



HAL
open science

Origines et évolution des voies de synthèse des phospholipides dans les trois domaines du vivant. Implications pour la nature des membranes du cenancêtre

Jonathan Lombard

► **To cite this version:**

Jonathan Lombard. Origines et évolution des voies de synthèse des phospholipides dans les trois domaines du vivant. Implications pour la nature des membranes du cenancêtre. Sciences agricoles. Université Paris Sud - Paris XI, 2012. Français. NNT : 2012PA112369 . tel-00819686

HAL Id: tel-00819686

<https://theses.hal.science/tel-00819686v1>

Submitted on 2 May 2013

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.

THESE DE DOCTORAT DE L'UNIVERSITE PARIS SUD

Ecole doctorale Gènes, Génomes, Cellules (ED 426)

Présentée par :

Jonathan LOMBARD

Pour l'obtention du grade de :

Docteur en Sciences de l'Université Paris Sud

**ORIGINES ET EVOLUTION DES VOIES DE SYNTHÈSE DES
PHOSPHOLIPIDES DANS LES TROIS DOMAINES DU VIVANT.
IMPLICATIONS POUR LA NATURE DES MEMBRANES DU
CENANCETRE**

Thèse soutenue le 17 décembre 2012 devant le jury composé de :

Vincent DAUBIN	Rapporteur
Simonetta GRIBALDO	Rapporteur
Pierre CAPY	Examineur
Olivier LESPINET	Examineur
Michel MORANGE	Examineur
David MOREIRA	Directeur de thèse

RÉSUMÉ

Les bases fondamentales de la biologie suggèrent que tous les organismes actuels partagent un dernier ancêtre commun, le cenancêtre. Dès que la comparaison moléculaire des organismes des trois domaines du vivant (archées, bactéries et eucaryotes) est devenue possible, d'importants débats ont émergé sur l'habitat du cenancêtre, son rapprochement des origines de la vie, sa nature unique ou communautaire et ses relations avec les trois domaines du vivant. Cependant, jusqu'à il y a peu les informations disponibles sur les organismes modernes n'étaient pas suffisantes pour décrire précisément sa biologie.

Notamment, la découverte chez les archées de membranes dont les composants principaux, les phospholipides, sont synthétisés par des mécanismes très différents de ceux des bactéries et les eucaryotes a conduit à proposer que chaque mécanisme de synthèse des phospholipides soit apparu indépendamment dans les lignées modernes. Dans ces hypothèses le cenancêtre aurait été dépourvu de phospholipides et, donc, de membranes. Cela met en cause la nature cellulaire du cenancêtre, qui semblait pourtant soutenue par d'autres indices indirects. Ces contradictions posent la question de l'existence de traces dans les organismes modernes d'une synthèse des phospholipides chez le cenancêtre. Dans cette thèse j'ai profité de l'explosion récente des données génomiques pour répondre à cette question.

Il avait déjà été montré que des membres de deux superfamilles protéiques universelles pouvaient avoir synthétisé de façon non spécifique chez le cenancêtre les énantiomères de glycérol phosphate servant d'ossature aux phospholipides. Les phospholipides archéens sont composés d'isoprénoïdes et les bactériens et eucaryotes d'acides gras. J'ai donc étudié l'évolution des voies de synthèse de ces molécules ainsi que celle de l'assemblage de tous les composants dans des phospholipides. Mes résultats montrent que la voie de synthèse des isoprénoïdes des eucaryotes et une voie hypothétique de synthèse des acides gras chez les archées avaient probablement des ancêtres moins spécifiques chez le cenancêtre. Une partie au moins de la machinerie d'assemblage des phospholipides semble aussi avoir été présente chez le cenancêtre.

Ceci suggère que le cenancêtre avait probablement des mécanismes peu spécifiques de synthèse des phospholipides et que les différences entre les membranes actuelles sont dues à la spécialisation de la machinerie ancestrale dans chaque lignée. Mes observations soulignent aussi l'importance d'étudier le cenancêtre à partir des informations issues des organismes actuels pour éviter toute confusion avec les origines de la vie.

ABSTRACT

The main bases of Biology suggest that all extant organisms share a last common ancestor, namely the cenancestor. As soon as the comparison of molecular characters of organisms representative of the whole diversity of life became possible, hot debates emerged about the environmental conditions in which the cenancestor lived, its closeness to the origins of life, its single or community nature and its relationships with the three domains of life (Archaea, Bacteria and Eucarya). However, available information about current organisms was for a long time inadequate to precisely describe the biology of this organism.

For instance, the observation that the main archaeal membrane components, called phospholipids, are synthesized by different means than their bacterial/eukaryotic counterparts was proposed to reveal that modern phospholipid biosynthesis pathways emerged late in independent lineages and were, therefore, absent in the cenancestor. This hypothesis argued that the cenancestor had no lipid membranes, so it could not be a cellular organism although other indirect clues indicated the opposite. These contradictions raise the question of the presence in modern organisms of traces that the cenancestor had a phospholipid biosynthesis machinery.

