des évènements de transfert horizontal de gènes ont été mis en évidence au cours de son histoire évolutive. Par exemple, les deux SST3 de type SPI1 et SPI2, spécifiques des agents pathogènes animaux, ont été retrouvés chez des agents pathogènes de plantes appartenant au genre *Xanthomonas* et *Pantoea*, respectivement (Alavi et al. 2008, Marguerettaz et al. 2011, Kirzinger et al. 2015).

b. Les effecteurs de type 3, au cœur de l'interaction hôte-pathogène

Les effecteurs de type 3 (ET3) correspondent aux protéines sécrétées directement dans le cytoplasme de la cellule hôte grâce au système de SST3. Ils possèdent un signal de sécrétion situé en position N-terminale, tandis que le domaine fonctionnel est localisé au centre de la protéine et en position C-terminale (Sory et al. 1995). Certains ET3 ont pour rôle d'inhiber les défenses basales de la plante, la PTI, et d'autres ont pour rôle d'inhiber les protéines de résistance de l'hôte qui interviennent lors de l'ETI. Enfin des effecteurs spécifiques de type Transcription Activator-Like (TAL) ont pour rôle de modifier l'expression des gènes de la cellule hôte, en se liant directement à l'ADN de la cellule hôte (Boch et al. 2009 ; Moscou & Bogdanove, 2009).

L'origine de ces effecteurs peut être comparée au système de brassage d'exons chez les eucaryotes (Gilbert, 1978). Ce principe de génération de nouveaux ET3 correspond à des assemblages variés des différents domaines et est appelé " réassortiment terminal " (Stavrinides et al. 2006). Ce processus est un processus stochastique qui implique que la partie N-terminale va fusionner avec un autre ET3, une séquence codante, ou une séquence non codante (fig. I-20). Seules les combinaisons favorables sont ensuite conservées (McCann & Guttman 2007). Ensuite, la transmission des ET3 se fait par des évènements de transfert horizontal de gènes. En effet, ces gènes sont souvent associés à des éléments génétiques mobiles, tels que les plasmides, comme c'est le cas pour l'effecteur *xopE3* localisé sur un plasmide spécifique de *Xanthomonas arboricola* pv. *pruni* (Pothier et al. 2011). Chez *Ralstonia solanacearum*, 30 gènes codant des ET3 présentent un GC% plus faible que l'ensemble du génome et sont associés à des séquences d'insertion ou des prophages, ce qui constitue une caractéristique d'un transfert horizontal de gènes. (Peeters et al. 2013).

Bien que le SST3 ait un rôle dans la spécialisation à l'hôte, notamment en ce qui concerne la différenciation entre une cellule animale et une cellule végétale (Abby & Rocha, 2006), le répertoire d' ET3 détermine la spécificité d'hôte de manière plus fine, au niveau de l'espèce végétale voire même du cultivar, cela a été montré particulièrement pour l'espèce *Pseudomonas syringae* (Sakar et al. 2006; Lindeberd et al.

		Génotype de la plante	
		R	r
Génotype de l'agent pathogène	Avr	Résistance	Sensibilité
	avr	Sensibilité	Sensibilité

Figure I-21 : Le modèle gène pour gène de Flor 1971. Dans ce modèle, la résistance de la plante ne pourra être mise en place que si la plante possède le gène de résistance R, qui cible le gène d'avirulence, nommée Avr. Dans tous les autres cas, la plante sera sensible à l'infection par l'agent pathogène.

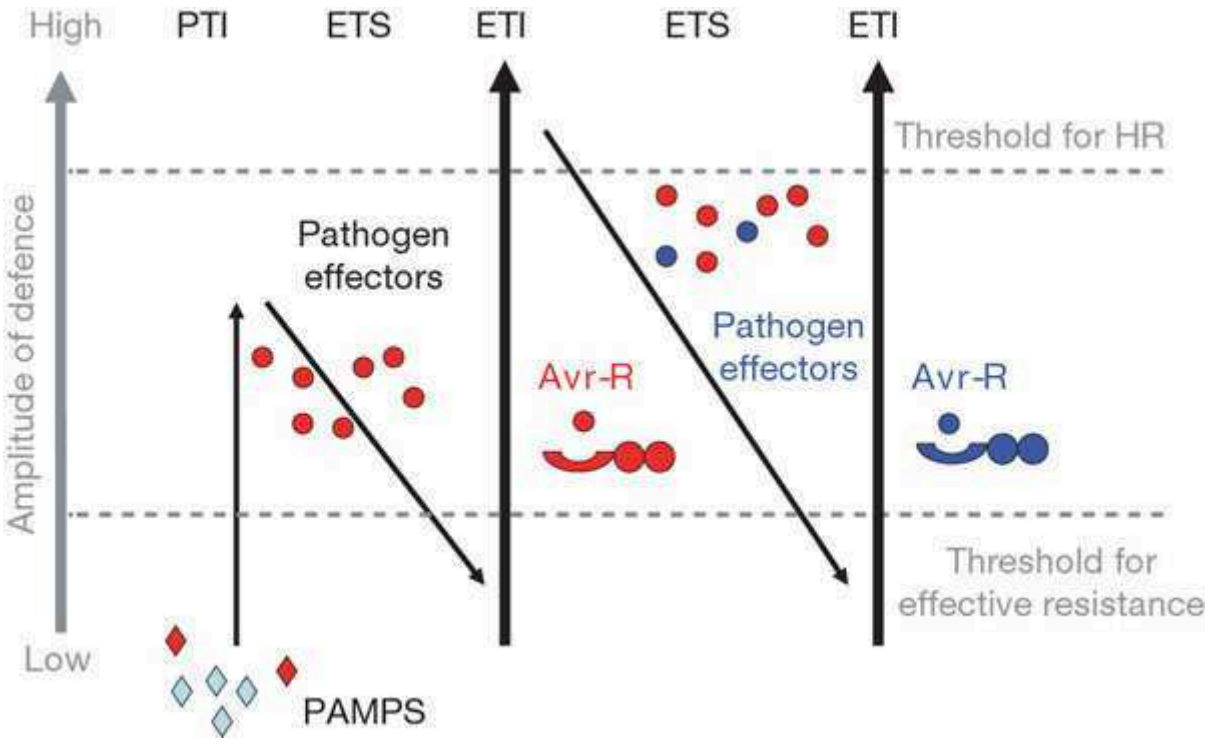


Figure I-22 : Modèle en zigzag de l'interaction plante-agent pathogène (Jones & Dangl, 2006). Le nom de ce modèle correspond aux variations d'amplitude des défenses de la plante au cours de l'infection. La reconnaissance des PAMP par la plante déclenche la PTI, qui correspond à un niveau de défense de faible amplitude. En retour, l'agent pathogène sécrète des effecteurs qui vont cibler les protéines impliquées dans la PTI, on entre alors dans une phase de sensibilité de la plante déclenchée par les effecteurs (phase ETS). Dans cette phase l'amplitude de défense de la plante diminue. Ensuite, les protéines de résistance de la plante reconnaissent les protéines d'avirulence, c'est la défense immunitaire ETI. L'amplitude des défenses de la plante est plus élevée que lors de la PTI, et cela déclenche une réaction hypersensible (HR) qui correspond à une mort cellulaire programmée des cellules végétales pour stopper l'infection.

2009), et différentes espèces appartenant au genre *Xanthomonas* : *X. axonopodis* et *X. arboricola* (Hajri et al. 2009; Hajri et al. 2012; Schwartz et al. 2015). Cependant, une analyse génomique réalisée sur 14 souches de l'espèce *Rastonia solanacearum* n'a pas permis d'établir clairement le lien entre le répertoire en ET3 propre à chaque souche et la spécificité d'hôtes (Peeters et al. 2013). Les auteurs suggèrent donc que les bases génétiques de la spécificité d'hôte de *R. solanacearum* ne reposent pas exclusivement sur les répertoires d'ET3. En effet, comme nous l'avons vu précédemment, les MCP jouent également un rôle dans cette spécificité d'hôte.

c. Défenses de la plante pendant la phase de multiplication : Le modèle en zigzag

A ce stade, l'agent pathogène se multiplie sur la plante hôte, et certains ET3 ont pour rôle d'inhiber la PTI. Cependant, un second mécanisme de défense est mis en place chez les plantes, afin d'arrêter l'infection dans les cas où l'agent pathogène a réussi à entrer dans la phase de multiplication. Cette seconde ligne de défense est nommée ETI, pour Effector-triggered immunity (Jones & Dangl, 2006). L'ETI reconnait les protéines effectrices sécrétées par la bactérie qui ciblent les défenses basales. Elle peut se déclencher soit suite à la reconnaissance directe des protéines de virulence ou suite à une reconnaissance indirecte (voir encadré II). La reconnaissance directe correspond à une relation " gène pour gène ", telle que définie par Flor en 1971 (fig. I-21). L'agent pathogène possède une protéine d'avirulence qui va être ciblée par une protéine de résistance. La résistance de la plante ne pourra s'exprimer que si la bactérie possède bien la protéine d'avirulence qui lui est associée.

Au cours de cette phase de multiplication, l'interaction hôte-pathogène va suivre une dynamique évolutive appelée le modèle en zigzag. Ce modèle d'évolution a été proposé pour expliquer en partie la coévolution entre l'agent pathogène et la plante hôte (fig. I-22) (Jones & Dangl, 2006). En effet, dans ce modèle, l'agent pathogène est reconnu par son hôte, qui déclenche alors ses défenses basales, à savoir la PTI. L'agent pathogène cible les protéines de la plante impliquées dans cette défense à l'aide des effecteurs de type 3. On observe alors une diminution des défenses de la plante. Cependant, au cours de son évolution, l'hôte va acquérir des protéines de résistance, qui vont lui permettre d'enclencher le second niveau de défense de leur système immunitaire, à savoir l'ETI. L'amplitude de l'ETI est plus importante que la PTI, car il s'agit d'une réponse spécifique à l'agent pathogène. Ce modèle en zigzag correspond donc à une alternance de phases de sensibilité et de résistance de la plante. Du fait de cette alternance, l'agent pathogène va acquérir de nouveaux ET3 pour contourner les défenses de la plante. En réalité ce modèle est plus complexe, et la distinction entre les MAMP et les ET3 n'est pas toujours

Encadré II : Modèles de reconnaissance et d'induction de l'ETI

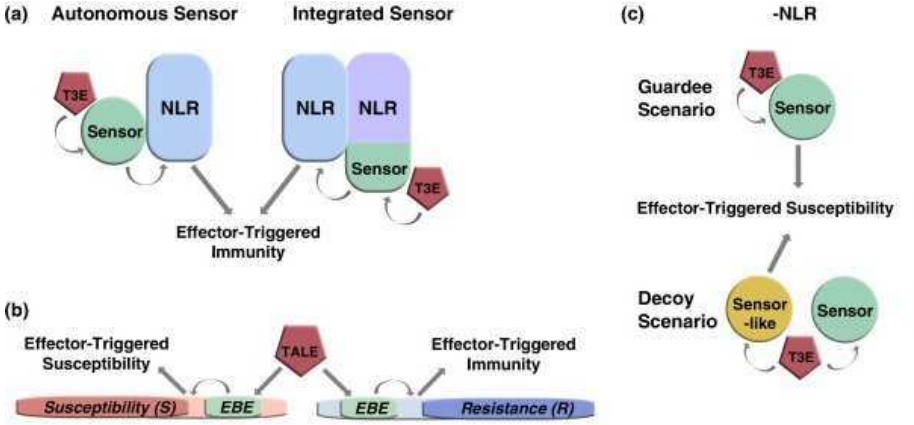


Figure représentant les différents mécanismes de perception des effecteurs de type 3 par les senseurs de la plante et conduisant à l'induction de l'ETI (d'après Khan et al. 2016)

Il existe 4 mécanismes de reconnaissance de l'effecteur de type 3 (ET3) par la plante :

Le premier mécanisme correspond à une **reconnaissance directe**, l'ET3 va se fixer à sa cible, ce qui va activer la reconnaissance du complexe par les protéines NLR qui induisent la réponse de résistance de la plante (fig. a).

Le second mécanisme correspond à **l'expression de gène de résistance induite par l'effecteur TAL** (fig. b). En effet, cet effecteur se fixe directement sur le promoteur d'un gène, il se lie donc à l'ADN chromosomique, ce qui a pour effet d'induire l'expression des gènes. Le gène exprimé correspond de manière générale à un gène de sensibilité, mais suite à des réarrangements chromosomiques dans l'ADN de l'hôte, il est possible de retrouver la cible de l'effecteur TAL dans le promoteur d'un gène de résistance.

Le troisième mécanisme correspond au **modèle de garde** (fig. c). Dans ce modèle, l'ET3 se fixe au senseur. Celui-ci change de conformation, il est alors reconnu par la protéine de résistance. Suite à cette reconnaissance, les mécanismes de défenses sont activés.

Le quatrième mécanisme correspond au **modèle de leurre** (fig. c). Selon ce modèle, la plante a acquis une protéine leurre (qui est similaire à la cible de l'ET3). Ainsi, lorsque l'ET3 se fixe sur cette protéine leurre, il n'y a pas d'expression du pouvoir pathogène. Mais cela permet à la plante de reconnaitre la bactérie et de déclencher les réactions de défenses de la plante.

très claire. En effet, les ET3 très répandus peuvent être reconnus par la plante dès la PTI, et donc correspondre à un MAMP. Les MAMP, quant à eux, peuvent avoir un rôle dans la virulence et pas uniquement dans la fitness de la bactérie (Thomma et al. 2011).

III. Le genre *Xanthomonas*

1. Morphologie, biochimie et physiologie

Les bactéries du genre *Xanthomonas* sont des bactéries à Gram-négatif. Les *Xanthomonas* possèdent un unique flagelle polaire qui leur confère une mobilité par nage dans un milieu liquide. Dans la grande majorité des cas, les colonies sont jaunes et mucoïdes. Cet aspect mucoïde est très caractéristique des *Xanthomonas*, il est le résultat de la production d'un exopolysaccharide particulier, le xanthane. Ce composé est utilisé comme épaississant dans l'industrie agroalimentaire et cosmétique. L'ensemble des souches connues dans ce genre est associé aux plantes et la majorité d'entre elles est phytopathogène. La gamme d'hôtes de ce genre est très variée puisqu'elle comprend environ 400 espèces végétales. Cependant, chaque souche possède une spécificité d'hôte très étroite.

2. Taxonomie et phylogénie du genre

Ce genre appartient à la famille des Xanthomonadaceae et à la classe des gamma-Proteobacteria. La taxonomie des espèces au sein du genre *Xanthomonas* a évolué en même temps que les méthodes permettant de définir une espèce bactérienne (voir encadré III). En effet, la première bactérie appartenant au genre *Xanthomonas* a été isolée en 1883 et était un agent pathogène des jacinthes (Wakker, 1883). Elle portait initialement le nom de *Bacterium hyacinthi*. Puis le genre *Xanthomonas* a été créé en 1939 par Dowson pour regrouper les souches présentant un unique flagelle polaire, des colonies jaunes, et le fait que la majorité des souches sont amylolitiques et produisent de l'acide à partir du lactose (Dowson, 1943). Lors de sa description initiale, le genre est divisé en 19 espèces dont par exemple, *Xanthomonas juglandis*. Puis jusque dans les années 70, on a assisté à la description de très nombreuses nouvelles espèces dans ce genre sur la base du principe qu'à chaque nouvel hôte-correspondait une nouvelle espèce. Cette stratégie " nouvel hôte – nouvelle espèce " a ainsi conduit à la description d'une centaine d'espèces dans ce genre sur la base pratiquement exclusive du caractère pathogène sur un nouvel hôte. La publication

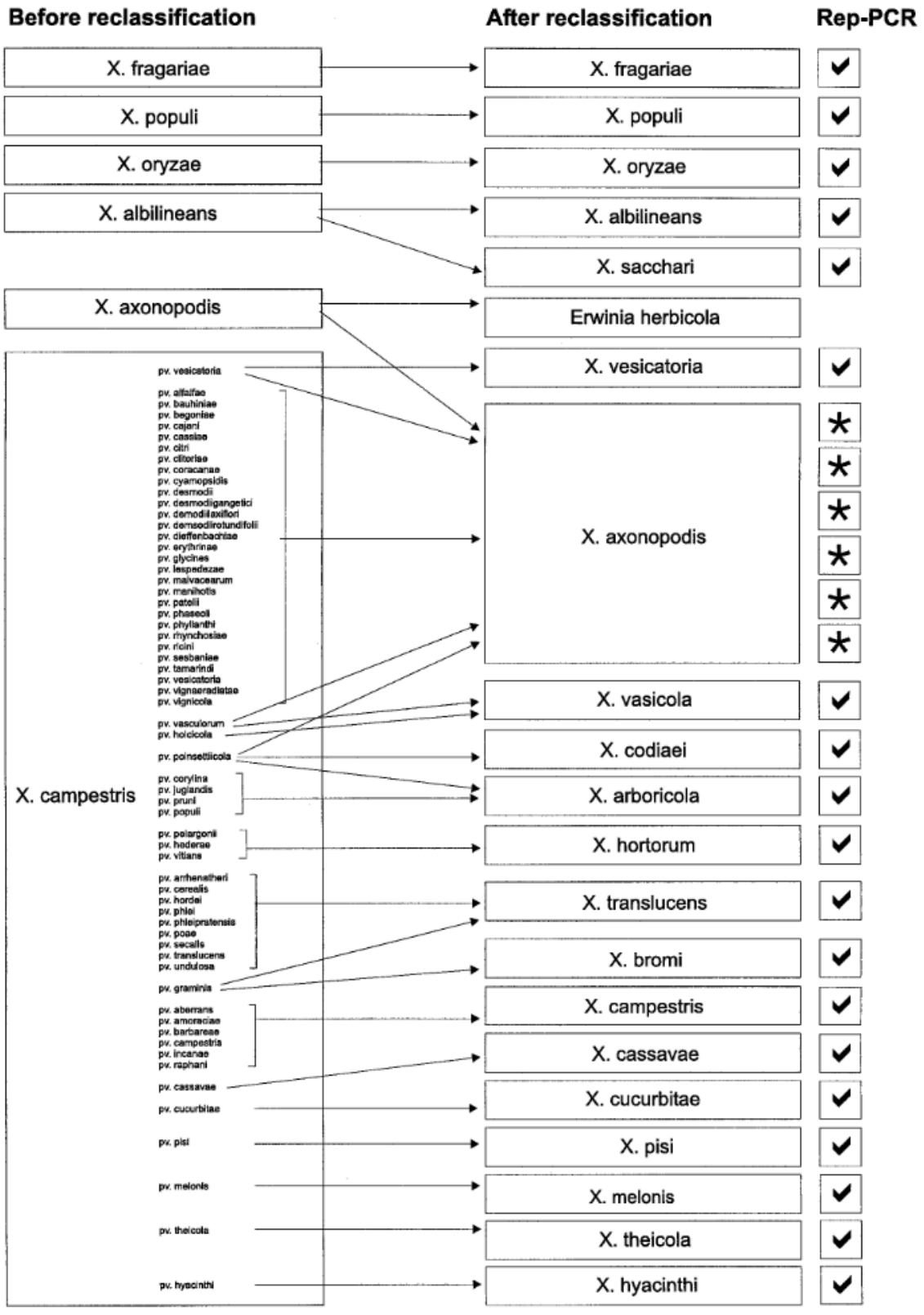


Figure I-23 : Reclassification des espèces au sein du genre *Xanthomonas* proposée par Vauterin et al. 1995 et validation par rep-PCR (Vauterin et al. 2000). A droite, est représentée la classification avant cette étude et à gauche celle proposée par les auteurs suite aux hybridations ADN:ADN effectuées sur 183 souches dans le genre et l'apport des analyses du polymorphisme d'éléments répétés (rep-PCR). Le nombre d'espèces proposées est de 20. Dans la colonne rep-PCR, les « √ » indiquent que la proposition est validée, tandis que les « * » indiquent qu'il existe une hétérogénéité génétique dans l'espèce. Le nombre de « * » indique le nombre de groupes observés au sein de l'espèce (Vauterin et al. 2000).

de l'Approved List (1980) a mis fin à cette pratique et seuls 5 noms d'espèce sont retenus : l'espèce type *X. campestris* et les espèces *X. albilineans*, *X. ampelina*, *X. axonopodis* et *X. fragariae*. Pour préparer cette révolution de la nomenclature bactérienne et éviter de perdre les noms de nombreux agents phytopathogènes, au cours des années 70, Dye et collaborateurs ont proposé de regrouper dans un niveau infra-sous-spécifique (pathovar) ces variants pathologiques (Dye & Lelliot, 1974 ; Young et al. 1978). C'est ainsi que l'espèce *X. campestris* s'est retrouvée subdivisée en 140 pathovars (Young et al. 1978). Il est également à noter que l'espèce *X. ampelina* avait été sortie de ce genre dès 1987 par Willems et collaborateurs pour former l'unique espèce d'un genre nouveau : *Xylophilus ampelinus*.

En 1995, sur la base d'un énorme travail de taxonomie formelle, Vauterin et collaborateurs (1995) ont reclassé une grande partie des pathovars et des espèces anciennement décrites au sein de ce genre *Xanthomonas* en 20 espèces (fig. I-23). Cette identification était basée sur l'hybridation ADN-ADN, méthode officielle permettant d'affilier à une même espèce l'ensemble des souches présentant des hybrides (>70%) stables (Δ Tm<5°C) (Wayne et al. 1987). Parmi ces 20 espèces, seulement 6 correspondent à des espèces déjà existantes, il s'agit de *X. albilineans*, *X. axonopodis*, *X. campestris*, *X. fragariae*, *X. populi*, et *X. oryzae*. Les 14 espèces nouvellement proposées correspondent au redéploiement de pathovars précédemment classés dans l'espèce *X. campestris*. Sept pathovars ont été élevés au rang d'espèce, il s'agit de *X. cassavae*, *X. cucurbitae*, *X. pisi*, *X. vesicatoria*, *X. hyacinthi*, *X. melonis* et *X. theicola* et de nouvelles espèces ont été créées pour regrouper un ensemble de pathovars, c'est le cas des espèces *X. arboricola*, *X. hortorum*, *X. vasicola* et *X. transluscens* (Vauterin et al. 1995 ; Vauterin et al. 2000) (fig. I-23). Des pathovars ont été subdivisés et ainsi trois espèces ont été créées *de novo*, il s'agit de *X. bromi*, *X. codiaei* et *X. sacchari*. Enfin, les contours de l'espèce type *X. campestris* ont été redéfinis au cours de cette étude puisqu'elle comportait un grand nombre de pathovars dont le positionnement phylogénétique a été revu. Sur une base phylogénétique, les auteurs proposent de ne conserver dans cette espèce qu'un petit nombre de pathovars (*aberrans*, *armoraciae*, *barbarae*, *campestris*, *incanae*, et *raphani*) qui s'avèrent tous être pathogènes de *Brassicaceae*. Cependant, cette étude n'était pas totalement exhaustive puisque 66 pathovars, précédemment décrits, n'avaient pas été pris en compte. Leur positionnement taxonomique n'était donc pas précisé, et leur nomenclature restait ambigüe puisque, formellement, ils appartenaient encore à *X. campestris*. En 2015, une nouvelle espèce a été décrite au sein du genre *Xanthomonas*, il s'agit de l'espèce *X. maliensis* (Triplett et al. 2015).

Actuellement le nombre d'espèces valides dans le genre *Xanthomonas* s'élève à 28, d'autres espèces non validées d'un point de vue nomenclatural ont récemment été proposées. En 2005, Schaad et collaborateurs ont proposé la création de trois

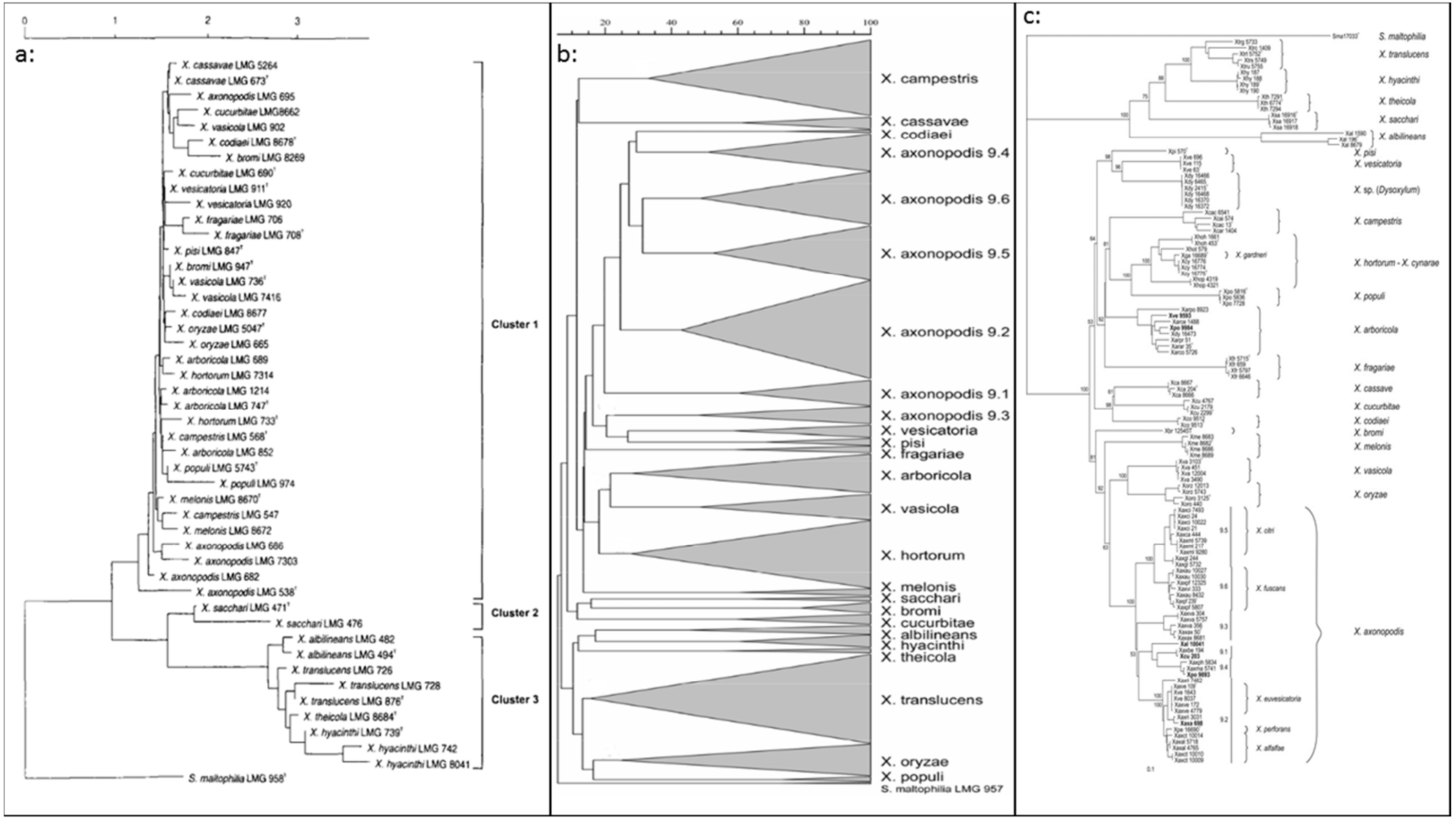
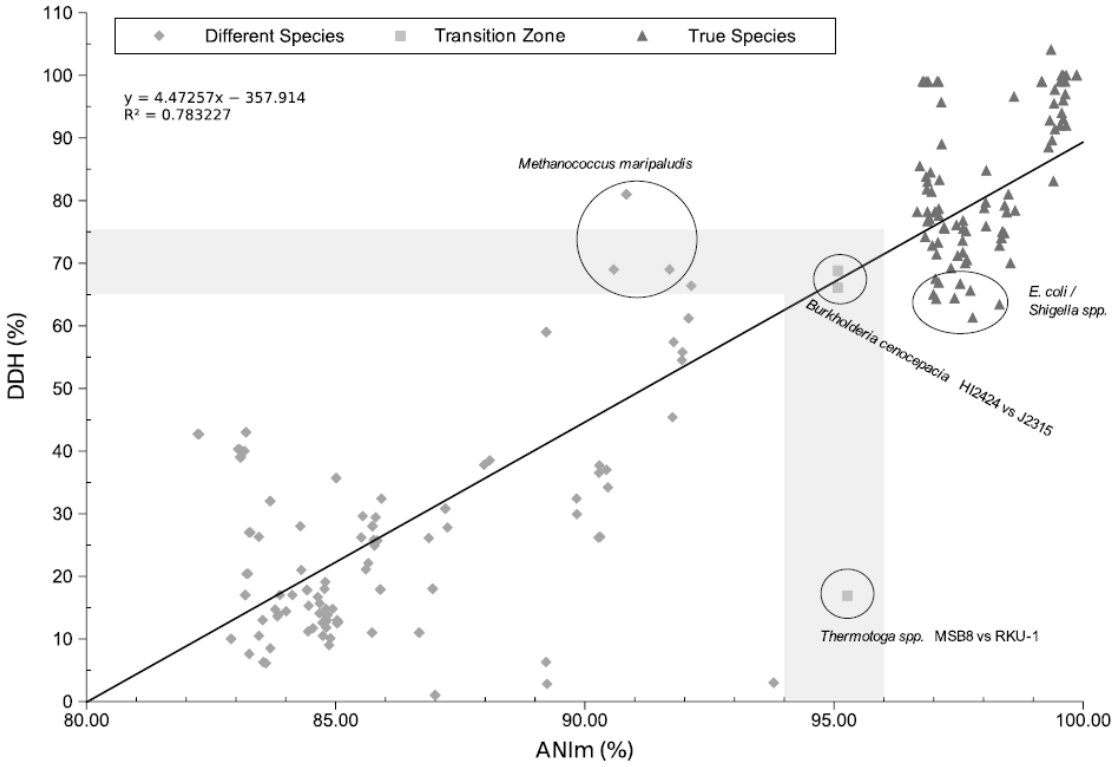


Figure I-24 : Représentations des relations phylogénétiques entre les différentes espèces du genre *Xanthomonas.* L'outgroup pour ces trois phylogénies correspond à une souche du genre *Stenotrophomonas.* a : phylogénie reconstruite en neighbor-joining en utilisant les séquences ADN du 16S (Hauden et al. 1997). b : Arbre UPGMA obtenu à partir d'une matrice de distance basée sur des profils génomiques obtenus en rep-PCR (couplant les éléments REP (Repetitive extragenic palindromic), ERIC (enterobacterial repetitive intergenic consensus) et BOX) (Rademaker et al. 2005). c : Phylogénie obtenue en neighbor-joining en utilisant les séquences partielles de 4 gènes de ménage (*dnaK*, *fyuA*, *gyrB* et *rpoD*) (Young et al. 2008)

espèces : *X. fuscans*, *X. alfalfae* et *X. citri* (initialement appelée *X. smithii*). Sur la base du polymorphisme des éléments répétés, Rademaker et collaborateurs (2005) ont proposé la division de l'espèce *X. axonopodis* en 6 groupes (9.1 à 9.6). Récemment, une proposition visant à remanier et élever la plupart de ces groupes définis au sein de l'espèce *X. axonopodis* au rang d'espèce a été publiée, mais ces nouvelles propositions ne sont pas encore valides du point de vue du Code international de la nomenclature (Constantin et al. 2016). Ainsi, les groupes 9.1 et 9.4 formeraient l'espèce *X. phaseoli*, le groupe 9.2, l'espèce *X. euvesicatoria*, le groupe 9.3, l'espèce *X. axonopodis* et les groupes 9.5 et 9.6 formeraient l'espèce *X. citri*. Bien que les espèces *X. cynarae* et *X. gardneri* seraient synonymes (Young et al. 2008), leur fusion n'a pas encore été formellement proposée. " *X. cannabis* " espèce dont le nom n'est pas encore validé, a été proposée pour regrouper les souches du pathovar *cannabis* (Jacobs et al. 2015) et la souche récemment décrite isolée de haricot (*Xanthomonas* sp. Strain Nyagatare, Aritua et al. 2015).

Les relations phylogénétiques entre les espèces du genre *Xanthomonas* ont ensuite été mises en évidence en utilisant différentes méthodologies. L'analyse basée sur une reconstruction en neighbor-joining des séquences nucléiques du gène codant l'ARNr16S (Hauben et al. 1997) divise le genre en trois groupes (fig. I-24), le groupe 1 est constitué de 15 espèces dont l'espèce type *X. campestris*, le groupe 2 constitué de l'espèce *X. sacchari* et le groupe 3 rassemble *X. albilineans*, *X. translucens*, *X. theicola* et *X. hyacinthi*. La taxonomie du genre a ensuite été revisitée en utilisant le polymorphisme des éléments répétés (analyse rep-PCR, Rademaker et al. 2005). Là encore on retrouve trois groupes dont la composition est différente de ceux obtenus avec l'ARNr16S (fig. I-24 b). Le premier à avoir divergé aurait été celui regroupant *X. albilineans*, *X. hyacinthi*, *X. theicola*, et *X. translucens*, mais également *X. oryzae* et *X. populi* ; puis un groupe comprenant *X. sacchari*, *X. bromi* et *X. cucurbitae* aurait divergé laissant ensemble les espèces *X. campestris*, *X. cassavae*, *X. melonis*, *X. codiaei*, *X. axonopodis*, *X. vesicatoria*, *X. pisi*, *X. fragariae*, *X. arboricola*, *X. vasicola* et *X. hortorum*. Les liens phylogénétiques entre ces différentes espèces ont ensuite été analysés sur la base des gènes de ménage (Parkinson et al. 2007 ; Young et al. 2008 ; Parkinson et al. 2009). Ainsi, Parkinson et al. (2007) divisent le genre en deux groupes, l'un correspond aux 5 espèces qui divergent le plus tôt dans la phylogénie (*X. theicola*, *X. translucens*, *X. hyacinthi*, *X. sacchari* et *X. albilineans*), et le second rassemble les autres espèces, il est lui-même subdivisé en 3 sous-groupes. Une étude basée sur quatre gènes de ménage (*dnaK*, *fyuA*, *gyrB* et *rpoD*) confirme la divergence entre ces deux groupes, qui sont alors nommés groupe 1 pour celui qui regroupe les 5 espèces qui divergent très tôt, et groupe 2 pour l'ensemble des autres espèces du genre (Young et al. 2008). La divergence entre les deux groupes est si importante que les auteurs suggèrent qu'il pourrait s'agir

Encadré III : Méthode de délimitation de l'espèce bactérienne



Graphique représentant la correspondance entre les pourcentages d'hybridation ADN-ADN (en ordonnée) et les ANI (identités nucléotidiques moyennes) (en abscisse). La zone grise représente le seuil au-dessus duquel les souches comparées appartiennent à la même espèce (Konstantinidis & Tiedje, 2005).

La définition de l'espèce bactérienne est basée sur la distance génétique entre individus. Ainsi, l'espèce procaryote correspond à un ensemble d'individus ayant un pourcentage d'hybridation ADN-ADN supérieur à 70 % avec un ΔTm inférieur à 5°C et partageant des caractères phénotypiques permettant de les différentier des espèces voisines (Stackebrandt et al. 2002). Les hybridations ADN-ADN sont difficiles à réaliser en routine. Les avancées en génomique ont permis de donner un équivalent aux hybridations ADN-ADN basé sur le pourcentage d'identité nucléotidique moyen (ANI) calculé *in silico* sur les séquences génomiques entières. Basée sur une analyse bibliographique de souches pour lesquelles les valeurs d'ANI et d'hybridation ADN-ADN étaient disponibles, Konstantinidis & Tiedje, 2005 ont montré que la valeur équivalente aux 70 % d'hybridation ADN-ADN correspond à une valeur d'ANI de 95 % (voir figure). Les ANI ont été utilisés pour définir les genomospecies chez différents agents phytopathogènes tels que *Pseudomonas* (Scortichini et al. 2013), mais également pour définir une nouvelle espèce chez *Xanthomonas*, *X. maliensis* (Triplett et al. 2015), qui n'est constituée jusqu'à présent que d'individus commensaux.

de deux genres différents. Cette étude est la dernière à avoir étudié de manière exhaustive les relations phylogénétiques entre toutes les espèces du genre (fig. I-24). La phylogénie basée sur sept gènes de ménage a montré que l'espèce *X. maliensis* se place à la racine du groupe 2 du genre *Xanthomonas* (Triplett et al. 2015). D'autres études phylogénomiques ont ensuite été réalisées, en incluant, cependant, un nombre limité d'espèces (Rodriguez et al. 2012 ; Naushad et al. 2015). Ces études ont confirmé la séparation du genre *Xanthomonas* en deux groupes bien distincts, mais ont également montré que d'autres genres bactériens (*Xylella* et *Stenotrophomonas*) se plaçaient entre ces deux groupes. Dans leur étude sur l'ordre des Xanthomonadales basée sur les séquences de 15 gènes de ménage et ribosomaux, Naushad et collaborateurs (2015) montrent que le plus proche voisin phylogénétique du groupe 1 des *Xanthomonas* est *Xylella*, alors que celui du groupe 2 est *Stenotrophomonas*.

Les souches et espèces commensales

Des souches appartenant au genre *Xanthomonas* sans pouvoir pathogène avéré ont été isolées à partir de différentes espèces végétales ; ces souches sont le plus souvent appelées " souches non-pathogènes ". Cependant, cette dénomination n'est pas satisfaisante car elle suppose que ces souches, dont l'écologie est encore très mal connue et qui sont très peu étudiées ne sont pathogènes sur aucune espèce végétale. Cette hypothèse étant difficile à tester, j'ai pris le parti d'adopter la dénomination " souches commensales ". Le commensalisme est défini comme une interaction hôte-microorganisme qui ne produit pas de dommages perceptibles ou persistants chez l'hôte (Casadevall & Pirofski, 2000). Le terme " commensal " peut désigner les microorganismes qui peuvent s'établir sur un hôte sans causer de dommages (White & Timbury, 1973), les organismes de la flore normale d'un hôte (Stewart, 1968) ou des agents pathogènes inoffensifs (Park & Williams, 1917).

La première étude comportant un grand nombre de souches commensales au sein du genre *Xanthomonas* a été réalisée en 1996 par Vauterin et al. En effet, la collection de souches de cette étude s'élevait à 70 individus, isolés d'un grand nombre d'espèces végétales ayant un intérêt agronomique ou ornemental. L'objectif de cette étude était d'identifier les espèces auxquelles appartenaient ces souches commensales, en comparant les profils protéiques et d'acides gras. Au moment où cette étude a été réalisée, le genre *Xanthomonas* était alors composé de 20 espèces (celles définies par Vauterin et al. 1995), ce qui correspond à la base de la taxonomie actuelle du genre. Parmi ces 70 souches, l'identification des espèces auxquelles elles appartenaient ont pu être faites pour 42 souches car les profils protéiques et d'acides gras donnaient un résultat exploitable et connu. Ces 42 souches ont finalement été assignées majoritairement aux deux espèces : *X. arboricola* et *X. axonopodis*.

Trois cas de figures ont empêché l'identification des 28 autres souches, soit aucune des deux techniques ne permettait d'assigner l'individu à une espèce connue, soit une des deux techniques ne permettait pas de l'assigner à une espèce, soit les deux techniques donnaient des résultats différents.

La nouvelle espèce *X. maliensis* est composée exclusivement de souches commensales (Triplett, et al. 2015). Les 4 souches connues aujourd'hui pour appartenir à cette espèce ont été isolées soit du riz (*Oryza sativa*) soit d'une graminée (*Ischaemum rugosum*). Ces souches sont capables de croitre sur leur hôte d'isolement sans provoquer de symptômes. Par ailleurs, lors d'une étude épidémiologique sur la résistance au cuivre des souches de *X. arboricola* pv. *juglandis*, menée dans les deux bassins français de production de la noix entre 2001 et 2009, un grand nombre de souches commensales appartenant à l'espèce *X. arboricola* a été isolé de noyer (Essakhi et al. 2015). De même, des souches appartenant au genre *Xanthomonas* ont été isolées de semences de haricot, mais après ré-inoculation, aucun symptôme n'a pu être obtenu, ces souches ont donc été considérées comme commensales (Boureau et al. 2013 ; Grimault et al. 2014). De manière générale, les souches sont génétiquement diverses et se positionnent dans différentes espèces du genre *Xanthomonas* : *X. campestris*, *X. axonopodis* et *X. arboricola*. (Guimbaud, 2012). Elles peuvent également former une nouvelle espèce, comme c'est le cas pour *X. maliensis*.

3. Caractéristiques génomiques des *Xanthomonas*

La taille moyenne des génomes de souches appartenant au genre *Xanthomonas* est de 5 Mb, seule une espèce possède un génome de taille réduite comparativement aux autres espèces, il s'agit de *X. albilineans* qui appartient au groupe 1 (environ 3Mb) (Pieretti et al. 2009). Actuellement, environ 400 génomes sont séquencés et disponibles sur des bases de données publiques (Jacques et al. 2016b). Cependant, certaines espèces ou pathovars sont plus représentés que d'autres. Le nombre de gènes au sein d'un génome est d'environ 4 000 et leur longueur moyenne d'environ 1 000 pb. Il existe une hétérogénéité du nombre de gènes au sein d'une même espèce. Par exemple, chez *X. campestris* la comparaison de 13 génomes a montré que le nombre de gènes pouvait aller de 3481 à 4290 (Roux et al. 2015). Le core génome de 9 souches appartenant aux espèces *X. campestris* (4 souches) et *X. citri* (5 souches) est composé de 2851 gènes orthologues (Huang et al. 2015). Le core génome de l'espèce *X. arboricola* basé sur 18 souches est de 2525 gènes (Garita-Cambronero et al. 2016a). Ces chiffres sont susceptibles de diminuer si l'ensemble des espèces du genre sont incluses dans l'analyse.

Au sein des génomes du genre *Xanthomonas*, on peut retrouver des éléments

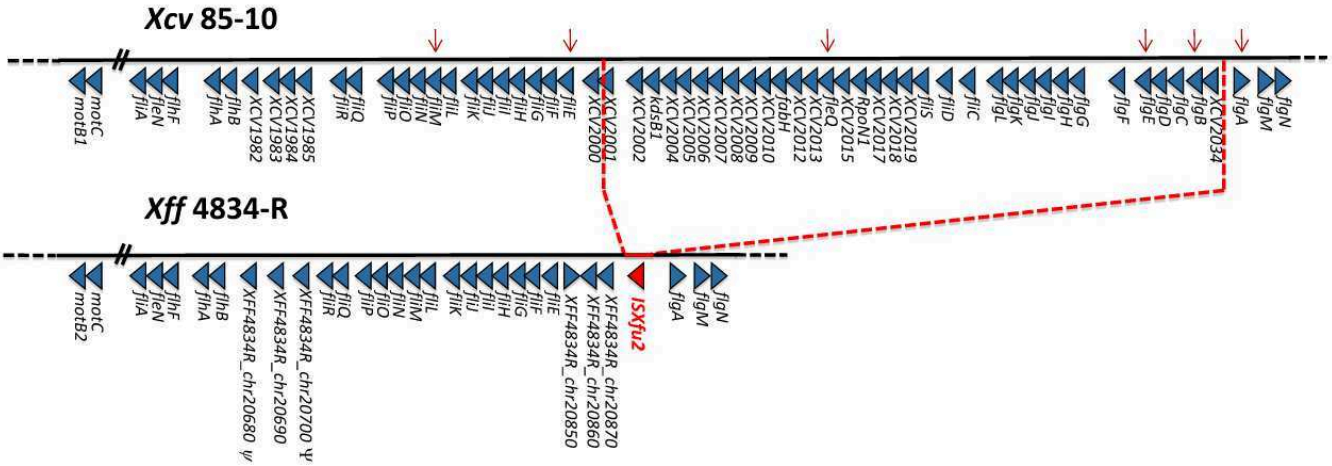


Figure I-25: Schéma représentant la délétion des gènes dans le cluster flagellaire de la souche CFBP 4834, appartenant à l'espèce *X. fuscans* pv. *fuscans* (*Xff*) comparé à la souche 85-10 appartenant à *X. campestris* pv. *vesicatoria (Xcv)*. Chaque flèche bleue représente les gènes du cluster. Les pointillés en rouge représentent l'ensemble des gènes délétés dans la souche de *Xff*. Enfin le gène coloré en rouge correspond à la transposase retrouvée à la place des gènes délétés.

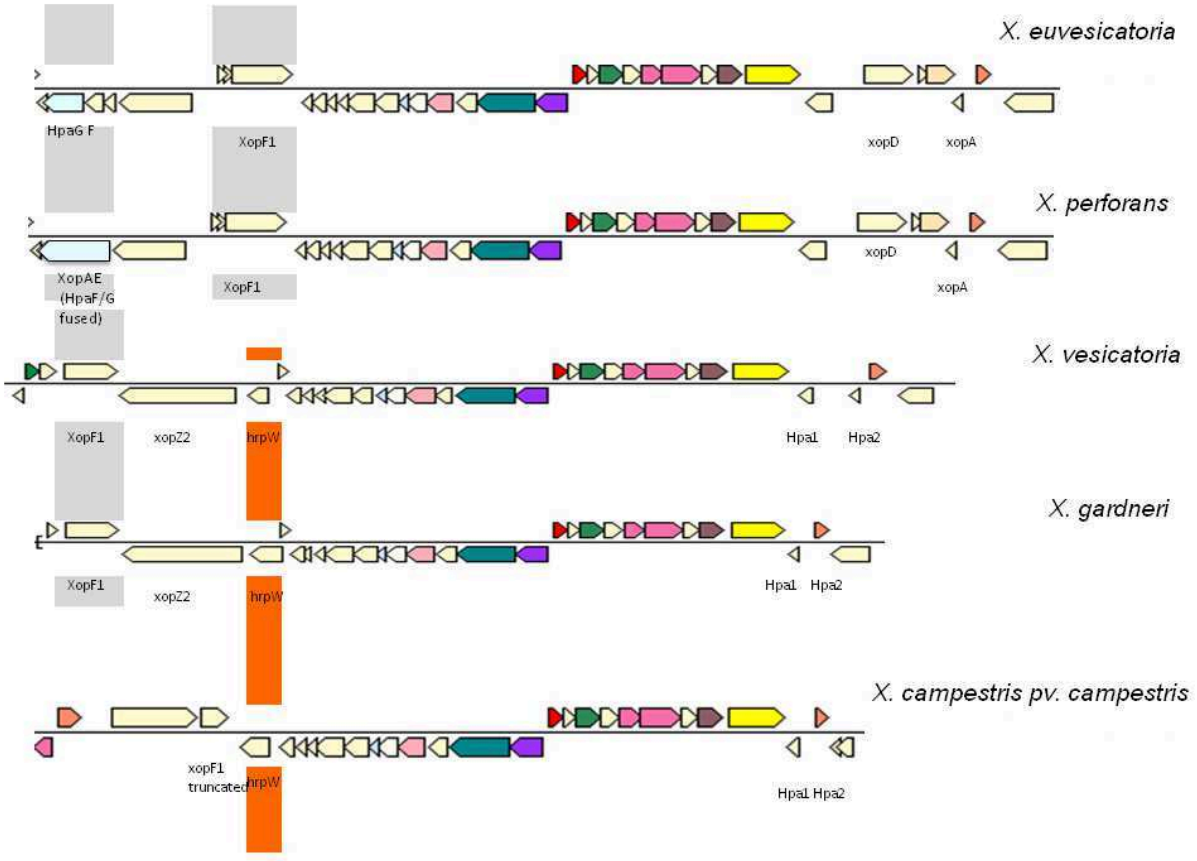


Figure I-26 : Comparaison du cluster de gènes du système de sécrétion de type 3 (et des effecteurs) situés à proximité chez 5 espèces de *Xanthomonas*. Chaque flèche représente un gène et les flèches de la même couleur entre les 5 souches montrent qu'ils sont homologues entre les souches (d'après Potnis et al. 2011).

génétiques mobiles, comme des plasmides, des éléments intégratifs conjugatifs (ICE), des séquences d'insertion (IS), des bactériophages ou des intégrons. Ainsi, un intégron aurait été acquis dans l'ancêtre commun du genre, et maintenu au cours de la diversification en espèces (Gillings et al. 2005). Cet intégron serait désormais inactif. La comparaison des éléments génétiques mobiles entre deux souches de *Xanthomonas* (*X. axonopodis* et *X. campestris*) et 4 souches de *Xylella* a montré que le nombre d'IS était plus important dans les génomes de *Xanthomonas* que chez les xylelles (44 et 73 IS pour les génomes de *Xanthomonas* et 1 ou 0 pour les génomes de *Xylella*) (Monteiro-Vitorello et al. 2005). L'espèce *X. oryzae* pv. *oryzae* est caractérisée par un très grand nombre d'IS (611 pour le génome de la souche MAFF311018) représentant 10% du génome (Ochiai et al. 2005). Un nombre limité de plasmides a été retrouvé au sein des génomes de *Xanthomonas* (de 2 à 4 en fonction des souches). Le plasmide pXOCgx01 décrit chez le pathovar *oryzicola* de *X. oryzae* est composé de 64 gènes, ce plasmide est similaire à pXAC64 décrit chez *X. axonopodis* pv. *citri* (Niu et al. 2015). Un autre plasmide porteur d'une résistance à la streptomycine a été décrit chez *X. campestris* pv. *vesicatoria* (Minsavage et al. 1990). Enfin, un plasmide a été décrit chez l'espèce *X. arboricola*, pXap41, ce plasmide est spécifique du pathovar *pruni*, agent pathogène des *Prunus*, et porte des gènes codant des ET3 (Pothier et al. 2011). Un ICE a été décrit chez *X. arboricola* pv. *juglandis*, il représente un ilot génomique de 94.8 kb et contient des gènes impliqués dans la résistance au cuivre (Cesbron et al. 2015). Cette étude a également permis de détecter des prophages au sein des génomes de souches pathogènes (Cesbron et al. 2015).

La comparaison génomique de deux souches de *Xanthomonas*, *X. axonopodis* pv. *citri* et *X. campestris* pv. *campestris* a montré que les gènes impliqués dans la biogenèse du flagelle et le chimiotactisme était organisé en 4 clusters de gènes, d'environ 150kb (DaSilva, et al. 2002). L'étude du cluster flagellaire chez *X. campestris* pv. *campestris* a montré que ce cluster était composé d'environ 40 séquences codantes (Hu et al. 2005). Ce cluster n'est cependant pas conservé chez toutes les souches du genre, en effet, des délétions de gènes au sein de ce cluster ont été observées chez plusieurs pathovars et espèces (Darasse et al. 2013). Ces délétions ont été reliées à l'absence de mobilité (Darasse et al. 2013) (fig. I-25).

Chez les *Xanthomonas*, le principal SST3 est de type Hrp2 (fig. I-26). La comparaison du cluster entre 5 espèces du genre a montré que l'organisation du cluster est conservée entre les souches et que des gènes codants des ET3 étaient localisés dans l'environnement proche de ce cluster (Potnis et al. 2011). Cependant, un SST3 de type SPI-1 a été retrouvé chez l'espèce du groupe 1, *X. albilineans* (Marguerettaz et al. 2011). De plus, l'absence d'un cluster de type Hrp a été rapportée pour des souches appartenant à l'espèce " *X. cannabis* " (Jacobs et al.

Figure I-27 : Inoculation de souches appartenant à l'espèce *X. arboricola* sur un milieu succinate-quinate. Le halo vert autour de l'inoculum montre la métabolisation du quinate, caractéristique de cette espèce.

2015), au pathovar *X. arboricola* pv. *fragariae* (Vandroemme et al. 2013) et à des souches commensales appartenant à l'espèce *X. arboricola* (Essakhi et al. 2015 ; Cesbron et al. 2015). Environ 50 ET3 ont été décrits dans le genre *Xanthomonas* (http://www.xanthomonas.org/). Parmi toutes les espèces de *Xanthomonas*, 12 possèdent des TAL effecteurs (Transcription Activator-Like) qui sont des effecteurs spécifiques (Jacques et al. 2016b). En effet, ils se lient au promoteur de gènes de la cellule végétale et permettent l'expression de ces gènes (Moscou & Bogdanove, 2009 ; Boch et al., 2009).

IV. *Xanthomonas arboricola* et ses hôtes

Le caractère phénotypique qui regroupe l'ensemble des souches appartenant à cette espèce est la capacité à métaboliser le quinate (Lee et al. 1992) (fig. I-27). Cette espèce est actuellement composée de 9 pathovars et de souches pathogènes du piment identifiées en 2010 par Myung et al, et sur *Zizyphus jujuba* (Myung et al. 2010) et *Vitis vinifera* (Sawada et al. 2011). Cette espèce comprend également des souches non pathogènes sur leur hôte d'isolement, autrement dit commensales.

1. *Xanthomonas arboricola* pv. *pruni*

Les souches appartenant à ce pathovar sont responsables de bactérioses sur les arbres fruitiers à noyau, tel que l'abricotier, le pêcher ou l'amandier dont les symptômes forment des tâches anguleuses sur feuille, donnant un aspect de tache d'huile. L'identification de cet agent pathogène a été faite en 1903 par Smith, en Amérique du Nord. Depuis cette première identification, ce pathovar semble fortement implanté et dispersé en Chine, Afrique du Sud et Uruguay, alors que sa présence ne serait que localisée en Bulgarie, Roumanie, Moldavie, Russie, Ukraine, inde, Pakistan, Japon, Corée, Canada, USA, Mexique, Argentine, Brésil et Australie (EPPO/OEPP, 2006). La détection de la bactérie en Europe a été faite pour la première fois en 1934, en Italie et Slovénie (Petri, 1934). Cette bactérie a provoqué une maladie au fort impact économique en Italie du milieu des années 1970 à la fin des années 1980 (Stefani, 2010). En Italie, cette épidémie serait consécutive à l'introduction d'un nouveau cultivar de prunier japonais (*Prunus salicina* cv. calita) qui s'est avéré extrêmement sensible à *X. arboricola* pv. *pruni* sous les conditions climatiques du nord de l'Italie. En Espagne, les symptômes caractéristiques de cette

Figure I-28 : Photos de symptômes du chancre vertical suintant sur le noyer (*Juglans regia*) induits par l'agent pathogène *X. arboricola* pv. *juglandis* (Hajri et al. 2010)

bactériose ont été observés entre 2006 et 2009 (Palacio-Bielsa, 2010). En France, cette maladie est devenue économiquement dommageable dès la fin des années 90. Grâce à l'étude d'une collection de 64 souches sur la base des séquences de l'ITS et de quatre gènes de ménage, Boudon et al. (2005) ont montré que les souches françaises seraient originaires d'Amérique du Nord et d'Italie.

Les hôtes des souches appartenant à ce pathovar sont les arbres du genre *Prunus*. Ce genre appartient à la famille des Rosacées. Le centre d'origine de ces arbres fruitiers est en Asie. En effet, une étude de biogéographie et de datation moléculaire a permis de mettre en évidence que ce genre serait apparu il y a 61 millions d'années dans l'Est de l'Asie, puis se serait diversifié au début de l'Eocène (Chin et al. 2014). Le début de la domestication dans ce genre aurait commencé il y a environ 4 000 ans avec la domestication des pêchers (Faust & Timon, 1995). Ensuite, le pêcher a été introduit en Europe puis aux Etats-Unis entre le 16ème et le 19ème siècle (Verde et al. 2013). La Californie représente désormais l'un des plus grands centre de production de pêches (DeJong et al. 1999) et d'abricots (Zaurov et al. 2013).

2. *Xanthomonas arboricola* pv. *juglandis*

L'identification des souches appartenant à ce pathovar a été faite pour la première en 1901, par Pierce, en Californie. Ensuite, des souches appartenant à ce pathovar ont été détectées dans plusieurs pays comme la Nouvelle-Zélande, l'Angleterre, la Roumanie, l'Iran, la France, la Hollande, la Hongrie, la Grèce, l'Espagne, le Portugal et l'Italie (Scortichini et al. 2000). Les symptômes de la bactériose du noyer sont des taches foliaires et des nécroses. Une partie importante de la lutte contre cet agent pathogène repose sur l'utilisation de composés cupriques et immanquablement la résistance au cuivre est apparue chez cet agent pathogène (Scortichini et al. 2000). En 2010, une nouvelle lignée appartenant à ce pathovar a été mise en évidence en France (Hajri et al. 2010). Cette lignée regroupe l'ensemble des souches qui sont responsables d'un chancre vertical suintant sur le bois des arbres. Il existe donc deux types de symptômes causés par cet agent pathogène, d'une part les symptômes de la bactériose classique et ceux du chancre vertical suintant (fig. I-28). Il a été supposé que l'acquisition de ce nouveau trait par les souches capables de provoquer le chancre vertical suintant serait dû à l'acquisition d'un ET3, *xopB* (Hajri et al. 2012). En effet, l'étude des répertoires d'effecteurs de type 3 a montré qu'au sein du pathovar *juglandis*, seules les souches capables de provoquer un chancre sur bois possédaient cet effecteur. Cette hypothèse a également été émise par Cesbron et al. 2015 suite à la comparaison des génomes d'une souche provoquant

a :

b :

Figure I-29 : Photos de symptômes de la bactériose du noisetier (*Corylus avellana*) induits par la bactérie *X. arboricola* pv. *corylina*. a : symptômes sur feuille ; b : symptômes sur tronc (Lamichhane & Varvaro, 2014)

un chancre et d'une autre provoquant une bactériose. L'analyse de la présence des effecteurs a montré que *xopB* était présent uniquement chez la souche responsable du chancre, et qu'une intégrase était localisée à proximité, suggérant un transfert horizontal de gènes.

Les souches du pathovar *juglandis* sont des agents pathogènes des noyers, *Juglans regia* et *Juglans nigra*, ce dernier étant cultivé principalement pour le bois (Stanford et al. 2000). Ces deux espèces ont un fort impact économique, principalement en Chine et aux Etats-Unis (Gunn et al. 2010). L'aire d'origine à partir de laquelle a été domestiqué *Juglans regia* serait probablement l'Iran ou l'Afghanistan (Vahdati et al. 2015). A partir de cette aire, cette espèce a été introduite dans le reste de l'Asie, en Europe et en Amérique du Nord. L'aire d'origine de l'espèce *Juglans nigra* serait l'Est des Etats Unis et le sud du Canada (Woeste & Michier, 2011). Cette espèce est également cultivée en Europe (Kremer et al. 2008).

3. *Xanthomonas arboricola* pv. *corylina*

La première identification de l'agent pathogène responsable de la bactériose du noisetier a été réalisée en 1913 par Barss dans l'Oregon en Amérique du Nord. Aujourd'hui, la répartition de ce pathovar est mondiale (Scortichini et al. 2002). En effet, le nombre de pays touchés par la bactériose du noisetier était de 13 dans les années 1990 et de 30 en 2010 (Lamichhane & Varvaro, 2014). La bactérie provoque des symptômes sur les feuilles, les bourgeons, les branches, le tronc et parfois les fruits (fig. I-29) (Lamichhane & Varvaro, 2014). Les souches du pathovar *corylina* sont pathogènes sur le noisetier, *Corylus avellana*. Il semblerait que sa domestication soit issue de trois processus indépendants, dans les pays méditerranéens (Italie et Espagne), en Turquie et en Iran (Boccacci & Botta, 2009) qui correspondent aux refuges glaciaires de cette espèce et à son aire d'origine (Palmé & Vendramin, 2002). Les trois plus grands pays producteurs de noisettes sont la Turquie, l'Italie et les Etats-Unis, plus principalement l'Oregon qui fournit 99% de la production américaine (Mehlenbacher & Olsen, 1997 ; Cristofori et al. 2014).

Les pathovars *pruni* et *corylina* sont règlementés comme organismes nuisibles de quarantaine[5] en Europe et sont classés dans la liste d'alerte A2 de l'OEPP qui regroupe les organismes déjà présents sur le territoire ; c'est également le cas du pathovar *juglandis* dans de nombreux pays.

[5]Un organisme de quarantaine correspond à un organisme nuisible qui a une importance potentielle pour l'économie de la zone menacée et qui n'est pas encore présent dans cette zone, ou bien qui y est présent, mais à distribution restreintre, et faisant l'objet d'une lutte officielle (FAO, 1990 ; FAO 1995).

4. Les pathovars *arracaciae*, *guizotiae* et *zantedeschiae*

Ces trois pathovars ont été récemment reclassés au sein de l'espèce *X. arboricola* (Fischer-Le Saux et al. 2015). Les souches du pathovar *arracaciae* sont pathogènes sur l'arracacha (la pomme de terre-céleri), plante originaire des Andes et cultivée en Amérique du Sud. La première identification de cet agent pathogène a été faite en 1971 par Pereira et al. au Brésil où il fait partie des agents pathogènes bactériens les plus importants avec les *Erwinia spp.* (Henz, 2002). Les souches du pathovar *guizotiae* sont pathogènes sur *Guizotia abyssinica*, le nyger (Yirgou, 1964). Cette plante est cultivée en Inde et en Ethiopie pour ses graines et son huile. Des introductions en Europe comme plante de couvert ont également été récemment tentées. Enfin, le pathovar *zantedeschiae* est l'agent pathogène de l'Arum d'Ethiopie, qui est une plante ornementale (Joubert & Truter, 1972). Il a été décrit pour la première fois en Afrique du Sud, pays d'origine de cette plante et à Taiwan en 2004 (Lee et al. 2004).

5. *Xanthomonas arboricola* pv. *populi*

Ce pathovar a été décrit en 1983 par De Kam pour regrouper des souches pathogènes sur peuplier (*Populus spp.*), de couleur jaune sur milieu Nutrient Agar et poussant à 30C ; c'est-à-dire divergentes par rapport à un autre agent pathogène du peuplier déjà bien connu *X. populi* (souches de couleur blanche dont la température optimale de croissance est 18C). Ces souches initialement incluses dans *X. campestris* ont été transférées dans *X. arboricola* pv. *populi* par Vauterin et al. 1995. Bien que l'hybridation ADN-ADN entre la souche type de l'espèce *X. arboricola* et les souches du pathovar *populi* était de 57%, soit en dessous des 70% du seuil de l'espèce, les auteurs ont placé ce pathovar au sein de *X. arboricola*. Cette divergence a également été observée lors d'analyses phylogénétiques de l'espèce basée sur un ou sept gènes de ménage (Hajri et al. 2011 ; Fischer-Le Saux et al. 2015). L'agressivité des souches de ce pathovar est faible et des conditions environnementales particulières sont supposées être nécessaires pour que se développent des infections par cette bactérie qualifiée d'opportuniste (Haworth & Spiers, 1992). Ces auteurs considèrent cette maladie comme non préoccupante.

6. *Xanthomonas arboricola* pv. *celebensis*

Les souches du pathovar *celebensis* ont été isolées en Indonésie du bananier appartenant au genre *Musa* (Gäumann, 1921). Cet agent pathogène semble présenter une faible incidence puisqu'aucun rapport ultérieur ne le mentionne et que seules deux souches appartenant à ce pathovar sont disponibles dans les collections de bactéries associées aux plantes (www.straininfo.net). Au sein du genre *Xanthomonas*, un autre pathovar est responsable de maladie sur le bananier, il s'agit de *X. vasicola* pv. *musacearum*. Une comparaison de la pathogénie des souches pathogènes sur banane, maïs et canne à sucre a été réalisée (Karamura et al. 2015). Les souches de *X. vasicola* pv. *musacearum* sont très agressives sur banane et canne à sucre alors que dans les mêmes conditions aucun symptôme ne s'est développé suite à l'inoculation par une souche de *X. arboricola* pv. *celebensis*. Les auteurs ont supposé que cette absence de symptômes pouvait être consécutive à une conservation inadéquate de la souche.

7. *Xanthomonas arboricola* pv. *fragariae*

Ce pathovar a été décrit par Janse en 2001. Les souches appartenant à ce pathovar seraient pathogènes sur fraisier mais les symptômes n'ont été observés que plus d'un mois après inoculation. Par ailleurs, en 2013, une étude génomique et de pouvoir pathogène a permis de mettre en évidence que les souches appartenant à *X. arboricola* pv. *fragariae* ne provoquaient pas de symptômes sur le fraisier, leur hôte d'isolement, contrairement aux souches de l'espèce *X. fragariae*, dont l'hôte est également le fraisier (Vandroemme et al. 2013). Ces auteurs suggèrent que ces souches appartiendraient à la flore microbienne communément retrouvée sur fraisier.

Ainsi, la pathogénie des trois pathovars *fragariae*, *populi* et *celebensis* peut être remise en cause, l'expression de symptômes n'étant apparemment possible que dans des conditions exceptionnelles pour le pathovar *populi* (Haworth & Spiers, 1992), et difficile sinon impossible à reproduire dans le cas des pathovars *fragariae* et *celebensis* (Vandroemme et al. 2013 ; Karamura et al. 2015).

8. Les souches pathogènes n'appartenant pas à un pathovar

En 2007, des souches pathogènes sur poivron ont été détectées en Corée (Myung et al. 2010). La phylogénie réalisée avec le gène de ménage *gyrB* a montré que ces

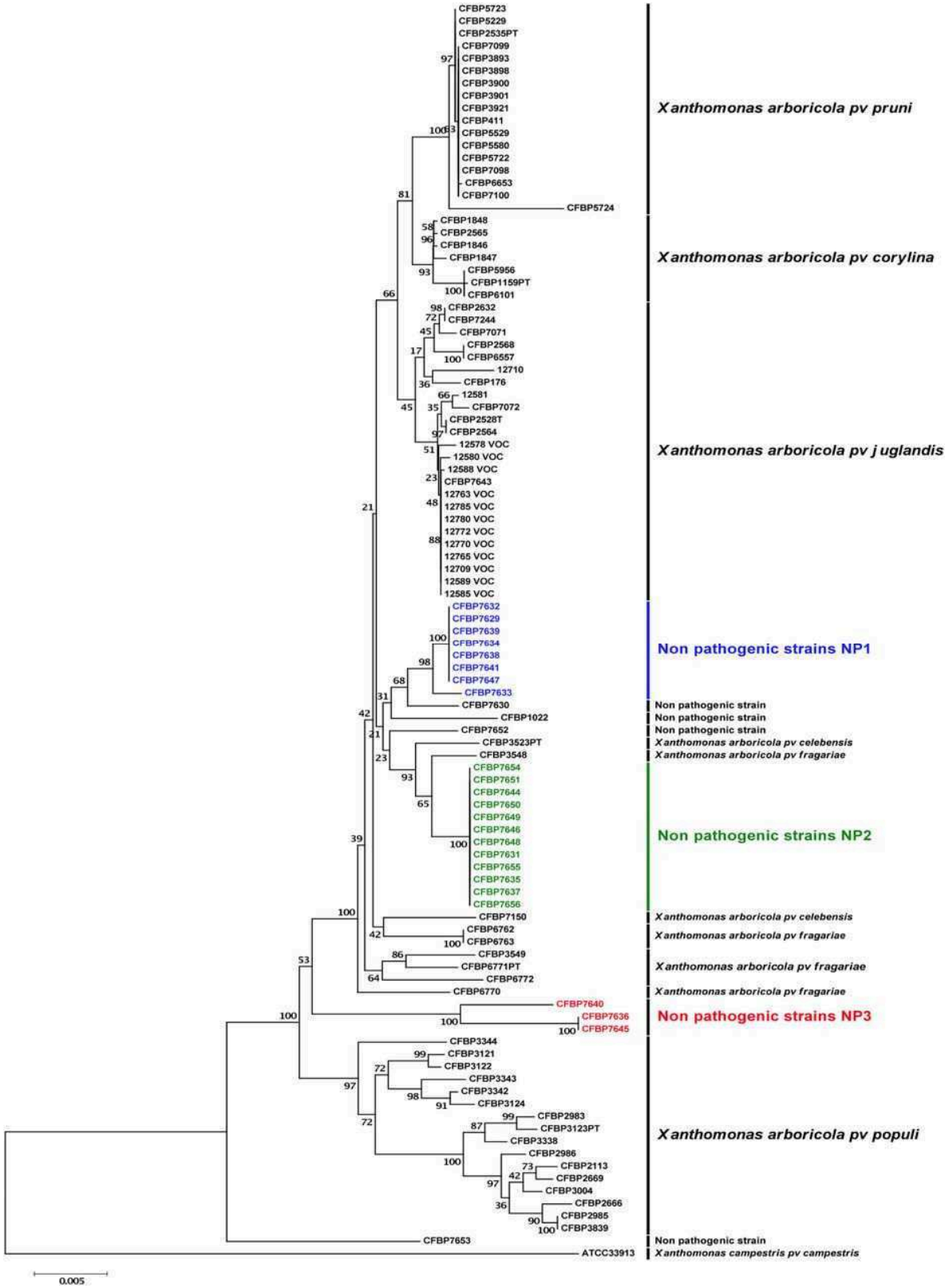


Figure I-30 : Arbre phylogénétique reconstruit en utilisant la méthode de maximum de vraisemblance et les séquences concaténées de 7 gènes de ménage (*atpD*, *dnaK*, *efP*, *fyuA*, *glnA*, *gyrB* et *rpoD*). Les trois lignées de souches non pathogènes isolées de noyer sont représentées en bleu pour la lignée NP1, en vert pour la lignée (NP2) et en rouge pour la lignée (NP3) (Essakhi et al. 2015).

souches sont différentes des espèces de *Xanthomonas* déjà connues pour attaquer le poivron, mais appartiennent à l'espèce *X. arboricola*. Au sein de cette espèce, ces souches ne se regroupent pas avec des pathovars connus. En Tanzanie, une enquête épidémiologique des souches responsables de la maladie des taches bactériennes sur feuilles de tomate a identifié 18 souches sur 30 comme appartenant à l'espèce *X. arboricola* à l'aide du gène *fyuA* (Mbega et al. 2012). D'autres souches appartenant à l'espèce *X. arboricola* mais qui ne sont pas classées dans un pathovar sont pathogènes sur tournesol (Ignatov et al. 2010), vigne (Sawada et al. 2011), chêne (Ishihahra et al. 2015) et jujube (Myung et al. 2010). Enfin, une souche isolée de l'orge et pathogène sur cet hôte a été isolée, et son génome séquencé récemment (Ignatov et al. 2015). Cette souche est dépourvue de SST3 et d'ET3.

9. Les souches commensales

Les hôtes d'origine des souches commensales de l'espèce *X. arboricola* sont très variés. La majorité des souches commensales étudiées par Vauterin et al. 1996 proviennent de plantes hôtes phylogénétiquement diverses (70 souches, 13 espèces hôtes) et qui ne correspondent généralement pas aux hôtes des souches pathogènes des pathovars inclus dans *X. arboricola*. Comme indiqué précédemment, des souches commensales phylogénétiquement diversifiées ont également été isolées de bourgeons de noyer (fig. I-30) (Cesbron et al. 2015 ; (Essakhi et al. 2015), mais également de semences de haricot (Boureau et al. 2013 ; Grimault et al. 2014). Une souche ne présentant pas la capacité à développer des symptômes et appartenant à l'espèce *X. arboricola* a également été isolée de *Prunus*, hôte des souches du pathovar *X. arboricola* pv. *pruni* (Garita-Cambronero et al. 2016a). La phylogénomique de 18 souches de l'espèce, incluant des souches pathogènes et commensale montre que cette souche se situe en dehors des souches du pathovar *pruni* (Garita-Cambronero et al. 2016b). Une analyse phylogénétique des souches commensales issues de l'étude de Vauterin et al. 1996, et de semences de haricot a montré que la majorité de ces souches appartenaient à l'espèce *X. arboricola* (Guimbaud, 2012).

L'établissement de répertoires de gènes impliqués dans la virulence a été entreprise par comparaisons génomiques ou recherche de gènes candidats par PCR entre les souches pathogènes et commensales afin de déterminer les facteurs génétiques impliqués dans cette différence phénotypique (Vandroemme et al. 2013 ; Essakhi et al. 2015 ; Cesbron et al. 2015 ; Garita-Cambronero et al. 2016b). Certaines souches commensales semblent ne posséder ni SST3 ni effecteur associé, c'est le cas pour la souche isolée de *Prunus* (Garita-Cambronero et al. 2016a) et certaines souches isolées de noyer tandis que d'autres souches isolées de ce même hôte ne

possèdent pas de SST3, mais quelques ET3, ou un SST3 et quelques ET3 (Cesbron et al. 2015 ; Essakhi et al. 2015). Une situation similaire est observée chez les souches appartenant au pathovar *fragariae* qui possèdent soit les 2 gènes recherchés, à savoir un gène codant un effecteur et un gène codant une protéine du SST3, soit aucun de ces deux gènes (Vandroemme et al. 2013).

Par ailleurs, dans les travaux conduits pour la description de l'espèce *X. dyei*, les auteurs ont rapporté l'existence de souches apparemment non-pathogènes sur leurs hôtes d'isolement et les ont assignées à l'espèce *X. arboricola* (Young et al. 2010). Ces souches ont été isolées de 5 espèces hôtes différentes et sur la base de 4 gènes de ménage ces souches ont pu être assignées à cette espèce, mais il n'a pas été possible de les assigner à un pathovar déjà connu. Ces souches pourraient être considérées comme des souches commensales, cependant il y a peu d'information sur leur pouvoir pathogène.

V. Questions de recherche

L'objectif principal de cette thèse était d'inférer l'histoire évolutive de l'espèce *X. arboricola* en incluant les souches commensales. En effet, ces souches sont très proches phylogénétiquement des individus pathogènes et ont été retrouvées en sympatrie avec celles-ci, on peut donc s'interroger sur les flux de gènes existant entre elles. La question de recherche que je me suis posée est donc : Ces souches commensales représentent-elles un risque pour de nouvelles émergences ?

Le premier chapitre présente l'étude de structuration génétique de l'espèce *X. arboricola*, ainsi que l'établissement et l'évolution des répertoires de facteurs impliqués dans les traits d'histoire de vie (ET3, MCP, TonB-Dependent Transporters (TBDT), et adhésines). Ici, l'objectif était tout d'abord de connaitre les relations génétiques entre souches commensales et pathogènes et de déterminer l'intensité des forces évolutives qui s'exercent au sein des populations identifiées, c'est-à-dire de comparer l'intensité de la recombinaison homologue par rapport à la mutation. Le second objectif était d'étudier le contenu en gènes impliqués dans les traits d'histoire de vie chez les différentes populations et d'identifier les événements de gains et de pertes de facteurs de virulence au sein de cette espèce. Cette étude a été réalisée en utilisant une collection de 187 souches, pathogènes et commensales, dont les séquences partielles de 7 gènes de ménage étaient disponibles. De plus, les répertoires de facteurs de virulence ont été établis par PCR[6] dans une sous-collection de 47 souches. Dans ce chapitre, nous avons montré l'existence de trois

[6]PCR : Polymerase Chain Reaction

groupes génétiques au sein de l'espèce *X. arboricola* : le groupe A, composé des pathovars *pruni*, *juglandis* et *corylina*, le groupe B, composé d'un grand nombre de souches commensales et des pathovars *arracaciae*, *guizotiae*, *fragariae*, *celebensis* et *zantedeschiae* ainsi que les souches pathogènes du piment, et le groupe C, composé de quelques souches commensales et du pathovar *populi*.

Dans le deuxième chapitre, nous nous sommes interrogés sur l'existence d'un flux de gènes entre populations pathogènes et commensales. Nous avons étudié, en particulier, les événements de recombinaison homologue et hétérologue, par le biais de la recherche des éléments génétiques mobiles entre les groupes génétiques identifiés au sein de l'espèce *X. arboricola* sur la base de séquences génomiques. L'objectif de cette étude était de déterminer le flux génique entre les différents groupes génétiques ainsi que son intensité et son orientation. Cette étude a été réalisée sur une collection de 44 génomes de souches de *X. arboricola* comprenant 36 génomes de souches caractérisées dans le chapitre 1, et les 8 génomes disponibles publiquement en décembre 2014. Dans ce chapitre, nous avons montré que le groupe C n'appartenait pas à l'espèce *X. arboricola* et donc le flux génique a été étudié entre les groupes A et B, uniquement. Nous avons montré la persistance d'un flux de gènes au cours de la divergence entre ces deux groupes, ainsi qu'une asymétrie du flux de gènes du groupe A vers le groupe B.

La persistance du flux de gènes au cours de la divergence suggère que les souches des groupes A et B vivent en sympatrie. L'objectif du troisième chapitre était de mieux comprendre l'écologie des souches commensales sur la plante et de déterminer si elles partagent les mêmes niches écologiques que les souches pathogènes. Pour cela, les principaux mécanismes impliqués dans les phases de colonisation et d'installation à savoir la mobilité, l'adhésion et la formation de biofilm, ont été recherchés dans une collection de 40 souches pour lesquelles les génomes ont été séquencés pour ce projet de thèse. De plus, des analyses d'antibiose ont été réalisées en confrontant deux à deux les souches pathogènes et les souches commensales, afin de déterminer si elles peuvent coexister sur une même plante. Dans ce chapitre, nous avons montré que les souches commensales sont capables de réaliser les phases précoces du cycle infectieux bien que, contrairement aux souches pathogènes, elles soient incapables d'engager un processus infectieux. Enfin, les analyses d'antibiose ont montré que les deux types de souches peuvent coexister.

L'importance du SST3 et des ET3 dans l'émergence des clones épidémiques a été mise en évidence dans le chapitre 1. Ces analyses suggèrent aussi le caractère ancestral du SST3 chez l'espèce *X. arboricola* et sa perte chez certaines souches commensales. Dans le quatrième chapitre de cette thèse, nous avons retracé l'histoire évolutive du SST3 qui est le principal facteur de virulence, au sein du genre *Xanthomonas* et de ses effecteurs au sein de l'espèce *X. arboricola*. L'objectif de cette

étude était de déterminer si ce système avait été acquis chez l'ancêtre commun des *Xanthomonas* puis perdu, et éventuellement ré-acquis chez certaines souches ou s'il avait été acquis de manière indépendante dans plusieurs lignées de ce genre. Cette étude a été réalisée en utilisant une collection de 82 génomes représentant l'ensemble des espèces appartenant à ce genre, telles que définies par Young et al. 2008. Nous montrons que le système de sécrétion a été acquis de manière ancestrale dans l'histoire évolutive du genre, puis il a été perdu dans certaines lignées de souches commensales.

Enfin, dans le dernier chapitre, je discuterai ces résultats dans leur globalité et essaierai de montrer en quoi ils contribuent à une meilleure compréhension des émergences d'agents pathogènes chez les *Xanthomonas*. Je préciserai également comment ils peuvent contribuer à la prédiction des émergences de nouveaux agents pathogènes dans ce genre.

Chapitre I

Introduction

Afin de déterminer le rôle que peuvent avoir les bactéries commensales dans l'émergence de nouvelles maladies, l'objectif de ce premier chapitre[7] était de déterminer la structure de population de l'espèce *X. arboricola* en intégrant sa composante commensale et d'inférer les événements de gains et de pertes de gènes liés à l'interaction avec la plante au cours de l'histoire évolutive de cette espèce. En effet, *X. arboricola* est composée à la fois de souches commensales et pathogènes (Vauterin et al. 1996 ; Vandroemme et al. 2013, Essakhi et al. 2015 ; Cesbron et al. 2015). Une première analyse de diversité de grande ampleur de l'espèce a porté principalement sur la composante pathogène ; elle n'incluait pas de souches commensales (Fischer-Le Saux et al. 2015). L'étude de la diversité génétique des souches pathogènes et commensales isolées de noyer a montré que ces dernières étaient différentes des souches pathogènes et formaient plusieurs groupes monophylétiques (Essakhi et al. 2015). Des souches commensales isolées d'un grand nombre d'espèces hôtes ont été incluses dans l'étude présentée ici ; de plus, l'ensemble de la diversité des souches pathogènes est représenté. Cette étude porte donc, sur une collection représentative de la diversité connue de cette espèce (187 souches) pour laquelle une analyse de génétique de population basée sur 7 gènes de ménage a été réalisée. Elle a permis de mettre en évidence que les souches commensales forment un réseau de souches recombinantes. Au sein de ce réseau, on retrouve certaines souches pathogènes

[7]Différentes personnes ont participé à l'élaboration de ce travail. Sophie Bonneau a réalisé le séquençage des gènes de ménage d'une partie de la collection ainsi que la recherche des effecteurs de type 3 par PCR pour l'ensemble de la collection, et a réalisé les premières analyses d'inférence de gains et de pertes au cours de l'histoire évolutive. Au cours de son stage de Master 2 réalisé avant mon arrivée dans l'équipe, Jean-François Guimbaud a séquencé les gènes de ménage d'un grand nombre de souches commensales, ce qui a permis d'identifier celles appartenant à l'espèce *X. arboricola*. Karine Durand et Chrystelle Brin ont réalisé la recherche par PCR de l'ensemble des facteurs impliqués dans les phases précoces de l'interaction (MCP et adhésines), Tristan Boureau a aidé pour la réalisation des PCR et l'interprétation des résultats liés aux effecteurs de type 3. Christophe Lemaire a participé aux discussions concernant les analyses de génétique des populations, leur interprétation et la rédaction du manuscrit. Marie-Agnès Jacques et Marion Fischer-Le Saux m'ont encadrée tout au long de ce projet pour les questions biologiques, l'analyse des données, leur interprétation et la rédaction du manuscrit. Pour ma part, j'ai réalisé l'ensemble des analyses de génétique des populations, j'ai continué celles concernant les facteurs impliqués dans les traits d'histoire de vie, et rédigé le manuscrit.

appartenant à des pathovars à faible incidence (unsuccessful pathovars) car leur expansion est limitée ainsi que leur impact économique. Les pathovars responsables d'épidémies à l'échelle mondiale avec un impact économique majeur, quant à eux, forment des clones épidémiques. Ainsi, l'espèce *X. arboricola* présente une structure épidémique, telle que définie par Maynard-Smith et al. en 1993. La seconde partie de l'étude est basée sur l'analyse d'une matrice de présence/absence des gènes impliqués dans les traits d'histoire de vie (gènes du système de sécrétion de type 3 et ses effecteurs et gènes codants des Methyl-accepting Chemotaxis Proteins (MCP), adhésines et TonB Dependent Transporters (TBDT). Ces données ont été acquises par PCR et les profils de certaines souches ont été validés en utilisant les génomes disponibles publiquement. Les répertoires d'effecteurs de type 3 sont très contrastés entre les souches commensales et les pathovars non performants d'une part, et les pathovars performants d'autre part. Au contraire, les répertoires de gènes impliqués dans les phases précoces de l'interaction (MCP, adhésines, TBDT) n'ont pas montré de différence notable suivant le profil pathologique des souches. Ainsi, l'inférence de gains et de pertes de gènes au cours de l'histoire évolutive de l'espèce a montré que l'ancêtre commun de l'espèce possédait un système de sécrétion de type 3 et un nombre restreint d'effecteurs, à partir de celui-ci les souches pathogènes ont acquis de nouveaux effecteurs, tandis que certaines souches commensales ont perdu ces gènes. D'autres souches commensales, quant à elles, ont conservé le répertoire ancestral. De plus, il a été montré que huit gènes d'effecteurs avaient été acquis par l'ancêtre commun des trois clones épidémiques et pourraient donc être à l'origine de leur caractère épidémique. Cette étude a été publiée dans " Environmental Microbiology Report ". Dans ce chapitre, les souches commensales[8] sont appelées "souches non pathogènes", ce qui correspond au terme utilisé au début de ma thèse pour ces souches.

[8]Parmi les souches commensales, nous avons inclus les souches du pathovar *fragariae*. Cependant, l'article de Karamura et al. (2015) a été publié plus tard, donc les souches du pathovar *celebensis* sont considérées comme pathogènes

Recombination-prone bacterial strains form a reservoir from which epidemic clones emerge in agroecosystems

Merda D.*, Bonneau S.*, Guimbaud J F., Durand K., Brin C., Boureau T., Lemaire C., Jacques $MA^{\$}$., Fischer-Le Saux $M^{\$}$.

* These authors contributed equally to this work.

$ These authors contributed equally to this work.

Abstract

The acquisition of virulence-related genes through horizontal gene transfer can modify the pathogenic profiles of strains and lead to the emergence of new diseases. *Xanthomonas arboricola* is a bacterial species largely known for the damage it causes to stone and nut fruit trees worldwide. In addition to these host-specific populations called pathovars, many nonpathogenic strains have been identified in this species. Their evolutionary significance in the context of pathogen emergence is unknown. We looked at seven housekeeping genes amplified from 187 pathogenic and nonpathogenic strains isolated from various plants worldwide to analyze population genetics and recombination dynamics. We also examined the dynamics of the gains and losses of genes associated with life history traits (LHTs) during *X. arboricola* evolution. We discovered that *X. arboricola* presents an epidemic population structure. Successful pathovars of trees (i.e., *pruni*, *corylina*, and *juglandis*) are epidemic clones whose emergence appears to be linked to the acquisition of eight genes coding for Type III effectors. The other strains of this species are part of a recombinant network, within which LHT-associated genes might have been lost.

We suggest that nonpathogenic strains, because of their high genetic diversity and propensity for recombination, may promote the emergence of pathogenic strains.

I. Introduction

Deciphering the population structure of bacterial species should help to predict the emergence of infectious diseases. Different structural models have been defined; they range from fully clonal to panmictic, depending on the relative roles of homologous recombination and mutation in shaping polymorphism within the population (Maynard-Smith et al. 1993). In the clonal model, polymorphism arises only by mutation. In contrast, in the panmictic model, the recombination rate within populations is high, leading to random associations among alleles. Epidemic population structure falls somewhere in between: it involves a recombinant network with high haplotype diversity and between-strain recombination, from which emerge epidemic clones composed of a limited number of highly frequent haplotypes.

The standard approach so far used to detect epidemic population structure is multilocus sequence typing (MLST), which highlights the presence of clonal complexes within haplotype networks. Recombination is an evolutionary force that encompasses homologous and non-homologous recombination (Didelot et al. 2012). Whereas homologous recombination can be estimated from allelic data, non-homologous recombination is characterized by the gain or loss of accessory genes that occurs via horizontal gene transfers (HGTs). HGTs have played a crucial role in the evolutionary history of human pathogens, as the exchange of genetic material related to pathogenicity has allowed the emergence and spread of new pathogen lineages (Schubert et al. 2009).

Initially described in *Neisseria meningitidis* (Maynard-Smith et al. 1993), epidemic population structures have largely been observed in human pathogens with environmental reservoirs (Pirnay et al. 2009; Yan et al. 2011; Theethakaew et al. 2013). Clinical isolates display a limited number of highly common genotypes, namely epidemic clones, whereas strains isolated from environmental reservoirs are diverse and form a recombinant network. In addition to human pathogens, a plant pathogen, *Pseudomonas syringae*, also appears to demonstrate epidemic population structure (Monteil et al. 2013). While epidemic clones are more frequent in agroecosystems, highly recombinant genetic lineages have also been found in non-agricultural settings. It has been hypothesized that these environments serve as environmental reservoirs for emerging crop pathogens (Monteil et al. 2013). Indeed, they could allow the emergence and spread of a new epidemic clone given favorable environmental conditions (Engering et al. 2013).

In phytopathogenic bacteria, genes associated with life history traits (LHTs) are expressed during interactions with the plant host. Methyl-accepting chemotaxis proteins (MCPs), TonB-dependent transporters (TBDTs), and adhesins are mostly involved in earlier interactions, such as when the bacteria arrive on the plant and colonize it (Mhedbi-Hajri et al. 2011a). Type III effectors (T3Es), secreted through the type III secretion system (T3SS), are mainly involved in escaping the host immune system and promoting nutrient uptake, which allows efficient growth and, ultimately, results in the development of symptoms in the host (Stavrinides et al. 2008; Xin & He, 2013). In phytopathogenic members of the genera *Pseudomonas* and *Xanthomonas*, such LHT-associated genes are involved in host specificity; for instance, repertoires of these genes correlates with host range (Sarkar et al. 2006; Lindeberg et al. 2009; Hajri et al. 2009; Mhedbi-Hajri et al. 2011b; Hajri et al. 2012; Mhedbi-Hajri et al. 2013; Schwartz et al. 2015). LHT-associated genes are often located on mobile genetic elements, such as plasmids or transposons, and can easily be lost or gained via HGT (Alfano & Collmer, 1996; Ferreira et al. 2015; Huang et al. 2015). If we wish to understand the evolutionary epidemiology of bacterial plant diseases, it will be crucial to characterize gene flows among nonpathogenic and pathogenic strains and examine the potential influence of nonpathogenic strains on the emergence of epidemic clones.

In this study, we used *X. arboricola* as our model species because, among the xanthomonads, this species is the only one for which numerous nonpathogenic isolates are available. Nonpathogenic strains are defined as strains that do not induce symptoms in their host of isolation following artificial inoculation and incubation under optimal conditions. Indeed, it is striking that many nonpathogenic strains have been isolated from symptomatic and asymptomatic host tissues (e.g., walnut buds, bean seeds) in this species (Vauterin et al. 1996; Boureau et al. 2013; ISTA METHOD, 2014; Essakhi et al. 2015). In addition to these nonpathogenic strains, *X. arboricola* is also known to encompass three pathovars: pathovars *pruni*, *juglandis*, and *corylina*, which cause bacterial spot on stone fruit trees, walnut blight, and bacterial blight on hazelnut, respectively. Over the last decade, these three pathovars have emerged in several countries in different parts of the globe, where they have had major economic impacts (Palacio-Bielsa et al. 2010; Ibarra-Caballero et al. 2013; Shen et al. 2013; Palacio-Bielsa et al. 2014). As a consequence, we will hereafter refer to them as "successful" pathovars. In addition to these successful pathovars, five other pathovars have demonstrated more limited expansion, as evidenced by the occasional reports of occurrence; from this point on, they will be referred to as "unsuccessful" pathovars. Their limited expansion could be explained by host or environmental factors (e.g. a restricted geographical area of their hosts, cultural practices less conducive for disease development) or by a weak aggressive-

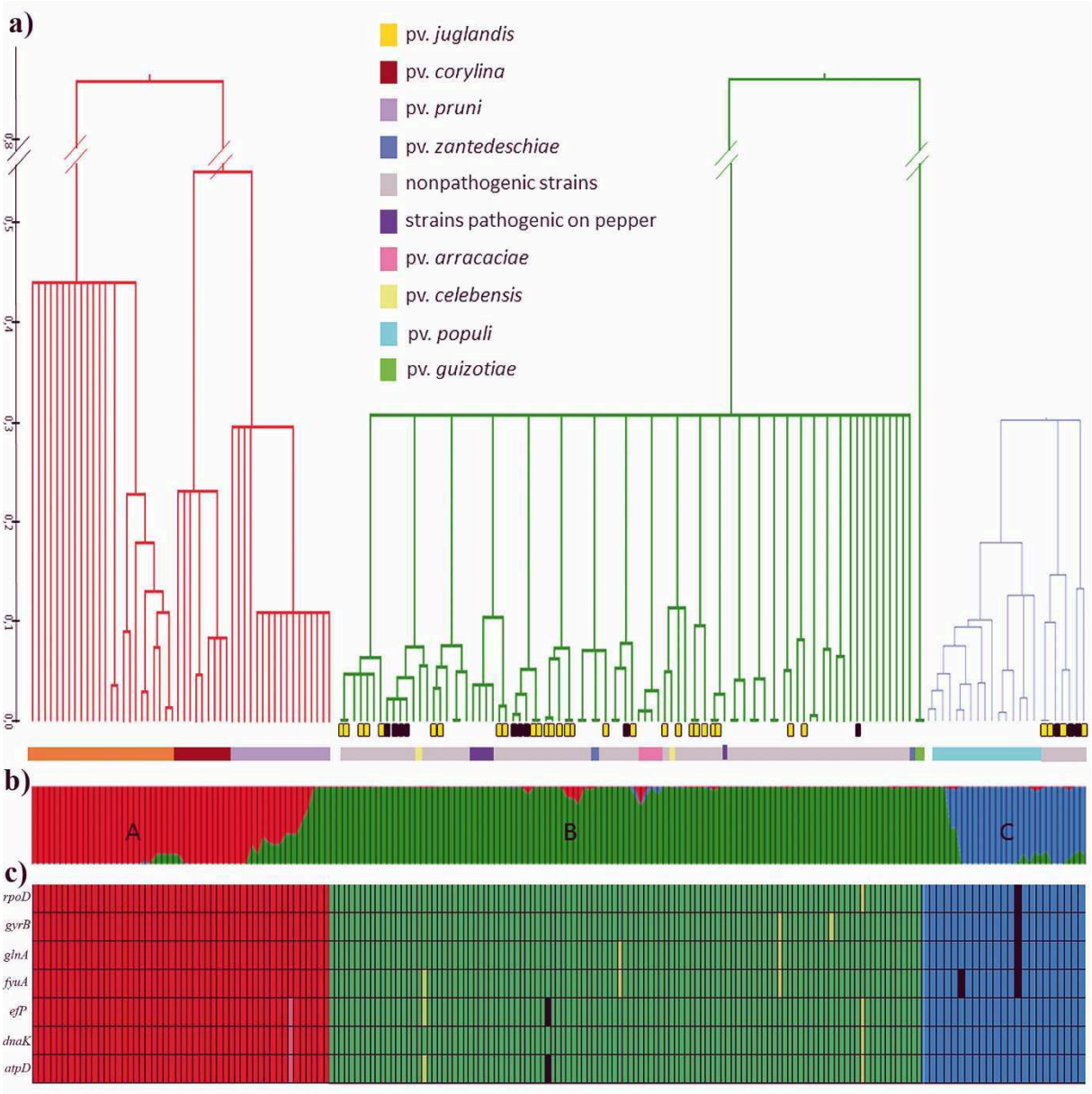


Fig. II-1: Genealogies (a), results of Bayesian clustering (b), and depiction of recombination events (c) for the 187 strains in collection I. (a) Genealogies were inferred for each of the groups identified by the Bayesian clustering analysis (Fig. II-1b). The vertical axis corresponds to the coalescence time. The black squares represent strains isolated from walnut and the yellow squares represent strains isolated from bean seeds. The colors of the bar under the genealogies correspond to the different pathovars (see the legend). (b) In the Bayesian clustering analysis, each group (K=3) is represented by a different color and each vertical bar represents a single strain. Within bars, the proportional coloration pattern indicates the posterior probability of an assignment to a given group. (c) Occurrence of recombination events within each strain, as identified using RDP 3.44. Each horizontal line corresponds to a housekeeping gene; from bottom to top, they are *atpD*, *dnaK*, *efP*, *fyuA*, *glnA*, *gyrB*, and *rpoD*. Group A strains are in red, group B strains are in green, and group C strains are in blue. The strains are in the same order as in the Bayesian clustering analysis. Recombination events are colored according to the group membership of the donor (minor parent). Recombinant fragments of unknown origin are in black. Recombinant fragments from group A and group B are in light red and green, respectively.

ness due to a limited number of virulence-associated factors (Collard et al. 2007; Nunney et al. 2014). These unsuccessful pathovars have been described (Fischer-Le Saux et al. 2015) on banana (pv. *celebensis*), poplar (pv. *populi*), arracacha (pv. *arracaciae*), niger (pv. *guizotiae*), and arum lily (pv. *zantedeschiae*). Finally, although *X. arboricola* pv. *fragariae* has been classified as a pathovar in the past, it failed to induce leaf blight on strawberries (its host of isolation) (Vandroemme et al. 2013; our unpublished results). As a consequence, we treat it as nonpathogenic in this study.