In this dissertation, I took advantage from the recent accumulation of genomic data to address this issue. Previous work had shown that the members of two universal protein superfamilies could be present in the cenancestor to carry out the non-specific synthesis of the glycerol phosphate enantiomers that are the backbones of modern phospholipids. Bacterial and eukaryotic phospholipids use fatty acids whereas archaeal phospholipids are made up of isoprenoids. Thus, I studied the evolution of the metabolic pathways that synthesize these molecules and build up the phospholipids from their components. My results show that the eukaryotic isoprenoid biosynthesis pathway and a hypothetical archaeal fatty acid biosynthesis pathway are likely to have had less specific ancestors in the cenancestor. In addition, the phospholipid assembly machinery was also probably present in the cenancestor.

These results suggest that the cenancestor was likely able to enzymatically synthesize its phospholipids by means less specific than modern ones. Dissimilarities in modern membrane phospholipids would result from the specialization of each biosynthesis system in each lineage. My work also stresses the fact that the cenancestor should be described on the basis of the comparison of modern organisms to avoid frequent confusions between the cenancestor and the origins of life.

REMERCIEMENTS

Je me souviens d'un après-midi de l'été 2005 au cours duquel Kris et moi nous baladions le long de notre parcours habituel entre l'avenue de Bayonne et le parc de la Vaguada, à Pampelune. Je rentrais en Espagne après ma première année universitaire en France et je lui expliquais le système académique Licence-Master-Doctorat qui était entré en vigueur l'année précédente. Je me souviens qu'il m'a dit « alors, dans 7 ans, tu auras 26 ans et tu pourrais déjà être docteur ? ». Ensuite la conversation a tourné autour de nos rêveries sur ce terme très suggestif mais qui en réalité ne voulait pas dire grande chose pour nous, ainsi que sur les sujets de recherche qui m'intéressaient et parmi lesquels l'évolution ancienne était ma principale aspiration.

Nous voici sept ans plus tard, je n'ai pas vu le temps passer et j'ai du mal à réaliser que cet objectif très vague et rêveur soit devenu une réalité. Je voudrais remercier ici tous les amis, enseignants et membres de ma famille qui m'ont permis de construire ce rêve, qui m'ont encouragé à le poursuivre et/ou aidé à le réaliser. Malheureusement, je pense que si je faisais la liste de ces personnes et des raisons pour lesquelles je leur suis reconnaissant, ces remerciements seraient aussi longs que les crédits de fin d'un film hollywoodien. Je vais donc me limiter à citer juste quelques noms importants en m'excusant de ne pas rentrer plus dans les détails et d'oublier certainement des gens. Afri, Sophie, Nacho, Kris, Gloria, Rubén, Marina Q, Laura EI, Esther, Cristina, Mariaje, Anne-Marie, Alain, Anaïs, Céline P, Julie, Doris, Valérie, Dominique CB, Céline B, Michel T, Jean, David N et Estelle, je vous remercie de m'avoir éveillé au monde, de vos conseils et de vos encouragements qui m'ont conduit à tenter cette thèse.

Je dois également remercier toutes les personnes qui ont permis à cette thèse d'avoir lieu et qui l'ont vécu « en direct ». Tout d'abord, David M, qui a été un excellent directeur de thèse, me donnant toujours de la liberté tout en étant disponible et à l'écoute. Merci beaucoup aussi à Puri de m'avoir donné l'opportunité de réaliser ce travail et d'être toujours une source de nouvelles idées à explorer. Merci à l'ensemble de l'équipe microbio (David, Puri, Estelle, Charles, Philippe, Céline P, Marianne, Marie, Ludwig, Paola) de leur bonne humeur et d'avoir été capables de créer une atmosphère très agréable et conviviale dans laquelle il est un plaisir de travailler. Merci aussi aux nombreuses autres personnes de l'ESE qui participent à cette ambiance et qui permettent d'apprendre des choses nouvelles tous les jours, à savoir Jacqui,

Gwendal, Julien, Boris, Hervé, Amandine, Jean-Tristan, Marta, Juan, Yann, Alexis, Claire, Lucie, Alodie, Damien H, Alex, Delphine R, Romain, Jane, Nathalie FL, Marina P et Juerguen. Merci à l'école doctorale GGC de m'avoir attribué la bourse qui a permis à ce travail d'avoir lieu et d'être toujours disponible pour les doctorants.

Merci à tous mes amis et à ma famille pour m'avoir supporté pendant ces trois dernières années, car on sait qu'il est parfois difficile de vivre le comportement étrange du thésard commun. Cela vaut spécialement pour Céline P, qui m'a supporté tous les jours à la fois à la maison, au travail et durant nos temps libres. Je dois un remerciement supplémentaire à David N pour m'avoir encouragé dès le départ dans les voies parfois étranges que cette thèse m'a permis d'explorer tout en me faisant découvrir un monde toujours plus large que, j'espère, nous aurons toute une vie pour parcourir.

Enfin, je remercie les personnes qui m'ont aidé dans la rédaction de ce manuscrit. Merci à Laura E et à quelques bibliothécaires anonymes de m'avoir aidé à trouver des références pas toujours évidentes à obtenir, ainsi qu'à David M, Anne, David N, Céline, Marianne et Philippe pour leurs commentaires et corrections qui m'ont permis de traduire le texte vers un français que j'espère compréhensible. Et merci encore à Céline P et David N de l'aide de dernière minute pour la mise en page. Enfin, merci à Juli, Simonetta et Olivier de leurs conseils lors de mon comité de thèse et à l'ensemble de jury d'avoir accepté d'évaluer mon travail.

RÉSUMÉ	1
ABSTRACT	2
PRÉFACE	9
INTRODUCTION	10
Première partie : Le cenancêtre	10
1. L'unicité du vivant	11
<i>Tous les organismes partagent un ancêtre commun</i>	11
<i>Le dernier ancêtre commun de tous les organismes : définitions</i>	14
<i>A la rencontre du cenancêtre</i>	16
2. Woese : le bouleversement de l'histoire du vivant	18
<i>Les trois « royaumes primaires »</i>	18
<i>Relations entre les trois « royaumes primaires »</i>	22
<i>Une vision particulière du cenancêtre : le progénote</i>	24
3. Débats sur le cenancêtre	27
<i>Le cenancêtre : ni ancêtre totipotent, ni communauté</i>	27
<i>Interprétations de l'arbre : cenancêtre et température</i>	32
4. Vacillations de l'arbre du vivant	38
<i>La racine</i>	38
<i>Arbre ou forêt du vivant ?</i>	43
<i>Origine des eucaryotes et relation avec le cenancêtre</i>	49
5. Portrait robot du cenancêtre	52
<i>A quoi ressemblait le cenancêtre</i>	52
<i>Génome à ADN ou à ARN</i>	54
<i>Le cenancêtre était-il entouré par des membranes lipidiques ?</i>	57
Deuxième partie : les membranes lipidiques	58
1. Importance physiologique et structure des membranes cellulaires	59
2. Une histoire des membranes plasmiques.	61
<i>De l'existence des membranes plasmiques</i>	61
<i>Les premiers modèles de structure membranaire</i>	67
<i>La crise d'un paradigme</i>	70
<i>Naissance et vie de la mosaïque fluide</i>	71
<i>Perspectives de l'analyse historique des membranes cellulaires pour celle des membranes anciennes</i>	73
3. Quand la diversité membranaire actuelle met en cause l'existence de membranes chez le cenancêtre	74
<i>Des phospholipides exotiques chez les archées</i>	75
<i>Un cenancêtre acellulaire</i>	77
<i>Une compartimentation minérale</i>	78
<i>Bilan des hypothèses acellulaires</i>	82

4. Avant le cenancêtre : le principe de continuité des membranes	83
<i>Premiers travaux sur les compartiments primordiaux</i>	83
<i>Des molécules amphiphiles dans la Terre primitive : synthèse endogène</i>	84
<i>Des molécules amphiphiles dans la Terre primitive : synthèse exogène</i>	86
<i>Auto-assemblage spontané de vésicules dans l'environnement prébiotique</i>	90
<i>Croissance et division de vésicules prébiotiques</i>	93
<i>De la vésicule à la protocellule</i>	95
<i>Propriétés des membranes protocellulaires</i>	96
<i>Membranes prébiotiques et principe de continuité</i>	99
5. Des protéines membranaires chez le cenancêtre	100
<i>Fonctions énergétiques de la membrane</i>	100
<i>Adressage, insertion et translocation de protéines</i>	103
<i>Protéines et fonctions membranaires chez le cenancêtre</i>	105
6. Quelle composition possible pour des membranes lipidiques chez le cenancêtre ?	105
<i>Des membranes de type bactérien chez le cenancêtre</i>	106
<i>Des membranes hétérochirales chez le cenancêtre</i>	108
<i>Synthèse enzymatique de glycérol phosphate chez le cenancêtre</i>	111
<i>Synthèse enzymatique des autres composants phospholipidiques chez le cenancêtre</i>	114
OBJECTIFS	115
MATÉRIEL ET MÉTHODES	116
1. <i>Vue d'ensemble de la stratégie utilisée</i>	116
2. <i>Construction des jeux de données</i>	119
3. <i>Analyses phylogénétiques</i>	121
RÉSULTATS ET DISCUSSION	123
Première partie : Synthèse d'isoprénoïdes dans les trois domaines du vivant.	123
1. <i>Contexte : Voies de synthèse des isoprénoïdes</i>	123
2. <i>Distribution des enzymes des voies du MVA et du MEP chez les trois domaines du vivant</i>	125
<i>La voie du mévalonate chez les eucaryotes</i>	125
<i>La voie du mévalonate chez les archées</i>	126
<i>Les voies du MVA et du MEP chez les bactéries</i>	127
<i>La voie du MEP chez les archées et les eucaryotes</i>	129
3. <i>Evolution des enzymes de la voie du mévalonate</i>	129
<i>Enzymes communes aux deux voies</i>	130
<i>Enzymes de la voie eucaryote</i>	135
<i>Enzymes de la voie archéenne</i>	136
4. <i>Evolution des enzymes de la voie du méthylerythritol phosphate</i>	138
<i>Origine des gènes de la voie du MEP chez les eucaryotes</i>	139
<i>Origine de la voie du MEP</i>	142