Nonpathogenic and pathogenic strains of *X. arboricola* occur sympatrically on hosts, as demonstrated by their simultaneous isolation from walnut buds (Essakhi et al. 2015), which could facilitate HGTs. As a result, this species is a relevant model, with which to study the roles of nonpathogenic isolates in evolutionary mechanisms that may lead to the emergence of epidemic clones in *Xanthomonas* genera. First, we studied the population structure of *X. arboricola* to determine the extent of genetic differentiation between pathogenic and nonpathogenic strains, and we evaluated the intensity of the evolutionary forces (i.e., recombination vs. mutation) shaping polymorphism. Second, we analyzed the dynamics of the gains and losses of LHT-associated genes. Taken together, these data provide clues about the population structure and evolutionary history of *X. arboricola* and allow us to propose a scenario in which epidemic clone emergence is associated with gains of T3E genes. Our results strongly suggest that, because nonpathogenic strains are highly genetically diverse and demonstrate a propensity for recombination, they may promote the emergence of pathogenic strains.

II. Results

Xanthomonas arboricola encompasses three genetic groups

The genetic structure of the 187 strains of collection I (Table II-S1) was studied by Bayesian clustering. The most likely number of clusters (K) was three: groups A, B, and C (fig. II-1b). Group A (n=50) was exclusively composed of all the strains belonging to the three successful pathovars. Group B (n=114) was composed of 43 pathogenic strains belonging to unsuccessful pathovars and 70 nonpathogenic strains isolated from diverse hosts. Group C (n=23) encompassed all strains of pathovar *populi* and seven nonpathogenic strains. Overall haplotype diversity in collection I was high, with a total of 115 sequence types (STs) detected among the 187 *X. arboricola* strains (Table II-S2). The splitgraph topology demonstrated a star-like topology composed of groups A and B from which four lineages (group C) diverged with long branches (in blue; fig. II-2a). The differentiation of these four

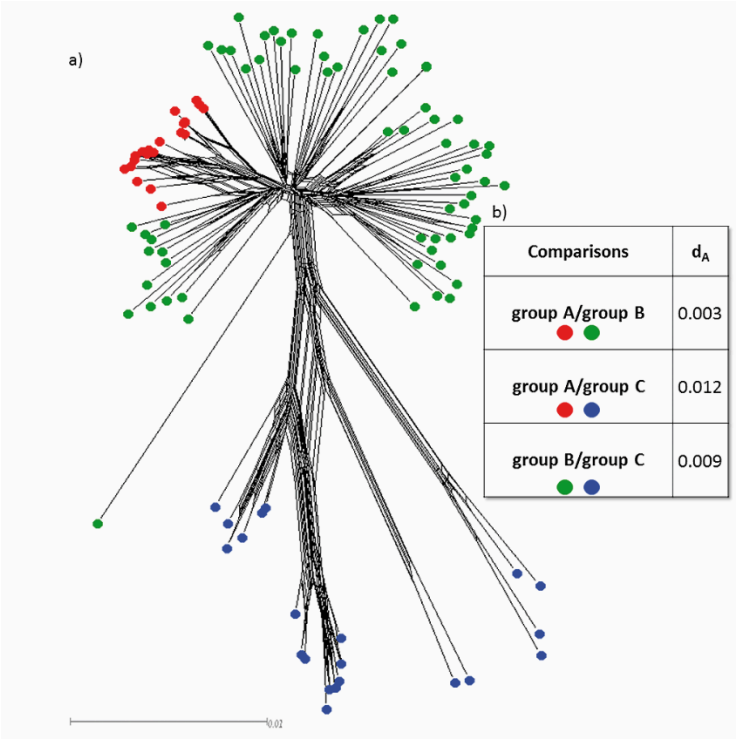


Fig. II-2: a—Splitgraph of the concatenated dataset (*atpD*, *dnaK*, *efP*, *fyuA*, *glnA*, *gyrB*, and *rpoD*) for the 187 *X. arboricola* strains studied (collection I). Strains in group A are in red, strains in group B are in green, and the four divergent lineages (group C) are in blue. b—Pairwise genetic distances (dA) among the three groups.

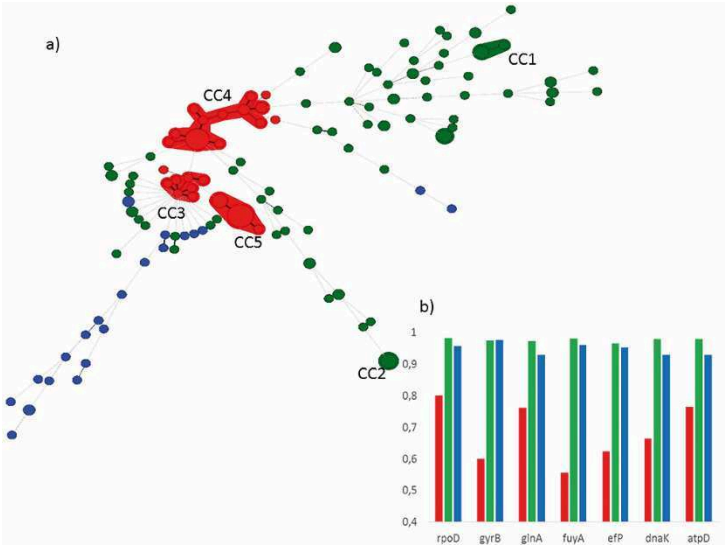


Fig. II-3: (a) Minimum spanning tree created using the allelic profiles of the 187 strains in collection I. Each color represents the groups identified in the Bayesian clustering analysis where K=3 (group A is in red, group B is in green, and group C is in blue). Clonal complexes (CCs) are highlighted. The circles represent the sequence types, with the size indicating their frequency. A CC was defined as a set of STs representing at least three strains and that shared at least five haplotypes. CC1: pathovar *arracaciae*; CC2: pathogenic strains on pepper CC3: pathovar *corylina*; CC4: pathovar *juglandis*; and CC5: pathovar *pruni*. (b) Histogram of haplotype diversity (Hd) within each group (red = group A, green = group B, and blue = group C) and for each housekeeping gene.

lineages from the main group of strains was also found in the ML tree (fig. II-S1), and was strongly supported by the Bayesian clustering results. Results obtained for K=2 underscored the differentiation between group C and the other strains in the collection (data not shown). Genetic distances between group C and groups A and B were at least three times higher than the genetic distance between groups A and B (fig. II-2b).

Xanthomonas arboricola is composed of three clonal complexes of successful pathovars and an unstructured network of highly diverse strains

The coalescence analysis of group A strains revealed host-based clustering (fig. II-1a). This result was also supported by the phylogeny (fig. II-S1) and by allelic profile comparisons: these strains formed three clonal complexes (CCs) (fig. II-3a). Compared to the other strains in the collection, group A strains exhibited very low haplotype diversity (fig. II-3b; Table II-S3); the number of segregating sites in group A corresponds to the quarter of the other groups (Table II-1). Two CCs, pv. *juglandis* and pv. *pruni*, were each composed of a single major ST (41% and 72% occurrence, respectively) and a few low-frequency STs, which is a classic signature of a bottleneck followed by recent expansion. Although the three neutrality tests failed to detect significant departures from the null hypothesis of mutation/drift equilibrium (Table II-S3), slightly negative values at some loci suggest an excess of singletons (i.e., substitution present in only one sequence) in polymorphism. The lack of significant values could be due to the large number of isolates belonging to the same ST, which could lead to a lack of pairwise differences.

In contrast to group A, no host-based clustering was observed in group B, with the exception of pathovar *arracaciae* and a group of four strains that are pathogenic on pepper (fig. II-1a; fig. II-3a). Within group B, most STs were singletons—each was composed of a single strain that differed from its nearest neighbors by three or more loci (fig. II-3a). Haplotype diversity (mean Hd = 0.976; fig. II-3b; Table II-S3) and the number of segregating sites (Table II-1) were higher in group B than in the other two groups.

Among the four lineages observed in group C (in blue, fig. II-2a), two of them corresponded to strains of the pathovar *populi*. Genealogy reconstruction showed that, although pv. *populi* strains formed two separate lineages, they all shared a private common ancestor (fig. II-1a). The two remaining divergent lineages encompassed seven nonpathogenic strains isolated from walnut buds and bean seeds that did not cluster according to their host of isolation (fig. II-1a).

Table II-1: Estimates of recombination and mutation

Groups and pathovars	N	No. of segregating sites [a]	θ_w [a]	R [b]	ρ/θ [b]	r/m [b]
all groups	187	649	126.57	83.23 (56.79–115.04)	0.66 (0.49–0.91)	2.02 (1.5–2.61)
group A	50	103	23.21	3.42 (0.88–6.78)	0.15 (0.04–0.297)	0.59 (0.21–1.03)
group B (with pv. *guizotiae*)	114	450	89.342	109.82 (35.32–312.08)	0.59 (0.17–1.34)	1.79 (0.61–3.71)
group B (without pv. *guizotiae*)	112	435	91.11	147.45 (72.87–251.81)	5.22 (1.94–9.52)	4.49 (2.49–7.20)
group C	23	358	100.52	90.23 (25.29–229.28)	0.97 (0.21–2.86)	3.19 (1.22–6.87)

N: number of sequences

[a]: parameter estimated using DnaSP; the value obtained was used to fix the θ value in ClonalFrame, as per Wicker *et al.* (2012)

[b]: parameter estimated using ClonalFrame; it was not estimated for the pathovars because they showed no Markov chain convergence

ρ^s: rho per site

θ^s: theta per site

θ_w:Watterson's theta

ρ or R: number of recombination events

ρ/θ: ratio of recombination events to mutation events

r/m: ratio of the impact of recombination vs. mutation at each site

Recombination drives diversification in nonpathogenic strains

The reticulate structures on the splitgraph (fig. II-2a) and the incongruence between the single-locus phylogenies (data not shown) suggested that recombination events have occurred in the strains sampled. Estimates of the pairwise homoplasy index (Bruen et al. 2006), $\Phi_w = 0.232$ (P-value=0.0), confirmed the presence of recombination events. Twenty such events were identified, but none of them occurred between the three groups (fig. II-1c). Indeed, the parental sequences of the recombinant fragments were either identified within the group to which the recombinant strains belonged or were of unknown origin. Recombination clearly appears to be the major evolutionary force that shaped nucleotide polymorphism in groups B and C (Table II-1). In fact, the r/m ratio was higher than one, which indicates that recombination has had a large influence within these two groups, even though mutation has occurred more frequently ($\rho/\Theta < 1$). In contrast, r/m ratios were far less than one in group A (Table II-1). Thus, mutation is the main evolutionary force that has been operating in group A. These results fit with the finding that the majority of the recombination events occurred in groups B and C (fig. II-1c).

Frequent recombination could break association between alleles. The extent of linkage between alleles at different loci can be determined using the standardized index of association (I_A^s). A zero value for I_A^s means that alleles are at linkage equilibrium in a non-clonal population. In contrast, values that are significantly different from zero indicate that a linkage disequilibrium exists (i.e., a nonrandom association between alleles at different loci). The I_A^s value for collection I was 0.6516 (P-value<0.05), which indicates strong linkage disequilibrium. When only one strain was used to represent each ST (clone correction), I_A^s decreased sharply ($I_A^s = 0.3558$, P-value<0.05) but remained significantly different from zero.

Pathogenic strains acquired LHT-associated genes, and nonpathogenic strains lost LHT-associated genes

Presence of three genes in the *hrp* gene cluster that encodes T3SS and of its master regulatory protein (*hrpX*) and repertoires of genes coding for 22 T3Es were defined for a subset of 73 strains of collection I (collection II; Table II-S4). While *hrpX* was present in all strains, T3SS and T3E encoding genes were successfully amplified in only 30 out of the 61 nonpathogenic strains. There were no more than six genes in the gene repertoire coding for T3Es in nonpathogenic strains, a number similar to that of unsuccessful pathovar strains in group B. Unexpectedly, two T3E genes (*xopR* and *avrBs2*) were detected in six nonpathogenic strains that lacked the T3SS genes. Twenty-four nonpathogenic strains and the CFBP 3123^{PT} strain

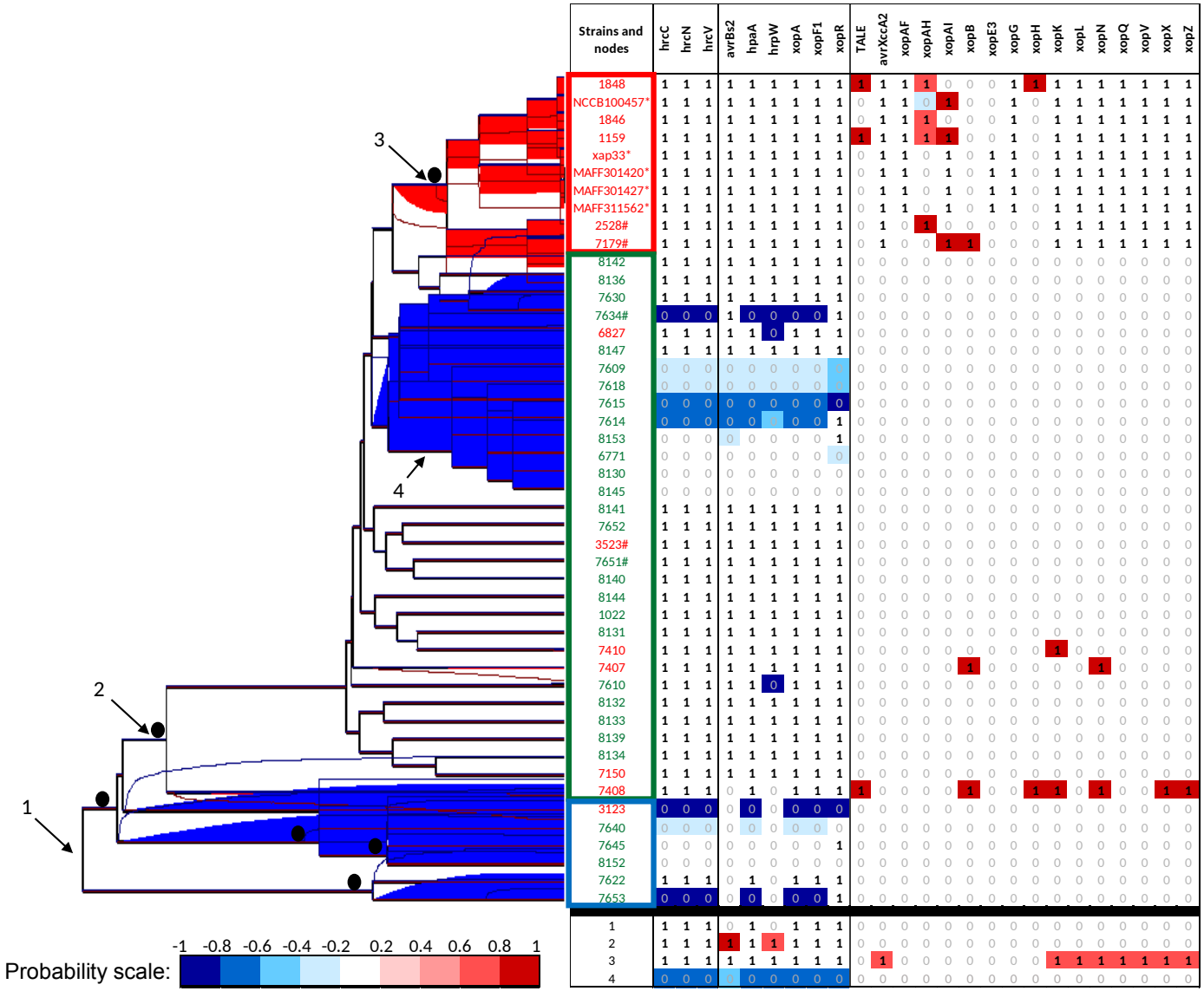


Fig. II-4: Inference of evolutionary gains and losses of LHT-associated genes. The figure depicts strain-specific gains and losses that have occurred since divergence from the sister strain. Losses are in blue, and gains are in red. The darker the color, the stronger the event probability. The strain numbers of the pathogenic and nonpathogenic strains are in red and green, respectively. The groups identified via Bayesian clustering analysis are represented by rectangles (red = group A, green = group B, and blue = group C). Group C strains were used as an outgroup. Bootstrap values higher than 85% are represented by black circles. In the figure, 1 = gene presence and 0 = gene absence. The ancestral repertoires for phylogenetic nodes 1 to 4 are indicated in the rows labeled 1 to 4 at the bottom of the figure. For strains indicates by *, T3Es repertoires were determined by BLAST analysis. For strains indicates by #, T3Es repertoires were validated by BLAST analysis.

(pathovar *populi*) harbored neither T3SS- nor T3E-encoding genes. The sizes of the repertoires of the T3E-encoding genes ranged from 15 to 19 in pathovars *pruni*, *corylina*, and *juglandis*. These values were far greater than those for nonpathogenic strains and unsuccessful pathovars, which ranged from 0 to 11. Pathogenic and nonpathogenic strains did not show differences in their repertoires of genes coding for adhesins, TBDTs, and MCPs (Table II-S5).

We inferred gains and losses of LHT-associated genes during *X. arboricola* evolution and only found significant events in the dynamic of T3SS- and T3E gene repertoires (data not shown). The analysis restricted to these latter repertoires showed that the T3SS genes and four T3E genes (*xopA*, *hpaA*, *xopF1*, and *xopR*) were found in the common ancestor of the *X. arboricola* strains (node 1; fig. II-4). It is highly probable that these genes were lost in group C strains CFBP 3123 (pathovar *populi*) and CFBP 7653 (nonpathogenic strain) (fig. II-4).

We found that T3E genes were acquired numerous times over evolutionary time by the successful pathovars (P-value ≈ 1). Eight T3E-encoding genes were probably acquired by the common ancestor of group A (node 3; fig. II-4). Following this acquisition, each strain in group A independently obtained additional T3E genes. It is also highly probable that T3E genes were acquired by certain group B pathogenic strains (pathovars *guizotiae*, *arracaciae*, and *zantedeschiae*). In group B, T3E genes appear to have been lost mainly in nonpathogenic strains. It is highly probable that the T3SS genes and six T3E genes were lost in the common ancestor of four nonpathogenic strains (node 4; fig. II-4). Similar loss events were also predicted in five other nonpathogenic strains. Indeed, neither T3SS nor T3E genes were found in these nonpathogenic strains with the exception of *xopR* detected in CFBP 7634, CFBP 7614 and CFBP 8153 and *avrBs2* in CFBP 7634.

xopR: an ancestral T3E gene in *X. arboricola* that has been lost in nonpathogenic strains

Because *xopR* was found in four nonpathogenic strains that otherwise lacked the T3SS genes and any known T3E genes and because the probability of gaining T3E genes was not strongly supported in those strains, we wondered if the presence of *xopR* reflected ancestral gene content or a recent HGT. The topologies of the *xopR* phylogeny and the strain phylogeny did not differ (P-value of SH test = 0.48; fig. II-5). The *xopR* sequence in CFBP 7653 was similar to that in CFBP 7622, its nearest phylogenetic neighbor, and these two sequences diverged greatly ($d_A = 0.13$) from those in other strains, which fits with the divergent phylogenetic position of these strains in the strain phylogeny. Thus, the *xopR* gene in CFBP 7653 was most likely inherited vertically rather than horizontally. In addition, the *xopR* genomic

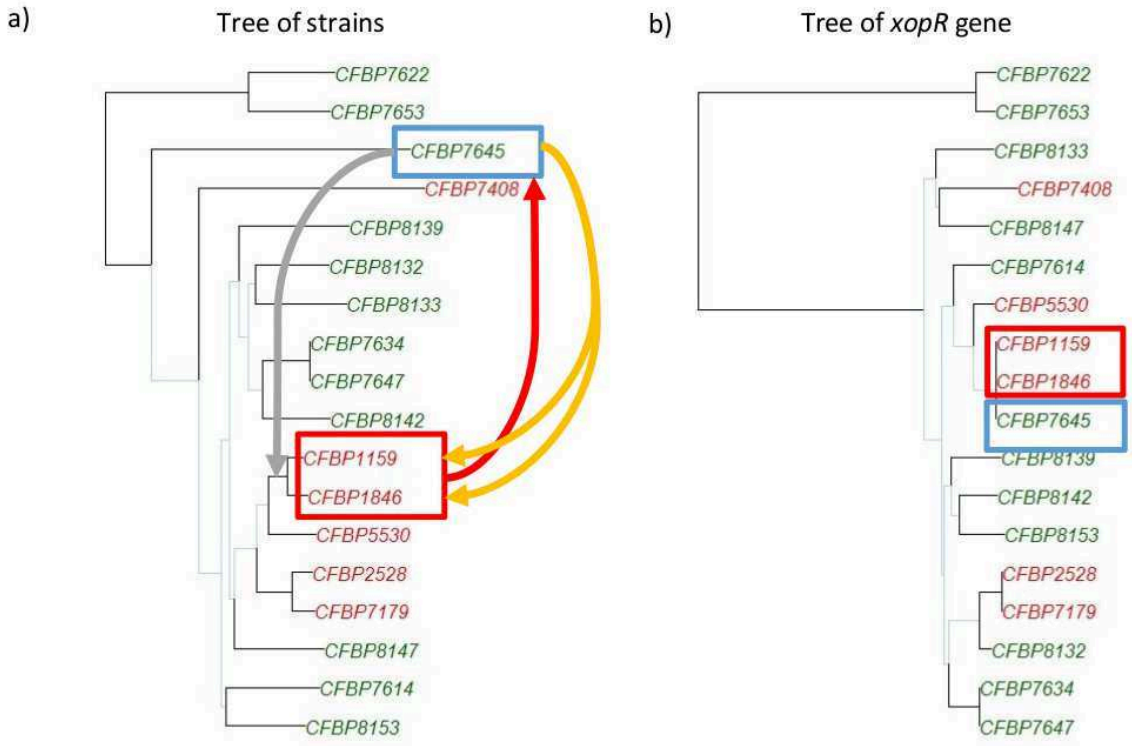


Fig II-5: Maximum likelihood phylogenies generated using the concatenated sequences of 7 housekeeping genes (a) and *xopR* sequences (b). The outgroups for the two phylogenies are CFBP 7622 and CFBP 7653. Pathogenic strains are in red and nonpathogenic strains are in green. Differences between the two topologies are represented by grey branches. The arrows on the tree of strains represent the three scenario for the recombinant events of *xopR*. The first scenario is in yellow, the second in grey and the third in red. The first two would involve a transfer of the *xopR* allele from nonpathogenic to pathogenic bacteria. In the first scenario, this transfer would result from two independent events - a transfer from the nonpathogenic strain to each of the pathogenic strains (yellow arrows). This scenario, which involves two evolutionary events, is less parcimonious than the other scenario. In the second scenario, this transfer would have occurred from the nonpathogenic strains to the common ancestor of the two pathogenic strains (grey arrows). In the third scenario, a transfer would have occurred from one pathogenic strain to nonpathogenic strain CFBP 7645 (red arrow).

environment has been highly conserved in all *X. arboricola* strains, whether they possess this gene or not (fig. II-6). There was no evidence of insertion sequences or tRNA genes 20 kb upstream or downstream from *xopR*, which could have served as evidence of a recent HGT. Therefore, it is more likely that *xopR* acquisition resulted from a one-time ancestral event and that this gene has been lost several times over *X. arboricola* history. This hypothesis is further supported by the GenoPlast results (fig. II-4). We can underline that CFBP 7645 (group C) and the two pathogenic strain CFBP 1846 and CFBP 1159 (group A) share the same *xopR* haplotype (fig. II-5). So, homologous recombination within the species might have occurred, as is the case for CFBP 7645.

III. Discussion

McDonald & Linde (2002) defined the risks posed by pathogen populations in relation to the forces underlying their evolutionary potential. Indeed, gene flow or genotype flow can result in transfers of favorable genes or allele combinations and thus give rise to highly aggressive strains. The higher the recombination rate, the greater the potential risk presented by the pathogen. *X. arboricola* is known as a plant pathogenic species and is unique in that it also comprises numerous nonpathogenic strains. Furthermore, nonpathogenic strains can occur sympatrically with pathogenic xanthomonads (Essakhi et al. 2015; Scortichini & Rossi, 2003). To characterize the evolutionary potential of *X. arboricola* and evaluate the risks of disease emergence presented by nonpathogenic strains, we studied the strength of two major evolutionary forces—homologous recombination and HGT events—in this species.

Xanthomonas arboricola displays an epidemic population structure

We found that *X. arboricola* displayed an epidemic population structure, which was originally described by Maynard-Smith et al. (1993). First, the numerous nonpathogenic strains and the unsuccessful pathovar strains (groups B and C) exhibited high haplotype diversity; most STs were represented by a single strain. The main evolutionary force responsible for polymorphism in these two groups was recombination; they thus form a recombinant network. Second, we identified three major clonal complexes; these correspond to the three successful pathovars. Evidence for an epidemic population structure also came from the signature dramatic drop in the I_A^s value when each ST was treated as a unit (Maynard-Smith et al. 1993; Theethakaew et al. 2013).

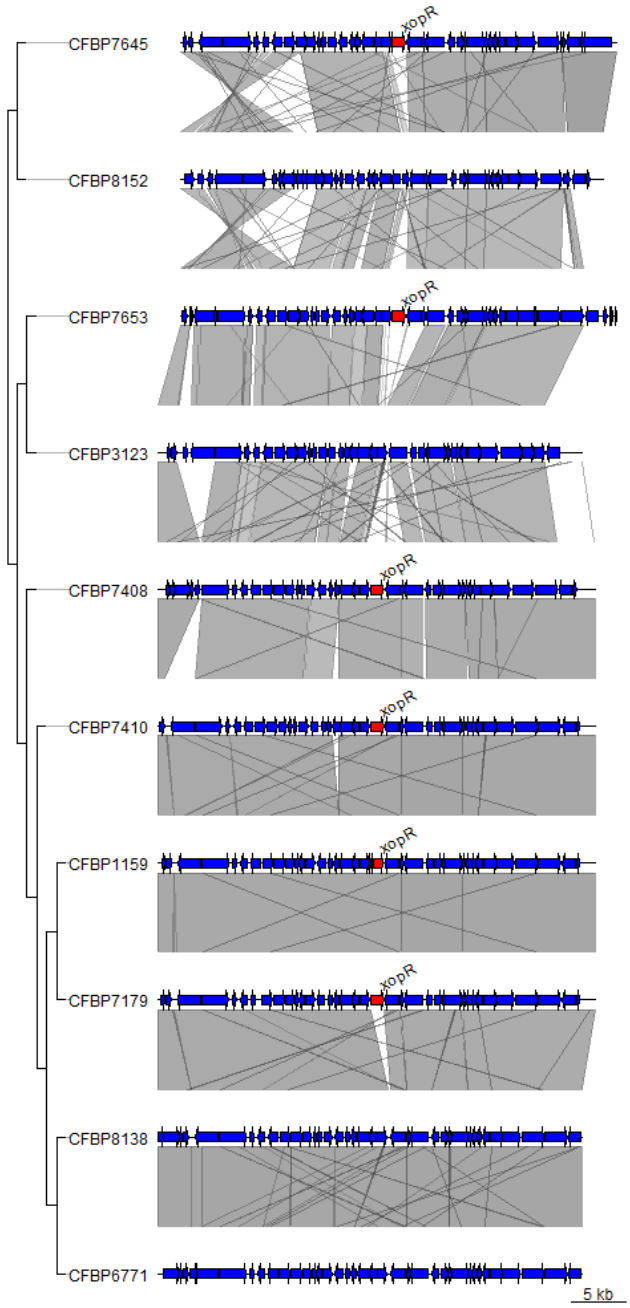


Figure II-6: Genomic environments of *xopR*. On the left, the dendrogram corresponds to the phylogenetic relationship between strains inferred from the seven housekeeping genes fragment using matrix distance method. On the right, genomic environments of *xopR*, within a windows of 20 kb upstream and downstream of *xopR*. Red arrows represent *xopR*. Blue arrows represent genes within genomic environment. Grey polygons represent the conservation degree between two strains

Recombination can introduce more polymorphism into a population than can mutation, depending on the length and nucleotide diversity of the imported fragment. In groups B and C, mutation occurred more frequently than recombination (ρ/Θ <1), although recombination had a greater impact on polymorphism (r/m>1). This result was further supported by the large recombinant fragments (341 bp to 2 780 bp) found in these groups. The r/m of *X. arboricola* recombinant network (r/m = 4.49) is equivalent to that of a species with an intermediate to high recombination rate, and the r/m of the *X. arboricola* epidemic clones (r/m = 0.49) is equivalent to that of a species with a low recombination rate (Vos & Didelot, 2009). In contrast to what has been observed in the recombinant network, mutation was the major evolutionary force responsible for polymorphism in group A, which comprised epidemic clones. This observation contrasts with what has been observed in *Escherichia coli*, a species that also displays epidemic population structure (Wirth et al. 2006). In *E. coli*, highly virulent epidemic clones have higher recombination rates than do nonpathogenic lineages. Wirth et al. (2006) hypothesized that this pattern results from the episodic selection of strains that are able to escape the host immune system. The difference between *X. arboricola* and *E. coli* might be explained by differences in host genetic diversity and within-host selection pressures. Agroecosystems are essentially large areas of land covered by genetically homogeneous plants. Consequently, environmental pressures on the pathogen are released (Stukenbrock & McDonald, 2008; Clay & Kover, 1996). It may also be that recombination occurs in agroecosystem settings, but only between individuals that are very closely related as a result of host specialization and recent emergence, such that the apparent effective recombination rate is low.

T3E genes are acquired in epidemic clones and lost in nonpathogenic strains from the recombinant network

The analysis of gene gains and losses indicated that 16 acquisition events of T3Es genes occurred over evolutionary history of the three epidemic clones: an initial acquisition of eight T3E genes most probably occurred in their common ancestor and subsequently, additional T3E genes were gained or lost. These results are supported by T3E phylogenetic data (Hajri et al. 2012). This founder event (i.e., the acquisition of the eight T3E genes), combined with favorable conditions, such as the horticultural expansion that took place in the late 19th century, could have selected for a clone adapted to stone and nut fruit trees. This hypothesis is supported by the fact that the pathovars have the same life cycle on their respective hosts and that they were first described in the same geographical area during the same time period—North America at the beginning of the 20th century (Pierce, 1901; Smith,

1903; Barss, 1915).

Conversely, in the nonpathogenic strains making up the recombinant network, either the LHT-associated gene repertoires have remained unchanged since the occurrence of the strains common ancestor or the T3E genes have been lost over evolutionary time. The selection pressure to maintain pathogenicity might have been released in nonpathogenic strains and could have led to the loss of the T3E genes. As the nonpathogenic strains were isolated from a large range of hosts and did not demonstrate host-related structuring, it may be that they are generalists.

Genetic material is exchanged between pathogenic and nonpathogenic strains

Gene flow occurred between pathogenic and nonpathogenic strains, and it involved LHT-associated genes. We showed that *xopR* is an ancestral T3E gene in *X. arboricola*. However, the nonpathogenic strain CFBP 7645 (group C), has the same *xopR* haplotype as two group A pathogenic strains. This pattern can be explained by homologous recombination and three scenarii of genetic exchanges are possible (fig. II-5). The most parsimonious one involves a transfer from a pathogenic strain to nonpathogenic strain CFBP 7645. Indeed, when a homologous recombination event occurs, each haplotype evolves independently and mutations may appear. However, in the case of these three haplotypes, the absence of mutations suggests recombination was recent. Thus, it is more likely that homologous recombination occurred between a group A pathogenic strain and the group C strain. As *xopR* product was shown to suppress innate immunity (Akimoto-Tomiyama et al. 2012), it is reasonable to assume that this kind of genetic exchange between pathogenic and nonpathogenic populations could represent seminal event of pathoadaptation.

IV. Conclusion

Before this study, *X. arboricola* was essentially known from its epidemic clones. Here, we have examined nonpathogenic strains, which make this biological model unique. We discovered that *X. arboricola* has an epidemic population structure. Epidemic clones that experience limited recombination are represented by the three successful pathovars (pathovars *juglandis*, *pruni*, and *corylina*), and the recombinant network is composed of nonpathogenic strains isolated from various hosts and unsuccessful pathovars (pathovars *zantedeschiae*, *guizotiae*, *arracaciae*, *celebensis*, and *populi*). We also looked at the relationship between LHT-associated genetic material and the evolutionary history of the species and found that the emergence of epidemic clones is linked to the acquisition of T3E genes, whereas nonpathogenic strains have

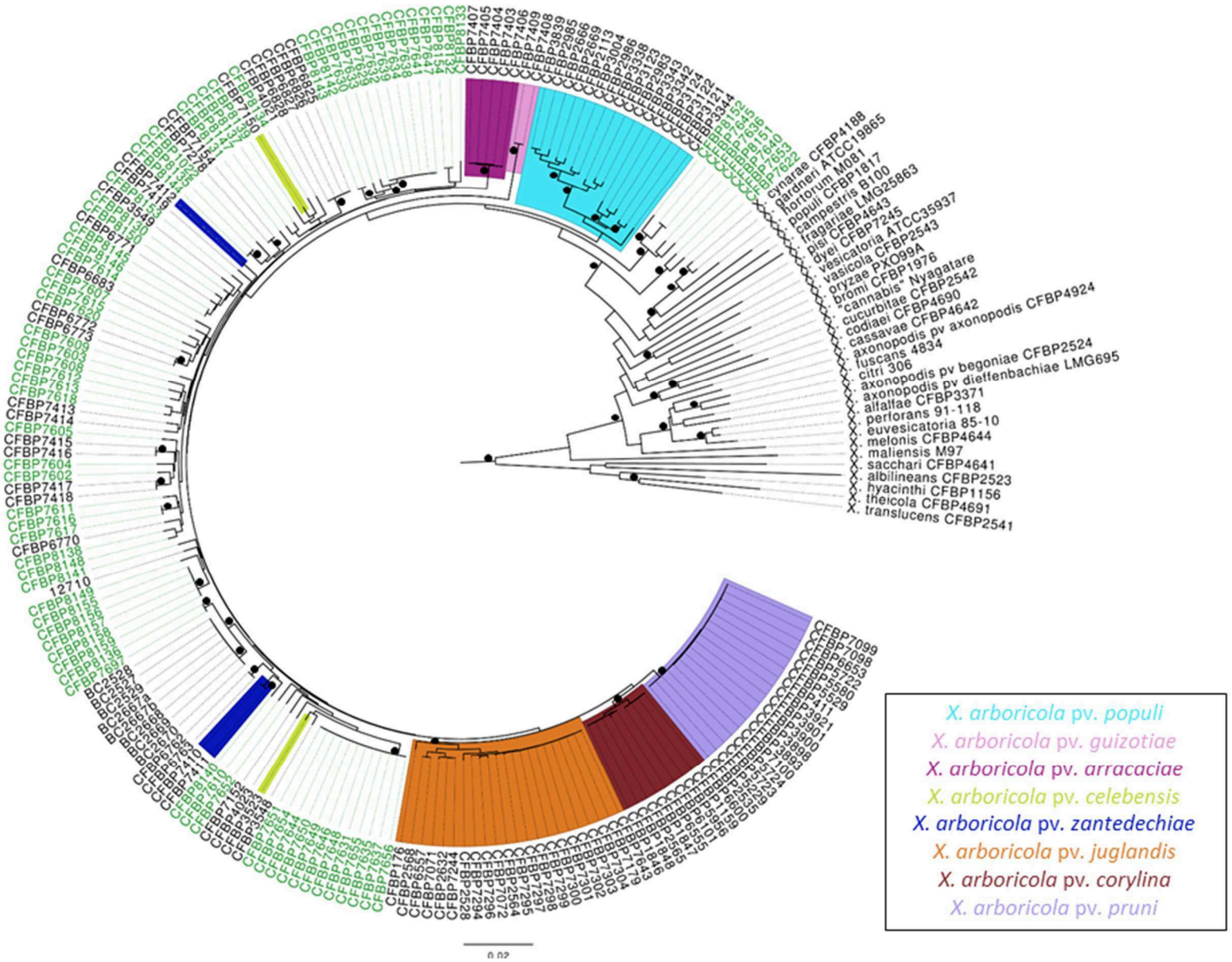


Figure II-8 : Assignation of nonpathogenic strains to *Xanthomonas arboricola* species. This phylogeny was obtained using neighbor-joining method and a concatenated dataset of partial sequences of *atpD*, *dnaK*, *efP*, *fyuA*, *glnA*, *gyrB* and *rpoD*. Bootstrap values were calculated with 1000 replicates and the bootstrap values higher than 85% are represented by black rings. All species of *Xanthomonas* genera are represented by the type strain of species. The tree is rooted with the sequence of *Xanthomonas translucens*. Each pathovar of *X. arboricola* is represented by a specific color. The nonpathogenic strains are represented by a green label.

retained the ancestral LHT repertoire or lost genes encoding these traits. By expanding our focus beyond the pathogenic strains, we discovered an unsuspected level of diversity in the nonpathogenic strains, similar to what has been observed in non- agricultural environments for *Pseudomonas syringae* and a fungal plant pathogen, *Venturia inaequalis* (Monteil et al. 2013, Leroy et al. 2014). We propose that the recombinant network of nonpathogenic strains represents a natural laboratory for pathogen emergence, within which homologous recombination and HGTs shuffle pathogenicity-related genetic material through various genetic backgrounds. Thanks to favorable conditions, a strain with an appropriate genetic combination may successfully emerge from this pool of variants as a new plant pathogen. In this context, *X. arboricola* can be viewed as a case study as is *Pseudomonas syringae*. It would be of interest to determine whether this proposed scenario of pathogen emergence functions equally well in other plant pathogenic species, such as *Clavibacter michiganensis*, that are composed of both epidemic clones and many nonpathogenic strains (Jacques et al. 2012).

V. Experimental procedures

Bacterial strains and growth conditions

To infer the evolutionary history of *X. arboricola*, we used a collection of 187 strains (collection I; Table II-S1) that included representatives from the species' eight known pathovars (*pruni*, *corylina*, *juglandis*, *celebensis*, *populi*, *arracaciae*, *guizotiae*, and *zantedeschiae*) and nonpathogenic strains isolated from diverse hosts. These latter were shown to be nonpathogenic on their host of isolation in previous studies: (i) strains previously classified as pathovar *fragariae* did not induced symptoms on strawberry plants after spray and vein inoculations (Vandroemme et al. 2013; our unpublished results). (ii) Strains isolated from walnut buds failed to induce symptoms when sprayed on walnut seedlings and, for a subset of strains, it was shown that they were also nonpathogenic on peach, radish, tomato and pepper (Essakhi et al. 2015). (iii) Absence of pathogenicity of strains isolated from bean seeds was demonstrated in the framework of developing a novel method for identification of strains responsible for common bacterial blight of beans (Boureau et al. 2013, ISTA Method, 2014). The remaining nonpathogenic strains were shown to be nonpathogenic on their host of isolation by their isolator (Vauterin et al. 1996). A phylogenetic analysis including representatives of all *Xanthomonas* species showed that all strains from collection I were closer to *X. arboricola* species than to any other species (fig. II-7). A subset of collection I (called collection II) was used in the gene repertoire analysis. Collection II was composed of 73 strains (12 pathogenic strains

representing the eight pathovars of *X. arboricola* and 61 nonpathogenic strains). The strains were freeze dried and stored at the CIRM-CFBP (International Center for Microbial Resources–French Collection of Plant-Associated Bacteria, Angers, France). Prior to analysis, the strains were cultured on YPGA medium (yeast extract, 7 g liter^{-1}; peptone, 7 g liter^{-1}; glucose, 7 g liter^{-1}; agar, 18 g liter^{-1}) for 2–4 days at 28°C or on Trypticase soy agar (TSA; tryptone, 15 g liter^{-1}; Soytone, 5 g liter^{-1}; sodium chloride, 5 g liter^{-1}; agar, 15 g liter^{-1}).

Housekeeping gene sequencing

Partial sequences of seven housekeeping genes (*atpD* [ATP synthase chain], *dnaK* [70 kDa heat shock protein], *efP* [elongation factor P], *fyuA* [transmembrane protein, Ton-B dependent transporter], *glnA* [glutamine synthetase I], *gyrB* [DNA gyrase subunit], and *rpoD* [RNA polymerase sigma 70 factor]) were obtained for all strains of collection I as previously described (Fischer-Le Saux et al. 2015).

Phylogenetic analysis

The partial sequences of the seven housekeeping genes were aligned, trimmed, and concatenated, which yielded fragments of 4,260 nucleotides. The concatenated dataset served as the basis for a maximum likelihood-based phylogenetic tree built using phyML (Guindon et al. 2010) and the GTR+I+gamma model selected with JModelTest (Posada, 2008). A splitgraph was constructed using nucleotide sequences from each sequence type (ST) and the Neighbornet algorithm in Splitstree4 v. 4.6 (Huson & Bryant, 2006). This method highlights phylogenetic conflicts caused by recombination events.

Genetic structure of collection I

To estimate the number of groups in collection I, Bayesian clustering was performed using STRUCTURE v. 2.3 (Pritchard et al. 2000; Falush et al. 2003; Falush et al. 2007) and the partial sequences of the seven housekeeping genes; the linkage model was employed. Five independent runs were performed for each number of populations (K) (K=1 to K=16); there was a burn-in period of 500,000 iterations followed by 1 500,000 Markov chain Monte Carlo (MCMC) iterations. The most likely value for K was found using an analysis of ΔK, a method developed by Evanno et al. (2005) and implemented in STRUCTURE HARVESTER (Earl & vonHoldt, 2011). A consensus of the different runs was obtained using CLUMPP (Jakobsson & Rosenberg, 2007). Results for the other values of K were analyzed using the same method to highlight evolutionary mechanisms. The divergence between the

different groups was estimated using DnaSP (Librado & Rozas, 2009), by employing the average number of net nucleotide substitutions per site between populations (d_A) (Nei, 1987).

Minimum spanning tree

Multilocus sequence typing was performed to characterize diversity within populations using BioNumerics software (Applied Maths, Belgium) and the MLST plugin. This method utilizes allelic data (Table II-S2). For each housekeeping gene, an allele number was assigned to each unique sequence. STs were defined as unique combinations of seven allele numbers. A minimum spanning tree was calculated to explore the relationships among STs. A clonal complex (CC) was defined as a set of at least three strains that shared at least five haplotypes. Singletons were STs that did not group into any CC. For the three groups identified via Bayesian clustering and for each pathovar, DnaSP (Librado & Rozas, 2009) was used to estimate haplotype diversity and molecular diversity indices, as well as to conduct tests of mutation/drift equilibrium (Tajima's D, Fu and Li's D, and Fu and Li's F). These analyses were performed for the seven housekeeping loci.

Estimation of evolutionary parameters

Clonal genealogies were constructed using ClonalFrame (Didelot & Falush, 2007) for the groups identified via Bayesian clustering. Strains that were genetically identical and had been isolated in the same geographical area in the same year were removed from the analysis in order to eliminate the potential bias of sampling. For all genealogies, three independent runs were performed using an MCMC simulation with a burn-in period of 500,000 iterations followed by 1 000,000 MCMC iterations. To obtain a better convergence of the Markov chains, Θ_w was estimated using DnaSP and fixed in the ClonalFrame runs. The convergence among the three runs within a given group was confirmed using the Gelman and Rubin convergence test.

Evolutionary parameters were inferred for each group to identify the major forces contributing to polymorphism. These parameters were estimated using ClonalFrame and were the following: the intragenic recombination rate R; the r/m ratio, which represents the impact of recombination (r) vs. mutation (m) on group diversity; and the ρ/Θ ratio, which is the ratio of the per-site recombination rate and the mutation rate).

To detect recombination events, we calculated the pairwise homoplasy index (PHI), Φ_w, using Splitstree4. When significant, it indicates that recombination has occurred. Recombination events were detected by calculating the standardized index

of association, I_A^s, for the allelic data using LIAN v. 3.5 (Haubold & Hudson, 2000). This statistic reveals the degree of linkage between alleles relative to a null model of random association ($I_A^s = 0$).

Recombinant sequences were identified by examining the concatenated sequences in collection I using RDP 3, GENECONV, BOOTSCAN, MAXIMUM CHI SQUARE, CHIMAERA, SISCAN, and 3SEQ implemented in RDP v. 3.44 (Martin & Rybicki, 2000). We considered that a recombination event was statistically supported when it was detected by two or more methods.

Amplification of genes coding for T3SS, T3E, MCP, TBDT, and adhesins

PCR amplification was used to determine the presence or absence of the following key genes: three genes in the hrp gene cluster (*hrcC*, *hrcN*, and *hrcV*) that encode T3SS; *hrpX*, the gene that encodes a T3SS master regulatory protein; and the 22 genes that encode T3Es. Primers and PCR conditions are provided in Hajri et al. (2012) and in Table II-S6.

The presence of a 35-gene set that codes for MCPs, TBDTs, and adhesins was determined via PCR amplification (Table II-S7). The set of target genes was identified by BLASTn on the genome sequences of four strains of *X. arboricola* (Cesbron et al. 2015) and is composed of 18 genes encoding MCPs, 14 genes encoding TBDTs, and 3 genes encoding adhesins. A set of 14 primer pairs specifically adapted to *X. arboricola* were designed using the alignment of orthologous gene sequences in *X. arboricola*, other primer pairs were described in Mhedbi-Hajri et al. (2011b).The same reaction mixture was used for all primer pairs as described in Mhedbi-Hajri et al. (2011b). The PCR conditions employed were the following: an initial denaturation step at 94°C for 5 min; 30 cycles of denaturation at 94°C for 30 s; an annealing step at 60°C or 65°C, depending on the Tm of the primers (Table II-S7); an extension step at 72°C (which lasted 30 s for PCR products $<$ 1 Kb and 1 min for those $>$ 1 Kb); and a final extension step at 72°C for 7 min.

The PCR approach was validated using the public genomic sequences of 11 strains: one pathovar *corylina* (Ibarra-Cabarello et al. 2013), four pathovar *pruni* (Garita-Cambronero et al. (2014); accession number: PRJDB1518; PRJDB1524; PRJDB1525), two pathovar *celebensis* (accession number: PRJNA254393; PRJNA 252657), two pathovar *juglandis* and two nonpathogenic strains (Cesbron et al. 2015). Among these 11 strains, six were present in our collection (CFBP 7150, CFBP 3523, CFBP 2528, CFBP 7179, CFBP 7634 and CFBP 7651) allowing us to validate the PCR results by BLASTn search in their genomic sequences. In addition, the T3E gene repertoire of the four pathovar *pruni* genome sequences, determined

by BLASTn search was identical to the ones determined by PCR for the pathovar *pruni* strain of collection II. The same observation was made for pathovar *corylina* strains.

Inference of gains and losses of candidate genes

Using GenoPlast (Didelot et al. 2009) and a subset of collection II, evolutionary gains and losses of genes coding for T3SS, T3Es, MCPs, TBDTs, and adhesins were inferred. To construct this collection, a single strain was chosen to represent each unique combination of phylogenetic position and LHT-associated gene repertoire. We included in this analysis the repertoires constructed by BLASTn analysis for the pv. *corylina* strain and the four pv. *pruni* strains. Phylogeny robustness was checked using a ML phylogenetic tree constructed based on the concatenated dataset. The phylogenetic tree used as input in GenoPlast was ultrametric. Three independent runs were performed using a burn-in period of 500,000 iterations followed by 2,000,000 MCMC iterations.

Analyses of *xopR* sequences and genomic environments

To sequence the *xopR* gene step by step from start to finish, we designed *xopR* sequencing primers (Table II-S8). PCR amplifications were performed in 40 μl reaction mixtures containing 1X Go Taq Buffer, 125 μM dNTP each, 0.2 μM of each primer (Table II-S8), 0.8 U of Go Taq polymerase, and 5 μl of bacterial suspension with the following conditions: 5 min of initial denaturation at 94°C; 35 cycles of 94°C for 1 min; an annealing step at 60°C for 30 s; an extension step at 72°C for 1 min; and a final extension step at 72°C for 7 min. Specific conditions (30 cycles; an annealing step at 55°C for 1 min; and an extension step at 72°C for 2 min) were needed when using the primers *xopR*-full-F and *xopR*-full-R. The PCR products were sequenced (Genoscreen).

The *xopR* phylogeny was built using the ML approach described above and congruency of *xopR* and the organismal trees was tested with the Shimodaira-Hasegawa test (Shimodaira & Hasegawa, 1999). The genomic environment of *xopR*, located by BLASTn in the genome sequences, was analyzed using the R package GenoplotR (Guy et al. 2010) and an in-house R script. A subset of 10 strains, chosen so as to represent the diversity of T3E repertoires and strains, was used in this analysis.

Nucleotide accession numbers

The sequences obtained in the frame of this study were deposited in Genbank: for *atpD*, KT874348 to KT874411; for *dnaK*, KT874284 to KT874347; for *efP*,

KT874220 to KT874283; for *fyuA*, KT874156 to KT874219; for *glnA*, KT874092 to KT874155; for *gyrB*, KT874028 to KT874091; for *rpoD*, KT873964 to KT874027; for *xopR*, KT873946 to KT873963. We used the sequences obtained by Fischer-Le Saux et al. (2015): for *atpD*, KP669177 to KP669273; for *dnaK*, KP669274 to KP669370; for *efP*, KP669371 to KP669467; for *fyuA*, KP669468 to KP669564; for *glnA*, KP669565 to KP669661; for *gyrB*, KP669662 to KP669918; for *rpoD*, KP669919 to KP670175. And, we also used the sequences obtained by Essakhi et al. (2015): for *atpD*, KF904442 to KF904417, KF904356; for *dnaK*, KF904543 to KF904518, KF904457; for *efP*, KF904644 to KF904619, KF904558; for *fyuA*, KF904744 to KF904726, KF904724, KF904724; for *glnA*, KF904844 to KF904821, KF904760; for *gyrB*, KF904947 to KF904922, KF904861; for *rpoD*, KF905048 to KF905023, KF905962.

Acknowledgments

We thank Céline Rousseau for the genomic environment analysis and Géraldine Taghouti and Perrine Portier at CIRM-CFBP (INRA, France) for preserving and supplying the strains. Inn-Shik Myung is acknowledged for providing strains from Korea. Jessica Pearce is acknowledged for editing the English. This work was funded by the INRA Plant Health and Environment Division and the regional government of the Pays de la Loire, France.

Supplemental materials

Table II-S1. Strains used in this study and their characteristics.

Nomenclature	CFBP code[a]	Code in other collections[b]	Host of isolation	Geographic origin	Year of isolation	Collection	ST	group
X. arboricola pv. arracaciae	7407^{PT}	NA[d]	Arracacia xanthorrhiza	Brazil	1969	I,II,III	41	B
X. arboricola pv. arracaciae	7403	NA	A. xanthorrhiza	Brazil	1992	I	41	B
X. arboricola pv. arracaciae	7404	NA	A. xanthorrhiza	Brazil	1995	I	41	B
X. arboricola pv. arracaciae	7405	NA	A. xanthorrhiza	Brazil	1995	I	41	B
X. arboricola pv. arracaciae	7406	NA	A. xanthorrhiza	Brazil	2001	I	42	B
X. arboricola pv. celebensis	3523^{PT}	LMG 677, NCPPB 1832, ATCC 19045	Musa acuminata	New Zealand	1960	I,II	25	B
X. arboricola pv. celebensis	7150	NA	M. acuminata			I	73	B
X. arboricola pv. corylina	1159^{PT}	ATCC19313, ICPB XC12, LMG 689, NCPPB 935	Corylus maxima	United States	1939	I,II,III	3	A
X. arboricola pv. corylina	1846	NA	Corylus avellana	France	1975	I,II,III	4	A
X. arboricola pv. corylina	1847	NA	C. avellana	Algeria	1977	I	60	A
X. arboricola pv. corylina	1848	NA	C. avellana	United Kingdom	1977	I,II,III	61	A
X. arboricola pv. corylina	2565	NA	C. avellana	France	1985	I	69	A
X. arboricola pv. corylina	5956	NA	C. avellana	France	1979	I	5	A
X. arboricola pv. corylina	6101	NA	C. avellana	France	1979	I	2	A
X. arboricola	6555	NA	C. avellana			I	59	A

pv. corylina								
X. arboricola pv. corylina	6600	NA	C. avellana	France	2002	I	6	A
X. arboricola pv.guizotiae	7408	NCPPB 1932, ICMP 5743, LMG 731	Guizotia abyssinica	Ethiopy	NA	I,II,III	43	B
X. arboricola pv.guizotiae	7409	NA	G. abyssinica			I	43	B
X. arboricola pv. juglandis	2528 [T]	ATCC 49083, ICMP35, LMG 747, NCPPB 411	Juglans regia	New Zealand	1956	I,II,III	11	A
X. arboricola pv. juglandis	7179[c]	NA	J.regia	France	2002	I,II,III	9	A
X. arboricola pv. juglandis	176	NA	J.regia	France	1961	I	56	A
X. arboricola pv. juglandis	2564	NA	J.regia	Italy	1965	I	10	A
X. arboricola pv. juglandis	2568	NA	J.regia	Italy	1985	I	53	A
X. arboricola pv. juglandis	2632	NA	J.regia	Spain	1984	I	52	A
X. arboricola pv. juglandis	6557	NA	J.regia	Italy	1999	I	53	A
X. arboricola pv. juglandis	7071	NA	Juglans sp.	Spain	1993	I	54	A
X. arboricola pv. juglandis	7072	NA	Juglans sp.	Spain	1993	I	38	A
X. arboricola pv. juglandis	7244	NA	J.regia	France	1978	I	52	A
X. arboricola pv. juglandis	7294	NA	J.regia	France	2001	I	7	A
X. arboricola pv. juglandis	7295	NA	J.regia	France	2002	I	12	A
X. arboricola	7296	NA	J.regia	France	2001	I	8	A

pv. juglandis								
X. arboricola pv. juglandis	7297	NA	J.regia	France	2003	I	9	A
X. arboricola pv. juglandis	7298	NA	J.regia	France	2001	I	13	A
X. arboricola pv. juglandis	7299	NA	J.regia	France	2002	I	9	A
X. arboricola pv. juglandis	7300	NA	J.regia	France	2001	I	9	A
X. arboricola pv. juglandis	7301	NA	J.regia	France	2003	I	9	A
X. arboricola pv. juglandis	7302	NA	J.regia	France	2003	I	9	A
X. arboricola pv. juglandis	7303	NA	J.regia	France	2003	I	9	A
X. arboricola pv. juglandis	7304	NA	J.regia	France	2003	I	9	A
X. arboricola pv. juglandis	7643	NA	J.regia	France	2003	I	9	A
X. arboricola pv. juglandis	12710	NA	J.regia	France	2002	I	55	A
X. arboricola pv. juglandis	12785	NA	J.regia	France	2003	I	9	A
X. arboricola pv. populi	3123^{PT}	ICMP 8923, LMG 12141	Populus X canadensis	Netherlands	1979	I,II,III	64	C
X. arboricola pv. populi	2113	NA	Populus X interamericana	Netherlands	1980	I	62	C
X. arboricola pv. populi	2666	NA	Populus X interamericana	France	1983	I	63	C
X. arboricola pv. populi	2669	NA	Populus X canadensis	France	1987	I	15	C
X. arboricola	2983	NA	Populus X canadensis	Italy	1989	I	16	C

pv. populi								
X. arboricola pv. populi	2985	NA	Populus X interamericana	Belgium	1989	I	65	C
X. arboricola pv. populi	2986	NA	Populus X interamericana	Belgium	1989	I	17	C
X. arboricola pv. populi	3004	NA	Populus X interamericana	France	1989	I	18	C
X. arboricola pv. populi	3121	NA	Salix alba	Netherlands	1980	I	19	C
X. arboricola pv. populi	3122	NA	Salix alba	Netherlands	1979	I	20	C
X. arboricola pv. populi	3124	NA	Populus X generosa	New Zealand	1986	I	21	C
X. arboricola pv. populi	3338	NA	Populus X interamericana	France	1991	I	14	C
X. arboricola pv. populi	3342	NA	Salixsp.	New Zealand	1988	I	22	C
X. arboricola pv. populi	3343	NA	Populussp.	New Zealand	1988	I	23	C
X. arboricola pv. populi	3344	NA	Salixsp.	Belgium	1984	I	24	C
X. arboricola pv. populi	3839	NA	Populus deltoides	Belgium	1984	I	65	C
X. arboricola pv. pruni	411	NA	Prunus persica	USA	1963	I	68	A
X. arboricola pv. pruni	2535	NA	Prunus salicina	New Zealand	1953	I	66	A
X. arboricola pv. pruni	3893	NA	P. persica	Italy	1989	I	68	A
X. arboricola pv. pruni	3898	NA	Prunus domestica	USA	1989	I	68	A
X. arboricola	3900	NA	P. persica	USA	1987	I	68	A

pv. pruni								
X. arboricola pv. pruni	3901	NA	P. armeniaca	USA	1987	I	68	A
X. arboricola pv. pruni	3921	NA	P. persica	Italy	1987	I	68	A
X. arboricola pv. pruni	5229	NA	P. sp.	Argentina	1996	I	66	A
X. arboricola pv. pruni	5529	NA	P. persica	Autralia	1964	I	68	A
X. arboricola pv. pruni	5530	NA	P. persica	Italy	1989	I,II,III	68	A
X. arboricola pv. pruni	5580	NA	P. japonica	France	2000	I	68	A
X. arboricola pv. pruni	5722	NA	P. persica	Brazil	1991	I	68	A
X. arboricola pv. pruni	5723	NA	P. sp.	Uruguay	NA	I	66	A
X. arboricola pv. pruni	5724	NA	P. amygalus	USA	NA	I	66	A
X. arboricola pv. pruni	6653	NA	P. persica	France	2000	I	67	A
X. arboricola pv. pruni	7098	NA	P. domestica	Spain	2001	I	68	A
X. arboricola pv. pruni	7099	NA	P. domestica	Spain	2003	I	68	A
X. arboricola pv. pruni	7100	NA	P. dulcis	Spain	2006	I	68	A
X. arboricola pv. zantedeschiae	7410^{PT}	NCPPB 2978	Zantedeschiae aethiopica	South Africa	1967	I,II,III	44	B
X. arboricola pv.	7411	NA	Z. aethiopica	South Africa	1966	I	44	B

zantedeschiae								
X. arboricola pv. zantedeschiae	7412	NA	Z. aethiopica	China	2002	I	46	B
X. arboricola	NA	BC2527	Capsicum annuum	Korea	2007	I	57	B
X. arboricola	NA	BC2528	Capsicum annuum	Korea	2007	I	57	B
X. arboricola	NA	BC2529	Capsicum annuum	Korea	2007	I	57	B
X. arboricola	NA	BC2667a	Capsicum annuum	Korea	2007	I	57	B
X. arboricola	NA	BC2667b	Capsicum annuum	Korea	2007	I	57	B
X. arboricola	NA	BC2668	Capsicum annuum	Korea	2007	I	57	B
X. arboricola	NA	BC2669	Capsicum annuum	Korea	2007	I	57	B
X. arboricola	NA	BC2670	Capsicum annuum	Korea	2007	I	57	B
X. arboricola	1022	NA	J. regia	France	1967	I,II,III	1	B
X. arboricola	3548*	NA	Fragaria sp.	France	1986	I	26	B
X. arboricola	3549*	NA	Fragaria sp.	France	1986	I	27	B
X. arboricola	4021	ICMP 8452	Magnolia sp.	New Zealand	1983	I	70	B
X. arboricola	4023	ICMP 8457	Magnolia stellata	New Zealand	1983	I	29	B
X. arboricola	6683	NA	Allium cepa	Cuba	NA	I	30	B
X. arboricola	6762*	NA	Fragaria X ananassa	Italy	NA	I	31	B
X. arboricola	6763*	NA	Fragaria X ananassa	Italy	NA	I	31	B
X. arboricola	6770*	NA	Fragaria X ananassa	Italy	1994	I	33	B
X. arboricola	6771*	PD 2780, LMG 19145	Fragaria X ananassa	Italy	NA[d]	I,II,III	34	B
X. arboricola	6772*	NA	Fragaria X ananassa	Italy	NA	I	37	B
X. arboricola	6773*	NA	Fragaria Xananassa	Spain	NA	I	37	B
X. arboricola	6825	NA	Capsicum annuum	USA	2001	I	71	B
X. arboricola	6826	NA	C annuum	USA	1990	I	71	B
X. arboricola	6827	NA	C.annuum	United States	2000	I,II,III	71	B
X. arboricola	6828	NA	C. annuum			I	71	B
X. arboricola	7152	LMG 5402, ICMP 3279	Euphorbia pulcherrima	New Zealand	1972	I	39	B
X. arboricola	7154	NA	E.pulcherrima	New Zealand	1972	I	40	B
X. arboricola	7278	LMG 8676, ICMP 7180	E.pulcherrima	New Zealand	1980	I	40	B
X. arboricola	7413	NCPPB 3200	Chrysanthemum	UK ex Kenya	1979	I	47	B

			morifolium					
X. arboricola	7414	NCPPB 3218	Syzygium1aromaticum	Indonesia	1980	I	48	B
X. arboricola	7415	NCPPB 2856	C. morifolium	Netherlands	1975	I	45	B
X. arboricola	7416	NCPPB 2864	C. morifolium	Netherlands	1975	I	45	B
X. arboricola	7417	NCPPB 2865	C. morifolium	Netherlands	1975	I	49	B
X. arboricola	7418	NCPPB 2866	C. morifolium	Netherlands	1975	I	49	B
X. arboricola	7419	NCPPB 1826	P.s domestica	UK	1966	I	51	B
X. arboricola	7602	LMG 8983	Solanum lycopersicum	NA	NA	I,II	50	B
X. arboricola	7603	LMG 8985	S. lycopersicum	NA	NA	I,II	74	B
X. arboricola	7604	LMG 8987	C. annuum	NA	NA	I,II	75	B
X. arboricola	7605	LMG 8990	S. lycopersicum	NA	NA	I,II	72	B
X. arboricola	7607	LMG 8998	Citrus sp.	NA	NA	I,II	76	B
X. arboricola	7608	LMG 12918	Anthurium andreanum	NA	NA	I,II	77	B
X. arboricola	7609	LMG 12920	A. andreanum	NA	NA	I,II,III	78	B
X. arboricola	7610	LMG 12940	Pueraria robata	NA	1981	I,II,III	79	B
X. arboricola	7611	LMG 12942	Glycine max	NA	NA	I,II	80	B
X. arboricola	7612	LMG 12944	Phaseolus sp.	NA	1986	I,II	81	B
X. arboricola	7613	LMG 12947	S. lycopersicum	NA	1986	I,II	82	B
X. arboricola	7614	LMG12948	Magnolia sp.	NA	NA	I,II,III	84	B
X. arboricola	7615	LMG 12953	Camellia sinensis	NA	1984	I,II,III	85	B
X. arboricola	7616	LMG 12954	Pisum sativum	NA	1985	I,II	86	B
X. arboricola	7617	LMG 12955	P. sativum	NA	1985	I,II	86	B
X. arboricola	7618	LMG 12956	C.coronarium	NA	1985	I,II,III	83	B
X. arboricola	7620	LMG 8995	S. lycopersicum	NA	NA	I,II	87	B
X. arboricola	7622	LMG 12455	Phaseolus vulgaris	United States	1985	I,II,III	88	C
X. arboricola	7629	NA	J. regia	France	2009	I	90	B
X. arboricola	7630	NA	J. regia	France	2009	I,II,III	91	B
X. arboricola	7631	NA	J. regia	France	2009	I,II	94	B
X. arboricola	7632	NA	J. regia	France	2009	I	90	B
X. arboricola	7633	NA	J. regia	France	2009	I	89	B
X. arboricola	7634	NA	J. regia	France	2009	I,II,III	90	B
X. arboricola	7635	NA	J. regia	France	2009	I	94	B

X. arboricola	7636	NA	J. regia	France	2009	I	92	C
X. arboricola	7637	NA	J. regia	France	2009	I	94	B
X. arboricola	7638	NA	J. regia	France	2009	I	90	B
X. arboricola	7639	NA	J. regia	France	2009	I	90	B
X. arboricola	7640	NA	J. regia	France	2009	I,II,III	93	C
X. arboricola	7641	NA	J. regia	France	2009	I	90	B
X. arboricola	7644	NA	J. regia	France	2009	I	94	B
X. arboricola	7645	NA	J. regia	France	2009	I,II,III	92	C
X. arboricola	7646	NA	J. regia	France	2009	I	94	B
X. arboricola	7647	NA	J. regia	France	2009	I,II	90	B
X. arboricola	7648	NA	J. regia	France	2009	I	94	B
X. arboricola	7649	NA	J. regia	France	2009	I,II	94	B
X. arboricola	7650	NA	J. regia	France	2009	I	94	B
X. arboricola	7651	NA	J. regia	France	2008	I,II,III	94	B
X. arboricola	7652	NA	J. regia	France	2008	I,II,III	95	B
X. arboricola	7653	NA	J. regia	France	2008	I,II,III	96	C
X. arboricola	7654	NA	J. regia	France	2008	I	94	B
X. arboricola	7655	NA	J. regia	France	2009	I	94	B
X. arboricola	7656	NA	J. regia	France	2009	I	94	B
X. arboricola	7697	NA	P. vulgaris	NA	2011	I,II	58	B
X. arboricola	8130	NA	P. vulgaris	France	2010	I,II,III	115	B
X. arboricola	8131	NA	P. vulgaris	France	2010	I,II,III	97	B
X. arboricola	8132	NA	P. vulgaris	France	2010	I,II,III	98	B
X. arboricola	8133	NA	P. vulgaris	France	2010	I,II,III	99	B
X. arboricola	8134	NA	P. vulgaris	France	2010	I,II,III	100	B
X. arboricola	8135	NA	NA	NA	NA	I,II	101	B
X. arboricola	8136	NA	P. vulgaris	NA	2011	I,II,III	58	B
X. arboricola	8137	NA	P. vulgaris	NA	2011	I,II	103	B
X. arboricola	8138	NA	P. vulgaris	NA	2010	I,II	105	B
X. arboricola	8139	NA	P. vulgaris	NA	NA	I,II,III	106	B
X. arboricola	8140	NA	P. vulgaris	NA	2010	I,II,III	107	B
X. arboricola	8141	NA	P. vulgaris	NA	2010	I,II,III	108	B

X. arboricola	8141	NA	*P. vulgaris*	NA	2010	I,II,III	108	B
X. arboricola	8142	NA	*P. vulgaris*	NA	2010	I,II,III	109	B
X. arboricola	8143	NA	*P. vulgaris*	NA	2010	I	109	B
X. arboricola	8144	NA	*P. vulgaris*	NA	2010	I,II,III	102	B
X. arboricola	8145	NA	*P. vulgaris*	NA	2010	I,II,III	35	B
X. arboricola	8146	NA	*P. vulgaris*	NA	2010	I,II	35	B
X. arboricola	8147	NA	*P. vulgaris*	NA	2010	I,II,III	104	B
X. arboricola	8148	NA	*P. vulgaris*	NA	2010	I,II	28	B
X. arboricola	8149	NA	*P. vulgaris*	NA	2010	I,II	110	B
X. arboricola	8150	NA	*P. vulgaris*	NA	2010	I,II	36	B
X. arboricola	8151	NA	*P. vulgaris*	NA	2010	I,II	111	C
X. arboricola	8152	NA	*P. vulgaris*	NA	2010	I,II,III	112	C
X. arboricola	8153	NA	*P. vulgaris*	NA	2010	I,II,III	32	B
X. arboricola	8154	NA	*P. vulgaris*	NA	2010	I,II	98	B
X. arboricola	8155	NA	*P. vulgaris*	NA	2009	I,II	113	B
X. arboricola	8156	NA	*P. vulgaris*	NA	2009	I,II	113	B
X. arboricola	8157	NA	*P. vulgaris*	NA	2009	I,II	113	B
X. arboricola	8158	NA	*P. vulgaris*	NA	2009	I,II	113	B
X. arboricola	8159	NA	*P. vulgaris*	NA	2009	I,II	114	B
X. campestris* pv.*campestris	5142[T]	LMG 568, ATCC 33913, ICMP 13, NCPPB 528, CFBP 5251	*Brassicaoleracea var. gemmifera*	United Kingdom	1957	I	NA	

[a] CFBP : CIRM-CFBP, Collection Française des Bactéries associées aux Plantes (INRA Angers, France)

[b] ICMP : International Collection of Micro-organisms from Plants, Auckland, New Zealand

LMG : Belgian Coordinated Collections of Micro-organisms, University of Gent, Belgium

NCPPB : National Collection of Plant Pathogenic Bacteria, York, UK

ATCC : American Type Culture Collection, Manassas, VA, USA

PD : Culture Collection of Plant Pathogenic, Plant Protection Service, Wageningen, Netherlands

[T] : Type strain; [PT] : pathotype strain

[d] NA : Unknown

* :strains previously classified as *X. arboricola* pv. *fragariae*

Table II-S2 : Allelic profiles of MLST analysis

ST	*atpD*	*dnaK*	*efP*	*fyuA*	*glnA*	*gyrB*	*rpoD*
1	1	1	1	1	1	1	1
2	2	2	2	2	2	2	2
3	2	2	2	32	2	2	2
4	2	2	2	2	2	38	2
5	2	2	32	2	2	2	2
6	2	15	2	2	2	2	2
7	3	3	3	2	3	3	3
8	3	3	3	2	4	3	40
9	3	3	25	2	3	3	3
10	3	3	25	2	3	3	40
11	3	3	25	2	3	40	40
12	3	3	25	2	4	3	4
13	3	4	25	2	3	3	3
14	4	35	6	5	9	13	8
15	4	35	27	37	35	5	5
16	4	38	29	5	35	6	6
17	5	35	30	6	5	8	8
18	5	35	30	6	35	8	5
19	6	5	31	7	6	9	9
20	7	5	31	7	6	10	9
21	8	6	5	8	8	12	10
22	8	7	5	8	8	12	10
23	8	8	6	9	10	12	11
24	9	9	6	10	11	14	9
25	10	10	7	4	12	15	12
26	11	11	8	11	12	16	13
27	12	12	9	12	13	17	14
28	12	72	13	64	67	73	78
29	13	14	11	14	15	20	16
30	14	16	25	15	16	21	18
31	15	17	12	16	17	22	19
32	15	39	9	58	33	77	19
33	16	18	13	17	18	23	20
34	17	39	10	12	19	17	21
35	17	39	10	58	19	63	77
36	17	74	55	69	19	35	80
37	18	19	14	18	20	21	22
38	19	3	3	2	4	3	24
39	20	22	16	20	22	25	26
40	21	23	17	21	23	15	27
41	22	24	18	22	24	26	28
42	22	24	18	23	24	27	28
43	23	25	19	24	25	28	29

44	24	26	20	1	17	29	30
45	24	30	24	28	28	33	34
46	25	27	21	25	26	30	31
47	26	28	22	26	17	31	32
48	27	29	23	27	27	32	33
49	28	31	13	29	30	34	35
50	28	41	35	39	39	43	44
51	29	32	34	30	31	35	36
52	30	3	3	2	4	3	4
53	30	3	3	35	4	3	4
54	30	3	3	2	4	3	23
55	30	3	3	3	32	4	3
56	30	3	3	4	34	37	4
57	31	33	26	31	33	36	37
58	31	62	26	53	57	36	66
59	32	2	2	2	2	3	2
60	32	34	2	2	2	3	38
61	33	2	2	2	2	3	2
62	34	35	27	33	35	39	39
63	34	37	28	36	37	41	42
64	34	38	4	5	7	11	5
65	34	38	28	36	35	7	7
66	35	36	2	34	36	3	41
67	35	36	2	34	36	18	17
68	35	36	2	34	36	18	41
69	36	2	2	2	2	3	2
70	37	13	10	13	14	19	15
71	38	40	33	38	38	42	43
72	38	44	24	42	42	46	47
73	39	20	15	19	21	24	25
74	40	42	36	40	40	44	45
75	41	43	23	41	41	45	46
76	42	45	37	43	29	47	48
77	43	46	38	44	43	48	49
78	44	47	38	45	44	49	50
79	45	48	10	46	45	50	51
80	46	49	8	47	46	51	52
81	47	50	13	44	47	52	53
82	48	51	39	45	44	4	54
83	48	51	39	45	44	55	55
84	49	52	40	48	48	53	56
85	50	53	41	49	49	21	57
86	51	52	13	50	50	54	58
87	52	54	42	51	51	56	59
88	53	55	43	52	52	57	60
89	54	11	44	53	53	58	61

90	54	56	10	53	53	58	61
91	54	57	17	35	1	59	62
92	55	58	45	54	54	60	63
93	56	59	46	55	55	61	64
94	57	11	47	56	34	15	13
95	58	60	10	4	32	58	61
96	59	61	48	57	56	62	65
97	60	64	25	59	58	64	68
98	61	65	49	60	59	65	69
99	62	66	50	61	60	66	70
100	63	21	25	62	23	67	71
101	64	67	51	63	61	68	72
102	64	67	51	68	61	68	72
103	65	66	1	62	62	69	73
104	65	71	1	62	66	69	73
105	66	44	52	64	18	70	14
106	67	13	53	65	63	71	74
107	68	68	54	31	33	29	20
108	69	69	20	66	64	72	75
109	70	70	53	67	65	58	76
110	71	73	1	66	66	74	79
111	72	75	46	44	68	75	64
112	73	76	56	70	69	76	81
113	74	77	8	71	17	78	82
114	75	78	57	72	70	79	83
115	17	63	10	58	24	63	67

Table II-S3: Estimates of diversity and neutrality for each group and each locus calculated using DnaSP

group names	locus	Hap	Hd	θ_w	π	Tajima's D	Fu and Li's D	Fu and Li's F
group A (n = 50)	*atpD*	8	0,77	2,9	0,005	0,98	0,49	0,77
	dnaK	6	0,67	2,46	0,005	1,38	-0,86	-0,14
	efP	3	0,62	0,89	0,004	0,97	1	1,16
	fuyA	5	0,56	5,13	0,008	0,62	0,08	0,32
	glnA	6	0,76	2,68	0,006	1,35	-0,14	0,42
	gyrB	7	0,6	5,36	0,004	-1,57	-2,42	-2,52 *
	rpoD	9	0,8	4,02	0,005	-0,73	-1,97	-1,83
	concatenate	23	0,9	23,44	0,005	NA	NA	NA
pathovar *corylina* (n = 9)	*atpD*	4	0,694	1,104	0,0011	-0,93	-0,73	-0,86
	dnaK	3	0,417	1,104	0,00088	-1,51	-1,68	-1,82
	efP	1	0	0	0	NA	NA	NA
	fuyA	1	0	0	0	NA	NA	NA
	glnA	1	0	0	0	NA	NA	NA
	gyrB	3	0,667	4,05	0,008	2,009 *	1,16	1,53
	rpoD	2	0,222	0,37	0,00036	-1,09	-1,19	-1,28
	concatenate	7	0,917	6,62	0,00143	NA	NA	NA
pathovar *juglandis* (n = 24)	*atpD*	3	0,489	1,87	0,004	2,02	1,29 *	1,77 *
	dnaK	2	0,083	0,27	0,0001	-1,16	-1,61	-1,7
	efP	2	0,507	0,8	0,004	2,24 *	0,98	1,54
	fuyA	4	0,308	5,62	0,005	-1,36	-0,59	-0,96
	glnA	4	0,569	2,14	0,002	-0,73	-0,46	-0,62

	gyrB	4	0,239	4,82	0,0026	-2,2 *	-2,5	-2,81 *
	rpoD	5	0,652	3,48	0,004	-1,19	-1,38	-1,54
	concatenate	13	0,83	19,01	0,003	NA	NA	NA
pathovar *pruni* (n = 17)	*atpD*	1	0	0	0	NA	NA	NA
	dnaK	1	0	0	0	NA	NA	NA
	efP	1	0	0	0	NA	NA	NA
	fuyA	1	0	0	0	NA	NA	NA
	glnA	1	0	0	0	NA	NA	NA
	gyrB	2	0,382	0,3	0,0005	0,57	0,68	0,74
	rpoD	2	0,118	0,3	0,00019	-1,16	-1,48	-1,59
	concatenate	3	0,471	0,59	0,00011	NA	NA	NA
group B (n = 114)	*atpD*	52	0,98	12,9	0,011	-1,13	-0,56	0,95
	dnaK	56	0,97	20,38	0,017	-1,22	-1,02	-1,33
	efP	37	0,97	6,06	0,011	-1,33	0,31	-0,49
	fuyA	52	0,98	14,47	0,013	-1,21	-1,15	-1,38
	glnA	48	0,97	9,58	0,007	-1,73	-1,32	-1,88
	gyrB	53	0,98	15,84	0,016	-0,92	-1,17	-1,29
	rpoD	57	0,98	15,44	0,014	-1,52	-1,5	-1,81
	concatenate	68	0,976	98,91	0,01278	NA	NA	NA
pathovar *arracaciae* (n = 5)	*atpD*	1	0	0	0	NA	NA	NA
	dnaK	1	0	0	0	NA	NA	NA
	efP	1	0	0	0	NA	NA	NA
	fuyA	2	0,4	0,48	0,0005	-0,81	-0,82	-0,77
	glnA	1	0	0	0	NA	NA	NA

	gyrB	2	0,4	0,48	0,0005	-0,82	-0,82	-0,77
	rpoD	1	0	0	0	NA	NA	NA
	concatenate	2	0,4	0,96	0,00017	NA	NA	NA
pathovar *celebensis* (n= 2)	*atpD*	2	1	3	0,004	NA	NA	NA
	dnaK	2	1	21	0,03	NA	NA	NA
	efP	2	1	3	0,009	NA	NA	NA
	fuyA	2	1	9	0,012	NA	NA	NA
	glnA	2	1	7	0,01	NA	NA	NA
	gyrB	2	1	11	0,015	NA	NA	NA
	rpoD	2	1	12	0,02	NA	NA	NA
	concatenate	2	1	66	0,01429	NA	NA	NA
pathovar *guizotiae* (n=2)	*atpD*	1	0	0	0	NA	NA	NA
	dnaK	1	0	0	0	NA	NA	NA
	efP	1	0	0	0	NA	NA	NA
	fuyA	1	0	0	0	NA	NA	NA
	glnA	1	0	0	0	NA	NA	NA
	gyrB	1	0	0	0	NA	NA	NA
	rpoD	1	0	0	0	NA	NA	NA
	concatenate	1	0	0	0	NA	NA	NA
pathovar *zantedeschiae* (n = 3)	*atpD*	2	0,667	5,33	0,007	NA	NA	NA
	dnaK	2	0,667	8	0,01	NA	NA	NA
	efP	2	0,667	2	0,006	NA	NA	NA
	fuyA	2	0,667	8	0,011	NA	NA	NA

	glnA	2	0,667	3,33	0,005	NA	NA	NA
	gyrB	2	0,667	10,67	0,015	NA	NA	NA
	rpoD	2	0,667	8,667	0,014	NA	NA	NA
	concatenate	2	0,667	46	0,01	NA	NA	NA
pathogenic on pepper (n=12)	*atpD*	2	0,485	2,98	0,006	1,88	1,41	1,74
	dnaK	2	0,485	3,97	0,008	1,95	1,47	1,81
	efP	2	0,485	1,32	0,006	1,63	1,19	1,48
	fuyA	2	0,485	3,31	0,006	1,91	1,43	1,77
	glnA	2	0,485	0,99	0,002	1,52	1,1	1,37
	gyrB	2	0,485	0,66	0,001	1,36	0,97	1,21
	rpoD	2	0,485	2,31	0,006	1,82	1,35	1,67
	concatenate	2	0,462	15,15	0,004	NA	NA	NA
nonpathogenic strains (n=90)	*atpD*	48	0,965	13,21	0,01	-1,11	-0,61	-0,98
	dnaK	49	0,964	18,18	0,02	-1,08	-1,73	-1,76
	efP	32	0,94	5,92	0,009	-1,42	-0,006	-0,66
	fuyA	47	0,964	14,79	0,012	-1,22	-1,54	-1,69
	glnA	46	0,964	10,06	0,007	-1,76	-1,87	-2,19
	gyrB	48	0,956	15,38	0,015	-0,87	-0,81	-0,99
	rpoD	50	0,965	14,39	0,013	-1,49	-1,48	-1,78
	concatenate	61	0,974	85,8	0,01	NA	NA	NA
group C (n = 23)	*atpD*	13	0,93	14,63	0,019	-0,27	0,29	0,13
	dnaK	14	0,93	13,28	0,018	0,05	-0,68	-0,53
	efP	13	0,95	6,5	0,014	-0,96	0,19	-0,19
	fuyA	15	0,96	21,95	0,03	-0,16	0,14	0,05

	glnA	15	0,93	8,4	0,013	0,15	0,39	0,37
	gyrB	18	0,98	21,68	0,029	-0,22	0,33	0,18
	rpoD	14	0,96	14,09	0,024	0,12	0,001	0,044
	concatenate	22	0,996	108,92	0,023	NA	NA	NA
pathovar *populi* (n = 16)	*atpD*	7	0,858	6,03	0,009	0,57	1,08	1,08
	dnaK	8	0,858	8,74	0,014	0,75	0,4	0,58
	efP	8	0,917	2,71	0,008	-0,2	0,91	0,7
	fuyA	9	0,925	11,75	0,022	1,77	1,39 *	1,73 *
	glnA	9	0,858	4,52	0,008	0,94	0,88	1,04
	gyrB	12	0,958	11,75	0,013	-0,78	-1,25	-1,29
	rpoD	9	0,925	5,12	0,01	0,79	0,41	0,6
	concatenate	16	1	63,59	0,014	NA	NA	NA
nonpathogenic strains (n = 7)	*atpD*	6	0,952	15,92	0,022	0,17	0,32	0,32
	dnaK	6	0,952	13,47	0,017	-0,14	0,075	0,03
	efP	5	0,905	7,34	0,02	0,28	0,484	0,483
	fuyA	6	0,952	23,84	0,03	-0,16	0,082	0,016
	glnA	6	0,952	6,12	0,008	-0,45	-0,192	-0,279
	gyrB	6	0,952	20,81	0,033	0,96	1,4 *	1,44
	rpoD	5	0,905	11,84	0,019	-0,06	0,22	0,17
	concatenate	6	0,952	101,63	0,023	NA	NA	NA

Hap : Number of haplotypes

Hd : Haplotype diversity

θ_w : Watterson estimator

π : Nucleotide diversity (Nei 1987)

* : Significance $p < 0.05$

Table II-S4: PCR results for presence (in grey) and absence (in white) of T3SS and T3E genes.

*ubiquitous T3Es as defined by Hajri *et al.,* 2012

[a]results obtained using hrcC-F and hrcC-R

[b]results obtained using hrcCa1F and hrcC-R

[c]full amplification of the target

[d] results obtained using avrBs2-int-F and avrBs2-int-R

[e]results obtained using avrBs2-Spe-F and avrBs2-Spe-R

[f]2 sets of primers used to confirm the presence of the target (for nonpathogenic strains CFBP7634, CFBP7647, CFBP7614, CFBP7645, CFBP7653, CFBP8153)

[g]results obtained using xopB(1) arbo i F et xopB(1) arbo i R

[h]results obtained using xopB(2) arbo i F et xopB(2) arbo i R

PCR results validate by BLAST analysis

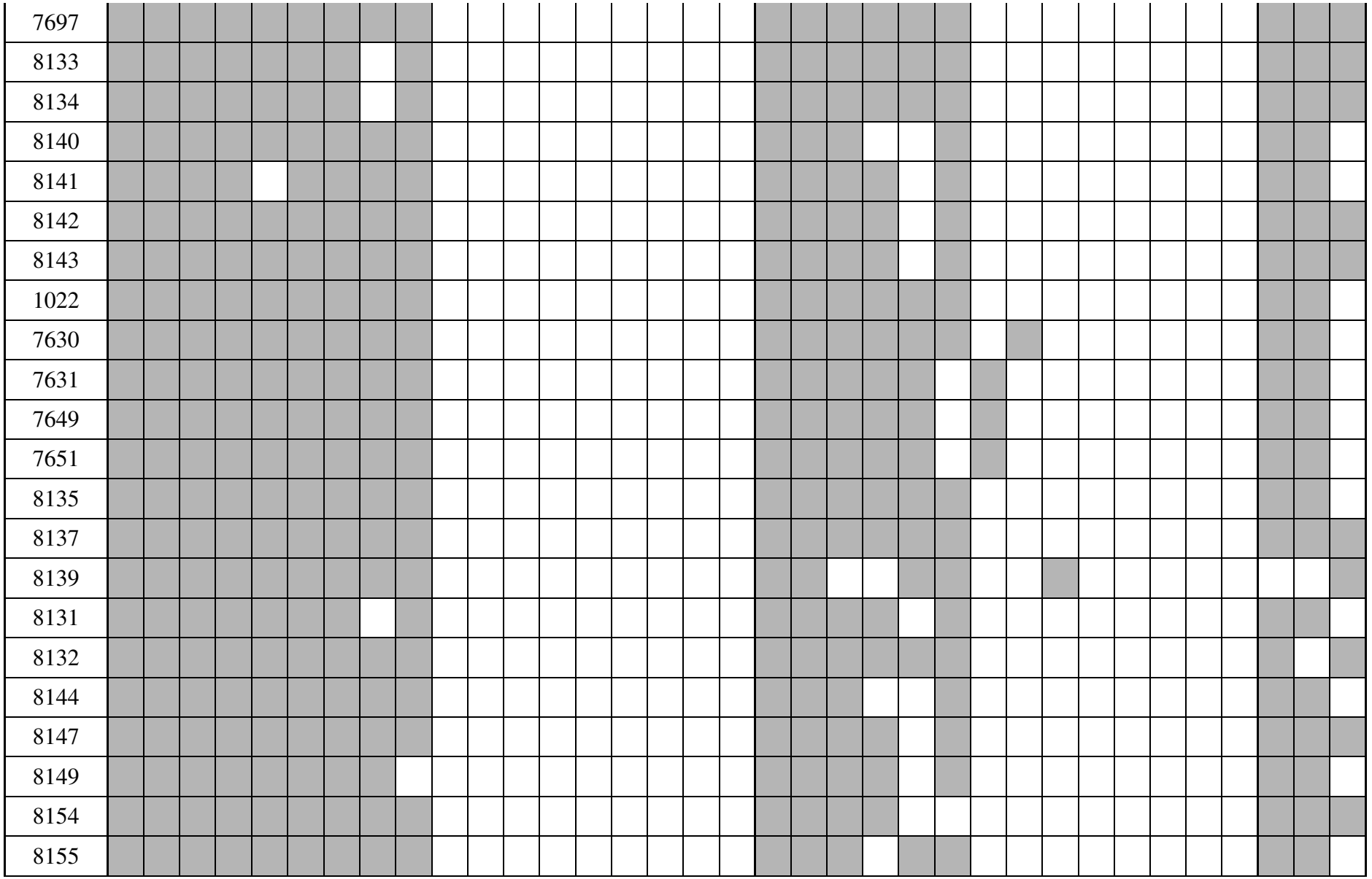
	T3SS					T3Es																								
N° CFBP	*hrpX*	*hrcV*	*hrcN*	*hrcC*[a]	*hrcC*[b]	*xopA**	*hpaA*[c] *	*xopF1**	*avrBs2*[d]*	*avrBs2e**	*xopR**	*hrpW*[c]*	*xopK*	*xopV*	*xopN*	*xopL*	*xopQ*	*xopAI*	*xopAF*	*xopG*[c]	*xopAH*	*xopE3*	*avrXccA2*	*avrBs3*	*xopH*	*xopX*	*xopZ*	*xopB*[g]	*xopB*[h]
5530	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■			■	■		
2528#	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■				■		■			■	■		
7179#	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■					■			■	■	■	■
1159	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■		■	■		■	■		
1846	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■		■	■	■		■			■	■		
1848	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■		■	■	■		■	■	■	■	■		
3523#	■	■	■	■	■	■	■	■	■	■	■	■																	
6771	■																												
3123	■																												
6827	■	■	■	■	■	■	■	■	■	■	■																		
7407	■	■	■	■	■	■	■	■	■	■	■	■			■														■
7408	■	■	■	■	■	■	■	■			■		■		■									■	■	■	■	■	■
7410	■	■	■	■	■	■	■	■	■	■	■	■	■																
7652	■	■	■	■	■	■	■	■	■	■	■	■																	
8136	■	■	■		■	■	■	■	■	■	■	■																	
7697	■	■	■		■	■	■	■	■	■	■	■																	
8133	■	■	■	■	■	■	■	■	■	■	■	■																	

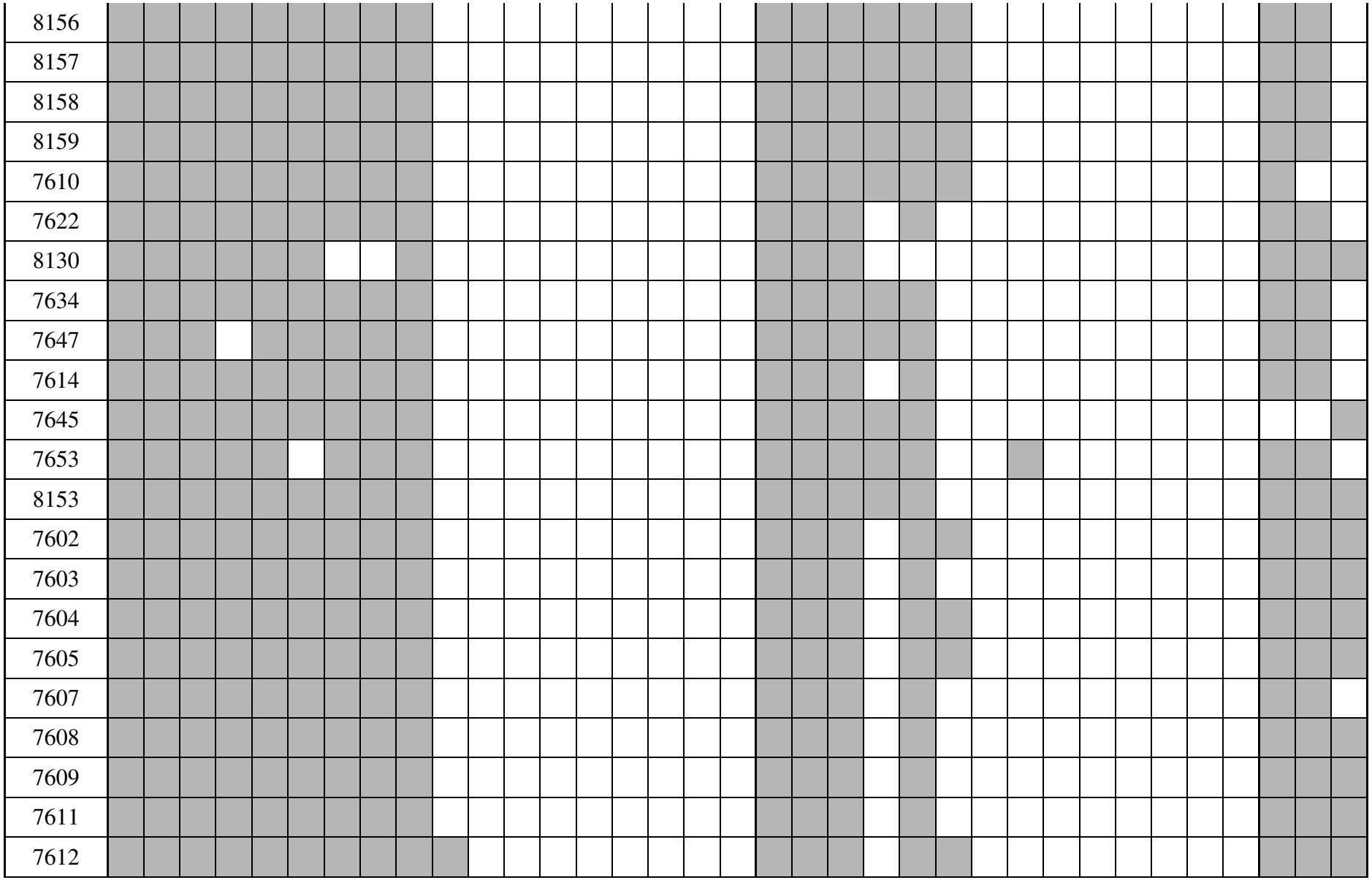
8134																														
8140																														
8141																														
8142																														
8143																														
1022																														
7630																														
7631																														
7649																														
7651#																														
8135																														
8137																														
8139																														
8131																														
8132																														
8144																														
8147																														
8149																														
8154																														
8155																														
8156																														
8157																														
8158																														
8159																														
7610																														
7622																														
8130																														
7634#											f																			
7647											f																			
7614											f																			
7645											f																			
7653											f																			
8153											f																			
7602																														
7603																														
7604																														
7605																														
7607																														
7608																														
7609																														
7611																														
7612																														

7613																												
7615																												
7616																												
7617																												
7618																												
7620																												
7640																												
8138																												
8145																												
8146																												
8148																												
8150																												
8151																												
8152																												

Table II-S5 : PCR results for presence (in grey) and absence (in white) of MCP, TBDT and adhesin genes

CFBP	XCV1933n	XCV1702n	XCV0669n	XCV1952n	XAC3498	XCV3261	XCC0276	XCV1942	XAC3768	XCV1944 n	XCV1955n	XAC3271n	XCC0324n	XCV2152n	XCV2155n	XCC2030n	XCC0108n	XAC2192n	XAC3201	XAC3613	XAC3620	XCC4162n	XCC0120	XAC3077	XCV3187n	XAC3050n	XAC4062n	XCC0394n	XCC2867n	XAC2193 n	XAC0291n	XAC0852n	XadA2	XCV2103	fhaB
5530	■	■	■	■	■	■	■	■	■										■	■	■	■											■	■	■
2528	■	■	■	■	■	■	■	■	■										■	■	■	■		■									■	■	■
7179	■	■	■	■	■	■	■	■	■										■	■	■	■											■	■	■
1159	■	■	■	■	■	■	■	■	■										■	■	■	■	■	■									■	■	■
1846	■	■	■	■	■	■	■	■	■										■	■	■	■		■									■	■	■
1848	■	■	■	■	■	■	■	■	■										■	■	■	■		■									■	■	■
3523	■	■	■	■	■	■	■	■	■										■	■	■	■	■										■	■	
6771	■	■	■	■	■	■	■	■	■										■	■	■		■										■	■	■
3123	■	■	■	■	■		■	■											■	■	■	■	■				■						■		
6827	■	■	■	■	■	■	■	■	■										■	■	■	■	■	■									■	■	
7407	■	■	■	■	■	■	■	■											■	■	■	■	■										■	■	
7408	■	■	■	■	■	■		■											■	■	■	■	■	■	■									■	■
7410	■	■	■	■	■	■	■	■	■										■	■	■	■	■	■									■	■	
7652	■	■	■	■	■	■	■	■	■										■	■	■	■	■										■	■	
8136	■	■	■	■	■	■	■	■	■										■	■	■	■	■	■									■	■	■


7697
8133
8134
8140
8141
8142
8143
1022
7630
7631
7649
7651
8135
8137
8139
8131
8132
8144
8147
8149
8154
8155


8156
8157
8158
8159
7610
7622
8130
7634
7647
7614
7645
7653
8153
7602
7603
7604
7605
7607
7608
7609
7611
7612

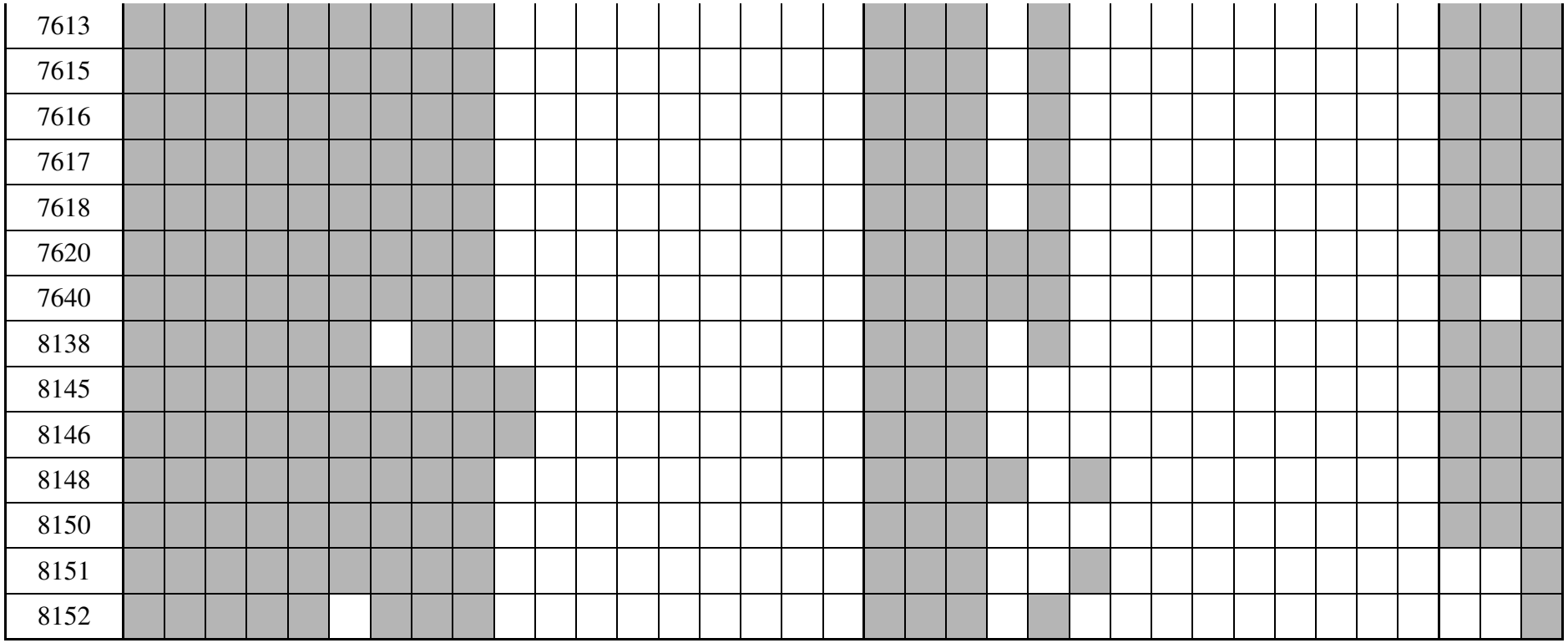

7613
7615
7616
7617
7618
7620
7640
8138
8145
8146
8148
8150
8151
8152

Table II-S6. PCR primers used to amplify T3SS and T3E genes.