5. Evolution de la synthèse d'isoprénoïdes à partir de leurs précurseurs	143
<i>L'évolution des isoprényl diphosphate synthases dans la littérature</i>	143
<i>Mise à jour de la phylogénie des isoprényl diphosphate synthases</i>	145
6. Conclusions	147
<i>Une voie de synthèse des précurseurs caractéristique pour chaque domaine du vivant</i>	147
<i>Possibilité d'une voie de synthèse des précurseurs des isoprénoïdes chez le cenancêtre</i>	148
<i>Des isoprényl diphosphate synthases chez le cenancêtre</i>	149
Deuxième partie : Synthèse d'acides gras chez les archées et implications pour le cenancêtre.	150
1. Contexte	150
<i>Des acides gras chez les archées</i>	150
<i>Les acides gras, représentants élémentaires d'une famille moléculaire très riche</i>	151
<i>Mécanismes de synthèse des acides gras et des polycétides secondaires</i>	153
<i>Evolution des machineries de synthèse des polycétides</i>	155
2. Evolution des gènes de synthèse des acides gras FAS-II chez les archées	159
<i>Etapas préalables à la synthèse des acides gras par la FAS-II bactérienne</i>	159
<i>La superfamille des thiolases</i>	162
<i>Les β-cétoacyl reductases</i>	166
<i>Les hydroxyacyl déshydratases</i>	169
<i>Les enoyl reductases</i>	173
3. Une synthèse d'acides gras indépendante de l'ACP ?	175
<i>Synthèse des acides gras chez les archées</i>	175
<i>Implications pour le cenancêtre</i>	176
Troisième partie : Synthèse des phospholipides dans les trois domaines du vivant	178
1. Contexte	178
<i>Liaisons ester, liaisons éther</i>	178
<i>Têtes polaires</i>	180
<i>Modifications des chaînes hydrophobes</i>	181
2. Evolution des enzymes de synthèse des phospholipides	182
<i>Les acyltransférases</i>	182
<i>Les prenyltransférases</i>	183
<i>La CDP-diacylglycérol synthase</i>	187
<i>Les CDP-alcool phosphatidyltransférases</i>	188
<i>Modifications des chaînes hydrophobes.</i>	189
3. Synthèse enzymatique de phospholipides chez le cenancêtre	190
<i>Enzymes de la synthèse des phospholipides présentes chez le cenancêtre</i>	190
<i>Implications sur la capacité du cenancêtre à fabriquer des membranes lipidiques</i>	191

CONCLUSION ET PERSPECTIVES	193
1. Sur les descriptions générales du cenancêtre	193
<i>L'Echelle des Etres et la génération spontanée</i>	195
<i>Biais expérimental</i>	198
<i>Définition du « complexe »</i>	201
<i>Tabula rasa</i>	203
2. Le statut des membranes chez le cenancêtre et l'évolution dans les trois domaines du vivant	205
<i>Le cenancêtre avait un mécanisme enzymatique de synthèse des phospholipides</i>	206
<i>Le bricolage de l'évolution</i>	207
<i>Quel type de membrane dans le cenancêtre ?</i>	210
<i>Implications des membranes mixtes du cenancêtre pour les trois domaines</i>	211
<i>Perspectives</i>	213
<i>Le cenancêtre, un organisme ancien</i>	215
BIBLIOGRAPHIE	217
TABLE DES ANNEXES	258

PRÉFACE

Dans la biologie actuelle, on a l'habitude de vivre dans le temps de l'article scientifique. Un article a normalement le mérite de délimiter clairement une question, de proposer une méthode pour y répondre, de l'appliquer, de présenter les résultats et de les discuter. C'est très efficace et, le plus souvent, assez court. Il existe des travaux de synthèse, mais ils ne peuvent pas servir à la fois pour résumer un contexte et expliquer un positionnement, et la plupart des livres qui s'écrivent sont soit des compilations d'articles de différents auteurs, soit de la vulgarisation des sciences.

Mon travail de doctorat a comporté une composante bibliographique très importante, qui m'a permis de mieux comprendre à la fois la biologie des organismes, l'histoire d'un certain nombre de découvertes et leurs implications. Ainsi, j'ai eu envie d'expliquer ce contexte au sens large, mon point de vue et comment l'ensemble de mon travail s'y replace. J'ai décidé de présenter ce travail dans mon manuscrit de thèse car il s'agit du format le plus flexible et le plus adapté. Le résultat est un peu inhabituel, notamment à cause de la structure du texte. L'introduction constitue la partie la plus longue car j'y ai développé des aspects historiques d'une façon assez détaillée. Cependant, j'ai essayé de donner des titres explicites aux différentes sections pour que le lecteur averti puisse se diriger directement vers les parties du manuscrit qui l'intéressent le plus.