Target	Primer Code	F / R	Sequence	Length of the amplified fragment (bp)	Temperature (°C)	PCR conditions	Reference sequences used to design primers
hrpX	hrpXaiF552	F	GCGCCTGGACATGATTGCGC	875	63	initial denaturation step at 95°C for 5 min, 35 cycles : denaturation at 95°C for 40 seconds, annealing at 63°C for 30 seconds and extension at 72°C for 1 min and a final extension step at 72°C for 7 min	hrpX conserved region, in genomes of the 5 strains of X. arboricola pv juglandis CFBP7179, CFBP2528, CFBP7651, CFBP7634 and NCPPB1447.
	hrpXafR1427	R	CATCGGTGCCTCATTGAAATGCTG				
	hrpXafF30	F	CTACTTCGCGGCGATCTCTGCG	1397	53	initial denaturation step at 95°C for 5 min, 35 cycles : denaturation at 95°C for 30 seconds, annealing at 53°C for 30 seconds and extension at 72°C for 1 min and a final extension step at 72°C for 7 min	hrpX conserved region, in genomes of the 5 strains of X. arboricolapvjuglandis CFBP7179, CFBP2528, CFBP7651, CFBP7634 and NCPPB1447.
	hrpXafR1427	R	CATCGGTGCCTCATTGAAATGCTG				
hrcV	hrcV-F	F	GCGCCATGAAATTCGTCAAGG	367	60	Hajri et al, 2012	"conserved regions in Xanthomonas sequenced genomes"
	hrcV-R	R	GCCAGCAGCAGGAACAGC				
hrcN	hrcN-F	F	ATGTCAACGTGATCGTGC	524	60	Hajri et al, 2012	"conserved regions in Xanthomonas sequenced genomes"
	hrcN-R	R	CTGGCTCATCACCCGGCTC				
hrcC	hrcC-F	F	ACCGAAGTGCAGGTGTTTC	575	60	Hajri et al, 2012	"conserved regions in Xanthomonas sequenced genomes"
	hrcC-R	R	ATCTCGATGATGGTGGCATCGAT				

	hrcCa1F	F	TGCTCAGGATCTGGGGCGC	828	55	initial denaturation step at 94°C for 5 min, 30 cycles : denaturation at 94°C for 30 seconds, annealing at 55°C for 30 seconds and extension at 72°C for 30 seconds and a final extension step at 72°C for 7 min and 1 µl of formamide added per 100 µl of PCR reaction	hrcCa1F conserved region of the 4 strains of X. arboricolapvjuglandis CFBP7179, CFBP2528, CFBP7651 and NCPPB1447.
	hrcC-R	R	ATCTCGATGATGGTGGCATCGAT				
xopA	xopA-int-F	F	TGCAGACGATGGGCATCG	239	60	as in Hajri et al, 2012 with PCR conditions as modifications : annealing at 60°C for 30 seconds	CFBP5530
	xopA-int-R	R	CTGCATCAGCTGCATCACGATC				
hpaA	hpaA-full-F	F	ATGATCCGGCGCATTTCGCCAG	816	60	Hajri et al, 2012	CFBP5530
	hpaA-full-F	R	TCATGCACGAATCTCCTGAGCGGC				
xopF1	xopF1-int-F	F	AGGCCATCGACCCCAAGATCC	779	60	Hajri et al, 2012	CFBP5530
	xopF1-int-R	R	GTTCTTGGCCTTGAGCGCATTCC				
avrBs2	avrBs2-int-F	F	TGCCGGTGTTGATGCACGA	850	60	Hajri et al, 2012	CFBP5530
	avrBs2-int-R	R	TCGGTCAGCAGGCTTTC				

	avrBs2-Spe-F	F	GGTGTTGATGCACGATTTCAGCGTTGGCCGCA	838	60	initial denaturation step at 94°C for 5 min, 35 cycles : denaturation at 94°C for 30 seconds, annealing at 60°C for 30 seconds and extension at 72°C for 1 min and a final extension step at 72°C for 7 min. (David Perrine unpublished results)	X. axonopodis
	avrBs2-Spe-R	R	CCTCGGGCTGATCGGTCAACAGGCTTTC				
xopR	xopR-int-F	F	CGTGCGGCCCTGATCGC	303	60	Hajri et al, 2012	CFBP5530
	xopR-int-R	R	GTAGCCCTGCATCATGCGTT				
	xopR-full-F	F	ATGCGCCTGAGTCAGTTGTTT	1230	55	as in Hajri et al, 2012 with PCR conditions as modifications : annealing at 55°C for 1 min	CFBP5530
	xopR-Full-R	R	GTAGCCGTTGTCGATTGCCTCTT				
hrpW	hrpW-full-F	F	ATGCAACGCATGCTCAGCGACAT	905	65	Hajri et al, 2012	CFBP5530
	hrpW-full-R	R	GTCTTCAGGTTCGCCAGCTTCAC				
xopK	xopK-int-F	F	CTCGGCATCCAGGGC	357	60	Hajri et al, 2012	CFBP5530
	xopK-int-R	R	GACAAAGCCCTTGTTCCA				
xopV	xopV-int-F	F	ACACGCCTGTTCGTCTC	236	60	Hajri et al, 2012	CFBP5530
	xopV-int-R	R	GCGATGTTCCATTTGTA				
xopN	xopN-int-F	F	GTCATGACCCAGGGCGC	864	60	Hajri et al, 2012	CFBP5530
	xopN-int-R	R	GGTGATGGCGGTGTGCTG				
xopL	xopL-int-F	F	CCACCGACCGTGGGCGCTTCATCATTA	1324	65	Hajri et al, 2012	CFBP5530
	xopL-int-F	R	ACATCTGCACTGCCTTGGCCAGC				

xopQ	xopQ-int-F	F	ACCCCGACGATGT	484	55	as in Hajri et al, 2012 with PCR conditions as modifications : 35 cyles, annealing at 55°C for 30 seconds and extension at 72°C for 1 min.	CFBP5530
	xopQ-int-R	R	TTGTTGTAGGCGCG				
xopAI	xopAI-int-F	F	AGAGCAGACCACGCCCTCTACG	507	65	Hajri et al, 2012	CFBP5530
	xopAI-int-R	R	GAATATTCTTCGGGAAGCGAGTGC				
xopAF	xopAF-int-F	F	ATAAGCCAATCAAGGGCGTTGG	205	60	Hajri et al, 2012	CFBP5530
	xopAF-int-R	R	ATTTCCCACTGCGCTCTGCGA				
xopG	xopG-full-F	F	ATCGCAGCGAGGATTACCCG	553	60	Hajri et al, 2012	CFBP5530
	xopG-full-R	R	AGACCATTCTCGGCACGAATAG				
xopAH	xopAH-int-F	F	GTGGCTCGCTGGCCCGATC	346	60	Hajri et al, 2012	CFBP5530
	xopAH-int-R	R	GCTGTAGCAGTTTCACTTGCTG				
xopE3	xopE3-int-F	F	CCGACATTGCCGTCAGCGATCACG	381	60	Hajri et al, 2012	CFBP5530
	xopE3-int-R	R	AGCGTTCTTGGGTGTGTTGAGCATTTG				
avrXcca2	avrXcca2-int-F	F	ACGGCCCGTTCTTTCCGCAAAGCC	371	65	Hajri et al, 2012	CFBP5530
	avrXcca2-int-R	R	CAACGGGCGCTCCGGCGACG				
TALE	avrBs3-int-F	F	AGAGCATTGTTGCCCAGTTATCTC	333	55	as in Hajri et al, 2012 with PCR conditions as modifications : denaturation at 94°C for 30 seconds, annealing at 55°C for 30 seconds and extension at 72°C for 30 seconds.	CFBP5530
	avrBs3int-R	R	GTTCGGTGACGCCCACTCT				
xopH	xopH-int-F	F	ATTGTGGTATGGGCCTAGGC	220	60	Hajri et al, 2012	CFBP5530

	xopH-int-R	R	TGCTTGGCGTACTCGTAGAAT				
xopX	xopXarbo i F5	F	GCGATGGTGCTGACGCCCTT	787	65	initial denaturation step at 94°C for 5 min, 30 cycles : denaturation at 94°C for 1 min, annealing at 65°C for 30 seconds and extension at 72°C for 2 min and a final extension step at 72°C for 7 min	CFBP2528
	xopXarbo i R5	R	GACCGCCGCCCCAGTTCATC				
xopZ	XopZarbo_iF	F	ACACGAGCAGAACGCGACCG	534	65	initial denaturation step at 94°C for 5 min, 30 cycles : denaturation at 94°C for 1 min, annealing at 65°C for 30 seconds and extension at 72°C for 2 min and a final extension step at 72°C for 7 min	xopZ conserved region of X. arboricolapvjuglandis CFBP7179 and X. arboricola CFBP2528
	XopZarbo_iR	R	CCGCCCGCGGAGAGTTTCAG				
xopB	XopB(1) arbo i F	F	GGCACCTGACGCGCAGCTTA	587	65	initial denaturation step at 94°C for 5 min, 30 cycles : denaturation at 94°C for 1 min, annealing at 65°C for 30 seconds and extension at 72°C for 2 min and a final extension step at 72°C for 7 min	CFBP7179
	XopB(1) arbo i R	R	GGGCACACGAGGCGAGGAAC				
	xopB(2) arbo i F	F	ACGCGAGCAAGCGAGCACAT	256	65	initial denaturation step at 94°C for 5 min, 30 cycles : denaturation at 94°C for 1 min, annealing at 65°C for 30 seconds and extension at 72°C for 2 min and a final extension step at 72°C for 7 min	CFBP7179
	xopB(2) arbo i R	~~G~~R	TTACGGCTCAGGGTGCGGGT				

Table II-S7. PCR primers used to amplify MCP, TBDT and adhesin genes

Target	Primer code	Sequence	PCRproduct size	Temperature (°C)	Reference	Reference sequences used to design primers[a]
XCV1933	XCV1933-Fwd1	ACTTCGATGGTGCGGTCGCGCCATTC	1142	65	Mhedbi-Hajri *et al,* 2011	Xav 85-10, Xac306, Xcc ATCC33913 and Xoo KACC10331
	XCV1933-Rev1	GCTGGCGTTCGGCATCGTGCTGCTGAT				
XCV1702	XCV1702-Fwd1	TTGACGGTGGAGGTCAGCTCTTCCATC	1352	65	Mhedbi-Hajri *et al,* 2011	Xav 85-10, Xac306, Xcc ATCC33913 and Xoo KACC10331
	XCV1702-Rev1	GTCAACAGCGTCAAGCAACGCTACG				
XCV0669	XCV0669-Fwd1	TCTCGGCCATGATGTCGGTCACGCGCTT	1623	60	Mhedbi-Hajri *et al,* 2011	Xav 85-10, Xac306, Xcc ATCC33913 and Xoo KACC10331
	XCV0669-Rev1	ATTACTTCTTCATCGTCGACCCGCAGATGC				
XCV1952	XCV1952-Fwd1	GTCGATGACGCTGATGATCTCGGCGATCTT	857	65	Mhedbi-Hajri *et al,* 2011	Xav 85-10, Xac306, Xcc ATCC33913 and Xoo KACC10331
	XCV1952-Rev1	CGCTTCTCCGCGCAGACCAAGGTGAT				
XCV1944	XCV1944-Fwd1	CCTGATCGTGCCGGCGCTTCATTTCCT	524	60	Mhedbi-Hajri *et al,* 2011	Xav 85-10, Xac306, Xcc ATCC33913 and Xoo KACC10331
	XCV1944-Rev1	CCGACAACCCGGTGCAGCAGGAACG				
XCV1955	XCV1955-Fwd1	GGCGAGTTGGTTGGCCTGGCGCGCAT	1190	65	Mhedbi-Hajri *et al,* 2011	Xav 85-10, Xac306, Xcc ATCC33913 and Xoo KACC10331
	XCV1955-Rev1	ACCGAGTTGTACGACGAGGTCAAACTGAAACTG				
XAC3271	XAC3271-Fwd1	CCGAGGCGGGGCATGGCGCCAAGCTGGTGGAGATAG	454	65	Mhedbi-Hajri *et al,* 2011	Xav 85-10, Xac306, Xcc ATCC33913 and Xoo KACC10331
	XAC3271-Rev1	GAGCCGCGGCCAGCGAACCCCTGGTGCGCCACTGCGT				
XCC0324	XCC0324-Fwd1	TGGGCGCTGCGTCTTCCACAAAAAGG	1770	65	Mhedbi-Hajri *et al,* 2011	Xav 85-10, Xac306, Xcc ATCC33913 and Xoo KACC10331
	XCC0324-Rev1	ATGCACTTCGGCGTCTTCGTGATCCTG				
XCV2152	XCV2152-Fwd1	TGACACCCTTATGGATGCGCTTCATCCTG	1996	65	Mhedbi-Hajri *et al,* 2011	Xav 85-10, Xac306, Xcc ATCC33913 and Xoo KACC10331
	XCV2152-Rev1	CAATTGCAGCAACGCTCCGGGCAAAAAC				
XCV2155	XCV2155-Fwd1	AACTCGGACTGGCGCTGTACCACCAAG	1220	65	Mhedbi-Hajri *et al,* 2011	Xav 85-10, Xac306, Xcc ATCC33913 and Xoo KACC10331
	XCV2155-Rev1	GCAACGCCTTGGCGCCGTAATGCAGT				
XCC2030	XCC2030-Fwd1	TAAAGGTAAAGGTGGTGCCGGCCAACA	926	65	Mhedbi-Hajri *et al,* 2011	Xav 85-10, Xac306, Xcc ATCC33913 and Xoo KACC10331

	XCC2030-Rev1	CGGTCATCTTCAGCCATGCCAGCCA				
XCC0108	XCC0108-Fwd1	CAACTCTCCTGCGTGATCTGCTCAAA	399	58	Mhedbi-Hajri *et al,* 2011	Xav 85-10, Xac306, Xcc ATCC33913 and Xoo KACC10331
	XCC0108-Rev1	GCCACCATCAGGAAGGGACG				
XAC2192	XAC2192-Fwd1	AACGGCATATCGCGCTACAACGAGGC	380	65	Mhedbi-Hajri *et al,* 2011	Xav 85-10, Xac306, Xcc ATCC33913 and Xoo KACC10331
	XAC2192-Rev1	TGTTGAACCGGTTGAACTCGGCCACC				
XCC4162	XCC4162-Fwd1	CGCAGCAGCAGCGCCAGTTGGTA	912	65	Mhedbi-Hajri *et al,* 2011	Xav 85-10, Xac306, Xcc ATCC33913 and Xoo KACC10331
	XCC4162-Rev1	ACACCGGCCCGTTGCGCCATCA				
XCC0394	XCC0394-Fwd1	AAGTTGCAACGTATTGGCAGTGTCCGCATA	837	65	Mhedbi-Hajri *et al,* 2011	Xav 85-10, Xac306, Xcc ATCC33913 and Xoo KACC10331
	XCC0394-Rev1	CGTAGATGCCAACCGAGCCCACTTC				
XCC2867	XCC2867-Fwd1	AACTGCGCAATGGTGTTCCGGTCTGCGTTT	729	65	Mhedbi-Hajri *et al,* 2011	Xav 85-10, Xac306, Xcc ATCC33913 and Xoo KACC10331
	XCC2867-Rev1	GACGTCCAGGTGAAGGGCTCGAGTG				
XAC2193	XAC2193-Fwd1	GCCTGGCGGCGGTCTCCCAACAA	790	65	Mhedbi-Hajri *et al,* 2011	Xav 85-10, Xac306, Xcc ATCC33913 and Xoo KACC10331
	XAC2193-Rev1	GGCGCGGCGCAAGGGCGTATAGA				
XAC0291	XAC0291-Fwd1	AGCTTGCCATCGGCAATGTCGGGACG	1231	65	Mhedbi-Hajri *et al,* 2011	Xav 85-10, Xac306, Xcc ATCC33913 and Xoo KACC10331
	XAC0291-Rev1	GAAGACGGCTATTACGTGCGCAGATACATC				
XAC4062	XAC4062-Fwd1	CCGCCTGCGTTCGCTGGAGCAATTCA	1442	65	Mhedbi-Hajri*et al,* 2011	Xav 85-10, Xac306, Xcc ATCC33913 and Xoo KACC10331
	XAC4062-Rev1	CCAAAATCCTGCCCCGCATTGACGATGAAGATATT				
XAC0852	XAC0852-Fwd1	CAAGTTTCTGTTCGACGGTCTGCCGATCA	1470	65	Mhedbi-Hajri *et al*, 2011	Xav 85-10, Xac306, Xcc ATCC33913 and Xoo KACC10331
	XAC0852-Rev1	AAAGCCGCTGCCATGCGCGTATTGCAACT				
XAC3050	XAC3050-Fwd1	GCTTTTTTCGGAACCCCCGGCAGGCAAA	831	65	Mhedbi-Hajri *et al,* 2011	Xav 85-10, Xac306, Xcc ATCC33913 and Xoo KACC10331
	XAC3050-Rev1	GGTGCCATTGCAAAGCGCAAAGCGAC				
XCV3187	XCV3187-Fwd1	TCCAACCGCAGGCTCGCCGGCAATTA	1285	65	Mhedbi-Hajri *et al,* 2011	Xav 85-10, Xac306, Xcc ATCC33913 and Xoo KACC10331
	XCV3187-Rev1	GGCCGGGCTGATACAGGTAAGCC				
XCV1942	XCV1942-Fwd1	ATCACCCGCAGCCTGACCCA	1327	60	This work	Xav85-10, CFBP2528
	XCV1942-Rev1	GGTTGACCTGCTCGATGCCCG				
XCV3261	XCV3261-Fwd1	GACGCCAACACCAACGTGGCA	1113	60	This work	Xav85-10, CFBP2528

	XCV3261-Rev1	GCGCATCTGTGCGAACACGC				
XAC3768	XAC3768-Fwd1	CGTGATCCAGGCGCAGACCG	1052	60	This work	Xac306 , CFBP2528
	XAC3768-Rev1	CCGAGCCTTCGGCCACCTTG				
XCC0276	XCC0276-Fwd1	GCATGCACGGCAACCAGCAA	635	60	This work	Xcc ATCC33913, CFBP2528
	XCC0276-Rev1	CTGCTGCACCGACTGCACGA				
XCC0120	XCC0120-Fwd1	CGCGGTTTGGGCTCGGACTT	1843	60	This work	Xcc ATCC33913, CFBP7651
	XCC0120-Rev1	TGGCCACGCCGAAGAACGAC				
XAC3620	XAC3620-Fwd1	GGCCGATTCCTGCTTGCCGT	1698	60	This work	Xac306 , CFBP2528
	XAC3620-Rev1	ATCGATCTGCGCGGCATGGG				
XAC3201	XAC3201-Fwd1	ACGCCGTTTTCTGCGGTGCT	691	60	This work	Xac306 , CFBP2528
	XAC3201-Rev1	ACCCTGGCTTGAACGCGACG				
XAC3077	XAC3077-Fwd1	AACATCGGCCGCATCGCCTC	1514	60	This work	Xac306 , CFBP2528
	XAC3077-Rev1	GCCCACAGGTCGTTCCACGG				
XAC3498	XAC3498-Fwd1	CGACTGCCAGCTCAGGCCAC	1899	60	This work	Xac306 , CFBP2528
	XAC3498-Rev1	CTCCCGCTGGCGTACTCCCT				
XAC3613	XAC3613-Fwd1	TGGTCGTAGCCGCCGGTGTA	1838	60	This work	Xac306 , CFBP2528
	XAC3613-Rev1	ACCTCCGATCAGGGCGGCAT				
XadA2	xadA2-Fwd1	CATTGACGATGCGGCGCGTG	728	60	This work	CFBP2528, CFBP7179, CFBP7651, CFBP7634
	xadA2-Rev1	TGGCGAGAGCAGTGTTGCCG				
XCV2103	XCV2103-Fwd1	CGCCATCGGCCAGCAACTCAA	1057	60	This work	Xav85-10, CFBP2528
	XCV2103-Rev1	ATTGCTGCCGAGCGAGAGCC				
fhaB	fhab2-Fwd1	GGTCGCACCTGTCGCGTCCAACG	1744	58	This work	CFBP2528, CFBP7179
	fhab2-Rev1	GAACTCGGTGCCGTGGACGGTTT				
	fhaB2-bis-Fwd1	CTGCAGCTGAGCATGGGCGT	1177	60	This work	Xav85-10, CFBP2528
	fhaB2-bis-Rev1	TGCCGTGGACGGTTTCGTCG				

Table II- S8 : primers used to sequence *xopR*.

Primer code	F R	Sequence	Length	Coordinates *
xopR-200F	F	GCAAGCTGGCATTTGGCGAGCG	22	97-118
xopR-164F	F	GCTGGKTGGGGCTGTTGCTGG	21	133-154
XopR-43F	F	CAGCGCTGGCTATCATGCGTGC	22	254-276
xopR-full-F	F	ATGCGCCTGAGTCAGTTGTTT	21	298-319
XopR-int-F	F	CGTGCGGCCCTGATCGC	17	1060-1076
XopR910F	F	ACCGGCGCGAAGTGCTCAAGG	21	1208-1228
xopRia2R	R	CGAGGGCTGCCGTCCGGG	18	925-942
XopR747R	R	TTGCTCCTGCGGAGTTGCCTC	22	1023-1044
xopRfullR	R	GTAGCCGTTGTCGATTGCCTCTT	23	1505-1527
xopR342R	R	ACGCTGATCGACGCGGGCC	19	621-639
xopR1439R	R	ATCCGTTTCCGCACTGGCAGCG	22	1715-1730

F : Forward primer

R : reverse primer

*: The coordinates correspond to primer position on CFBP 2528 *xopR* sequence

Chapitre II

Introduction

La génétique des populations basée sur sept gènes de ménage a montré que l'espèce *X. arboricola* était composée de trois groupes génétiques, un groupe composé des trois clones épidémiques et deux groupes composés de pathovars à faible incidence et de souches commensales. Aucun événement de flux de gènes n'a pu être détecté entre ces groupes. Afin de déterminer si cette absence de flux de gènes concerne l'ensemble des génomes, nous avons réalisé une étude de génomique des populations. L'objectif de ce chapitre[9] était donc d'étudier la recombinaison homologue entre les clones épidémiques et le réseau recombinant composé majoritairement de souches commensales, et également de déterminer l'importance du transfert horizontal de gènes chez ces deux types de population en comparant leurs contenus en éléments génétiques mobiles. En effet, ces deux mécanismes moléculaires peuvent être impliqués dans l'émergence de nouveaux agents pathogènes (Bartoli et al. 2016). L'ensemble des analyses a été réalisé sur une collection de 44 séquences génomiques. Afin d'étudier la recombinaison homologue, nous avons tout d'abord étudié le polymorphisme partagé entre les clones épidémiques et le réseau recombinant. Cette analyse a mis en évidence l'existence d'un polymorphisme partagé entre les deux populations et a montré que le sens du flux de gènes se réalisait des clones épidémiques vers le réseau recombinant. Ensuite, une analyse de scénarios de divergence a permis de valider le sens du flux de gènes et a également montré que ce flux de gènes était persistant tout au long de la divergence entre les clones épidémiques et le réseau. De plus, cette analyse a montré que les événements de flux génique concernent l'ensemble du génome des souches recombinantes, tandis que les rares événements de recombinaison homologue provenant des souches recombinantes vers

[9]Pour ce travail, Martial Briand m'a guidée pour l'utilisation des outils bio-informatiques liés à la génomique des populations. Adrien Rieux a réalisé les analyses d'estimation de taux de mutations et de datations moléculaires. Louis Legoff a réalisé les recherches d'éléments génétiques mobiles dans les génomes. Mahendra Mariadassou a écrit les scripts R permettant d'identifier les coordonnées des régions recombinantes dans les génomes. Marie-Agnès Jacques et Marion Fischer-Le Saux m'ont encadrée pour poser les questions biologiques, pour l'analyse des données, pour leur interprétation et la rédaction du manuscrit. J'ai réalisé l'ensemble des analyses, excepté celles concernant la datation moléculaire et j'ai également encadré Louis Legoff en stage de Master 1 pour la recherche des éléments génétiques mobiles.

les clones épidémiques ne concernent que des ilots génomiques. Enfin, la comparaison des contenus en éléments génétiques mobiles a montré qu'ils étaient présents en plus grand nombre dans les clones épidémiques qu'au sein du réseau recombinant. Cela suggère donc que les souches pathogènes ont évolué grâce à l'acquisition de matériel génétique par transfert horizontal. Ces résultats correspondent à ce qui a déjà été observé chez les agents pathogènes de l'homme, et les individus commensaux proches génétiquement, comme *Chlamydia trachomatis* (Joseph et al. 2012), et *Escherichia coli* (Didelot et al. 2012). Cet article est rédigé afin d'être soumis prochainement à ISME journal.

Persistence of an asymmetric gene flow from pathogenic strains to commensal ones along divergence of epidemic clones and recombinogenic background.

Merda D., Briand M., Rieux A., Legoff L., Mariadassou M., Jacques $MA^{\$}$., Fischer-Le Saux $M^{\$}$.

$ These authors contributed equally to this work.

I. Introduction

Genetic recombination has a major role in disease emergence. Engering et al. (2013) proposed a conceptual framework to study disease emergence based on three scenarii: (i) expansion of the pathogen geographical area, (ii) colonization of a new host, and (iii) acquisition of novel traits (e.g. xenobiotic resistance). Recombination has been shown to be involved in the two latter scenarii. This evolutionary force is composed of two mechanisms, homologous recombination on the one hand and gene acquisition by Horizontal Gene Transfer (HGT) on the other hand (Didelot et al. 2012). Homologous recombination, which could be assimilated to introgression, plays a major role in emergence of new fungal pathogens (Stuckenbrock, 2016). In bacteria, homologous recombination can also drastically change the allelic profile and be responsible for colonization of a new host. For instance the plant pathogenic bacteria *Xylella fastidiosa* subsp. *morus* would result from hybridization through homologous recombination between the subspecies *fastidiosa* and *multiplex* (Nunney et al. 2014). A host shift has been associated to this homologous recombination

event leading *X. fastidiosa* subsp. *morus* to be pathogenic on mulberry. HGT can also rapidly change bacterial phenotype through acquisition of a new trait and lead to host adaptation. This was demonstrated in the bacterial plant pathogen *Pantoea agglomerans*, for which acquisition of a plasmid encoding virulence genes by a commensal strain led to the emergence of a host-specialized pathogen (Barash & Manulis-Sasson, 2009).

Recombination impacts population structure and three bacterial population structures were defined according to the intensity of recombination: clonal, panmictic and epidemic (Maynard-Smith et al. 1993, 2001). An epidemic population structure is composed of clones responsible for epidemics presenting few recombination events in their genomes and of a recombinant network of freely recombining strains. This recombinant network could be considered as a natural laboratory for emergence of new pathogens within which homologous recombination and HGT of virulence factors represent the main evolutionary force (chapter 1). For instance, the human pathogen *Vibrio parahaemolyticus*, presents an epidemic population structure (Gonzalez-Escalona et al. 2008) within which two pandemic ecotypes have emerged thanks to lateral acquisition of 10 new genes involved in adhesion and adaptation to extracellular stimuli (Turner et al. 2016).

Mobile Genetic Elements (MGE) are gene carriers responsible for DNA exchanges among bacteria. Plasmids have the ability to self-replicate and could contain genes coding for adaptive functions such as antibiotic resistance or virulence. For example, the plasmid responsible for pathogen emergence in *Pantoea agglomerans* harbors a Pathogenicity Island (Manulis & Barash, 2003; Barash & Manulis-Sasson, 2007). However, plasmid stability in bacterial cells is dependent on fitness gain and absence of incompatibility (Hyland et al. 2014). Indeed, incompatibilities could exist between the newly acquired plasmid and existing plasmids or chromosomic genome (Bennett, 2008), as Gram-negative plasmids are not maintained in Gram-positive bacteria or eukaryotic cells. ISs represent another class of MGEs. ISs are small DNA molecules containing one transposase encoding gene and delimited on both sides by inverted repeated regions. ISs are classified in 25 families based on transposase homology (Chandler & Mahillon, 2002). They play a role in acquisition or loss of genes as a consequence of homologous recombination and HGT, and they could play a role in gene expression modification following genomic rearrangements (Sousa et al. 1997; Bennet et al. 2004). Finally, prophages are bacteriophages (i.e. virus infecting bacteria) that integrate into bacterial genomes. The prophages have a role in HGT because they carry genes called "passenger compartments" (Pallen & Wren, 2007). So, they could allow bacteria to become pathogenic thanks to virulence factor acquisition. Indeed, genes encoding toxins, virulence and resistance factors are among the genes carried by prophages (Canchaya et al. 2004; McCann

Encadré IV : ClonalFrame et la détection des séquences recombinantes

ClonalFrame est un programme d'inférence bayésienne permettant d'étudier la microévolution des bactéries, en utilisant des données MultiLocus Sequence Typing (MLST), souvent basées sur des gènes de ménage (Didelot & Falush, 2007). Une version plus rapide a été développée pour l'utilisation de génomes entiers, ClonalFrameML (Didelot & Wilson, 2015). Contrairement aux modèles phylogénétiques, cette méthode prend en compte les événements de mutation et de recombinaison homologue pour reconstruire une généalogie clonale de l'échantillon. Le modèle de coalescence utilisé correspond au modèle de Kingman 1982, dans lequel la taille de l'échantillon est supposée constante au cours du temps.

Cette méthode peut être utilisée pour détecter les séquences recombinantes dans un alignement nucléotidique. En effet, il permet d'obtenir les coordonnées des séquences recombinantes au sein de l'alignement. Pour cela, le logiciel calcule la probabilité d'un événement de recombinaison en fonction de la localisation des sites polymorphes le long de la séquence d'ADN. Ainsi, plus les sites polymorphes sont proches, plus la probabilité que ce polymorphisme soit dû à un événement de recombinaison est importante (figure).

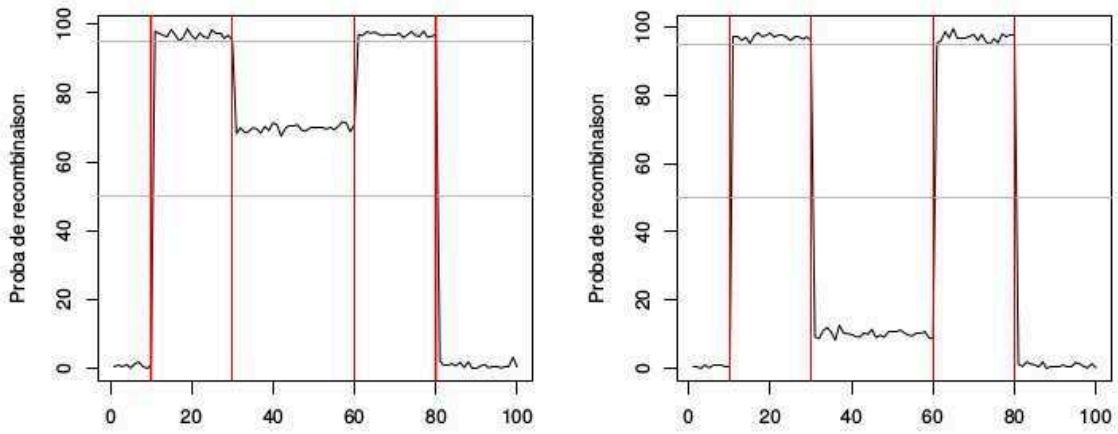


Figure : A gauche, un événement de recombinaison. A droite, deux événements de recombinaison (Mariadassou, com pers). L'axe des ordonnées correspond à la probabilité de recombinaison. L'axe des abscisses correspond aux sites nucléotiques de l'alignement d'ADN.

Pour la détection des séquences recombinantes dans le chapitre 2, ClonalFrame a été utilisé. Le script R (Mariadassou, com pers) a permis d'extraire les coordonnées de ces séquences. Pour cela, nous avons considéré qu'un événement correspondait à de la recombinaison homologue, si sa probabilité était de plus de 95%. De plus, nous avons considéré qu'il ne s'agissait que d'un seul événement si deux événements étaient séparés par un événement ayant une probabilité de plus de 50 % (à gauche sur la figure). Au contraire, si ces deux évènements de recombinaison homologue sont séparés par un événement ayant une probabilité inférieur à 50 %, les deux événements sont considérés indépendants.

& Guttman, 2008).

Xanthomonas arboricola is a bacterial species, which encompasses both pathogenic and commensal strains. It presents an epidemic population structure (chapter 1) and represents a model of choice to disentangle the evolutionary forces and mechanisms acting in the transition between commensalism and pathogenicity in this major genus of plant associated bacteria. *X. arboricola* pathogenic strains are responsible for plant diseases and cause economically important damages in stone and nut fruit trees. These pathogenic strains are host-specialized and grouped in an infrasubspecific taxonomic range called pathovar, which gathers strains responsible for the same symptoms on the same host range (Dye, 1980). Among the eight pathovars belonging to *X. arboricola*, three are considered successful because they are worldwide disseminated and epidemics are frequently reported (chapter 1). This is the case of pathovars *corylina*, *juglandis* and *pruni*, responsible for bacterial blight on hazelnut, walnut blight, and bacterial spot on stone fruit trees, respectively. The five other pathovars are pathogenic on banana (pv. *celebensis*), poplar (pv. *populi*), arracacha (pv. *arracaciae*), niger (pv. *guizotiae*) and arum lily (pv. *zantedeschiae* (Fischer-Le Saux et al. 2015). These five pathovars are considered unsuccessful as they present a limited geographical expansion and epidemics are merely rarely reported (chapter 1). *X. arboricola* is also characterized by the presence of numerous commensal strains isolated from diverse plants. Commensal strains and unsuccessful pathovars form a recombinant network, from which have emerged the successful pathovars representing epidemic clones (chapter 1).

Several elements suggest the importance of recombination in *X. arboricola* evolution. Firstly, homologous recombination was shown to be the major evolutionary force acting in the recombinant network of *X. arboricola* (chapter 1). Secondly, a type three effector, *xopR*, known to suppress plant innate immunity (Akimoto-Tomiyama et al. 2012) was shown to be horizontally transferred from an epidemic clone (i.e. the pathovar *corylina*) to a commensal strain (chapter 1). Thirdly, epidemic clones have sequentially accumulated numerous type three effectors (T3Es) and would hence be HGT-prone (chapter 4). Lastly, comparative genomics between four pathogenic and commensal strains suggested that ISs would be major players of epidemic clones genome plasticity (Cesbron et al. 2015). Does gene flow massively occur between epidemic clones and commensal strains and could it contribute to acquisition of virulent traits by commensal strains that could favor future emergences? To answer these questions, we deciphered the recombinant sequences in core genomes of 44 *X. arboricola* strains and determined their origin. The role of HGT and relative importance of different MGEs in shaping *X. arboricola* diversity was elucidated by comparing the mobilomes of epidemic clones and of the background commensal population.

Encadré V : Inférence des scénarios de divergence avec δaδi

δaδi (Gutenkunst et al. 2009) est une méthode d'inférence de scénarios évolutifs basée sur une méthode de diffusion. Cette méthode permet d'obtenir des spectres de fréquence joints dont les profils diffèrent en fonction des différents scénarios évolutifs. En effet, les événements démographiques affectent le spectre de fréquence d'une population. Les spectres de fréquence joints correspondent à une matrice représentant la fréquence de chaque allèle dérivé au sein des populations testées (figure).

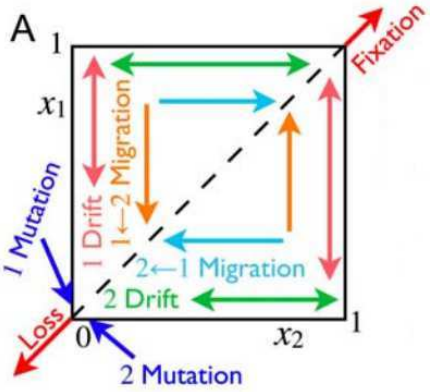


Figure représentant l'action des différentes forces évolutives sur le spectre de fréquence joint (d'après Gutenkunst et al. 2009).

Sur cette figure, chaque axe représente la fréquence d'un allèle dérivé dans les deux populations étudiées. Ainsi, les allèles partagés entre les deux populations se retrouvent proches de la diagonale sur la figure, tandis que les allèles propres à chaque population vont être représentés proches des axes. La migration (représentée par la recombinaison homologue chez les bactéries) aura donc tendance à homogénéiser les populations et donc à augmenter le nombre de sites situés sur la diagonale. Au contraire, la dérive génétique qui augmente la différenciation entre deux populations aura tendance à augmenter le nombre de sites situés proches des axes. Ainsi, il est possible de simuler des spectres de fréquence joints selon différents scénarios de divergence et de les comparer avec le spectre observé avec nos données empiriques. Par une approche de comparaison des vraisemblances, il est possible de retrouver le scénario le plus probable qui explique notre jeu de données.

Cette méthode présente une limite majeure, puisqu'elle est dépendante des scénarios que l'on teste. Ainsi, on obtient le scénario le plus probable parmi ceux que l'on teste mais pas nécessairement le scénario ayant conduit aux données observées.

II. Materials and methods

Genome sequences

A collection of 44 genome sequences representing the diversity of the *X. arboricola* species was used (Table III-S1). It encompasses a set of 23 genomes representing the commensal strains and 21 genomes representing pathogenic strains that belong to all pathovars described in this species (Fischer-Le Saux et al. 2015). This collection was composed of 11, 20 and 6 genomes belonging to groups A, B and C respectively as described in chapter 1 (Table III-S1).

ANI calculations

To study homologous recombination in *X. arboricola* species we needed a dataset that complies with species delimitation criteria. The ANI (average nucleotide identity) values are considered as an alternative to DNA-DNA hybridizations to delineate bacterial species (Richter & Rossello-Mora 2009). ANIs were calculated for each pairwise genome comparisons in the collection, using Jspecies software (Richter & Rossello-Mora 2009). By definition, two strains having an ANIb value above the 95% threshold belong to the same species. A distance matrix based on ANI – d = 100 – (ANI values) – was calculated including all genomes and its visualization was performed using a multidimensional scaling software implemented in Declic (Laizet et al. 2014).

Core genome analysis

Only strains complying with the species delineation criteria were used in a Mauve alignment to obtain *X. arboricola* core genome, including coding and non-coding regions (Darling et al. 2004). A principal components analysis (PCA) was performed using all SNPs present within the core genome alignment and using the R package adegenet (Jombart, 2008). Detection of recombinant sequences within the core genome alignment was performed using ClonalFrame (Didelot et al. 2007), and an R script with ape (Paradis et al. 2004) and coda packages (Plummer et al. 2015) allowed us to discard the recombination events detected[10] (Mariadassou, comm. pers). Three ClonalFrame runs were realized with a Markov chain Monte Carlo (MCMC) of 100,000 iterations, a burn-in length of 50,000 iterations, and iteration samples every 100 iterations.

[10] voir encadré IV

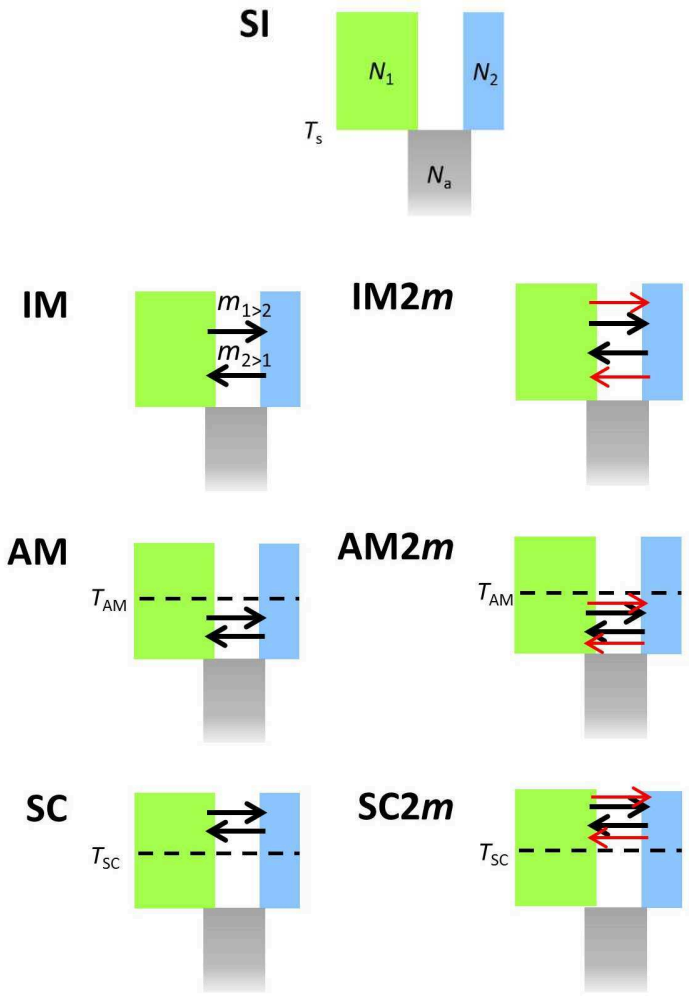


Figure III-1 : The seven tested model in δaδi. SI : Strict isolation; IM: isolation with migration; AM: ancient migration; SC : secondary contact. 2M corresponds to models tested with two migration rates different, one along genome and one in genomic island.

One phylogeny was built using all SNPs of the core genome and another robust one using only the non-recombinant SNPs detected as described above. Phylogenies were built in maximum likelihood using phyML software (Guindon et al. 2010), and the GTR + I + gamma model.

Analysis of gene flow between *X. arboricola* strains

To characterize the gene flow between *X. arboricola* strains (orientation of the flow and size of the recombinant fragments), fineSTRUCTURE v.2 pipeline was used using all SNPs present in the core genome alignment obtained by MAUVE (Darling et al. 2004). This pipeline allows to run ChromoPainter and fineSTRUCTURE (Lawson et al. 2012), this method is based on chromosome painting method, and allows highlighting for each genome the sum of segments received from other genomes included in the analysis. The fineSTRUCTURE run was performed using a MCMC of 1,000,000 iterations, a burn-in of 50,000 and chain samples every 100 iterations. The heat map was obtained using fineSTRUCTURE interface.

To infer the most likely divergence model between groups identified, we used a diffusion-based approach implemented in δaδi[11] (Gutenkunst et al. 2009; Tine et al. 2014). In this analysis, sites were supposed unlinked, so only SNPs distant of at least 1,000 bp were kept. The ancestral state for each SNP was determined using CFBP 7645 genome, a commensal strain belonging to group C, a sister genomospecies of *X. arboricola* that was excluded from *X. arboricola sensu stricto* (this work). Seven models were tested (fig. III-1): (i) strict isolation (SI), which supposed that no gene flow occurred between the two populations, (ii) isolation with migration (IM), which supposed gene flow during divergence, (iii) ancestral migration (AM), which supposed gene flow followed by strict isolation, and (iv) secondary contact (SC), which supposed gene flow after strict isolation. The three models IM, AM and SC were also tested adding a distinction between two kinds of migration rates (2M), corresponding to homogeneous migration in entire genomes and migration occurring only at hotspots in genomic islands. These seven models were compared using likelihood ratio test (LRT) and parameters were estimated from the most likely model. Recombination hotspots were estimated using interval program available in LDhat program (Auton & McVean, 2007). This analysis was performed on the core genome alignment obtained by MAUVE.

[11] voir encadré V

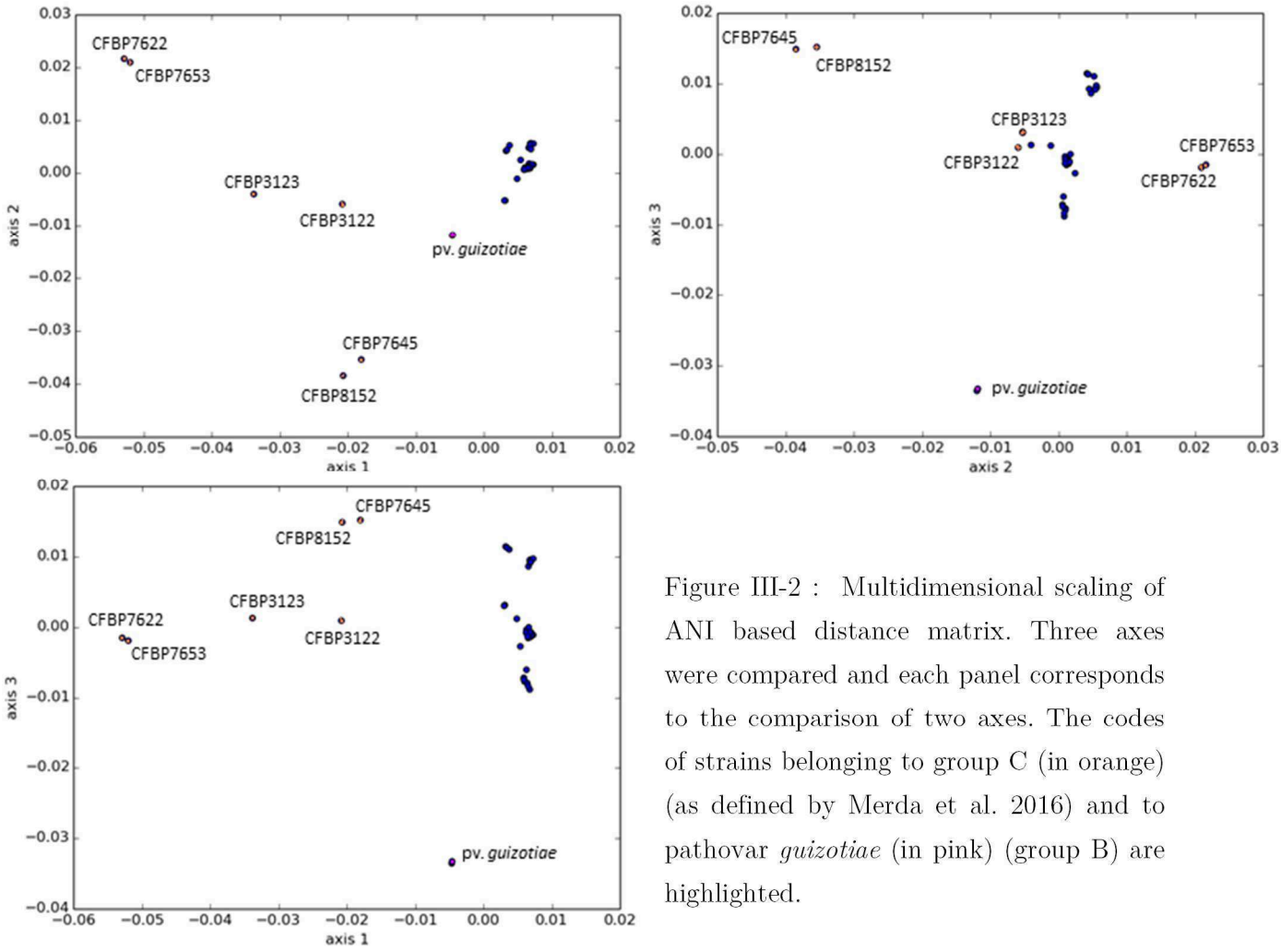


Figure III-2 : Multidimensional scaling of ANI based distance matrix. Three axes were compared and each panel corresponds to the comparison of two axes. The codes of strains belonging to group C (in orange) (as defined by Merda et al. 2016) and to pathovar *guizotiae* (in pink) (group B) are highlighted.

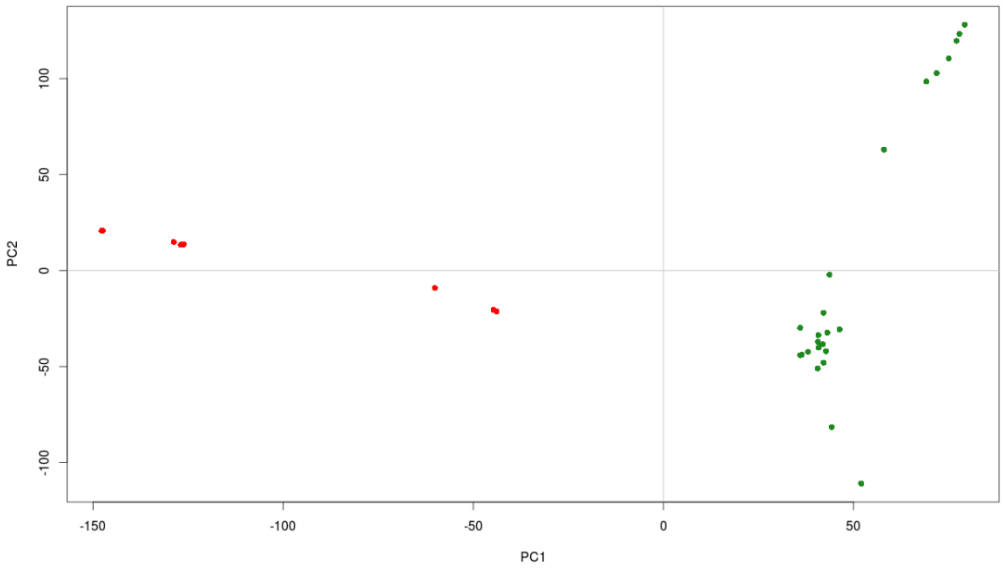


Fig. III-3: Principal component analysis based on 393,610 SNPs found in core genome of 36 strains representing *X. arboricola* species. Axis 1 explains 42 % and axis 2 explains 25 % of variance. Strains belonging to group A are represented by red dots, while groupB strains appear s green dots.

Analysis of gene flow originating from outside of *X. arboricola*

To determine if incoming gene flow occurred from other bacterial genera, a blastn was performed with Blast + (NCBI) using the recombinant sequences detected by ClonalFrame and the NCBI nucleotide database (date of consultation: March 2016). Only the sequences having at least two different best hits (with an e-value inferior to 0.01) from other sources than the *Xanthomonas* genus were kept for the analysis and were considered as having an origin outside of *X. arboricola*. Recombinant sequences originating from outside *Xanthomonas* genus were analyzed using genome annotations to determine if they corresponded to coding sequences.

Detection of mobile genetic elements (MGEs) in *X. arboricola* genomes

Plasmids search was performed using all plasmid sequences publicly available in NCBI. Each retrieved plasmid sequence from the database (date of release: May 2016) was cut in short sequences of 500 bp. These short sequences were compared (blastn) to *X. arboricola* genomes in order to detect sequences with probable plasmidic origin in genome sequences. A short sequence was supposed to be present if both the percentages of sequence identity and length were above 80%. If more than a half of the short sequences of one plasmid were found in a genome, the target sequence was considered as plasmidic origin and was classified according to the plasmid coverage and to a score based on identity percentages and lengths of aligned short fragments. The score was calculated using the average of the products of identity percentage and length percentage obtained for each short sequence complying with the criteria mentioned above : almost identical (AI) corresponded to target sequences with a score superior to 9800 (meaning for instance that the target sequence has 99% of sequence identity with a plasmid sequence over 99 % of length), strong homology (SH) to plasmids with a score between 9000 and 9800 and homologous (H) to plasmids with a score below 9000.

Detection of insertion sequences (IS) was performed with the ISsaga pipeline (Varani et al. 2011), which uses ISfinder database (Siguier et al. 2006). The parameters used for the search were an identity above 97%, word-size of 3 and an e-value of 1e-05. This pipeline only detect complete ISs. As genomic sequences of our collection were assembled in contigs, and as ISs are known to be responsible for assemblage difficulties, the number of ISs detected by ISsaga pipeline was necessarily underestimated. In order to address this bias, we estimated the frequency of each IS detected in *X. arboricola* genomes by counting its inverted repeat regions (IRR) as

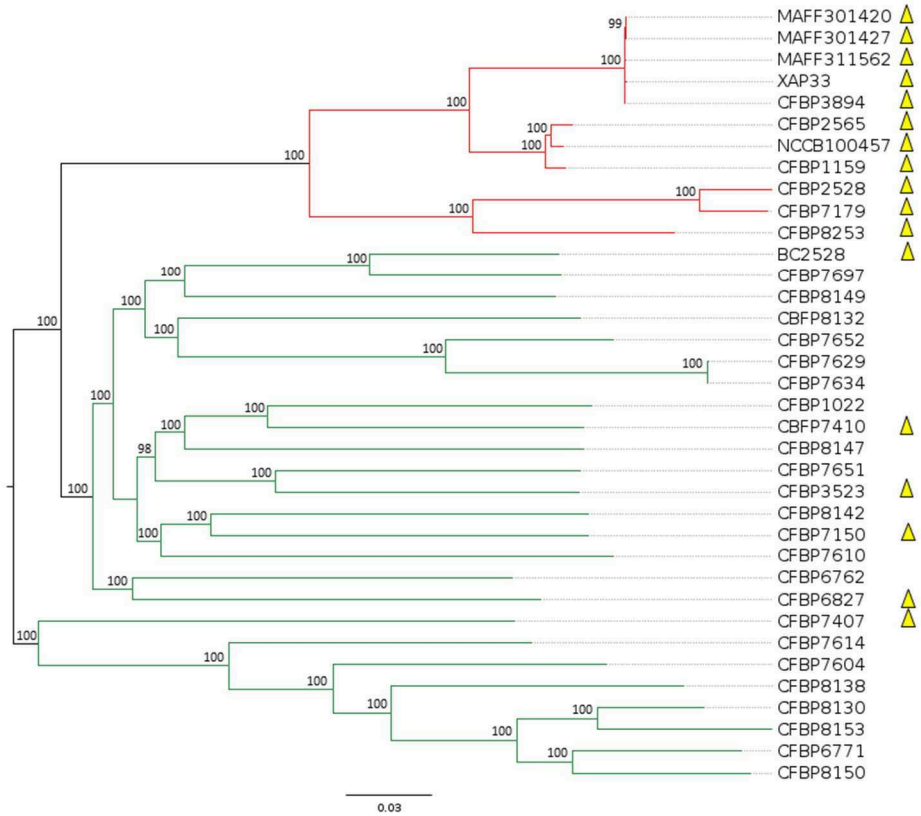


Figure III-4: Maximum likelihood phylogeny of 36 strains representing *X. arboricola* based on non-recombinant SNPs of the core genome (1,983,99 SNPs). The red branches corresponds to group A strains and the green branches corresponds to group B. The yellow triangle indicates pathogenic strains. The bootstrap values were calculated for 100 replicates, and only values superior to 85 were represented.

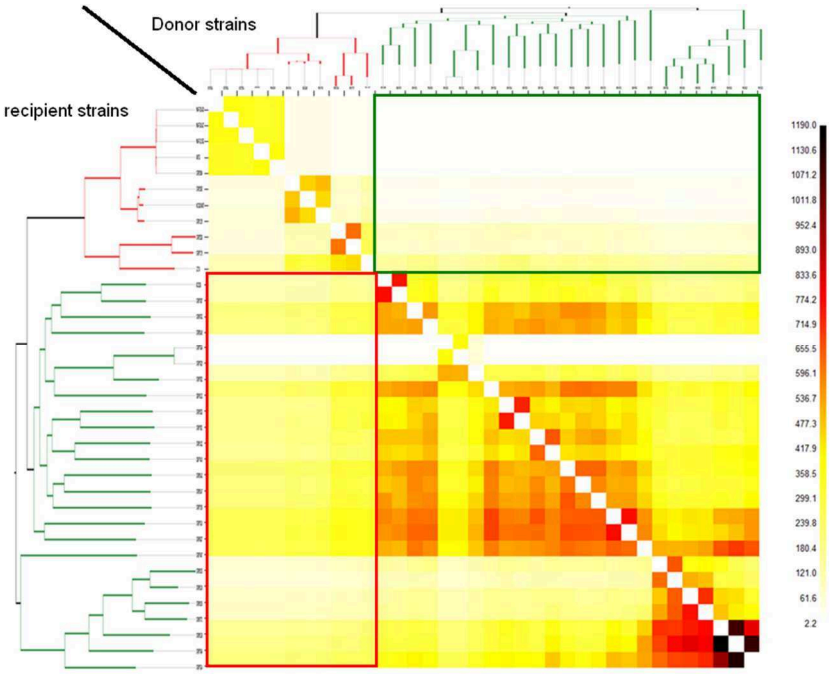


Figure III-5: Representation of gene flow between strains of groups A and B. The phylogeny was built in Bayesian on non-recombinant SNPs, posterior probabilities are represented for each node. The red branches represent group A strains and the green branches represent group B strains. The square represents the sum of shared genomic regions between the donor and the recipient strains. A scale of number of shared genomic regions is represented and correspond to the square color. The green rectangle represents gene flow when group B strains give to group A. The red rectangle gene flow when group A strains give to group B.

previously described (Darrasse et al. 2013). Briefly, the 50 bp extremities of detected IS corresponding to one motif of IRR were extracted from previous analysis and used as query in a blastn against *X. arboricola* genomes. Only results with identity above 95% were conserved. The total number of ISs was estimated as the number of complete ISs plus the number of pairs of IRRs.

The phage search was performed using the PHAST tool (Zhou et al. 2011). A based-score classification was obtained for each phage cluster detected. The score corresponded to the number of phage coding genes found in the cluster. A phage was considered intact if the score was superior to 90 (the maximal score can be 150), questionable for a score between 90 and 60 and incomplete for a score inferior to 60.

A Mann-Whitney test was performed to compare the number of each MGE in each of the two groups identified in *X. arboricola*. The null hypothesis was absence of differences between the two groups. The alternative hypothesis was "it exists a difference between the two groups". These tests were performed using R software, and the alternative hypothesis was selected if the p-value was inferior to 5%.

III. Results

Delineation and genetic structure of *X. arboricola* species

To restrict our dataset to strains strictly belonging to the same species, we performed a thorough analysis of *X. arboricola* species contours. Pairwise genome comparisons showed that the six strains belonging to group C as defined by multilocus sequence analysis (MLSA) in chapter 1 and the two strains from pathovar *guizotiae* did not fulfil the species delineation criterion (encadré III) and thus should be excluded from *X. arboricola* species. Indeed, these eight strains exhibited ANIb values below 95% when they were compared with CFBP 2528, the type strain of *X. arboricola* species (Table III-S2). These results were highlighted using a multidimensional scaling reflecting genetic distances between strains in a two dimensional graph, which separated these eight strains from the 36 other strains for each comparison axis (fig. III-2). A heterogeneity was observed among these strains as they clustered in three pairs and two singletons (fig. III-2). These clusters were highly distant from each other with ANI values that were below the 95% threshold. Thus these eight strains would represent five novel *Xanthomonas* species that remain to be formally described. In the remainder of this study, the dataset will only comprise the 36 strains, which complied with the ANI species delineation criterion and thus *X. arboricola* designates the species excluding these eight strains.

Comparative genomics and SNPs analyses identified two main groups in *X. arboricola*, one corresponding to the three successful pathovars, and the other one to

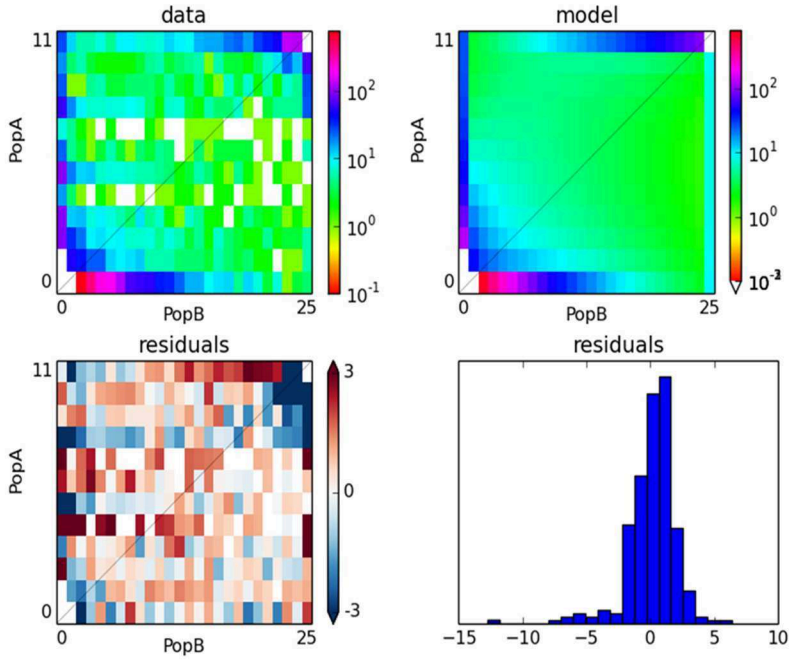


Figure III-6: Distribution of derived alleles in group A and B of the best model IM2M. The derived alleles were estimated using the group C strain CFBP 7645. The first HeatMap corresponds to the joint allele frequency spectrum obtained for group A in vertical axis and group B in horizontal axis. The second HeatMap corresponds to that inferred according the IM2M model. The two last graphs correspond to the residuals distribution corresponding to the difference in number of alleles in our dataset compared to that expected in the model (in blue, the number is higher in our dataset, in red, the number is higher in the model)

Table III-1: likelihood obtained for the seven tested models. SI : Strict isolation; IM: isolation with migration; AM: ancient migration; SC : secondary contact. 2M corresponds to models tested with two migration rates different, one along genome and one in genomic island. In red, are highlighted the two best model compared with LRT.

MODEL TESTED	LOG LIKELIHOOD
SI	-1459,85
IM	-1374,49
IM2M	-1144,46
AM	-1456,76
AM2M	-1459,97
SC	-1390,05
SC2M	-1362,81

the commensal strains and unsuccessful pathovars. The core genome of the 36 *X. arboricola* strains based on coding and noncoding regions was composed of 3,254,906 bp representing 393,610 SNPs. PCA based on all SNPs occurring in the core genome delineated two main groups along axis 1, which explained 42 % of variance (fig. III-3). These two groups corresponded to groups A and B previously described based on MLSA (chapter 1). The clonal phylogenetic tree based on 198,399 non recombinant SNPs (fig. III-4) evidenced three divergent branches, one corresponding to group A and two corresponding to group B, in which pathogenic and commensal strains were interspersed.

Detection of recombinant sequences and gene flows within *X. arboricola*

The coalescence analysis detected 20,992 recombinant sequences in the 2,020 coding sequences from the core genome of the 36 *X. arboricola* strains. We studied the gene flow within *X. arboricola* species using a chromosome painting approach (fig. III-5). The gene flow intensity was asymmetric from epidemic clones towards commensal strains. Moreover, we observed four groups of strains with high intensity of gene flow along the matrix diagonal. Two groups corresponded to the three successful pathovars in group A and the two subgroups in group B, had high to moderate intra-group exchanges and reduced ones with other groups. Overall, the intensity of genetic exchanges is higher in the two B subgroups than in the successful pathovars. Origin analysis of the shared polymorphism allowed us to highlight that gene flow was more frequent for group A strains being donors and group B strains being recipients (fig. III-5). The extreme example was pathovar pruni that received no gene flow from group B. A weak gene flow was observed towards pathovar *corylina* and it was a little bit more intense in pathovar *juglandis*. Similarly, two commensal strains, CFBP 7634 and CFBP 7629, received no DNA from other strains.

Divergence models based on site frequency spectrum (SFS) were analyzed between epidemic clones (group A) on the one hand and the background recombinogenic population (group B) on the other hand. The LRT between the two best models, IM and IM2M, was tested with a χ^2 of 10.6, a df of 2, and a probability of 0.05. The more likelihood model (Log-likelihood = -1144,46) was the IM2M model suggesting a gene flow during the divergence with a heterogeneity in genomic sequence, that is genomic regions more or less submitted to homologous recombination (fig. III-6, fig III-S1). Parameters were estimated from the best model. The migration intensity parameter (M) concerning the entire genome was higher in group A to group B (M = 9,42) than in group B to group A (M = 0,1). In contrast, the migration parameter in genomic island (Me) was higher in group B to group A (Me

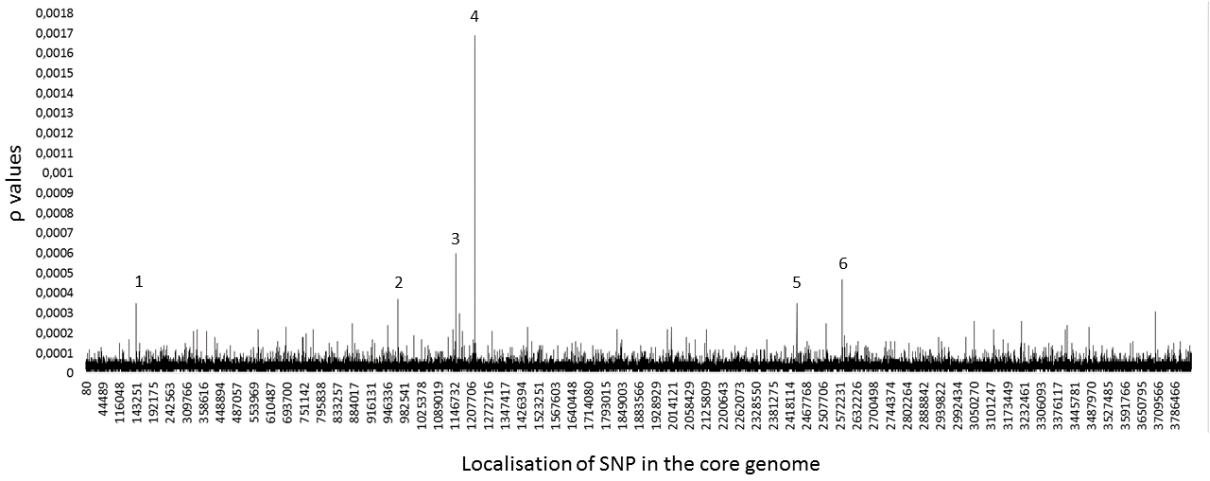


Figure III-7: ρ value for each SNP belonging to group A core genome. The vertical axis represents the ρ value, and the horizontal axis corresponds to localization in core genome for each SNP. The 6 outliers are indicating by a number.

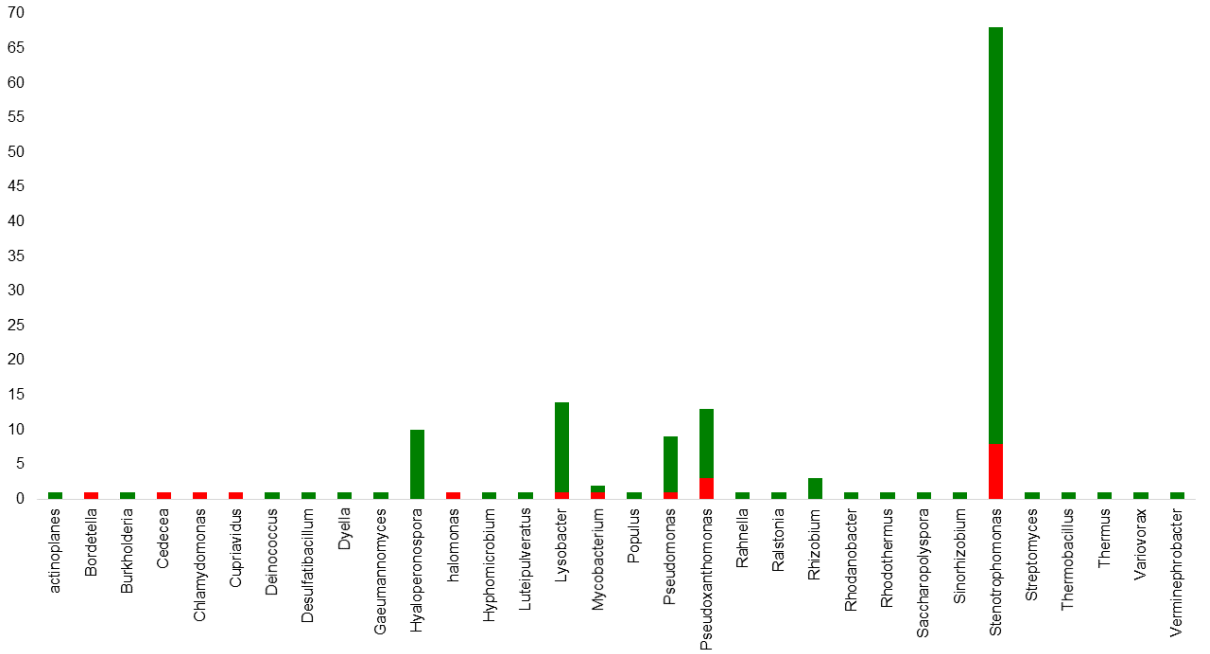


Figure III-8: Histogram representing the origins of recombinant sequences. The recombinant sequences in group A and B strains are represented in red and in green, respectively in green.

= 4,84) than in group A to group B (Me = 1,64*10-17). This suggested a more important gene flow from group A to group B that evenly affected the different parts of the genomes of group B members. In contrast, when group B were donors, the gene flow concerned essentially genomic islands and targeted hotspots in the genomes of epidemic clones (group A).

Analysis of recombination hotspots in the core genome of epidemic clones revealed six SNPs as outliers among the 149,552 SNPs of group A core genome (3,841,529 bp) (fig. III-7). The mean representing the recombination rate for each SNP was 3.10-5. The distribution of ρ values for each SNP along the genomic sequence, as done in Yahara et al. (2015) highlighted six SNPs with a ρ value at least ten times higher than the ρ mean, with SNP 4 having the highest ρ value of $1.7*10^{-3}$. Among these six SNPs, two were located in intergenic regions (SNPs 2 and 4), and four in gene coding regions (SNPs 1, 3, 5 and 6). These coding regions were annotated as *rim* (an ABC transporter), short-chain dehydrogenase, major facilitator superfamily and hypothetical protein, respectively. This suggested that these four genes exhibited a higher recombination rate than other genome regions in group A and thus represent recombination hotspots.

Origin of recombinant sequences from outside *Xanthomonas* genus

Among the 20,992 recombinant sequences 144 came from other genera. *Stenotrophomonas* genus was a major donor providing 68 recombinant sequences (fig. III-8). Four other genera were highly represented in this donor list, i.e. *Pseudoxanthomonas* and *Lysobacter*, which belonged to Xanthomonadaceae, *Pseudomonas*, and *Hyaloperonospora* an oomycete. The majority of these 144 recombinant sequences were identified in group B. No recombinant sequences having an external origin to *Xanthomonas* was found in pathovar *pruni*.

Comparison of MGE contents in epidemic clones and recombinogenic background

The three searched categories of MGEs (phages, plasmids and ISs) were more frequent in epidemic clones than in background population, suggesting that epidemic clones were more efficiently submitted to HGT than commensal strains (fig. III-9). Indeed, a Mann-Whitney test was performed to compare the MGE numbers in group A and in group B and was significant for the three comparisons (Table III-2). In epidemic clones, most strains (9/11) had plasmids (up to 6) whereas in group B, only four strains (out of 25) harbored plasmids (1 to 2) (fig. III-9). The only group

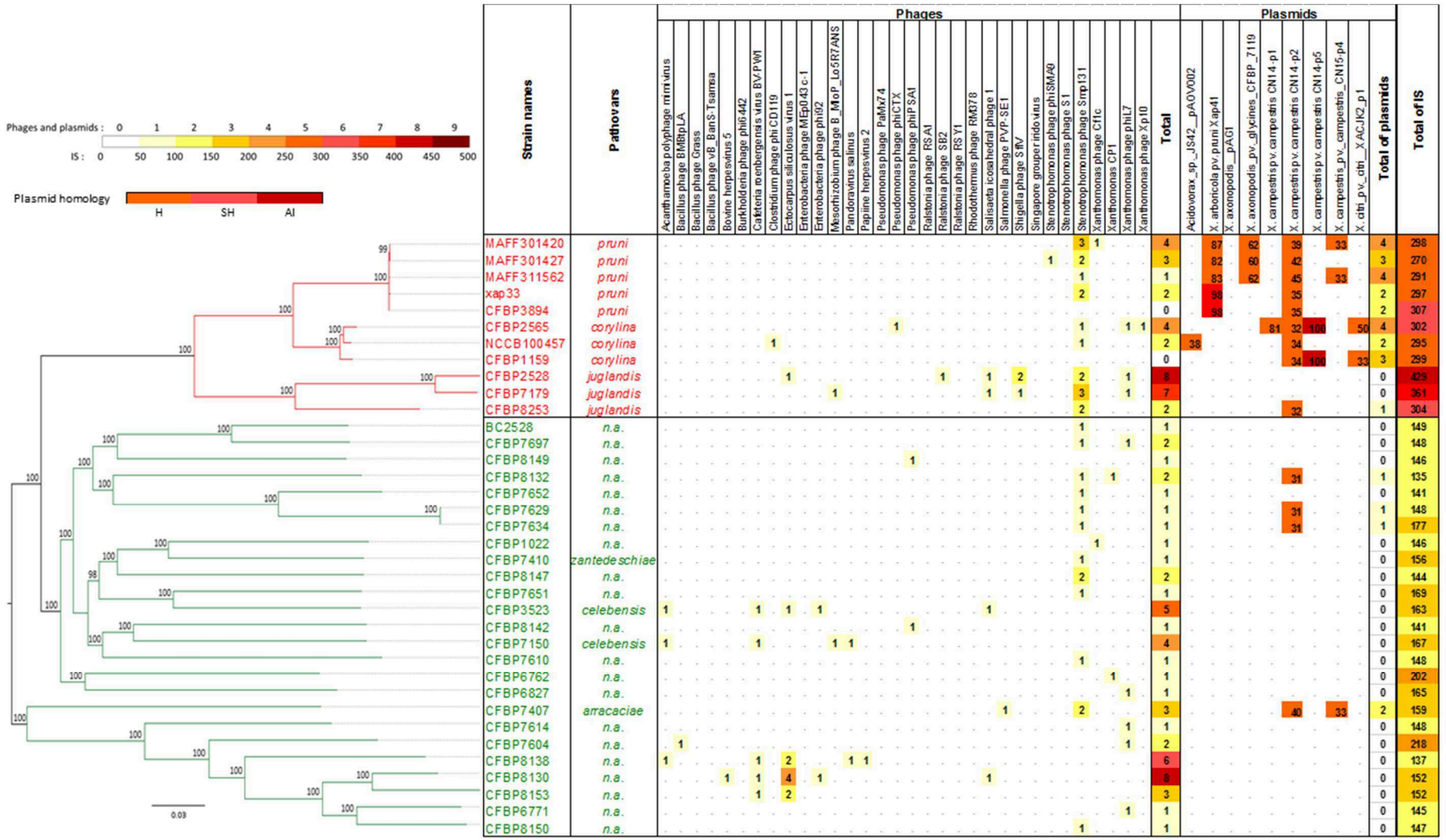


Figure III-9: Representation of MGE repertoires (phages, plasmids and IS) in the 36 strains belonging to *X. arboricola*. Group A strains are represented in red and group B strains are represented in green. The absence of MGE is represented by a grey dot and the number correspond to the copy number of MGE when it was present. Each square is colored according the copy number of MGE, according the scale in left bottom. The phylogenetic tree was built in maximum likelihood and using the non-recombinant SNP of 36 strains of *X. arboricola*. The bootstrap values were calculates to 100 replicates, and only the values superior to 85 are represented.

B strain that had two plasmids belonged to pathovar *arracaciae*. Concerning the ISs, twice more ISs were found in epidemic clones than in background population. In epidemic clones, IS numbers ranged from 270 to 429 with the highest numbers in pathovar *juglandis*. In group B, it ranged from 135 to 218. The difference between groups A and B was less pronounced for phage numbers than for other MGEs.

Concerning ISs, three families (ISL3, IS3_ssgr_IS407, IS5_ssgr_IS5 and IS5) represented 50% of the ISs found in *X. arboricola* genomes and one family was exclusively present in group A strains (IS3_ssgr_IS150).

Concerning the plasmid diversity, eight different plasmids were found in epidemic clones whereas only two different plasmids were observed in group B. Among the two plasmids present in group B, CN14 p2 from *X. campestris* pv. *campestris* seemed ubiquist as it was present in all *X. arboricola* strains having a plasmid sequence. Most plasmids identified in group A were previously described in other *Xanthomonas* species (*X. fuscans*, *X. campestris*, *X. oryzae*, *X. axonopodis* and *X. arboricola*) and one originated from *Acidovorax* genera. Plasmid Xap41 was strictly found in the five *X. arboricola* pv. *pruni* genomes which suggests their acquisition in pathovar *pruni* common ancestor.

Concerning the phages, 32 different phages were identified. Among them, *Stenotrophomonas* phage Smp131 was detected in most strains with a highest copy number of three. In contrast, the majority of phages was present in only few strains; 10 were specific to group B and 6 to group A. Despite these heterogeneous repertoires, no correlation was observed between phage repertoires, strain phylogenetic relationships or pathovar classification which suggests independent phage acquisitions.

IV. Discussion

Devastating bacterial crop diseases are mostly caused by host specialized epidemic clones. These clones represent highly frequent genotypes which often emerge from a single highly adaptive ancestor. In bacterial species with an epidemic population structure these clones are superimposed upon a network of highly diverse freely recombining strains. It was shown that HGT and homologous recombination are major players in new pathogen emergence (Supply et al. 2013). *X. arboricola* species presents an epidemic population structure, within which host-specialized successful pathovars represent epidemic clones and commensal strains and unsuccessful pathovars the recombinant network. We wonder if these commensal strains that are commonly found in agroecosystems present a risk for new disease emergences and which evolutionary force could drive this pathological adaptation. More precisely,

Table III-2: Mean number and variance of MGE (phages, plasmids and IS) found in 36 strains of *X. arboricola*. The *p-values* corresponds to the comparison of means between groups A and B with a Mann-Withney test.

		Group A	**Group B**
phages	Mean	3,09	2,08
	Confidence interval	0-8	0-8
	p-value	9.98e-02	
plasmids	Mean	3	0,2
	Confidence interval	0-6	0-2
	p-value	4,80e-04	
IS	Mean	313	156
	Confidence interval	270-429	135-218
	p-value	1,84e-06	

we wonder if epidemic clones could represent a reservoir of adapted genes and alleles for new emergences in the species or if inversely, host specialization has led to a barrier to gene flow with background population. We used a population genomics approach to address these questions and studied gene flow within the core genome and MGE contents.

Persistence of an asymmetric gene flow along divergence of epidemic clones

Shared polymorphisms were detected between epidemic clones and background population (fig. III-5). Because this polymorphism could result from diverse divergence scenarii, we compared different divergence models including migration between groups, based on SFS. This analysis allowed us to highlight that there are no barrier to gene flow between epidemic clones and background population as homologous recombination occurs during their divergence. This scenario could correspond to recurrent introgression observed in fungi (Stuckenbrock, 2016). In eukaryotic models, it was hypothesized that this introgression could speed up adaptative evolution, and lead to new emergence. Progressive introgression was also shown in evolutionary history of human pathogen *Campylobacter coli* that colonizes agricultural settings (Sheppard et al. 2013). It should be noted that SFS analysis was extensively used to study speciation in eukaryote models (Rougement et al. 2016) but is rarely used in bacterial models. In this latter case it was used to infer demographic history but not migration (Cornejo et al. 2012; Montano et al. 2015). For example, recent expansion of *Streptococcus mutans* was demonstrated using SFS (Cornejo et al. 2012). A previously population genetic analysis based on housekeeping gene fragments and using the same strains did not revealed gene flow between groups A and B (chapter 1). This highlights the limitation of methods based on few genes to infer complex evolutionary histories.

Asymmetry of gene flow from epidemic clones towards background population was observed both with chromosome painting and SFS approaches. This observation also concerns virulence associated-genes, as a detailed gene flow analysis of the type three secretion system gene cluster showed that homologous recombination was highly frequent within group B and that gene flow occurred more frequently from group A to group B (chapter 4). It is striking to note that similar results were obtained with *Chlamydia trachomatis*, an obligate intracellular human pathogen (Joseph et al. 2012). This species is composed of virulent and avirulent strains and was divided in four clades: clade 1 containing an invasive biovar, clades 2 and 4 containing strains more or less aggressive and clade 3 containing exclusively strains responsible for trachoma disease. As in *X. arboricola* the gene flow highlighted be-

tween these four clades was asymmetric with less aggressive strains exchanging a lot between them, and invasive strains giving to other clades but not receiving much. Similar results were obtained with pathogenic and commensal strains of *Escherichia coli* (Didelot et al. 2012). In this study, two pathogenic strains belonging to phylogroup B1 did not receive recombinant fragments from phylogroup A composed exclusively of commensal strains, whereas they are donors to phylogroup A. Difference of lifestyle was put forward to explain this gene flow pattern. Indeed, this gene flow asymmetry could be explained by difference of selection pressure within these two groups. Recombination events are dependent on DNA transfer between two bacteria cells, but also of selection pressure on genome (Yahara et al. 2016; Turner et al. 2016). Differences in selection pressures could also account for the different patterns of genome evolution observed in *X. arboricola* between epidemic clones and background population. In the former homologous recombination mostly occurs at hotspots whereas in the latter recombinant events are evenly distributed in the genomes.

Presence of gene flow between epidemic clones and background population suggests that pathogenic and commensal strains were in sympatry. Indeed, commensal strains have already been isolated from the same host or organs as pathogenic strains, but no data are available concerning niche sharing, i.e. simultaneous presence of commensal and pathogenic strains in the same niche. Globally, the ecology of commensals is still poorly understood. This hypothesis is supported by the fact that pathogenic and commensal strains seem to differ mostly by their capacity or incapacity to induce symptoms in the host plant as phenotypic tests showed no difference in the first step of host colonization (data not shown[12]).

HGT drives the evolution of epidemic clones

MGE content comparison showed that epidemic clones (group A) had more MGEs (in particular ISs and plasmids) than background population (group B), suggesting that group A strains were more submitted to HGT than group B ones. This emphasizes a previous result based on genomic comparison of two pathogenic strains with two commensal strains belonging to *X. arboricola* (Cesbron et al. 2015) showing much more ISs in pathogenic strains. Indeed, IS expansion is commonly observed in bacteria with a recent host-restricted lifestyle and may be seen as an early step before genome reduction that is observed in obligate intracellular bacteria (Siguier et al. 2014). According to this model, IS expansion and absence of genome reduction in the three host-specialized successful pathovars of *X. arboricola* are in favor of a recent emergence. High number of MGEs was also observed in *Erwinia*

[12]Ces résultats sont présentés dans le chapitre 3 de ce manuscrit de thèse

tracheiphila an obligate vector-transmitted plant pathogen, and the authors proposed that this species was in the first step of host specialization (Shapiro et al. 2016). The originality of our model species is to provide direct comparison of MGEs content between different host-specialized pathogens and commensal strains within the same species.

The plasmid pXap41, initially described in *X. arboricola* pv. *pruni* (Pothier et al. 2011) was exlusively found in pathovar *pruni* strains. It encompasses three virulence genes (*xopE2*, *xopE3* and *mltB*). Because it was found in all pv. *pruni* strains, this plasmid was probably acquired in their common ancestor. This acquisition could have led to the ability to colonize a new host belonging to *Prunus* genus as described by Engering et al. (2013), and could have driven the divergence between pv. *corylina* and pv. *pruni*. So, we can hypothesize that the plasmid transfer allowed the host shift of pv. *pruni* strains.

In conclusion, using genomics data and species delineation criterion *X. arboricola* species was restricted to the two genetic groups, A and B. The group B corresponds to the recombinant network and the group A corresponds to three epidemic clones. This study based on SFS analyses showed that DNA exchanges between groups A and B occurred during their divergence, with a gene flow more frequent from group A to group B than group B to group A. The study of MGEs involved in HGT (plasmids, ISs and prophages) showed that group A strains are more submitted to HGT than group B strains. This suggests that in group A, core genome of epidemic clones is strongly conserved with hotspots of recombination whereas MGEs belonging to accessory genome are highly variable, and that in group B, core genome is evenly prone to homologous recombination whereas MGEs are less.

Datation phylogénétique de l'émergence des trois clones épidémiques chez *X. arboricola*

Ce travail a été réalisé en collaboration avec Adrien Rieux (UMR PVBMT, CIRAD-La Réunion)

I. Introduction

Les résultats des études de génétique et génomique des populations indiquent que les trois pathovars épidémiques *pruni*, *corylina* et *juglandis* auraient une origine commune. En effet, ils appartiennent tous les trois au même groupe génétique identifié sur la base de 7 gènes de ménage (chapitre 1) ou des séquences génomiques (chapitre 2). Le fait qu'ils partagent aussi un répertoire d'effecteurs de type 3 très similaire renforce cette hypothèse d'une histoire évolutive commune. Les observations biologiques et historiques sur ces trois pathovars corroborent aussi cette hypothèse. En effet, ces agents pathogènes sont responsables des mêmes symptômes sur leurs hôtes respectifs et partagent de nombreux traits biologiques. De plus, les rapports de 1ère description de ces agents pathogènes ont été réalisés aux Etats-Unis en 1903 par Smith pour le pathovar *pruni*, en 1913 par Barss pour le pathovar *corylina* et 1901 par Pierce pour le pathovar *juglandis*. Ceci suggère qu'ils auraient émergé en Amérique du Nord au début du $20^{ème}$ siècle. Cependant, ces rapports sont également le signe manifeste de l'essor d'une nouvelle discipline, la phytobactériologie née avec la première description du rôle d'une bactérie dans une maladie d'une plante (Burril, 1878, cité par Lepoivre, 2007). Ils sont également concomitants de la 1ère description de très nombreuses autres maladies des plantes, lesquelles préexistaient mais n'avaient pas été formellement décrites.

La datation phylogénétique permet d'estimer l'âge des nœuds internes (les ancêtres communs) au sein d'un arbre phylogénétique. Dans le cas d'agents pathogènes, l'âge d'un nœud ancestral peut nous renseigner sur la période à laquelle leur

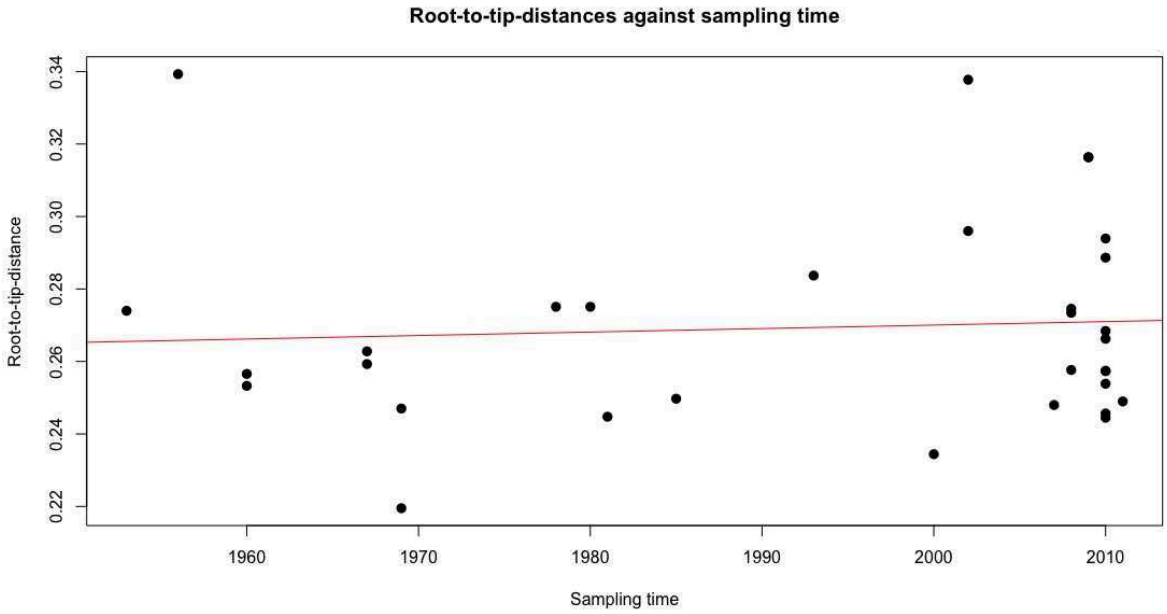


Figure III-10 : Régression linéaire (en rouge) des distances de chaque échantillon à la racine en fonction de la date d'isolement, pour le jeu de données complet.