La question principale de ce travail est de savoir si le dernier ancêtre commun de tous les organismes (le cenancêtre) était entouré d'une membrane lipidique, comme les organismes actuels. En effet, notre connaissance de cet organisme est encore très réduite et la conception que l'on peut avoir de sa membrane change radicalement notre façon d'envisager sa biologie. Ainsi, j'ai exposé dans la première partie de l'introduction les principaux débats qui existent autour du cenancêtre et de l'arbre de la vie. Dans la deuxième partie j'explique pourquoi la membrane est essentielle à la compréhension des cellules actuelles et comment les doutes sur la présence de membranes chez le cenancêtre affectent notre vision de cet organisme. A la fin de l'introduction il en ressort que le cenancêtre devait avoir une membrane, mais que son mécanisme de synthèse reste inconnu. C'est donc à la question de l'origine et l'évolution des voies de synthèse des lipides membranaires que mon travail scientifique a essayé de donner une réponse.

INTRODUCTION

Première partie : Le cenancêtre

“It may be asked how far I extend the doctrine of the modification of species [...]. All the members of whole classes can be connected together by chains of affinities, and all can be classified on the same principle, in groups subordinate to groups [...]. Therefore I cannot doubt that the theory of descent with modification embraces all the members of the same class. I believe that animals have descended from at most only four or five progenitors, and plants from an equal or lesser number.

Analogy would lead me one step further, namely, to the belief that all animals and plants have descended from some one prototype. But analogy may be a deceitful guide. Nevertheless all living things have much in common, in their chemical composition, their germinal vesicles, their cellular structure, and their laws of growth and reproduction [...]. Therefore I should infer from analogy that probably all the organic beings which have ever lived on this earth have descended from some one primordial form, into which life was first breathed.¹”

Charles Darwin, *On the Origins of Species*, 1859

¹ « Jusqu’où, pourra-t-on me demander, poussez-vous votre doctrine de la modification des espèces ? [...] Tous les membres de classes entières sont reliés les uns aux autres par une chaîne d’affinités, et peuvent tous, d’après un même principe, être classés en groupes subordonnés à d’autres groupes [...]. Je ne puis donc douter que la théorie de la descendance avec modifications ne doive comprendre tous les membres d’une même classe ou d’un même règne. Je crois que tous les animaux descendent de quatre ou cinq formes primitives tout au plus, et toutes les plantes d’un nombre égal ou même moindre.

L’analogie me conduirait à faire un pas de plus, et je serais disposé à croire que tous les animaux et toutes les plantes descendent d’un prototype unique ; mais l’analogie peut-être un guide trompeur. Toutefois, toutes les formes de la vie ont beaucoup de caractères communs : la composition chimique, la structure cellulaire, les lois de croissance [...]. L’analogie me pousse donc à penser que tous les êtres organisés qui ont vécu sur la terre descendent probablement d’une même forme primordiale dans laquelle la vie a été insufflée à l’origine » [Darwin 1859].

1. L'unicité du vivant

Tous les organismes partagent un ancêtre commun

La biologie actuelle est construite sur la base de quatre piliers fondamentaux. Chronologiquement, le premier est la théorie cellulaire qui postule que tous les organismes vivants sont constitués d'unités structurales et fonctionnelles que l'on appelle des cellules, qui apparaissent par division de cellules antérieures. Le deuxième est la théorie darwinienne de l'évolution qui explique l'apparition de la diversité biologique comme le résultat de l'accumulation, au cours de nombreuses générations, de modifications ayant privilégié d'une manière ou d'une autre la propagation des organismes porteurs de ces caractéristiques par rapport à d'autres organismes. Le troisième pilier est la génétique, qui étudie le fait que les organismes vivants engendrent des organismes qui leur ressemblent. Enfin, la biologie moléculaire et la biochimie ont permis de découvrir les molécules qui interviennent dans le fonctionnement des cellules et celles qui codent leur information génétique.

Réunis, ces quatre piliers convergent vers l'idée d'unicité du vivant. D'abord, comme l'indique l'épigraphe, l'une des conséquences de l'évolution darwinienne est que les similarités partagées entre organismes d'une même espèce, d'un même genre, d'une même famille, etc. sont considérées comme le résultat d'une parenté commune. On peut donc remonter de proche en proche à des groupes de plus en plus larges d'organismes possédant des caractéristiques communes jusqu'à arriver potentiellement à un groupe rassemblant l'ensemble du vivant, et conclure que ces groupes ont évolué à partir d'un ancêtre commun. L'hérédité joue évidemment un rôle très important dans l'évolution biologique car celle-ci dépend de l'existence d'un mécanisme de transmission des caractères à la descendance. L'hypothèse darwinienne d'une ascendance commune coïncide avec l'affirmation de la théorie cellulaire selon laquelle toute cellule se forme à partir d'une cellule précédente, *Omnis cellula e cellula*.

Bien que l'ascendance unique soit potentiellement incluse dans ces hypothèses, elle reste subordonnée à la démonstration qu'il existe des caractéristiques communes à tous les organismes. Au contraire, si l'on démontrait qu'il existe des organismes ne partageant aucune caractéristique avec les autres organismes, il faudrait conclure à l'existence de plusieurs ancêtres non liés entre eux. Pendant le XIX^{ème} siècle et la première moitié du XX^{ème} la biochimie semblait indiquer que tous les organismes étaient composés des mêmes molécules [Kluyver 1959], mais l'étude comparative restait irréalisable dans la pratique. Il a fallu

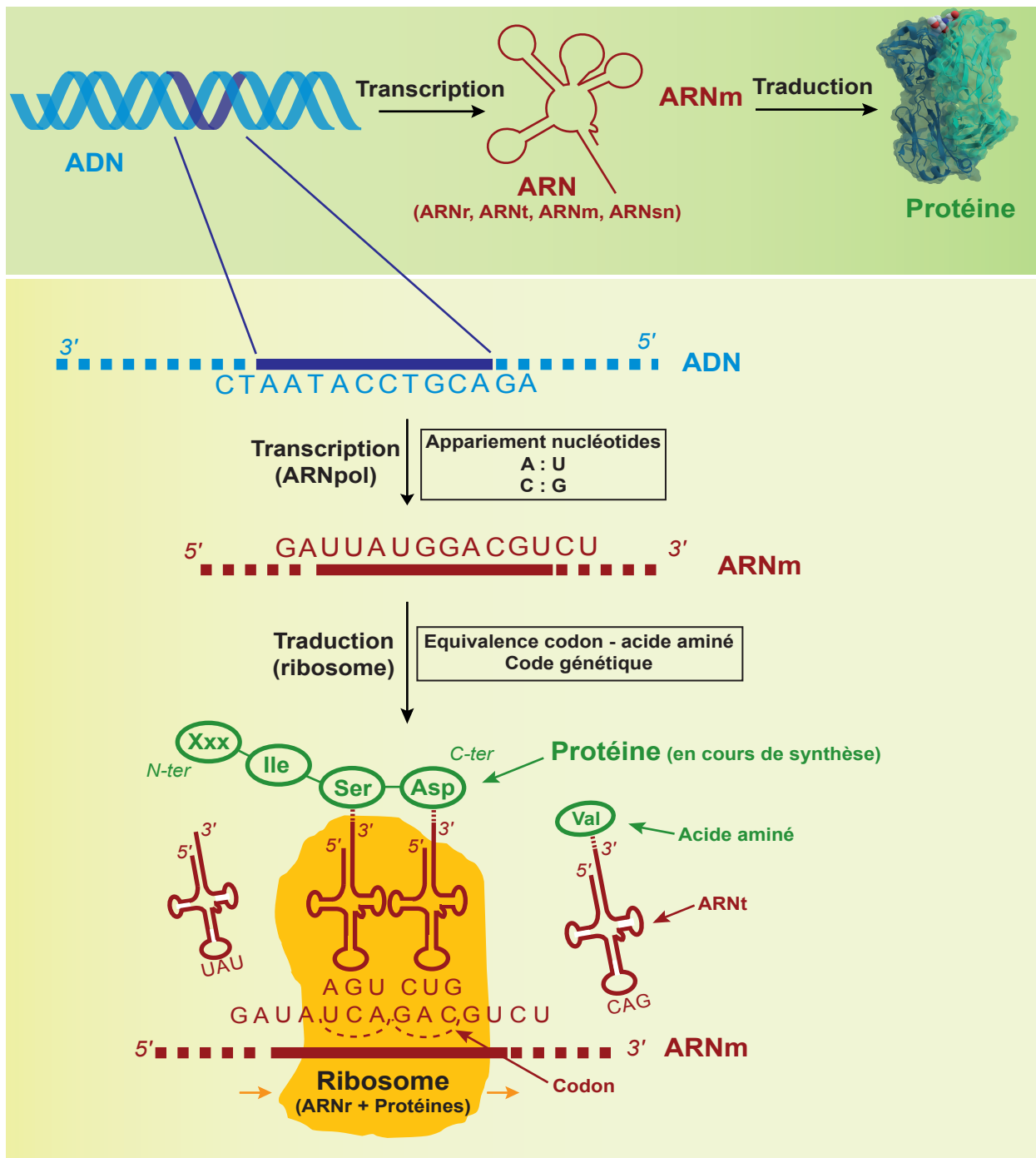


Figure 1. Système informationnel universel.

La partie supérieure présente les trois types principaux de macromolécules du vivant et les relations entre elles : 1) la transcription permet de synthétiser l'ARN à partir de l'ADN ; 2) les ARN peuvent avoir différentes fonctions, seuls les ARNm seront traduits en protéines. La partie inférieure schématise uniquement le mécanisme qui permet la synthèse d'une protéine à partir d'un gène donné.

atteindre le développement de la biologie moléculaire pour accéder plus facilement à une information comparable dans tous les organismes.