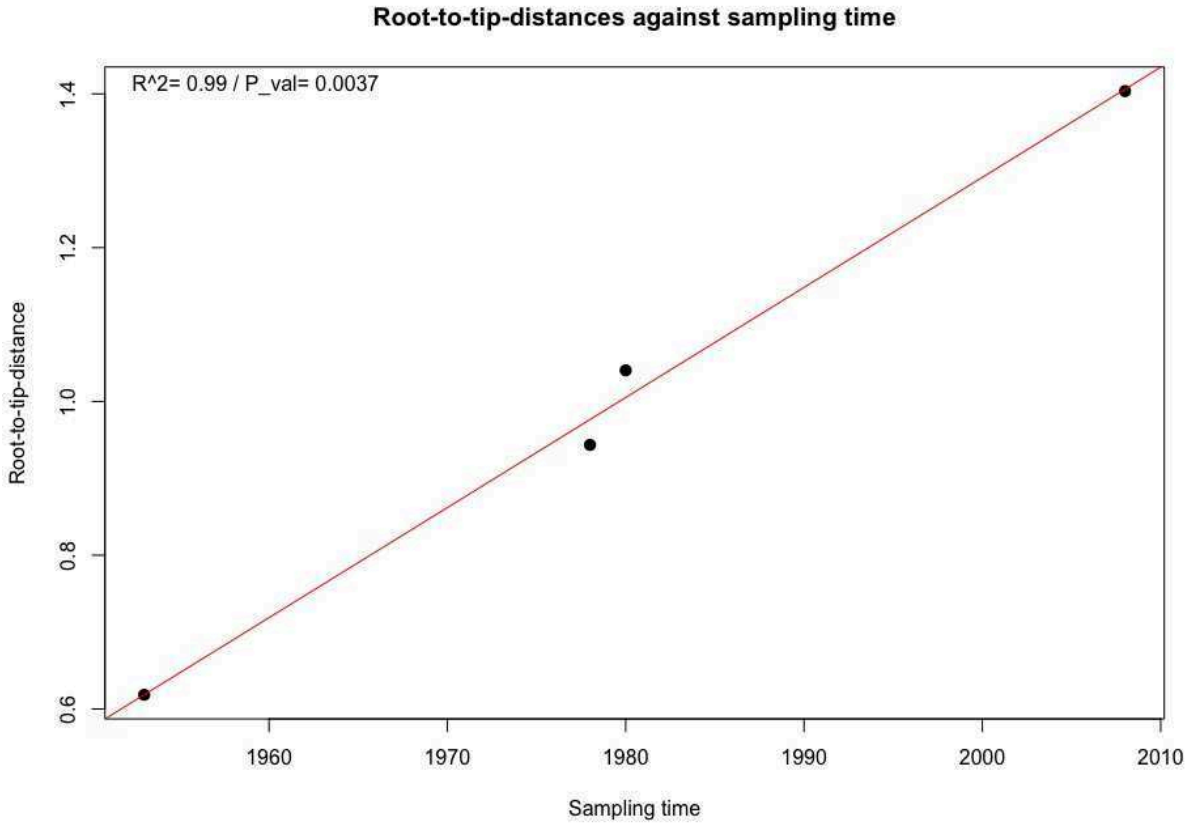


Figure III-11 : Régression linéaire (en rouge) des distances de chaque échantillon à la racine, en fonction de la date d'isolement pour les 4 souches du pathovar *pruni*.

émergence a pu survenir (Rieux & Balloux 2016). Afin de dater une phylogénie, il faut estimer la vitesse à laquelle les mutations se fixent dans les génomes (le taux de mutation). Pour cela, les méthodes phylogénétiques requièrent de convertir la divergence moléculaire observée en temps absolu (cette procédure est appelée calibration de l'arbre phylogénétique). Cela peut tout d'abord se faire en associant aux nœuds internes de l'arbre (les ancêtres communs associés à nos échantillons) une date provenant de l'information historique, par exemple de type fossile associé à l'organisme en question (Sauquet, 2013), ou de son hôte, dans le cas d'une coévolution entre individus (Santos-Garcia et al. 2015). La seconde approche, appelée " tip-dating ", consiste à directement utiliser pour la calibration de l'arbre les dates auxquelles ont été récoltés les échantillons (Rieux & Balloux, 2016). Chez les bactéries, il est difficile d'obtenir des données fossiles, donc la méthode la plus utilisée pour calibrer un arbre est celle du " tip-dating ". Cette méthode a, par exemple, été utilisée pour dater l'acquisition de la résistance aux antibiotiques chez *Mycobacterium tuberculosis* (Eldhom et al. 2015).

Chez *X. arboricola*, la collection de séquences génomiques étudiées comprend des souches qui ont été isolées à des dates différentes, la plus ancienne étant la souche appartenant au pathovar *corylina* isolée en 1939 et la plus récente, la souche CFBP 7697 isolée en 2011. Pour son application, la méthode dite du " tip-dating " nécessite de bénéficier d'échantillons isolés durant une période de temps suffisamment grande pour pouvoir y observer l'accumulation progressive et temporelle des substitutions. Nous avons donc, dans un premier temps, recherché le signal temporel existant dans cette collection pour envisager de réaliser une datation moléculaire, et dans un second temps nous avons utilisé la méthode du "tip-dating" pour dater l'émergence des trois clones épidémiques.

II. Matériel et Méthodes

La collection est composée de 32 séquences génomiques (10 sont disponibles publiquement et 22 ont été séquencées dans le cadre de ce travail) (Table III-S1). Il s'agit de la collection utilisée dans le chapitre précédent en ayant toutefois exclu les souches pour lesquelles la date d'isolement n'est pas disponible. Elle comprend 10 génomes appartenant aux clones épidémiques et 22 génomes appartenant au réseau recombinant.

Afin de réaliser la datation phylogénétique, une phylogénie clonale de l'ensemble de ces 32 souches a été reconstruite. Pour cela, le core génome comportant les régions codantes et non codantes a été construit en utilisant le logiciel Mauve (Darling et al. 2004). Les séquences recombinantes ont été détectées en utilisant ClonalFrame

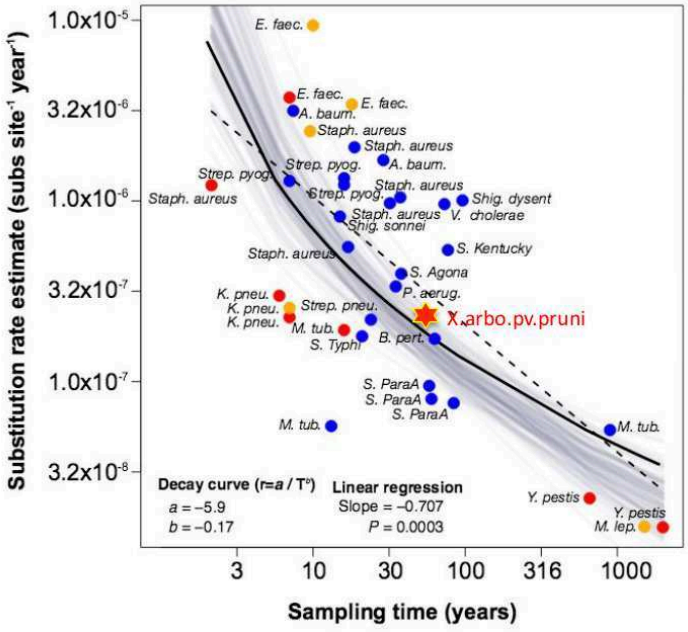


Fig. III-12 : Régression linéaire (en pointillé) et la courbe de décroissance (en trait plein) des taux de mutations estimés par site et par années obtenus pour les agents pathogènes humains en fonction de la date des échantillons en années. Les lignes grises représentent le degré d'incertitude de la courbe de décroissance. Le taux de mutation de *X. arboricola* pv. *pruni* est représenté par une étoile rouge (D'après Duchêne et al. in press ; Rieux, comm. pers).

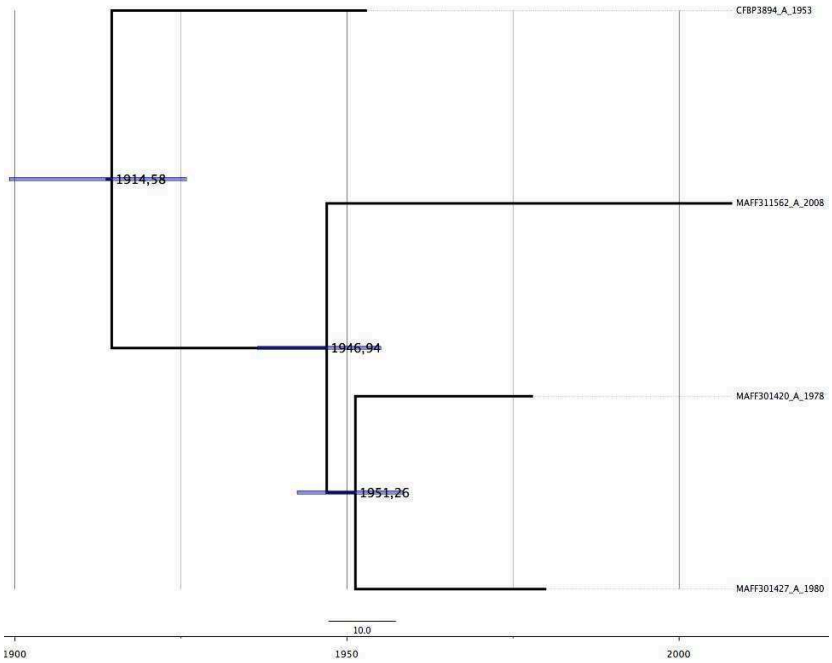


Figure III-13 : Phylogénie basée sur les SNP non recombinants des 4 souches du pathovar *pruni* et reconstruite en utilisant la méthode bayésienne. Chaque nœud est daté en années et l'intervalle de confiance est indiqué par les barres bleues.

(Didelot et al. 2007) et extraites du core génome en utilisant un script R.

La recherche de signal temporel dans le jeu de données a nécessité de construire une phylogénie en maximum de vraisemblance sans contraindre les feuilles selon la date d'isolement en utilisant phyML (Guindon et al. 2010) et les 198 397 SNP non recombinants issus de l'alignement du core génome purgé de la recombinaison. Pour chaque souche, la distance génétique des feuilles jusqu'à la racine (correspondant au nombre de substitutions s'étant accumulés dans les génomes depuis l'ancêtre commun) a été calculée. Une régression linéaire entre ces distances et la date d'isolement des souches a ensuite été réalisée. Si la régression est positive, c'est-à-dire que les souches les plus anciennes ont accumulés moins de mutations que les souches plus récentes, on considère qu'un signal temporel existe au sein du jeu de données.

Dans un second temps, la datation moléculaire a été réalisée en utilisant BEAST (Drummond et al. 2007) avec un modèle démographique Bayesian skyline (Drummond et al. 2005). Cette analyse de type Bayésienne permet de conjointement estimer le taux de mutation et l'âge de chacun des nœuds de la phylogénie.

Lorsque l'on cherche à extrapoler le taux de mutation estimé sur un ensemble de souches à un panel d'isolats élargi il est nécessaire de vérifier que les taux de substitutions entre les différents clades ne sont pas trop différents.

III. Résultats et Discussion

Identification d'un signal temporel dans le jeu de données

La régression linéaire des feuilles à la racine (fig. III-10) montre qu'il n'y a pas de signal temporel dans notre jeu de données complet ($R^2 = 0.11$, p-value > 0.5), et donc que la méthode du tip-dating n'est pas applicable à l'ensemble de la collection. Cependant, cette régression linéaire a également été réalisée au niveau de chaque nœud, afin de détecter un éventuel signal temporel dans un sous-groupe de la phylogénie. Un seul nœud a montré un signal temporel ($R^2 = 0.99$; p-value $= 0.0037$), il s'agit de celui correspondant à l'ancêtre commun des 4 souches du pathovar *pruni* (fig. III-11)

Application du "tip-dating" sur les quatre souches du pathovar *pruni*

Le taux de mutation estimé pour ce pathovar est de 3.15×10^{-7} mutations par site et par année, avec un intervalle de confiance situé entre 2.4×10^{-7} et 3.9×10^{-7} mutations par site et par année. Ce taux de mutation correspond à l'attendu en fonction du pas de temps entre la souche la plus récente et la plus ancienne (Duchêne et al. in press). Il est proche de celui de *Pseudomonas aeruginosa*, bactérie associée

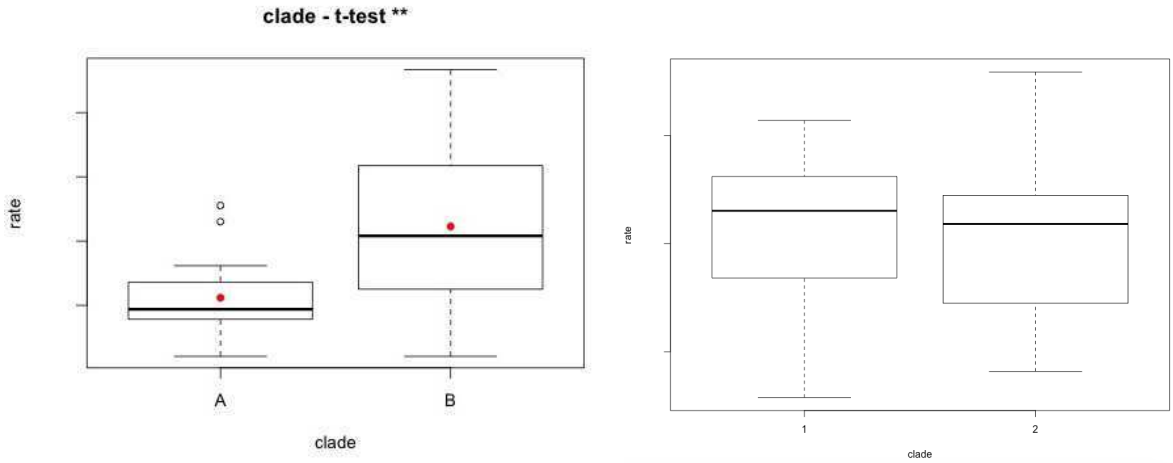


Figure III-14 : Comparaison des taux de substitutions entre les groupes A et B (à gauche), (groupes génétiques identifiés par Merda et al. 2016 et dans le chapitre 2) et entre le pathovar *pruni* (1) et les deux pathovars *juglandis* et *corylina* (2) (à droite).

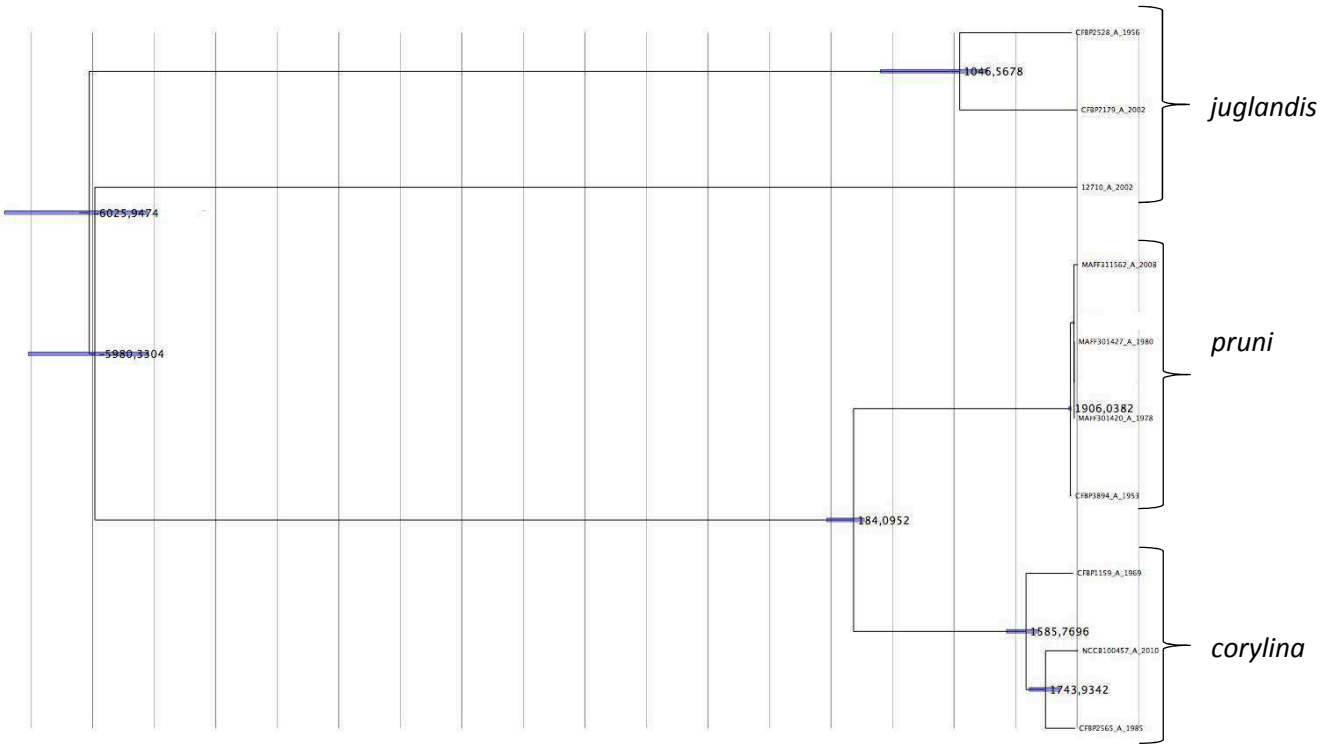


Figure III-15 : Phylogénie basée sur les SNP non recombinants des 10 souches appartenant aux trois pathovars *pruni, juglandis* et *corylina* et reconstruite en utilisant la méthode bayésienne. Chaque nœud est daté en années et l'intervalle de confiance est indiqué par les barres bleues.

à la fois aux plantes et aux animaux (fig. III-12). Cette estimation a permis de dater les trois nœuds de la phylogénie de ces quatre souches (fig. III-13). La divergence par rapport à l'ancêtre commun à ces 4 souches daterait de 1914 [1898 – 1927]. Cette date est en accord avec la première identification de la maladie en Amérique du Nord, puisque celle-ci a été découverte en 1903 par Smith. Ensuite, la souche MAFF 311562 aurait divergé en 1946 et les deux souches MAFF 301420 et 301427 auraient divergé en 1951.

Afin de savoir si ce taux de mutation peut être utilisé pour calibrer l'arbre entier, les taux de substitution au sein des groupes A et B ont été comparés en utilisant une horloge moléculaire relâchée (fig. III-14a). Les taux étant significativement plus élevés dans le groupe B, le taux de mutation obtenu à partir des 4 souches du pathovar *pruni* ne peut pas être appliqué à l'ensemble de la phylogénie. La même approche a été réalisée pour comparer les taux de substitution entre le pathovar *pruni* et les pathovars *corylina* et *juglandis* (fig. III-14b). Dans ce cas, il n'y a pas de différence significative entre les deux groupes, le taux de mutation obtenu chez le pathovar *pruni* peut donc être utilisé pour calibrer une phylogénie composée uniquement des trois pathovars *pruni*, *juglandis* et *corylina*.

Sur cette phylogénie datée (fig. III-15), on voit que la date correspondant à l'ancêtre commun des 4 souches du pathovar pruni est la même que celle obtenue précédemment, et donc est en accord avec les données historiques concernant la première description de cet agent pathogène. Cependant, pour les deux autres pathovars, les différentes dates obtenues sont bien antérieures aux dates de leur première description. En effet, concernant le pathovar *corylina*, les trois souches auraient divergé de leur ancêtre commun aux alentours de 1585. Cette date pourrait coïncider avec l'introduction du noisetier en Amérique du Nord, puisque ce dernier a été importé depuis l'Europe vers le continent américain à la fin du XV^{ème} siècle. Le pathovar *juglandis*, quant à lui, serait encore plus ancien puisque les deux souches CFBP 2528 et CFBP 7179 auraient divergé de leur ancêtre commun vers 1046. Ces trois pathovars auraient divergé de leur ancêtre commun il y a environ 6 000 ans, soit 4 000 ans avant J. C. Cette période correspond au néolithique et est marquée par une intensification de la domestication des plantes. Enfin, il est également intéressant de noter que la domestication du noyer aurait débuté il y a environ 12 000 ans, et cette date correspond à la date estimée pour la divergence entre le groupe A et le groupe B, lorsque la datation a été réalisé sur l'ensemble de la collection (résultats non présentés car les groupes A et B évoluant de manière différente, cette analyse n'est pas robuste). En conclusion, ces différentes émergences pourraient donc avoir eu lieu suite à des actions anthropiques, telle que la domestication et l'introduction d'organismes dans une nouvelle aire.

IV. Perspectives

Cette étude correspond à la première étude dans laquelle un taux de mutation a été estimé de manière précise chez une bactérie pathogène de plante. Afin d'affiner ce taux de mutation calculé à partir des 4 génomes du pathovar *pruni*, il est possible de réaliser les analyses en partitionnant les SNP en sites synonymes et non synonymes. En effet, les sites synonymes sont supposés être moins soumis à la sélection purifiante que les sites non synonymes, puisqu'ils ne modifient pas la séquence protéique. En conséquence, le taux de mutation est supposé plus élevé pour les sites synonymes que pour les sites non synonymes. De plus, afin d'affiner la datation, il faudrait avoir un nombre de souches plus important par pathovars, puisque nous avons en moyenne 3 génomes par pathovars, ce qui n'est probablement pas représentatif de la diversité. Initialement, nous n'avions pas comme objectif de réaliser cette analyse lorsque nous avons fait le choix de séquencer les génomes, et donc les dates d'isolement ne faisaient pas partie de nos critères de sélection des souches. Il pourrait donc être intéressant de séquencer des souches plus anciennes et plus récentes, disponibles dans les collections de souches afin d'obtenir un meilleur signal temporel et de (in)valider ces résultats et ce, pour les trois clones épidémiques. Enfin, cette méthode pourrait être utilisée pour dater les événements d'acquisition des ET3 spécifiques à ces trois pathovars, comme cela a été réalisé pour l'acquisition d'éléments intégratifs conjugatifs (ICE) chez *Vibrio cholerae* (Spagnoletti et al. 2014).

Supplemental materials

Table III-S1 : Collection used for these study. * indicates the 36 strains used for all analysis after the ANI analysis. In “PATHOGENIC” column, o indicates pathogenic strains and n indicates commensal strains. The “GENETIC GROUP” indicates the group which belong each strains (chapter 1). $ indicates the 32 strains used for the “tip-dating”

NAME STRAINS	PATHOVAR	PATHOGENIC	GENETIC GROUP	HOST OF ISOLATION	YEAR OF ISOLATION	CONTIG	GENOME SIZE (MB)	CDS NUMBER
CFBP1159*$	corylina	o	A	Corylus maxima	1969	124	5,11	4396
CFBP2565*$	corylina	o	A	Corylus avellana	1985	101	5,06	4433
NCCB100457*$	corylina	o	A	Corylus colurna	2010	281	5,23	4455
CFBP8253*$	juglandis	o	A	Juglans regia	2002	92	5	4265
CFBP2528*$	juglandis	o	A	Juglans regia	1956	8	5,09	4325
CFBP7179*$	juglandis	o	A	Juglans regia	2002	14	5,16	4399
CFBP3894*$	pruni	o	A	Prunus salicina	1953	77	5,06	4337
MAFF301420*$	pruni	o	A	NA	1978	396	5	4511
MAFF301427*$	pruni	o	A	NA	1980	356	4,91	4373
MAFF311562*$	pruni	o	A	NA	2008	432	5,09	4681
XAP33*	pruni	o	A	Prunus amygdalus cv. Guara	NA	474	5,1	4565
CFBP7410*$	zantedeschiae	o	A	Zantedeschiae aethiopica	1967	37	4,98	4201
CFBP7407*$	arracaciae	o	B	Arracacia xanthorrhiza	1969	232	5,16	4462
CFBP3523*$	celebensis	o	B	Musa acuminata	1960	3	4,91	4055
CFBP7150*$	celebensis	o	B	Musa acuminata	1960	7	4,98	4098
CFBP7408	guizotiae	o	B	Guizotia abyssinica	<1966	162	4,94	4299
CFBP7409	guizotiae	o	B	Guizotia abyssinica	<1966	220	4,16	4199
CFBP7604*	NA	n	B	capsicum cv. Japaleno	NA	6	4,85	4050
CFBP6762*	NA	n	B	fragaria X ananassa	NA	103	4,91	4215

CFBP6771*$	NA	n	B	fragaria X ananassa	1993	56	4,91	4124
CFBP7697*$	NA	n	B	Phaseolus vulgaris	2011	8	4,95	4141
CFBP8130*$	NA	n	B	Phaseolus vulgaris	2010	78	4,77	3985
CFBP8132*$	NA	n	B	Phaseolus vulgaris	2010	44	4,93	4179
CFBP8138*$	NA	n	B	Phaseolus vulgaris	2010	57	4,76	3959
CFBP8142*$	NA	n	B	Phaseolus vulgaris	2010	55	4,85	4097
CFBP8147*$	NA	n	B	Phaseolus vulgaris	2010	105	4,89	4183
CFBP8149*$	NA	n	B	Phaseolus vulgaris	2010	33	4,88	4117
CFBP8150*$	NA	n	B	Phaseolus vulgaris	2010	351	4,92	4496
CFBP8153*$	NA	n	B	Phaseolus vulgaris	2010	78	4,84	4045
CFBP7614*	NA	n	B	Magnolia sp.	NA	46	4,92	4202
CFBP1022*$	NA	n	B	Juglans regia	1967	43	5,1	4268
CFBP7629*$	NA	n	B	Juglans regia	2009	62	4,95	4238
CFBP7634*$	NA	n	B	Juglans regia	2009	4	4,94	4141
CFBP7651*$	NA	n	B	Juglans regia	2008	6	5,03	4229
CFBP7652*$	NA	n	B	Juglans regia	2008	78	5,23	4439
BC2528*$	NA	o	B	piment	2007	132	4,9	4167
CFBP6827*$	NA	o	B	piment	2000	5	4,95	4107
CFBP7610*$	NA	n	B	pueraria lobata	1981	46	4,96	4211
CFBP7622	NA	n	C	Phaseolus vulgaris	1985	106	4,75	4187
CFBP8152	NA	n	C	Phaseolus vulgaris	2010	199	4,73	4188
CFBP7645	NA	n	C	Juglans regia	2009	14	4,76	4087
CFBP7653	NA	n	C	Juglans regia	2008	77	4,87	4152
CFBP3122	populi	o	C	Salix alba	1980	92	4,79	4125
CFBP3123	populi	o	C	Populus X canadensis	1979	28	4,65	4050

Table III-S2 : ANI values for each pairwise genome comparisons of 44 genomes composing the initial collection. For the three tables, ANI values inferior to 95, corresponding to bacterial species limit are highlighted in lightgreen. a : corresponds to group B strains compared to all 44 strains, b: corresponds to group C strains compared to all 44 strains, and c: corresponds to group A strains compared to all 44 strains

a :

	CFBP8149	CFBP8147	CFBP7407	CFBP7651	CFBP8138	CFBP8132	CFBP8142	CFBP8153	CFBP7697	CFBP8150	CFBP6762	CFBP7408	CFBP6827	CFBP7652	CFBP3523	CFBP7410	CFBP6771	CFBP8130	CFBP7614	BC2528	CFBP7629	CFBP7634	CFBP7409	CFBP1022	CBFP7150	CFBP7610	CFBP7604
CFBP8149	---																									
CFBP8147	97,26	---																								
CFBP7407	96,82	96,75	---																							
CFBP7651	97,28	97,16	96,59	---																						
CFBP8138	96,74	96,7	96,72	96,66	---																					
CFBP8132	97,22	97,2	96,72	97,09	96,7	---																				
CFBP8142	97,24	97,23	96,79	97,07	96,77	97,25	---																			
CFBP8153	96,96	96,8	96,83	96,82	97,69	96,83	96,84	---																		
CFBP7697	97,36	97,25	96,77	97,15	96,75	97,21	97,21	96,84	---																	
CFBP8150	96,64	96,55	96,54	96,5	97,61	96,57	96,64	97,99	96,58	---																
CFBP6762	97,21	97,14	96,79	97,07	96,65	97,14	97,12	96,8	97,24	96,55	---															
CFBP7408	95,01	94,94	94,69	94,95	95,02	94,97	94,97	95,12	94,97	94,96	94,9	---														
CFBP6827	97,13	97,15	96,83	97,05	96,79	97,1	97,12	96,89	97,08	96,66	97,09	95,02	---													
CFBP7652	97,07	96,99	96,46	96,91	96,56	96,99	97,01	96,69	97,01	96,42	97,01	94,87	96,96	---												
CFBP3523	97,28	97,2	96,77	97,62	96,74	97,15	97,13	96,91	97,16	96,64	97,12	95,01	97,06	97,02	---											
CFBP7410	97,26	97,14	96,66	97,03	96,79	97,13	97,16	96,91	97,25	96,65	97,15	94,92	97,07	96,91	97,07	---										
CFBP6771	96,8	96,75	96,75	96,7	97,72	96,76	96,81	98,13	96,76	98,29	96,68	95,05	96,81	96,64	96,8	96,84	---									
CFBP8130	96,8	96,76	96,76	96,71	97,71	96,76	96,79	98,59	96,76	98,22	96,74	95,11	96,81	96,61	96,78	96,84	98,32	---								
CFBP7614	96,83	96,78	96,72	96,75	97,46	96,8	96,8	97,33	96,86	97,23	96,77	95,07	96,9	96,69	96,84	96,91	97,35	97,41	---							
BC2528	97,34	97,22	96,73	97,09	96,7	97,19	97,17	96,81	98,38	96,56	97,29	94,97	97,09	97,02	97,16	97,26	96,78	96,75	96,83	---						
CFBP7629	96,85	96,74	96,32	96,73	96,38	96,79	96,8	96,47	96,78	96,24	96,83	94,75	96,78	98,27	96,79	96,76	96,41	96,41	96,47	96,75	---					
CFBP7634	96,89	96,81	96,38	96,78	96,43	96,85	96,84	96,5	96,82	96,28	96,86	94,79	96,82	98,32	96,83	96,8	96,46	96,46	96,5	96,8	99,83	---				
CFBP7409	95,02	94,94	94,7	94,96	95,03	94,98	94,97	95,12	94,98	94,96	94,9	99,9	95	94,88	95,01	94,93	95,06	95,1	95,07	94,96	94,75	94,8	---			
CFBP1022	97,33	97,21	96,81	97,15	96,69	97,22	97,17	96,8	97,33	96,61	97,19	94,99	97,09	97,05	97,18	97,55	96,75	96,74	96,8	97,33	96,84	96,89	95	---		
CBFP7150	97,31	97,28	96,85	97,19	96,78	97,27	97,28	96,93	97,28	96,69	97,22	95,02	97,14	97,09	97,22	97,24	96,86	96,84	96,85	97,31	96,87	96,89	95,02	97,3	---	
CFBP7610	97,1	97,04	96,68	97,01	96,83	97,07	97,08	96,86	97,07	96,59	97,01	94,95	97,1	96,96	97,06	97,04	96,76	96,8	96,86	97,06	96,69	96,73	94,96	97,11	97,15	---
CFBP7604	96,78	96,69	96,73	96,73	97,77	96,71	96,73	97,58	96,72	97,54	96,65	95,08	96,8	96,6	96,74	96,84	97,64	97,67	97,58	96,77	96,39	96,42	95,08	96,74	96,78	96,83
CFBP7653	93,28	93,21	93,18	93,25	93,26	93,24	93,25	93,31	93,22	93,13	93,25	92,78	93,23	93,35	93,26	93,24	93,26	93,24	93,24	93,22	93,36	93,38	92,78	93,23	93,3	93,24
CFBP8152	94,27	94,24	94,11	94,27	94,21	94,23	94,21	94,23	94,21	94,1	94,19	93,49	94,22	94,35	94,21	94,25	94,19	94,19	94,2	94,22	94,64	94,66	93,49	94,26	94,34	94,24
CFBP7645	94,57	94,51	94,32	94,59	94,53	94,51	94,55	94,56	94,53	94,38	94,55	93,72	94,56	94,76	94,53	94,53	94,5	94,51	94,52	94,52	95,12	95,16	93,73	94,57	94,62	94,57
CFBP7622	93,18	93,12	93,2	93,15	93,15	93,13	93,15	93,22	93,12	93,07	93,14	92,66	93,15	93,23	93,12	93,13	93,15	93,15	93,13	93,15	93,18	93,21	92,67	93,13	93,21	93,13
CFBP3123	93,91	93,88	93,78	93,87	93,84	93,86	93,87	93,89	93,87	93,73	93,88	93,25	93,9	93,86	93,85	93,88	93,85	93,85	93,84	93,89	93,83	93,87	93,25	93,93	93,94	93,85
CFBP3122	94,83	94,73	94,56	94,76	94,7	94,77	94,78	94,73	94,76	94,56	94,79	93,79	94,79	94,7	94,75	94,73	94,69	94,68	94,72	94,76	94,6	94,65	93,79	94,82	94,82	94,76
ncppb1447	96,74	96,68	96,29	96,75	96,37	96,66	96,67	96,45	96,72	96,22	96,68	94,68	96,66	96,65	96,77	96,61	96,4	96,39	96,45	96,7	96,45	96,5	94,7	96,69	96,72	96,6
12710	96,65	96,53	96,22	96,64	96,38	96,53	96,59	96,43	96,58	96,2	96,57	94,7	96,58	96,54	n,a	96,47	96,37	96,35	96,4	96,56	96,37	96,41	94,72	96,56	96,61	96,49
maff301427	96,94	96,84	96,43	96,83	96,62	96,85	96,9	96,74	96,88	96,46	96,88	94,94	96,89	96,72	96,87	96,81	96,66	96,69	96,73	96,84	96,51	96,55	94,96	96,91	96,96	96,82
CFBP1159	96,9	96,83	96,45	96,79	96,62	96,81	96,84	96,7	96,82	96,48	96,83	94,93	96,83	96,71	n,a	96,82	96,65	96,65	96,66	96,85	96,5	96,54	94,94	96,88	96,9	96,8
CFBP7179	96,66	96,61	96,26	96,67	96,31	96,58	96,59	96,36	96,62	96,15	96,6	94,66	96,53	96,57	96,67	96,53	96,32	96,32	96,39	96,61	96,39	96,44	94,67	96,61	96,64	96,53
maff301420	96,93	96,82	96,39	96,8	96,6	96,84	96,88	96,72	96,86	96,43	96,88	94,93	96,88	96,68	96,85	96,76	96,65	96,67	96,71	96,8	96,49	96,54	94,95	96,9	96,95	96,8
xap33	96,93	96,85	96,43	96,83	96,61	96,88	96,88	96,73	96,88	96,48	96,87	94,94	96,89	96,71	96,83	96,8	96,66	96,67	96,69	96,86	96,53	96,57	94,98	96,91	96,92	96,83
maff311562	96,9	96,8	96,33	96,77	96,56	96,8	96,83	96,69	96,81	96,41	96,87	94,9	96,85	96,65	96,81	96,75	96,63	96,65	96,68	96,77	96,43	96,48	94,94	96,86	96,92	96,75
CFBP2528	96,69	96,62	96,29	96,71	96,34	96,6	96,62	96,4	96,64	96,19	96,62	94,68	96,59	96,6	96,71	96,57	96,36	96,35	96,42	96,65	96,4	96,45	94,68	96,61	96,67	96,58
nccb100457	96,91	96,78	96,45	96,77	96,58	96,81	96,84	96,69	96,81	96,43	96,84	94,91	96,84	96,67	96,81	96,78	96,63	96,62	96,64	96,78	96,43	96,48	94,93	96,88	96,89	96,78
CFBP3894	96,9	96,82	96,4	96,8	96,58	96,84	96,84	96,69	96,84	96,44	96,84	94,9	96,85	96,66	96,8	96,76	96,62	96,63	96,67	96,82	96,47	96,52	94,94	96,87	96,91	96,8
CFBP2565	96,89	96,81	96,51	96,79	96,62	96,79	96,87	96,69	96,8	96,48	96,8	94,92	96,83	96,71	96,8	96,83	96,61	96,66	96,64	96,84	96,48	96,53	94,94	96,89	96,92	96,81

b :

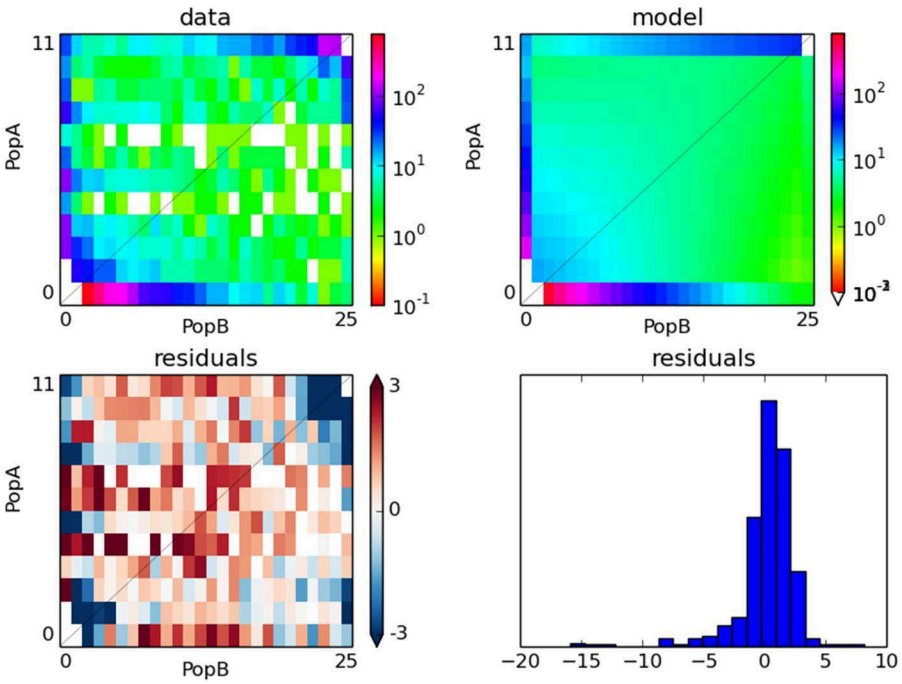
	CFBP7653	CFBP8152	CFBP7645	CFBP7622	CFBP3123	CFBP3122
CFBP8149	93,28	94,27	94,57	93,18	93,91	94,83
CFBP8147	93,21	94,24	94,51	93,12	93,88	94,73
CFBP7407	93,18	94,11	94,32	93,2	93,78	94,56
CFBP7651	93,25	94,27	94,59	93,15	93,87	94,76
CFBP8138	93,26	94,21	94,53	93,15	93,84	94,7
CFBP8132	93,24	94,23	94,51	93,13	93,86	94,77
CFBP8142	93,25	94,21	94,55	93,15	93,87	94,78
CFBP8153	93,31	94,23	94,56	93,22	93,89	94,73
CFBP7697	93,22	94,21	94,53	93,12	93,87	94,76
CFBP8150	93,13	94,1	94,38	93,07	93,73	94,56
CFBP6762	93,25	94,19	94,55	93,14	93,88	94,79
CFBP7408	92,78	93,49	93,72	92,66	93,25	93,79
CFBP6827	93,23	94,22	94,56	93,15	93,9	94,79
CFBP7652	93,35	94,35	94,76	93,23	93,86	94,7
CFBP3523	93,26	94,21	94,53	93,12	93,85	94,75
CFBP7410	93,24	94,25	94,53	93,13	93,88	94,73
CFBP6771	93,26	94,19	94,5	93,15	93,85	94,69
CFBP8130	93,24	94,19	94,51	93,15	93,85	94,68
CFBP7614	93,24	94,2	94,52	93,13	93,84	94,72
BC2528	93,22	94,22	94,52	93,15	93,89	94,76
CFBP7629	93,36	94,64	95,12	93,18	93,83	94,6
CFBP7634	93,38	94,66	95,16	93,21	93,87	94,65
CFBP7409	92,78	93,49	93,73	92,67	93,25	93,79
CFBP1022	93,23	94,26	94,57	93,13	93,93	94,82
CBFP7150	93,3	94,34	94,62	93,21	93,94	94,82
CFBP7610	93,24	94,24	94,57	93,13	93,85	94,76
CFBP7604	93,3	94,23	94,52	93,15	93,8	94,66
CFBP7653	---	92,84	93,01	97,88	93,74	93,42
CFBP8152	92,84	---	97,74	92,8	93,5	93,9
CFBP7645	93,01	97,74	---	92,91	93,58	94,03
CFBP7622	97,88	92,8	92,91	---	93,64	93,34
CFBP3123	93,74	93,5	93,58	93,64	---	95,68
CFBP3122	93,42	93,9	94,03	93,34	95,68	---
ncppb1447	93,47	94,29	94,71	93,38	93,94	94,68
12710	93,47	94,23	94,76	93,41	93,93	94,86
maff301427	93,32	94,28	94,66	93,24	93,97	94,75
CFBP1159	93,31	94,26	94,68	93,18	93,9	94,73
CFBP7179	93,44	94,35	94,78	93,36	93,93	94,67
maff301420	93,32	94,29	94,66	93,24	93,95	94,73
xap33	93,34	94,29	94,68	93,24	93,98	94,75
maff311562	93,27	94,27	94,63	93,19	93,91	94,67
CFBP2528	93,45	94,35	94,76	93,35	93,97	94,7
nccb100457	93,29	94,28	94,66	93,21	93,89	94,72
CFBP3894	93,31	94,26	94,64	93,21	93,94	94,73
CFBP2565	93,3	94,32	94,71	93,17	93,87	94,75

c :

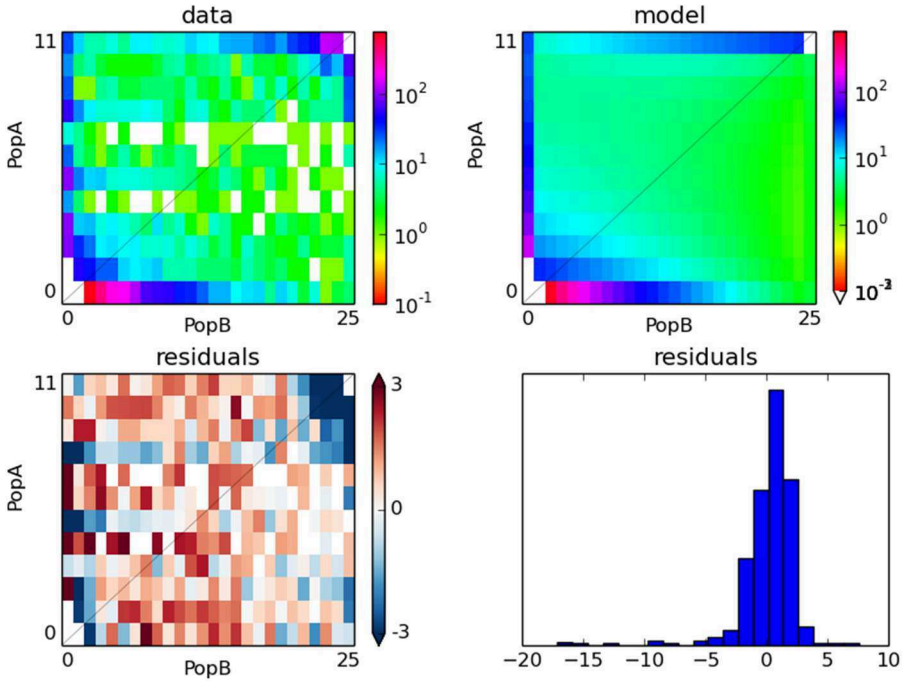
	ncppb1447	12710	maff301427	CFBP1159	CFBP7179	maff301420	xap33	maff311562	CFBP2528	nccb100457	CFBP3894	CFBP2565
CFBP8149	96,74	96,65	96,94	96,9	96,66	96,93	96,93	96,9	96,69	96,91	96,9	96,89
CFBP8147	96,68	96,53	96,84	96,83	96,61	96,82	96,85	96,8	96,62	96,78	96,82	96,81
CFBP7407	96,29	96,22	96,43	96,45	96,26	96,39	96,43	96,33	96,29	96,45	96,4	96,51
CFBP7651	96,75	96,64	96,83	96,79	96,67	96,8	96,83	96,77	96,71	96,77	96,8	96,79
CFBP8138	96,37	96,38	96,62	96,62	96,31	96,6	96,61	96,56	96,34	96,58	96,58	96,62
CFBP8132	96,66	96,53	96,85	96,81	96,58	96,84	96,88	96,8	96,6	96,81	96,84	96,79
CFBP8142	96,67	96,59	96,9	96,84	96,59	96,88	96,88	96,83	96,62	96,84	96,84	96,87
CFBP8153	96,45	96,43	96,74	96,7	96,36	96,72	96,73	96,69	96,4	96,69	96,69	96,69
CFBP7697	96,72	96,58	96,88	96,82	96,62	96,86	96,88	96,81	96,64	96,81	96,84	96,8
CFBP8150	96,22	96,2	96,46	96,48	96,15	96,43	96,48	96,41	96,19	96,43	96,44	96,48
CFBP6762	96,68	96,57	96,88	96,83	96,6	96,88	96,87	96,87	96,62	96,84	96,84	96,8
CFBP7408	94,68	94,7	94,94	94,93	94,66	94,93	94,94	94,9	94,68	94,91	94,9	94,92
CFBP6827	96,66	96,58	96,89	96,83	96,53	96,88	96,89	96,85	96,59	96,84	96,85	96,83
CFBP7652	96,65	96,54	96,72	96,71	96,57	96,68	96,71	96,65	96,6	96,67	96,66	96,71
CFBP3523	96,77	n,a	96,87	n,a	96,67	96,85	96,83	96,81	96,71	96,81	96,8	96,8
CFBP7410	96,61	96,47	96,81	96,82	96,53	96,76	96,8	96,75	96,57	96,78	96,76	96,83
CFBP6771	96,4	96,37	96,66	96,65	96,32	96,65	96,66	96,63	96,36	96,63	96,62	96,61
CFBP8130	96,39	96,35	96,69	96,65	96,32	96,67	96,67	96,65	96,35	96,62	96,63	96,66
CFBP7614	96,45	96,4	96,73	96,66	96,39	96,71	96,69	96,68	96,42	96,64	96,67	96,64
BC2528	96,7	96,56	96,84	96,85	96,61	96,8	96,86	96,77	96,65	96,78	96,82	96,84
CFBP7629	96,45	96,37	96,51	96,5	96,39	96,49	96,53	96,43	96,4	96,43	96,47	96,48
CFBP7634	96,5	96,41	96,55	96,54	96,44	96,54	96,57	96,48	96,45	96,48	96,52	96,53
CFBP7409	94,7	94,72	94,96	94,94	94,67	94,95	94,98	94,94	94,68	94,93	94,94	94,94
CFBP1022	96,69	96,56	96,91	96,88	96,61	96,9	96,91	96,86	96,61	96,88	96,87	96,89
CBFP7150	96,72	96,61	96,96	96,9	96,64	96,95	96,92	96,92	96,67	96,89	96,91	96,92
CFBP7610	96,6	96,49	96,82	96,8	96,53	96,8	96,83	96,75	96,58	96,78	96,8	96,81
CFBP7604	96,36	96,33	96,64	96,62	96,28	96,65	96,64	96,6	96,33	96,61	96,61	96,6
CFBP7653	93,47	93,47	93,32	93,31	93,44	93,32	93,34	93,27	93,45	93,29	93,31	93,3
CFBP8152	94,29	94,23	94,28	94,26	94,35	94,29	94,29	94,27	94,35	94,28	94,26	94,32
CFBP7645	94,71	94,76	94,66	94,68	94,78	94,66	94,68	94,63	94,76	94,66	94,64	94,71
CFBP7622	93,38	93,41	93,24	93,18	93,36	93,24	93,24	93,19	93,35	93,21	93,21	93,17
CFBP3123	93,94	93,93	93,97	93,9	93,93	93,95	93,98	93,91	93,97	93,89	93,94	93,87
CFBP3122	94,68	94,86	94,75	94,73	94,67	94,73	94,75	94,67	94,7	94,72	94,73	94,75
ncppb1447	---	98,1	97,36	97,68	99,17	97,36	97,38	97,33	99,39	97,7	97,36	97,69
12710	98,1	---	97,46	97,89	98,07	97,44	97,49	97,43	98,09	97,9	97,46	97,9
maff301427	97,36	97,46	---	98,7	97,28	99,85	99,84	99,81	97,32	98,7	99,83	98,7
CFBP1159	97,68	97,89	98,7	---	97,51	98,69	98,75	98,62	97,57	99,57	98,72	99,5
CFBP7179	99,17	98,07	97,28	97,51	---	97,28	97,28	97,24	99,21	97,57	97,25	97,49
maff301420	97,36	97,44	99,85	98,69	97,28	---	99,85	99,85	97,31	98,68	99,81	98,7
xap33	97,38	97,49	99,84	98,75	97,28	99,85	---	99,81	97,35	98,79	99,87	98,78
maff311562	97,33	97,43	99,81	98,62	97,24	99,85	99,81	---	97,29	98,64	99,79	98,66
CFBP2528	99,39	98,09	97,32	97,57	99,21	97,31	97,35	97,29	---	97,62	97,29	97,56
nccb100457	97,7	97,9	98,7	99,57	97,57	98,68	98,79	98,64	97,62	---	97,3	97,56
CFBP3894	97,36	97,46	99,83	98,72	97,25	99,81	99,87	99,79	97,29	97,3	---	98,74
CFBP2565	97,69	97,9	98,7	99,5	97,49	98,7	98,78	98,66	97,56	97,56	98,74	---

Figure III-S1 : Results obtained with δaδi for the 6 other models. . SI : Strict isolation; IM: isolation with migration; AM: ancient migration; SC : secondary contact. 2M corresponds to models tested with two migration rates different, one along genome and one in genomic island. For each result, the first HeatMap corresponds to the joint allele frequency spectrum obtained for group A in vertical axis and group B in horizontal axis. The second HeatMap corresponds to that inferred according the model. The two last graphs correspond to the residuals distribution, that are the difference in number of alleles in our dataset compared to that expected in the model (in blue, the number is higher in our dataset, in red, the number is higher in the model)

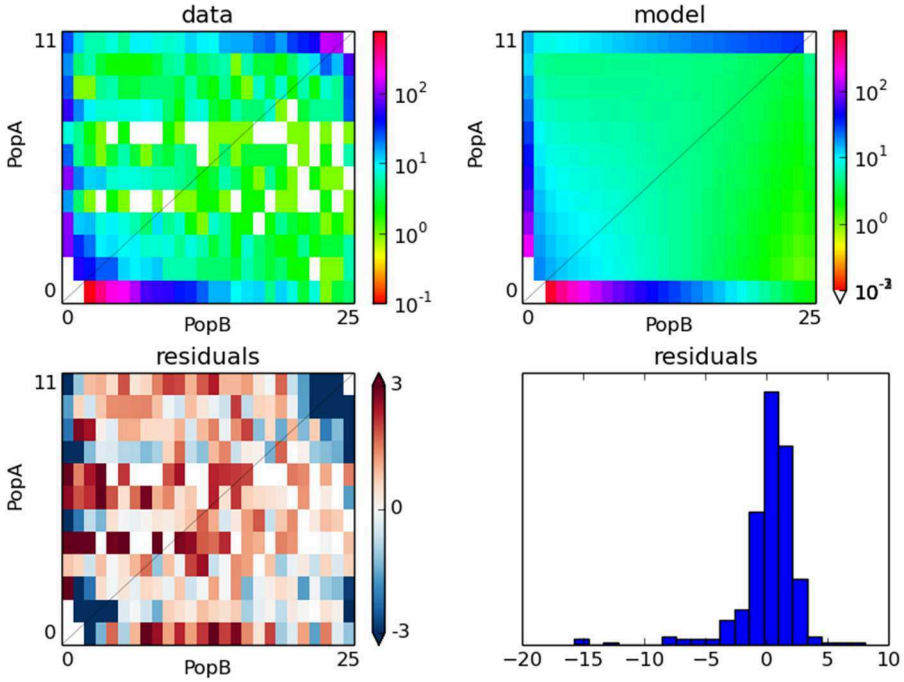
SI:



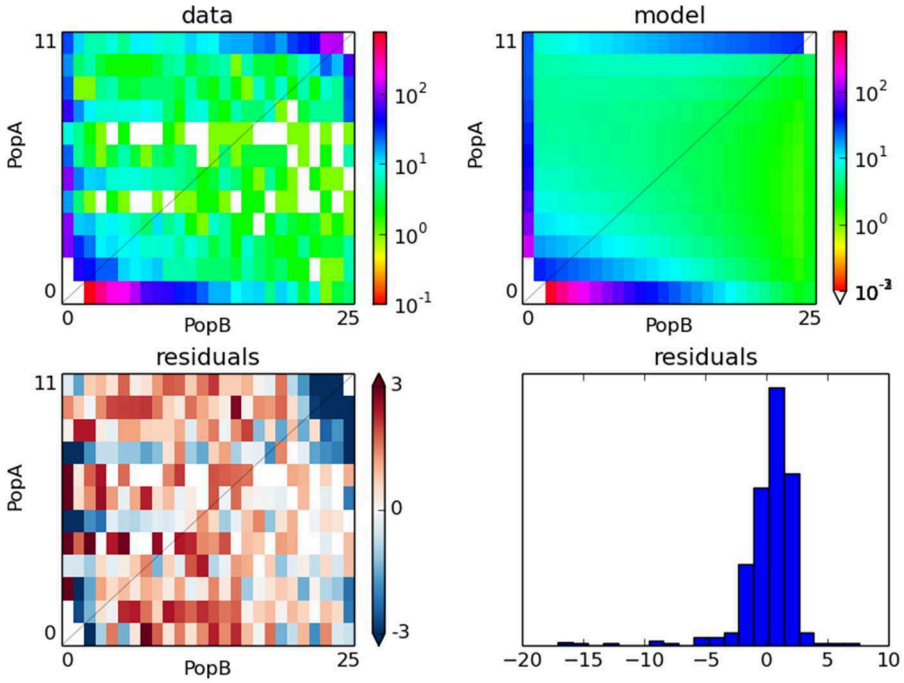
IM:


data
model
residuals
residuals
PopA
PopB
11
0
25
10^2
10^1
10^0
10^{-1}
10^{-2}
3
-3
−20
−15
−10
−5
5
10

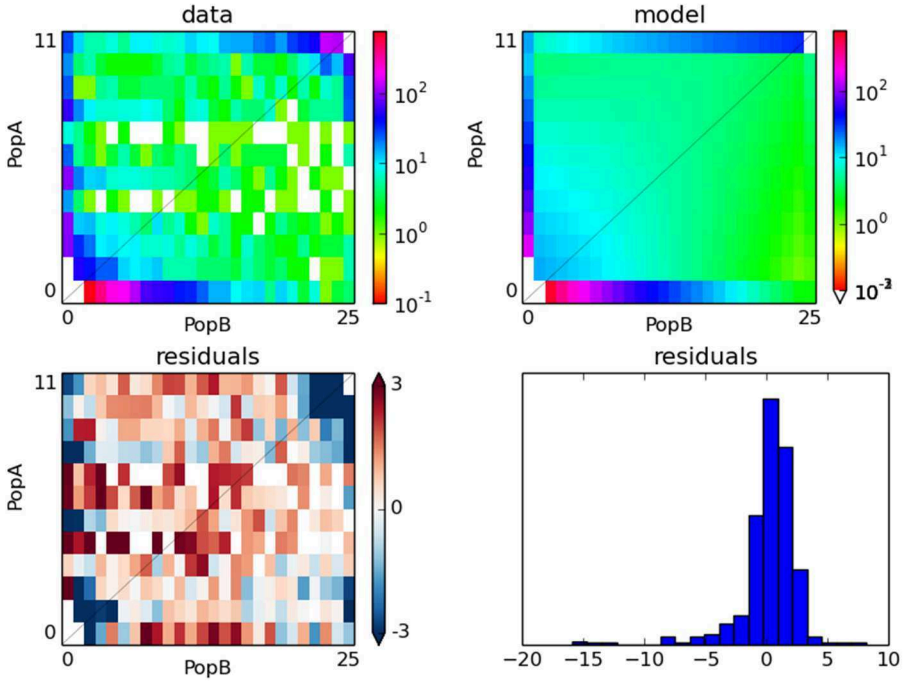
AM:


data
model
residuals
residuals
PopA
PopB
11
0
25
10^2
10^1
10^0
10^{-1}
10^{-2}
3
-3
−20
−15
−10
−5
5
10

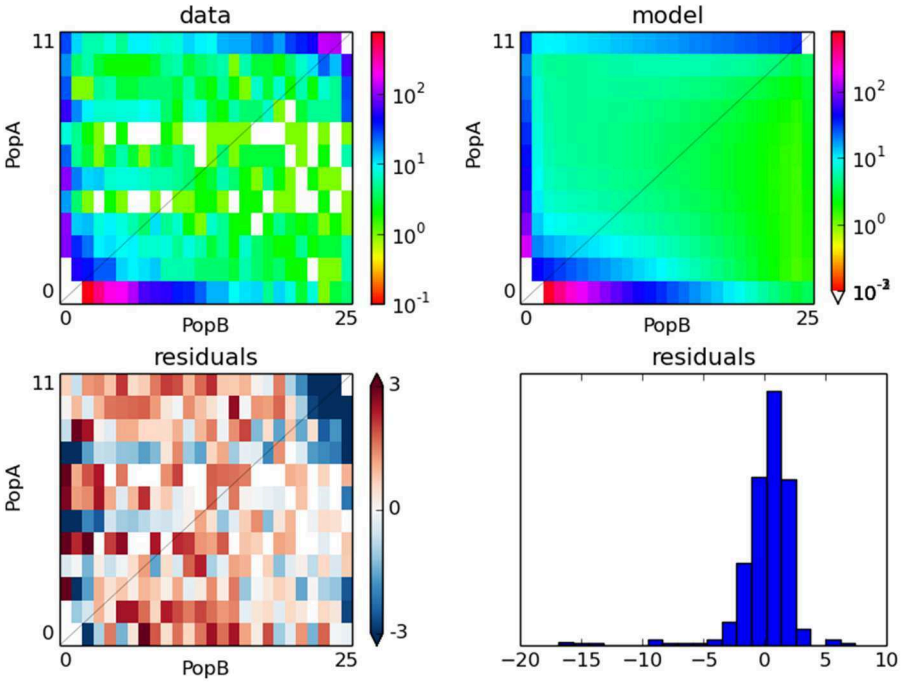
SC:


data
model
residuals
residuals
PopA
PopB
11
0
25
10^2
10^1
10^0
10^{-1}
3
-3
−20
−15
−10
−5
5
10

AM2M:


data
model
residuals
residuals
PopA
PopB
11
0
25
10^2
10^1
10^0
10^{-1}
3
-3
−20
−15
−10
−5
5
10

SC2M:


data
model
residuals
residuals
PopA
PopB
11
0
25
10^2
10^1
10^0
10^{-1}
3
-3
−20
−15
−10
−5
0
5
10

Chapitre III

Etude phénotypique des souches commensales et pathogènes appartenant à l'espèce *Xanthomonas arboricola*[13]

Merda D., De Wever A., Preveaux A., Jacques $MA^{\$}$., Fischer-Le Saux $M^{\$}$.

I. Introduction

L'écologie des souches commensales appartenant au genre *Xanthomonas* est très mal connue, elles sont en effet caractérisées uniquement par leur incapacité à provoquer des symptômes sur leur plante hôte d'isolement. Cependant, le flux de gènes détecté au cours de la divergence des souches commensales et pathogènes dans l'espèce *X. arboricola* laisse supposer que ces souches commensales et pathogènes vivent en sympatrie (chapitre 2). Les souches commensales ont été isolées de plantes hôtes communes aux souches pathogènes telles que le noyer ou divers *Prunus* (Essakhi et al. 2015 ; Garita-Cambronero et al. 2016b), et également de plantes hôtes d'autres espèces de *Xanthomonas* comme le haricot (Boureau et al. 2013 ; Grimault et al. 2014). Il est cependant difficile de savoir si au sein de la plante, ces deux types de souches colonisent les mêmes sites voire même si elles coexistent au sein des mêmes biofilms. Au sein de ces structures, les échanges de matériel génétique sont intensifiés (Danhorn & Fuqua, 2007) et la cohabitation de souches commensales

[13]Pour ce travail, les analyses de phénotypage et de comparaison de taille population entière et internalisée ont été réalisées par Alexis DeWever, au cours de son stage de Master 1. Anne Préveaux a participé à la réalisation des manipulations de dynamiques de taille de population sur haricot et aux analyses statistiques de ces dynamiques. Marie-Agnès Jacques et Marion Fischer-Le Saux m'ont encadrée pour poser les questions biologiques, pour l'analyse des données, pour leur interprétation et la rédaction du manuscrit. Pour ma part, j'ai encadré Alexis DeWever pendant son stage de Master 1, j'ai participé aux expériences de phénotypages et de manipulations *in planta*, et j'ai rédigé le manuscrit.

et pathogènes au sein de même biofilm pourrait expliquer ce flux de gènes persistant au cours de l'histoire évolutive.

En dehors de leur capacité à provoquer des symptômes, les agents pathogènes sont également capables de détecter, coloniser et s'établir sur des sites favorables à leur survie et vie sur la plante. Ces étapes font intervenir le chimiotactisme, la mobilité et la formation de biofilm. Dans le chimiotactisme, la bactérie détecte des molécules dans son environnement grâce aux Methyl-accepting Chemotaxis Proteins (MCP) qui vont transduire un signal jusqu'au flagelle et l'activer (Alexander & Zhulin, 2007). Il existe trois types de mobilité, la nage (swimming) et l'essaimage (swarming), qui font intervenir le flagelle et la contraction (twitching) qui fait intervenir le pilus de type IV (Kearns, 2010). La formation de biofilm fait intervenir des adhésines (Danhorn & Fuqua, 2007).

La comparaison de quelques phénotypes (mobilité, résistance au cuivre, activité pectinolytique) a été réalisée pour quatre souches pathogènes et commensales isolées de noyer (Cesbron et al. 2015). Cette étude a montré qu'il existait des différences entre les souches pathogènes et commensales, suggérant des niches écologiques différentes. De plus, une autre étude phénotypique a été réalisée en comparant une souche commensale isolée de faux merisier (*Prunus mahaleb*) (CITA 44) et les différents pathovars de l'espèce *X. arboricola* (Garita-Cambronero et al. 2016a). Les phénotypes comparés étaient l'utilisation de sources de carbone, les sources attractives et déclenchant la cascade de transduction du chimiotactisme, et la mobilité. Les auteurs ont montré que cette souche était mobile comme les souches pathogènes testées, mais ils ont noté des différences concernant les sources de carbone utilisées et les sources attractives entre les souches pathogènes et la souche commensale. La multiplication des souches commensales composant l'espèce *X. maliensis* sur leur hôte d'isolement, le riz a été comparée à celle l'agent pathogène *X. oryzae* pv. *oryzae*. Des différences de tailles de population ont été observées, l'agent pathogène étant capable d'atteindre des tailles de population de 10^9 UFC/feuilles à 14 jours après inoculation et la souche commensale atteignant seulement 10^7 UFC/feuilles (Triplett et al. 2016).

L'espèce *Xanthomonas arboricola* est composée de 8 pathovars, capables de provoquer des maladies sur différentes espèces hôtes, comme les arbres fruitiers du genre *Prunus*, le noyer ou le noisetier. L'existence de souches commensales appartenant à cette espèce a été mise en évidence par Vauterin et al. 1996, puis d'autres études ont permis de mettre en évidence que cette espèce était caractérisée par un grand nombre de souches sans pouvoir pathogène avéré, représentant une grande diversité, et isolées d'un grand nombre d'espèces végétales (Guimbaud, 2012 ; Essakhi et al. 2015 ; Vandroemme et al. 2013 ; Cesbron et al. 2015). La ca-ractérisation de ces souches ne se fait que sur leur incapacité à induire des symptômes sur leur

hôte d'isolement, mais sont-elles capables de coloniser une espèce végétale? Pour répondre à cette question, nous avons étudié les phénotypes connus pour être impliqués dans la colonisation de l'hôte, à savoir la mobilité, l'adhésion et la formation de biofilm, ainsi que leur capacité à se multiplier sur l'hôte à partir duquel elles ont été isolées. Nous avons étudié l'inférence de gains et de perte de MCP au cours de l'histoire évolutive de ces souches, afin de déterminer s'il existait des différences entre les souches commensales et pathogènes. Enfin, dans le but de déterminer si les souches commensales et pathogènes peuvent survivre au sein d'un même biofilm, nous avons réalisé des tests d'antibiose en confrontant deux à deux souches pathogènes et souches commensales.

II. Matériel et Méthodes

1. Matériel biologique

Une collection de 40 souches a été utilisée pour étudier les différents phénotypes (Table IV-S1). Parmi ces 40 souches, 15 sont des agents pathogènes et 25 sont des souches commensales. Parmi ces souches, 39 font partie de la collection de génomes utilisée dans le chapitre précédent, et appartiennent aux 3 groupes génétiques A, B et C (chapitre 1). La dernière souche appartient au pathovar *juglandis*, CFBP 8350, sa séquence génomique est disponible publiquement, cependant, sa mauvaise qualité nous a obligé à la retirer des analyses de génomiques. Une sous collection de 6 souches commensales isolées de haricot a été sélectionnée pour étudier leur dynamique de population sur l'hôte d'isolement. Les souches sont conservées dans une solution de glycérol à 40% à -80°C.

2. Méthodes

a. Etude des phénotypes impliqués dans la phase asymptomatique du cycle infectieux

Quantification et caractérisation de la mobilité bactérienne

La mobilité des souches a été testée en gélose molle (Darrasse et al. 2013). Les souches ont été cultivées à 28°C pendant 4 jours dans du milieu MOKA (extrait de levure 4 g/L; acides casaminés 8 g/L; KH_2PO_4 2 g/L; $MgSO_4$, $7H_2O$ 0,3 g/L) contenant 0,2% ou 0,6% d'agar avec ou sans 0,05% de 2,3,5-triphényltétrazolium chloride (TTC). L'addition de chlorure de triphényltétrazolium (TTC) permet une meilleure visualisation de l'envahissement de la gélose par les bactéries mobiles. En

effet, les bactéries réduisent le TTC en formazan insoluble de couleur rouge, facilitant la lecture des résultats. Cependant certaines souches peuvent être sensibles à ce composé d'où la nécessité d'utiliser en parallèle les milieux sans TTC. Les témoins utilisés sont : CFBP 2528 (non-mobile), et CFBP 7179, CFBP 7651 et CFBP 7634 connues pour être mobiles (Cesbron et al. 2015). Une goutte de 10 μl d'une suspension bactérienne à 1 x 10^8 UFC/ml a été déposée au centre de la boite et le rayon de la zone colonisée par la bactérie a été mesuré 4 jours après inoculation. La lecture du swarming se fait dans les milieux avec 0,6% d'agar, et celles du swimming et du twitching dans les milieux à 0.2% d'agar. Le twitching peut être différencié du swarming, car le déplacement par twitching se fait au fond de la boite de Pétri, contrairement au swarming où elle envahit l'ensemble de la gélose.

Quantification de l'adhésion bactérienne

La capacité d'adhésion des souches au polypropylène a été testée en milieu MME (KH_2PO_4 2 10.5 g/L; KH_2PO_4 2 4.5 g/L, $(NH_4)_2SO_4$ à 1 g/L, glutamate 1M 20 ml/L, $MgSO_4$ 1M 1 ml/L et acides casaminés 0,15 g/L) en plaque 96 puits (Darsonval et al. 2009). Un volume de 10 μl d'une suspension bactérienne à 10^8 UFC/ml a été déposé dans 190 μL de MME et incubé à 28°C pendant 4 jours. La culture liquide a été éliminée à 4 jours et sa densité optique (DO) a été mesurée au spectrophotomètre à 600 nm avec le logiciel KC4 (Bio-Tek Instruments Inc., Winooski, VT, USA), pour vérifier la croissance bactérienne. Les bactéries adhérentes ont été colorées avec 200 μl de cristal violet (CV) à 1% pendant 15 min. Le CV non fixé a été éliminé puis les puits rincés trois fois avec de l'eau distillée. Enfin, 200 μl d'éthanol à 90% ont été ajoutés pour solubilisé le CV fixé et DO mesurée à 600 nm.

Quantification du biofilm bactérien

La capacité des souches à former un biofilm et des pellicules a été testée en milieu SoBG dans des tubes en verre (tryptone 20 g/L; extrait de levure 5 g/L; NaCl 0,5 g/L; $MgSO_4$ 2,4 g/L; KCl 0,186 g/L; glycérol 2%) (Darsonval et al. 2009). Un volume de 500 μL d'une suspension à 10^8 UFC/ml a été ajouté à 5 ml de milieu SoBG et incubé pendant 11 jours à 28°C sans agitation. La présence de pellicule (biofilm non adhérant situé à l'interface entre le milieu et l'air) a été notée. La culture liquide a été éliminée à 11 jours et sa DO mesurée au spectrophotomètre à 600 nm, dans le but de vérifier la croissance bactérienne. Le biofilm adhérant à la paroi du tube a été coloré avec 5 mL de CV à 1% pendant 15min. Les 5 ml de CV ont été éliminés puis les tubes rincés trois fois avec de l'eau distillée. Enfin, le biofilm a été solubilisé dans 5 ml d'éthanol à 90% puis la DO à 600 nm mesurée. Le

témoin utilisé est un tube de SoBG non inoculé.

b. Analyses des répertoires de gènes associés aux phases précoces de l'interaction

La recherche des MCP, TonB-dependent transporters (TBDT) et adhésines par PCR est décrite dans le chapitre 1, pour une collection de 41 souches (12 souches pathogènes et 29 commensales). Parmi ces 41 souches, 24 sont incluses dans la collection utilisée pour les analyses phénotypiques. La phylogénie de ces 41 souches a été réalisée en utilisant 7 gènes de ménage (*atpD*, *dnaK*, *efP*, *fyuA*, *glnA*, *gyrB* et *rpoD*), et la méthode de maximum de vraisemblance. Les inférences de gain et de perte de ces gènes au cours de l'histoire évolutive ont été réalisées avec Genoplast (Didelot et al. 2009). Trois analyses indépendantes ont été réalisées, avec pour chacune une période de pré-chauffage de la chaine de Markov Monte Carlo (MCMC) composée de 500 000 itérations, suivi de 2 000 000 d'itérations.

c. Tests d'antibiose

Afin de tester si les souches commensales et pathogènes peuvent coexister sur un même hôte, nous avons réalisé des tests d'antibiose in vitro. Pour cela, nous avons réalisé l'inoculation d'une goutte de 10 μL de suspension de la souche commensale à 10^8 UFC/ml sur un milieu TSA 10% contenant la souche pathogène à 10^8 UFC/ml. L'incubation des souches a été réalisée pendant 6 jours à 28°C. Le témoin de croissance des souches correspond à l'inoculation des souches commensales et pathogènes sur un milieu TSA 10% seul. Trois types d'interaction peuvent être observés : (i) Les deux souches se développent, il n'y a pas d'interaction, (ii) la croissance de l'une des deux souches est inhibée par la présence de l'autre, il y a inhibition de la souche commensale ou pathogène, (iii) La croissance des deux souches est inhibée, il y a double inhibition. Ce test a été répété trois fois, afin de valider les résultats, et chaque test était composé de trois réplicats par souche.

d. Dynamique de taille de populations des souches commensales in planta

Les 6 souches utilisées ici ont été isolées de semences de haricot et appartiennent toutes à l'espèce *X. arboricola*. La souche pathogène utilisée en témoin positif est *X. fuscans* subsp. *fuscans* 4834-R. En effet, il n'existe pas de souche pathogène du haricot dans l'espèce *X. arboricola* ; nous n'avons donc pas pu utiliser en témoin positif une souche phylogénétiquement proche des souches commensales que nous

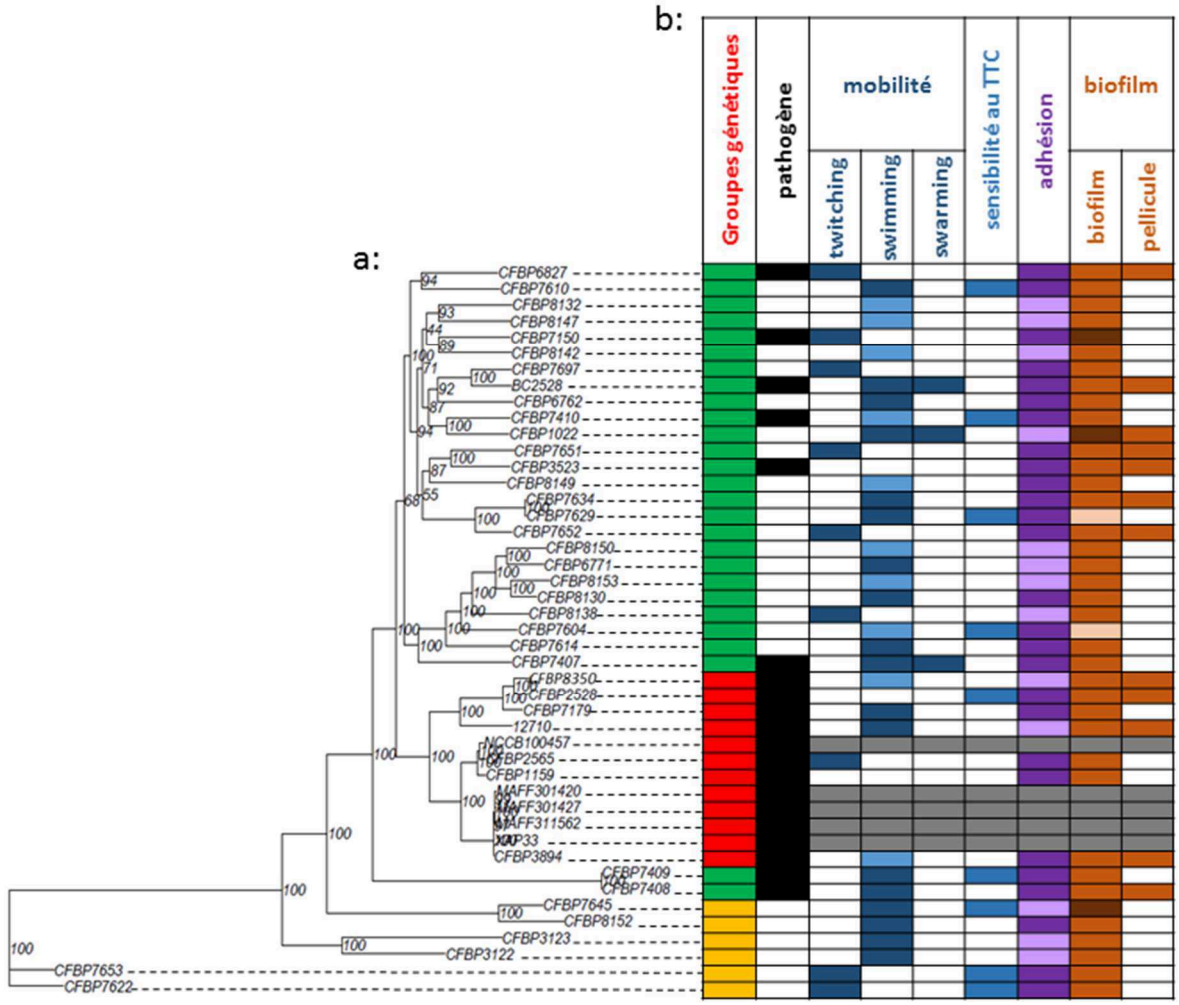


Figure IV-1 : Représentation des phénotypes des souches en fonction de leur positionnement phylogénétique.

a : Phylogénie en maximum de vraisemblance obtenue avec le concaténat des 1703 CDS constituant le core génome de l'ensemble des souches.

b : Tableau résumant les résultats des différents phénotypes étudiés. La première colonne représente l'appartenance des souches aux trois groupes génétiques obtenus par l'assignation bayésienne, basée sur les 7 gènes de ménage (en rouge, le groupe A, en vert, le groupe B et en jaune le groupe C) (voir chapitre 1). La deuxième colonne indique le caractère phytopathogène (en noir) et commensal (en blanc). Dans les colonnes suivantes, l'absence du phénotype est représentée par une case blanche, et son intensité est représentée par un dégradé (clair : pour un phénotype faiblement exprimé ; foncé : pour un phénotype fortement exprimé). Les cases grises correspondent aux souches qui n'ont pu être étudiées.

voulions tester. Ces dynamiques de populations ont été réalisées sur l'espèce hôte à partir de laquelle les souches ont été isolées.

Les suspensions bactériennes à la concentration de 1 x 10^6 UFC/ml ont été inoculées par pulvérisation sur le feuillage de 20 haricots (*Phaseolus vulgaris* cultivar flavert), âgés de 16 jours. Des prélèvements sont effectués 2h et 24h, 4 et 11 jours post-inoculation. A chaque date de prélèvement, la $1^{ère}$ trifoliée est prélevée sur cinq plantes par traitement. La taille des populations est quantifiée à chaque date de prélèvement, par des dilutions/étalements sur boites.

e. Statistiques

Pour le test de mobilité, un test exact de Fisher avec un degré de liberté (ddl) de 1 a été réalisé, pour comparer les répétitions 1 et 2 (3 réplicats par souches) des tests de mobilité. Deux classes d'effectifs ont été choisies : souches mobiles et souches non mobiles. Pour ce test, l'hypothèse nulle est " il n'y a pas de différence significative entre les 2 répétitions ". Un test exact de Fisher avec un ddl de 5 a ensuite été réalisé pour comparer les souches pathogènes et commensales. Six classes d'effectifs ont été choisies : " swarming, faiblement mobile ", " swarming, fortement mobile ", " swimming, faiblement mobile ", " swimming, fortement mobile " et " twitching ", " non mobile ". L'ensemble des tests a été réalisé sur les données sans TTC car certaines souches peuvent être sensibles à ce composé ce qui peut altérer leur mobilité. Pour ces tests, l'hypothèse nulle est " il n'y a pas de différence significative entre les souches pathogènes et commensales ".

Pour le test d'adhésion, un test exact de Fisher avec un ddl de 2 a été réalisé pour comparer les répétitions 1 et 2 (3 réplicats par souches) des tests d'adhésion. Trois classes d'effectifs ont été choisies : " souches non adhérentes ", " souches faiblement adhérentes " et " souches très adhérentes ". Le choix des classes a été fait en prenant comme témoin la souche pathogène CFBP 7179 (très adhérente) et la souche commensale CFBP 7651 (non adhérente). Un test exact de Fisher avec un ddl de 2 a été réalisé pour comparer les souches pathogènes et commensales. Les classes utilisées pour ce test étaient les mêmes que précédemment.

Pour la formation de biofilm, un test exact de Fisher avec un ddl de 2 a été réalisé pour comparer les répétitions 1 et 2 (3 réplicats par souches) des tests de formation de biofilm. Trois classes d'effectifs ont été choisies : " biofilm faible (DO<0.1) ", " biofilm intermédiaire (DO comprise entre 0.1 et 0.2) " et biofilm important (DO>0.2) . Un test exact de Fisher avec un ddl de 2 a été réalisé pour comparer les souches pathogènes et commensales, en utilisant les mêmes classes que précédemment. Enfin, pour déterminer s'il y a des différences significatives de formation de pellicules entre les souches pathogènes et commensales, un test exact

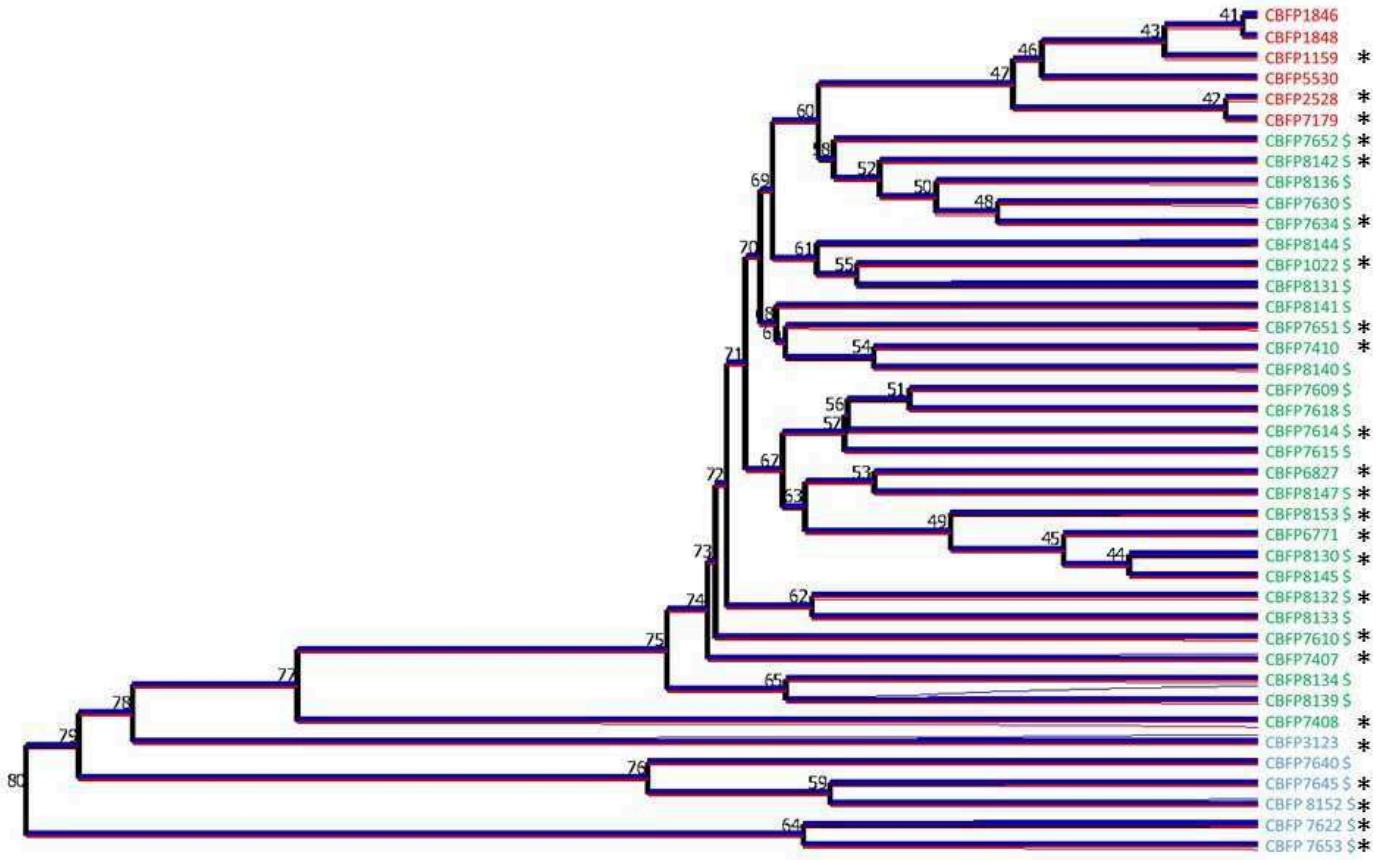


Figure IV-2 : Inférence de gain et de perte de MCP, TBDT et adhésines au cours de l'histoire évolutive de l'espèce *X. arboricola*. Les événements de pertes sont représentés en bleu et les gains en rouge. Les répertoires correspondent à ceux présentés dans Merda et al. (2016). La phylogénie a été reconstruite en utilisant 7 gènes de ménage (*atpD*, *dnaK*, *efP*, *fyuA*, *glnA*, *gyrB* et *rpoD*), et la méthode de maximum de vraisemblance. Les valeurs à chaque nœud correspondent aux valeurs de boostrap calculés pour 1000 réplicats. La couleur du nom des souches correspond à leur appartenance aux groupes génétiques décrits par Merda et al. 2016 (en rouge, le groupe A, en vert, le groupe B et en bleu, le groupe C). $: indique les souches commensales. * : indique les souches pour lesquelles les analyses phénotypiques et le séquencage de génomes ont été réalisés.

de Fischer avec un ddl de 1 a été réalisé. Pour ce test, les classes d'effectifs utilisées étaient : “ présence ” ou “ absence ” de pellicule.

Pour la multiplication sur haricot, la taille d'échantillon (N=5) étant petite, nous avons réalisé un test non paramétrique de Mann-Whitney. Ce test a permis de comparer pour chacune des souches s'il existait une différence significative de la taille de la population en fonction du temps.

Un test de χ^2 avec un ddl de 6 a été réalisé, pour comparer les répétitions 1, 2 et 3 (3 réplicats par souches). Pour ce test, quatre classes d'effectifs ont été choisies : “ pas d'interaction ”, “ inhibition de la souche commensale ”, “ inhibition de la souche pathogène ” et “ double inhibition ”. Si ce test est non significatif, les données des trois répétitions sont traitées dans un même jeu de données. Ensuite, pour savoir s'il y a des différences significatives entre le nombre d'interactions des souches non pathogènes et des souches pathogènes, un test exact de Fisher a été réalisé avec un ddl de 3 en utilisant les mêmes classes que précédemment.

Tous les tests statistiques ont été réalisés sur le site internet biostatGV. Pour chaque test, nous n'avons pas trouvé de différence significative entre les répétitions, et donc l'ensemble des données a été traité comme un jeu de données unique.

III. Résultats

1. Mobilité, adhésion et formation de biofilms chez les souches commensales et phytopathogènes

L'analyse des phénotypes impliqués dans les étapes précoces de l'interaction ne montre pas de différence significative entre les agents pathogènes et les souches commensales. En effet, un test de Fisher a été réalisé pour chaque phénotype afin de comparer les souches pathogènes et commensales, les p-value étaient de 0.31, 0.08 et 0.78 pour la mobilité, l'adhésion et la formation de biofilm, respectivement. La capacité des souches à être fortement ou faiblement adhérentes ne semble pas liée à leur position phylogénétique. En effet, au sein du pathovar *juglandis*, les souches CFBP 8350 et 12710 présentent une faible adhésion tandis que la souche CFBP 2528 est fortement adhérente (fig. IV-1). Concernant les biofilms, la majorité des souches pathogènes ou commensales forment des biofilms intermédiaires, c'est-à-dire dont la DO du biofilm était comprise entre 0.1 et 0.2.

Le type de mobilité le plus fréquemment observé est la nage (fig. IV-1). En effet, 28 souches sur 40 présentent cette capacité. Il existe, cependant, des différences d'expression de ce phénotype, puisque 18 souches sur les 28 sont très mobiles et ont un diamètre de dispersion supérieur à celui observé chez la souche CFBP 7179

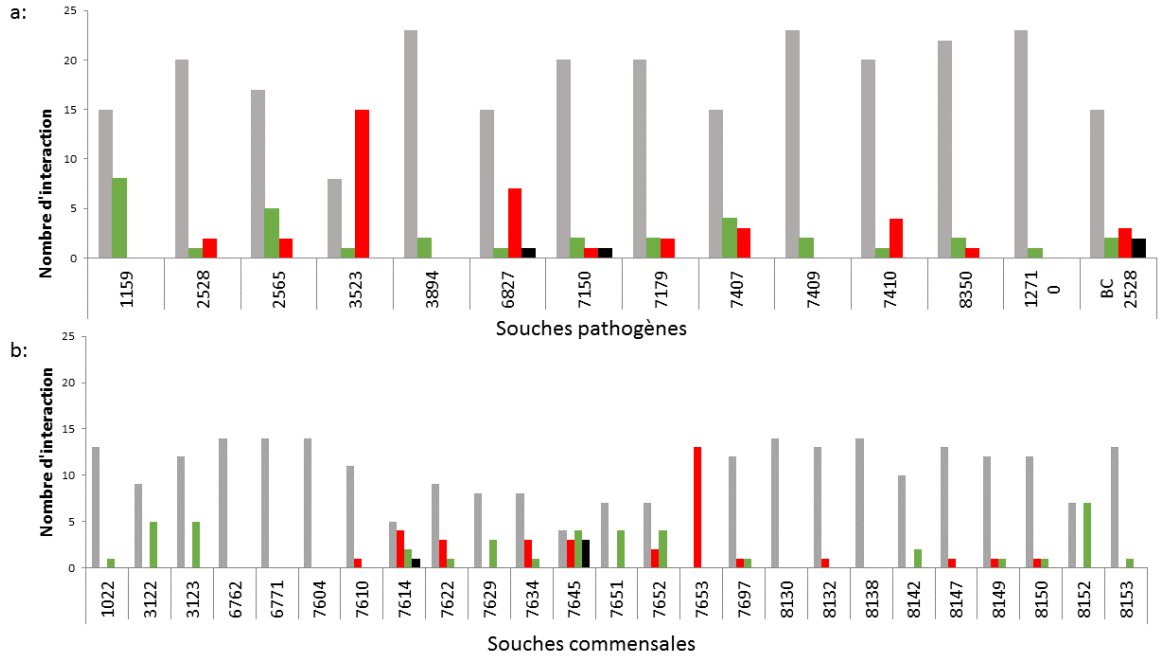


Figure IV-3 : Résultats des analyses d'antibiose suite à la confrontation d'une souche pathogène et une souche commensale. a : Résultats concernant les réponses des souches pathogènes en confrontation avec une commensale. b : Résultats concernant les réponses des souches commensales en confrontation avec une souche pathogène. En gris, il n'y a pas d'interactions, en rouge la souche inhibe les souches avec lesquelles elle est confrontée, en vert, la souche est inhibée, et en noir, les deux souches sont inhibées.

(témoin positif utilisé). Il existe des différences au sein d'un même pathovar. En effet, les deux souches CFBP 7408 et CFBP 7409, appartenant au pathovar *guizotiae* sont faiblement et fortement mobiles, respectivement.

Le twitching n'a pu être observé que chez les souches sans mobilité flagellaire. Sur les 11 souches sans mobilité flagellaire, 8 souches (3 pathogènes et 5 commensales) présentent un phénotype de twitching. Il n'y a pas de lien entre la phylogénie et ce mode de mobilité puisque les souches faisant du "twitching" sont dispersées dans la phylogénie. Le " swarming " est le mode de mobilité le plus rare dans les souches de notre collection puisque seulement 3 souches l'utilisent (une souche commensale et deux souches pathogènes). L'ensemble de ces 3 souches appartient au même groupe génétique, le groupe B.

Lors de l'analyse de la capacité à former des biofilms, 12 souches parmi les 40 de la collection formaient des pellicules à l'interface entre le milieu et l'air. Parmi celles-ci, 8 sont des souches pathogènes et 4 sont des souches commensales. Les souches pathogènes forment significativement (p-value = 0.03) plus de pellicules que les commensales, cependant ce caractère n'est pas commun à la majorité des souches pathogènes et il est généralement variable au sein d'un même pathovar. Parmi les souches pathogènes, seules les deux souches du pathovar *corylina* ne forment pas de pellicule.

2. Coexistence des souches pathogènes et commensales

Les souches pathogènes et commensales présentent les mêmes phénotypes associés aux phases précoces de l'interaction. Le choix d'un site d'installation est dépendant du chimiotactisme. Les répertoires de gènes codants les MCP mais aussi les TBDT et adhésines ont été établis au chapitre 1. L'inférence de gain et de perte de ces gènes au cours de l'histoire évolutive de l'espèce *X. arboricola* n'a pas mis en évidence d'évènements de modification de ces répertoires au cours de l'histoire de cette espèce (fig. IV-2). Ceci semble indiquer qu'il n'y a pas de différence entre les souches commensales et pathogènes dans les phases précoces de l'interaction, et probablement qu'elles colonisent les mêmes sites sur la plante.

Afin de déterminer si les souches commensales et pathogènes peuvent coexister sur un même hôte, des tests d'antibisose ont été réalisé entre un grand nombre de souches en mettant en contact une souche pathogène et une souche commensale (fig. IV-3). Le test de χ^2 ne montre pas de différence significative entre les souches pathogènes et commensales. De manière générale, aucune antibiose n'a été observée entre les souches commensales et pathogènes. Il existe tout de même deux exceptions, il s'agit des deux souches CFBP 3523 et CFBP 7653 qui inhibent la croissance des souches avec lesquelles elles se retrouvent en contact. La souche CFBP 7653 est

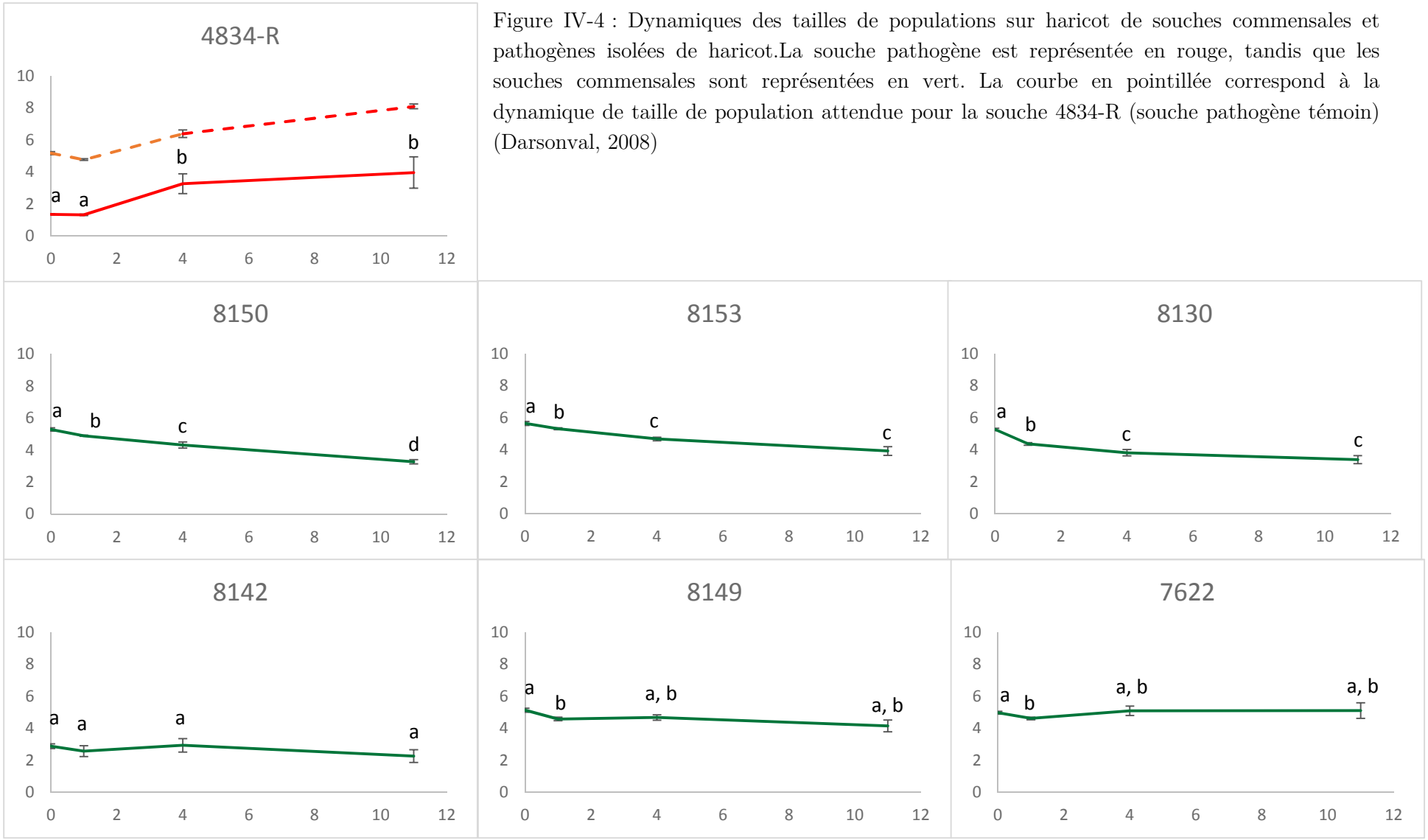


Figure IV-4 : Dynamiques des tailles de populations sur haricot de souches commensales et pathogènes isolées de haricot.La souche pathogène est représentée en rouge, tandis que les souches commensales sont représentées en vert. La courbe en pointillée correspond à la dynamique de taille de population attendue pour la souche 4834-R (souche pathogène témoin) (Darsonval, 2008)

une souche commensale tandis que la souche CFBP 3523 appartient au pathovar *celebensis*.

3. Multiplication *in planta* des souches commensales

Chez les souches commensales isolées du haricot, une analyse de la dynamique de la taille de population sur feuille de haricot jusqu'à 11 jours a montré une différence significative entre les souches commensales et la souche pathogène (*Xff* 4834-R). L'analyse des inoculums a montré que la concentration initiale de la souche *Xff* 4834-R était de 1 x 10^2 UFC/ml au lieu de 1 x 10^6 UFC/ml prévu pour ce test. On observe cependant bien pour cette souche une augmentation de la taille de population à partir de J4 indiquant des conditions favorables à sa multiplication. Pour information, nous avons représenté sur la figure IV-4 une cinétique provenant de Darsonval et collaborateurs (2008). En effet, cette souche étant utilisée en routine au sein de l'équipe, sa cinétique de multiplication sur le haricot dans ces conditions expérimentales est bien connue. La dynamique de population des souches commensales, quant à elle, est marquée par une diminution de la taille de population jusqu'à J4, puis cette taille de population s'est stabilisée jusqu'à J11. Cela est observé pour 5 souches sur les 6 inoculées. Ainsi, à J11 les tailles de populations sont autour de 1 x 10^5 UFC/g pour CFBP 8149 et CFBP 7622, et en moyenne de 3.17 x 10^4 à 9.4 x 10^3 UFC/g pour CFBP 8130, CFBP 8153. Bien qu'inoculée avec une suspension concentrée à 1 x 10^6 UFC/ml, la souche CFBP 8142 a une taille de population constante au cours du temps mais nettement plus faible que celle de l'ensemble des souches inoculées dans cette étude (9.28 x 10^2 UFC/g contre 3.17 x 10^4 UFC/g en moyenne pour les autres souches). Seule la population de la souche CFBP 8150 décroit au cours du temps, en effet à J4, sa taille de population est d'environ 10^4 UFC/g, tandis qu'à J11, elle est de 10^3 UFC/g.

IV. Discussion

Les souches pathogènes et les souches commensales diffèrent évidemment par leur capacité à engendrer des symptômes sur plante hôte sensible, cependant elles échangent du matériel génétique ce qui implique une sympatrie sur au moins une partie de leur cycle biologique. Le but de cette étude est de mieux comprendre l'écologie des souches commensales en comparant des phénotypes liés aux capacités de colonisation des sites favorables au sein de la plante et en déterminant si ces deux types de souches peuvent coexister (absence d'antibiose). Les phénotypes étudiés sont la mobilité, l'adhésion et la formation de biofilm. De plus la multiplication des souches *in planta* a été analysée. Des tests d'antibiose entre souches pathogènes

et commensales ont été réalisés *in vitro*. Pour l'ensemble des phénotypes testés au cours de cette étude, nous n'avons pas mis en évidence de différence significative entre les souches pathogènes et commensales, sauf en ce qui concerne la capacité à former des pellicules qui est plus fréquemment observée chez les souches pathogènes que chez les souches commensales. Les souches commensales semblent donc capables d'exprimer les mêmes phénotypes que les souches pathogènes, phénotypes essentiels à la capacité de colonisation des parties aériennes des plantes.

La première étape de la colonisation de la feuille correspond à la recherche de sites favorables pour l'installation. Lors de cette étape, l'un des éléments qui intervient est la mobilité de la souche. Nos résultats indiquent que la majorité des souches (32/35) de la collection, qu'elles soient pathogènes ou non, est mobile. Cela indique qu'elles possèdent les structures génétiques permettant ce phénotype, c'est-à-dire le flagelle et/ou le pilus de type IV. Le chimiotactisme est le mécanisme amont de la mobilité permettant l'attraction vers les sites favorables pour la colonisation. Lors de cette étude, nous n'avons pas essayé de différencier les souches selon leur profil chimiotactique. En effet, la caractérisation *in vitro* des capacités chimiotactiques des souches est très lourde et difficilement envisageable sur une telle collection. Cependant, le profil chimiotactique de la souche commensale isolée de *Prunus*, CITA 44, a été comparé à ceux de deux souches pathogènes isolées de *Prunus* (Garita-Cambronero et al. 2016a). Ces profils sont variables entre les trois souches, en effet, sur 23 composés carbonés, la souche commensale présente le même profil pour 18 composés qu'une des deux souches pathogènes, et seulement pour 11 avec la deuxième souche. Les deux souches pathogènes, ne partagent que 14 composés entre elles. Leurs résultats semblent donc indiquer que les profils de chimiotactisme ne sont pas dépendants de la pathogénie des souches. Ceci est en accord avec nos analyses de répertoires de MCP. et l'inférence de gains et de pertes de ces gènes au cours de l'histoire évolutive de cette espèce qui ont montré qu'il n'y avait pas d'événements majeurs de pertes ou de gains liés à l'émergence des souches pathogènes. Cela peut donc laisser supposer que les souches commensales et pathogènes colonisent le même type de site sur leur plante hôte, et pourraient échanger du matériel génétique au niveau de ces sites. Cette étude est basée sur des amplifications PCR réalisées avec a priori, puisque nous avons recherché uniquement les MCP connus chez les *Xanthomonas*. La recherche de MCP sans a priori dans les génomes de *X. arboricola* est en cours. Elle se base sur la recherche du domaine conservé de ces gènes, le domaine MA, tel que réalisé dans la thèse d'Indiana (2014). Cela nous permettra donc d'établir les répertoires complets des MCP présents chez ces souches et d'identifier si les souches colonisant le même hôte ont des répertoires similaires. En effet, ces gènes, comme les effecteurs de types 3 sont impliqués dans la spécificité d'hôtes chez *X. axonopodis* (Mhedbi-Hajri et al. 2011). Aucune spécificité

d'hôte n'a pu être mise en évidence sur la base des répertoires d'ET3 chez les souches commensales (chapitre 1 ; chapitre 4), mais qu'en est-il des répertoires de MCP ? Cette analyse sans *a priori* nous permettra de répondre à cette question.

Nos résultats ont permis de mettre en évidence que trois souches de notre collection ne sont pas mobiles, et huit présentent exclusivement une mobilité par "twitching". Parmi les trois souches non mobiles, on retrouve la souche CFBP 2528 connue pour ce caractère et les souches CFBP 1159 et CFBP 3523, ces trois souches sont des souches pathogènes. Garita-Cambronero et al. 2016 ont montré que la souche commensale CITA 44 est mobile. L'absence de mobilité de la souche CFBP 1159 a déjà été rapportée, trois gènes du cluster flagellaire sur les 7 recherchés par PCR étaient absents chez cette souche (Darrasse et al. 2013). L'analyse génomique de la souche CFBP 2528 montre une délétion de 9 gènes contigus dans son cluster flagellaire (Cesbron et al. 2015). Ainsi, l'absence de mobilité de la souche CFBP 3523 et des 8 souches mobiles exclusivement par " twitching ", parmi lesquelles on retrouve 6 souches commensales, pourrait également être expliquée par une délétion dans le cluster flagellaire. La recherche de ces gènes dans les génomes nous permettra de valider ou non cette hypothèse mais également de déterminer la taille de la délétion dans le cluster flagellaire pour la souche CFBP 1159. L'absence de flagelle a déjà été mise en évidence chez un autre agent phytopathogène, *Xanthomonas fuscans* subsp. *fuscans*, responsable de la graisse commune du haricot (Darasse et al. 2013). En effet, la moitié des souches isolées au cours d'une épidémie provoquée par cet agent phytopathogène était non-mobile (Darrasse et al. 2013). L'hypothèse émise par les auteurs est qu'il s'agirait d'une stratégie populationnelle pour limiter la réaction de défense de l'hôte. En effet le flagelle possède un motif reconnu (Flg22) par un récepteur de l'hôte (FLS2) qui induit une cascade de réaction de défense. L'absence de flagelle dans une partie de la population responsable de cette épidémie n'engendrerait pas les réactions de défense de l'hôte proportionnée à la taille réelle de la population (Darrasse et al. 2013).

L'adhésion permet le positionnement stable des cellules bactériennes sur les sites favorables de l'environnement. Nos résultats montrent que l'ensemble des souches de la collection est capable d'adhérer. Les souches commensales, comme les souches pathogènes seraient potentiellement capables à cette étape de se diriger vers des sites favorables et d'y adhérer. On peut donc supposer que les deux types de souches ont dans leur génome les répertoires de gènes codant pour les structures d'adhésion (Pilus de Type IV, exopolysaccharides et adhésines) (Mhedbi-Hajri et al. 2011). Des différences dans le répertoire d'adhésines ont été observées chez les quatre souches de l'étude comparative de Cesbron et al. (2015) entre les souches pathogènes et commensales. Cependant, aucun lien entre contenu en gènes et observations phénotypiques n'a pu être mis en évidence. En utilisant une collection

de souches plus importante, comme celle que nous avons utilisée ici, il sera peut être possible d'établir ce lien. Contrairement à ce qui a été observé chez Cesbron et al. (2015), nos résultats montrent que la souche CFBP 7651 adhère fortement, alors qu'elle présentait les valeurs d'adhésion les plus faibles dans l'étude précédente. Nos résultats sont cependant corroborés par ceux obtenus lors d'un essai précédent dans lequel CFBP 7651 montrait les mêmes valeurs d'adhésion que CFBP 7179 (Biessy, 2013). Ce manque de cohérence entre les résultats des deux études pourrait s'expliquer par le fait que la coloration au cristal violet est une étape critique dans la manipulation, ce qui la rend particulièrement difficile à répéter. En effet, cette étape a pour but de colorer les bactéries ayant adhéré et de mesurer la DO de la solution obtenue. Cette mesure de DO peut être biaisée en cas de rinçage incomplet ou trop agressif, rendant par conséquent, des valeurs d'adhesion sous- ou sur-estimée.

Dans le cycle biologique d'une bactérie phytopathogène, l'adhésion est suivie de la formation de biofilm. Nos résultats montrent que les souches pathogènes et commensales sont capables de former des biofilms de taille quasiment identique. Cela semble donc indiquer que chacun des deux types est capable de se diriger vers des sites favorables, d'y adhérer et de s'agréger en biofilm. La capacité à former un biofilm est très importante lors de l'installation de la bactérie sur la plante. On pourrait supposer que les deux types de souches peuvent former, ensemble, des biofilms. Les résultats des tests d'antibiose montrent l'absence d'antagonisme entre ces deux populations ce qui rend possible leur cohabitation au sein d'un même biofilm. Les biofilms sont un lieu privilégié d'échange de matériel génétique (Danhorn & Fuqua, 2007), ils pourraient être le siège des flux géniques observés entre les populations pathogènes et commensales de *X. arboricola* (Chapitre 2). Une étude de compétition *in planta* a été réalisée sur haricot en utilisant différents couples de souches pathogènes du haricot et des souches commensales isolées de cette plante hôte (Montier, 2014). Pour l'ensemble des couples il a été montré que les deux souches coexistaient sur la plante et étaient capables de se transmettre à la plantule après germination des graines inoculées.

La dynamique des tailles de population sur l'hôte permet indiquer si la bactérie est capable de survivre sur cet hôte et son degré d'adaptation à cet hôte. Il a été montré que les populations d'agents pathogènes peuvent atteindre une taille de 10^8 UFC/g (unités formant colonies) en conditions expérimentales contrôlées, tandis que les bactéries de la flore indigène se maintiennent à des tailles de 10^4 UFC/g au bout de 11 jours (Darsonval et al. 2010). Concernant les souches mises en situation nulle, comme *Escherichia coli*, leur taille de population décroît au cours du temps pour ne pratiquement plus être détectables dès 11j post-inoculation. Nos résultats montrent des dynamiques de taille de populations des souches commensales isolées de haricot équivalentes à celles des souches indigènes. En effet, les tailles de populations se

maintiennent entre 4 et 11 jours après inoculation, après avoir diminué entre 1 et 4 jours après inoculation. Cette diminution est également observée pour les souches indigènes. En effet, entre J1 et J7, la taille de population passe de 10^5 UFC/g à 10^4 UFC/g (Darsonval et al. 2010). La souche CFBP 8150 fait figure d'exception car elle présente la même dynamique qu'une souche en situation nulle, c'est-à-dire inadaptée à l'environnement avec une taille de population en diminution depuis l'inoculation et une taille d'environ 10^3 UFC/g au bout de 11 jours, ce qui correspond à la taille de population obtenue dans les mêmes conditions pour *Escherichia coli*, inoculée sur haricot (Darsonval et al. 2010).

En conclusion, contrairement aux deux études phénotypiques précédentes (Cesbron et al. 2015 ; Garita-Cambronero et al. 2016a) dont le nombre de souches était de deux et une, respectivement, notre collection de souches commensales était beaucoup plus importante (25 souches). Alors que ces deux études pointaient des différences phénotypiques sur des caractères liés aux phases précoces de l'interaction avec la plante entre les souches commensales et les souches pathogènes, dans cette étude, nous n'observons pas de différence significative entre ces deux types de souches. L'avantage d'avoir une collection de souches plus complète est que l'on s'affranchit de la variabilité individuelle et que l'on obtient des résultats généraux. Les souches commensales ont une dynamique de taille de populations similaire à celle des souches indigènes du haricot. Elles ne diffèrent pas non plus des souches pathogènes pour les caractères de mobilité, adhésion et formation de biofilm. Ces souches semblent donc présenter les mêmes aptitudes à coloniser leur plante hôte dans les étapes précoces de l'interaction. De plus, les tests d'antibiose ont montré que les souches commensales et pathogènes pouvaient coexister. Il semble donc qu'elles pourraient coloniser les mêmes niches écologiques et se retrouver en sympatrie. Ceci pourrait expliquer le flux de gènes observés entre ces deux types de souches lors de la reconstruction de scénarios de divergence (chapitre 2). Des tests de localisation in planta avaient été entrepris pour cette étude, en étudiant les tailles de populations internalisées dans les tissus de la plante et les tailles de populations entières. Cependant, la manipulation du noyer en serre est assez difficile car ce sont des plantes fragiles. Nous avions donc réalisé ces tests en utilisant des feuilles en survie dans de l'eau gélosée. Ces feuilles n'ont pu être maintenues que 4 jours, et à l'issue de ces 4 jours, le nombre de feuilles analysables n'était pas suffisant pour obtenir des résultats robustes, qui ne sont donc pas présentés ici. Afin de déterminer la localisation épiphyte et/ou endophyte des souches pathogènes et commensales la mise au point de ce test devra être poursuivie. Le noyer avait été choisi car nous disposions d'une importante collection de souches épidémiques et commensales sur cet hôte. Face aux difficultés de manipulation de cette plante, une alternative pourrait être de choisir un autre hôte.

Données supplémentaires

Table IV-1 : Collection de souches

n° souche	pathovar	pathogenie	groupe génétique	hôte d'isolement	année d'isolement	phenotype *in vitro*	MCP	dynamique de population
CFBP7407	arracaciae	o	B	Arracacia xanthorrhiza	1969	x	x	
CFBP3523	celebensis	o	B	Musa acuminata	1960	x		
CFBP7150	celebensis	o	B	Musa acuminata	1960	x		
CFBP1159	corylina	o	A	Corylus maxima	1969	x		
CFBP1846	corylina	o	A	Corylus avellana	1975		x	
CFBP1848	corylina	o	A	Corylus avellana	1977		x	
CFBP2565	corylina	o	A	Corylus avellana	1985	x		
nccb100457	corylina	o	A	Corylus colurna	2010			
CFBP7408	guizotiae	o	B	Guizotia abyssinica	<1966	x	x	
CFBP7409	guizotiae	o	B	Guizotia abyssinica	<1966	x		
12710	juglandis	o	A	Juglans regia	2002	x		
CFBP2528	juglandis	o	A	Juglans regia	1956	x		
CFBP7179	juglandis	o	A	Juglans regia	2002	x		
CFBP8350	juglandis	o	A	NA	NA	x		
bc2528	NA	o	B	piment	2007	x		
CFBP1022	NA	n	B	Juglans regia	1967	x	x	
CFBP6762	NA	n	B	fragaria X ananassa	NA	x		
CFBP6771	NA	n	B	fragaria X ananassa	1993	x	x	
CFBP6827	NA	o	B	piment	2000	x	x	
CFBP7604	NA	n	B	capsicum cv. Japaleno	NA	x		
CFBP7609	NA	n	B	Anthurium andreanum	NA		x	
CFBP7610	NA	n	B	pueraria lobata	1981	x	x	
CFBP7614	NA	n	B	Magnolia sp.	NA	x	x	
CFBP7615	NA	n	B	Camellia sinensis	1984		x	
CFBP7618	NA	n	B	C.coronarium	1985		x	
CFBP7622	NA	n	C	Phaseolus vulgaris	1985	x	x	x

CFBP7629	NA	n	B	Juglans regia	2009	x		
CFBP7630	NA	n	B	Juglans regia	2009		x	
CFBP7634	NA	n	B	Juglans regia	2009	x	x	
CFBP7640	NA	n	C	Juglans regia	2009		x	
CFBP7645	NA	n	C	Juglans regia	2009	x	x	
CFBP7651	NA	n	B	Juglans regia	2008	x	x	
CFBP7652	NA	n	B	Juglans regia	2008	x	x	
CFBP7653	NA	n	C	Juglans regia	2008	x	x	
CFBP7697	NA	n	B	Phaseolus vulgaris	2011	x		
CFBP8130	NA	n	B	Phaseolus vulgaris	2010	x	x	x
CFBP8131	NA	n	B	Phaseolus vulgaris	2010		x	
CFBP8132	NA	n	B	Phaseolus vulgaris	2010	x	x	
CFBP8133	NA	n	B	Phaseolus vulgaris	2010		x	
CFBP8134	NA	n	B	Phaseolus vulgaris	2010		x	
CFBP8136	NA	n	B	Phaseolus vulgaris	2011		x	
CFBP8138	NA	n	B	Phaseolus vulgaris	2010	x		
CFBP8139	NA	n	B	Phaseolus vulgaris	2010		x	
CFBP8140	NA	n	B	Phaseolus vulgaris	2010		x	
CFBP8141	NA	n	B	Phaseolus vulgaris	2010		x	
CFBP8142	NA	n	B	Phaseolus vulgaris	2010	x	x	x
CFBP8144	NA	n	B	Phaseolus vulgaris	2010		x	
CFBP8145	NA	n	B	Phaseolus vulgaris	2010		x	
CFBP8147	NA	n	B	Phaseolus vulgaris	2010	x	x	
CFBP8149	NA	n	B	Phaseolus vulgaris	2010	x		x
CFBP8150	NA	n	B	Phaseolus vulgaris	2010	x		x
CFBP8152	NA	n	C	Phaseolus vulgaris	2010	x	x	
CFBP8153	NA	n	B	Phaseolus vulgaris	2010	x	x	x
CFBP3122	populi	o	C	Salix alba	1980	x		
CFBP3123	populi	o	C	Populus X canadensis	1979	x	x	
CFBP3894	pruni	o	A	Prunus salicina	1953	x		

CFBP5530	pruni	o	A	Prunus persica	1989		x
maff301420	pruni	o	A	NA	1978		
maff301427	pruni	o	A	NA	1980		
maff311562	pruni	o	A	NA	2008		
xap33	pruni	o	A	Prunus amygdalus cv. Guara	NA		
CFBP7410	zantedeschiae	o	A	Zantedeschiae aethiopica	1967	x	x

Chapitre IV

Introduction

Les résultats des chapitres 1 et 3 ont montré que les seuls gènes impliqués dans les traits d'histoire de vie qui présentaient des distributions différentes entre les souches commensales et pathogènes étaient les gènes du système de sécrétion de type 3 (SST3) et les effecteurs. L'objectif de ce chapitre[14] était donc de déterminer l'histoire évolutive du SST3 au sein du genre *Xanthomonas*. En effet, la présence de souches sans SST3, au sein de ce genre, laisse supposer qu'il y a eu des événements d'acquisition ou de perte de ce système au cours de l'histoire évolutive du genre. En étudiant son histoire évolutive, nous cherchons à déterminer s'il a été fréquemment transmis par transfert horizontal de gènes ou au contraire s'il a été transmis verticalement. Pour cela, nous avons étudié ce cluster de gènes chez l'ensemble des espèces du genre, en utilisant 82 séquences génomiques. Tout d'abord, une phylogénie robuste du genre basée sur le core génome a été réalisée, afin de déterminer à quel nœud ont eu lieu les différents événements identifiés. Ensuite, nous avons analysé les environnements génomiques dans lequel est situé le cluster de gènes du SST3. De manière générale, le cluster de gènes du SST3 est retrouvé dans un environnement génomique relativement conservé au sein du genre, suggérant une acquisition ancestrale, suivie soit de réarrangements génomiques dans les régions bordantes du cluster, soit de la perte du cluster en entier. Enfin, nous avons mis en évidence que la recombinaison homologue a lieu au sein des gènes du SST3. Ainsi, la position génomique du cluster de gènes du SST3 est ancestrale, cependant, l'évolution du cluster se fait par recombinaison homologue. L'évolution des répertoires en effecteurs de type 3 a été étudiée au sein des 44 génomes de l'espèce *X. arboricola*. Les résultats sont en accord avec ceux obtenus dans le chapitre 1. Un nombre de 10 ET3 est commun à l'ensemble des

[14]Martial Briand m'a guidée pour l'utilisation des outils bio-informatiques. Eran Bosis a réalisé les recherches des effecteurs de type 3 grâce à l'approche de machine learning. Céline Rousseau a rédigé le script permettant d'automatiser l'analyse des environnements génomiques. Perrine Portier a participé au choix des souches du genre *Xanthomonas* à séquencer. Mathieu Barret a réalisé le séquençage des souches en MiSeq, et a aidé à la rédaction du manuscrit. Enfin, Marie-Agnès Jacques et Marion Fischer-Le Saux m'ont encadrée pour le choix des souches, les questions biologiques, les analyses à réaliser, l'interprétation des résultats et pour la rédaction du manuscrit. J'ai réalisé l'ensemble des analyses phylogénétiques et des environnements génomiques, ainsi que les analyses statistiques concernant les effecteurs de type 3, j'ai également rédigé le manuscrit.

souches formant les clones épidémiques. Les environnements génomiques de ces ET3 montrent que l'ancêtre commun aux clones épidémiques auraient acquis ces ET3 de manière indépendante. Ensuite, chaque clone épidémique a acquis des ET3. Concernant les souches commensales, nous avons mis en évidence que certaines souches ont conservé le répertoire ancestral de l'ancêtre commun de l'espèce *X. arboricola* composé de 7 ET3, dont 5 sont inclus dans le cluster de gènes du SST3. Certaines souches commensales, quant à elles, ont perdu ce cluster de gènes du SST3 et les ET3 du répertoire ancestral au cours de l'évolution. Cet article a été soumis à Molecular Biology and Evolution.

Evolution of the type three secretion system and type three effector repertoires in *Xanthomonas*

Merda D., Briand M., Bosis E., Rousseau C., Portier P., Barret M., Jacques $MA^{\$}$., Fischer-Le Saux $M^{\$}$.

$ These authors contributed equally to this work.

Abstract

Deciphering evolution and transmission patterns of virulence determinants is necessary to understand pathogen emergence. *Xanthomonas* are bacteria responsible for numerous epidemics in agroecosystems worldwide and represent a major threat to plant health. The main virulence factor of *Xanthomonas* is the type three secretion system (T3SS) of the Hrp2 family. Since some *Xanthomonas* strains are deprived of T3SS, we wondered if the distribution of T3SS in this bacterial genus resulted from losses or independent acquisitions. Based on comparative genomics of 82 genomes representing the diversity of the genus, we inferred three ancestral acquisitions of the Hrp2 cluster during *Xanthomonas* evolution. These ancestral acquisitions have been followed by subsequent losses in commensal strains and reacquisition in some species. While mutation is the main driving force that brings polymorphism at the gene level, inter-species homologous recombination of large fragments expanding through several genes shapes Hrp2 cluster polymorphism, especially within *X. arboricola* species. Horizontal gene transfer of the entire Hrp2 cluster also occurred. Using machine learning method, we showed a reduced core effectome in commensal strains composed of *xopF1*, *xopM*, *avrBs2* and *xopR* and

stepwise accumulation of numerous type 3 effector genes in successful pathogens. Our data suggest that capacity to intimately interact with plants through T3SS would be an ancestral trait of xanthomonads as it is the case for other major plant pathogenic clades. Since then T3SS have experienced a highly dynamic evolutionary history characterized by an intense gene flux between species that may reflect its role in pathogen adaptability and evolvability.

I. Introduction

Unraveling evolution and transmission of virulence factors is crucial to understanding how pathogens emerge. In Gram negative bacteria, a major virulence factor is the set of type three effectors (T3Es) secreted directly in host cells by the type three secretion system (T3SS). This system is a complex protein structure anchored in the bacterial membrane whose role is to inject effector proteins into host cells, such as animal, plant or human cells (Diepold & Armitage 2015). The T3Es interfere with a variety of host functions and in pathogenic interactions they either trigger disease (via neutralization of plant defenses) or host resistance (through recognition of microbial-associated molecular patterns inducing hypersensitive response). Some specific T3Es called transcription activator like effectors (TALEs), first identified in *Xanthomonas* genus, even manipulate host gene expression to promote bacterial multiplication (Boch et al. 2009). The T3SS is also widespread in mutualistic and commensal bacteria (Preston et al. 2007, Coombes 2009, Barret et al. 2013, Lackner et al. 2011, Abby & Rocha, 2012) where it is involved in interactions with protist, fungi, animal or plant host cells. Given the central role of T3SS and T3Es in prokaryote-eukaryote interactions, it is crucial to unveil their origin and evolution in order to understand microbial ecology in regards to host adaptation.

The origin and evolution of T3SS has been extensively studied, and it was shown that it has a common evolutionary history with the flagellar cluster, which is required for swimming motility, and with which it shares ten homologous genes (McCann & Guttman 2008). However there is still a debate whether T3SS derives from the flagellar cluster or if the two systems have a common ancestor and have undergone parallel evolution. Abby & Rocha (2012) recently proposed that T3SS evolved by exaptation from the flagellar cluster. T3SS diversification then occurred and seven families of T3SS capable of translocating effectors in eukaryotic cells are recognized in Gram-negative bacteria (Ysc, SPI-1, SPI-2, Chlamy, Hrp1, Hrp2 and Rhizo) (Troisfontaines & Cornelis 2005). Diversification of T3SS does not follow vertical bacterial evolution. Indeed, it was shown that the T3SS has been submitted to numerous horizontal gene transfer (HGT) (Gophna et al. 2003; Troisfontaines

& Cornelis 2005; Naum et al. 2009; Kirzinger et al. 2015). These HGT events are facilitated by T3SS localization in plasmidic or genomic pathogenicity islands (Groisman & Ochman 1996; Hacker et al. 1997; Oh et al. 2005; Toth et al. 2006).

Extensive diversification of T3SS seems to be driven by bacteria ecology, as T3SS families correlate with the host type (e.g. animal, plant, fungi, amoeba). Indeed, Rhizo, Hrp1 and Hrp2 seem to be more frequently involved in interactions with plants whereas Ysc, Chlamy, SPI-1 and SPI-2 appear to be specific to animal pathogens, insects and amoeba (Abby & Rocha 2012). However, some exceptions to these rules were found; for instance a Hrp2 T3SS was also described in an endofungal bacterium (Lackner et al. 2011). In general, each bacterial genus harbors a T3SS from a specific family (Diepold & Armitage 2015; Troisfontaines & Cornelis 2005) but some bacteria with complex lifestyles harbor several T3SS from different families (Zhao et al. 2009; Barret et al. 2013, Abby et al. 2016). These contrasting patterns of T3SS content could be explained by ancient vertical inheritance mixed with T3SS gain and loss events (Kirzinger et al 2015). The main T3SS families found in phytopathogens are Hrp1 (in *Pseudomonas* and *Erwinia*) and Hrp2 (in *Xanthomonas*, *Ralstonia*, *Acidovorax*, and *Burkholderia*) however alternative non-canonical T3SS were also described in some plant-associated bacteria such as commensal pseudomonads associated with plants (Barret et al. 2013; Mavrodi et al. 2011; Preston et al. 2001). These two main families of T3SS clusters, Hrp1 and Hrp2, differ in gene content, synteny and transcriptional regulation. Approximately 20 protein-coding genes, called *hrp* (hypersensitive reaction and pathogenicity) and *hpa* (*hrp*-associated) genes, are involved in the biogenesis of T3SS. Among them, 9 genes, which were renamed *hrc* (*hrp* conserved) genes, are highly conserved in plant and animal pathogens and 8 have homologs in the flagellar cluster (Tampakaki et al. 2010).

T3E repertoires are highly diverse within a given genus and even within a bacterial species (McCann & Guttman 2008). They vary both in terms of content and size with, for instance, 6 to 26 T3E genes found in *X. axonopodis strains* (Hajri et al. 2009). It has been speculated that this large variability could be the consequence of host adaptation process. Indeed, in *Pseudomonas* and *Xanthomonas*, whose pathogenic strains are highly host specific, T3E repertoire composition is correlated with host range (Sakar et al. 2006; Hajri et al. 2009; Lindeberg et al. 2009; Hajri et al. 2012; Schwartz et al. 2015). The plasticity of T3E repertoires within a species could be explained by frequent HGTs (McCann & Guttman 2008) and indeed many T3E genes were found associated with mobile genetic elements. As T3Es are modular proteins, acquisition of new T3E genes can also result from shuffling process, called "terminal reassortment", analogous to exon shuffling in eukaryotes (Stavrinides et al. 2006). Determining diversity and evolution of T3E

repertoires in pathogenic bacteria is essential to get insights into host adaptation mechanisms. However, identification of T3E genes in whole genome sequences is still a challenge as T3Es are structurally and functionally highly diverse with more than 50 families identified so far in *Xanthomonas* and Pseudomonas (Ryan et al. 2011; Lindeberg et al. 2012). Recently, machine learning approaches were developed. They rely on multiple criteria such as the presence of a secretion signal necessary for recognition by T3SS machinery that is found in N-terminal region of T3E (McDermott et al. 2011) or specific amino-acid composition (Lower & Schneider 2009). In *Xanthomonas*, these approaches allowed identification of seven novel T3Es in the reference strain 85-10 (Teper et al. 2016), and thus represent great promise to decipher T3E repertoires in exponentially growing genomic data.

Xanthomonas are major plant pathogens devastating crops worldwide. Their main pathogenicity determinants are the Hrp2 type-T3SS and its effectors (White et al. 2009). *X. albilineans* is an exception in the genus as it encompasses no Hrp2 family gene cluster, but instead a SPI-1 T3SS (Marguerettaz et al. 2011). Four *Xanthomonas* species lacking any Hrp-T3SSs and associated T3Es, namely *X. sacchari* (Studholme et al. 2011), "*X. cannabis*" (Jacobs et al. 2015), "*X. pseudoalbilineans*" (Pieretti et al. 2015) and *X. maliensis* (Triplett et al. 2015) were recently described. Moreover, some strains of the *X. arboricola* species were also found without any T3SS. These *X. arboricola* strains lacking any T3SS are considered as commensal since no pathogenicity on their respective hosts was observed (Essakhi et al. 2015; Cesbron et al. 2015; Garita-Cambronero et al. 2016 ; chapter 1). The *X. arboricola* species presents an epidemic population structure as defined by Maynard-Smith et al. (1993), where epidemic clones are represented by successful pathovars attacking stone and nut fruit trees worldwide and the recombinant network is represented by unsuccessful pathovars and commensal strains (chapter 1). Epidemic clone emergence seems correlated with acquisition of T3Es whereas in recombinant network, strains would have lost T3E coding genes and T3SS cluster (chapter 1).

Recent evidences of several *Xanthomonas* species and strains lacking the Hrp2 cluster pose questions about the evolution of virulence and T3SS origin in this genus. Are pathogenicity and Hrp2 cluster ancestral features of *Xanthomonas* that have been vertically inherited and lost in some species or do they represent more recent acquisitions? Contrasting with our deep knowledge of ancient evolutionary history of T3SS, few is known about recent origin and evolution of T3SS and its role in plant pathogen emergence. Given the pivotal role of T3SS and T3Es in xanthomonads pathogenicity and host specificity and given their heterogeneous distribution at genus and species scales, the Hrp2 cluster and its effectors in *Xanthomonas* genus appear to be a good model to study T3SS origin and evolution at a fine evolutionary scale. Recently, Jacobs et al. (2015) addressed this research topic focusing on

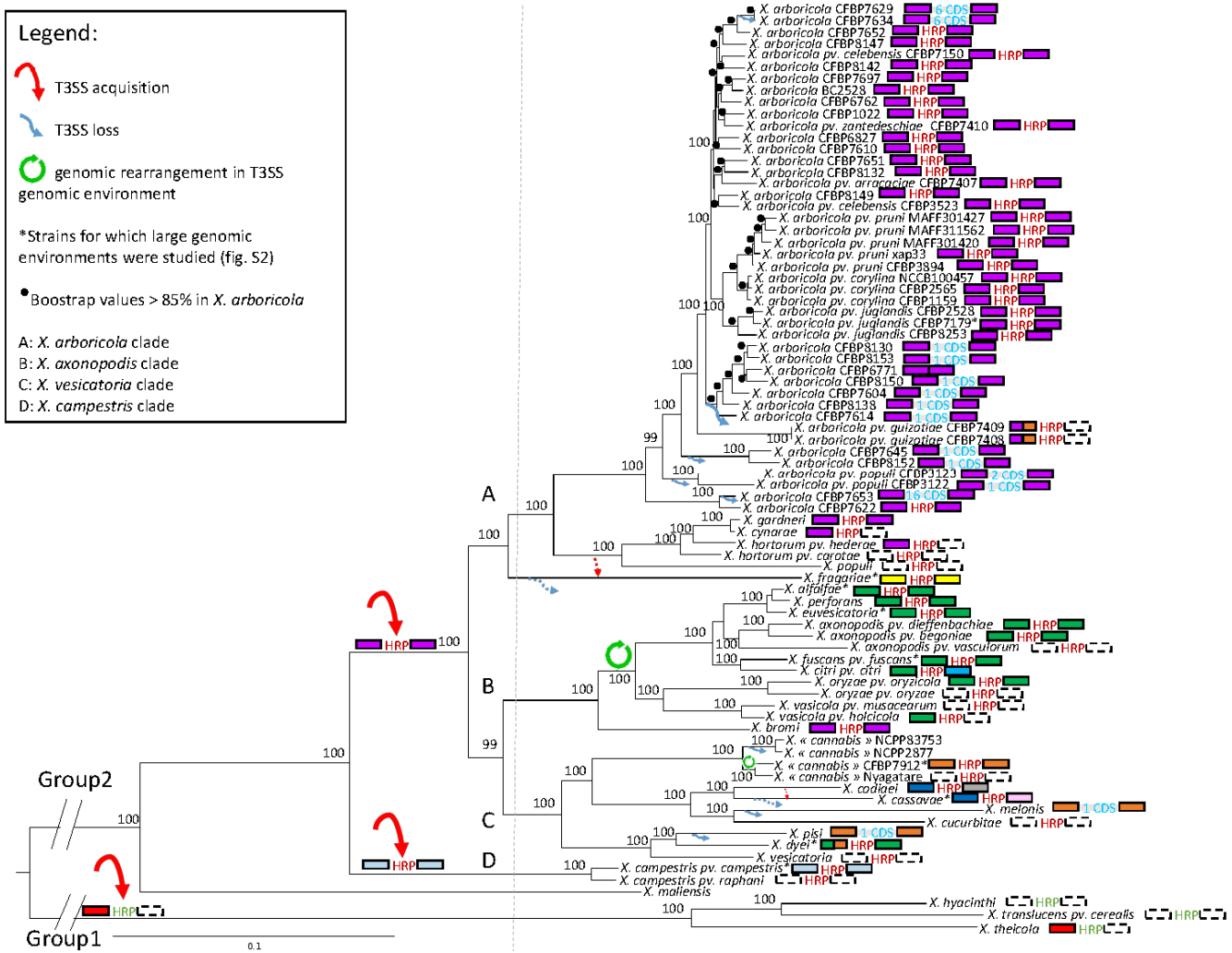


Figure V-1: Maximum likelihood phylogeny based on the concatenated sequences of the core proteome (993 proteins) of 80 strains representing the entire *Xanthomonas* genus and schematic representation of T3SS genomic environments. The T3SS cluster is represented by the letters HRP and its genomic environments (20 kb on each side) by colored rectangles; different colors correspond to different genomic environments (according to Supplementary fig. S1, Supplementary Material online). Dotted line represents absence of information due to contig interruption. In Hrp-negative strains, HRP letters are replaced by the number of CDS found in place of T3SS cluster at the putative T3SS cluster insertion site. Genomic environments of the insertion site are represented as described above. * indicates the strain for which large genomic environments (200 kb) were studied (see supplementary fig. S2, Supplementary Material online). Probable T3SS acquisition events are represented by red arrow, loss events by blue arrow, and genomic rearrangements by green circled arrow. A dotted arrow represents hypothesis of T3SS loss and re-acquisition. Bootstrap scores (100 bootstraps) higher than 85% are displayed at each node.

"*X. cannabis*" and related strains. In this study, we conducted our analyses on a strain collection representing all valid species in the genus, inferring their phylogenetic relatedness based on the whole genus core genome. Moreover, to get insights into T3SS evolution, we studied not only cluster synteny and *hrc* gene phylogeny but we also considered the genomic environment of the T3SS cluster. Finally, to understand mechanisms responsible for variability in T3E repertoires, we determined the T3E repertoires in a collection of 44 *X. arboricola* genomes representing both commensal and pathogenic strains using a machine learning approach designed to detect T3E coding genes in *Xanthomonas* genome sequences.

II. Results

Organism- and T3SS-evolutionary histories in *Xanthomonas*

The presence of T3SS clusters was investigated in 82 genome sequences of *Xanthomonas* strains (Table V-S1) through BLASTp searches of 295 proteins representing the diversity of T3SS families (Table V-S2). Hrp2 cluster was detected in 61 genome sequences, SPI-1 was associated to the genome sequence of *X. albilineans*, and 20 genome sequences were free of any T3SS encoding genes. Most strains devoid of any T3SS belonged to *X. arboricola* (Table V-S1). Among these 14 strains, 12 were described as commensal strains and two belonged to the pathovar populi. T3SS cluster was also missing in *X. pisi* and *X. melonis* genomes, and as previously shown in some "*X. cannabis*" strains, *X. maliensis* and the group 1 species *X. sacchari*.

A determinant part of further analyses was the reconstruction of a robust phylogenetic tree of the *Xanthomonas* genus in order to infer evolutionary scenarios of T3SS gains and losses. Phylogenetic relationships among *Xanthomonas* species were reconstructed using their predicted core proteomes (fig. V-1). Group 2 species were the main focus of this study and thus *X. translucens*, *X. hyacinthi* and *X. theicola* belonging to group 1 were used to root the tree. According to this phylogenetic reconstruction whose branches were supported by very high bootstrap values, *X. maliensis* and *X. campestris* diverged very early from other species in group 2. This isolated phylogenetic position was unexpected for *X. campestris* as previous MLSA studies on whole genus diversity placed *X. campestris* in the core of group 2 (Young et al. 2008, Triplett et al. 2015). The isolated position of *X. campestris* was previously shown in genome-based phylogeny (Rodriguez et al. 2012; Naushad et al. 2013), but as their trees included only a limited number of species, the isolated position of *X. campestris* from all group 2 species could not be suspected. Three major clades supported by 100% bootstrap values could be distinguished among all other species in group 2: (i) "*X. arboricola*" clade encompassing *X. arboricola*,

X. gardneri, *X. cynarae*, *X. hortorum* and *X. populi*, (ii) "*X. axonopodis*" clade encompassing species of the *X. axonopodis* complex (*X. axonopodis*, *X. alfalfae*, *X. perforans*, *X. euvesicatoria*, *X. fuscans*, and *X citri*), *X. oryzae*, *X. vasicola* and *X. bromi*, and (iii) "*X. vesicatoria*" clade encompassing "*X. cannabis*", *X. codiaei*, *X. cassavae*, *X. melonis*, *X. curcurbitae*, *X. pisi*, *X. dyei* and *X. vesicatoria* (fig. V-1). Hrp2-negative strains were interspersed in the phylogenetic tree of the genus; a distribution pattern that could be explained by ancestral acquisition and subsequent losses or by recent independent Hrp2 acquisitions.

Analysis of the genomic environments of the T3SS clusters revealed three ancestral acquisitions of this cluster. One of these acquisitions would have occurred in the ancestor of the three group 2 clades ("*X. arboricola*", "*X. axonopodis*", and "*X. vesicatoria*"). The same genomic environments (20 kb flanking regions) were highlighted for Hrp2 clusters belonging to "*X. arboricola*" clade and *X. bromi*. Given the relative divergence between *X. arboricola* and *X. bromi*, it is tempting to speculate that Hrp2 cluster was acquired through one acquisition event in their common ancestor which corresponds to the node after *X. campestris* divergence. For other group 2 species a variety of T3SS flanking regions that were mainly clade-specifics were observed (fig. V-1, fig. V-S1). Two scenarios could be put forth to explain the different flanking regions of the Hrp2 cluster in the various group 2 clades: the first scenario involves the loss of the ancestral Hrp2 cluster and re-acquisition at a different location, and the second scenario includes rearrangements in the flanking regions of the ancestral Hrp2 cluster. To decipher which scenario was the most probable, broader genomic environments were compared in strains selected for the quality of their genome assembly and representing the different clades. Using a window of 200 kb upstream and downstream of the T3SS cluster, we showed that synteny with the T3SS genomic environment of *X. bromi* and "*X. arboricola*" clade could be retrieved in strains from "*X. axonopodis*" and "*X. vesicatoria*" clades. This observation supports the second scenario (fig. V-S2). In contrast, synteny could not be detected using 20 kb windows because of the insertion of large fragments in direct T3SS flanking regions. For instance, large T3SS genomic environments of *X. euvesicatoria*, *X. alfalfae* and *X. fuscans* presented rearrangements in comparison to the ones in "*X. arboricola*" clade with an 80 kb fragment inserted upstream to the T3SS cluster in the formers and downstream in the latter. A similar observation was made for "*X. cannabis*" strain CFBP 7912, for which synteny with *X. arboricola* T3SS flanking regions was retrieved 50 kb upstream and downstream of the T3SS cluster. Thus, genomic rearrangement events and gene insertions affected the direct flanking regions of the T3SS cluster but not its location in the genome; these rearrangements were different from one another in the different clades and were sometimes supported by the presence of insertion sequences (ISs) and tRNAs (fig.

V-S2). Altogether, these results supported the scenario of an ancestral acquisition of the T3SS cluster in the common ancestor of the three group 2 clades and subsequent rearrangements in flanking regions of the Hrp2 cluster.

The second acquisition would have occurred in *X. campestris* ancestor. In this clade no synteny could be retrieved in the genomic environment of the T3SS clusters even when flanking regions as broad as 200 kb were compared with those of other *Xanthomonas* species (fig. V-S2). While the genetic organization of the T3SS cluster is similar within all group 2 strains, the T3SS encoding genes experienced homologous recombination (see below) that shuffled the phylogenetic signal of *hrp* genes. Thus, we could not reinforce the hypothesis of ancestral and independent acquisition of T3SS by *X. campestris* with T3SS gene divergence from other group 2 T3SS genes. We favored this parsimonious scenario in contrast to the one, which consisted in an acquisition by the common ancestor of all group 2 strains followed by the loss of this T3SS cluster in *X. campestris* and its re-acquisition in a different genomic context.

The third acquisition of T3SS cluster would have occurred in group 1. The genetic organization of T3SS clusters in group 1 species was different from the one of group 2 species with the genes encoding the master regulators HrpX and HrpG located within the T3SS cluster. Moreover, all T3SS encoding genes are divergent from those of group 2 strains, except for *hrcC* from *X. translucens* pv. *cerealis* which is nearly identical to those of *X. cynarae* and *X. gardneri* (Fig. V-2B and fig. V-S2). Due to the limited size of the contigs, T3SS flanking regions of group 1 species could not be analyzed in detail (fig. V-1). The only flanking region that could be analyzed (in *X. theicola*) differed greatly from all group 2 T3SS flanking regions. Altogether, this suggests an independent acquisition event of T3SS in the common ancestor of *X. translucens*, *X. hyacinthi* and *X. theicola*.

Ancestral acquisitions of T3SS clusters implied that the absence of Hrp2 clusters in the 19 strains of *Xanthomonas* group 2 was the result of multiple loss events. The *X. arboricola* strains lacking T3SS were dispersed in five monophyletic lineages within a cluster encompassing all *X. arboricola* strains indicating at least five loss events (fig. V-1) during diversification of this species. One loss event could have occurred in the common ancestor of the two strains belonging to "*X. cannabis*" and two independent loss events could have occurred in *X. melonis* and *X. pisi*. As *X. maliensis* was the most divergent species of group 2, it was impossible to hypothesize event responsible for absence of T3SS cluster, and it might be that this species never had any T3SS cluster. In these 19 strains, the entire T3SS cluster was missing as no single gene could be detected. No traces of pseudogenization, which could be identified by a weak homology with T3SS coding genes with a blast approach, were detected. Blast search of the flanking regions of Hrp2 cluster in strains

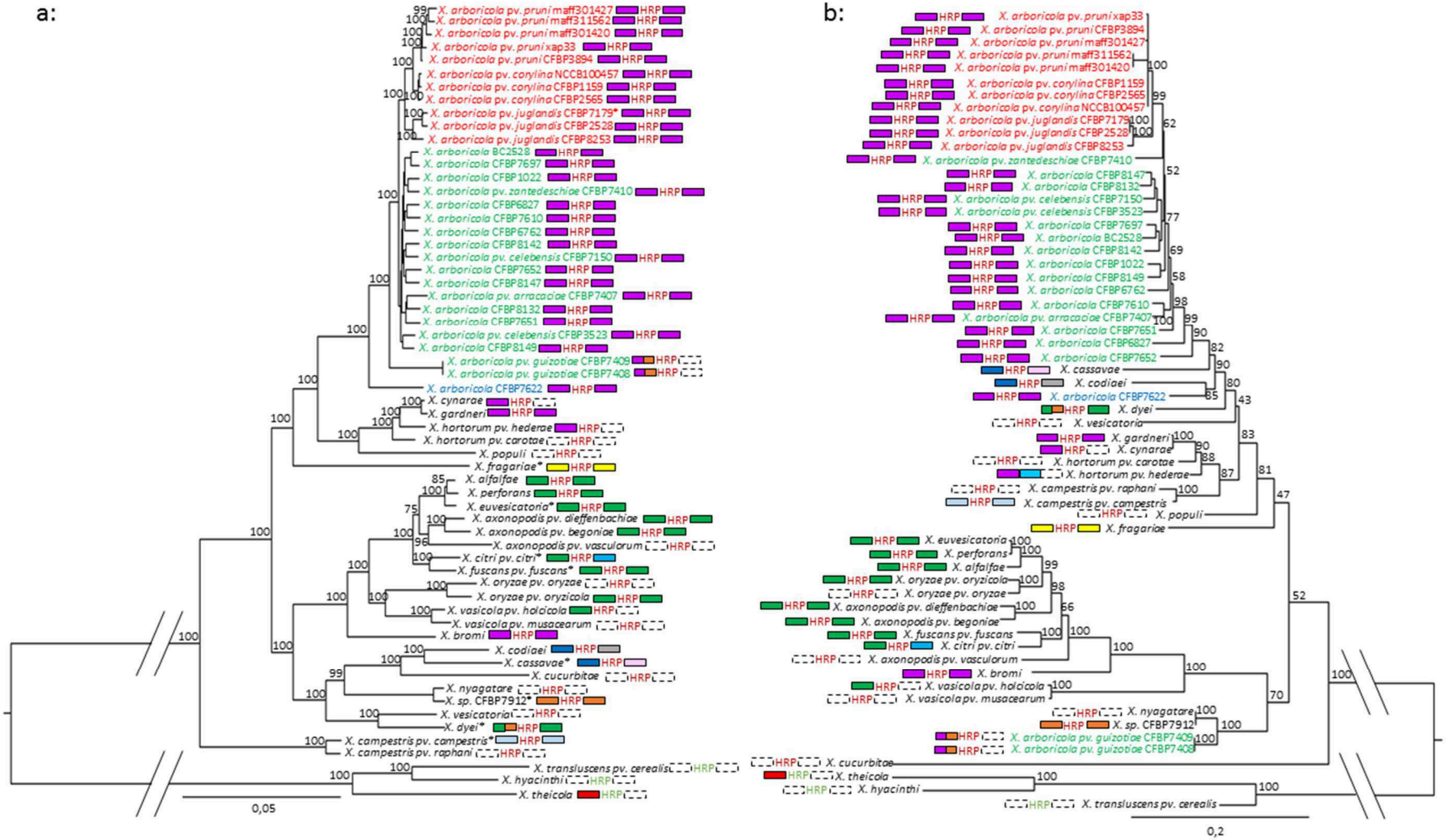


Figure V-2: Comparison of organism and T3SS phylogenies and schematic representation of T3SS genomic environments. These two phylogenies were constructed in maximum likelihood. A, the organism phylogeny is based on the concatenated sequences of the core proteome (1135 proteins) of 61 *Xanthomonas* spp. strains harboring a T3SS cluster. B, The T3SS phylogeny is based on the concatenated sequences of 10 hrc genes. For strains belonging to *X. arboricola* a color code was used to represent the three groups defined in Merda et al. (2016). Group A strains are indicated in red, group B strains in green, and group C strains in blue. Strains of all other species of *Xanthomonas* are indicated in black. Bootstrap scores (100 bootstraps) higher than 85% are displayed at each node. T3SS clusters and their genomic environments (20 kb on each side of T3SS cluster) are represented as explained in the legend of fig. V-1.

lacking it and synteny analysis allowed us to retrieve the probable excision sites of the DNA fragment containing the Hrp2 encoding genes (fig. V-S1). Excision would have occurred between trpG and ltaE. Nevertheless, we did not find any mobile genetic elements between these loci.

A loss followed by a re-acquisition of the whole T3SS cluster should have happened in *X. fragariae* and *X. cassavae* evolutionary history. Despite their phylogenetic positions in *Xanthomonas* group 2 a lack of synteny in T3SS flanking regions was observed among *X. fragariae*, *X. cassavae* and other group 2 *Xanthomonas* species, even when genomic environments as large as 200 kb were studied (fig. V-S1, fig. V-S2). Another element in favor of this hypothesis is the presence of four ISs found upstream and downstream of the T3SS cluster in *X. fragariae genome*. No remnants of the ancestral T3SS cluster were detected in the genomes of *X. fragariae* and *X. cassavae* at the expected location of the T3SS cluster. Because these species were localized in two different phylogenetic positions within group 2, this suggested independent re-acquisition events for these two species (fig. V-1). *X. codiaei* shared the same left border of T3SS cluster than *X. cassavae*, its sister species, suggesting that the loss and re-acquisition of T3SS cluster might have occurred in their common private ancestor, but unfortunately contig interruption in *X. codiaei* precluded the analysis of a large genomic environment to confirm this hypothesis.

In conclusion, at least three independent acquisition events of the Hrp2 cluster were evidenced in *Xanthomonas* evolution, as well as subsequent whole cluster losses and re-acquisitions, and also rearrangements in T3SS flanking regions (fig. V-1). One event occurred in group 1 ancestor, another event in *X. campestris* ancestor and the last in the common ancestor of all other species. Five species within group 2 (*X. pisi*, *X. melonis*, *X. cassavae*, *"X. cannabis"* and *X. fragariae*) would have lost the ancestral T3SS cluster. Among them, *X. fragariae* and *X. cassavae* would have re-acquired independently a T3SS cluster at different genomic locations. T3SS loss events were also evidenced in five lineages within *X. arboricola* but the majority of strains belonging to this species retained their ancestral T3SS cluster. Finally, genomic rearrangements occurred in T3SS genomic environments, as those evidenced in *"X. axonopodis"* clade and in *"X. cannabis"* strain CFBP 7912.

Homologous recombination in the T3SS coding genes

To determine if the T3SS cluster follows the same evolutionary history as the species, phylogeny based on concatenated *hrc* coding genes was compared to the organism phylogeny based on the core proteome of Hrp2-positive strains (fig. V-2A and B). The two topologies were different, as attested by the SH test (p-value = 0.00092). Indeed, numerous incongruences were observed. For instance, while

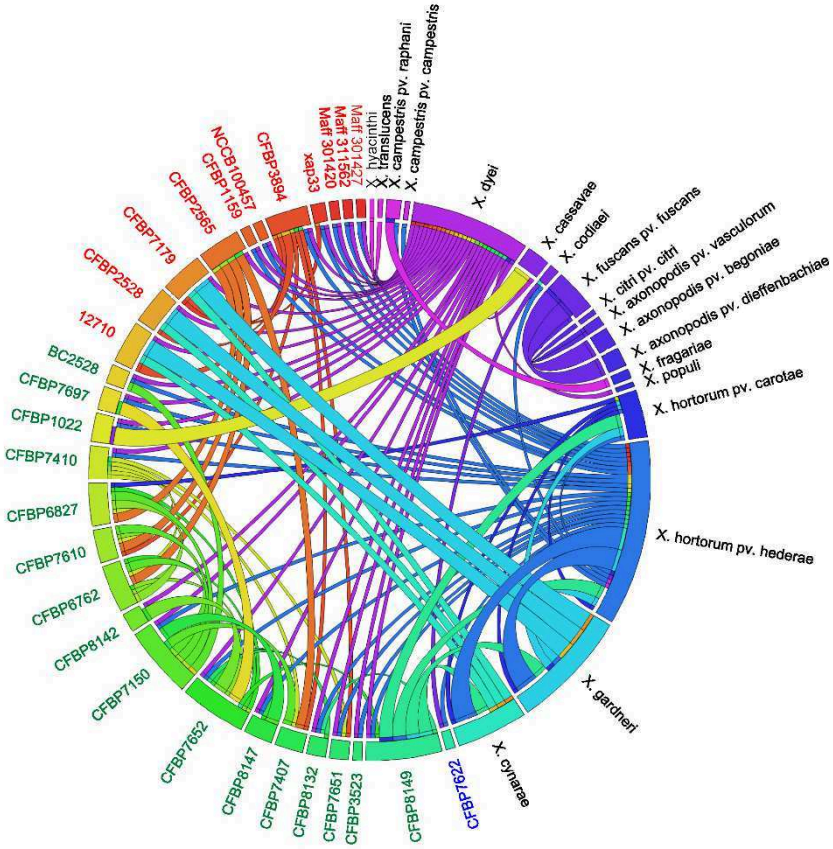


Figure V-3: Representation of recombination events affecting 16 *hrc/hrp* genes of the T3SS cluster in *Xanthomonas* genus. The donor and recipient strains were identified with RDP software, and the representation was obtained using Circos. Each strain is represented by a rectangle of a different color. The recombination events are represented by a link between donor and recipient strains. The color of the link corresponds to the color of the donor and indicates the direction of recombination event. The width of the links represents the number of genes (one or two) involved in the recombination events. For strains belonging to *X. arboricola* a color code was used to represent the three groups defined in Merda et al., 2016. Group A strains are indicated in red, group B strains in green, and group C strains in blue.

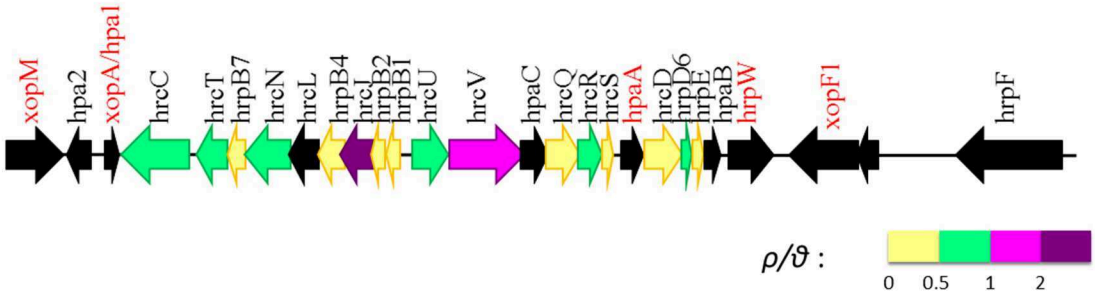


Figure V-4: Representation of ratios of recombination rate vs mutation rate (ρ/θ) along the T3SS cluster using 61 genomes representing the genus diversity. Ratios were calculated for the 16 *hrc/hrp* genes of the core region of T3SS cluster using RDP sotfware. Arrows are colored according to the color scale which indicates the range of the ρ/θ value. The black arrows represent *hpa* and T3E genes for which ρ/θ was not calculated. Genetic organization of the cluster is based on the sequence of CFBP 2528. In red are represented the five core T3E genes located in the T3SS cluster of X. arboricola strains.

most *X. arboricola* strains exhibited a monophyly for the T3SS coding genes, T3SS coding genes from the *X. arboricola* pv. *guizotiae* diverged from those of other *X. arboricola* strains (fig. V-2B). These sequences were closely related to those of *"X. cannabis"* strains CFBP 7912 and Nyagatare in *"X. vesicatoria"* clade. Similar incongruences were observed for *X. campestris*, *X. cassavae*, *X. codiaei*, *X. dyei* and *X. vesicatoria* whose T3SS coding genes were phylogenetically related to those of *"X. arboricola"* clade despite the high divergence among these species in the organism phylogeny. These incongruences between *hrc* phylogeny and organism phylogeny can be explained by homologous recombination occurring during T3SS evolution.

To determine if recombination events affected the whole cluster or only some genes individual phylogenies were built for each *hrc/hrp* gene coding for the secretion machinery (fig. V-S3). In all phylogenies, *X. arboricola* pv. *guizotiae* strains (CFBP 7408 and CFBP 7409) did not group with other *X. arboricola* strains, but with *"X. cannabis"* strains CFBP 7912 and Nyagatare, suggesting that their whole T3SS cluster was acquired through a unique homologous recombination event with these phylogenetically distant strains. Individual *hrc/hrp* phylogenies were compared in pairs using a SH test (Table V-S3). For half of the comparisons, the p-values were below 0.05, indicating that most topologies of these trees were different, and that recombination occurred in *hrc* genes. For instance T3SS coding genes from *X. campestris* clustered with genes from *"X. arboricola"* and *"X. vesicatoria"* clades that did not reflect the intermediate position of *X. campestris* between group 1 and group 2 in the organism phylogeny (fig. V-1, fig. V-S3). In contrast to strains of *"X. axonopodis"* clade, which clustered together in most hrc phylogenies, strains from *"X. arboricola"* and *"X. vesicatoria"* clades were interspersed. This suggested numerous HGT between these two latter clades.

The recombinant sequences in T3SS cluster of *Xanthomonas* spp. strains were detected and their origins were identified to characterize the gene flow. The identified exchanges concerned the entire sequence of one or two adjacent genes in the concatenated *hrc* genes. The two genes for which the number of exchanges were the highest were *hrpE* and *hrpB2* (32 and 26 events, respectively) and the two genes for which the smallest number of exchange events were observed were *hrcC* and *hrcT* (3 and 0 events, respectively). *X. arboricola* strains were the main recipients of recombination events (fig. V-3). They mostly received genes from *X. dyei* and *X. hortorum* pv. *hederae*. The genetic structure of *X. arboricola* was previously deciphered; it encompasses a group of strains that are prone to recombination (group B) and epidemic clones representing three successful pathovars (group A) (chapter 1). Noteworthy, most exchanges were detected between group B strains and epidemic clones (group A) gave *hrp/hrc* alleles to strains belonging to the recombinogenic

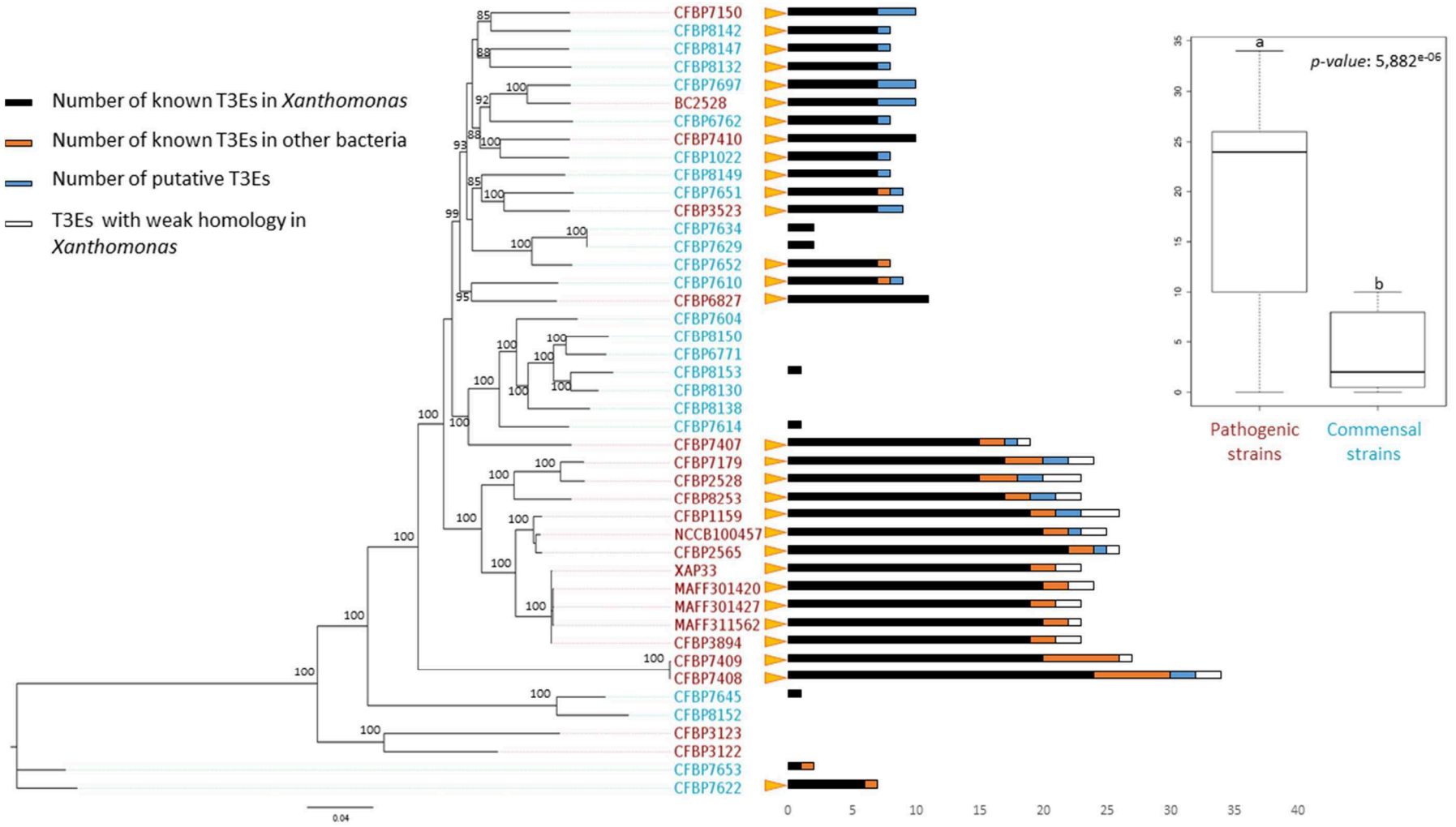


Figure V-5: Representation of T3E repertoires in *X. arboricola* strains. A. The phylogeny was performed in maximum likelihood using the concatenated sequences of 1,705 CDS composing the core genome of these 44 strains. Bootstrap values higher than 85 are indicated at each node. Pathogenic strains are represented in red and commensal ones in blue. At the tip of each branch the orange triangles represent the presence of T3SS cluster and bars represent the composition of the T3E repertoire according to the legend. B. The boxplot represents the mean of T3E number in pathogenic strains (on the left) and in commensal strains (on the right). A Mann-Whitney test was used for comparison.

background (group B), but no reverse events were detected. In contrast, only two *X. arboricola* strains were donors for other species (CFBP 1022 was donor for *X. cassavae* and CFBP 8149 for *X. hortorum* strains, *X. gardneri* and *X. cynarae*). T3SS gene exchanges were also detected between strains of *X. axonopodis* species complex but noticeably no flow of genes occurred between this clade and other clades.

Finally, we explored for each individual T3SS coding gene, the evolutionary force responsible for observed polymorphism using the ρ/Θ ratio. For most genes (14 out of 16) the ρ/Θ ratio was smaller than 1 indicating that mutation had more impact than recombination on generating new alleles (fig. V-4). Only two genes had a $\rho/\Theta > 1$, *hrcJ* ($\rho/\Theta > 2$) and *hrcV* ($1 > \rho/\Theta > 2$). In conclusion, within *hrc/hrp* genes mutation was the major evolutionary force that have brought polymorphism and generated allelic variants at gene scale. This polymorphism was distributed within organisms through homologous recombination of entire genes or contiguous genes and most of the genetic exchanges occurred within and between *"X. arboricola"* and *"X. vesicatoria"* clades.

T3E repertoires in *X. arboricola*

The T3E coding genes were searched for in all genome sequences belonging to *X. arboricola* species by a machine-learning approach dedicated to *Xanthomonas* organisms to decipher the diversity of T3E repertoires in this species. It allowed us to define a set of seven ancestral core T3E genes. The T3E repertoires in this collection were highly variable with strains having no T3E and others having 34 T3Es (fig. V-5). Eight of the 14 strains lacking the T3SS cluster, were also deprived of T3E coding genes. In contrast, between one and two T3E genes were identified in the remaining six strains (Table V-S4). For instance, three Hrp2-negative strains had *xopR* and *avrBs2*. While synteny of *xopR* genomic environments was previously shown (chapter 1), we showed here that genomic environments of *avrBs2* were also highly syntenic between all strains with the exception of strains CFBP 7408 and CFBP 7409 (fig. V-S4). This element is in favor of an ancestral acquisition and subsequent losses of these T3Es during *X. arboricola* evolution. It is worthy to note that *avrBs2* was systematically accompanied by three CDS, namely *xylR*, a TonB-dependent receptor and a hypothetical protein that presented the same distribution in our collection whatever the genomic context (fig. V-S4). In addition to *xopR* and *avrBs2*, five other T3E genes (already described in *Xanthomonas*) were found in all *X. arboricola* strains having a T3SS. These five T3E genes were located within the T3SS cluster (fig. V-4) and their distribution strictly followed the distribution of the T3SS cluster. Synteny in the flanking regions of *avrBs2*, *xopR*, and of the five T3E genes associated with the T3SS cluster suggested that they were most

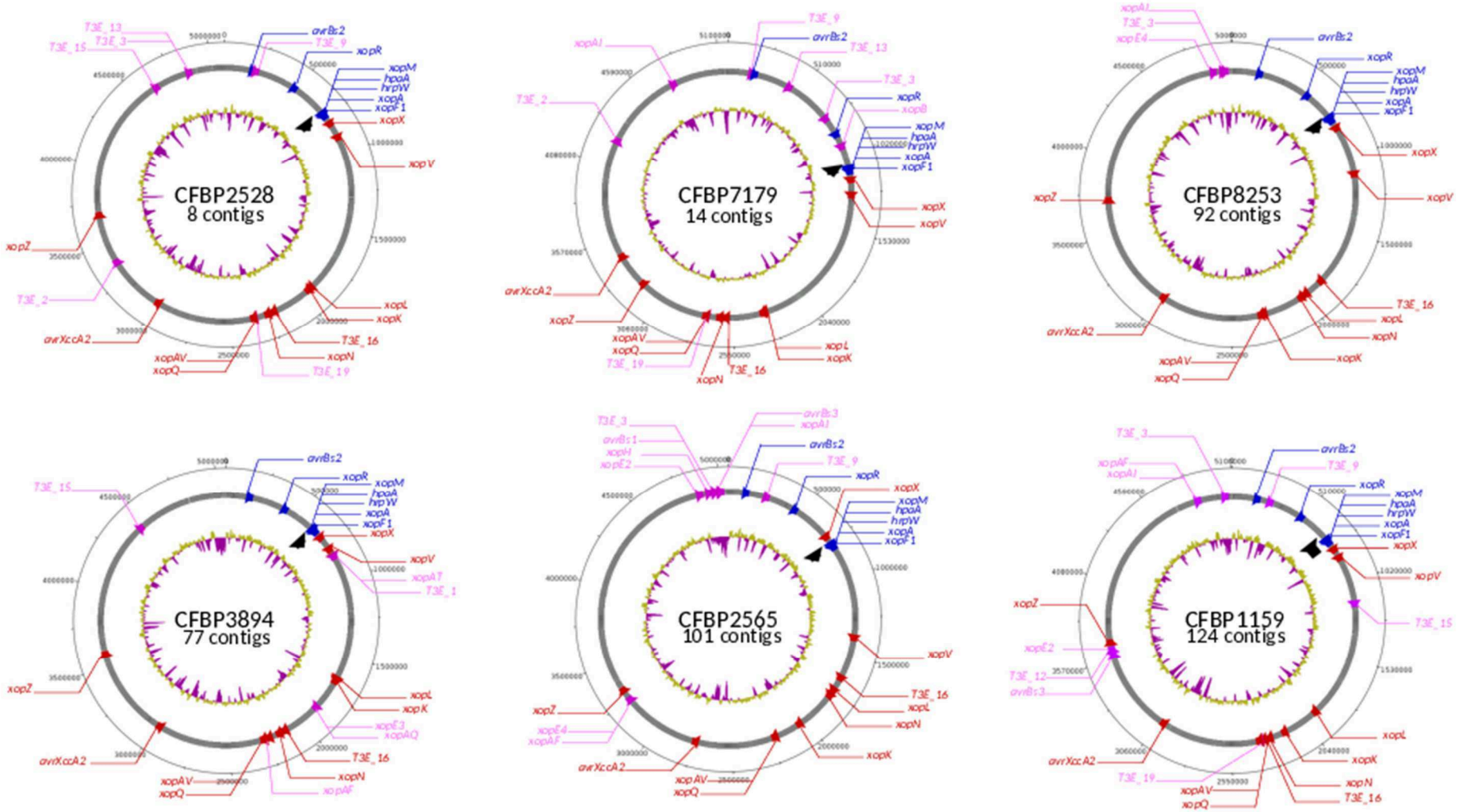


Figure V-6: Graphical circular representation of the draft genomes of *X. arboricola* strains belonging to the three successful pathovars : pv. *juglandis* (CFBP 2528, CFBP 7179, CFBP 8253), pv. *pruni* (CFBP 3894) and pv. *corylina* (CFBP 2565, CFBP 1159). The contigs were ordered by Mauve using the CFBP 2528 genome sequence as reference. The blue arrows represent the core T3E genes, the red arrows represent the 10 TE3 genes acquired by the common ancestor of these strains, the pink arrows represent the T3E genes independently acquired by the strains. The black arrows represent the T3SS coding genes.

probably acquired by an ancient *X. arboricola* strain, thus these seven T3E genes will be designated as the ancestral repertoire thereafter.

Pathogenic strains had a higher number of T3E genes than commensal strains (fig. V-5). The T3E repertoire of pathogenic strains encompassed the ancestral repertoire and a large number of additional T3E genes. Indeed, the pan-T3 effectome in *X. arboricola* was composed of 57 T3E genes, among them only 11 were found in both pathogenic and commensal strains, and the 46 others were present exclusively in pathogenic strains. Most T3E genes identified in *X. arboricola* were already described in other *Xanthomonas* spp. This is the case for the seven T3E genes of the ancestral repertoire and 31 other T3E genes (Table V-S4). Among the 19 remaining T3E genes composing the pan-effectome, seven were known in other bacteria (*Ralstonia* and *Pseudomonas*), six putative new T3E genes were identified and six (T3E_14 to T3E_19) other putative new T3E genes had a weak homology (less than 30% of homology) with T3E genes known in *Xanthomonas* (*xopAH*, *xopJ1*, *xopAO*, *xopAV*, *xopG*, and *xopM*, respectively) and thus are not considered as orthologous of these genes, but are candidate T3E genes as they could derived from these genes. The strains having a large number of known T3E genes (more than 11) had also putative new T3E genes; this is the case for CFBP 7407 (pv. *arracaciae*), CFBP 7408 and CFBP 7409 (pv. *guizotiae*) and all strains belonging to successful pathovars *pruni*, *corylina* and *juglandis* (group A according to chapter 1).

The successful pathovars shared ten T3E genes that were sequentially acquired in their common ancestor. These 10 T3E genes were *xopX*, *xopV*, *xopL*, *xopK*, *xopN*, *xopAV*, *xopQ*, *avrXccA2*, *xopZ* and T3E_16 which presented 27.2% of homology with *xopJ1*. To determine if their ancestral acquisition resulted from one or several events, we analyzed more precisely their location in genome sequences. Contig alignment using CFBP 2528 as reference revealed that these T3E genes were dispersed along the chromosome, except *xopAV* and *xopQ* that effectively colocalized (fig. V-6, fig. V-S5). The dispersion of these T3E genes along the genome sequences suggested that they were acquired following several acquisition events. Because of the synteny evidenced in the flanking regions of each of these T3E genes in the genomes of group A strains, these independent acquisition events probably occurred in their ancestor before separation into three distinct pathovars. Few traces of these HGT events were retrieved in the vicinity of these T3E genes (fig. V-S5). Indeed, ISs were found in the genomic environments of *xopQ* and *xopAV* for strains CFBP 7179 and CFBP 3894.

III. Discussion

The T3SS is among the most complex cellular machinery of bacteria. T3SS acquisition and evolution has played a major role in ecological adaptation of plant pathogens, and HGT has been a driving force in T3SS evolutionary history at multiple evolutionary time scales. The advent of next generation sequencing technologies provides invaluable data to thoroughly decipher its evolution and offer clues to understand host adaptation and pathogen emergence. Collectively, the *Xanthomonas* genus attacks c. 400 different plants including economically important crops (Ryan et al. 2011). Until recently this genus was mostly known through its pathogenic counterpart whose main virulence determinant is the Hrp2 T3SS and its effectors (Jacques et al. 2016). However, recent studies revealed new species and strains devoid of T3SS and numerous commensal strains in species otherwise well-known for their aggressive host-specialized pathogens (Triplett et al. 2015; Jacobs et al. 2015; chapter 1). These results raise intriguing questions: at which time point in xanthomonads evolution Hrp2 T3SS was acquired? Is it of ancestral origin and was it subsequently lost in some species and strains or was it acquired several time during *Xanthomonas* evolution? As commensal xanthomonads were found in sympatry with other *Xanthomonas* pathogenic strains, may HGT of the T3SS cluster have occurred during xanthomonads evolutionary history? More generally we considered the role of HGT and homologous recombination in shaping T3SS cluster polymorphism. We addressed these questions using phylogenomic analysis and detailed analyses of the genomic environments of T3SS cluster.

Comparative genomics analyses of a collection of 82 xanthomonad strains allowed us to infer three ancestral acquisitions of Hrp2 cluster during *Xanthomonas* evolution, two in group 2 and one in group 1 strains. Indeed, analysis of the chromosomal context of the T3SS cluster coupled with phylogenomic analysis highlighted an ancestral acquisition in the common ancestor of all group 2 species excluding *X. campestris*. This latter species, which diverged early in group 2, has a T3SS cluster at a different chromosomal location, which supports an independent acquisition. The third ancestral T3SS acquisition occurred in group 1. A different genetic organization of the T3SS cluster and low similarities with T3SS coding genes from the group 2 species indicated that group 1 strains acquired independently a different Hrp2 cluster. Until now *X. transluscens* was the only group 1 species known to harbor a Hrp2 cluster (Whichmann et al. 2013). Our results indicate that this atypical Hrp2 cluster is shared with *X. theicola* and *X. hyacinthi* and that it was probably acquired by their common ancestor. The high divergence based on phylogenomics and T3SS phylogeny between group 1 and group 2 species evidenced in this study is in accordance with the paraphyly of *Xanthomonas* group 1 and group 2 showed by

Naushad et al. (2013) and support the split of the genus as previously proposed by Young et al. (2008). We can thus consider the two T3SS cluster acquisitions at the root of group 2 as ancestral acquisitions as they occurred before speciation of most xanthomonads. Capacity to intimately interact with plants through translocation of T3Es would thus be an ancient trait of xanthomonads as it is the case for other important plant pathogenic clades. (Kirzinger et al. 2015; Diallo et al. 2012).

After ancestral acquisition T3SS got lost in some strains and species scattered in the *Xanthomonas* phylogenetic tree (fig. V-1). We had the opportunity to compare genome sequences of phylogenetically close strains with and without T3SS within two distant species, *"X. cannabis"* and *X. arboricola*. Scattering of strains without T3SS in the tree, the conservation of similar genomic environments in the T3SS positive strains, and a similar genome sequence in strains without T3SS are three lines of evidence in favor of the T3SS loss hypothesis. Absence of pseudogenes or remnants of T3SS encoding genes might be surprising, but a similar observation was made in nonpathogenic *P. syringae* strains, from which the entire cluster has been excised (Mohr et al. 2008). Because of absence of remnants of T3SS and T3E genes in some *"X. cannabis"* strains, Jacobs et al. (2015) favored the T3SS gain hypothesis in *"X. cannabis"* Nyagatare strain. The genome quality of Nyagatare strain didn't allow us to study its T3SS genomic environment, but CFBP 7912 strain which belongs to the same species and harbors a T3SS cluster was analyzed (fig. V-1 and V-2). We highlighted that the T3SS cluster was localized in the same chromosomal environment as in other group 2 xanthomonads with acquisition of additional genes at the direct flanking regions. The loss hypothesis in commensal strains of *X. arboricola* species was previously proposed in chapter 1 based on bayesian inference of gene gains and losses. The loss hypothesis was also proposed in *Pantoea* genus (Kirzinger et al. 2015) and in *Pseudomonas syringae* species (Mohr et al. 2008 ; Diallo et al. 2012). T3SS loss can be followed by re-acquisition leading to emergence of pathogens, as demonstrated for *X. fragariae* and *X. cassavae* which are responsible for damaging diseases. Such complex scenarios with ancestral acquisition, losses and regains, were also proposed in *Pantoae* genus (Kirginzer et al. 2015) and *P. syringae* (Clarke et al. 2010).

Losses of T3SS could be explained by a loss of function, as suggested by Abby & Rocha (2012). Indeed it could be beneficial to lose this energetically costly machinery if it does not enhance bacterial fitness (Gophna et al. 2003). Undoubtedly, highly host-specialized pathogens like pathovars *pruni*, *corylina* and *juglandis* get a high benefit from T3SS and a large T3E repertoire in the arm race against plant immunity. However in commensal strains colonizing various plant hosts and with a limited T3E repertoire (like *X. arboricola* group B strains, chapter 1), the fitness cost provided by T3SS and a limited set of T3ES might be high and in consequence

it could be lost. Bacteria should also be viewed as communities, within which some subpopulations may act as profiteers. Likewise, T3SS-negative strains may benefit from the presence of T3SS-positive strains colonizing the same niche as demonstrated in murine infections by *Pseudomonas aeruginosa* (Czechowska et al 2014).

Once acquired, we showed that T3SS coding genes were submitted to homologous recombination events leading to replacement of large fragments encompassing one complete gene or adjacent genes or even of the entire cluster. However, within each gene allelic polymorphism is mostly generated by mutation, except for *hrcJ* and *hrcV* (fig. V-4). The study of genomic environments allowed us to distinguish two mechanisms of HGT: acquisition of a new cluster in a different chromosomal context as previously discussed, and homologous recombination within T3SS cluster leading to allelic replacement. One homologous recombination event leading to entire T3SS cluster replacement was shown for *X. arboricola* pv. *guizotiae*. In these strains, T3SS cluster was ancestral as shown by its genomic localization, but phylogenies for each individual gene composing T3SS cluster showed a high genetic relatedness with the phylogenetic distant strains CFBP 7912 and strain Nyagatare. Interestingly, all these strains originate from South-East Africa and were isolated from two crops (Niger seed and bean) used in mixed crop-livestock farming systems and as intercrops in maize production (Abera et al. 2009), making their co-occurrence plausible. Beside this whole Hrp2 cluster replacement, which seems quite rare (only detected once), numerous localized homologous recombination events between strains of different species have shuffled the vertical inheritance signal. This situation is similar to the ones observed within Hrp1 T3SS in pseudomonads and enterobacterial plant pathogens, for which conflicts between phylogenies of some *hrp/hrc* genes and with MLSA trees were evidenced (Naum et al. 2009; Tegli et al. 2011; Sarris et al. 2013). An HGT event expanding through several genes was previously suggested in pseudomonads (Sarris et al. 2013). Strikingly, despite the major role of recombination in shaping *Xanthomonas* T3SS diversity at cluster scale, polymorphism within each gene was mostly generated by mutation as previously seen in the HrpZ operon of *Pseudomonas syringae* species complex (Guttman et al. 2006).

Gene flow sheds light on bacterial ecology. The study of "donor" and "recipient" strains of recombinant fragments showed that *X. arboricola* strains were the main "recipient", in particular commensal strains, and *X. dyei* and *X. hortorum* the two main donors (fig. V-3). This suggests that commensal strains are found in sympatry with a large number of different *Xanthomonas* species because genetic material exchanges can only take place when individuals are in contact. *X. dyei* and *X. arboricola* strains were isolated from the endemic species *Dysoxylum spectabile* in New-Zealand (Young et al. 2010). These observations reinforce the hypothesis proposed in chapter 1 that the commensal strains in *X. arboricola* are generalist

organisms found in many different plants. In contrast, no gene flow at T3SS locus occurred between *X. axonopodis* clade and other species in group 2. A divergent evolution of T3SS cluster in this clade may have led to an optimized allelic combination. This result reinforces the major role of T3SS in this important clade hosting host-specialized pathogens and within which we have never found so far strains lacking T3SS.

The function of T3SS, often named injectisome is to deliver T3E in host cells. What happens to T3Es when T3SS got lost? In most strains devoid of T3SS, no T3E genes could be detected in their genomes using machine learning approach. Indeed, some T3E genes (five in *X. arboricola*) are housed in the T3SS cluster (fig. V-4) and thus got lost with it. *xopR* and *avrBs2*, which are not located in T3SS cluster were found in the genomes of some commensal *Xanthomonas* strains lacking any T3SS. Their conserved genomic environments, when compared to strains with T3SS, demonstrated that they are remnants of an ancestral T3E repertoire (fig. V-S4; chapter 1). An intriguing question is why they did not get lost if there is no longer T3SS to inject them into host cell. T3SS loss might be recent. Alternatively *xopR* and *avrBs2* secretion might be mediated by the flagellum apparatus as demonstrated for some non-flagellar proteins (Journet et al. 2004). They might also have an additional function independent of the T3SS.

We highlighted an extremely reduced ancestral core repertoire and stepwise acquisition of numerous additional T3Es in pathogenic strains of *X. arboricola*. Recent studies show that most T3Es are involved in suppression of innate immunity. The ancestral repertoire of *X. arboricola* is composed of *xopA*, *hpaA*, *hrpW*, *xopF1* and *xopM* located in the T3SS cluster, plus *xopR* and *avrBs2*. The localization of these five T3E genes in the T3SS cluster was previously observed in other *Xanthomonas* species (Noel et al. 2002, Potnis et al. 2011, da Silva et al. 2002, Teper et al. 2016). For numerous authors, XopA, HrpW and HpaA are not *bona fide* T3E and should be better considered as accessory or translocation proteins that help the translocation process. Taking this into account, *X. arboricola* core effectome comprises only four T3E genes (*xopF1*, *xopM*, *avrBs2* and *xopR*) and is comparable in size to the one of *X. campestris* (Roux et al. 2015). Together, these results challenge the list of ten core T3E genes (*avrBs2*, *xopF1*, *xopK*, *xopL*, *xopN*, *xopP*, *xopQ*, *xopR*, *xopX*, *xopZ*) previously proposed (Ryan et al. 2011; White et al. 2009). Considering that *xopR* and *avrBs2* were missing in only one and two strains respectively out of 13 *X. campestris* (Roux et al. 2015), we propose a list of three putative core *Xanthomonas* T3E: AvrBs2, XopR, and XopF1. Interestingly, AvrBs2 contributes to bacterial fitness in field conditions, including epiphytic survival (Whichmann & Bergelson 2004). It is required for full aggressiveness both in dicots and monocots and was shown to inhibit pathogen-associated molecular pattern-triggered immune

(PTI) responses in rice (Zhao et al. 2011; Li et al. 2015). Similarly, XopR inhibits plant basal defenses (Akimoto-Tomiyama et al. 2012). The role of XopF1 is unknown.

Beside this reduced ancestral core T3E repertoire, stepwise accumulation of additional T3Es occurred in pathogenic strains and in particular in successful pathovars of *X. arboricola*. This accumulation through sequential acquisition events appears to be a long-term evolutionary process as many T3Es were acquired before pathovar radiation in *X. arboricola* group A. This contrasts with recent acquisition of a unique pathogenic island leading to pathogen emergence in *Pantoea agglomerans* (Barash et al. 2009). Six out of the ten T3E genes acquired in the ancestor of the three group A epidemic clones are commonly found in pathogenic xanthomonads from other species (previously listed as core *Xanthomonas* effectome; Ryan et al. 2011; White et al. 2009) and among them XopL, XopN, XopQ, XopX, and XopZ were shown to suppress host innate immunity (see review in Macho & Zipfel, 2015; Song & Yang, 2010; Teper et al. 2015; Sinha et al. 2013). Accumulation of numerous T3Es targeting PTI at the basal steps of pathogen emergence in addition to the ancestral ones, also involved in PTI suppression, strength the idea that PTI suppression is crucial for pathogenic strains to achieve successful infection (Macho & Zipfel 2015). Few variable T3E genes contribute to differential repertoires between pathovars. They represent good candidates to explain host specialization. For instance it is worthy to note that in CFBP 6827, a pathogen of tomato and pepper, three out of the four variable T3E genes (*xopAD*, *xopP*, and *xopZ2*) are also retrieved in several other *Xanthomonas* species pathogenic on the same hosts (Potnis et al. 2011). Collectively, these observations fit with and fuel previous proposed models of T3E repertoire evolution in plant pathogens. Population genetics analysis of *X. axonopodis* predicted that acquisition of T3Es in a generalist pathogen common ancestor could explain the specialization of different pathovars (Mhedbi-Hajri et al. 2013). Using a *Xanthomonas* species with an epidemic clonal structure (chapter 1), we deciphered a generalist population with a reduced ancestral repertoire and pathogenic strains harboring numerous variable T3Es. A notable feature is that *P. syringae* strains which share many common traits with *Xanthomonas* lifestyle also subvert plant immunity using a reduced core effectome and many variable effectors (Lindeberg et al. 2012).

In conclusion, we showed three ancestral acquisitions of the Hrp2 cluster demonstrating that a tight interaction with plants is an ancestral trait of xanthomonads. During radiation most species retained this ancestral T3SS but some lost it and subsequently it could have been re-gained. Mutation is the main evolutionary force that generates new alleles in *hrc/hrp* genes. In group 2 *Xanthomonas*, inter- and intra-species homologous recombination of large fragments expanding through

one or more genes shuffles this polymorphism generating new allelic combinations in Hrp2 clusters. A set of four ancestral core T3E genes is found in commensal strains and pathogens in *X. arboricola* that may approximate the *Xanthomonas* ancestral core effectome. We propose that they may allow these strains to overcome basal plant immunity under specific environmental conditions, but could have a fitness cost explaining their losses in some strains. In contrast, some strains experienced a different evolutionary pathway with stepwise accumulation of T3Es that probably accounts for their efficacy to break plant immunity and explain high aggressiveness. *X. arboricola* represents the archetype of this evolutionary scenario, which seems to share similarities with the one proposed by Lindeberg et al. (2012) for *P. syringae* and culminates in a narrow host-specialization.

IV. Materials and Methods

Genome sequencing and annotation

We used a collection of 82 genome sequences (Table V-S1) representing the known diversity of *X. arboricola* (44 strains) and the diversity of *Xanthomonas* genus (36 additional strains belonging to *Xanthomonas* species). Among the 44 strains of *X. arboricola*, 23 strains were described as commensal and 21 as pathogenic (chapter 1). Genomes were sequenced using the Illumina 2500 technology (Genoscreen, Lille, France). Libraries of genomic DNA were performed using the Kit Nextera XT. Paired-end reads of 2*100 bp were assembled in contigs using SOAPdenovo 1.05 (Li et al. 2010) and Velvet 1.2.02 (Zerbino & Birney, 2008). Annotation was performed using EuGene-PP using similarities with known protein sequences (Sallet et al. 2014).

Prediction of T3SS cluster and T3E repertoires

The T3SS coding genes representing all T3SS families (Ysc, SPI-1, SPI-2, Chlamy, Hrp1, Rhizo) were detected in genome sequences using BLASTp searches. The query sequences used for the BLASTp search are presented in the Table V-S2. Genes were considered homologs when the percent of identity was higher than 80 on at least 80% of the length. Lower thresholds were used to highlight putative pseudogenes. The T3E genes detection was performed in all genomes of *X. arboricola* by a machine-learning approach (E. Bosis, unpublished results).

Genomic environments of genes

Genomic environments were analyzed using the R package GenoplotR (Guy et al. 2010). First, for T3SS coding genes, this analysis was performed using only strains having a T3SS cluster to select conserved flanking regions upstream and downstream of the cluster. A BLASTn search in the genomes of strains lacking T3SS cluster using the flanking conserved regions identified in the first analysis as queries allowed us to retrieve the putative insertion site of T3SS cluster in strains lacking it. To study the synteny in the genomic environments of T3SS insertion site, both strains with and without T3SS cluster were included in the final analysis.

For the genomic environments of T3E genes, these T3E genes were located by their locus tag obtained during the search with the machine learning approach. The same strategy as described above was used to study the genomic environment of *avrBs2* insertion site in *X. arboricola* strains having this gene or not.

Mapping T3E genes on whole genome sequences

To locate the T3E genes in the genomes of pathogenic strains of group A, the contigs of genome sequences were ordered using MAUVE (Darling et al. 2004). The sequence of CFBP 2528 was used as reference because the number of contigs was the lowest for this strain (8 contigs). For each genome, the contigs were concatenated using Geneious (Kearse et al. 2012) according to the order obtained with MAUVE. The circular representations were obtained using DNAPlotter (Carver et al. 2009). The localization of each T3E genes in pathogenic strains was identified using their locus tag.

Determination of core genomes

The core proteome of *Xanthomonas* was identified with family-companion (Carrere et al. 2015). The core proteome of *X. arboricola* was identified with orthoMCL V2.0.9 analyses on predicted full-length proteins (Li et al. 2003). OrthoMCL clustering analyses were performed using the following parameters: P-value Cut-off = 1*105; Percent Match Cut-off = 80; MCL Inflation = 1.5; Maximum Weight = 316.

Phylogenies of core and T3SS coding genes

Phylogenies were performed in maximum likelihood using phyML. The phylogeny of *Xanthomonas* genus was performed using the concatenated core proteome obtained with family-companion. For *X. arboricola* phylogeny, the concatenated orthologuous groups were used. Each orthologous group in *X. arboricola* was aligned

using Macse (Ranwez et al. 2011). Only alignments having a percentage of identity by sites higher than 75% were kept for the phylogeny reconstruction. These alignments were concatenated using Geneious (Kearse et al. 2012). For organism maximum likelihood phylogenies the JTT model was used. For the T3SS coding genes phylogeny, each *hrc* gene (with the exception of *hrcL* for which some CDS were truncated) was aligned using muscle (Edgar, 2004), taking into account sequence translation in proteins to conserve the reading frame. The 10 *hrc* genes were concatenated according to their order in T3SS cluster. Phylogeny was performed with the GTR + I + gamma model.