La biologie moléculaire a révélé que l'ADN était le support de l'information génétique. Pour servir dans la cellule, cette information doit être *transcrite* en une molécule complémentaire, l'ARN, qui est responsable de la *traduction*, i. e. de la synthèse de protéines avec une séquence d'acides aminés qui suit celle indiquée sur l'ADN (Figure 1). La traduction nécessite principalement trois types d'ARN : (1) l'*ARN messenger* (ARNm) porte l'information d'un gène codant, i. e. dont la séquence correspond à une protéine ; (2) l'*ARN ribosomal* (ARNr) fait partie des *ribosomes*, structures faisant correspondre à chaque triplet de nucléotides (codon) l'ARNt qui lui est complémentaire ; et (3) l'*ARN de transfert* (ARNt) porte un acide aminé qui dépend de son anticodon, i. e. un triplet complémentaire du codon, toujours avec la même relation, que l'on appelle le *code génétique* (Figure 2). Le ribosome, composé d'ARNr et de protéines, lit la séquence nucléotidique des ARNm trois par trois et fait correspondre à chaque codon l'ARNt transportant l'acide aminé qui lui est associé selon le code génétique. Ensuite, le ribosome attache les acides aminés dans l'ordre stipulé par l'ARNm (Figure 1), produisant des protéines qui, à leur tour, seront responsables de la plupart des activités catalytiques et régulatrices de la cellule, comme enzymes du métabolisme mais aussi participant à la synthèse d'ADN et d'ARN.

Le mécanisme ADN-ARN-protéines est présent chez tous les organismes connus à ce jour ainsi que le souligne la célèbre épigramme de Jacques Monod et François Jacob « Tout ce qui est vrai pour le colibacille est vrai pour l'éléphant » [Monod et Jacob 1961]². La caractéristique *universelle* (i. e. présente dans tous les organismes connus) la plus saisissante est probablement le code génétique qui régit la relation qui existe entre un codon de l'ARNm et un acide aminé. Le changement de cette relation aurait des répercussions gravissimes car il entraînerait virtuellement la modification de toutes les protéines de la cellule. On a donc

² L'histoire de cette citation est en réalité assez compliquée. Ce serait Albert Jan Kluver qui en 1926 aurait affirmé l'unicité de la biochimie par la phrase « Entre l'éléphant et la bactérie de l'acide butyrique, c'est du pareil au même » [Kluver 1959]. L'idée de l'unicité de la biochimie a été rapidement acceptée et s'est répandue dans toute la biologie. Dans leur article de 1961, Monod et Jacob ne font que rappeler de façon approximative la formule de Kluver pour étendre à la biologie moléculaire l'unité que l'on observait en biochimie. Pourtant c'est leur formulation qui a été retenue, spécialement en biologie moléculaire. Cet aphorisme a d'ailleurs souvent été critiqué pour avoir donné la sensation que l'étude d'*E. coli* et autres organismes modèles pouvait se faire au détriment de l'analyse de la diversité, mais il reste valable dans son message original : l'unicité du système ADN-ARN-protéines et, en incluant la notion de Kluver, de la biochimie.