Phylogeny comparisons and recombination analyses

The topology of the concatenated *hrc* tree was compared to the core proteome tree of *Xanthomonas* genus performing a Shimodaira-Hasegawa test (Shimodaira & Hasegawa, 1999) using R package phangorn (Schliep, 2011). In the same way, topologies of each individual *hrc/hrp* tree were compared to each other and to the topology of the concatenated *hrc/hrp* tree. The impact of recombination (r) relative to mutation (m) was analyzed with the ρ/Θ statistics using RDP v.3.44 (Martin & Rybicki, 2000) for each *hrc/hrp* gene located in the core region of the cluster (16 genes between *hrcC* and *hrpE*). The origins of recombinant sequences were identified by examining the concatenated sequences of *hrp* and *hrc* genes (concatenated according to their order in T3SS cluster) using RDP 3, GENECONV, BOOTSCAN, MAXIMUM CHI SQUARE, CHIMAERA, SISCAN, and 3SEQ implemented in RDP v. 3.44 (Martin & Rybicki, 2000). We considered that a recombination event was statistically supported when it was detected by at least two methods (chapter 1). The recombination event representation was realized using Circos (Krzywinski et al. 2009).

Acknowledgments

We thank G. Taghouti for DNA extraction, ANAN platform for genome sequencing and J. Gouzy for the genome assembling and annotation. We thank CIRM-CFBP (Beaucouzé, INRA, France) for strain preservation and supply. This work was part of a program funded by the French Agence Nationale de la Recherche (ANR) and by INRA (AIP Bioresources project "Taxomic"). We thank members of the French Network of *Xanthomonas* (FNX) for fruitful discussions. We thank the INRA Plant Health and Environment Division and the regional government of the Pays de la Loire, France for funding the doctoral fellowship of DM. This work benefited from interactions promoted by COST Action FA 1208 (https://www.cost-sustain.org).

Supplemental materials

Table V-S1 : Collection of genomes

Nomenclature	code	group of *X. arboricola*[§]	Genome Accession	Host of isolation	Geographic origin	Year of isolation	contig number	size (Mb)	% GC	Gene number
X. albilineans	CFBP2523	NA[&]	NA	Saccharum officinarum	Fidji	1961	148	3,68	63,1	3129
X. alfalfae subsp. citrumelonis	CFBP3371	NA	NA	NA	NA	1989	46	4,9	65	4255
X. arboricola	bc2528	B	NA	*Capsicum annuum*	Korea	2007	132	4,9	65,7	4167
X. arboricola	CFBP1022	B	NA	*Juglans regia*	France	1967	43	5,1	65,6	4268
X. arboricola	CFBP6762	B	NA	*fragaria X ananassa*	Italy	NA	103	4,91	65,74	4215
X. arboricola	CFBP6771	B	NA	*fragaria X ananassa*	Italy	1993	56	4,91	65,87	4124
X. arboricola	CFBP6827	B	NA	*Capsicum annuum*	United States	2000	5	4,95	65,67	4107
X. arboricola	CFBP7604	B	NA	*Capsicum annuum*	NA	NA	6	4,85	65,78	4050
X. arboricola	CFBP7610	B	NA	*Pueraria lobata*	NA	1981	46	4,96	65,63	4211
X. arboricola	CFBP7614	B	NA	*Magnolia sp.*	NA	NA	46	4,92	65,54	4202
X. arboricola	CFBP7622	C	NA	*Phaseolus vulgaris*	United States	1985	106	4,75	66,18	4187
X. arboricola	CFBP7629	B	NA	*Juglans regia*	France	2009	62	4,95	65,59	4238
X. arboricola	CFBP7634*	B	PRJNA275693	*Juglans regia*	France	2009	4	4,94	66,04	4141
X. arboricola	CFBP7645	C	NA	*Juglans regia*	France	2009	14	4,76	65,46	4087
X. arboricola	CFBP7651*	B	PRJNA275695	*Juglans regia*	France	2008	6	5,03	65,98	4229
X. arboricola	CFBP7652	B	NA	*Juglans regia*	France	2008	78	5,23	65,49	4439

X. arboricola	CFBP7653	C	NA	*Juglans regia*	France	2008	77	4,87	66,14	4152
X. arboricola	CFBP7697	B	NA	*Phaseolus vulgaris*	NA	2011	8	4,95	65,67	4141
X. arboricola	CFBP8130	B	NA	*Phaseolus vulgaris*	France	2010	78	4,77	65,99	3985
X. arboricola	CFBP8132	B	NA	*Phaseolus vulgaris*	France	2010	44	4,93	65,7	4179
X. arboricola	CFBP8138	B	NA	*Phaseolus vulgaris*	NA	2010	57	4,76	66,04	3959
X. arboricola	CFBP8142	B	NA	*Phaseolus vulgaris*	NA	2010	55	4,85	65,8	4097
X. arboricola	CFBP8147	B	NA	*Phaseolus vulgaris*	NA	2010	105	4,89	65,72	4183
X. arboricola	CFBP8149	B	NA	*Phaseolus vulgaris*	NA	2010	33	4,88	65,69	4117
X. arboricola	CFBP8150	B	NA	*Phaseolus vulgaris*	NA	2010	351	4,92	65,85	4496
X. arboricola	CFBP8152	C	NA	*Phaseolus vulgaris*	NA	2010	199	4,73	65,52	4188
X. arboricola	CFBP8153	B	NA	*Phaseolus vulgaris*	NA	2010	78	4,84	65,9	4045
X. arboricola pv. arracaciae	CFBP7407	B	NA	*Arracacia xanthorrhiza*	Brazil	1969	232	5,16	65,21	4462
X. arboricola pv. celebensis	CFBP3523*	B	PRJNA254393	*Musa acuminata*	New Zealand	1960	3	4,91	65,6	4055
X. arboricola pv. celebensis	CFBP7150*	B	PRJNA252657	*Musa acuminata*	NA	NA	7	4,98	65,6	4098
X. arboricola pv. corylina	CFBP1159	A	NA	*Corylus maxima*	United States	1939	124	5,11	65,54	4396
X. arboricola pv. corylina	CFBP2565	A	NA	*Corylus avellana*	France	1985	101	5,06	65,55	4433
X. arboricola pv. corylina	NCCB100457*	A	PRJNA193452	*Corylus colurna*	NA	1985	281	5,23	65,5	4455
X. arboricola pv. guizotiae	CFBP7408	B	NA	*Guizotia abyssinica*	Ethiopy	NA	162	4,94	65,6	4299
X. arboricola pv. guizotiae	CFBP7409	B	NA	*Guizotia abyssinica*	NA	NA	220	4,16	65,92	4199

X. arboricola pv. juglandis	CFBP8253	A	NA	*Juglans regia*	France	2002	92	5	65,55	4265
X. arboricola pv. juglandis	CFBP2528*	A	PRJNA275685	*Juglans regia*	New Zealand	1956	8	5,09	65,92	4325
X. arboricola pv. juglandis	CFBP7179*	A	PRJNA275691	*Juglans regia*	France	2002	14	5,16	65,93	4399
X. arboricola pv. populi	CFBP3122	C	NA	*Salix alba*	Netherlands	1980	92	4,79	65,36	4125
X. arboricola pv. populi	CFBP3123	C	NA	*Salix alba*	Netherlands	1979	28	4,65	65,48	4050
X. arboricola pv. pruni	CFBP3894	A	NA	*Prunus salicina*	Netherlands	1953	77	5,06	65,42	4337
X. arboricola pv. pruni	MAFF301420*	A	PRJDB1524	NA	NA	1978	396	5	65,3	4511
X. arboricola pv. pruni	MAFF301427*	A	PRJDB1525	NA	NA	1980	356	4,91	65,4	4373
X. arboricola pv. pruni	MAFF311562*	A	PRJDB1518	NA	NA	2008	432	5,09	65,3	4681
X. arboricola pv. pruni	XAP33*	A	PRJNA243321	*Prunus amygdalus cv. Guara*	NA	NA	474	5,1	65,4	4565
X. arboricola pv. zantedeschiae	CFBP7410	B	NA	*Zantedeschiae aethiopica*	South Africa	1967	37	4,98	65,53	4201
X. axonopodis pv. begoniae	CFBP2524	NA	NA	*Begonia sp*	New Zealand	1962	108	5,07	64,8	4348
X. axonopodis pv. dieffenbachiae	LMG695*	NA	PRJNA254465	*Anthurium sp*	Brazil	1965	1	5,036	64,9	4417
X. axonopodis pv. vasculorum	CFBP5823	NA	NA	*Saccharum officinarum*	Mauritius island	1979	328	4,69	64,2	4357
X. bromi	CFBP1976	NA	NA	*Bromus carinatus*	France	1980	222	4,9	64,8	4272
X. campestris pv. campestris	B100*	NA	PRJNA29801	NA	Germany	1939	1	5,079	65	4471
X. campestris pv. raphani	756C*	NA	PRJNA63187	NA	NA	NA	1	4,94	65,3	4520
"X. cannabis"	Nyagatare*	NA	PRJNA262934	*Phaseolus vulgaris*	Rwanda	2013	109	4,88	65,7	3947
"X. cannabis"	NCPPB2877*	NA	PRJNA266578	*Cannabis sativa*	Romania	1974	257	4,75	65,8	3910

"X. cannabis"	NCPPB3753*	NA	PRJNA266587	*Cannabis sativa*	Japan	1982	260	4,8	65,6	3976
"X. cannabis"	CFBP7912#	NA	NA	*Phaseolus vulgaris*			69	4,9	65,7	4180
X. cassavae	CFBP4642*	NA	PRJNA172046	*Manihot esculenta*	Malawi	1951	7	5,2	65,2	4774
X. citri pv. citri	306*	NA	PRJNA177640	NA	NA	NA	1	5,17	64,8	4427
X. codiaei	CFBP4690	NA	NA	*Codiacum variegatum*	United States	1987	110	5,04	66	4293
X. cucurbitae	CFBP2542	NA	NA	*Cucurbita maxima*	New Zealand	1968	181	4,42	65,4	3696
X. cynarae	CFBP4188	NA	NA	*Cynara scolymus*	France	1996	102	5,06	63,7	4375
X. dyei	CFBP7245	NA	NA	*Metrosideros excelsa*	New Zealand	1993	152	5,31	64,3	4501
X. euvesicatoria	85-10*	NA	PRJNA13649	NA	NA	NA	1	5,18	64,7	4726
X. fragariae	LMG25863*	NA	PRJNA80793	*fragaria X ananassa*	Belgium	2002	96	4,18	62,2	3786
X. fuscans pv. fuscans	CFBP4834*	NA	PRJNA254240	*Phaseolus vulgaris cv. Michelet*	France	1998	1	4,98	64,7	4586
X. gardneri	ATCC19865*	NA	PRJNA60023	NA	NA	NA	552	5,23	63,7	5028
X. hortorum pv. carotae	M081*	NA	PRJNA50443	*Daucus carotae*	United States	NA	154	5,03	63,8	4419
X. hortorum pv. hederae	CFBP4925	NA	NA	*Hedera helix*	United States	1944	313	5,32	63,8	4501
X. hyacinthi	CFBP1156	NA	NA	*Hyacinthus orientalis*	Netherlands	1958	104	4,91	68,1	4128
X maliensis	M97*	NA	PRJNA196946	*Oryza sativa*	Mali	2009	1285	4,96	66,7	3762
X melonis	CFBP4644	NA	NA	*Cucumis melo*	Brazil	1974	37	4,68	66,1	3961
X pisi	CFBP4643	NA	NA	*Pisum sativum*	Japan	1957	81	4,25	64,7	4450
X. oryzae pv. oryzae	PXO99*	NA	PRJNA286129	NA	NA	NA	1	5,24	63,6	4988
X. oryzae pv. oryzicola	BAI3*	NA		NA	NA	NA	1	4,64	63,9	4715
X. perforans	91-118*	NA	PRJNA60021	NA	NA	NA	291	5,26	65	4638
X. populi	CFBP1817	NA	NA	*Populus x canadensis cv. Regenerata*	France	1957	407	4,33	63,3	4192
X. theicola	CFBP4691	NA	NA	*Camellia sinensis*	Japan	1974	430	4,54	68,4	3926

X. translucens pv. cerealis	CFBP2541 *	NA	PRJNA268946	*Bromus inermis*	United States	1941	10	4,52	67,4	3823
X. sacchari	CFBP4641	NA	NA	*Saccharum officinarum*	France	1998	57	4,89	69	4253
X. vasicola pv. holcicola	CFBP2543	NA	NA	*Sorghum vulgare*	New Zealand	1969	135	4,92	63,3	4428
X. vasicola pv. musacearum	NCPPB4394*	NA	PRJNA73883	NA	NA	NA	85	4,76	63,5	3827
X. vesicatoria	ATCC35937*	NA	PRJNA60019	NA	NA	NA	296	5,53	64,1	4928

* genome sequence publicly available

strain CFBP 7912 is classified in "*X. cannabis*"according JSpecies calculations that revealed ANI values of 96,4 % with strains NCPPB 2877 and NCPPB 3753 (95-96 % is the proposed threshold for species delineation).

§ groups as determined by Merda et al. (2016) DOI: 10.1111/1758-2229.12397

[&]NA : not available

Table V-S2 : T3SS coding genes used as query blast searches

Function	Organism	gene name	product	Accession
T3SS-Rhi	Rhizobiales	Bra-RhcC1	Outer membrane ring	YP_005612846.1
T3SS-Rhi	Rhizobiales	Bra-RhcC2	Outer membrane ring	YP_005612817.1
T3SS-Rhi	Rhizobiales	Bra-RhcJ	Inner membrane ring	YP_005612844.1
T3SS-Rhi	Rhizobiales	Bra-Y4yQ	Inner membrane ring	CCD89717.1
T3SS-Rhi	Rhizobiales	Bra-RhcQ	Cytoplasmic ring	YP_005612839.1
T3SS-Rhi	Rhizobiales	Bra-RhcR	Export apparatus	YP_005612838.1
T3SS-Rhi	Rhizobiales	Bra-RhcS	Export apparatus	YP_005612837.1
T3SS-Rhi	Rhizobiales	Bra-RhcT	Export apparatus	YP_005612836.1
T3SS-Rhi	Rhizobiales	Bra-RhcU	Export apparatus	YP_005612835.1
T3SS-Rhi	Rhizobiales	Bra-RhcV	Export apparatus	YP_005612858.1
T3SS-Rhi	Rhizobiales	Bra-NolU	Needle assembly	YP_005612843.1
T3SS-Rhi	Rhizobiales	Bra-NolV	Needle assembly	YP_005612842.1
T3SS-Rhi	Rhizobiales	Bra-NopX	Translocon	CAQ57561.1
T3SS-Rhi	Rhizobiales	Bra-NopA	Translocon	CAQ57567.1
T3SS-Rhi	Rhizobiales	Bra-NopB	Translocon	AGH09934.1
T3SS-Rhi	Rhizobiales	Bra-RhcN	ATPase	YP_005612841.1
T3SS-Rhi	Rhizobiales	RhcL		AAB91947.1
T3SS-Bur-T3SS1	Burkholderiales	Bur-T3S1-BpscC	Outer membrane ring	YP_111402.1
T3SS-Bur-T3SS1	Burkholderiales	Bur-T3S1-BpscJ	Inner membrane ring	YP_111409.1
T3SS-Bur-T3SS1	Burkholderiales	Bur-T3S1-BpscD	Inner membrane ring	YP_111419.1
T3SS-Bur-T3SS1	Burkholderiales	Bur-T3S1-BpscQ	Cytoplasmic ring	YP_111415.1
T3SS-Bur-T3SS1	Burkholderiales	Bur-T3S1-BpscR	Export apparatus	YP_111416.1
T3SS-Bur-T3SS1	Burkholderiales	Bur-T3S1-BpscS	Export apparatus	YP_111417.1

T3SS-Bur-T3SS1	Burkholderiales	Bur-T3S1-BpscT	Export apparatus	YP_111404.1
T3SS-Bur-T3SS1	Burkholderiales	Bur-T3S1-BpscU	Export apparatus	YP_111412.1
T3SS-Bur-T3SS1	Burkholderiales	Bur-T3S1-BpscV	Export apparatus	YP_111413.1
T3SS-Bur-T3SS1	Burkholderiales	Bur-T3S1-BPSS1393	Needle assembly	YP_111405.1
T3SS-Bur-T3SS1	Burkholderiales	Bur-T3S1-BPSS1396	Needle assembly	YP_111408.1
T3SS-Bur-T3SS1	Burkholderiales	Bur-T3S1-BpscL	Needle assembly	YP_111407.1
T3SS-Bur-T3SS1	Burkholderiales	Bur-T3S1-BPSS1409	Needle major subunit	YP_111421.1
T3SS-Bur-T3SS1	Burkholderiales	Bur-T3S1-BPSS1398	Needle minor subunit	YP_111410.1
T3SS-Bur-T3SS1	Burkholderiales	Bur-T3S1-BpscN	ATPase	YP_111406.1
T3SS-Bur-T3SS2	Burkholderiales	Bur-T3S2-BpscC2a|Bur-T3S2-BpscC2b	Outer membrane ring	YP_001583717.1
T3SS-Bur-T3SS2	Burkholderiales	Bur-T3S2-BpscJ2	Inner membrane ring	YP_001583720.1
T3SS-Bur-T3SS2	Burkholderiales	Bur-T3S2-BpscQ2	Cytoplasmic ring	YP_001583731.1
T3SS-Bur-T3SS2	Burkholderiales	Bur-T3S2-BpscT2	Export apparatus	YP_001583725.1
T3SS-Bur-T3SS2	Burkholderiales	Bur-T3S2-BpscV2	Export apparatus	YP_001583726.1
T3SS-Bur-T3SS2		Burkholderiales		YP_001583730.1
T3SS-Bur-T3SS2		Burkholderiales	Needle assembly	YP_001583722.1
T3SS-Bur-T3SS2		Burkholderiales	Needle major subunit	YP_001583716.1
T3SS-Bur-T3SS2		Burkholderiales	Needle minor subunit	YP_001583715.1
T3SS-Bur-T3SS2	Burkholderiales	Bur-T3S2-BpscN2	ATPase	YP_001583723.1
T3SS-Bur-T3SS2	Burkholderiales	Bur-T3S3-BsaP	Effector export	YP_111550.1
T3SS-Bur-T3SS3	Burkholderiales	Bur-T3S3-BsaO	Outer membrane ring	YP_111551.1
T3SS-Bur-T3SS3	Burkholderiales	Bur-T3S3-BsaV	Cytoplasmic ring	YP_111544.1

T3SS-Bur-T3SS3	Burkholderiales	Bur-T3S3-BsaX	Export apparatus	YP_111542.1
T3SS-Bur-T3SS3	Burkholderiales	Bur-T3S3-BsaY	Export apparatus	YP_111541.1
T3SS-Bur-T3SS3	Burkholderiales	Bur-T3S3-BsaZ	Export apparatus	YP_111540.1
T3SS-Bur-T3SS3	Burkholderiales	Bur-T3S3-BsaQ	Export apparatus	YP_111549.1
T3SS-Bur-T3SS3	Burkholderiales	Bur-T3S3-BsaT	Needle assembly	YP_111546.1
T3SS-Bur-T3SS3	Burkholderiales	Bur-T3S3-BsaU	Needle assembly	YP_111553.1
T3SS-Bur-T3SS3	Burkholderiales	Bur-T3S3-OrgA	Needle assembly	YP_111554.1
T3SS-Bur-T3SS3	Burkholderiales	Bur-T3S3-OrgB	Needle assembly	YP_111555.1
T3SS-Bur-T3SS3	Burkholderiales	Bur-T3S3-BsaL	Needle major subunit	YP_111556.1
T3SS-Bur-T3SS3	Burkholderiales	Bur-T3S3-BsaK	Needle minor subunit	YP_111557.1
T3SS-Bur-T3SS3	Burkholderiales	Bur-T3S3-BipB	Translocon	YP_111558.1
T3SS-Bur-T3SS3	Burkholderiales		Translocon	YP_111600.1
T3SS-Bur-T3SS3	Burkholderiales	Bur-T3S3-BipC	Translocon	YP_111625.1
T3SS-Bur-T3SS3	Burkholderiales	Bur-T3S3-BipD	Translocon	YP_111629.1
T3SS-Bur-T3SS3	Burkholderiales		Translocon	YP_111634.1
T3SS-Bur-T3SS3	Burkholderiales		Translocon	YP_111635.1
T3SS-Ral-Hrp2	Ralstonia	Ral-HrcC	Outer membrane ring	NP_522435.1
T3SS-Ral-Hrp2	Ralstonia	Ral-HrcJ	Inner membrane ring	NP_522428.1
T3SS-Ral-Hrp2	Ralstonia	Ral-HrcQ	Cytoplasmic ring	NP_522422.1
T3SS-Ral-Hrp2	Ralstonia	Ral-HrcR	Export apparatus	NP_522421.1
T3SS-Ral-Hrp2	Ralstonia	Ral-HrcS	Export apparatus	NP_522420.1
T3SS-Ral-Hrp2	Ralstonia	Ral-HrcT	Export apparatus	NP_522433.1

T3SS-Ral-Hrp2	Ralstonia	Ral-HrcU	Export apparatus	NP_522425.1
T3SS-Ral-Hrp2	Ralstonia	Ral-HrcV	Export apparatus	NP_522424.1
T3SS-Ral-Hrp2	Ralstonia	Ral-HrpD	Needle assembly	NP_522432.1
T3SS-Ral-Hrp2	Ralstonia	Ral-HrpH	Needle assembly	NP_522429.1
T3SS-Ral-Hrp2	Ralstonia	Ral-HrpF	Needle assembly	NP_522430.1
T3SS-Ral-Hrp2	Ralstonia	hrpY	Needle major subunit	NP_522416.1
T3SS-Ral-Hrp2	Ralstonia	Ral-HrpJ	Needle minor subunit	NP_522427.1
T3SS-Ral-Hrp2	Ralstonia	Ral-PopF1|Ral-PopF2	Translocon	NP_523114.1
T3SS-Ral-Hrp2	Ralstonia		Translocon	NP_522461.1
T3SS-Ral-Hrp2	Ralstonia	Ral-HrcN	ATPase	NP_522431.1
T3SS-Bsc	Bordetella	Bor-BscC	Outer membrane ring	NP_880879.1
T3SS-Bsc	Bordetella	Bor-BscJ	Inner membrane ring	NP_880891.1
T3SS-Bsc	Bordetella	Bor-BscD	Inner membrane ring	NP_880905.1
T3SS-Bsc	Bordetella	Bor-BscQ	Cytoplasmic ring	NP_880885.1
T3SS-Bsc	Bordetella	Bor-BscR	Export apparatus	NP_880884.1
T3SS-Bsc	Bordetella	Bor-BscS	Export apparatus	NP_880883.1
T3SS-Bsc	Bordetella	Bor-BscT	Export apparatus	NP_880882.1
T3SS-Bsc	Bordetella	Bor-BscU	Export apparatus	NP_880881.1
T3SS-Bsc	Bordetella	Bor-BcrD	Export apparatus	NP_880904.1
T3SS-Bsc	Bordetella	Bor-BscO	Needle assembly	NP_880887.1
T3SS-Bsc	Bordetella	Bor-BscK	Needle assembly	NP_880890.1
T3SS-Bsc	Bordetella	Bor-BscL	Needle assembly	NP_880889.1

T3SS-Bsc	Bordetella	Bor-BscF	Needle major subunit	NP_880907.1
T3SS-Bsc	Bordetella	Bor-BscI	Needle minor subunit	NP_880892.1
T3SS-Bsc	Bordetella	Bor-BopB	Translocon	NP_880895.1
T3SS-Bsc	Bordetella	Bor-BopD	Translocon	NP_880896.1
T3SS-Bsc	Bordetella	Bor-Bsp22	Translocon	NP_880899.1
T3SS-Bsc	Bordetella	Bor-BscN	ATPase	NP_880888.1
T3SS-Xan-Hrp2	Xanthomonas	Xan-HrcC	Outer membrane ring	
T3SS-Xan-Hrp2	Xanthomonas	Xan-HrpB	Periplasmic Ring	
T3SS-Xan-Hrp2	Xanthomonas	hrpB2	Periplasmic Ring	
T3SS-Xan-Hrp2	Xanthomonas	hrpB5	Periplasmic Ring	
T3SS-Xan-Hrp2	Xanthomonas	Xan-HrcJ	Inner membrane ring	
T3SS-Xan-Hrp2	Xanthomonas	Xan-HrcD	Inner membrane ring	
T3SS-Xan-Hrp2	Xanthomonas	hrpD6	Inner membrane ring	
T3SS-Xan-Hrp2	Xanthomonas	Xan-HrcQ	Cytoplasmic ring	
T3SS-Xan-Hrp2	Xanthomonas	Xan-HrcR	Export apparatus	
T3SS-Xan-Hrp2	Xanthomonas	Xan-HrcS	Export apparatus	
T3SS-Xan-Hrp2	Xanthomonas	Xan-HrcT	Export apparatus	
T3SS-Xan-Hrp2	Xanthomonas	Xan-HrcU	Export apparatus	
T3SS-Xan-Hrp2	Xanthomonas	Xan-HrcV	Export apparatus	
T3SS-Xan-Hrp2	Xanthomonas	Xan-HrpB7	Needle assembly	
T3SS-Xan-Hrp2	Xanthomonas	Xan-HrpB4	Needle assembly	
T3SS-Xan-Hrp2	Xanthomonas	Xan-HrcL	Needle assembly	YP_362162.1

T3SS-Xan-Hrp2	Xanthomonas	Xan-HrpE	Needle major subunit	
T3SS-Xan-Hrp2	Xanthomonas	Xan-HrpF	Translocon	
T3SS-Xan-Hrp2	Xanthomonas	hpa3	Translocon	
T3SS-Xan-Hrp2	Xanthomonas	hpaB	Translocon	
T3SS-Xan-Hrp2	Xanthomonas	hpaA	Translocon	
T3SS-Xan-Hrp2	Xanthomonas	hpaP	Translocon	
T3SS-Xan-Hrp2	Xanthomonas	HpaH	Translocon	
T3SS-Xan-Hrp2	Xanthomonas	Xan-XopA	Translocon	
T3SS-Xan-Hrp2	Xanthomonas	Xan-HrcN	ATPase	
T3SS-Pseudo-Psc	Pseudomonas	Pse-PscC	Outer membrane ring	YP_234292.1
T3SS-Pseudo-Psc	Pseudomonas	Pse-PscJ	Inner membrane ring	YP_234287.1
T3SS-Pseudo-Psc	Pseudomonas	Pse-PscQ	Cytoplasmic ring	YP_234302.1
T3SS-Pseudo-Psc	Pseudomonas	Pse-PscR	Export apparatus	YP_234300.1
T3SS-Pseudo-Psc	Pseudomonas	Pse-PscS	Export apparatus	YP_234299.1
T3SS-Pseudo-Psc	Pseudomonas	Pse-PscT	Export apparatus	YP_234298.1
T3SS-Pseudo-Psc	Pseudomonas	Pse-PscU	Export apparatus	YP_234297.1
T3SS-Pseudo-Psc	Pseudomonas	Pse-PcrD	Export apparatus	YP_003529903.1
T3SS-Pseudo-Psc	Pseudomonas	Pse-PscP	Needle assembly	YP_234303.1
T3SS-Pseudo-Psc	Pseudomonas			YP_003529900.1
T3SS-Pseudo-Psc	Pseudomonas	Pse-PscF	Needle major subunit	NP_791208.1
T3SS-Pseudo-Psc	Pseudomonas			YP_234284.1
T3SS-Pseudo-Psc	Pseudomonas	Pse-PopB	Translocon	YP_234286.1

T3SS-Pseudo-Psc		Pseudomonas	Translocon	YP_003529901.1
T3SS-Pseudo-Psc		Pseudomonas	Translocon	YP_234289.1
T3SS-Pseudo-Psc	Pseudomonas	Pse-PopD	Translocon	YP_234288.1
T3SS-Pseudo-Psc	Pseudomonas	Pse-PcrV	Translocon	YP_234307.1
T3SS-Pseudo-Psc		Pseudomonas	Translocon	YP_234276.1
T3SS-Pseudo-Psc		Pseudomonas	Translocon	YP_234308.1
T3SS-Pseudo-Psc		Pseudomonas	Translocon	YP_003529904.1
T3SS-Pseudo-Psc		Pseudomonas	Translocon	YP_003529893.1
T3SS-Pseudo-Psc	Pseudomonas	Pse-PscN	ATPase	YP_234305.1
T3SS-Pseudo-Psc	Pseudomonas	Pse-PopN	Effector export	ABQ88365.1
T3SS-LEE-Esc	E. Coli	Esc-LEE-EscC	Outer membrane ring	NP_312602.1
T3SS-LEE-Esc	E. Coli	Esc-LEE-EscJ	Inner membrane ring	NP_312600.1
T3SS-LEE-Esc	E. Coli	Esc-LEE-EscD	Inner membrane ring	NP_312585.1
T3SS-LEE-Esc	E. Coli	Esc-LEE-EscS	Export apparatus	NP_312609.1
T3SS-LEE-Esc	E. Coli	Esc-LEE-EscT	Export apparatus	NP_312608.1
T3SS-LEE-Esc	E. Coli	Esc-LEE-EscF	Needle major subunit	NP_312579.1
T3SS-LEE-Esc	E. Coli	Esc-LEE-EspD	Translocon	NP_312582.1
T3SS-LEE-Esc	E. Coli	Esc-LEE-EspB	Translocon	NP_312581.1
T3SS-LEE-Esc	E. Coli	Esc-LEE-EspA	Translocon	NP_312583.1
T3SS-LEE-Esc	E. Coli	Esc-LEE-EscN	ATPase	NP_312595.1
T3SS-ETT2-Eiv-Epr	E. Coli	Esc-ETT2-EivG	Outer membrane ring	NP_311760.1
T3SS-ETT2-Eiv-Epr	E. Coli	Esc-ETT2-EprK	Inner membrane ring	NP_311743.1

T3SS-ETT2-Eiv-Epr	E. Coli	Esc-ETT2-EprH	Inner membrane ring	NP_311746.1
T3SS-ETT2-Eiv-Epr	E. Coli	Esc-ETT2-EpaO	Cytoplasmic ring	EEC28302.1
T3SS-ETT2-Eiv-Epr	E. Coli	Esc-ETT2-EpaP	Export apparatus	ABV07255.1
T3SS-ETT2-Eiv-Epr	E. Coli	Esc-ETT2-EpaQ	Export apparatus	NP_311751.1
T3SS-ETT2-Eiv-Epr	E. Coli	Esc-ETT2-EpaR1|Esc-ETT2-EpaR2	Export apparatus	YP_001459636.1
T3SS-ETT2-Eiv-Epr	E. Coli	Esc-ETT2-EpaS	Export apparatus	EKH99126.1
T3SS-ETT2-Eiv-Epr	E. Coli	Esc-ETT2-EivA	Export apparatus	NP_311758.1
T3SS-ETT2-Eiv-Epr	E. Coli	Esc-ETT2-EivI	Needle assembly	NP_311756.1
T3SS-ETT2-Eiv-Epr	E. Coli	Esc-ETT2-EivJ	Needle assembly	NP_311754.1
T3SS-ETT2-Eiv-Epr	E. Coli	Esc-ETT2-OrgA	Needle assembly	EFJ74085.1
T3SS-ETT2-Eiv-Epr	E. Coli	Esc-ETT2-EprI	Needle major subunit	NP_311745.1
T3SS-ETT2-Eiv-Epr	E. Coli	Esc-ETT2-EprJ	Needle minor subunit	NP_311744.1
T3SS-ETT2-Eiv-Epr	E. Coli	Esc-ETT2-EivC	ATPase	BAB37153.1
T3SS-ETT2-Eiv-Epr	E. Coli			NP_311761.1
T3SS-ETT2-Eiv-Epr	E. Coli	Esc-ETT2-EivE	Effector export	NP_311759.1
T3SS-Hrp1	Erwinia	Erw-Hrp-HrcC	Outer membrane ring	YP_003529907.1
T3SS-Hrp1	Erwinia	Erw-Hrp-HrcJ	Inner membrane ring	YP_003529902.1
T3SS-Hrp1	Erwinia	Erw-Hrp-HrcQ	Cytoplasmic ring	YP_003529887.1
T3SS-Hrp1	Erwinia	Erw-Hrp-HrcR	Export apparatus	YP_003529886.1
T3SS-Hrp1	Erwinia	Erw-Hrp-HrcS	Export apparatus	YP_003529885.1
T3SS-Hrp1	Erwinia	Erw-Hrp-HrcT	Export apparatus	YP_003529884.1
T3SS-Hrp1	Erwinia	Erw-Hrp-HrcU	Export apparatus	YP_003529883.1

T3SS-Hrp1	Erwinia	Erw-Hrp-HrcV	Export apparatus	YP_003529892.1
T3SS-Hrp1	Erwinia	Erw-Hrp-HrpP	Needle assembly	YP_003529888.1
T3SS-Hrp1	Erwinia		Translocon	YP_003529905.1
T3SS-Hrp1	Erwinia		Translocon	YP_003529889.1
T3SS-Hrp1	Erwinia		Translocon	YP_003529891.1
T3SS-Hrp1	Erwinia		Translocon	YP_003529906.1
T3SS-Hrp1	Erwinia		Translocon	YP_003529908.1
T3SS-Hrp1	Erwinia	Erw-Hrp-HrpN	Translocon	YP_003529910.1
T3SS-Hrp1	Erwinia	Erw-Hrp-HrpW	Translocon	YP_003529914.1
T3SS-Hrp1	Erwinia		Translocon	YP_003529909.1
T3SS-Hrp1	Erwinia	Erw-Hrp-HrcN	ATPase	YP_003529890.1
T3SS-SPI-1	Erwinia	Erw-PAI2-PrgH1	Inner membrane ring	YP_003530133.1
T3SS-SPI-1	Erwinia	Erw-PAI2-InvA1	Export apparatus	YP_003530137.1
T3SS-SPI-1	Erwinia	Erw-PAI2-OrgAa	Needle assembly	YP_003530129.1
T3SS-SPI-1	Erwinia	Erw-PAI2-PrgI	Needle major subunit	YP_003530132.1
T3SS-SPI-1	Erwinia	Erw-PAI2-PrgJ1	Needle minor subunit	YP_003530933.1
T3SS-SPI-1	Erwinia		Translocon	YP_003530134.1
T3SS-SPI-1	Erwinia		Translocon	YP_003530147.1
T3SS-SPI-1	Erwinia	Erw-PAI2-SpaI1	ATPase	YP_003530139.1
T3SS-SPI-1	Erwinia	Erw-PAI2-InvE1	Effector export	YP_003530940.1
T3SS-SPI-1	Erwinia	Erw-PAI3-PrgH3	Inner membrane ring	YP_003530934.1
T3SS-SPI-1	Erwinia	Erw-PAI3-InvA3	Export apparatus	YP_003530938.1

T3SS-SPI-1	Erwinia		Needle major subunit	YP_003530948.1
T3SS-SPI-1	Erwinia			YP_003530935.1
T3SS-SPI-1	Salmonella	Sal-SPI1-InvG	Outer membrane ring	NP_461819.1
T3SS-SPI-1	Salmonella	Sal-SPI1-PrgK	Inner membrane ring	NP_461792.1
T3SS-SPI-1	Salmonella	Sal-SPI1-PrgH	Inner membrane ring	NP_461795.1
T3SS-SPI-1	Salmonella	Sal-SPI1-SpaO	Cytoplasmic ring	NP_461812.1
T3SS-SPI-1	Salmonella	Sal-SPI1-SpaP	Export apparatus	NP_461811.1
T3SS-SPI-1	Salmonella	Sal-SPI1-SpaQ	Export apparatus	NP_461810.1
T3SS-SPI-1	Salmonella	Sal-SPI1-SpaR	Export apparatus	NP_461809.1
T3SS-SPI-1	Salmonella	Sal-SPI1-SpaS	Export apparatus	NP_461808.1
T3SS-SPI-1	Salmonella	Sal-SPI1-InvA	Export apparatus	NP_461817.1
T3SS-SPI-1	Salmonella	Sal-SPI1-InvI	Needle assembly	NP_461814.1
T3SS-SPI-1	Salmonella	Sal-SPI1-InvJ	Needle assembly	NP_461813.1
T3SS-SPI-1	Salmonella	Sal-SPI1-OrgA	Needle assembly	NP_461791.1
T3SS-SPI-1	Salmonella	Sal-SPI1-OrgB	Needle assembly	NP_461790.1
T3SS-SPI-1	Salmonella	Sal-SPI1-PrgI	Needle major subunit	NP_461794.1
T3SS-SPI-1	Salmonella	Sal-SPI1-PrgJ	Needle minor subunit	NP_461793.1
T3SS-SPI-1	Salmonella	Sal-SPI1-SipB	Translocon	NP_461806.1
T3SS-SPI-1	Salmonella	Sal-SPI1-SipC	Translocon	NP_461805.1
T3SS-SPI-1	Salmonella	Sal-SPI1-SipD	Translocon	NP_461804.1
T3SS-SPI-1	Salmonella	Sal-SPI1-InvC	ATPase	NP_460380.1
T3SS-SPI-2	Salmonella	Sal-SPI2-SsaC	Outer membrane ring	NP_460359.1

T3SS-SPI-2	Salmonella	Sal-SPI2-SsaJ	Inner membrane ring	NP_460374.1
T3SS-SPI-2	Salmonella	Sal-SPI2-SsaD	Inner membrane ring	NP_460360.1
T3SS-SPI-2	Salmonella	Sal-SPI2-SsaQ	Cytoplasmic ring	NP_460383.1
T3SS-SPI-2	Salmonella	Sal-SPI2-SsaR	Export apparatus	NP_460384.1
T3SS-SPI-2	Salmonella	Sal-SPI2-SsaS	Export apparatus	NP_460385.1
T3SS-SPI-2	Salmonella	Sal-SPI2-SsaT	Export apparatus	NP_460386.1
T3SS-SPI-2	Salmonella	Sal-SPI2-SsaU	Export apparatus	NP_460387.1
T3SS-SPI-2	Salmonella	Sal-SPI2-SsaV	Export apparatus	NP_460379.1
T3SS-SPI-2	Salmonella	Sal-SPI2-SsaO	Needle assembly	NP_460381.1
T3SS-SPI-2	Salmonella	Sal-SPI2-SsaP	Needle assembly	NP_460382.1
T3SS-SPI-2	Salmonella	Sal-SPI2-SsaK	Needle assembly	NP_460376.1
T3SS-SPI-2	Salmonella	Sal-SPI2-SsaL	Needle assembly	NP_460377.1
T3SS-SPI-2	Salmonella			NP_460378.1
T3SS-SPI-2	Salmonella	Sal-SPI2-SsaG	Needle major subunit	NP_460371.1
T3SS-SPI-2	Salmonella			NP_460372.1
T3SS-SPI-2	Salmonella	Sal-SPI2-SsaI	Needle minor subunit	NP_460373.1
T3SS-SPI-2	Salmonella	Sal-SPI2-SseC	Translocon	NP_460365.1
T3SS-SPI-2	Salmonella	Sal-SPI2-SseD	Translocon	NP_460366.1
T3SS-SPI-2	Salmonella	Sal-SPI2-SseB	Translocon	NP_460363.1
T3SS-Mxi	Shigella	Shi-MxiD	Outer membrane ring	NP_858278.1
T3SS-Mxi	Shigella	Shi-MxiJ	Inner membrane ring	NP_858272.1
T3SS-Mxi	Shigella	Shi-MxiG	Inner membrane ring	NP_858269.1

T3SS-Mxi	Shigella	Shi-Spa33	Cytoplasmic ring	NP_858285.1
T3SS-Mxi	Shigella	Shi-Spa9	Export apparatus	NP_858287.1
T3SS-Mxi	Shigella	Shi-Spa29	Export apparatus	NP_858288.2
T3SS-Mxi	Shigella	Shi-Spa40	Export apparatus	NP_858289.1
T3SS-Mxi	Shigella	Shi-MxiA	Export apparatus	NP_858280.1
T3SS-Mxi	Shigella	Shi-Spa13	Needle assembly	NP_858283.1
T3SS-Mxi	Shigella	Shi-Spa32	Needle assembly	NP_858284.1
T3SS-Mxi	Shigella	Shi-MxiK	Needle assembly	NP_858273.1
T3SS-Mxi	Shigella	Shi-MxiN	Needle assembly	NP_858274.2
T3SS-Mxi	Shigella	Shi-MxiH	Needle major subunit	NP_858270.1
T3SS-Mxi	Shigella	Shi-MxiI	Needle minor subunit	NP_858271.1
T3SS-Mxi	Shigella	Shi-IpaB	Translocon	NP_858261.1
T3SS-Mxi	Shigella	Shi-IpaC	Translocon	NP_858260.2
T3SS-Mxi	Shigella	Shi-IpaD	Translocon	NP_858259.1
T3SS-Mxi	Shigella	Shi-Spa47	ATPase	NP_858282.1
T3SS-Mxi	Shigella	Shi-MxiC	Effector export	NP_858279.1
T3SS-Ysc	Yersinia		Outer membrane ring	NP_395185.1
T3SS-Ysc	Yersinia	Yer-Ysc-YscJ	Inner membrane ring	YP_002345344.1
T3SS-Ysc	Yersinia	Yer-Ysc-YscQ	Cytoplasmic ring	NP_395177.1
T3SS-Ysc	Yersinia	Yer-Ysc-YscR	Export apparatus	YP_002345352.1
T3SS-Ysc	Yersinia			NP_395178.1
T3SS-Ysc	Yersinia	Yer-Ysc-YscS	Export apparatus	YP_002345353.1

T3SS-Ysc	Yersinia	Yer-Ysc-YscT	Export apparatus	YP_002345354.1
T3SS-Ysc	Yersinia			NP_395190.1
T3SS-Ysc	Yersinia	Yer-Ysc-YscO	Needle assembly	NP_395175.1
T3SS-Ysc	Yersinia			NP_395192.1
T3SS-Ysc	Yersinia	Yer-Ysc-YscK	Needle assembly	NP_395194.1
T3SS-Ysc	Yersinia	Yer-Ysc-YscL	Needle assembly	YP_002345346.1
T3SS-Ysc	Yersinia			NP_395195.1
T3SS-Ysc	Yersinia			NP_395188.1
T3SS-Ysc	Yersinia			YP_002345339.1
T3SS-Ysc	Yersinia	Yer-Ysc-YopB	Translocon	NP_395163.1
T3SS-Ysc	Yersinia	Yer-Ysc-YopD	Translocon	NP_395162.1
T3SS-Ysc	Yersinia	Yer-Ysc-LcrV	Translocon	NP_395165.1
T3SS-Ysc	Yersinia		Translocon	YP_002345349.1
T3SS-Ysc	Yersinia			NP_395176.1
T3SS-Ysc	Yersinia	Yer-Ysc-YscN	ATPase	NP_395174.1
T3SS-Ysc	Yersinia	Yer-Ysa-YsaC	Outer membrane ring	YP_002345337.1
T3SS-Ysc	Yersinia	Yer-Ysa-YsaQ	Cytoplasmic ring	YP_002345351.1
T3SS-Ysc	Yersinia	Yer-Ysa-YsaI	Needle minor subunit	YP_002345343.1
T3SS-Ysc	Yersinia		Translocon	YP_002345341.1
T3SS-Ysc	Yersinia		Translocon	YP_002345342.1
T3SS-Ysc	Yersinia	Yer-Ysa-YsaN	ATPase	YP_002345348.1
T3SS-Ch-Ysc	Chlamydia	Chldo-YscR	Export apparatus	NP_220077.1

T3SS-Ch-Ysc	Chlamydia	Chldo-YscU	Export apparatus	NP_219594.1
T3SS-Ch-Ysc	Chlamydia	Chldo-YscO	Needle assembly	NP_220191.1
T3SS-Ch-Ysc	Chlamydia	Chldo-YscL	Needle assembly	NP_220076.1
T3SS-Ch-Ysc	Chlamydia		Translocon	NP_219563.1
T3SS-Ch-Ysc	Chlamydia		Translocon	NP_220238.1
T3SS-Ch-Ysc	Chlamydia			NP_220236.1
T3SS-Ch-Ysc	Chlamydia	Chldo-YscN	ATPase	NP_220188.1
T3SS-Ch-Ysc	Chlamydia			NP_220384.1

Table V-S3 : Results of SH tests comparing phylogenies of T3SS coding genes in *Xanthomonas*

		H1 : topologies alternatives															
		hrc_concat	hrcC	hrcD	hrcJ	hrcN	hrcQ	hrcR	hrcS	hrcT	hrcU	hrcV	hrpE	hrpB1	hrpB2	hrpB4	hrpB7
H0: alignement	hrc_concat	0,96034	0	0,88536	0,04023	0	0,00083	0,16132	0,00578	0,20897	0,0206	0,00684	0,02441	0,03873	0,11108	0,3825	0,01626
	hrcC	0,37406	0,91395	0,34218	0,00832	0,52414	0,72117	0,65727	0,00259	0,26576	0,0002	0,46417	0	0,94543	0,89079	0,82633	0,7081
	hrcD	0,00025	0	0,94864	0,43053	0	0,04018	0,00005	0	0,00233	0	0,12352	0	0	0,41333	0,3802	0,00036
	hrcJ	0,05185	0	0	0,92988	0,07593	0	0,02137	0,04262	0	0,29572	0,43796	0	0,29101	0,73003	0,46092	0,02644
	hrcN	0	0,00001	0,00012	0,05916	0,91767	0,23817	0,03422	0,33393	0,0009	0,00322	0,27644	0	0,53221	0,42616	0,16445	0,09918
	hrcQ	0	0	0,85718	0,94507	0,9462	0,95422	0,69799	0,1154	0	0,08249	0,8301	0	0,76662	0,75029	0,90246	0,78262
	hrcR	0	0,01155	0,19194	0,35578	0,66565	0,28707	0,90366	0,27447	0,73229	0,08124	0,21551	0	0,32863	0,97158	0,02542	0,55015
	hrcS	0	0	0,00005	0,00029	0,00041	0,05747	0,0254	0,76005	0,0018	0,11881	0,01921	0	0,00016	0,00014	0,00008	0,0001
	hrcT	0	0	0,03912	0,00017	0,00778	0,06186	0,0903	0,17226	0,90986	0,30794	0,51963	0,00227	0,66663	0,19137	0,80863	0,72965
	hrcU	0	0	0,00026	0,03987	0,00043	0	0,00025	0,92947	0,00792	0,8608	0,03076	0,92051	0,62439	0,20338	0,0045	0,00021
	hrcV	0	0	0,46891	0,13923	0,06054	0,03686	0,04388	0,49159	0,00934	0,00053	0,94393	0,05171	0,02328	0,00002	0,91582	0,46595
	hrpE	0	0,00187	0,01405	0	0,02687	0,00386	0	0,00001	0,00011	0,02048	0	0,8257	0,01079	0,08082	0	0,00646
	hrpB1	0	0,00572	0,03533	0,48392	0,2558	0,04064	0,30653	0,00634	0,00264	0,19341	0,01187	0	0,92934	0,11034	0,50696	0,68073
	hrpB2	0	0	0,06997	0,956	0,00082	0,02319	0,70404	0,00021	0,00054	0,92429	0,97765	0,03508	0,00551	0,96611	0,9265	0,62853
	hrpB4	0	0	0	0,95477	0,28745	0,00553	0,03089	0	0	0,02795	0,00873	0	0,44886	0,0748	0,86887	0,29673
	hrpB7	0	0	0,00349	0,00318	0,38469	0,6936	0,81701	0,90497	0,02205	0,56477	0,00167	0	0,67116	0,94118	0,01916	0,9435

Table V-S4 : T3E repertoires in *X. arboricola*

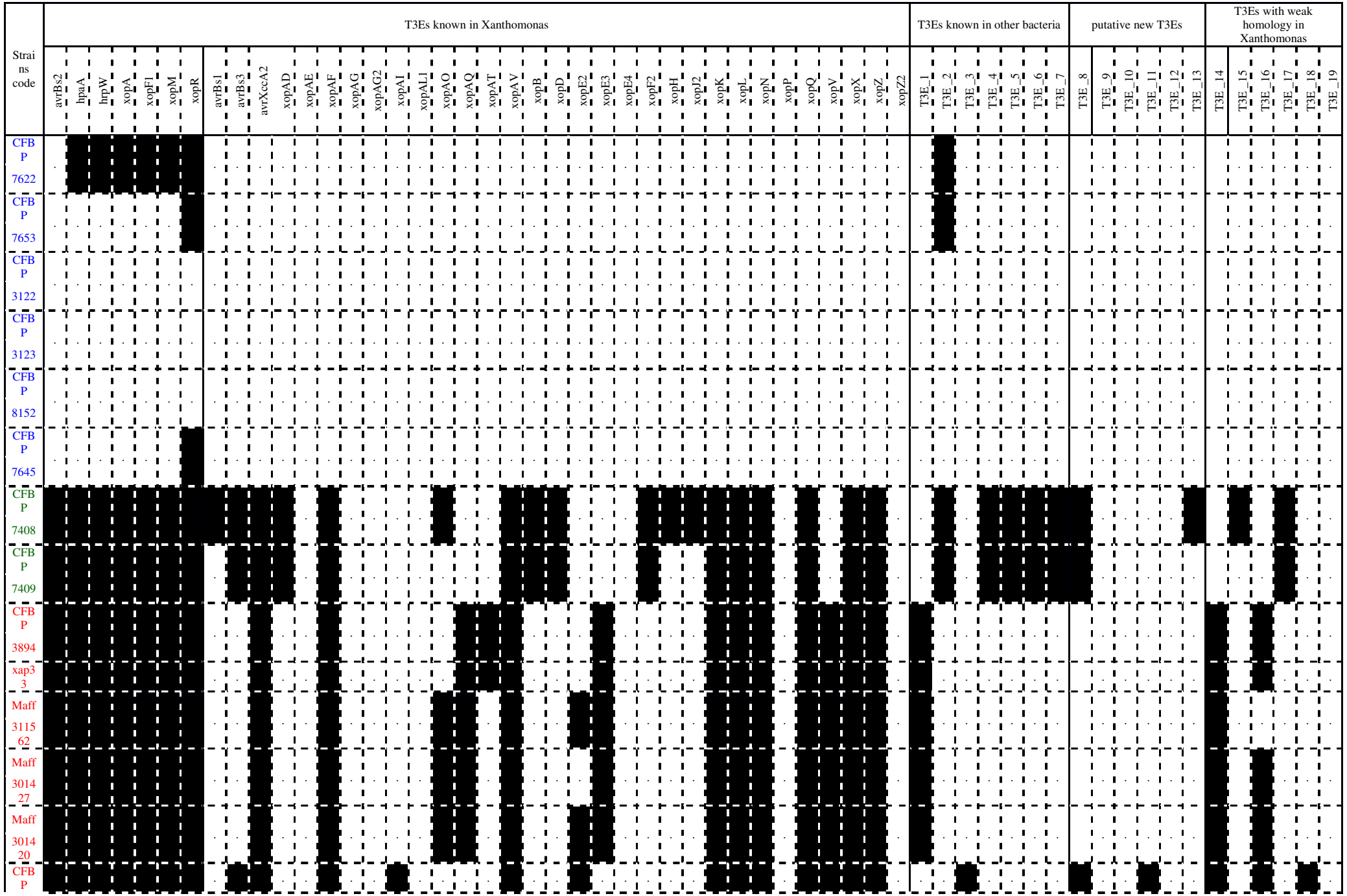


Table V-S4 : T3E repertoires in *X. arboricola*

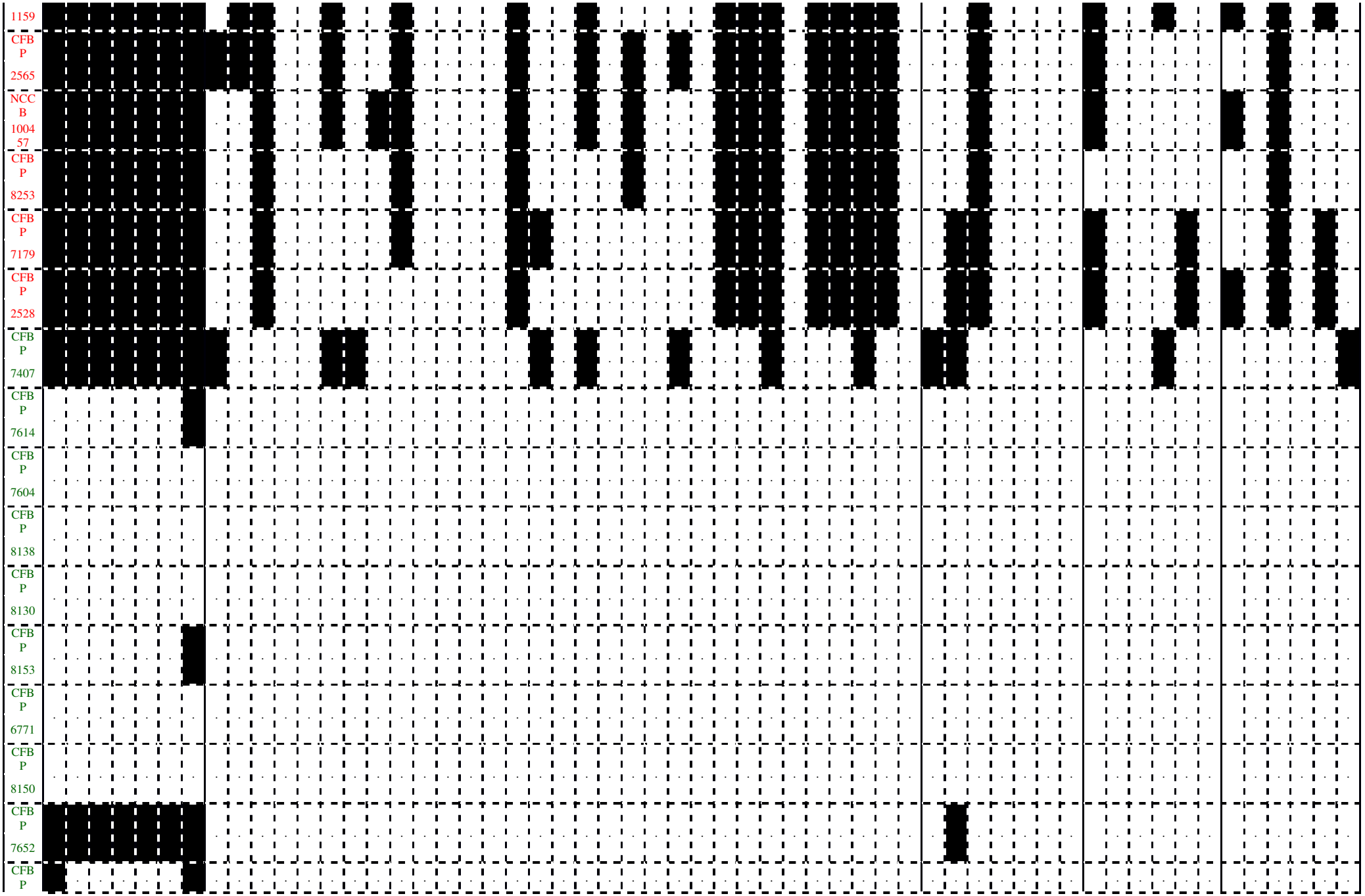


Table V-S4 : T3E repertoires in *X. arboricola*

Fig V-S1 : Genomic environnements of T3SS cluster

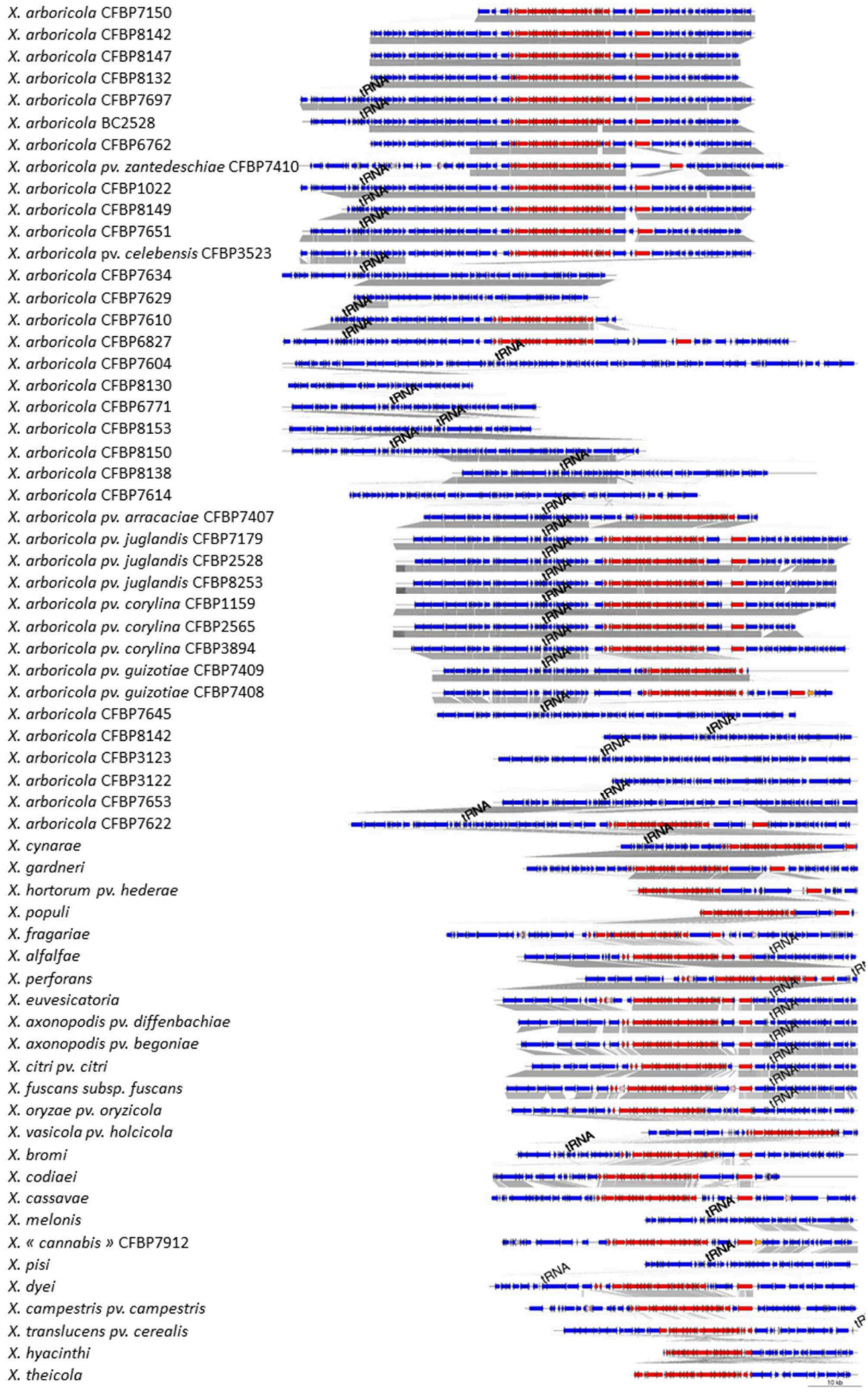


Fig V-S1 : Genomic environnements of T3SS cluster

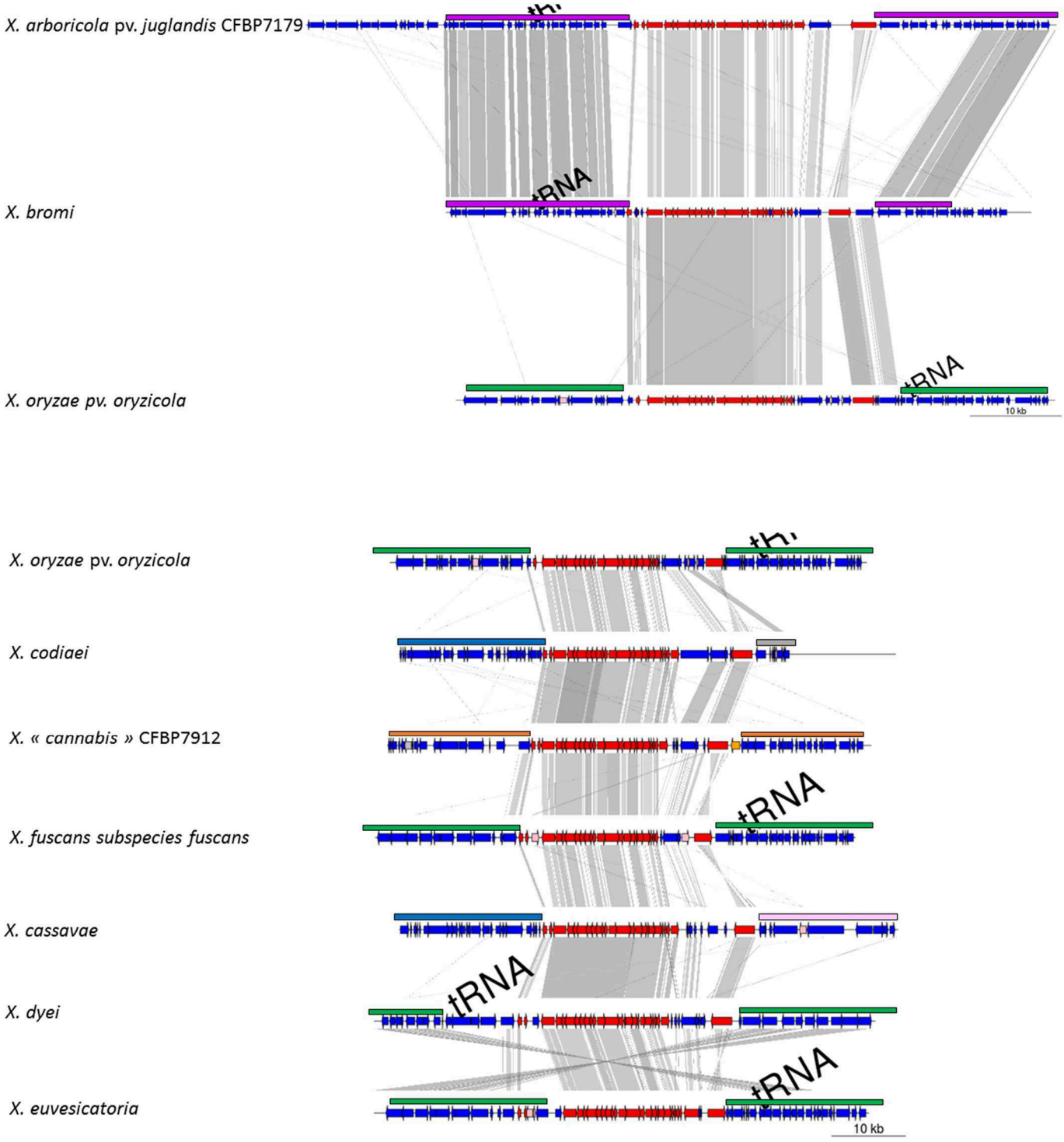


Fig V-S1 : Genomic environnements of T3SS cluster

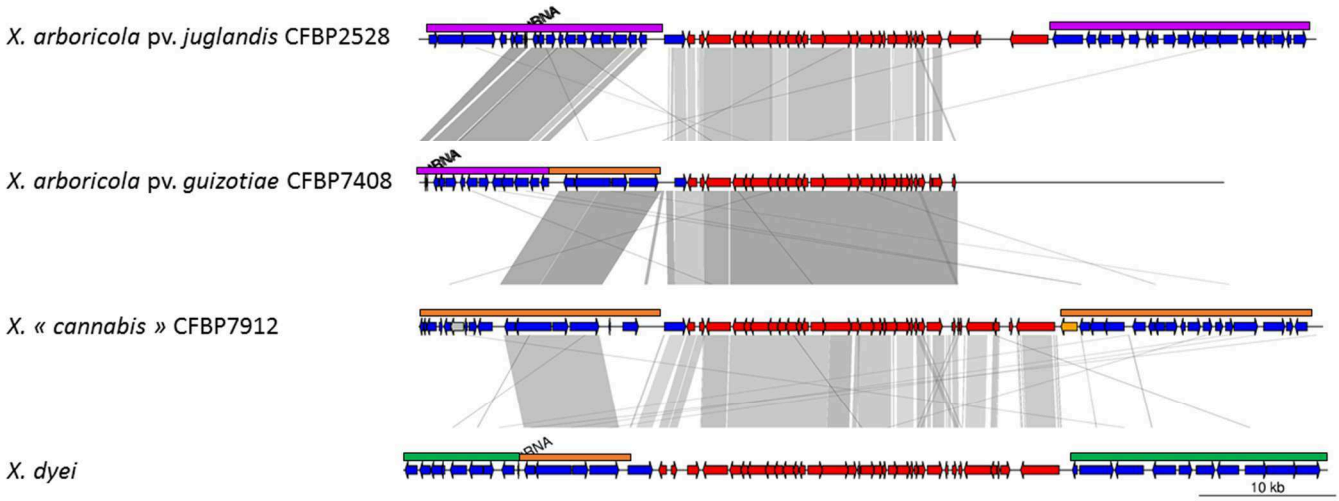
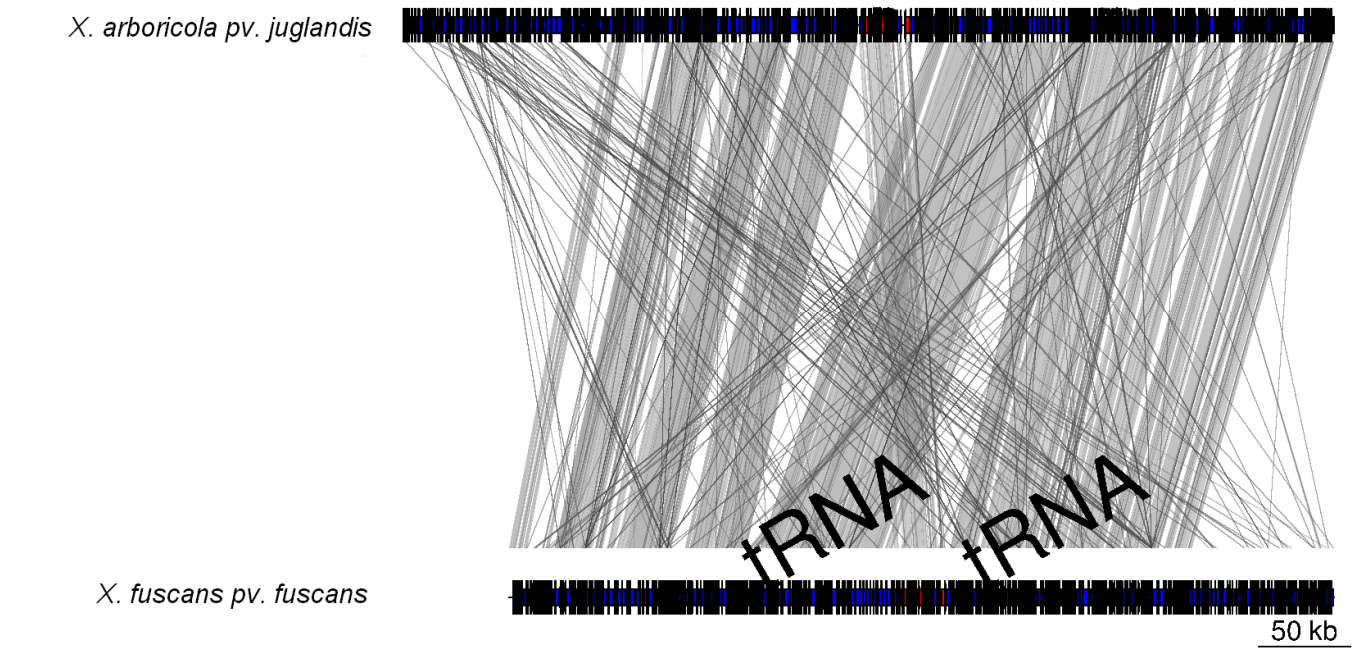


Fig V-S2 : Genomic environnements of a large region of T3SS cluster



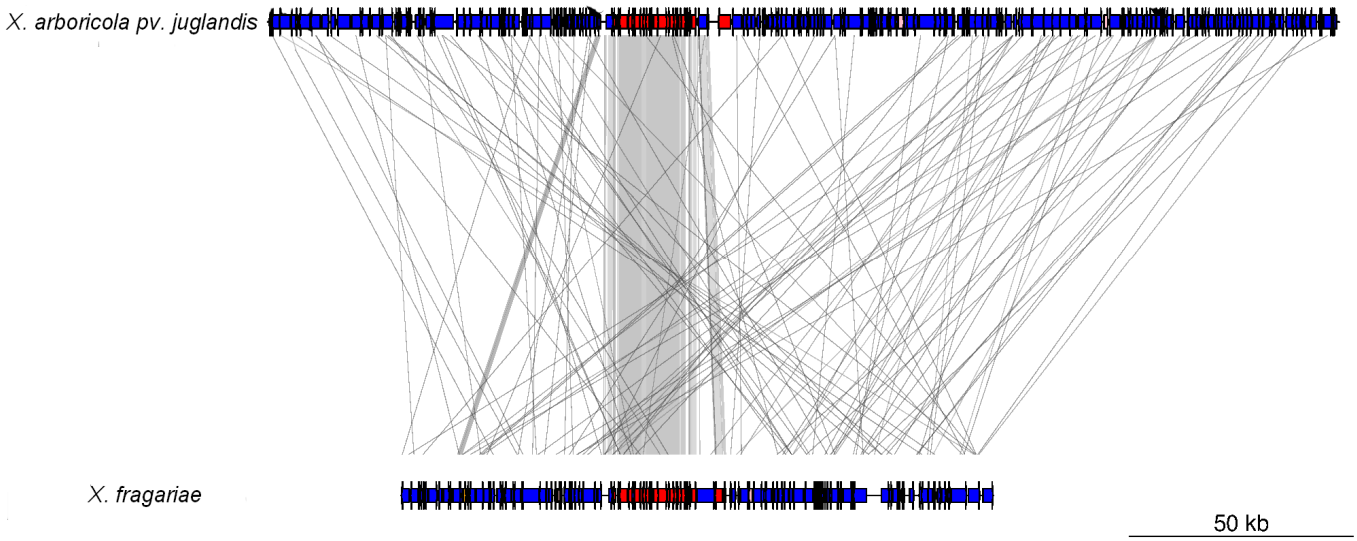
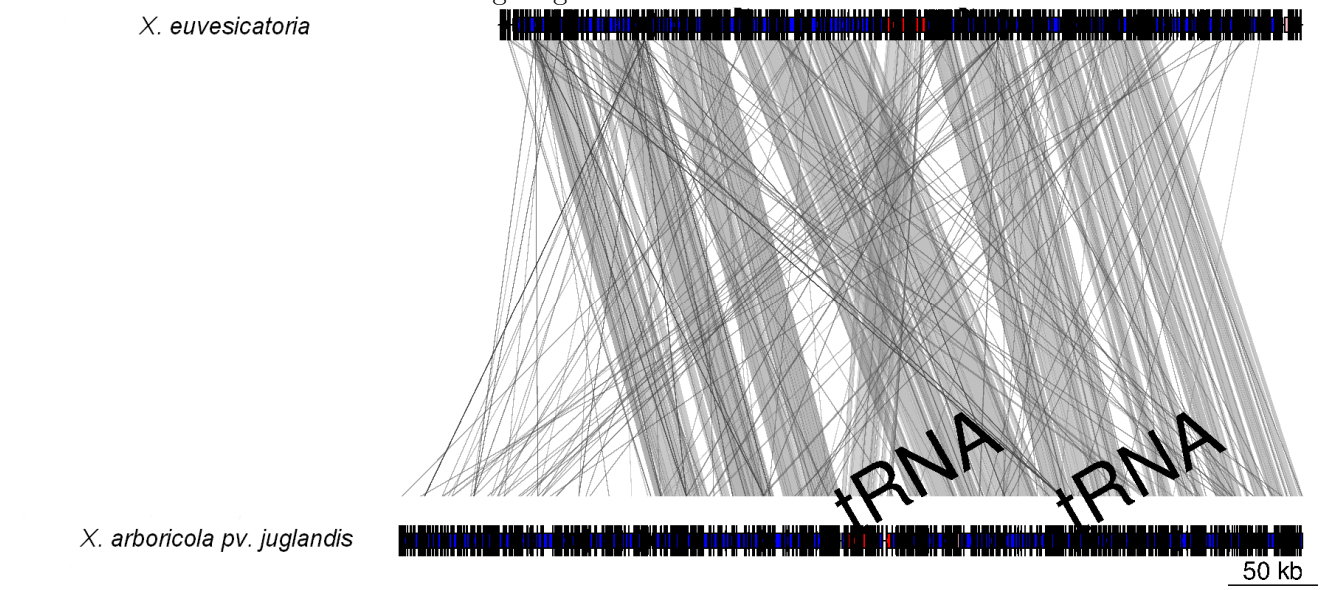
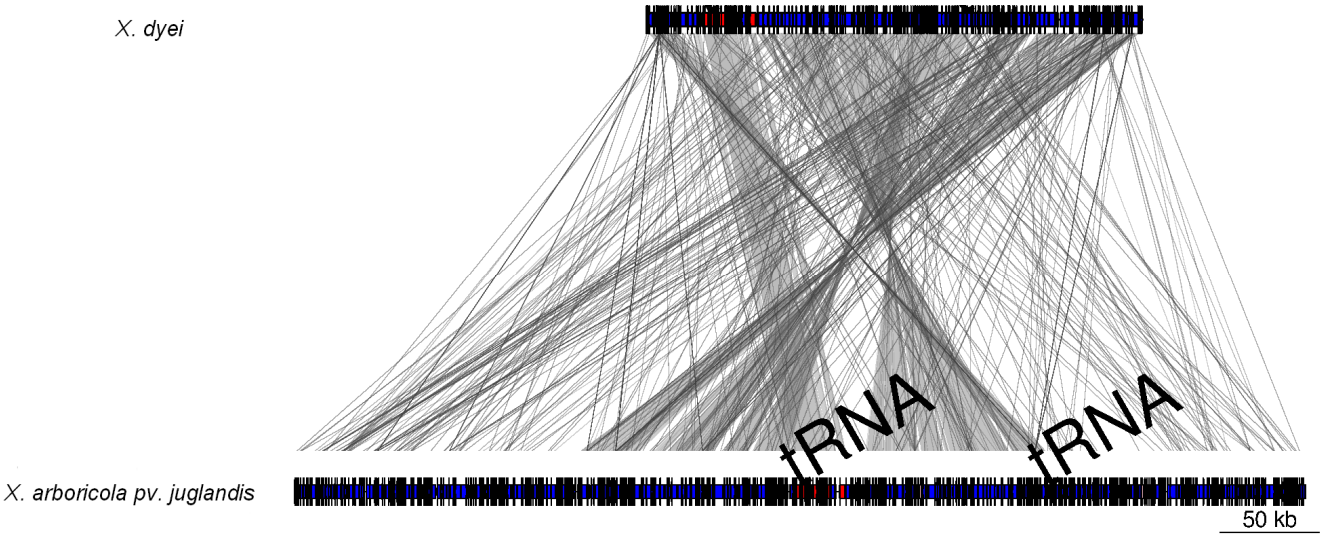


Fig V-S2 : Genomic environnements of a large region of T3SS cluster





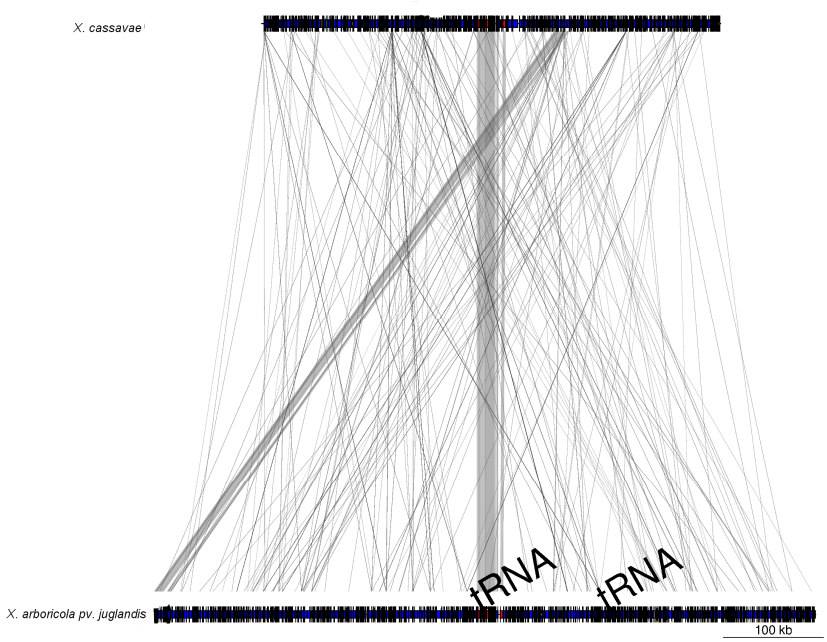


Fig V-S2 : Genomic environnements of a large region of T3SS cluster

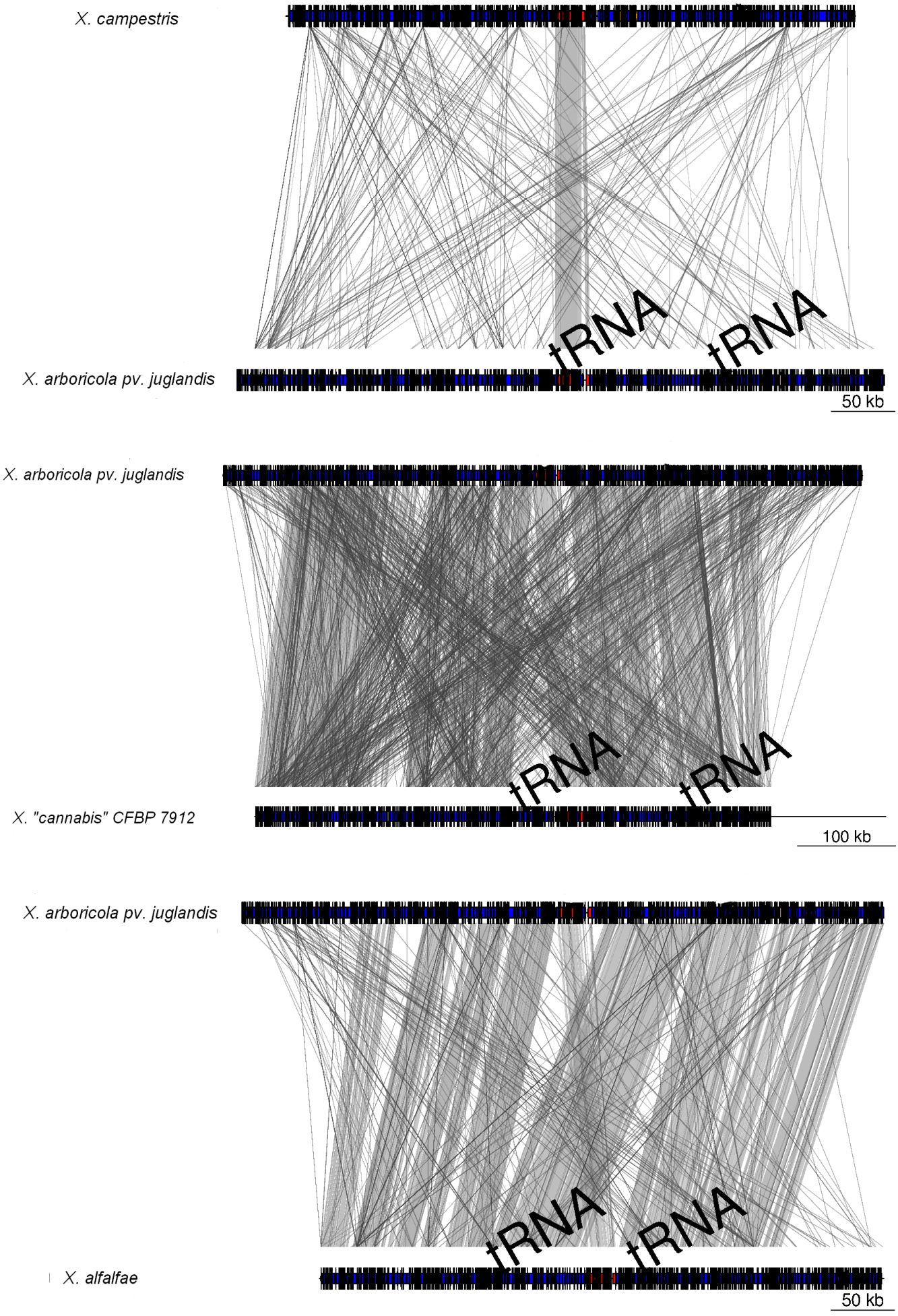


Fig V-S4 : Genomics environments *avrBs2*

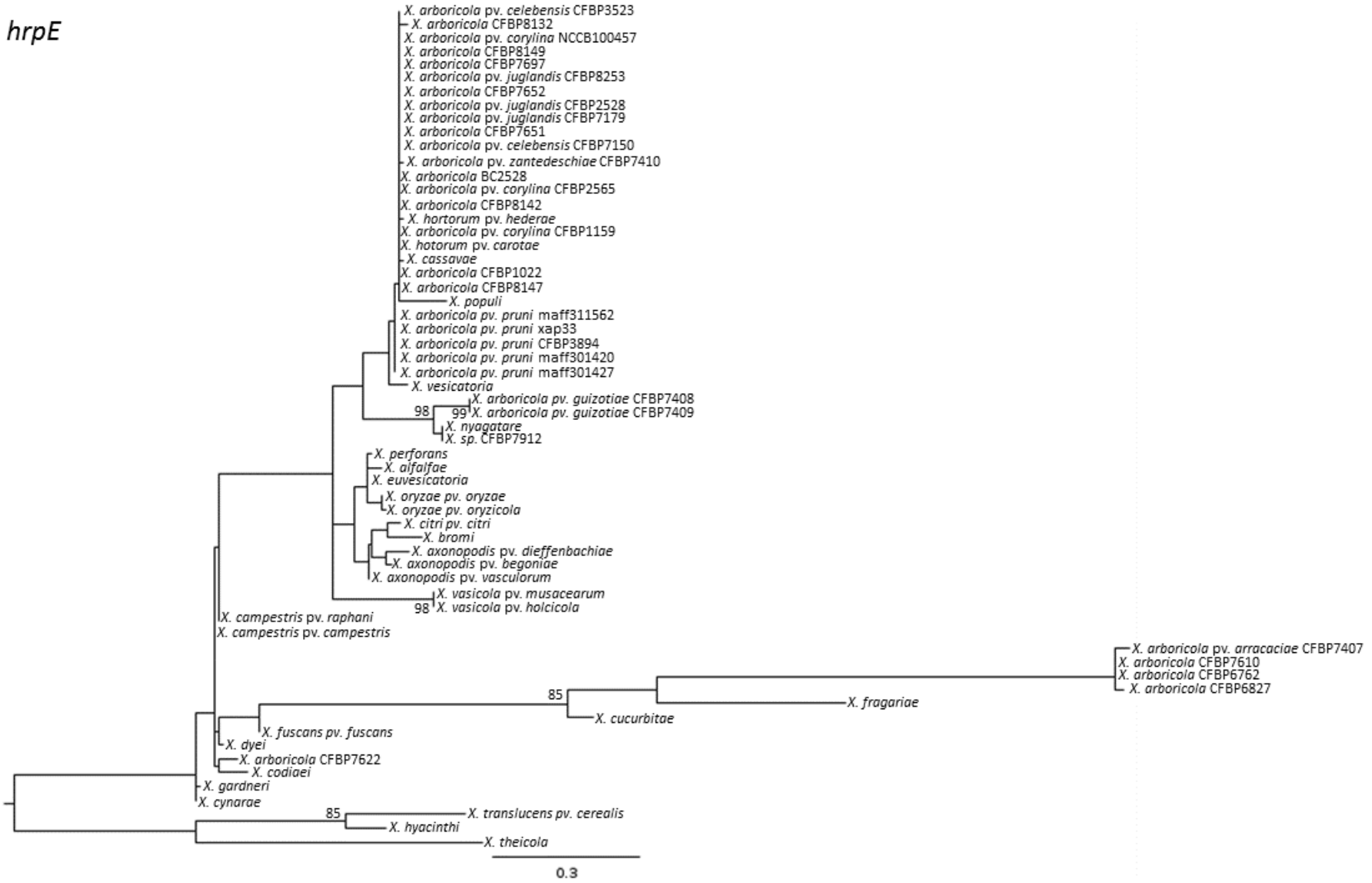


Fig V-S4 : Genomics environments *avrBs2*

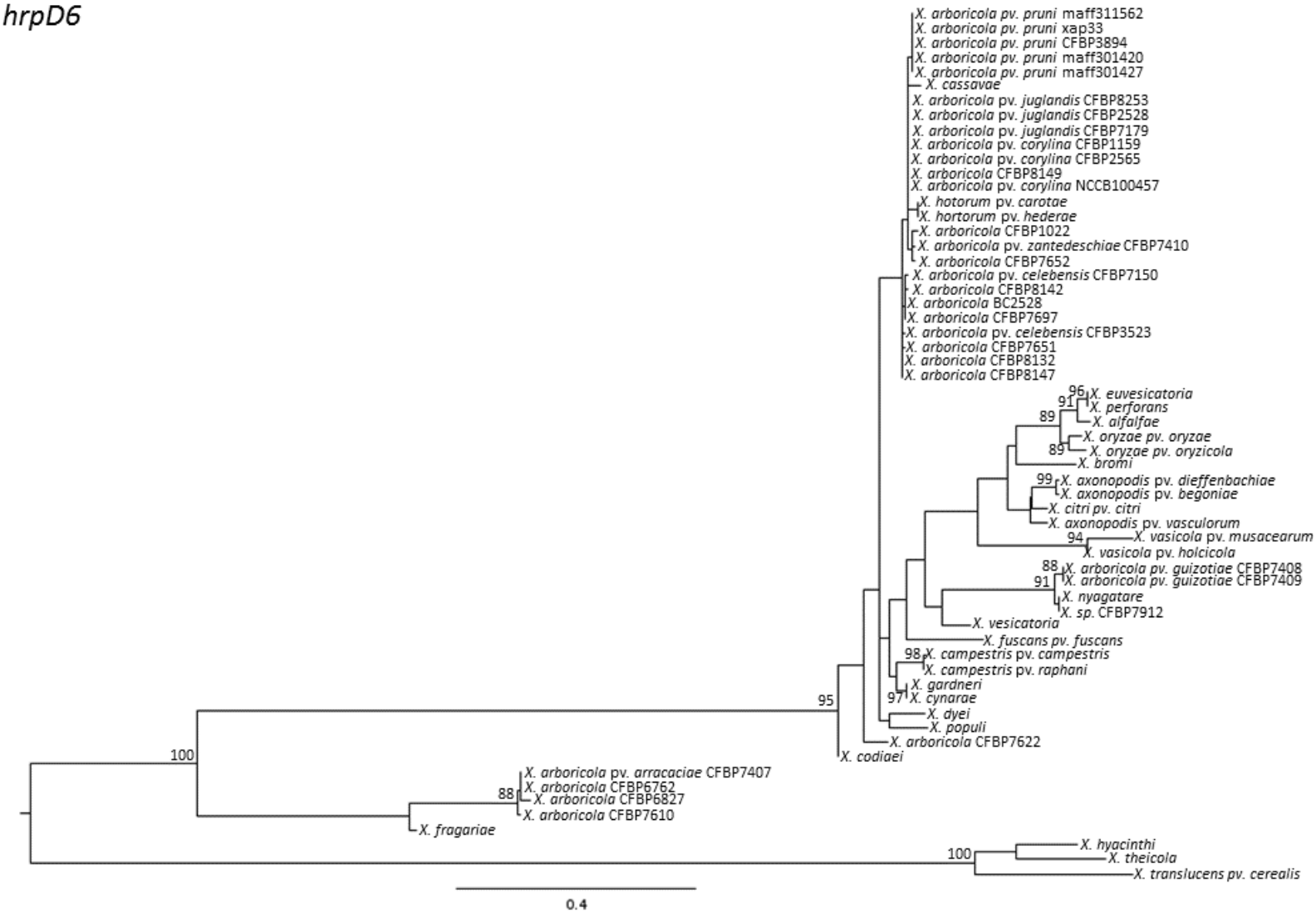


Fig V-S4 : Genomics environments *avrBs2*

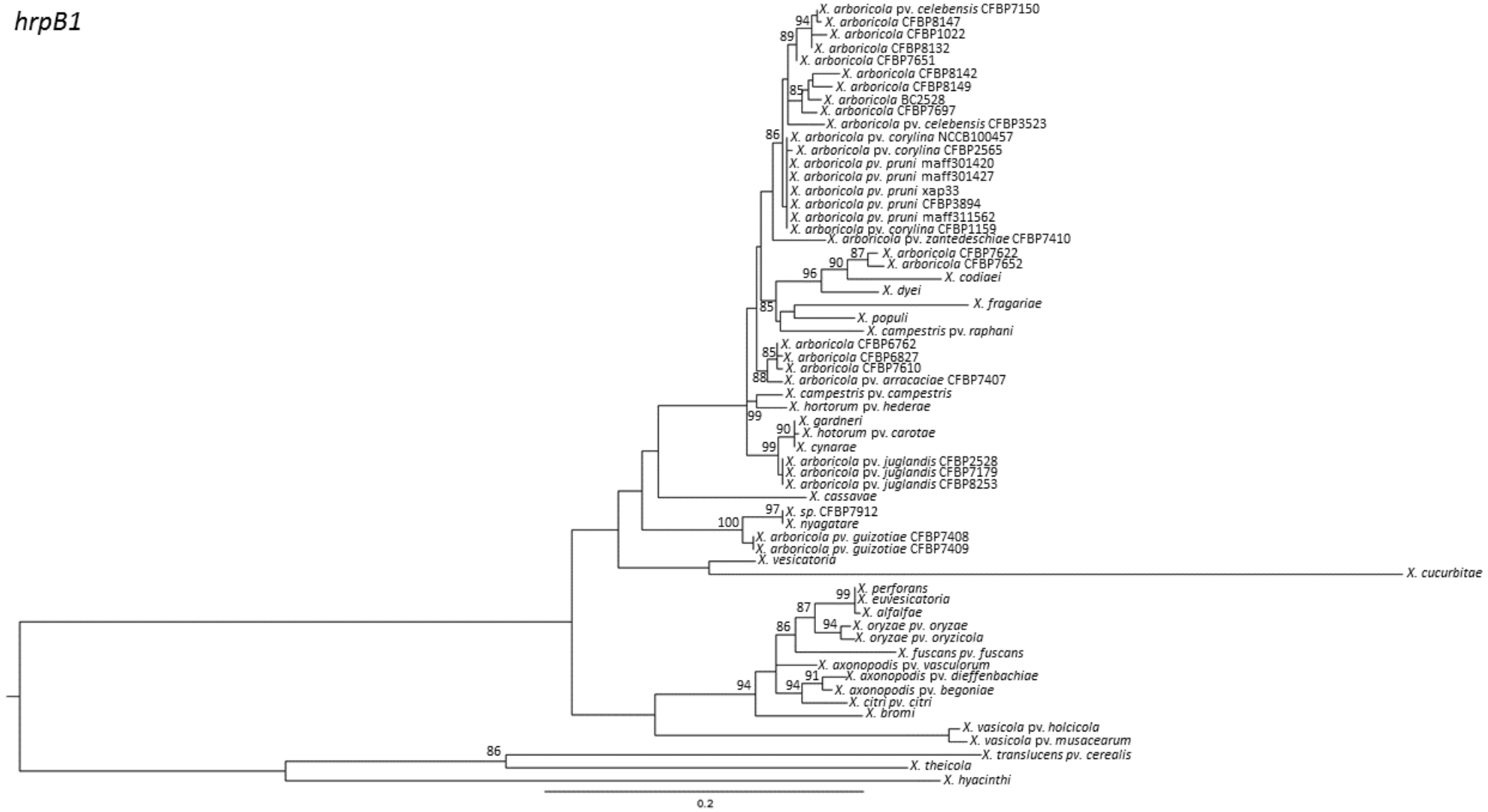


Fig V-S4 : Genomics environments *avrBs2*

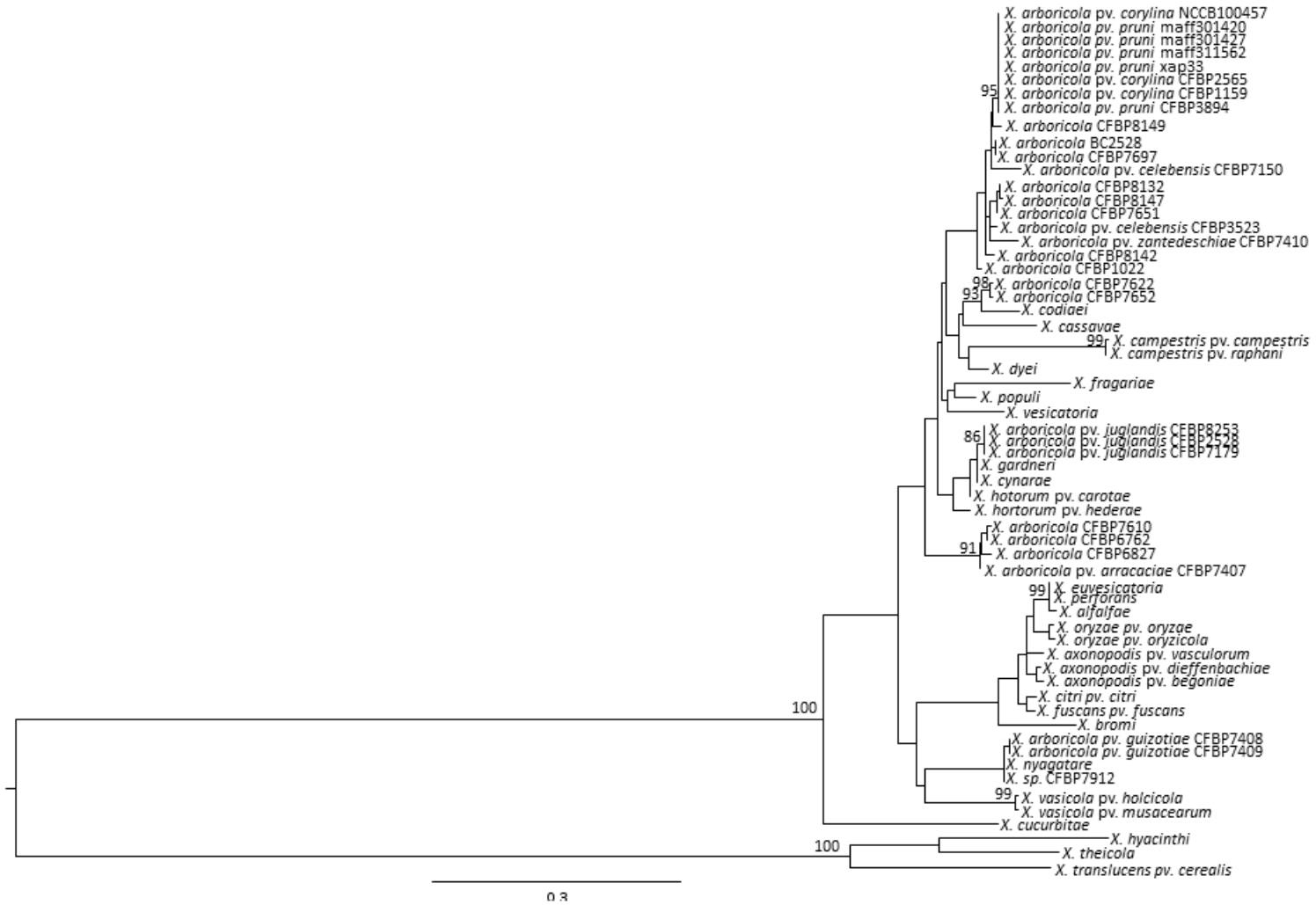


Fig V-S4 : Genomics environments *avrBs2*

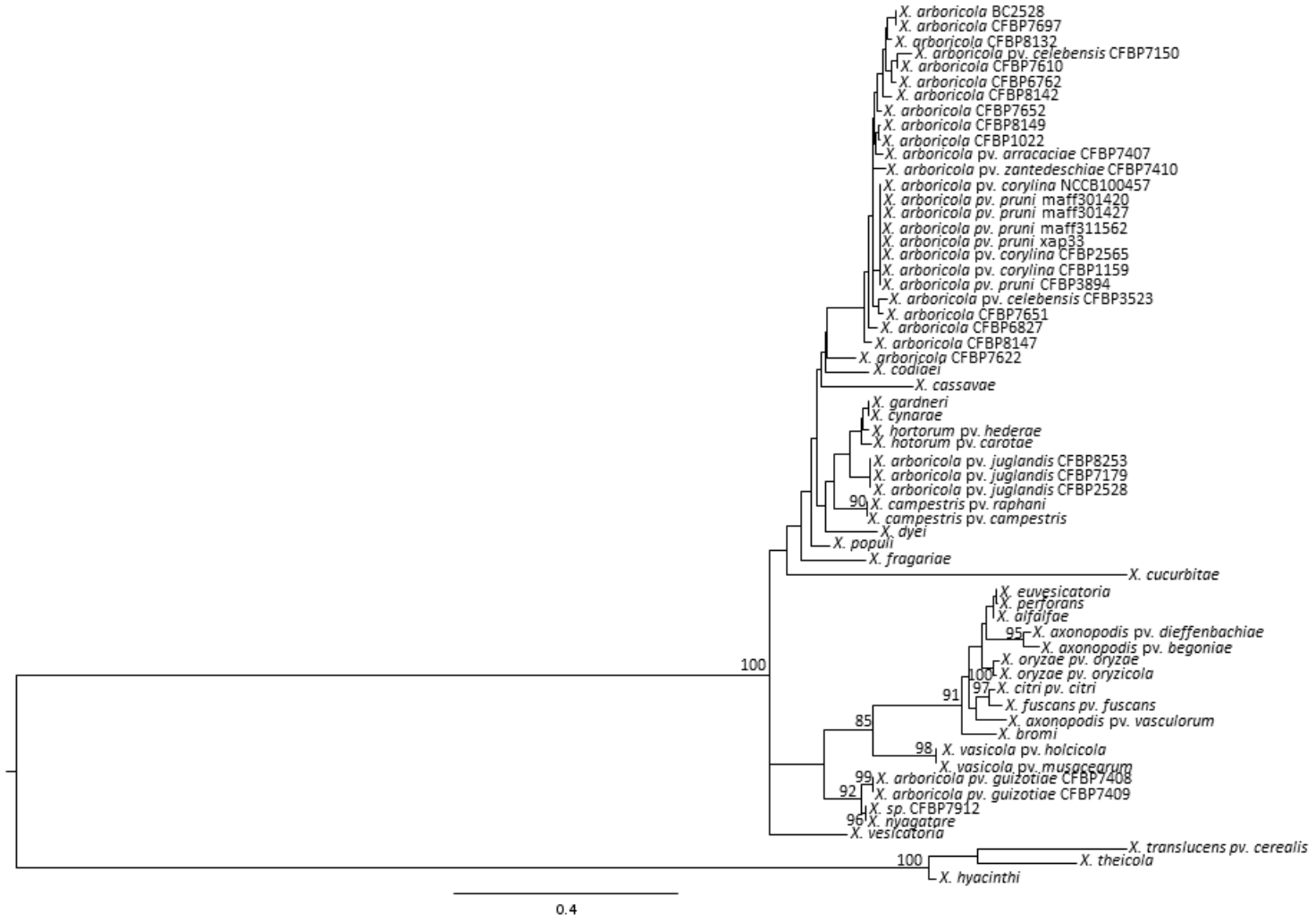


Fig V-S4 : Genomics environments *avrBs2*

Fig V-S4 : Genomics environments *avrBs2*

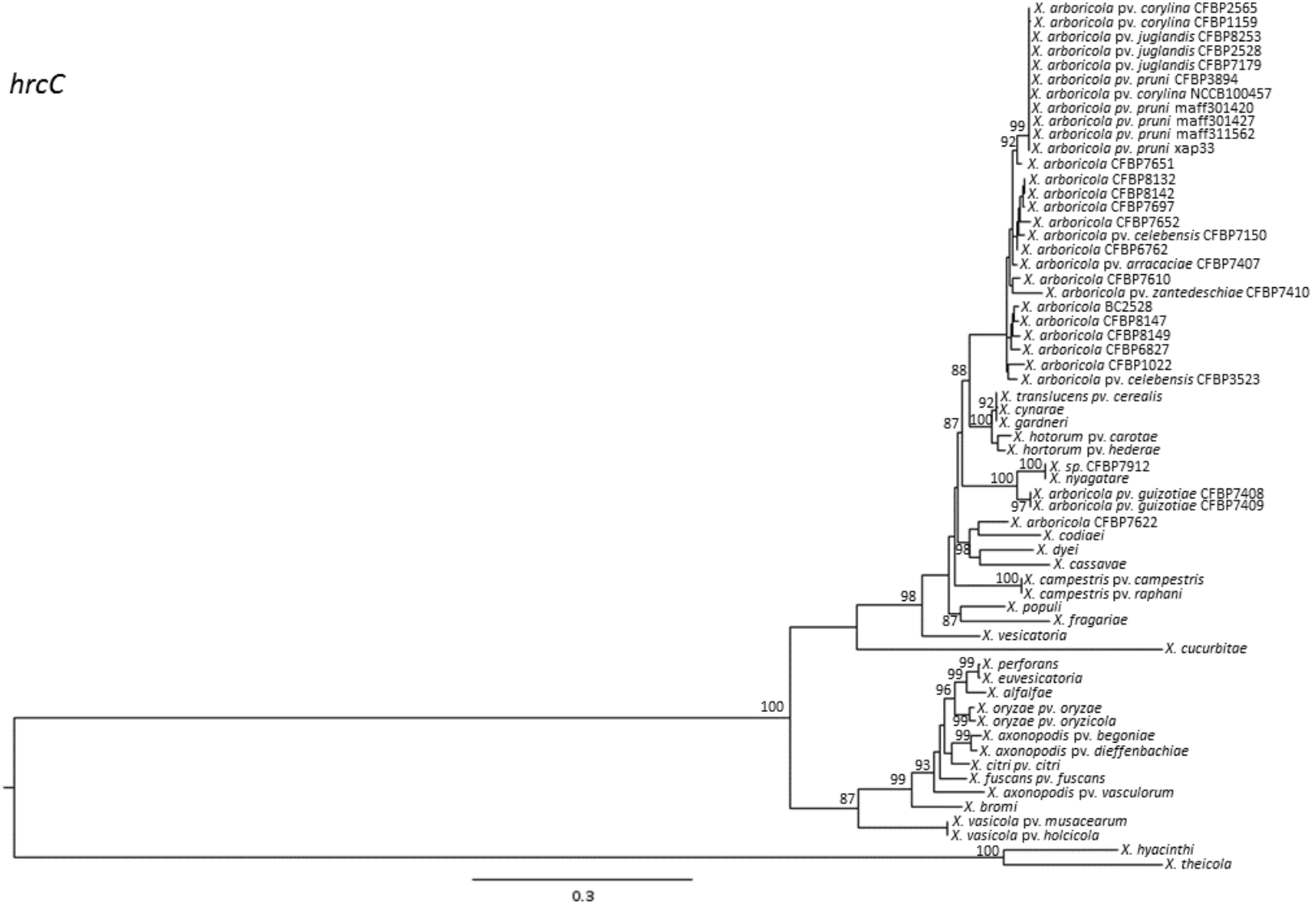


Fig V-S4 : Genomics environments *avrBs2*

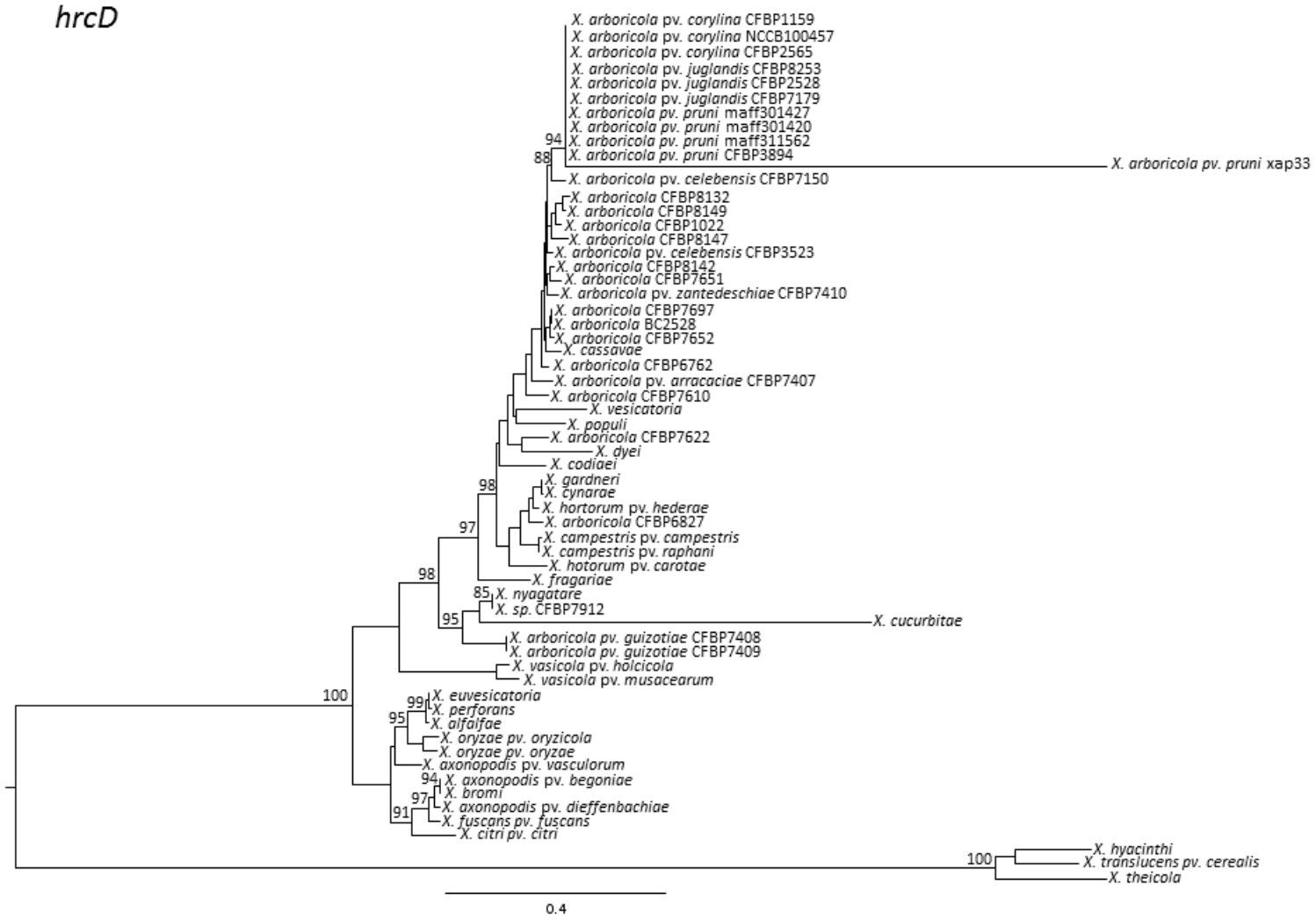


Fig V-S4 : Genomics environments *avrBs2*

Fig V-S4 : Genomics environments *avrBs2*

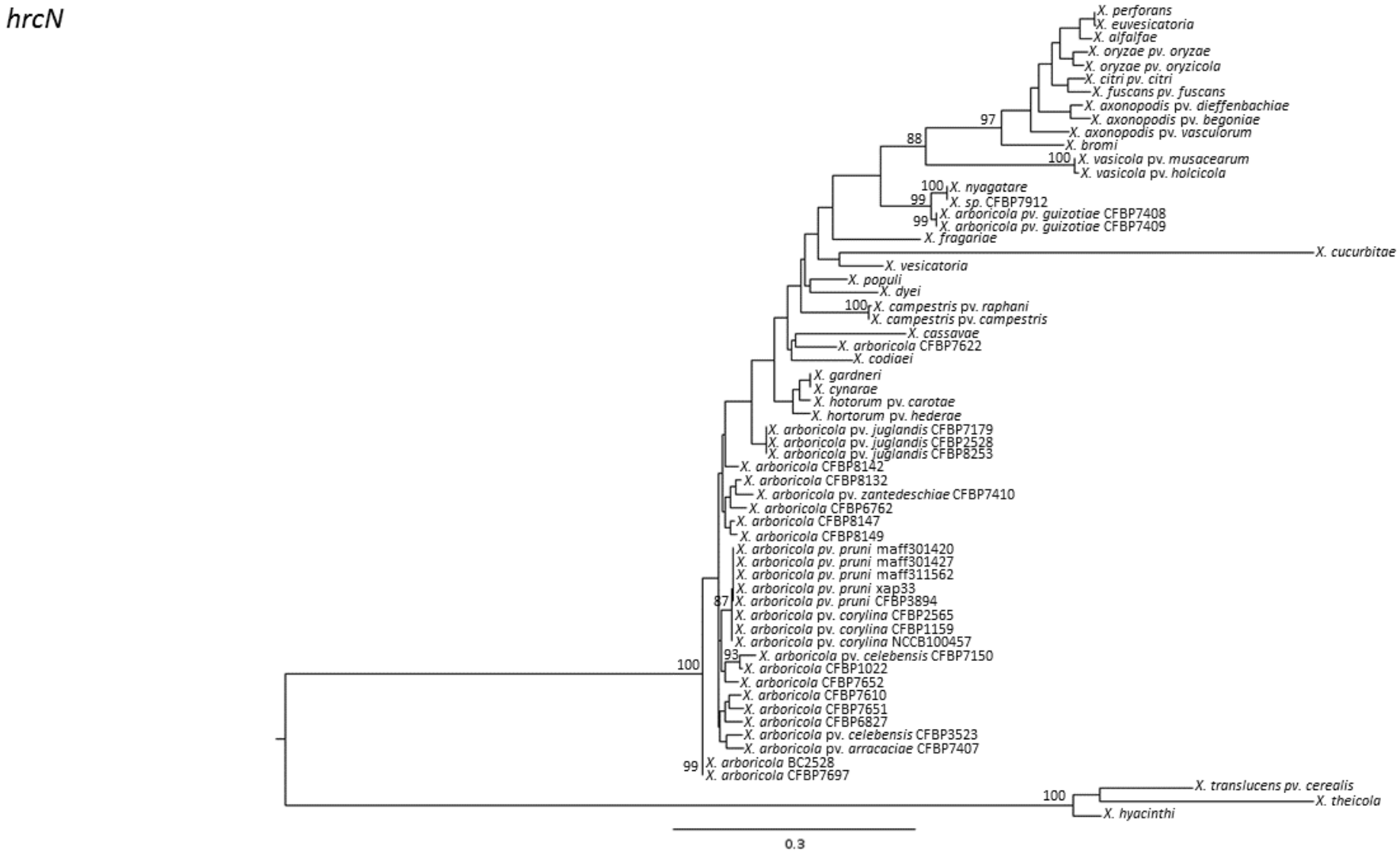


Fig V-S4 : Genomics environments *avrBs2*

Fig V-S4 : Genomics environments *avrBs2*

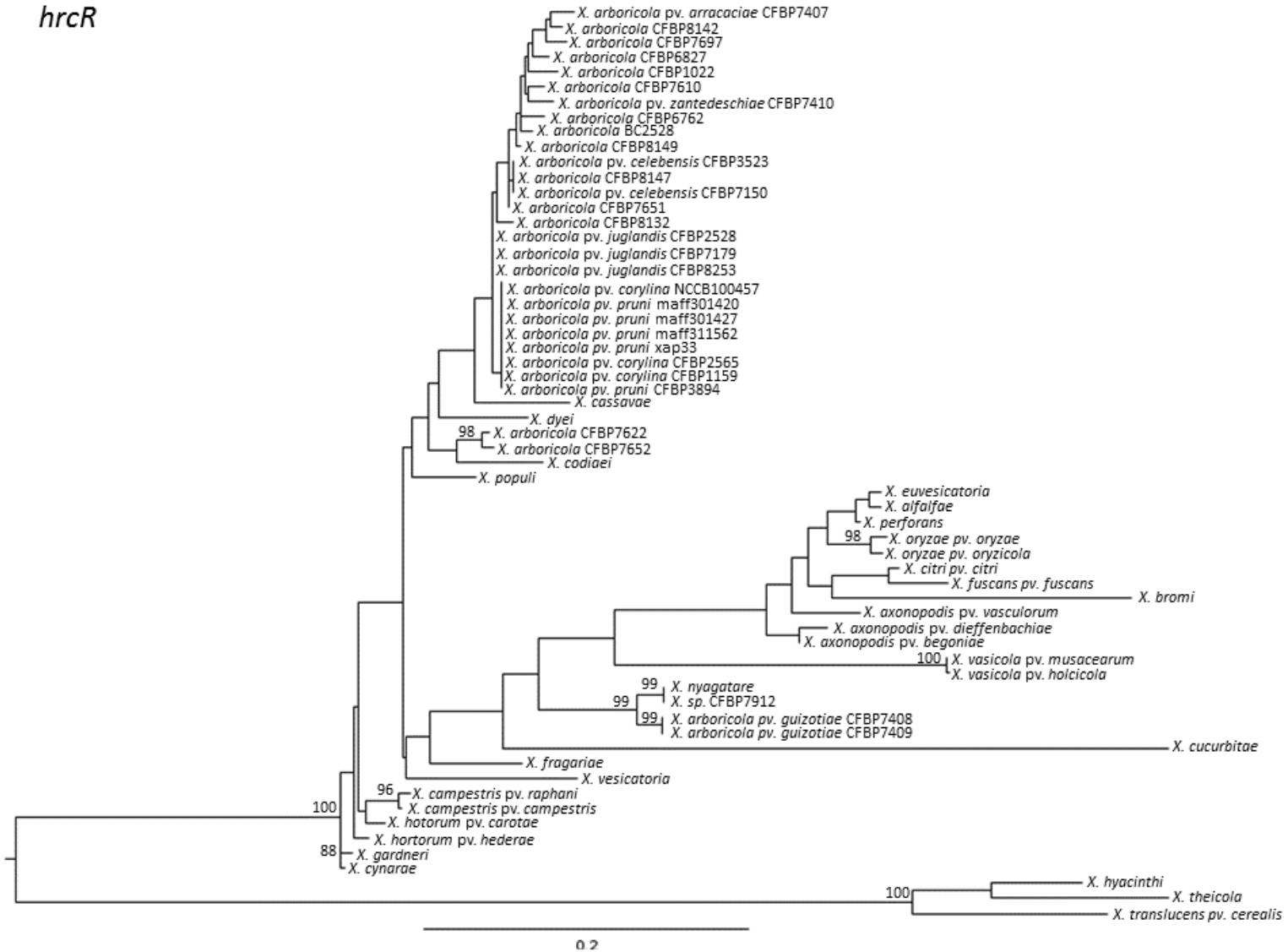


Fig V-S4 : Genomics environments *avrBs2*

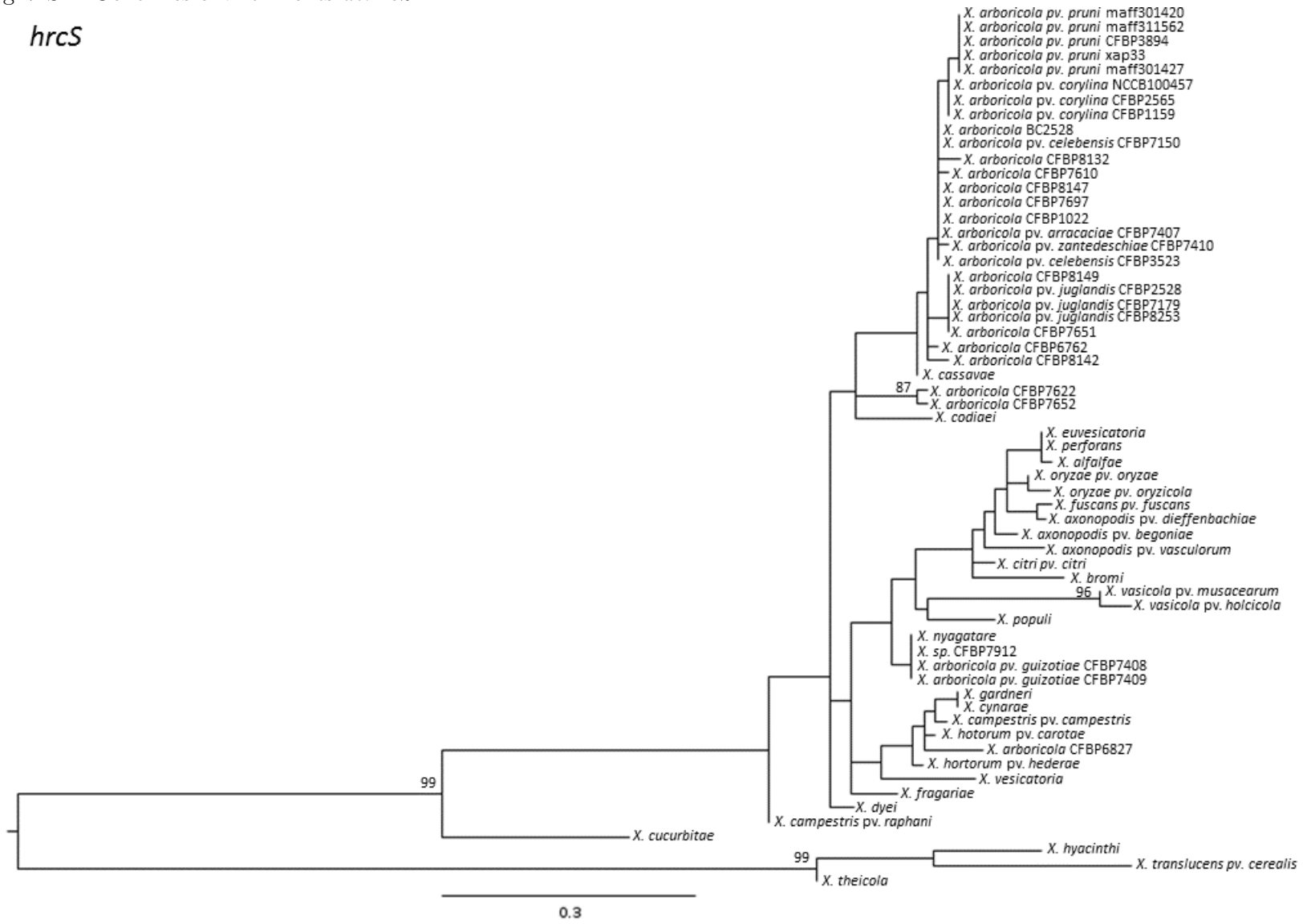


Fig V-S4 : Genomics environments *avrBs2*

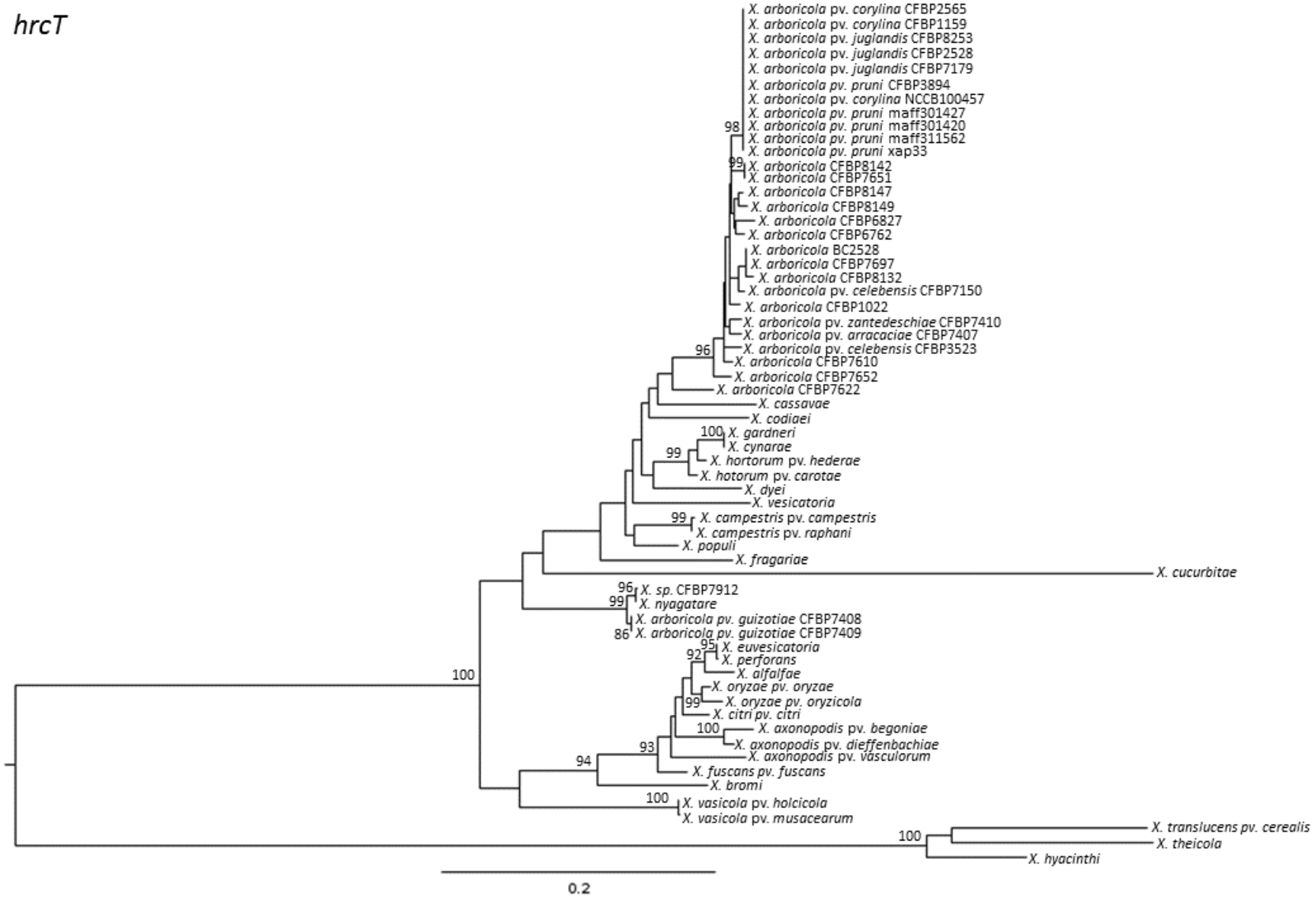


Fig V-S4 : Genomics environments *avrBs2*

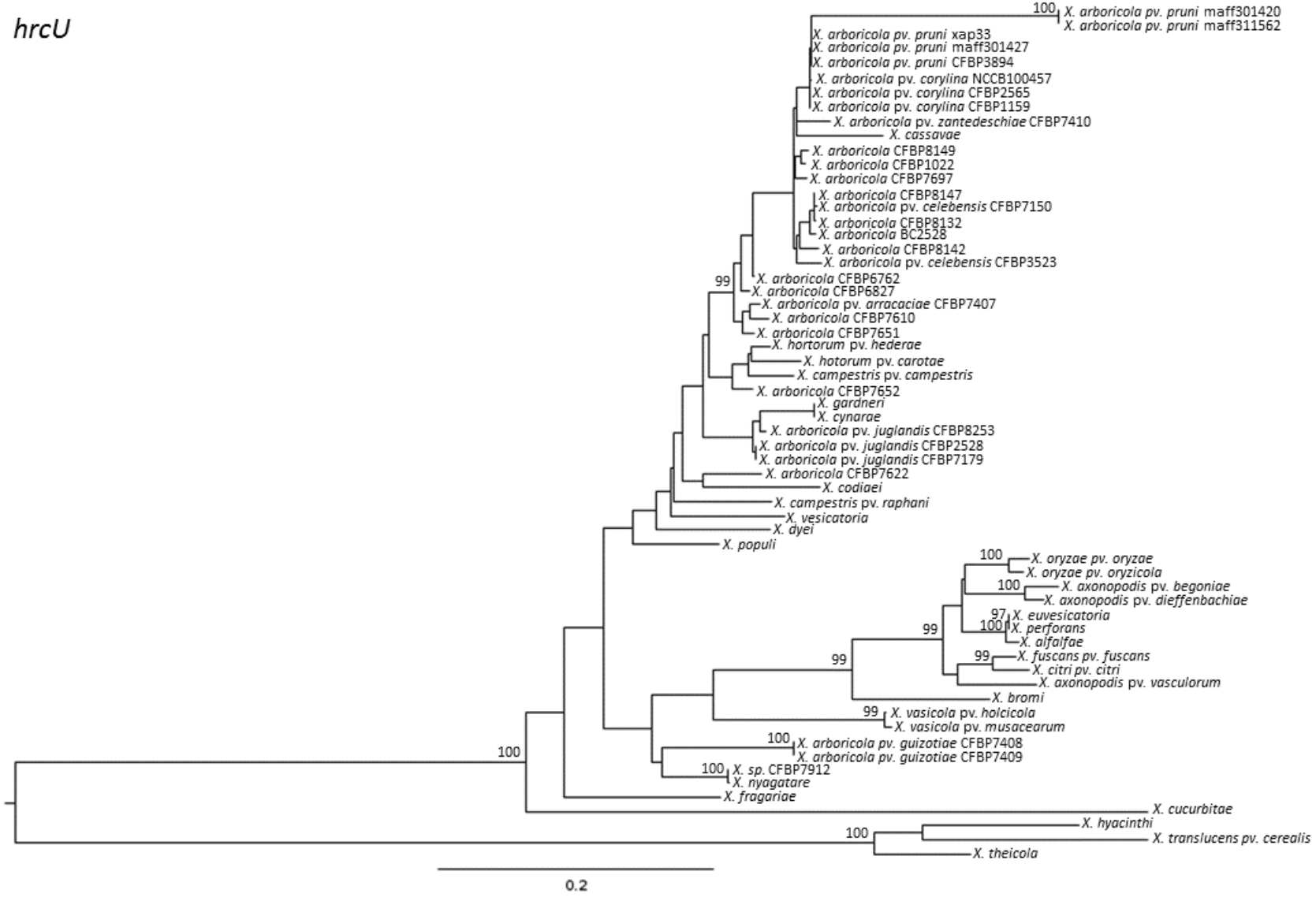


Fig V-S4 : Genomics environments *avrBs2*

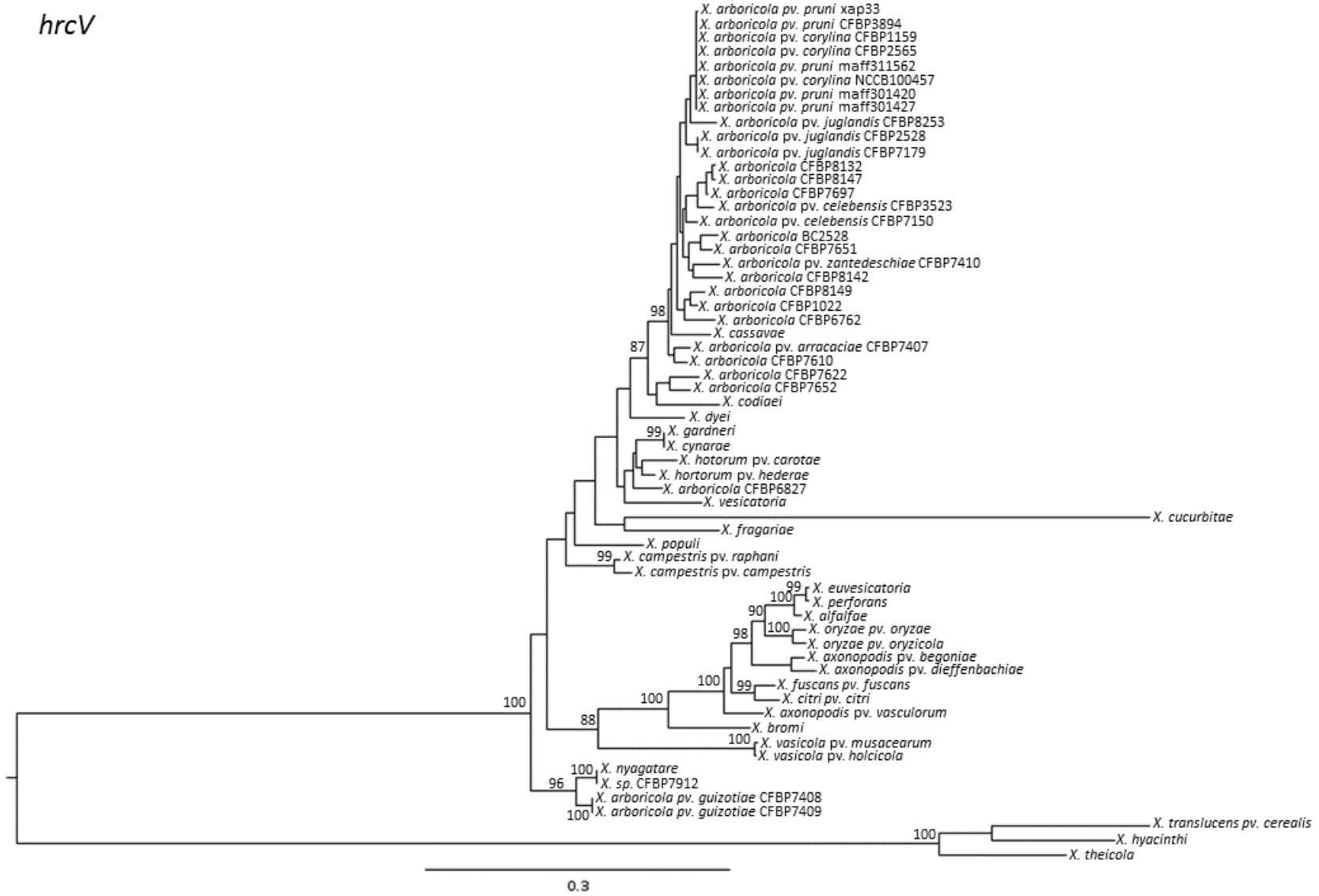


Fig V-S5 : Genomics environments of 10 T3Es acquired in *X. arboricola* group A strains

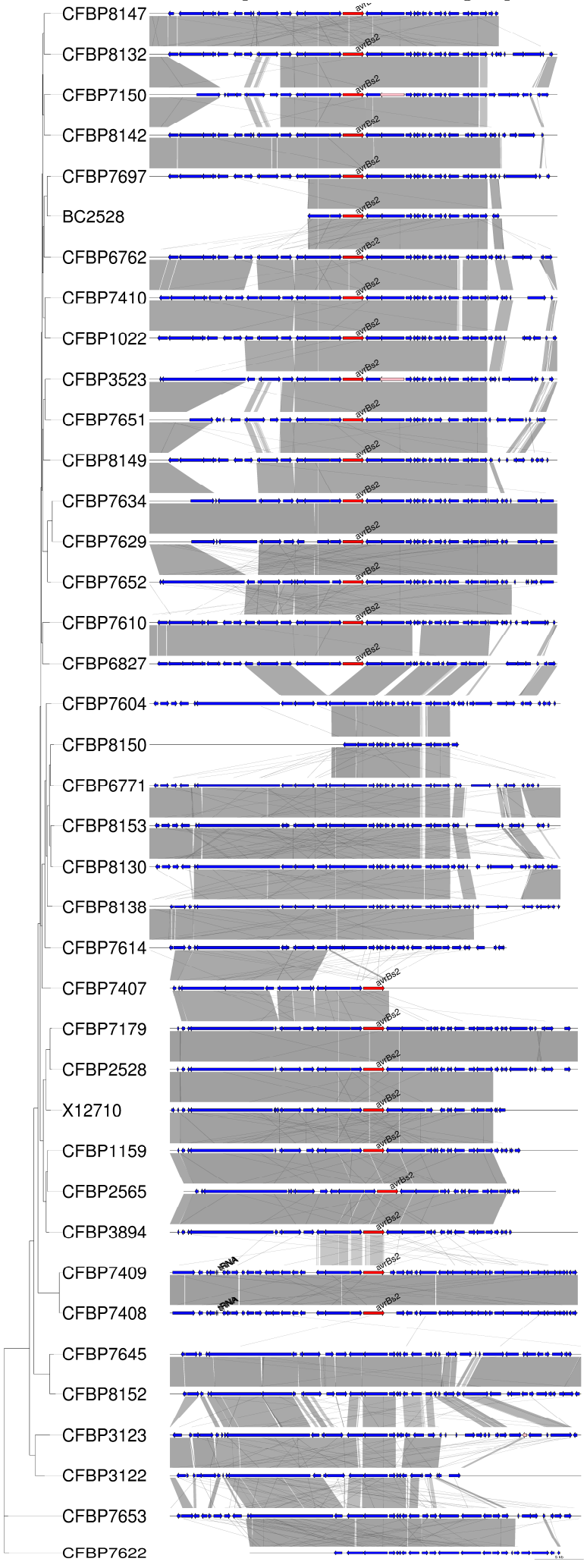
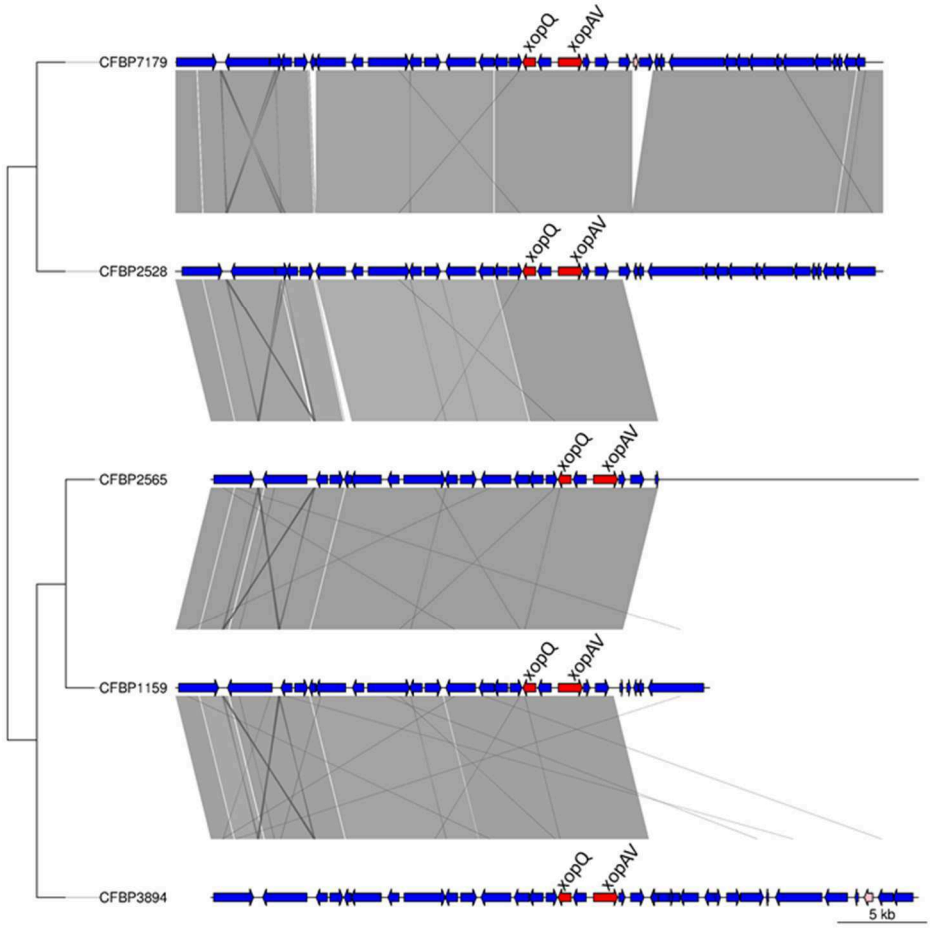


Fig V-S5 : Genomics environments of 10 T3Es acquired in *X. arboricola* group A strains



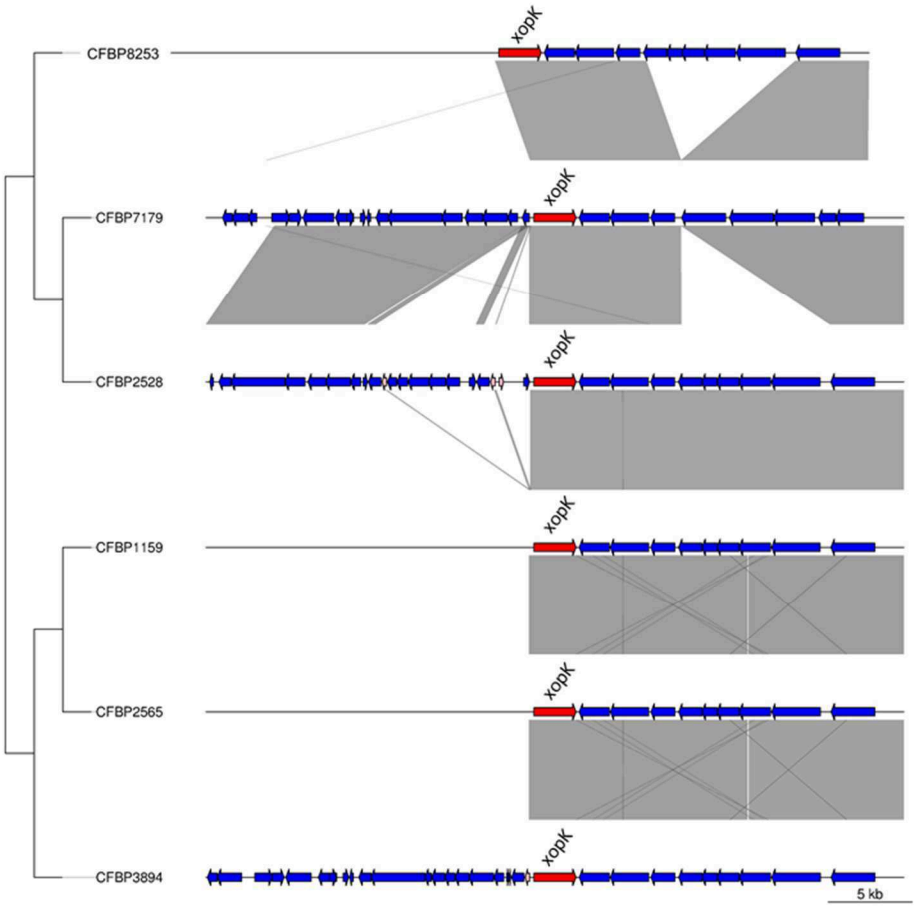


Fig V-S5 : Genomics environments of 10 T3Es acquired in *X. arboricola* group A strains

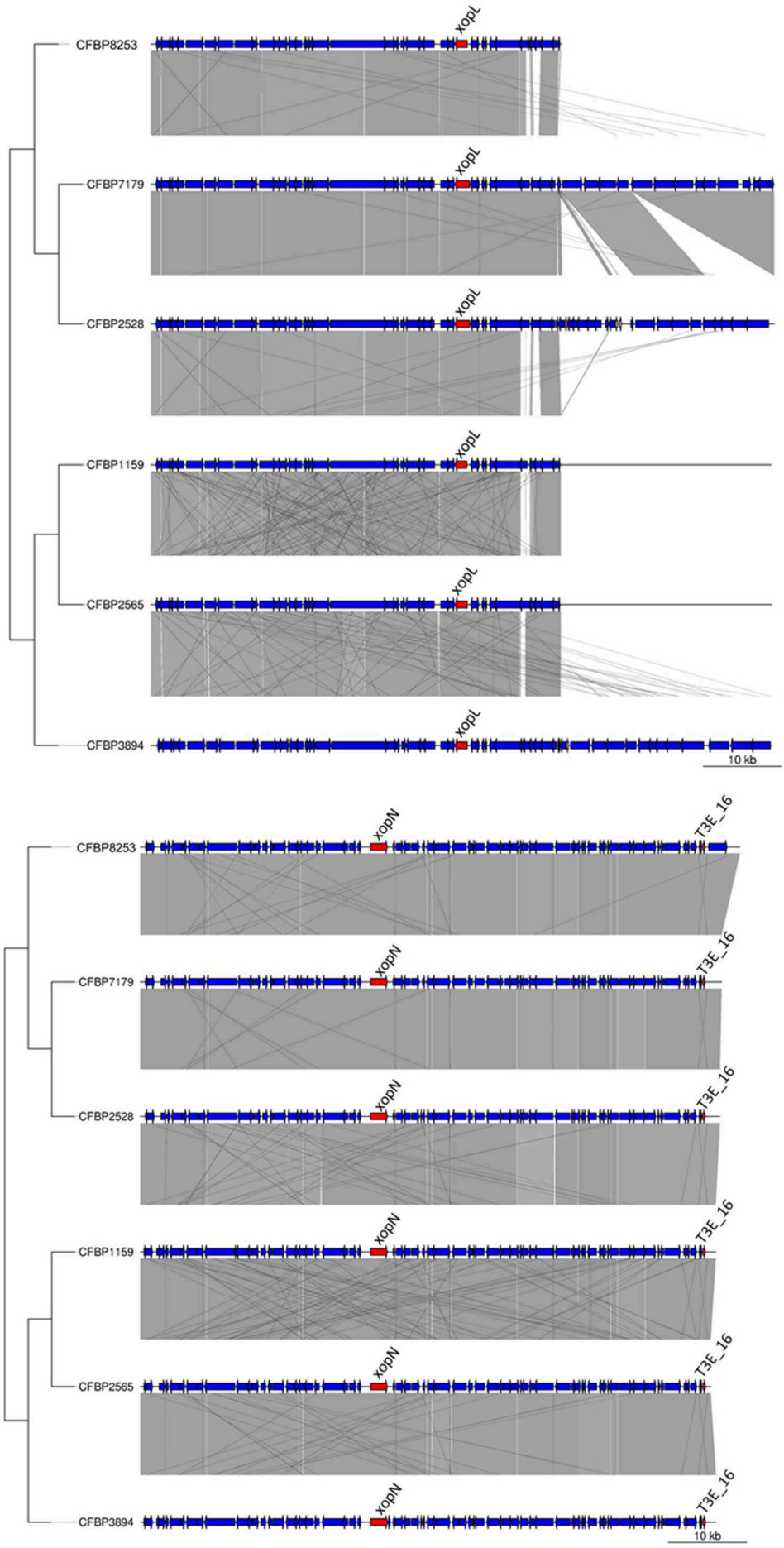


Fig V-S5 : Genomics environments of 10 T3Es acquired in *X. arboricola* group A strains

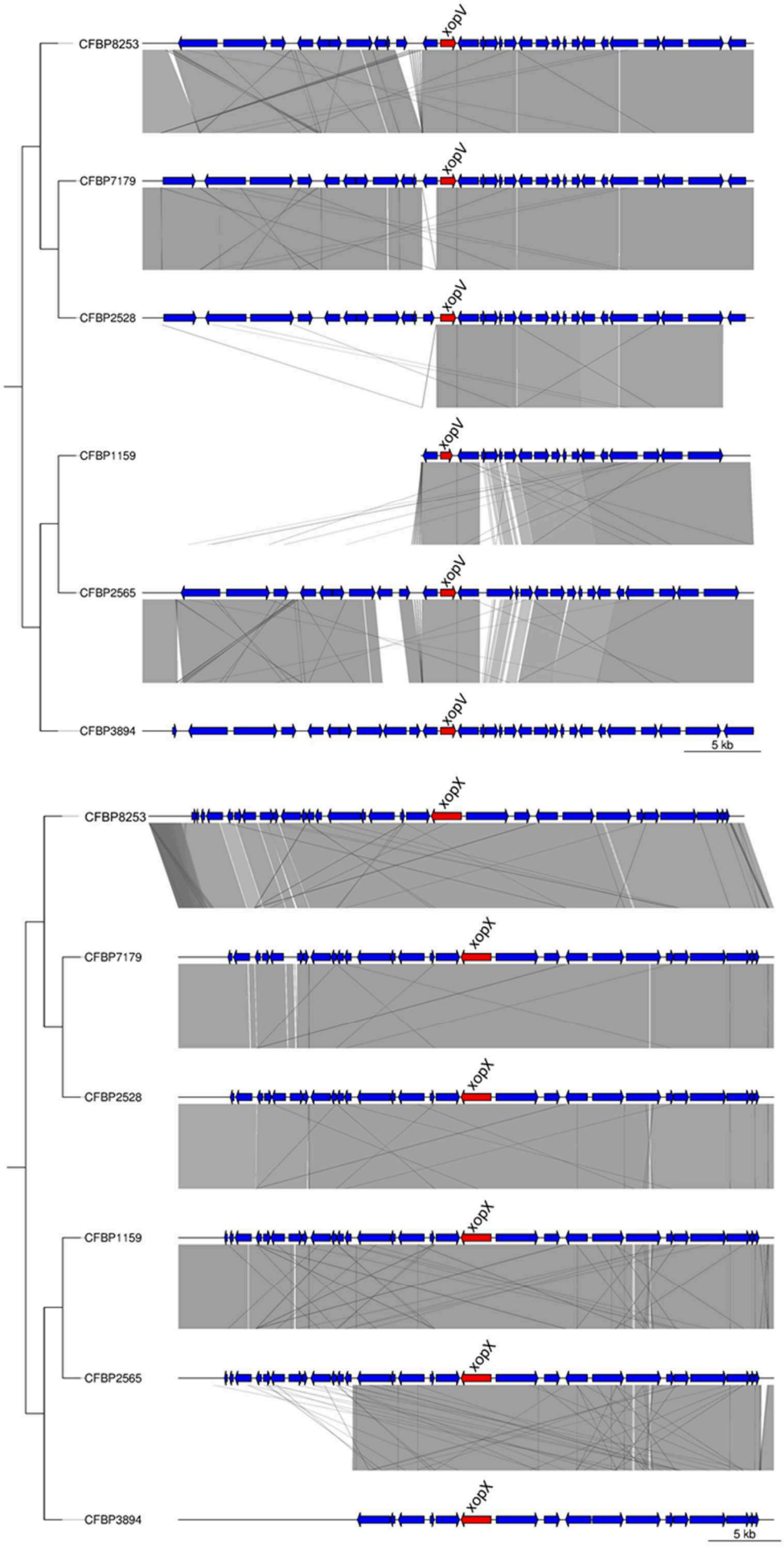
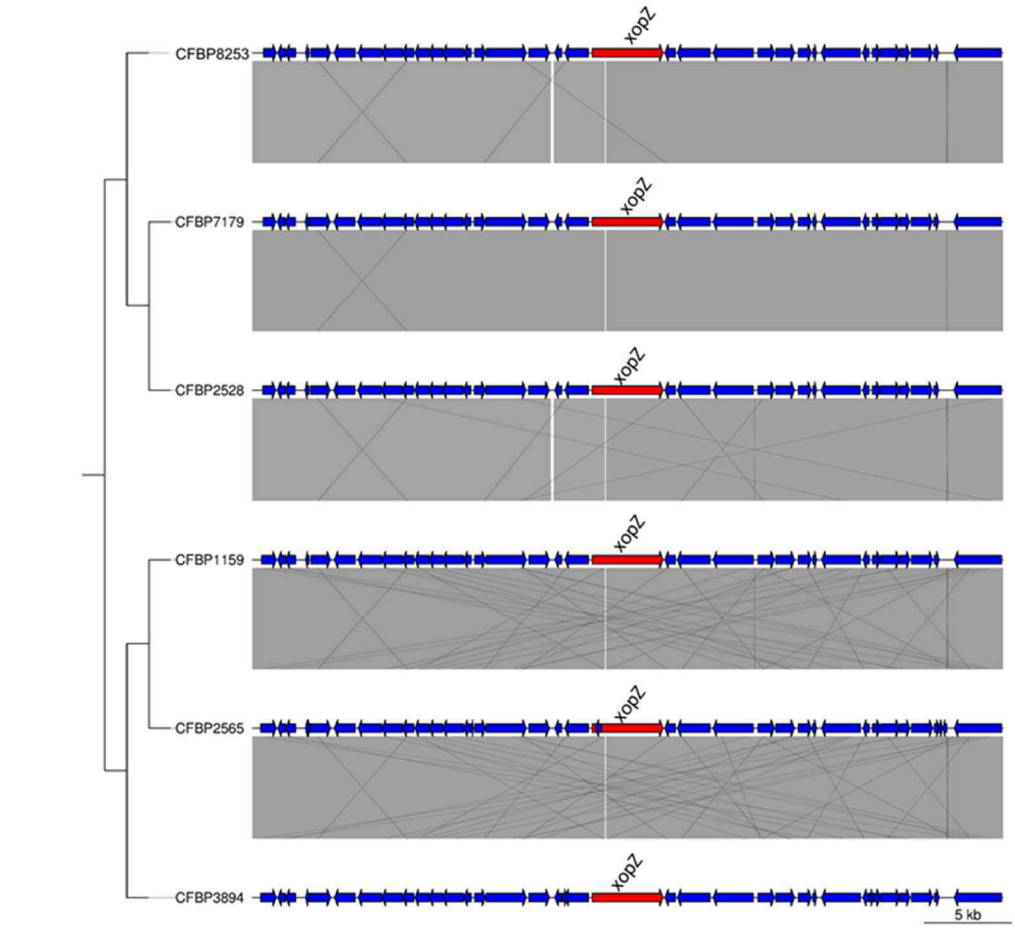
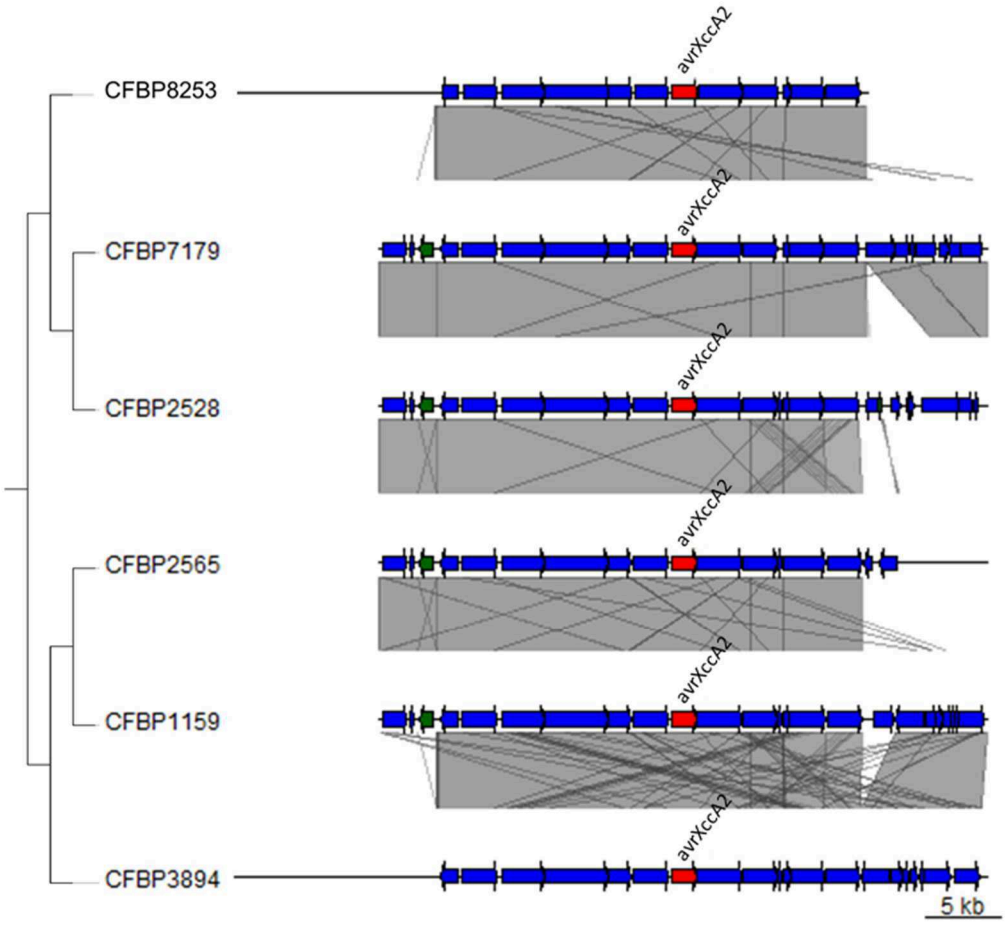


Fig V-S5 : Genomics environments of 10 T3Es acquired in *X. arboricola* group A strains





Discussion générale

Ces travaux de thèse avaient pour objectif de mieux comprendre le rôle que pourraient représenter les souches commensales dans l'émergence de nouvelles maladies bactériennes dans les agroécosystèmes. Pour cela, nous avons étudié l'histoire évolutive d'une espèce bactérienne, *Xanthomonas arboricola*. En effet, cette espèce est composée à la fois de souches commensales et pathogènes. L'étude de Vauterin et al. (1996), complétée par du typage moléculaire réalisé dans mon équipe d'accueil (Guimbaud, 2012), a montré que des souches commensales isolées d'hôtes variés se plaçaient dans cette espèce. De plus, une importante collection de souches commensales de *X. arboricola* issues de semences de haricot et de bourgeons de noyer avait été rassemblée par cette équipe avant mon arrivée et avait donné lieu à une étude de diversité (Essakhi et al. 2015). Deux études de génomique comparative et de phénotypage avaient été réalisées afin de mieux connaitre l'écologie de ces souches commensales que ce soit dans mon équipe d'accueil (Cesbron et al. 2015), ou dans une autre équipe en Espagne (Garita-Cambronero et al. 2016b). Parmi les agents pathogènes, on retrouve les agents pathogènes responsables de maladies, tels que les trois pathovars épidémiques (*pruni*, *corylina* et *juglandis*) et ceux ayant une faible incidence (*guizotiae*, *arracaciae* et *zantedeschiae*). Récemment, la pathogénie de certains pathovars précédemment décrits a été remise en question. C'est le cas des pathovars *fragariae* (Vandroemme et al. 2013) et *celebensis* (Karamura et al. 2015) pour lesquels il est impossible de reproduire des symptômes lors d'inoculations artificielles sur les hôtes supposés de ces deux pathovars. Pour répondre à notre objectif, nous avons étudié la structure de populations de l'espèce *X. arboricola*, réalisé des inférences de scenarios évolutifs, comparé les phénotypes impliqués dans l'interaction avec la plante et enfin étudié l'histoire évolutive du principal facteur de virulence, le système de sécrétion de type 3 et ses effecteurs. L'ensemble des résultats obtenus avec ces différentes analyses est discuté au cours de ce dernier chapitre.

I. Le scénario évolutif de l'espèce *X. arboricola*

1. Les souches commensales dans l'émergence de nouveaux agents pathogènes

En incluant les souches commensales dans notre étude de génétique des populations appliquée à *X. arboricola*, nous avons pu décrire le scénario évolutif de cette espèce (fig. VI-1). Cette espèce présente une structure de population épidémique, telle que définie par Maynard-Smith et al. (1993), au sein de laquelle le réseau recombinant est composé principalement de souches commensales et des pathovars ayant une faible incidence. Les clones épidémiques, quant à eux, correspondent au trois pathovars performants, *pruni*, *corylina* et *juglandis*. L'histoire évolutive du SST3 montre qu'il a été acquis chez l'ancêtre commun d'un grand nombre d'espèces du genre *Xanthomonas*, puis a été perdu chez certaines souches commensales ou pathogènes (chez *X. albilineans*, par exemple). On peut

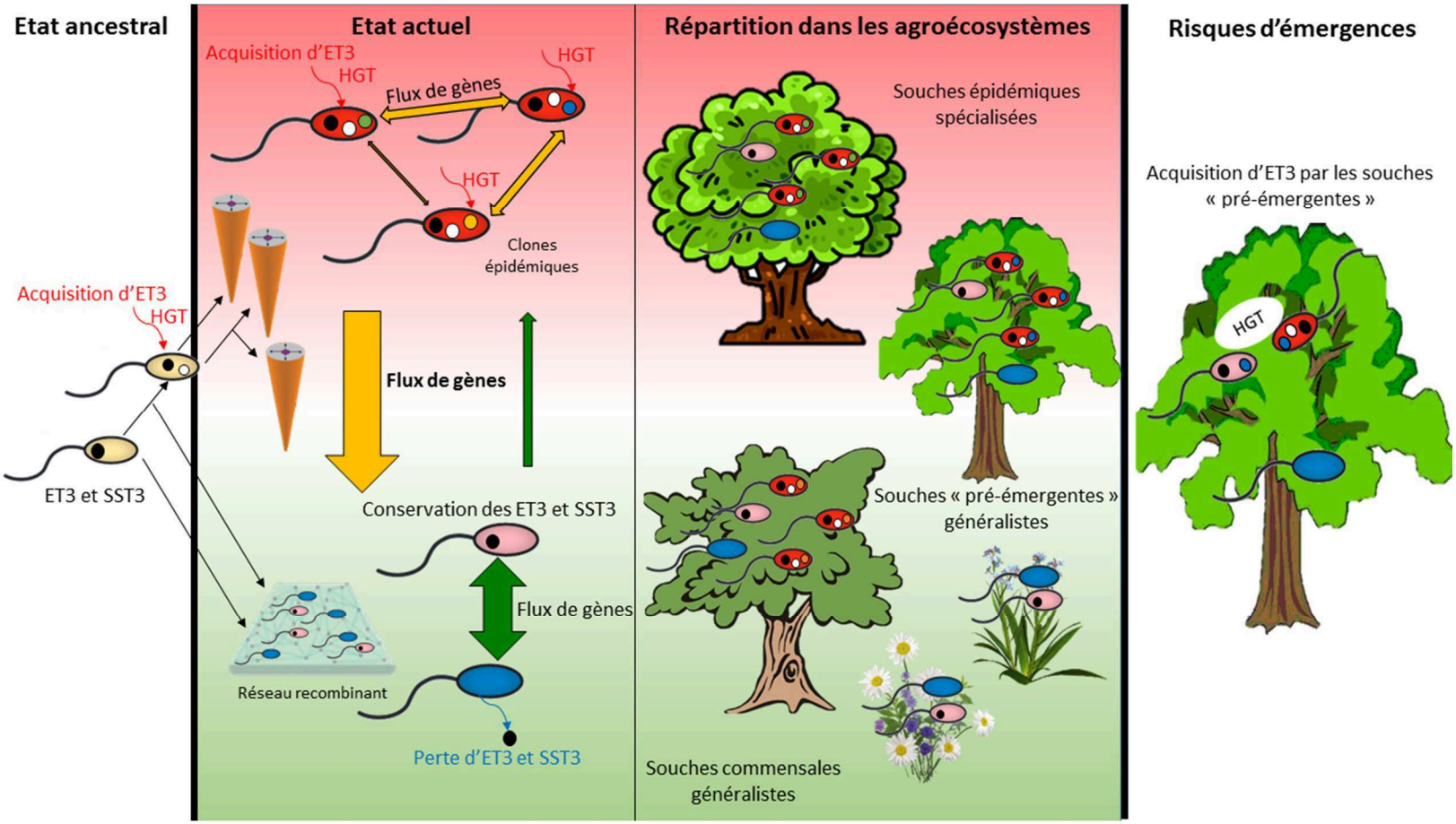


Figure VI-1 : Modèle retraçant l'histoire évolutive de l'espèce *X. arboricola*. Depuis un ancêtre commun (en beige) possédant un SST3 et des ET3 (rond noir), deux groupes ont divergé : un groupe représentant les souches commensales sans SST3 et ET3 (en bleu), et un ayant conservé ces facteurs. Le deuxième groupe a, par la suite, donné, d'une part, les souches « pré-émergentes » (en rose) et l'ancêtre commun des trois clones épidémiques qui a acquis des ET3. Les clones se sont ensuite diversifiés en trois clones épidémiques, avec des acquisitions d'ET3 indépendantes. Les flèches (jaunes lorsqu'une souche pathogène est donneuse, et verte lorsque c'est une souche commensale) représentent le flux de gènes entre les souches, et leur largeur représente l'intensité du flux. L'évolution des souches pathogènes est marquée par des événements d'acquisition (flèches rouges) d'ET3 (ronds blancs, bleus, oranges et verts) par transfert horizontal de gènes (HGT), alors qu'un groupe de souches commensales ont perdu le SST3 et les ET3 de l'ancêtre commun (flèche bleue), et les autres ont conservé le répertoire ancestral. Dans les agroécosystèmes, les souches commensales et pathogènes vivent en sympatrie, les souches pathogènes étant retrouvées exclusivement sur leur hôte, et les souches commensales sur un nombre varié d'espèces végétales. Cette sympatrie et le fait que certaines souches commensales possèdent des facteurs de virulence représentent un risque pour l'émergence d'agents pathogènes qui peuvent acquérir de nouveaux ET3.

donc supposer qu'il y a eu une perte ou modification de la pathogénie pour ces souches. Ces résultats sont à rapprocher de résultats obtenus grâce à la comparaison génomique entre les deux espèces *Streptococcus pneumoniae* et *S. mitis*, la première groupant uniquement des souches pathogènes et la seconde composée de souches commensales. L'évolution de ces dernières vers le commensalisme était associée à des pertes de facteurs de virulence (Killian et al. 2008). Il semblerait donc que les souches commensales de l'espèce *X. arboricola* sans SST3 ne représentent pas un risque majeur pour l'émergence de nouvelles maladies. En effet, comme dans le cas de *Clavibacter michiganensis* (Zaluga et al. 2014), pour que cette émergence se produise, il faudrait cumuler différents évènements d'acquisition, comme l'acquisition d'un SST3 puis de plusieurs ET3. Or, au cours de l'histoire évolutive du SST3 au sein du genre *Xanthomonas*, nous avons détecté peu d'événements de transfert horizontal. Il est donc peu probable que cette succession d'évènements puisse survenir dans ces souches commensales qui auraient déjà perdu ce SST3 au cours de leur histoire évolutive.

Cependant, certaines souches que nous qualifions de commensales, car elles ne provoquent pas de symptômes sur leur hôte d'isolement après inoculation artificielle, ont conservé le répertoire ancestral de facteurs de virulence, à savoir un SST3 et sept ET3. Ces souches diffèrent peu de certaines souches pathogènes qui n'ont qu'un ou deux ET3 de plus. On peut supposer que cette acquisition d'ET3 conditionne l'acquisition d'un pouvoir pathogène et d'une spécialité d'hôte pour ces dernières. Ceci suppose que les souches commensales possédant le répertoire ancestral pourraient représenter un risque pour de nouvelles émergences de maladie. En effet, l'acquisition d'ET3 est assez fréquente car ils sont souvent liés à des éléments génétiques mobiles (McCann & Guttman 2007). Par exemple, l'acquisition du gène codant l'ET3, *sopE*, chez *Salmonella enterica* subspecies 1 serovar Typhimurium aurait eu lieu par l'intermédiaire d'un phage (Ehrbar & Hardt, 2005). Cette acquisition aurait augmenté la fitness de ce clone épidémique (Ehrbar & Hardt, 2005 ; Lopez et al. 2012). De plus, l'histoire évolutive du SST3 a montré qu'il existait un flux de gènes des clones épidémiques vers le réseau recombinant. Ces événements pourraient conduire potentiellement à l'acquisition d'un allèle favorable au niveau des gènes de virulence, comme cela a été montré pour le clone Hajj chez *Neisseria meningitidis* (Mustapha et al. 2015).

Nous avons donc mis en évidence l'existence de deux types de souches commensales, celles qui ont perdu le SST3 et les ET3 présents chez l'ancêtre commun de cette espèce, et d'autres qui ont conservé ce répertoire. Nous qualifions ces dernières de "souches pré-émergentes", bien que contrairement à l'étude de Aviv et al. 2013, nous n'ayons pas fait de tests de génomique fonctionnelle afin de savoir si l'acquisition d'un ou plusieurs ET3 leur permettrait de devenir pathogènes. Nous faisons l'hypothèse que ce type d'acquisition par ces souches est possible du fait que les ET3 sont fréquemment soumis à des transferts horizontaux. De plus, les auteurs avaient qualifié leurs souches de "pré-émergentes" car elles étaient proches génétiquement des souches pathogènes et ne différaient d'elles que par l'absence d'un mégaplasmide, pESI. Ici, nos souches sont proches génétiquement de

certaines souches pathogènes mais ont un répertoire en ET3 différent. Par conséquent, nous considérons qu'elles représentent un risque réel dans l'émergence de nouveaux agents pathogènes.

Dans le cadre de l'utilisation de ces souches commensales pour une potentielle lutte biologique, il apparait clairement qu'elles ne feraient pas de bonnes candidates. Les souches sans SST3 et ET3 présentent un risque moindre d'émergence d'un nouvel agent pathogène que celles qui possèdent déjà cette machinerie cellulaire, et seraient à privilégier pour la mise au point d'une éventuelle lutte biologique. Dans cette perspective, la souche CFBP 7653, dont les tests d'antagonisme ont montré qu'elle inhibait la croissance de toutes les souches pathogènes auxquelles elle a été confrontée (chapitre 3) et qui est dépourvue de SST3 (chapitre 1 et 4), serait une bonne candidate. Cependant, étant donné que des souches de *Xanthomonas* dépourvues de SST3 auraient quand même montré des capacités pathogènes de type opportuniste, telles que *X. arboricola* pv. *populi* (Haworth and Spiers 1992), une telle démarche semble tout de même présenter des risques.

2. L'émergence des trois pathovars performants

Les trois pathovars *juglandis*, *pruni* et *corylina* sont inclus dans un même groupe génétique et semblent évoluer de manière similaire. En effet, l'étude des taux de substitution montre qu'ils accumulent des mutations avec des taux similaires et l'inférence des gains et des pertes des ET3 montre que ces trois clones ont acquis des facteurs en commun au cours de leur histoire évolutive. Les premiers rapports de description de ces agents pathogènes (Pierce, 1901 ; Smith, 1903 ; Bars, 1913) pourraient laisser supposer qu'ils ont émergé à la fin du XIX^{ème}, début XX^{ème} siècle en Amérique du Nord. Cependant, la datation moléculaire montre que le pathovar *juglandis*, d'une part, et les pathovars *pruni* et *corylina*, d'autre part, auraient divergé il y a 6 000 ans. Les pathovars *pruni* et *corylina* auraient, quant à eux, divergé il y a 1 800 ans. On peut noter enfin, que seul le pathovar *pruni* dont aucune espèce hôte n'est originaire d'Amérique du Nord aurait émergé récemment, au début du 20ème siècle.

Il est probable que les agents pathogènes appartenant aux pathovars *juglandis* et *corylina* pré-existaient aux premières descriptions de maladies et que l'émergence de ces maladies ait été causée par l'expansion de l'arboriculture en Californie à la fin du 19ème siècle. En effet, bien que les espèces domestiquées du noyer et du noisetier soient originaires du Moyen Orient (Vahdati et al. 2014 ; Boccacci & Botta, 2009), il existe des espèces endémiques d'Amérique du Nord proches génétiquement des espèces domestiquées, il s'agit de *Juglans nigra*, le noyer noir (Victory et al. 2006) et *Corylus americana* (Coyne et al. 1998). On peut supposer que ces espèces endémiques ont constitué un réservoir à partir duquel les agents pathogènes ont migré vers les cultures. En effet, la migration d'agents pathogènes peut se réaliser des environnements sauvages vers le compartiment cultivé (Burdon & Thrall, 2008 ; Leroy et al. 2014). Par exemple, chez *Pseudomonas syringae* pv. *tomato* il a été montré que des souches provenant de l'environnement étaient proches phy-

logénétiquement de souches présentes dans les agroécosystèmes suggérant des événements de migration entre les deux compartiments (Monteil et al. 2016). Les agroécosystèmes sont des environnements peu variables avec une densité en espèce hôte très importante. Cela élimine la pression de sélection, présente dans les milieux sauvages, sur l'agent pathogène et favorise la dissémination de clones épidémiques hautement spécialisés à leur hôte (Burdon & Thrall, 2008 ; Stuckenbrock & McDonald, 2008). Cependant, il existe tout de même des pressions de sélection imposées par l'homme sur l'agent pathogène dans ces milieux, telles que les traitements au cuivre, ou le déploiement de gènes de résistance, par exemple.

Un autre argument en faveur d'une migration depuis les espèces endémiques vers le compartiment domestiqué est que *J. nigra* possède une résistance naturelle aux souches du pathovar juglandis (Frutos & Lopez, 2012). Belisario et al. 1999 ont inoculé par pulvérisation une souche du pathovar *juglandis* sur 6 espèces appartenant au genre *Juglans* et ont montré que certaines espèces étaient plus ou moins sensibles à la bactérie. Cependant, aucun symptôme ne se développe sur l'espèce *J. nigra*. Cela pourrait suggérer une coévolution entre l'agent pathogène et cet hôte. En effet, la coévolution dans les milieux naturels correspond à l'acquisition de résistances contre des attaques d'agents pathogènes au cours du temps, et au contournement de ces résistances par les agents pathogènes. Or, ce processus n'a pas lieu dans les agroécosystèmes, dans lesquels le patrimoine génétique des espèces hôtes est dépendant des cultivars qui ont été plantés (Raucher, 2001). Les gènes de résistance présents dans les agroécosystèmes sont le plus souvent le résultat de l'amélioration variétale et non d'une coévolution entre l'hôte et l'agent pathogène.

La diversification de ces trois pathovars ne suit pas la diversification des trois hôtes (Boudon, 2005). En effet, le pathovar *juglandis* diverge en premier, suivi par le pathovar *corylina* et enfin par le pathovar *pruni*. Les hôtes des pathovars *corylina* et *juglandis*, quant à eux, appartiennent au même ordre, les Fagales, tandis que le genre *Prunus* (hôte du pathovar *pruni*) appartient à l'ordre des Rosales. Ainsi, les relations phylogénétiques des agents pathogènes ne reflètent pas celles de leurs plante-hôtes puisque les hôtes des pathovars *corylina* et *juglandis* sont plus proches phylogénétiquement que ceux du pathovar *pruni*. De plus, les datations phylogénétiques montrent une émergence récente pour les pathovars *corylina* et *pruni* incompatible avec les temps évolutifs ayant conduits à la radiation entre Rosales et Fagales. Il semble donc plus probable que les scénarios d'émergence de ces trois maladies soient marqués, soit par des changements d'hôtes ou, potentiellement, des acquisitions de nouveaux traits, ou les deux, plutôt que par une coévolution de ces trois pathovars avec leur hôte domestiqué respectif, bien qu'une coévolution entre *J. nigra* et le pathovar *juglandis* soit envisageable dans les milieux sauvages. Il est probable, par exemple, que l'acquisition du plasmide pXap41 par le pathovar *pruni* ait entrainé son émergence. En effet, ce plasmide est spécifique de ce pathovar et porte deux ET3. De même, les pressions de sélection induites par les traitements au cuivre en noyeraies ont pu favoriser l'acquisition d'un ICE (integrative and conjugative element) porteur de gènes de résistance au cuivre et entrainer l'émergence de la lignée responsable du chancre vertical suintant au sein du pathovar *juglandis*. En effet, la recherche par PCR des gènes de cet

a :

b :

Fig VI-2 : Résultats des tests de pouvoir pathogène sur (a) piment (variété Yollo wonder) et (b) tomate de la souche pathogène sur piment BC 2528 (à gauche) et de la souche « pré-émergente » isolée de haricot CFBP 7697 (à droite), 7 jours après inoculation (Fischer-Le Saux, comm. pers.)

ICE dans une collection de souches provoquant la bactériose classique du noyer et d'autres provoquant le chancre vertical suintant a montré que seules les souches responsables du chancre possédaient cet ICE (Cesbron et al. 2015).

3. L'émergence du pathovar *guizotiae*

L'histoire évolutive des souches du pathovar *guizotiae* est marquée par des événements de recombinaison homologue interspécifique, au niveau des gènes de ménage, mais, également au sein de ceux codant le SST3. En effet, la phylogénie des gènes du SST3 a mis en évidence que les allèles du pathovar *guizotiae* étaient plus proches de ceux de l'espèce *"X. cannabis"* que de ceux de *X. arboricola*. De plus, l'histoire évolutive de ce pathovar est également marquée par des événements de transfert horizontaux d'ET3. L'une des deux souches appartenant à ce pathovar présente le nombre le plus important d'ET3 au sein de notre collection, dont un nombre important est commun avec ceux de bactéries d'autres genres et aurait probablement été acquis depuis ces dernières. Cette recombinaison homologue interspécifique pour de larges fragments génomiques tels que le cluster de gènes codant le SST3 peut laisser supposer qu'il y a eu un relâchement du système de réparation des mésappariements de l'ADN (MMR).

Cet agent pathogène est responsable de maladie sur le nyger en Ethiopie et en Inde. Cette plante est cultivée pour son huile qui est la principale huile végétale utilisée en Ethiopie (Geleta & Ortiz, 2013). Elle a également été commercialisée en Europe afin d'effectuer des cultures en rotation et pour sa capacité restructurante des sols. La maladie causée par cet agent pathogène correspond à l'une des principales maladies du nyger, avec celle causée par l'agent pathogène fongique, *Alternaria sp.* (Geleta, 2007). En 2005, cette maladie était caractérisée comme sévère en Ethiopie et moins importante en Inde (Saharan et al. 2005). Cependant, elle est peu étudiée et son incidence est mal connue. Dans le chapitre 1, l'agent pathogène responsable de cette maladie a été classé en tant que pathovar à faible incidence, mais il est probable que ce pathovar représente à l'avenir un clone épidémique, par exemple si cette culture s'intensifie.

La structuration génétique basée sur 7 gènes de ménage montrait que ce pathovar était inclus dans le réseau recombinant. Cependant, les valeurs d'identité nucléotidique moyennes basées sur les génomes sont inférieures à la valeur seuil de la délimitation de l'espèce bactérienne, déterminée par Konstantinidis & Tiedje, 2005. On peut donc supposer que l'émergence de ce pathovar a eu lieu à partir du réseau recombinant et qu'elle a entrainé sa différenciation avec l'espèce *X. arboricola*. En effet, il a été montré que la différenciation écologique pouvait entrainer la divergence entre deux sous-populations à partir d'une population recombinante (Shapiro et al. 2012). Ce modèle est basé sur un équilibre recombinaison et sélection, et selon lequel le balayage sélectif ne concernerait pas un seul gène, mais le génome entier comme décrit dans la théorie de l'écotype de Cohan & Perry, 2007.

II. L'écologie des souches commensales

Il est à noter que le caractère "commensal" a été défini par rapport à l'hôte d'isolement. Or, il est également envisageable qu'un agent pathogène se retrouve en situation non hôte, c'est-à-dire isolé d'un hôte sur lequel il n'est pas capable de provoquer des symptômes. Etant donné qu'il est impossible de tester le pouvoir pathogène sur l'ensemble des hôtes possible, dans ce cas ce sont les analyses génétiques et génomiques qui peuvent permettre de suspecter un caractère pathogène (proximité génétique avec un agent pathogène connu et répertoire de gènes codant des facteurs de virulence). Par exemple, la souche CFBP 7697 isolée de semence de haricot présentait une proximité phylogénétique par MLSA avec les souches coréennes pathogènes de piment. Le séquençage des génomes a confirmé leur proximité phylogénétique et a révélé que la souche CFBP 7697 possédait le même répertoire en ET3 que la souche coréenne pathogène sur piment (BC 2528) (chapitre 4). Un test de pouvoir pathogène de CFBP 7697 a été réalisé, et a montré qu'elle provoquait des symptômes sur tomate et poivron (fig. VI-2 ; Fischer- Le Saux comm pers.). Cependant, on peut noter une agressivité plus faible sur la tomate comparativement aux symptômes provoqués par la souche BC 2528. La souche BC 2528, agent pathogène du poivron, semble avirulente sur la variété de poivron testée (Yolo wonder). En effet, on observe une réaction hypersensible (HR) de la plante et non une maladie, lors de l'inoculation de cette souche. Pour cet exemple, le test était facile à réaliser car nous avions le profil d'une souche pathogène qui nous a indiqué l'espèce végétale sur laquelle réaliser le test de pouvoir pathogène. Cet exemple illustre les potentialités des analyses génomiques pour prédire le caractère pathogène d'une souche considérée *a priori* commensale, et l'importance de disposer de bases de données génomiques sur les agents pathogènes les plus exhaustives possible.

Au cours de l'étude de phénotypage des souches, nous avons réalisé des tests d'antibiose qui ont mis en évidence que dans la plupart des cas les souches commensales et pathogènes peuvent coexister. Cependant, il existe d'autres types d'interaction que nous n'avons pas testés. Chez *E. coli*, il a été montré que la présence de souches commensales augmentait l'agressivité de la souche pathogène, et donc qu'il existait une synergie entre les deux souches bactériennes (Tourret et al. 2011). Une interaction bénéfique a également été montrée entre des souches appartenant à l'agent pathogène *Pseudomonas aeruginosa* (Buckling et al. 2007). Un modèle de coopération a également été décrit, dans lequel les bactéries libèrent des nutriments dans le milieu grâce à leur secretome, et des bactéries dépourvues de ce secretome et dites " cheaters " (tricheuses) profitent de ces nutriments (Lawrence, 2009; Nogueira et al. 2009). Dans ce modèle qui fait appel à la théorie de la sélection de parentèle[15], il est suggéré qu'à terme, les souches coopératrices (celles permettant la libération des nutriments) recrutent les " cheaters ", en leur transférant les gènes du secretome nécessaire à la libération des nutriments dans le milieu. Il est pro-

[15]Théorie de sélection de parentèle correspond au processus favorisant certains caractères en raison de leurs effets bénéfiques sur la survie de la famille (Griffin & West, 2002). Cette théorie permet d'expliquer la coopération entre organismes.

bable que les souches commensales de *X. arboricola* profitent des nutriments libérés par les souches pathogènes. Et donc, si l'espèce *X. arboricola* suivait ce modèle décrit par Lawrence en 2009, cela augmenterait le risque potentiel d'émergence, puisque la seconde partie du modèle correspond au recrutement des " cheaters ".

Enfin, un autre type d'interaction, mais cette fois-ci avec la plante, pourrait être l'opportunisme. Les agents pathogènes opportunistes sont définis comme des organismes qui peuvent devenir pathogènes sur un hôte affaibli (Brown et al. 2012). Chez les agents phytopathogènes, cela peut se caractériser par une activité pectinolytique, qui permet de dégrader la pectine (Hugouvieux-Cotte-Pattat et al. 2014). Cette activité a été testée *in vitro* pour deux souches commensales et deux souches pathogènes (Cesbron et al. 2015), et seules les deux souches commensales possédaient cette capacitié. Ce type d'interaction pourrait également être testé pour l'ensemble de nos souches commensales sur un hôte présentant des blessures ou sur organes détachés. L'appellation " agent pathogène opportuniste " a déjà été proposée que le pathovar *populi* (Haworth & Spiers 1992), placé initialement dans l'espèce *X. arboricola* mais dont nos valeurs d'ANI sont en dessous du seuil de l'espèce, et pour des souches de *X. arboricola* isolées de symptômes sur la vigne (Sawada et al. 2011).

III. Apport de la génomique

Dans le premier chapitre de cette thèse nous avons recherché les ET3 par PCR. Cette recherche a donc été réalisée avec *a priori*, puisque seuls les ET3 connus chez *Xanthomonas* ont été recherchés. Ensuite, nous avons recherché les ET3 dans les génomes sans *a priori* par une approche d'apprentissage automatique qui est une méthode d'apprentissage d'un algorithme sur des données empiriques. Cette méthode se base essentiellement sur le signal de sécrétion, l'absence de structure secondaire (caractéristique des ET3), leur GC%, leur régulation par les régulateurs HrpX et HrpG et leur homologie avec des ET3 connus dans différents genres bactériens. Par PCR, nous trouvions 6 ET3 présents chez les souches " pré-émergentes ", l'approche sans *a priori* nous a permis d'en détecter 7 (6 sont communs à l'approche PCR). De plus, avec l'approche PCR, certaines souches pathogènes présentes dans le réseau recombinant, telles que la souche pathogène sur piment, CFBP 6827, et la souche du pathovar *celebensis* CFBP 3523, possédaient le même répertoire que des souches " pré-émergentes ". L'analyse des génomes a permis de mettre en évidence un nombre plus important d'ET3 (11 ET3) dans la souche CFBP 6827. La souche CFBP 3523 possède deux ET3 en plus du répertoire ancestral qui sont deux ET3 putatifs. Enfin, la différence la plus notable entre l'approche de recherche par PCR et l'approche génomique concerne la souche CFBP 7408 appartenant au pathovar *guizotiae*. En effet, cette souche possède le nombre d'ET3 le plus important (34 ET3) de notre collection alors que par PCR, seuls 14 ET3 ont été retrouvés, et son répertoire semblait plus petit que celui des trois clones épidémiques. Dans l'ensemble des 44 génomes de *X. arboricola* analysés par cette approche basée sur l'apprentissage automatique, 6 nouveaux ET3 putatifs ont été retrouvés. Afin de

déterminer si ces CDS codent réellement pour des protéines sécrétées par le SST3, et par conséquent de réels ET3, il faudrait entreprendre des travaux en génomique fonctionnelle.

L'étude des génomes a également permis de mettre en évidence des flux de gènes entre les clones épidémiques et le réseau recombinant, alors que cela n'était pas mis en évidence en étudiant seulement les gènes de ménage. Après avoir validé la structuration en 2 groupes avec les génomes, nous avons pu tester des scénarios de divergence, et le plus probable montre un flux de gènes asymétrique tout au long de la divergence entre les deux populations. Cependant, ces méthodes étant basées sur des modèles de taille de population constante au cours du temps et les fréquences alléliques, il faudrait faire varier la taille de la population épidémique dans le modèle. En effet, une émergence est souvent issue d'un goulot d'étranglement, soit une diminution de la taille efficace de la population, suivie d'une expansion. Dans le cas d'un balayage sélectif du génome entier, tel que défini dans le modèle de l'écotype de Cohan & Perry, 2007, l'allèle pourtant issu du réseau recombinant, et donc en faible fréquence dans celui-ci, se retrouve en haute fréquence dans les clones épidémiques. L'inférence, qui est basée sur des modèles de populations à taille constante au cours du temps, pourrait attribuer l'origine de l'allèle à la population où la fréquence est la plus élevée, à savoir les clones épidémiques.

Enfin, l'apport des génomes dans cette étude nous a permis de mettre en évidence des mécanismes évolutifs, tels que le flux de gènes et le transfert horizontal, auxquels nous n'avions pas accès par PCR ou avec les gènes de ménage. Cependant, ces génomes étaient sous forme de contigs et donc non circulaires comme l'est le chromosome bactérien. Cela a donc limité les analyses des environnements génomiques d'ET3, par exemple. En effet, les contigs sont interrompus fréquemment au niveau des séquences d'insertion qui empêchent un bon assemblage (Varani et al. 2011). Or, les ET3 sont souvent associés à ces éléments génétiques mobiles, et dans le génome des souches du pathovar *guizotiae* par exemple, certains ET3 étaient localisés sur des contigs ne portant que ce CDS. De plus, cela nous a également limités pour la détection des IS dans les génomes. Afin de déterminer la fréquence de chaque IS, nous avons donc recherché uniquement les séquences bordantes, comme réalisé chez Darasse et al. 2013, mais il est probable que cette approche ait faussé le nombre d'IS réellement présentes dans les génomes. De même, les effecteurs TAL possèdent de longues séquences répétées qui ne peuvent être assemblées correctement lors de séquençage par la méthode Illumina. Ils sont souvent non ou partiellement présents dans les génomes obtenus par cette méthode. Deux effecteurs TAL ont été détectés dans les génomes de *X. arboricola* (chapitre 4). Un séquençage par la méthode PacBio permettrait d'être plus exhaustif dans la recherche de ces effecteurs (Booher et al. 2015). Enfin, nous n'avons pas pu étudier les mécanismes de réarrangements génomiques qui peuvent également jouer un rôle dans l'émergence d'agent pathogène.

IV. Conclusion et Perspectives

Perspectives permettant d'améliorer nos connaissances de l'émergence de maladie

L'émergence des agents pathogènes chez *X. arboricola* est marquée par des événements de recombinaison homologue et de transfert horizontal de gènes. Les pressions de sélection n'ont pas été étudiées au cours de cette thèse, or elles peuvent également expliquer l'émergence de nouvelles maladies. En effet, elles pourraient expliquer l'asymétrie du flux génétique observée entre les souches commensales et pathogènes. La génomique évolutive de *X. campestris* et *X. citri* a montré que la recombinaison homologue et la sélection naturelle jouaient un rôle dans le changement d'hôte de ces deux espèces (Huang et al. 2015). Il est donc probable que la sélection ait joué un rôle dans l'émergence des pathovars, et il pourrait être intéressant d'étudier cette force évolutive.

L'étude des scénarios évolutifs semble indiquer que les trois pathovars *juglandis*, *corylina* et *pruni* auraient émergé suite à l'acquisition de facteurs de virulence. Afin de valider cette hypothèse, il pourrait être envisagé de dater les évènements d'acquisition de ces facteurs de virulence avec la méthode du " tip-dating " et de les comparer aux dates obtenues pour l'émergence de ces trois pathovars. Une analyse similaire a été réalisée chez *Vibrio cholerae* afin de dater l'acquisition des éléments intégratifs conjugatifs (Spagnoletti et al. 2014). Afin d'étudier plus précisément ces scénarios d'émergence, il pourrait être intéressant d'échantillonner dans les zones supposées d'origine de ces pathovars et de réaliser une étude de génomique des populations spécifiques à ces trois pathovars.

Les trois pathovars dont la pathogènie est remise en cause chez l'espèce *X. arboricola* : *X. arboricola* pv. *fragariae*, *X. arboricola* pv. *populi* et *X. arboricola* pv. *celebensis* présentent la particularité d'avoir des cousins pathogènes au sein du genre *Xanthomonas*, lesquels provoquent de forts symptômes sur leur plante hôte : *X. fragariae* sur le fraisier, *X. populi* sur le peuplier et *X. vasicola* pv. *musacearum*, sur la banane. Comme nous avons vu précédemment que des interactions synergiques peuvent exister entre une souche pathogène et des souches commensales ou des agents pathogènes opportunistes, des tests d'inoculation en couple pourraient aider à comprendre si les souches de *X. arboricola* jouent ce rôle pour les agents pathogènes du genre *Xanthomonas*.

Une analyse de diversité spécifique du pathovar *guizotiae* pourrait être intéressante à réaliser. En effet, ce pathovar semble particulier au sein de l'espèce *X. arboricola*, puisqu'il est caractérisé par des événements de recombinaison homologue inter-spécifique et d'acquisition de facteurs de virulence. Au cours de cette thèse, nous n'avons pas pu étudier plus spécifiquement ce pathovar puisque seules deux souches étaient disponibles dans les centres de ressources biologiques. Pour cela, il serait intéressant d'aller échantillonner dans les zones où la maladie sur nyger a été détectée, à savoir l'Ethiopie et l'Inde.

Ces travaux laissent supposer que les souches commensales dites " pré-émergentes " et les souches pathogènes diffèrent par leur répertoire en effecteurs de type 3. Il est

possible d'envisager des analyses de génomique fonctionnelle afin de valider l'hypothèse suivante : les souches " pré-émergentes " et les souches pathogènes présentent peu de différences en termes de répertoire en effecteurs, il pourrait être intéressant de voir si ces souches " pré-émergentes ", qui sont proches génétiquement de souches pathogènes peuvent devenir pathogène suite à l'acquisition d'un ou plusieurs effecteurs. Une telle manipulation validerait le risque que peuvent représenter ces souches commensales pour de nouvelles émergences. Par exemple, la souche du pathovar *zantedeschiae* CFBP 7410 est proche génétiquement de la souche CFBP 1022 isolée de noyer, et possède trois ET3 que la souche " pré-émergente " ne possède pas, *xopAD*, *xopAE* et *xopK*. Il serait possible de transférer ces effecteurs dans la souche " pré-émergente ", puis de réaliser un test de pouvoir pathogène sur l'arum d'Ethiopie, l'hôte des souches du pathovar *zantedeschiae*. L'acquisition de plusieurs ET3 suite à un seul événement de transfert horizontal de gènes est possible. En effet, le plasmide spécifique du pathovar *pruni* porte deux ET3, *xopE2* et *xopE3*. Un autre couple candidat pour ces expérimentations pourrait être la souche CFBP 6827 pathogène sur piment et la souche " pré-émergente " CFBP 7610. Cependant, le nombre d'ET3 différents entre ces deux souches est plus important que pour le premier couple proposé, puisque 4 sont spécifiques de la souche pathogène et deux de la souche " pré-émergente ". Il est également possible d'envisager de faire ces manipulations avec les souches commensales qui ne possèdent ni SST3 ni ET3. Cependant, l'étude de l'histoire évolutive du SST3 au sein des *Xanthomonas* a montré que les événements d'acquisition de ce dernier correspondent à des événements rares.

Perspectives permettant d'améliorer nos connaissances sur l'écologie des souches commensales

Ces travaux ont donc permis d'améliorer nos connaissances sur l'écologie des souches commensales de *X. arboricola*. En effet, nous savons désormais que ces souches présenteraient les mêmes phases d'interactions avec la plante, exceptée la phase de multiplication qui est spécifique aux agents pathogènes et leur permet d'atteindre de grandes tailles de populations. Afin d'approfondir ces connaissances sur l'écologie des souches commensales, une analyse des composés organiques sur lesquelles elles sont capables de se développer pourrait être envisagée. Cela a été réalisé pour la souche commensale isolée de *prunus*, en utilisant des microplaques GenII de Biolog (Garita-Cambronero et al. 2016a). Les auteurs ont montré des différences avec les souches pathogènes. Cette analyse avait été entreprise au cours de mes travaux de thèse. Cependant, l'analyse de la coloration des puits était difficile à réaliser sans l'équipement dédié (Omnilog), et donc les résultats étaient difficiles à interpréter. De plus certaines souches auraient un fort pouvoir réducteur entrainant un résultat faussement positif dans le puits contrôle. Il serait souhaitable de reprendre ces analyses avec les outils permettant leur interprétation et en adaptant les protocoles.

Concernant les interactions entre les souches commensales et pathogènes, nous avons étudié leur comportement *in vitro*. Nos résultats montrent que majoritairement, il n'y a

pas d'interaction négative entre les souches pathogènes et commensales. Ces résultats semblent indiquer que ces deux types de souches peuvent coexister au sein d'une même plante hôte. Cependant, il serait intéressant d'étudier les interactions de ces souches *in planta*. Ces tests n'ont pas pu être réalisés au cours de cette thèse, car ils nécessitent encore une importante étape de mise au point. En effet, dans la collection que nous possédons, la seule plante hôte à partir de laquelle des souches commensales et pathogènes ont été isolées est le noyer. La manipulation du noyer est assez difficile, et pour l'instant nous n'avons pas réussi à maintenir les feuilles en survie dans de l'eau gélosée plus de 4 jours, avec de plus une perte importante. De plus, la co-inoculation d'une souche commensale et d'une souche pathogène, *in vitro* dans un premier temps et *in planta* par la suite, pourrait être réalisée pour l'étude des biofilms puisque ces structures correspondent à des hotspots de recombinaison (Danhorn & Fuqua, 2007). Pour cela, il faudrait marquer les souches avec des protéines autofluorescentes afin de les identifier en microscopie. Certaines souches avaient été marquées par la GFP avant mon arrivée (Biessy, 2013), seulement la plupart ne conservaient pas le plasmide qui portait la GFP en absence de pression de sélection.

Nous avons également montré que ces souches commensales pouvaient être présentes sur un grand nombre d'espèces d'intérêt agronomique ou ornemental. Il pourrait être intéressant d'étudier la diversité que représente ce réservoir recombinant en étudiant la diversité des Xanthomonas en dehors des agroécosystèmes. En effet, une analyse du microbiote d'*Arabidopsis thaliana* dans les milieux sauvages a montré que cette plante hébergeait des individus appartenant à l'espèce *X. arboricola* (Roux, comm pers.). Il semble donc que l'on retrouve des individus appartenant à cette espèce en dehors des agroécosystèmes, et on peut émettre l'hypothèse que ces milieux sauvages pourraient héberger une diversité encore plus importante de souches commensales.

En conclusion, l'étude des souches commensales a montré l'importance que peuvent représenter les réservoirs environnementaux, à savoir la communauté microbienne dans laquelle évolue l'agent pathogène, dans l'émergence de nouvelles maladies dans les agroécosystèmes. En effet, en ne s'intéressant qu'aux clones épidémiques, c'est-à-dire qu'aux agents causant des maladies, on occulte le laboratoire naturel responsable de ces émergences. Ainsi, cette étude basée essentiellement sur *X. arboricola* nous a permis de détecter une diversité très importante dans l'espèce évoluant principalement par recombinaison homologue, et qui reçoit des allèles provenant d'agents pathogènes. Cette force évolutive est responsable de l'émergence de plusieurs agents pathogènes, tels que *Xylella fastidiosa* subspecies *morus* (Nunney et al. 2014). De plus, nous avons retracé l'histoire évolutive du SST3 et nous avons montré que la présence du SST3 et de 7 ET3 est un caractère ancestral. Cette étude nous a également permis de mettre en évidence deux types de souches commensales : celles qui ont conservé le SST3 et le répertoire ancestral d'ET3 et qualifiées de souches " pré-émergentes ", et celles qui les ont perdu au cours du temps. L'histoire évolutive de ces deux types de souches commensales n'était pas suspectée au début de ma thèse. La définition de ces deux types de souches commensales est importante pour prévenir les émergences de maladies, puisqu'elles ne représentent pas les

mêmes risques d'émergence. En effet, les souches " pré-émergentes " semblent présenter un réel risque comparativement à celles qui sont dépourvues de SST3 et ET3. Au sein du genre *Xanthomonas*, d'autres souches commensales ont été mises en évidence, certaines appartiennent à des espèces connues comme *X. campestris* et d'autres forment des espèces comme *X. maliensis*. D'autres espèces bactériennes associées aux plantes hébergent à la fois des agents pathogènes et commensaux, telles que *Clavibacter michiganensis* (Jacques et al. 2012). Il est donc probable que ce réseau recombinant formé principalement par des souches commensales soit encore plus diversifié que ce que nous avons mis en évidence chez *X. arboricola*, et leur étude pourrait permettre de généraliser nos résultats à l'ensemble du genre. Une structure épidémique est aussi montrée chez *Pseudomonas syringae* (Monteil et al. 2013). Dans ce cas les souches du réseau recombinant provenaient du compartiment non agricole lié au cycle de l'eau. Cependant on peut observer aussi chez cette espèce des souches isolées du compartiment cultivé ayant un faible pouvoir pathogène et génétiquement très polymorphes, comme *Pseudomonas syringae* pv. *actinidifoliorum* (Cunty et al. 2015), qui pourraient aussi représenter le réseau recombinant.

Enfin, l'étude du flux de gènes nous a montré l'importance que pouvaient avoir les études de génomique des populations pour comprendre l'écologie des souches. En effet, nous avons pu retracer des événements de flux de gènes, au niveau du SST3 notamment, entre un grand nombre d'espèces du genre *Xanthomonas*. Cela indique que ces souches ont pu être en sympatrie. Les études de métagénomique donnent ces informations mais, pour une situation actuelle, et ne permettent pas de formuler des hypothèses sur les éventuelles mises en sympatrie des bactéries au cours de leur histoire évolutive qui ont potentiellement façonné leur génome actuel. Cependant, l'étude de métagénomes anciens permettrait de valider ces différentes hypothèses de mise en contact des microorganismes au cours de leur histoire évolutive.

Références

Abby SS, Rocha EPC. 2012. The non-flagellar Type III Secretion System evolved from the bacterial flagellum and diversified into Host-cell adapted systems. PLoS Genet. 8:e1002983.

Abera T, Feyisa D, Friesen DK. 2009. Effects of Crop Rotation and N-P Fertilizer Rate on Grain Yield and Related Characteristics of Maize and Soil Fertility at Bako, Western Oromia, Ethiopia. East African J. Sci. 3:70–79.

Akimoto-Tomiyama C, Furutani A, Tsuge S, Washington EJ, Nishizawa Y, Minami E, Ochiai H. 2012. XopR, a type III effector secreted by *Xanthomonas oryzae* pv. *oryzae*, suppresses microbe-associated molecular pattern-triggered immunity in *Arabidopsis thaliana*. Mol. Plant. Microbe. Interact. 25:505–14.

Alavi SM, Sanjari S, Durand F, Brin C, Manceau C, Poussier S. 2008. Assessment of the Genetic Diversity of *Xanthomonas axonopodis* pv. *phaseoli* and *Xanthomonas fuscans* subsp. *fuscans* as a Basis To Identify Putative Pathogenicity Genes and a Type III Secretion System of the SPI-1 Family by Multiple Suppression Subtractive H. Appl. Environ. Microbiol. 74:3295–3301.

Alexander RP, Zhulin IB. 2007. Evolutionary genomics reveals conserved structural determinants of signaling and adaptation in microbial chemoreceptors. Proc. Natl. Acad. Sci. U. S. A. 104:2885–2890.

Alexandre G. 2010. Coupling metabolism and chemotaxis-dependent behaviours by energy taxis receptors. Microbiology 156:2283–2293.

Alfano JR, Collmer A. 1996. Bacterial Pathogens in Plants: Life up against the Wall. Plant Cell 8:1683–1698.

Altizer S, Ostfeld RS, Johnson PTJ, Kutz S, Harvell CD. 2013. Climate Change and Infectious Diseases: From Evidence to a Predictive Framework. Science 341:514–519.

Anderson PK, Cunningham AA, Patel NG, Morales FJ, Epstein PR, Daszak P. 2004. Emerging infectious diseases of plants: pathogen pollution, climate change and agrotechnology drivers. Trends Ecol. Evol. 19:535–544.

Auton A, McVean G A T. 2007. Recombination rate estimation in the presence of hotspots. Genome Res. 17:1219–1227.

Aviv G, Tsyba K, Steck N, Rahav G, Grassi GA, Gal-mor O. 2014. A unique megaplasmid contributes to stress tolerance and pathogenicity of an emergent *Salmonella enterica* serovar Infantis strain. 16:977–994.

Barash I, Manulis-Sasson S. 2007. Virulence mechanisms and host specificity of gall-forming *Pantoea agglomerans*. Trends Microbiol. 15:538–545.

Barash I, Manulis-Sasson S. 2009. Recent evolution of bacterial pathogens: the gall-forming *Pantoea agglomerans* case. Annu. Rev. Phytopathol. 47:133–152.

Bardoel BW, van der Ent S, Pel MJC, Tommassen J, Pieterse CMJ, van Kessel KPM, van Strijp J A G. 2011. *Pseudomonas* evades immune recognition of flagellin in both mammals and plants. PLoS Pathog. 7:e1002206.

Barret M, Egan F, O'Gara F. 2013. Distribution and diversity of bacterial

secretion systems across metagenomic datasets. Environ. Microbiol. Rep. 5:117–126.

Barss HP. 1913. A new filbert disease in Oregon. Oregon Agric. Exp. Stn. Bienn. Crop Pest Hortic. Rep. 14:213–223.

Bartoli C, Roux F, Lamichhane JR. 2016. Molecular mechanisms underlying the emergence of bacterial pathogens: an ecological perspective. Mol. Plant Pathol. 17:303–310.

Belisario BA, Zoina A, Pezza L, Luongo L. 1999. Susceptibility of species of Juglans to pathovars of *Xanthomonas campestris*. 29:75–80.

Bennett PM. 2004. Genome Plasticity. In: Woodford N, Johnson AP, editors. Genomics, Proteomics, and Clinical Bacteriology: Methods and Reviews. Totowa, NJ: Humana Press. p. 71–113.

Bennett PM. 2008. Plasmid encoded antibiotic resistance: acquisition and transfer of antibiotic resistance genes in bacteria. Br. J. Pharmacol. 153:S347–S357.

Berg HC. 2003. The rotary motor of bacterial flagella. Annu. Rev. Biochem. 72:19–54.

Biessy A. 2013. Pathoadaptation dans l'espèce *Xanthomonas arboricola*: Phénotypage par microscopie de souches pathogènes et commensales.

Boccacci P, Botta R. 2009. Investigating the origin of hazelnut (*Corylus avellana* L.) cultivars using chloroplast microsatellites. Genet. Resour. Crop Evol. 56:851–859.

Boch J, Scholze H, Schornack S, Landgraf A, Hahn S, Kay S, Lahaye T, Nickstadt A, Bonas U. 2009. Breaking the code of DNA binding specificity of TAL-Type III effectors. Science. 326:1509–1512.

Boller T, He SY. 2009. Innate Immunity in Plants: An Arms Race between Pattern Recognition Receptors in Plants and Effectors in Microbial Pathogens. 324:742–744.

Booher NJ, Carpenter SCD, Sebra RP, Wang L, Salzberg SL, Leach JE, Bogdanove AJ. 2015. Single molecule real-time sequencing of *Xanthomonas oryzae* genomes reveals a dynamic structure and complex TAL (transcription activator-like) effector gene relationships. Microb. Genomics 1:1–22.

Boto L. 2010. Horizontal gene transfer in evolution: facts and challenges. Proc. Biol. Sci. 277:819–827.

Boudon S, Manceau C, Nottéghem JL. 2005. Structure and Origin of *Xanthomonas arboricola* pv. *pruni* populations causing bacterial spot of stone fruit trees in Western Europe. Am. Phytopathol. Soc. 95:1081–1088.

Boureau T, Kerkoud M, Chhel F, Hunault G, Darasse A, Brin C, Durand K, Hajri A, Poussier S, Manceau C, et al. 2013. A multiplex-PCR assay for identification of the quarantine plant pathogen *Xanthomonas axonopodis* pv. *phaseoli*. J. Microbiol. Methods 92:42–50.

Bren A, Eisenbach M. 2000. How signals are heard during bacterial chemotaxis: protein-protein interactions in sensory signal propagation. J. Bacteriol. 182:6865–6873.

Brown SP, Cornforth DM, Mideo N. 2012. Evolution of virulence in opportunistic pathogens: Generalism, plasticity, and control. Trends Microbiol. 20:336–342.

Bruhn C a W, Audelin AM, Helleberg M, Bjorn-Mortensen K, Obel N, Gerstoft J, Nielsen C, Melbye M, Medstrand P, Gilbert MTP, et al. 2014. The origin and emergence of an HIV-1 epidemic: from introduction to endemicity. AIDS 28:1031–1040.

Buckling A, Harrison F, Vos M, Brockhurst M a., Gardner A, West S a., Griffin A. 2007. Siderophore-mediated cooperation and virulence in *Pseudomonas aeruginosa*. FEMS Microbiol. Ecol. 62:135–141.

Burdon JJ, Thrall PH. 2008. Pathogen evolution across the agro-ecological interface: implications for disease management. Evol. Appl. 1:57–65.

Caballero JD, Clark ST, Coburn B, Zhang Y, Wang PW, Donaldson SL, Tullis DE, Yau YCW, Waters VJ, Hwang DM, et al. 2015. Selective Sweeps and Parallel Pathoadaptation Drive *Pseudomonas aeruginosa* evolution in the cystic fibrosis lung. MBio 6:e00981–15.

Canchaya C, Fournous G, Brüssow H. 2004. The impact of prophages on bacterial chromosomes. Mol. Microbiol. 53:9–18.

Cardinale BJ, Duffy JE, Gonzalez A, Hooper DU, Perrings C, Venail P, Narwani A, Mace GM, Tilman D, Wardle DA, et al. 2012. Biodiversity loss and its impact on humanity. Nature 486:59–67.

Carrere S, Cottret L, Rancurel C, Briand M. 2015. Orthomcl-Companion: a user friendly tool to analyse protein families. F1000Research 4:489.

Carver T, Thomson N, Bleasby A, Berriman M, Parkhill J. 2009. DNAPlotter: Circular and linear interactive genome visualization. Bioinformatics 25:119–120.

Casadevall A, Pirofski LA. 2000. Host-pathogen interactions: Basic concepts of microbial commensalism, colonization, infection, and disease. Infect. Immun. 68:6511–6518.

Cesbron S, Briand M, Essakhi S, Gironde S, Boureau T, Manceau C, Fischer-Le Saux M, Jacques M-A. 2015. Comparative genomics of pathogenic and nonpathogenic strains of *Xanthomonas arboricola* unveil molecular and evolutionary events libked to pathoadaptation. Front. Plant Sci. 6:1126.

Chandler M, Mahillon J. Insertion Sequences Revisited. In: Mobile DNA II. American Society of Microbiology. p. 305–366.

Chang Q, Wang W, Regev-Yochay G, Lipsitch M, Hanage WP. 2015. Antibiotics in agriculture and the risk to human health: How worried should we be? Evol. Appl. 8:240–247.

Chin SW, Shaw J, Haberle R, Wen J, Potter D. 2014. Diversification of almonds, peaches, plums and cherries - Molecular systematics and biogeographic history of *Prunus* (Rosaceae). Mol. Phylogenet. Evol. 76:34–48.

Chinchilla D, Bauer Z, Regenass M, Boller T, Felix G. 2006. The *Arabidopsis* receptor kinase FLS2 binds flg22 and determines the specificity of flagellin perception. Plant Cell 18:465–476.

Clarke CR, Cai R, Studholme D, Guttman DS, Vinatzer B a. 2010. *Pseudomonas syringae* Strains Naturally Lacking the Classical *P. syringae* hrp/hrc Locus Are

Common Leaf Colonizers Equipped with an Atypical Type III Secretion System. Mol. plant-microbe Interact. 23:198–210.

Clay K, Kover PX. 1996. The Red Queen Hypothesis and plant/pathogen interactions. Annu. Rev. Phytopathol. 34:29–50.

Cohan FM, Perry EB. 2007. A Systematics for Discovering the Fundamental Units of Bacterial Diversity. Curr. Biol. 17:373–386.

Collard J-M, Place S, Denis O, Rodriguez-Villalobos H, Vrints M, Weill F-X, Baucheron S, Cloeckaert A, Struelens M, Bertrand S. 2007. Travel-acquired salmonellosis due to *Salmonella* Kentucky resistant to ciprofloxacin, ceftriaxone and cotrimoxazole and associated with treatment failure. J. Antimicrob. Chemother. 60:190–192.

Comas I, Coscolla M, Luo T, Borrell S, Holt KE, Kato-Maeda M, Parkhill J, Malla B, Berg S, Thwaites G, et al. 2013. Out-of-Africa migration and Neolithic coexpansion of *Mycobacterium tuberculosis* with modern humans. Nat. Genet. 45:1176–1182.

Constantin EC, Cleenwerck I, Maes M, Baeyen S, Van Malderghem C, De Vos P, Cottyn B. 2016. Genetic characterization of strains named as *Xanthomonas axonopodis* pv. *dieffenbachiae* leads to a taxonomic revision of the *X. axonopodis* species complex. Plant Pathol. 65:792–806.

Coombes BK. 2009. Type III secretion systems in symbiotic adaptation of pathogenic and non-pathogenic bacteria. Trends Microbiol. 17:89–94.

Cornejo OE, Lefébure T, Pavinski Bitar PD, Lang P, Richards VP, Eilertson K, Do T, Beighton D, Zeng L, Ahn SJ, et al. 2013. Evolutionary and population genomics of the cavity causing bacteria *streptococcus mutans*. Mol. Biol. Evol. 30:881–893.

Costerton JW, Lewandowski Z, Caldwell DE, Korber DR, Lappin-Scott HM. 1995. Microbial Biofilms. Annu. Rev. Microbiol. 49:711–745.

Coyne CJ, Mehlenbacher SA, Smith DC. 1998. Sources of Resistance to Eastern Filbert Blight in Hazelnut. J. Am. Soc. Hortic. scince 123:253–257.

Cristofori V, Bizzarri S, Silvestri C, Muleo R, Rugini E, Salvador FR De. 2014. First Evaluations on Vegetative and Productive Performance of Many Hazelnut Cultivars in Latium Region. Acta Hortic.:91–98.

Cunty A, Cesbron S, Poliakoff F, Jacques M a., Manceau C. 2015. Origin of the outbreak in France of *Pseudomonas syringae* pv. *actinidiae* biovar 3, the causal agent of bacterial canker of kiwifruit, revealed by a multilocus variable-number tandem-repeat analysis. Appl. Environ. Microbiol. 81:6773–6789.

Czechowska K, McKeithen-Mead S, Al Moussawi K, Kazmierczak BI. 2014. Cheating by type 3 secretion system-negative *Pseudomonas aeruginosa* during pulmonary infection. Proc. Natl. Acad. Sci. U. S. A. 111:7801–7806.

Danhorn T, Fuqua C. 2007. Biofilm Formation by Plant-Associated Bacteria. Annu. Rev. Microbiol. 61:401–422.

Darling ACE, Mau B, Blattner FR, Perna NT. 2004. Mauve: Multiple Align-

ment of Conserved Genomic Sequence With Rearrangements. Genome Res. 14:1394–1403.

Darrasse A, Carrère S, Barbe V, Boureau T, Arrieta-Ortiz ML, Bonneau S, Briand M, Brin C, Cociancich S, Durand K, et al. 2013. Genome sequence of *Xanthomonas fuscans* subsp. fuscans strain 4834-R reveals that flagellar motility is not a general feature of xanthomonads. BMC Genomics 14:761.

Darsonval a, Darrasse a, Durand K, Bureau C, Cesbron S, Jacques M. 2009. Adhesion and fitness in the bean phyllosphere and transmission to seed of *Xanthomonas fuscans* subsp. *fuscans*. Mol. Plant. Microbe. Interact. 22:747–57.

Darsonval a, Darrasse a, Meyer D, Demarty M, Durand K, Bureau C, Manceau C, Jacques M. 2008. The Type III secretion system of *Xanthomonas fuscans* subsp. fuscans is involved in the phyllosphere colonization process and in transmission to seeds of susceptible beans. Appl. Environ. Microbiol. 74:2669–78.

Darwin Charles. 1859. On the origin of species by means of natural selection, or, The preservation of favoured races in the struggle for life. London:John Murray,.

Daszak P, Cunningham A a, Hyatt AD. 2000. Emerging infectious diseases of wildlife - threats to biodiversity and human health. Science. 287:443–449.

DeJong TM, Tsuji W, Doyle JF, Grossman YL. 1999. Comparative economic efficiency of four peach production systems in California. HortScience 34:73–78.

Demba Diallo M, Monteil CL, Vinatzer B a., Clarke CR, Glaux C, Guilbaud C, Desbiez C, Morris CE. 2012. *Pseudomonas syringae* naturally lacking the canonical type III secretion system are ubiquitous in nonagricultural habitats, are phylogenetically diverse and can be pathogenic. ISME J. 6:1325–1335.

Denamur E, Matic I. 2006. Evolution of mutation rates in bacteria. Mol. Microbiol. 60:820–827.

Didelot X, Darling A, Falush D. 2009. Inferring genomic flux in bacteria. Genome Res. 19:306-317

Didelot X, Falush D. 2007. Inference of bacterial microevolution using multilocus sequence data. Genetics. 175:1251-1266

Didelot X, Guillaume M, Falush D, Darling AE. 2012. Impact of homologous and non-homologous recombination in the genomic evolution of *Escherichia coli*.

Didelot X, Maiden MCJ. 2010. Impact of recombination on bacterial evolution. Trends Microbiol. 18:315–322.

Didelot X, Walker AS, Peto TE, Crook DW, Wilson DJ. 2016. Within-host evolution of bacterial pathogens. Nat. Rev. 14:150–162.

Diepold A, Armitage JP. 2015. Type III secretion systems: the bacterial flagellum and its injectome. Philos. Trans. B 370:20150020.

Dowson WJ. 1943. On the generic names *Pseudomonas*, *Xanthomonas* and *Bacterium* for certain bacterial plant pathogens. Trans. Br. Mycol. Soc. 26:4–14.

Drummond a. J, Rambaut a., Shapiro B, Pybus OG. 2005. Bayesian coalescent inference of past population dynamics from molecular sequences. Mol. Biol. Evol. 22:1185–1192.

Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. 7:214.

Duchêne S, Holt KE, Weill F-X, Le Hello S, Hawkey J, Edwards DJ, Fourment M, Holmes EC. 2016. Genome-scale rates of evolutionary change in bacteria. Genome Res.:in press.

Dye DW, Bradbury JF, Goto M, Hayward AC, Lelliott RA, Schroth MN. 1980. International standards for naming pathovars of phytopathogenic bacteria and a list of pathovar names and pathotype strains. Rev. Plant Pathol. 59:153–168.

Dye DW, Lelliott RA. 1974. *Xanthomonas*. In: Bergey's manual of determinative bacteriology.

Earl D a., vonHoldt BM. 2011. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv. Genet. Resour. 4:359–361.

Eden-Green SJ, Billing E. 1972. Fireblight: Occurrence of Bacterial Strands on Various Hosts under Glasshouse Conditions. Plant Pathol. 21:121–123.

Edgar RC. 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32:1792–1797.

Ehrbar K, Hardt WD. 2005. Bacteriophage-encoded type III effectors in *Salmonella enterica* subspecies 1 serovar Typhimurium. Infect. Genet. Evol. 5:1–9.

Eldholm V, Monteserin J, Rieux A, Lopez B, Sobkowiak B, Ritacco V, Balloux F. 2015. Four decades of transmission of a multidrug-resistant *Mycobacterium tuberculosis* outbreak strain. Nat. Commun. 6:7119.

Engering A, Hogerwerf L, Slingenbergh J. 2013. Pathogen-host-environment interplay and disease emergence. Emerg. Microbes Infect. 2:e5.

Essakhi S, Cesbron S, Fischer-Le Saux M, Bonneau S, Jacques M-A, Manceau C. 2015. Phylogenetic and VNTR analysis identified non-pathogenic lineages within *Xanthomonas arboricola* lacking the canonical type three secretion system. Appl. Environ. Microbiol. 81:5395–5410.

Evanno G, Regnault S, Goudet J. 2005. Detecting the numbers of clusters of individuals using the software STRUCTURE: a simulation study. Mol. Ecol. 14:2611–2620.

Falush D, Stephens M, Pritchard JK. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164:1567–87.

Falush D, Stephens M, Pritchard JK. 2007. Inference of population structure using multilocus genotype data: dominant markers and null alleles. Mol. Ecol. Notes 7:574–578.

Falush D, Wirth T, Linz B, Pritchard JK, Stephens M, Kidd M, Blaser MJ, Graham DY, Vacher S, Perez-Perez GI, et al. 2003. Traces of Human migration in *Helicobacter pylori* populations. Science. 299:1582–1585.

Faust M, Timon B. 1995. Origin and dissemination of Peach. In: Horticultural reviews. p. 311–455.

Felix G, Duran JD, Volko S, Boller T. 1999. Plants have a sensitive perception system for the most conserved domain of bacterial flagellin. Plant J. 18:265–276.

Ferreira RM, Amanda Carolina P. de Oliveira, Leandro M. Moreira JB, Gourbeyre E, Siguier P, Ferro MIT, Ferro J a., Chandler M, Varania AM. 2015. A TALE of Transposition: Tn 3 -Like Transposons Play a Major Role in the Spread of Pathogenicity Determinants of *Xanthomonas citri*. 6:1–12.

Fischer-Le Saux M, Bonneau S, Essakhi S, Manceau C, Jacques M-A. 2015. Aggressive emerging pathovars of *Xanthomonas arboricola* represent widespread epidemic clones that are distinct from poorly pathogenic strains, as revealed by multilocus sequence typing. Appl. Environ. Microbiol. 81:4651–4668.

Flemming H-C, Wingender J, Szewzyk U, Steinberg P, Rice S a., Kjelleberg S. 2016. Biofilms: an emergent form of bacterial life. Nat. Rev. Microbiol. 14:563–575.

Flor HH. 1971. Current Status of the Gene-For-Gene Concept. Annu. Rev. Phytopathol. 9:275–296.

Friesen T, Stukenbrock E, Liu Z, Meinhardt S, Ling H, Faris J, Rasmussen J, Solomon P, McDonald B, Oliver R. 2006. Emergence of a new disease as a result of interspecific virulence gene transfer. Nat. Genet. 38:953–956.

Frutos D, Lopez G. 2012. SEARCH FOR JUGLANS REGIA GENOTYPES RESISTANT/TOLERANT TO *XANTHOMONAS ARBORICOLA* pv. *JUGLANDIS* IN THE FRAMEWORK OF COST ACTION 873. J. Plant Pathol. 94:37–46.

Galan JE, Collmer A. 1999. Type III Secretion Machines: Bacterial Devices for Protein Delivery into Host Cells. Science. 284:1322–1328.

Garita-cambronero J, Palacio-bielsa A, López MM, Cubero J. 2016a. Comparative Genomic and Phenotypic Characterization of Pathogenic and Non-Pathogenic Strains of *Xanthomonas arboricola* Reveals Insights into the Infection Process of Bacterial Spot Disease of Stone Fruits. PLoS One 11:e0161977.

Garita-Cambronero J, Palacio-Bielsa A, López MM, Cubero J. 2016b. Draft genome sequence for virulent and avirulent strains of *Xanthomonas arboricola* isolated from *Prunus* spp. in Spain. Stand. Genomic Sci. 11:12.

Garita-Cambronero J, Sena-Vélez M, Palacio-Bielsa a, Cubero J. 2014. Draft Genome Sequence of *Xanthomonas arboricola* pv. *pruni* Strain Xap33, Causal Agent of Bacterial Spot Disease on Almond. Genome Announc. 2:e00440–14.

Garrett K a, Dendy SP, Frank EE, Rouse MN, Travers SE. 2006. Climate change effects on plant disease: genomes to ecosystems. Annu. Rev. Phytopathol. 44:489–509.

Gaumann E. 1921. Investigation on the blood disease of banana in Celebes. Meded. van het Inst. voor Plantenziekten 50:47.

Gegner JA, Graham DR, Roth AF, Dahlquist FW. 1992. Assembly of an MCP receptor, CheW, and kinase CheA complex in the bacterial chemotaxis signal transduction pathway. Cell 70:975–982.

Geleta M. 2007. Genetic diversity, phylogenetics and molecular systematics of Guizotia Cass. (Asteraceae).

Geleta M, Ortiz R. 2013. The importance of *Guizotia abyssinica* (niger) for sustainable food security in Ethiopia. Genet. Resour. Crop Evol. 60:1763–1770.

Gilbert W. 1978. Why genes in pieces? Nature 271:501.

Gillings MR, Holley MP, Stokes HW, Holmes AJ. 2005. Integrons in *Xanthomonas*: A source of species genome diversity. Proc. Natl. Acad. Sci. 102:4419–4424.

Gladieux P, Zhang XG, Afoufa-Bastien D, Valdebenito Sanhueza RM, Sbaghi M, Le Cam B. 2008. On the origin and spread of the scab disease of apple: Out of central Asia. PLoS One 3:e1455.

Golubchik T, Batty EM, Miller RR, Farr H, Young BC, Larner-Svensson H, Fung R, Godwin H, Knox K, Votintseva A, et al. 2013. Within-Host Evolution of *Staphylococcus aureus* during Asymptomatic Carriage. PLoS One 8:e61319.

Gómez-Alpizar L, Carbone I, Ristaino JB. 2007. An Andean origin of *Phytophthora infestans* inferred from mitochondrial and nuclear gene genealogies. Proc. Natl. Acad. Sci. U. S. A. 104:3306–3311.

González-Escalona N, Martinez-Urtaza J, Romero J, Espejo RT, Jaykus LA, DePaola A. 2008. Determination of molecular phylogenetics of *Vibrio parahaemolyticus* strains by multilocus sequence typing. J. Bacteriol. 190:2831–2840.

Gophna U, Ron EZ, Graur D. 2003. Bacterial type III secretion systems are ancient and evolved by multiple horizontal-transfer events. Gene 312:151–163.

Green SK, Schroth MN, Cho JJ, Kominos SD, Vitanza-Jack VB. 1975. Agricultural Plants and Soil. Appl. Environ. Microbiol. 28:987–991.

Griffin AS, West S a. 2002. Kin selection: fact and fiction. Trends Ecol. Evol. 17:15–21.

Grimault V.1, Olivier V.2, Rolland M.3 DA. and JMA. 2014. Detection of *Xanthomonas axonopodis* pv. *phaseoli* and *Xanthomonas axonopodis* pv. *phaseoli* var. *fuscans* on *Phaseolus vulgaris* . Int. Rules Seed Test.

Groisman EA, Ochman H. 1996. Pathogenicity islands: bacterial evolution in quantum leaps. Cell 87:791–794.

Guimbaud J-F. 2012. Etude de l'adaptation à l'hôte de bactéries épiphytes non-pathogènes des espèces *Xanthomonas arboricola*, *X. axonopodis* et *X. campestris* par des approches génétiques et des tests *in planta*.

Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. 59:307–21.

Guiserix M. 2009. Mécanismes d'émergence des maladies infectieuses: étude par la modélisation du rôle de la protection de groupe , dans des populations hôtes homogènes ou structurées spatialement. :324.

Gunn BF, Aradhya M, Salick JM, Miller AJ, Yongping Y, Lin L, Xian H. 2010. Genetic variation in walnuts (*Juglans regia* and *J. sigillata*; Juglandaceae): Species

distinctions, human impacts, and the conservation of agrobiodiversity in Yunnan, China. Am. J. Bot. 97:660–671.

Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD. 2009. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. PLoS Genet. 5:e1000695.

Guttman DS, Gropp SJ, Morgan RL, Wang PW. 2006. Diversifying selection drives the evolution of the type III secretion system pilus of *Pseudomonas syringae*. Mol. Biol. Evol. 23:2342–2354.

Guy L, Roat Kultima J, Andersson SGE. 2010. genoPlotR: comparative gene and genome visualization in R. Bioinformatics 26:2334–2335.

Hacker J, Blum-Oehler G, Mühldorfer I, Tschäpe H. 1997. Pathogenicity islands of virulent bacteria: structure, function and impact on microbial evolution. Mol. Microbiol. 23:1089–1097.

Hajri A, Brin C, Hunault G, Lardeux F, Lemaire C, Manceau C, Boureau T, Poussier S. 2009. A "repertoire for repertoire" hypothesis: repertoires of type three effectors are candidate determinants of host specificity in *Xanthomonas*. PLoS One 4:e6632.

Hajri A, Meyer D, Delort F, Guillaumès J, Brin C, Manceau C. 2010. Identification of a genetic lineage within *Xanthomonas arboricola* pv. *juglandis* as the causal agent of vertical oozing canker of Persian (English) walnut in France. Plant Pathol. 59:1014–1022.

Hajri A, Pothier JF, Fischer-Le Saux M, Bonneau S, Poussier S, Boureau T, Duffy B, Manceau C. 2012. Type three effector gene distribution and sequence analysis provide new insights into the pathogenicity of plant-pathogenic *Xanthomonas arboricola*. Appl. Environ. Microbiol. 78:371–84.

Händel N, Hoeksema M, Mata F, Brul S, Kuile H. 2016. Effects of Stress , Reactive Oxygen Species , and the SOS Response on De Novo Acquisition of Antibiotic Resistance in *Escherichia coli*. 60:1319–1327.

Harwich MD, Alves JM, Buck GA, Strauss III JF, Patterson JL, Oki AT, Girerd PH, Jefferson KK. 2010. Drawing the line between commensal and pathogenic *Gardnerella vaginalis* through genome analysis and virulence studies. BMC Genomics 11:375.

HAUBEN L, VAUTERIN L, SWINGS J, MOORE ERB. 1997. Comparison of 16S Ribosomal DNA Sequences of All *Xanthomonas* Species. Int. J. Syst. Bacteriol. 47:328–335.

Haubold B, Hudson RR. 2000. LIAN 3.0: detecting linkage disequilibrium in multilocus data. Linkage Analysis. Bioinformatics 16:847–848.

Haworth RH, Spiers AG. 1992. Isolation of *Xanthomonas campestris* pv. *populi* from stem lesions on *Salix matsudana* X alba "Aokautere" in New Zealand. Eur J For. Pathol 22:247–251.

Hengge-Aronis R. 2002. Recent insights into the general stress response regulatory network in *Escherichia coli*. J. Mol. Microbiol. Biotechnol. 4:341–346.

Henz GP. 2002. Present situation of arracacha (*Arracacia xanthorrhiza*) diseases in Brazil. Hortic. Bras. 20:135–144.

Hirano SS, Upper CD. 2000. Bacteria in the leaf ecosystem with emphasis on *Pseudomonas syringae*! a pathogen, ice nucleus, and epiphyte. Microbiol. Mol. Biol. Rev. 64:624–653.

Hougouvieux-Cotte-Pattat N, Condemine G, Shevchik VE. 2014. Bacterial pectate lyases, structural and functional diversity. Environ. Microbiol. Rep. 6:427–440.

Huang C-L, Pu P-H, Huang H-J, Sung H-M, Liaw H-J, Chen Y-M, Chen C-M, Huang M-B, Osada N, Gojobori T, et al. 2015. Ecological genomics in *Xanthomonas*: the nature of genetic adaptation with homologous recombination and host shifts. BMC Genomics 16:188.

Hulko M, Berndt F, Gruber M, Linder JU, Truffault V, Schultz A, Martin J, Schultz JE, Lupas AN, Coles M. 2006. The HAMP Domain Structure Implies Helix Rotation in Transmembrane Signaling. Cell 126:929–940.

Huson DH, Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. 23:254–267.

Hyland EM, Wallace EWJ, Murray AW. 2014. A model for the evolution of biological specificity: A cross-reacting DNA-binding protein causes plasmid incompatibility. J. Bacteriol. 196:3002–3011.

Ibarra Caballero J, Zerillo M, Snelling J, Boucher C, Tisserat N. 2013. Genome sequence of *Xanthomonas arboricola* pv. *corylina*, isolated from Turkish Filbert in Colorado. Genome Announc.

Ignatov AN, Kyrova EI, Vinogradova S V, Kamionskaya AM, Schaad NW, Luster DG. 2014. Draft Genome Sequence of *Xanthomonas arboricola* Strain 3004, a Causal Agent of Bacterial Disease on Barley. Genome Announc. 3:e01572–14.

Ignatov AN, Punina N V, Matveeva E V. 2010. *Xanthomonas arboricola*-bacterial pathogen of crops in Russia. Zashchita i Karantin Rastenii 4:41–43.

Indiana A. 2014. Rôles du chimiotactisme et de la mobilité flagellaire dans la fitness des *Xanthomonas*.

Ishihara M, Takikawa Y, Akiba M, Kawabe Y. 2015. A new bacterial disease observed on *Quercus myrsinifolia*. For. Pathol. 45:459–466.

Izoré T, Job V, Dessen A. 2011. Biogenesis, regulation, and targeting of the type III secretion system. Structure 19:603–612.

Jacobs JM, Pesce C, Lefeuvre P, Koebnik R. 2015. Comparative genomics of a cannabis pathogen reveals insight into the evolution of pathogenicity in *Xanthomonas*. Front. Plant Sci. 6:431.

Jacques M, Josi K, Darrasse A, Samson R. 2005. *Xanthomonas axonopodis* pv. *phaseoli* var. *fuscans* Is Aggregated in Stable Biofilm Population Sizes in the Phyllosphere of Field- Grown Beans. 71:2008–2015.

Jacques M-A, Arlat M, Boulanger A, Boureau T, Carrère S, Cesbron S, Chen NWG, Cociancich S, Darrasse A, Denancé N, et al. 2016. Using Ecology,

Physiology, and Genomics to Understand Host Specificity in *Xanthomonas*. Annu. Rev. Phytopathol. 54:163–187.

Jacques M-A, Denancé N, Legendre B, Morel E, Briand M, Mississipi S, Durand K, Olivier V, Portier P, Poliakoff F, et al. 2016. New Coffee Plant-Infecting *Xylella fastidiosa* variants derived via homologous recombination. App 82:1556–1568.

Jacques M-A, Durand K, Orgeur G, Balidas S, Fricot C, Bonneau S, Quillévéré A, Audusseau C, Olivier V, Grimault V, et al. 2012. Phylogenetic analysis and polyphasic characterization of *Clavibacter michiganensis* strains isolated from tomato seeds reveal that nonpathogenic strains are distinct from *C. michiganensis* subsp. *michiganensis*. Appl. Environ. Microbiol. 78:8388–402.

Jakobsson M, Rosenberg N. 2007. CLUMMP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23:1801–1806.

Janse JD, Rossi MP, Gorkink RFJ, Derks JHJ, Swings J, Janssens D, Scortichini M. 2001. Bacterial leaf blight of strawberry (*Fragaria* (x) *ananassa*) caused by a pathovar of *Xanthomonas arboricola*, not similar to *Xanthomonas fragariae* Kennedy & King. Description of the causal organism as *Xanthomonas arboricola* pv. *fragariae* (pv. nov., comb. no. Plant Pathol. 50:653–665.

Jombart T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24:1403–5.

Jones JDG, Dangl JL. 2006. The plant immune system. Nature 44:323–329.

Joseph SJ, Didelot X, Rothschild J, De Vries HJC, Morré S a., Read TD, Dean D. 2012. Population genomics of *chlamydia trachomatis*: Insights on drift, selection, recombination, and population structure. Mol. Biol. Evol. 29:3933–3946.

Joubert JJ, Truter SJ. 1972. A variety of *Xanthomonas campestris* pathogenic to *Zantedeschia aethiopica*. Netherlands J Plant Pathol 78:212–217.

Journet L, Hughes KT, Cornelis GR. 2005. Type III secretion: a secretory pathway serving both motility and virulence. Mol. Membr. Biol. 22:41–50.

De Kam M. 1983. *Xanthomonas campestris* pv. *populi*, the causal agent of bark necrosis in poplar. Eur J Plant Pathol 90:13–22.

Karamura G, Smith J, Studholme D, Kubiriba J, Karamura E. 2015. Comparative pathogenicity studies of the *Xanthomonas vasicola* species on maize, sugarcane and banana. African J. Plant Sci. 9:385–400.

Kearns DB. 2010. A field guide to bacterial swarming motility. Nat Rev Micro 8:634–644.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markovitz S, Duran C, et al. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28:1647–1649.

Kilian M, Poulsen K, Blomqvist T, Håvarstein LS, Bek-Thomsen M, Tettelin H, Sørensen UBS. 2008. Evolution of *Streptococcus pneumoniae* and its close

commensal relatives. PLoS One 3:e2683.

Kirov SM. 2003. Bacteria that express lateral flagella enable dissection of the multifunctional roles of flagella in pathogenesis. FEMS Microbiol. Lett. 224:151–159.

Kirzinger MWB, Butz CJ, Stavrinides J. 2015. Inheritance of *Pantoea* type III secretion systems through both vertical and horizontal transfer. Mol. Genet. genomics 290:2075–2088.

Kirzinger MWB, Stavrinides J. 2012. Host specificity determinants as a genetic continuum. Trends Microbiol. 20:88–93.

Kong KF, Vuong C, Otto M. 2006. *Staphylococcus* quorum sensing in biofilm formation and infection. Int. J. Med. Microbiol. 296:133–139.

Konstantinidis KT, Tiedje JM. 2005. Genomic insights that advance the species definition for prokaryotes. Proc. Natl. Acad. Sci. U. S. A.

Kort EN, Goy MF, Larsen S, Adler J. 1975. Methylation of a membrane protein involved in bacterial chemotaxis. Proc. Natl. Acad. Sci. 72:3939–3943.

Krell T, Lacal J, Busch A, Silva-Jiménez H, Guazzaroni M-E, Ramos JL. 2010. Bacterial Sensor Kinases: Diversity in the Recognition of Environmental Signals. Annu. Rev. Microbiol. 64:539–559.

Kremer D, Cavlovic J, Bozic M, Dubravac T. 2008. Distribution and Management of Black Walnut (*Juglans nigra* L .) in Croatia. Period. Biol. 110:317–321.

Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra M a. 2009. Circos: an information esthetic for comparative genomics. Genome Res. 19:1639–1645.

Lackner G, Moebius N, Hertweck C. 2011. Endofungal bacterium controls its host by an *hrp* type III secretion system. ISME J. 5:252–261.

Laizet Y, Chaumeil P, Frigerio JM, Franc A. 2014. A primer for virtual Biodiversity Lab.

Lamichhane JR, Varvaro L. 2014. *Xanthomonas arboricola* disease of hazelnut: current status and future perspectives for its management. Plant Pathol. 63:243–254.

Lawrence JG. 2009. Microbial Evolution: Enforcing Cooperation by Partial Kin Selection. Curr. Biol. 19:943–945.

Lawson DJ, Hellenthal G, Myers S, Falush D. 2012. Inference of population structure using dense haplotype data. PLoS Genet. 8:e1002453.

Lee YA, Chen KP, Chang YC. 2004. First report of bacterial leaf blight of white-flowered calla lily caused by *Xanthomonas campestris* pv. *zantedeschiae* in Taiwan. Plant Pathol. 54:239.

Lemaire C, De Gracia M, Leroy T, Michalecka M, Lindhard-Pedersen H, Guerin F, Gladieux P, Le Cam B. 2016. Emergence of new virulent populations of apple scab from nonagricultural disease reservoirs. New Phytol. 209:1220–1229.

Leroy T, Le Cam B, Lemaire C. 2014 Jan 7. When virulence originates from non-agricultural hosts: New insights into plant breeding. Infect. Genet. Evol.

Li L, Stoeckert CJJ, Roos DS. 2003. OrthoMCL: Identification of Ortholog

Groups for Eukaryotic Genomes. Genome Res. 13:2178–2189.

Li S, Wang Y, Wang S, Fang A, Wang J, Liu L, Zhang K, Mao Y, Sun W. 2015. The Type III Effector AvrBs2 in *Xanthomonas oryzae* pv. *oryzicola* Suppresses Rice Immunity and Promotes Disease Development. Mol. Plant. Microbe. Interact. 28:869–880.

Li Y, Hu Y, Bolund L, Wang J. 2010. State of the art de novo assembly of human genomes from massively parallel sequencing data. Hum. Genomics 4:271–277.

Librado P, Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451–2.

Lindeberd M, Cunnas S, Collmer A. 2009. The evolution of *Pseudomonas syringae* host specificity and type III effector repertoires. Mol. Plant Pathol. 10:767–775.

Lindeberg M, Cunnac S, Collmer A. 2012. *Pseudomonas syringae* type III effector repertoires: Last words in endless arguments. Trends Microbiol. 20:199–208.

Lindow SE, Brandl MT. 2003. Microbiology of the Phyllosphere. Appl. Environ. Microbiol. 69:1875–1883.

Liu R, Ochman H. 2007. Stepwise formation of the bacterial flagellar system. Proc. Natl. Acad. Sci. 104:7116–7121.

Loconsole G, Saponari M, Boscia D, D'Attoma G, Morelli M, Martelli GP, Almeida RPP. 2016. Intercepted isolates of *Xylella fastidiosa* in Europe reveal novel genetic diversity. Eur. J. Plant Pathol. 146:85–94.

Lopez C a., Winter SE, Rivera-Chávez F, Xavier MN, Poon V, Nuccio SP, Tsolis RM, Bäumler AJ. 2012. Phage-mediated acquisition of a type III secreted effector protein boosts growth of *salmonella* by nitrate respiration. MBio 3:e00143–12.

Luisetti J, Jailloux F, Germain E, Prunier JP, Gardan L. 1975. Caractérisation de *Xanthomonas corylina* (Miller et al.) Starr et Burkholder responsable de la bactériose du noisetier récemment observée en France. CR Séances Acad Agric Fr. 61:845–849.

Macho AP, Zipfel C. 2015. Targeting of plant pattern recognition receptor-triggered immunity by bacterial type-III secretion system effectors. Curr. Opin. Microbiol. 23:14–22.

Macnab RM, Ornston MK. 1977. Normal-to-curly flagellar transitions and their role in bacterial tumbling. Stabilization of an alternative quaternary structure by mechanical force. J. Mol. Biol. 112:1–30.

Manulis S, Barash I. 2003. *Pantoea agglomerans* pv. *gypsophilae* and *betae*, recently evolved pathogens? Mol. Plant Pathol. 4:307–314.

Marguerettaz M, Pieretti I, Gayral P, Puig J, Brin C, Cociancich S, Poussier S, Rott P, Royer M. 2011. Genomic and evolutionary features of the SPI-1 type III secretion system that is present in *Xanthomonas albilineans* but is not essential for xylem colonization and symptom development of sugarcane leaf scald. Mol. plant-microbe Interact. 24:246–259.

Martelli GP, Boscia D, Porcelli F, Saponari M. 2016. The olive quick decline

syndrome in south-east Italy: a threatening phytosanitary emergency. Eur. J. Plant Pathol. 144:235–243.

Martinez-Urtaza J, Huapaya B, Gavilan RG, Blanco-Abad V, Ansede-Bermejo J, Cadarso-Suarez C, Figueiras A, Trinanes J. 2008. Emergence of Asiatic *Vibrio* diseases in South America in phase with El Niño. Epidemiology 19:829–837.

Martinez-Urtaza J, Trinanes J, Gonzalez-Escalona N, Baker-Austin C. 2016. Is El Niño a long-distance corridor for waterborne disease? Nat. Microbiol. 1:16018.

Mattick JS. 2002. Type IV Pili and Twitching Motility. Annu. Rev. Microbiol. 56:289–314.

Maurelli AT. 2007. Black holes, antivirulence genes, and gene inactivation in the evolution of bacterial pathogens. FEMS Microbiol. Lett. 267:1–8.

Mavrodi D V., Joe A, Mavrodi O V., Hassan K a., Weller DM, Paulsen IT, Loper JE, Alfano JR, Thomashow LS. 2011. Structural and functional analysis of the type III secretion system from *Pseudomonas fluorescens* Q8r1-96. J. Bacteriol. 193:177–189.

Maynard Smith J, Feil EJ, Smith NH. 2000. Population structure and evolutionary dynamics of pathogenic bacteria. Bioessays 22:1115–22.

Maynard-Smith J, Smith N, O'Rourke M, Spratt B. 1993. How clonal are bacteria? Proc. Natl. Acad. Sci. 90:4384–4388.

Mazzaglia A, Studholme DJ, Taratufolo MC, Cai R, Almeida NF, Goodman T, Guttman DS, Vinatzer B a., Balestra GM. 2012. *Pseudomonas syringae* pv. *actinidiae* (PSA) isolates from recent bacterial canker of kiwifruit outbreaks belong to the same genetic lineage. PLoS One 7:e36518.

Mbega ER, Wulff EG, Mabagala RB, Adriko J, Lund OS, Mortensen CN. 2012. Xanthomonads and other yellow-pigmented *Xanthomonas*-like bacteria associated with tomato seeds in Tanzania. African J. Biotechnol. 11:14303–14312.

McCann HC, Guttman DS. 2008. Evolution of the type III secretion system and its effectors in plant-microbe interactions. New Phytol. 177:33–47.

McDermott JE, Corrigan A, Peterson E, Oehmen C, Niemann G, Cambronne ED, Sharp D, Adkins JN, Samudrala R, Heffron F. 2011. Computational prediction of type III and IV secreted effectors in gram-negative bacteria. Infect. Immun. 79:23–32.

McDonald B a, Linde C. 2002. Pathogen population genetics, evolutionary potentiel, and durable resistance. Annu. Rev. Phytopathol. 40:349–379.

Mckenzie GJ, Lee PL, Lombardo M, Hastings PJ, Rosenberg SM. 2001. SOS Mutator DNA Polymerase IV Functions in Adaptive Mutation and Not Adaptive Amplification. 7:571–579.

Mehlenbacher SA, Olsen J. 1997. THE HAZELNUT INDUSTRY IN OREGON, USA. Acta Hortic.:337–346.

Mekalanos JJ. 1992. Environmental signals controlling expression of virulence determinants in bacteria. J. Bacteriol. 174:1–7.

Mhedbi-Hajri N, Darrasse A, Pigné S, Durand K, Fouteau S, Barbe V, Manceau C, Lemaire C, Jacques M-A. 2011. Sensing and adhesion are adaptive functions in the plant pathogenic xanthomonads. BMC Evol. Biol. 11:67.

Mhedbi-Hajri N, Hajri A, Boureau T, Darrasse A, Durand K, Brin C, Fischer-Le Saux M, Manceau C, Poussier S, Pruvost O, et al. 2013. Evolutionary history of the plant pathogenic bacterium *Xanthomonas axonopodis*. PLoS One 8:e58474.

Miller PW, Bollen WB, Simmons JE, Gross HN, Barss H. 1940. The pathogen of filbert bacteriosis compared with *Phytomonas juglandis*, the cause of walnut blight. Phytopathology 30:713–733.

Miskinyte M, Sousa A, Ramiro RS, Moura de Sousa JA, Kotlinowski J, Caramalho I, Magalhães S, Soares MP, Gordo I. 2013. The Genetic Basis of *Escherichia coli* Pathoadaptation to Macrophages. PLoS Pathog. 9:e1003802.

Mohr TJ, Liu H, Yan S, Morris CE, Castillo J a, Jelenska J, Vinatzer B a. 2008. Naturally occurring nonpathogenic isolates of the plant pathogen *Pseudomonas syringae* lack a type III secretion system and effector gene orthologues. J. Bacteriol. 190:2858–70.

Montano V, Didelot X, Foll M, Linz B, Reinhardt R, Suerbaum S, Moodley Y, Jensen JD. 2015. Worldwide Population Structure, Long Term Demography, and Local Adaptation of *Helicobacter pylori*. Genetics 200:947–963.

Monteil CL, Cai R, Liu H, Llontop MEM, Leman S, Studholme DJ, Morris CE, Vinatzer B a. 2013. Nonagricultural reservoirs contribute to emergence and evolution of Pseudomonas syringae crop pathogens. New Phytol. 199:800–11.

Monteil CL, Yahara K, Studholme DJ, Mageiros L, Swingle B, Morris CE, Vinatzer BA, Sheppard SK. 2016. Population-genomic insights into emergence, crop-adaptation, and dissemination of *Pseudomonas syringae* pathogens. Microb. Genomics:in press.

Monteiro-Vitorello CB, de Oliveira MC, Zerillo MM, Varani AM, Civerolo E, Van Sluys M-A. 2005. *Xylella* and *Xanthomonas* Mobil'omics. OMICS 9:146–159.

Montier MS. 2015. Etude de la transmission de bactéries du genre *Xanthomonas* à et par les graines de haricot.

Moore WEC, Stackebrandt E, Kandler O, Colwell RR, Krichevsky MI, Truper HG, Murray RGE, Wayne LG, Grimont PAD, Brenner DJ, et al. 1987. Report of the Ad Hoc Committee on Reconciliation of Approaches to Bacterial Systematics. Int. J. Syst. Evol. Microbiol. 37:463–464.

Morens DM, Folkers GK, Fauci AS. 2004. The challenge of emerging and re-emerging infectious diseases. Nature 430:242–249.

Morris CE, Monier J, Jacques MA. 1997. Methods for Observing Microbial Biofilms Directly on Leaf Surfaces and Recovering Them for Isolation of Culturable Microorganisms. 63:1570–1576.

Morris CE, Monier J-M. 2003. THE ECOLOGICAL SIGNIFICANCE OF BIOFILM FORMATION BY PLANT-ASSOCIATED BACTERIA. Annu. Rev. Phy-

topathol. 41:429–453.

Morse SS. 1995. Factors in the emergence of infectious diseases. Emerg. Infect. Dis. 1:7–15.

Moscou MJ, Bogdanove AJ. 2009. A Simple Cipher Governs DNA Recognition by TAL Effectors. Science. 326:1501–1501.

Mustapha MM, Marsh JW, Krauland MG, Fernandez JO, Lemos APS De, Hotopp JCD, Wang X, Mayer LW, Lawrence JG, Hiller NL, et al. 2015. Genomic Epidemiology of Hypervirulent Serogroup W, ST-11 *Neisseria meningitidis*. EBIOMedicine 2:1447–1455.

Myung IS, Moon SY, Jeong IH, Lee SW, Ra DS. 2010. A new disease, bacterial fruit rot of jujube, caused by *Xanthomonas arboricola* in Korea. New Dis. reports 22:12.

Naish S, Dale P, Mackenzie JS, McBride J, Mengersen K, Tong S. 2014. Climate change and dengue: a critical and systematic review of quantitative modelling approaches. BMC Infect. Dis. 14:167.

Naum M, Brown EW, Mason-Gamer RJ. 2009. Phylogenetic evidence for extensive horizontal gene transfer of type III secretion system genes among enterobacterial plant pathogens. Microbiology 155:3187–3199.

Naushad S, Adeolu M, Wong S, Sohail M, Schellhorn HE, Gupta RS. 2015. A phylogenomic and molecular marker based taxonomic framework for the order Xanthomonadales: proposal to transfer the families Algiphilaceae and Solimonadaceae to the order Nevskiales ord. nov. and to create a new family within the order Xanthomonadales, . Antonie Van Leeuwenhoek 107:467–485.

Nei M. 1987. Molecular Evolutionary Genetics. In: Columbia Univ. Press, New York.

Niu X-N, Wei Z-Q, Zou H-F, Xie G-G, Wu F, Li K-J, Jiang W, Tang J-L, He Y-Q. 2015. Complete sequence and detailed analysis of the first indigenous plasmid from *Xanthomonas oryzae* pv. *oryzicola*. BMC Microbiol. 15:233.

Nizan R, Barash I, Valinsky L, Lichter A, Manulis S. 1997. The presence of hrp genes on the pathogenicity-associated plasmid of the tumorigenic bacterium *Erwinia herbicola* pv. *gypsophilae* . Mol. plant-microbe Interact. 10:677–682.

Noel L, Thieme F, Nennstiel D, Bonas U. 2002. Two Novel type III-secreted proteins of *Xanthomonas campestris* pv. *vesicatoria* are encoded within the hrp pathogenicity island. J. Bacteriol. 184:1340–1348.

Nogueira T, Rankin DJ, Touchon M, Taddei F, Brown SP, Rocha EPC. 2009. Horizontal Gene Transfer of the Secretome Drives the Evolution of Bacterial Cooperation and Virulence. Curr. Biol. 19:1683–1691.

Nunney L, Schuenzel EL, Scally M, Bromley RE, Stouthamer R. 2014. Large scale intersubspecific recombination in the plant pathogenic bacterium *Xylella fastidiosa* is associated with the host shift to mulberry.

Appl. Environ. Microbiol. 80:3025–3033.

Nürnberger T, Brunner F, Kemmerling B, Piater L. 2004. Innate immunity in plants and animals: strinking similarities and obvious differences. Immunol. Rev. 198:249–266.

Ochiai H, Inoue Y, Takeya M, Sasaki A, Kaku H. 2005. Genome sequence of *Xanthomonas oryzae* pv. *oryzae* suggests contribution of large numbers of effector genes and insertion sequences to its race diversity. Japan Agric. Res. Q. 39:275–287.

Oh SC, Kim JF, Beer S V. 2005. The Hrp pathogenicity island of *Erwinia amylovora* and identification of three novel genes required for syntenic infection. Mol. Plant Pathol. 6:125–138.

Otto M. 2013. Staphylococcal Infections: Mechanisms of Biofilm Maturation and Detachment as Critical Determinants of Pathogenicity. Annu. Rev. Med. 64:175–188.

Palacio-Bielsa A, Cambra MA, Cubero J, Rossello M, Lopez M. 2014. Mancha bacteriana de los frutales de huesco y del almendro (Xanthomonas arboricola pv. pruni), una grave enfermedad emergente en Espana. Phytoma-Espana 259:36–40.

Palacio-Bielsa A, Rosello M, Cambra MA, Lopez M. 2010. First report on Almond in Europe of Bacterial Spot Disease of Stone Fruit caused by *Xanthomonas arboricola pv. pruni*. Plant Dis. 94:786–786.

Pallen MJ, Wren BW. 2007. Bacterial pathogenomics. Nature 449:835–42.

Palme AE, Vendramin GG. 2002. Chloroplast DNA variation, postglacial recolonization and hybridization in hazel, *Corylus avellana*. Mol. Ecol. 11:1769–1780.

Paradis E, Claude J, Strimmer K. 2004. APE: Analyses of phylogenetics and evolution in R language. Bioinformatics 20:289–290.

Park WH, Williams AW. 1917. Pathogenic microorganisms.

Parker IM, Gilbert GS. 2004. The Evolutionary Ecology of Novel Plant-Pathogen Interactions. Annu. Rev. Ecol. Evol. Syst. 35:675–700.

Parkinson N, Aritua V, Heeney J, Cowie C, Bew J, Stead D. 2007. Phylogenetic analysis of *Xanthomonas* species by comparison of partial gyrase B gene sequences. Int. J. Syst. Evol. Microbiol. 57:2881–2887.

Parkinson N, Cowie C, Heeney J, Stead D. 2009. Phylogenetic structure of *Xanthomonas* determined by comparison of gyrB sequences. Int. J. Syst. Evol. Microbiol. 59:264–274.

Peeters N, Carrère S, Anisimova M, Plener L, Cazalé A-C, Genin S. 2013. Repertoire, unified nomenclature and evolution of the Type III effector gene set in the *Ralstonia solanacearum* species complex. BMC Genomics 14:859.

Pereira ALG, Paradella Filho O, Zagatto AG. 1971. Uma nova doença bacteriana da mandioquinha-salsa (*Arracacia xanthorrhiza*) causada por *Xanthomonas arracaciae* n. sp. Arq. Inst. Biol. (Sao. Paulo). 38:99–108.

Pérez-Bueno ML, Pineda M, Díaz-Casado E, Barón M. 2015. Spatial and temporal dynamics of primary and secondary metabolism in *Phaseolus vulgaris* challenged by *Pseudomonas syringae*. Physiol. Plant. 153:161–174.

Perilla-Henao LM, Casteel CL. 2016. Vector-borne bacterial plant pathogens:

Interactions with hemipteran insects and plants. Front. Plant Sci. 7:1163.

Petri L. 1934. Rassegna dei casi fitopatologici osservati nel 1933. Boll. della Stn. di Patol. Veg. di Roma:328–330.

Pierce N. B. 1901. Walnut bacteriosis. Bot. Gaz. 31:272–273.

Pieretti I, Cociancich S, Bolot S, Carrèr S, Morisset A, Rott P, Royer M. 2015. Full genome sequence analysis of two isolates reveals a novel *xanthomonas* species close to the sugarcane pathogen *Xanthomonas albilineans*. Genes (Basel). 6:714–733.

Pieretti I, Royer M, Barbe V, Carrere S, Koebnik R, Cociancich S, Couloux A, Darrasse A, Gouzy J, Jacques M-A, et al. 2009. The complete genome sequence of *Xanthomonas albilineans* provides new insights into the reductive genome evolution of the xylem-limited Xanthomonadaceae. BMC Genomics 10:616.

Pirnay JP, Bilocq F, Pot B, Cornelis P, Zizi M, Van Eldere J, Deschaght P, Vaneechoutte M, Jennes S, Pitt T, et al. 2009. *Pseudomonas aeruginosa* population structure revisited. PLoS One 4.

Plummer M, Best N, Cowles K, Vines K. 2006. CODA: convergence diagnosis and output analysis for MCMC. R News 6:7–11.

Portaliou AG, Tsolis KC, Loos MS, Zorzini V, Economou A. 2016. Type III Secretion: Building and Operating a Remarkable Nanomachine. Trends Biochem. Sci. 41:175–189.

Posada D. 2008. jModelTest: phylogenetic model averaging. Mol. Biol. Evol. 25:1253–6.

Pothier JF, Vorhölter FJ, Blom J, Goesmann A, Pühler A, Smits THM, Duffy B. 2011. The ubiquitous plasmid pXap41 in the invasive phytopathogen *Xanthomonas arboricola pv. pruni*: Complete sequence and comparative genomic analysis. FEMS Microbiol. Lett. 323:52–60.

Potnis N, Krasileva K, Chow V, Almeida NF, Patil PB, Ryan RP, Sharlach M, Behlau F, Dow JM, Momol M, et al. 2011. Comparative genomics reveals diversity among xanthomonads infecting tomato and pepper. BMC Genomics 12:146.

Preston GM. 2007. Metropolitan Microbes: Type III Secretion in Multihost Symbionts. Cell Host Microbe 2:291–294.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945–59.

Rademaker JLW, Louws FJ, Schultz MH, Rossbach U, Vauterin L, Swings J, de Bruijn FJ. 2005. A Comprehensive Species to Strain Taxonomic Framework for *Xanthomonas*. Phytopathology 95:1098–1111.

Ranwez V, Harispe S, Delsuc F, Douzery EJP. 2011. MACSE: Multiple alignment of coding SEquences accounting for frameshifts and stop codons. PLoS One 6:e22594.

Rasko D a, Rosovitz MJ, Myers GS a, Mongodin EF, Fricke WF, Gajer P, Crabtree J, Sebaihia M, Thomson NR, Chaudhuri R, et al. 2008. The pangenome structure of *Escherichia coli*: comparative genomic analysis of *E. coli* commensal and

pathogenic isolates. J. Bacteriol. 190:6881–93.

Richter M, Rosselló-Móra R. 2009. Shifting the genomic gold standard for the prokaryotic species definition. Proc. Natl. Acad. Sci. U. S. A. 106:19126–19131.

Rieux A, Balloux F. 2016. Inferences from tip-calibrated phylogenies: A review and a practical guide. Mol. Ecol.

Rodriguez-R LM, Grajales A, Arrieta-Ortiz M, Salazar C, Restrepo S, Bernal A. 2012. Genomes-based phylogeny of the genus *Xanthomonas*. BMC Microbiol. 12:43.

Rohmer L, Guttman DS, Dangl JL. 2004. Diverse evolutionary mechanisms shape the type III effector virulence factor repertoire in the plant pathogen *Pseudomonas syringae*. Genetics 167:1341–1360.

Rossez Y, Wolfson EB, Holmes A, Gally DL, Holden NJ. 2015. Bacterial Flagella: Twist and Stick, or Dodge across the Kingdoms. PLoS Pathog. 11:1–15.

Rougemont Q, Gagnaire P-A, Perrier C, Genthon C, Besnard A-L, Launey S, Evanno G. 2016. Inferring the demographic history underlying parallel genomic divergence among pairs of parasitic and non-parasitic lamprey ecotypes. Mol. Ecol.

Roux B, Bolot S, Guy E, Denancé N, Lautier M, Jardinaud M-F, Fischer-Le Saux M, Portier P, Jacques M-A, Gagnevin L, et al. 2015. Genomics and transcriptomics of *Xanthomonas campestris* species challenge the concept of core type III effectome. BMC Genomics 16:975.

Ryan RP, Vorhölter F-J, Potnis N, Jones JB, Van Sluys M-A, Bogdanove AJ, Dow JM. 2011. Pathogenomics of *Xanthomonas*: understanding bacterium–plant interactions. Nat. Rev. Microbiol. 9:344–355.

Saharan GS, Mehta N, Sangwan MS. 2005. Diseases of oiseed crops.

Sakar SF, Gordon JS, Martin GB, Guttman DS. 2006. Comparative Genomics of host-specific virulence in *Pseudomonas syringae*. Genetics 174:1041–1056.

Sallet E, Gouzy J, Schiex T. 2014. EuGene-PP: a next generation automated annotation pipeline for prokaryotic genomes. Bioinformatics 30:2659–2661.

Santos-Garcia D, Vargas-Chavez C, Moya A, Latorre A, Silva FJ. 2015. Genome evolution in the primary endosymbiont of whiteflies sheds light on their divergence. Genome Biol. Evol. 7:873–888.

Saponari M, Boscia D, Nigro F, Martelli GP. 2013. Identification of DNA sequences related to *Xylella fastidiosa* in oleander, almond and olive trees exhibiting leaf scorch symptoms in Apulia (Southern Italy). J. plant Pathol. 95:668.

Sarris PF, Trantas EA, Baltrus DA, Bull CT, Wechter WP, Yan S, Ververidis F, Almeida NF, Jones CD, Dangl JL, et al. 2013. Comparative genomics of multiple strains of *Pseudomonas cannabina pv. alisalensis*, a potential model pathogen of both monocots and dicots. PLoS One 8:e59366.

Sauquet H. 2013. A practical guide to molecular dating. Comptes Rendus - Palevol 12:355–367.

Savic DJ, Ferretti JJ. 2003. Novel Genomic Rearrangement That Affects Expres-

sion of the *Streptococcus pyogenes* Streptolysin O (*slo*) gene. J. Bacteriol. 185:1857–1869.

Sawada H, Kunugi Y, Watauchi K, Kudo A, Sato T. 2011. Bacterial spot, a new disease of grapevine(*Vitis vinifera*) caused by *Xanthomonas arboricola*. Japanese J. Phytopathol. 77:7–22.

Schaad NW, Postnikova E, Lacy GH, Sechler A, Agarkova I, Stromberg PE, Stromberg VK, Vidaver AK. 2005. Reclassification of *Xanthomonas campestris pv. citri* (ex Hasse 1915) Dye 1978 forms A, B/C/D, and E as *X. smithii* subsp. *citri* (ex Hasse) sp. nov. nom. rev. comb. nov., *X. fuscans* subsp. *aurantifolii* (ex Gabriel 1989) sp. nov. nom. rev. comb. nov., and X. Syst. Appl. Microbiol. 28:494–518.

Schliep KP. 2011. phangorn: Phylogenetic analysis in R. Bioinformatics 27:592–593.

Schofield MJ, Hsieh P. 2003. DNA mismatch repair: molecular mechanisms and biological function. Annu. Rev. Microbiol. 57:579–608.

Schrag SJ, Wiener P. 1995. Emerging infectious disease: what are the relative roles of ecology and evolution? Tree 10:319–324.

Schubert S, Darlu P, Clermont O, Wieser A, Magistro G, Hoffmann C, Weinert K, Tenaillon O, Matic I, Denamur E. 2009. Role of intraspecies recombination in the spread of pathogenicity islands within the *Escherichia coli* species. PLoS Pathog. 5:e1000257.

Schwartz AR, Potnis N, Timilsina S, Wilson M, Patané J, Martins J, Minsavage G V, Dahlbeck D, Akhunova A, Vallad GE, et al. 2015. Phylogenomics of *Xanthomonas* field strains infecting pepper and tomato reveals diversity in effector repertoires and identifies determinants of host specificity. Front. Microbiol.

Scortichini M, Marcelletti S, Ferrante P, Petriccione M, Firrao G. 2012. *Pseudomonas syringae* pv. *actinidiae*: A re-emerging, multi-faceted, pandemic pathogen. Mol. Plant Pathol. 13:631–640.

Scortichini M, Marchesi U, Di Prospero P. 2001. Genetic diversity of *Xanthomonas arboricola* pv. *juglandis* (synonyms: *X. campestris* pv. *juglandis*; *X. juglandis* pv. *juglandis*) Strains from different geographical areas shown by repetitive polymerase chain reaction genomic fingerprinting. J. Phytopathol. 149:325–332.

Scortichini M, Rossi MP. 2003. Genetic diversity of *Xanthomonas arboricola* pv. *fragariae* strains and comparison with some other *X. arboricola* pathovars using repetitive PCR genomic fingerprinting. J. Phytopathol. 151:113–119.

Scortichini M, Rossi MP, Marchesi U. 2002. Genetic, phenotypic and pathogenic diversity of *Xanthomonas arboricola* pv. *corylina* strains question the representative nature of the type strain. Plant Pathol. 51:374–381.

Shapiro BJ, Friedman J, Cordero OX, Preheim SP, Timberlake SC, Szabó G, Polz MF, Alm EJ. 2012. Population Genomics of Early Events in the Ecological Differentiation of Bacteria. Science (80-.). 336:48–51.

Shapiro LR, Scully ED, Straub TJ, Park J, Stephenson AG, Beattie G a., Gleason ML, Kolter R, Coelho MC, De Moraes CM, et al. 2016. Horizontal Gene Acquisitions, Mobile Element Proliferation, and Genome Decay in the Host-Restricted

Plant Pathogen *Erwinia Tracheiphila*. Genome Biol. Evol. 8:649–664.

Shen YM, Huang TC, Chao CH, Liu HL. 2013. First Report of Bacterial Spot Caused by *Xanthomonas arboricola* pv. *pruni* on Japanese Plum in Taiwan. Plant Dis. 97:835–835.

Sheppard SK, Didelot X, Jolley K a., Darling AE, Pascoe B, Meric G, Kelly DJ, Cody A, Colles FM, Strachan NJC, et al. 2013. Progressive genome-wide introgression in agricultural *Campylobacter coli*. Mol. Ecol. 22:1051–1064.

Shimodaira H, Hasegawa M. 1999. Multiple Comparisons of Log-Likelihoods with Applications to Phylogenetic Inference. Mol. Biol. Evol. 16:1114–1116.

Siguier P, Gourbeyre E, Chandler M. 2014. Bacterial insertion sequences: Their genomic impact and diversity. FEMS Microbiol. Rev. 38:865–891.

Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M. 2006. ISfinder: the reference centre for bacterial insertion sequences. Nucleic Acids Res. 34:32–36.

da Silva a CR, Ferro J a, Reinach FC, Farah CS, Furlan LR, Quaggio RB, Monteiro-Vitorello CB, Van Sluys M a, Almeida NF, Alves LMC, et al. 2002. Comparison of the genomes of two *Xanthomonas* pathogens with differing host specificities. Nature 417:459–463.

Sinha D, Gupta MK, Patel HK, Ranjan A, Sonti R V. 2013. Cell Wall Degrading Enzyme Induced Rice Innate Immune Responses Are Suppressed by the Type 3 Secretion System Effectors XopN, XopQ, XopX and XopZ of *Xanthomonas oryzae pv. oryzae*. PLoS One 8:e75867.

Smith EF. 1903. Observations on a hitherto unreported bacterial disease, the cause of which enters the plant through ordinary stomata. Science. 17:456–457.

Sokurenko E V, Chesnokova V, Dykhuizen DE, Ofek I, Wu X, Krogfelt K, Struve C, Schembri M, Hasty D. 1998. Pathogenic adaptation of *Escherichia coli* by variation of FimH adhesin. Proc. Natl. Acad. Sci. 95:8922–8926.

Song C, Yang B. 2010. Mutagenesis of 18 type III effectors reveals virulence function of XopZ(PXO99) in *Xanthomonas oryzae pv. oryzae*. Mol. Plant. Microbe. Interact. 23:893–902.

Sory MP, Boland A, Lambermont I, Cornelis GR. 1995. Identification of the YopE and YopH domains required for secretion and internalization into the cytosol of macrophages, using the cyaA gene fusion approach. Proc. Natl. Acad. Sci. U. S. A. 92:11998–12002.

Spagnoletti M, Ceccarelli D, Rieux A, Fondi M, Taviani E, Fani R, Colombo MM, Colwell RR, Balloux F. 2014. Acquisition and evolution of SXT-R391 integrative conjugative elements in the seventh-pandemic *Vibrio cholerae* lineage. MBio 5:e01356–14.

Stanford AM, Harden R, Parks CR. 2000. Phylogeny and biogeography of *Juglans* (Juglandaceae) based on matK and ITS sequence data. Am. J. Bot. 87:872–882.

Stavrinides J, Ma W, Guttman DS. 2006. Terminal reassortment drives the quantum evolution of type III effectors in bacterial pathogens. PLoS Pathog. 2:e104.

Stavrinides J, McCann HC, Guttman DS. 2008. Host-pathogen interplay and the evolution of bacterial effectors. Cell. Microbiol. 10:285–292.

Stefani E. 2010. Economic significance and control of bacterial spot/canker of stone fruits caused by *Xanthomonas arboricola pv. Pruni*. J. Plant Pathol. 92:99–104.

Stewart FS. 1968. Bacteriology and immunology for students of medicine.

Studholme DJ, Wasukira A, Paszkiewicz K, Aritua V, Thwaites R, Smith J, Grant M. 2011. Draft genome sequences of *xanthomonas sacchari* and two banana-associated xanthomonads reveal insights into the *xanthomonas* group 1 clade. Genes (Basel). 2:1050–1065.

Stukenbrock EH. 2016. The Role of Hybridization in the Evolution and Emergence of New Fungal Plant Pathogens. Phytopathology 106:104–112.

Stukenbrock EH, McDonald BA. 2008. The origins of plant pathogens in agro-ecosystems. Annu. Rev. Phytopathol. 46:75–100.

Suerbaum S, Maynard-Smith J, Bapumia K, Morelli G, Smith NH, Kunstmann E, Dyrek I, Achtman M. 1998. Free recombination within *Helicobacter pylori*. Proc. Natl. Acad. Sci. 95:12619–12624.

Supply P, Marceau M, Mangenot S, Roche D, Rouanet C, Khanna V, Majlessi L, Criscuolo A, Tap J, Pawlik A, et al. 2013. Genomic analysis of smooth tubercle bacilli provides insights into ancestry and pathoadaptation of *Mycobacterium tuberculosis*. Nat. Genet. 45:172–179.

Tampakaki AP, Skandalis N, Gazi AD, Bastaki MN, Panagiotis F. S, Charova SN, Kokkinidis M, Panopoulos NJ. 2010. Playing the "Harp": Evolution of Our Understanding of hrp/hrc Genes 1. Annu. Rev. Phytopathol. 48:347–370.

Tegli S, Gori A, Cerboneschi M, Cipriani MG, Sisto A. 2011. Type three secretion system in *Pseudomonas savastanoi* pathovars: Does timing matter? Genes (Basel). 2:957–979.

Tenaillon O, Denamur E, Matic I. 2004. Evolutionary significance of stress-induced mutagenesis in bacteria. Trends Microbiol. 12:264–270.

Teper D, Burstein D, Salomon D, Gershovitz M, Pupko T, Sessa G. 2016. Identification of novel *Xanthomonas euvesicatoria* type III effector proteins by machine-learning approach. Mol. Plant Pathol. 17:395–411.

Theethakaew C, Feil EJ, Castillo-Ramírez S, Aanensen DM, Suthienkul O, Neil DM, Davies RL. 2013. Genetic relationships of *Vibrio parahaemolyticus* isolates from clinical, human carrier, and environmental sources in Thailand, determined by multilocus sequence analysis. Appl. Environ. Microbiol. 79:2358–2370.

Thomma BPHJ, Nürnberger T, Joosten MH a J. 2011. Of PAMPs and effectors: the blurred PTI-ETI dichotomy. Plant Cell 23:4–15.

Thung I, Aramin H, Vavinskaya V, Gupta S, Park JY, Crowe SE, Valasek MA. 2016. Review article: the global emergence of *Helicobacter pylori* antibiotic resistance. Aliment. Pharmacol. Ther. 46:514–533.

Tine M, Kuhl H, Gagnaire P-A, Louro B, Desmarais E, Martins RST,

Hecht J, Knaust F, Belkhir K, Klages S, et al. 2014. European sea bass genome and its variation provide insights into adaptation to euryhalinity and speciation. Nat. Commun. 5:5770.

Toft C, Andersson SGE. 2010. Evolutionary microbial genomics: insights into bacterial host adaptation. Nat. Rev. 11:465–475.

Toth IK, Pritchard L, Birch PRJ. 2006. Comparative genomics reveals what makes an enterobacterial plant pathogen. Annu. Rev. Phytopathol. 44:305–336.

Tourret J, Aloulou M, Garry L, Tenaillon O, Dion S, Ryffel B, Monteiro RC, Denamur E. 2011. The interaction between a non-pathogenic and a pathogenic strain synergistically enhances extra-intestinal virulence in *Escherichia coli*. Microbiology 157:774–785.

Triplett LR, Verdier V, Campillo T, Van Malderghem C, Cleenwerck I, Maes M, Deblais L, Corral R, Koita O, Cottyn B, et al. 2015. Characterization of a novel clade of *Xanthomonas* isolated from rice leaves in Mali and proposal of *Xanthomonas maliensis* sp. nov. Antonie Van Leeuwenhoek 107:869–881.

Troisfontaines P, Cornelis GR. 2005. Type III Secretion: More systems than you think. physiology 20:326–339.

Turner JW, Berthiaume CT, Morales R, Armbrust EV, Strom MS. 2016. Genomic evidence of adaptive evolution in emergent *Vibrio parahaemolyticus* ecotypes. Elem. Sci. Anthr. 4:000117.

Turner TR, James EK, Poole PS. 2013. The plant microbiome. Genome Biol. 14:209.

Vahdati K, Mohseni Pourtaklu S, Karimi R, Barzehkar R, Amiri R, Mozaffari M, Woeste K. 2014. Genetic diversity and gene flow of some Persian walnut populations in southeast of Iran revealed by SSR markers. Plant Syst. Evol. 301:691–699.

Vanderplank JE. 1963. Plant Diseases: Epidemics and Control.

Vandroemme J, Cottyn B, Pothier JF, Pflüger V, Duffy B, Maes M. 2013. *Xanthomonas arboricola pv. fragariae*: what's in a name? Plant Pathol. 62:1123–1131.

Vanneste JL, Yu J, Cornish D a, Tanner DJ, Windner R, Chapman JR, Taylor RK, Mackay J, Dowlut S. 2012. Identification, virulence and distribution of two biovars of *Pseudomonas syringae pv. actinidiae* in New Zealand. Plant Dis. 97:708–719.

Varani AM, Siguier P, Gourbeyre E, Charneau V, Chandler M. 2011. ISsaga is an ensemble of web-based methods for high throughput identification and semi-automatic annotation of insertion sequences in prokaryotic genomes. Genome Biol. 12:R30.

Vauterin L, Hoste B, Kersters K, Swings J. 1995. Reclassification of *Xanthomonas*. :472–489.

Vauterin L, Rademaker J, Swings J. 2000. Synopsis on the Taxonomy of the Genus *Xanthomonas*. Am. Phytopathol. Soc. 90:677–682.

Vauterin L, Yang P, Alvarez A, Takikawa Y, Roth D a., Vidaver AK, Stall RE, Kersters K, Swings J. 1996. Identification of non-pathogenic *Xanthomonas*

strains associated with plants. Syst. Appl. Microbiol. 19:96–105.

Verde I, Abbott AG, Scalabrin S, Jung S, Shu S, Marroni F, Zhebentyayeva T, Dettori MT, Grimwood J, Cattonaro F, et al. 2013. The high-quality draft genome of peach (*Prunus persica*) identifies unique patterns of genetic diversity, domestication and genome evolution. Nat. Genet. 45:487–494.

Vezzulli L, Pruzzo C, Huq A, Colwell RR. 2010. Environmental reservoirs of *Vibrio cholerae* and their role in cholera. Environ. Microbiol. Rep. 2:27–33.

Viana D, Comos M, McAdam PR, Ward MJ, Selva L, Guinane CM, González-Muñoz BM, Tristan A, Foster SJ, Fitzgerald JR, et al. 2015. A single natural nucleotide mutation alters bacterial pathogen host tropism. Nat. Genet. 47:361–366.

Victory ER, Glaubitz JC, Rhodes OE, Woeste KE. 2006. Genetic homogeneity in *Juglans nigra* (Juglandaceae) at nuclear microsatellites. Am. J. Bot. 93:118–126.

Vos M, Didelot X. 2009. A comparison of homologous recombination rates in bacteria and archaea. ISME J. 3:199–208.

Wadhams GH, Armitage JP. 2004. Making sense of it all: bacterial chemotaxis. Nat. Rev. Mol. Cell Biol. 5:1024–1037.

Wakker, JH. 1883. Vorläufige Mittheilungen über Hyacinthenkrankheiten. Bot. Cent.:315–317.

Walker TS, Walker TS, Bais HP, Bais HP, De E, De E, Schweizer HP, Schweizer HP, Rahme LG, Rahme LG, et al. 2004. *Pseudomonas aeruginosa* -Plant Root Interactions. Pathogenicity, Biofilm Formation, and Root Exudation. Plant Physiol. 134:320–331.

Warren JR, Marshall BJ. 1983. Unidentified curved bacilli on gastric epithelium in active chronic gastritis. Lancet 8336:1273–1275.

Weinert L a., Welch JJ, Suchard M a., Lemey P, Rambaut A, Fitzgerald JR. 2012. Molecular dating of human-to-bovid host jumps by *Staphylococcus aureus* reveals an association with the spread of domestication. Biol. Lett. 8:829–832.

Wellington EMH, Boxall AB a, Cross P, Feil EJ, Gaze WH, Hawkey PM, Johnson-Rollings AS, Jones DL, Lee NM, Otten W, et al. 2013. The role of the natural environment in the emergence of antibiotic resistance in Gram-negative bacteria. Lancet Infect. Dis. 13:155–165.

White FF, Potnis N, Jones JB, Koebnik R. 2009. The type III effectors of *Xanthomonas*. Mol. Plant Pathol. 10:749–766.

White RG, Timbury MC. 1973. Parasitism: the mechanisms of pathogenicity.

Wichmann F, Vorhölter FJ, Hersemann L, Widmer F, Blom J, Niehaus K, Reinhard S, Conradin C, Kölliker R. 2013. The noncanonical type III secretion system of *Xanthomonas translucens pv. graminis* is essential for forage grass infection. Mol. Plant Pathol. 14:576–588.

Wichmann G, Bergelson J. 2004. Effector Genes of *Xanthamonas axonopodis pv. vesicatoria* Promote Transmission and Enhance Other Fitness Traits in the Field.

Genetics 166:693–706.

Wirth T, Falush D, Lan R, Colles F, Mensa P, Wieler LH, Karch H, Reeves PR, Maiden MCJ, Ochman H, et al. 2006. Sex and virulence in *Escherichia coli*: An evolutionary perspective. Mol. Microbiol. 60:1136–1151.

Wittstock U, Gershenzon J. 2002. Constitutive plant toxins and their role in defense against herbivores and pathogens. Curr. Opin. Plant Biol. 5:300–307.

Woeste K, Michler C. 2011. Juglans. In: Wild Crop Relatives: Genomic and Breeding Resources. p. 77–88.

Woodford N, Ellington MJ. 2007. The emergence of antibiotic resistance by mutation. Clin. Microbiol. Infect. 13:5–18.

Woolhouse MEJ. 2002. Population biology of emerging and re-emerging pathogens. Trends Microbiol. 10:3–7.

Wuicher K, Zhulin IB. 2003. Molecular evolution of sensory domains in cyanobacterial chemoreceptors. Trends Microbiol. 11:200–203.

Wuichet K, Alexander R, Zhulin I. 2007. Comparative genomic and protein sequence analyses of a complex system controlling bacterial chemotaxis. Methods Enzym. 422:1–31.

Xin X-F, He SY. 2013. Pseudomonas syringae pv. tomato DC3000: A Model Pathogen for Probing Disease Susceptibility and Hormone Signaling in Plants. Annu. Rev. Phytopathol. 51:473–498.

Yahara K, Didelot X, Jolley K a., Kobayashi I, Maiden MCJ, Sheppard SK, Falush D. 2016. The landscape of realized homologous recombination in pathogenic bacteria. Mol. Biol. Evol. 33:456–471.

Yamaoka Y, Debakey ME, Affairs V. 2009. *Helicobacter pylori* typing as a tool for tracking human migration. Clin. Microbiol. Infect. 15:829–834.

Yan Y, Cui Y, Han H, Xiao X, Wong HC, Tan Y, Guo Z, Liu X, Yang R, Zhou D. 2011. Extended MLST-based population genetics and phylogeny of *Vibrio parahaemolyticus* with high levels of recombination. Int. J. Food Microbiol. 145:106–112.

Yao J, Allen C. 2006. Chemotaxis Is Required for Virulence and Competitive Fitness of the Bacterial Wilt Pathogen *Ralstonia solanacearum* Chemotaxis Is Required for Virulence and Competitive Fitness of the Bacterial Wilt Pathogen *Ralstonia solanacearum*. 188:3697–3708.

Yirgou D. 1964. *Xanthomonas guizotiae* sp. nov. on *Guizotia abyssinica*. Phytopathology 54:1490–1491.

Young GM, Schmiel DH, Miller VL. 1999. A new pathway for the secretion of virulence factors by bacteria: the flagellar export apparatus functions as a protein-secretion system. Proc. Natl. Acad. Sci. U. S. A. 96:6456–6461.

Young JM, Dye DW, Bradbury JF, Panagopoulos CG, Robbs CF. 1978. A proposed nomenclature and classification for plant pathogenic bacteria. New Zeal. J. Agric. Res. 21:153–177.

Young JM, Park D-C, Shearman HM, Fargier E. 2008. A multilocus sequence

analysis of the genus *Xanthomonas*. Syst. Appl. Microbiol. 31:366–77.

Young JM, Wilkie JP, Park D-C, Watson DRW. 2010. New Zealand strains of plant pathogenic bacteria classified by multi-locus sequence analysis; proposal of *Xanthomonas dyei* sp. nov. Plant Pathol. 59:270–281.

Załuga J, Stragier P, Baeyen S, Haegeman A, Van Vaerenbergh J, Maes M, De Vos P. 2014. Comparative genome analysis of pathogenic and non-pathogenic *Clavibacter* strains reveals adaptations to their lifestyle. BMC Genomics 15:392.

Zaurov DE, Molnar TJ, Eisenman SW, Ford TM, Mavlyanova RF, Capik JM, Funk CR, Goffreda JC. 2013. Genetic Resources of Apricots (*Prunus armeniaca* L.) in Central Asia. HortScience 48:681–691.

Zerbino DR, Birney E. 2008. Velvet Algorithms for de novo short read assembly using de Bruijn graphs. Genome Res.

Zhao Y, Sundin GWSW, Wang D. 2009. Construction and analysis of pathogenicity island deletion mutants of *Erwinia amylovora*. Can. Jouornal Microbiol. 55:457–464.

Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. 2011. PHAST: A Fast Phage Search Tool. Nucleic Acids Res. 39:1–6.

Zipfel C. 2008. Pattern-recognition receptors in plant innate immunity. Curr. Opin. Immunol. 20:10–16.

UNIVERSITE BRETAGNE LOIRE

INRA SCIENCE & IMPACT IRHS Institut de Recherche en Horticulture et Semences

université angers

Thèse de Doctorat

Déborah MERDA

Histoire évolutive de *Xanthomonas arboricola*, espèce bactérienne composée de souches pathogènes et commensales

Evolutionary history of *Xanthomonas arboricola*, bacterial species composed of pathogenic and commensal strains

Résumé

Comprendre l'émergence des maladies dans les agroécosystèmes nécessite d'étudier l'histoire évolutive des populations bactériennes associées aux plantes. L'objectif de ce travail était de déterminer les évènements évolutifs conduisant à l'émergence des lignées pathogènes ou pathovars dans l'espèce *Xanthomonas arboricola*. Une analyse de génétique des populations a été menée sur un panel de souches phytopathogènes et commensales et complétée par l'inférence des gains et pertes de facteurs de virulence. Cette espèce possède une structure de population épidémique ; les clones épidémiques ont émergé suite à l'acquisition de facteurs de virulence à partir d'un fond recombinant de souches commensales. Une analyse de génomique des populations et la reconstruction de scénarios de divergence entre ces clones et le réseau de souches recombinantes, a montré la persistance d'un flux de gènes asymétrique entre ces deux groupes, dans le sens souches pathogènes vers souches commensales. Enfin, l'histoire évolutive du principal facteur de virulence des *Xanthomonas*, le système de sécrétion de type 3, a été retracée au sein du genre, et a montré que celui-ci avait été acquis ancestralement puis perdu dans certaines souches commensales. En conclusion, l'ancêtre commun de *X. arboricola* possédait des facteurs de virulence et au sein des souches commensales, certaines ont perdu ces facteurs, tandis que d'autres ont conservé le répertoire ancestral. Ces dernières diffèrent peu de certains agents pathogènes, et pourraient représenter un risque pour de nouvelles émergences. Des travaux de génomique fonctionnelle permettraient de valider ces hypothèses.

Mots clés

Emergence, *Xanthomonas*, génomiques de populations, recombinaison, système de sécrétion de type 3, environnements génomiques, bactérie, phytopathogène

Abstract

Deciphering the evolutionary history of bacterial populations associated to plants is necessary to understand disease emergence in agroecosystems. The aim of this study is to unveil the evolutionary events responsible for pathogenic lineages or pathovar emergences in *Xanthomonas arboricola*. This species is composed of both plant pathogenic and commensal strains Population genetics analyses and gain and loss inferences of virulence factors showed that *X. arboricola* exhibits an epidemic population structure, within which epidemic clones emerged from a recombinogenic background population following virulence factor acquisition. Population genomics and inference of divergence scenarii between epidemic clones and the network of recombinant strains showed persistence of homologous recombination along divergence of these two groups, with an asymmetric gene flux from pathogenic strains to commensal ones. Finally, evolutionary history of the type three secretion system (T3SS), the main virulence factor in *Xanthomonas* genus, was studied at genus scale and showed that T3SS was ancestrally acquired and lost in commensal strains. Altogether these analyses allowed us to show that the common ancestor of *X.arboricola* had virulence factors, and that within commensal strains, some lost these virulence factors whereas others kept the ancestral repertoire. These latter strains have a similar repertoire to that of some pathogenic strains, and could represent a risk for new disease emergence. Functional genomics could allow us to validate these hypotheses.

Key Words

Emergence, *Xanthomonas*, population genomics, recombination, Type Three secretion system, genomic environments, bacteria, phytopathogenic