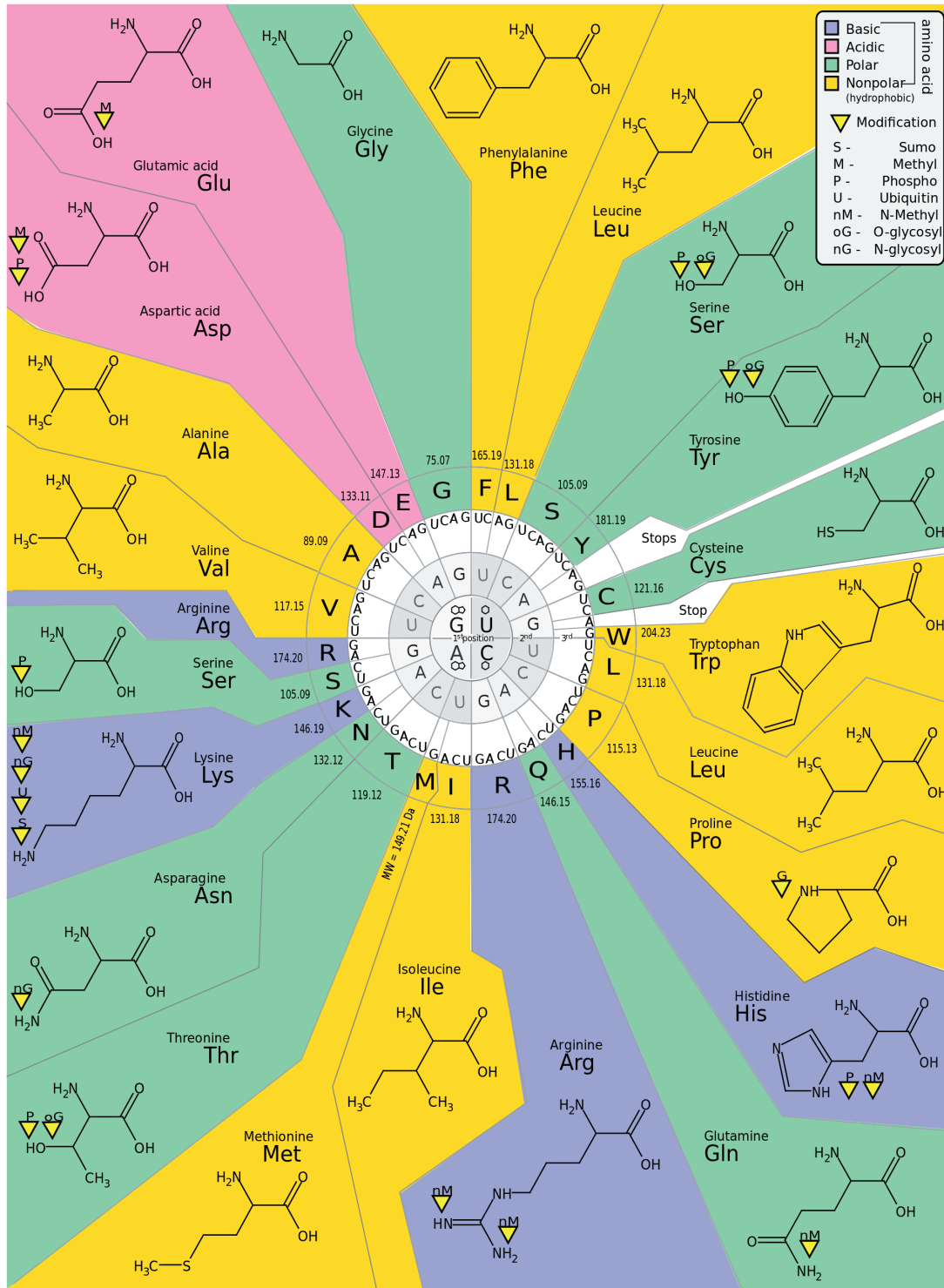


Figure 2. Code génétique universel.

Relation fixe qui existe entre une suite de trois nucléotides de l'ARNm (codon) et un acide aminé. La lecture du codon dans cette figure se fait du centre vers l'extérieur. Les couleurs correspondent aux propriétés chimiques des acides aminés. Image libre de droits (K. Gramatikoff).

postulé qu'une telle transformation devait être très rare. Une fois mis en place, ce code ne pourrait plus changer au sein d'une lignée, de telle manière que la détermination du nombre de codes génétiques existants dans les organismes vivants permettrait de déterminer le nombre de lignées d'origines différentes qui composent le vivant actuel [Hinegardner et Engelberg 1963]. Aujourd'hui on sait que le code est universel à quelques exceptions près, très rares, limitées et récentes à l'échelle de l'évolution des organismes [Knight *et al.* 2001 ; Miranda *et al.* 2006], ce qui logiquement soutient l'origine unique des organismes actuels.

En résumé, le système ADN-ARN-protéines (comprenant le code génétique et les ribosomes) est universel et fournit la base matérielle qui manquait à l'évolution darwinienne et à la théorie cellulaire pour confirmer l'ascendance commune des organismes actuels. Le fait que tous les organismes descendent d'un organisme commun est ainsi majoritairement accepté aujourd'hui, malgré le fait que certains auteurs continuent sporadiquement de mettre en cause cette hypothèse et développent l'idée alternative selon laquelle les organismes actuels auraient divergé à partir de plusieurs ancêtres sans relation entre eux [Sober et Steel 2002].

Certains travaux ont essayé de discriminer entre ces deux hypothèses en construisant des démonstrations mathématiques [Penny *et al.* 2003 ; Theobald 2010], mais il me semble important de souligner que la descendance à partir d'un ancêtre commun, tout en étant *a priori* réfutable, comme toute hypothèse scientifique, est intrinsèque à la biologie actuelle, et que son désaveu soulèverait aujourd'hui beaucoup plus de questions qu'il n'apporterait de réponses. Par définition, l'hypothèse de l'origine commune permet d'expliquer la plupart des similarités entre organismes comme étant le résultat intrinsèque de leur parenté. Dans l'hypothèse de l'existence de plusieurs ancêtres totalement indépendants les uns des autres, les similarités devraient être interprétées comme des convergences, i. e. des chemins évolutifs différents conduisant à des résultats qui se ressemblent. Des cas de convergence existent en biologie et en général on les détecte parce que, tout en ayant des caractéristiques ressemblantes, les éléments comparés portent des particularités qui manifestent leur origine et leur histoire évolutive différentes. Par exemple, on connaît des protéines dont le site actif a convergé à partir de séquences sans rapport évolutif parce que la configuration de ce site actif est thermodynamiquement la plus apte à réaliser leur fonction, mais le reste de la structure des protéines diffère [Elias et Tawfik 2012 ; Galperin et Koonin 2012]. En revanche, la machinerie de traduction semble être un système trop sophistiqué et similaire dans l'ensemble du vivant pour avoir convergé plusieurs fois de façon indépendante sans laisser de traces.

Autrement dit, pour que l'hypothèse de la convergence puisse remplacer celle de parenté commune il faudrait apporter des arguments supplémentaires qui expliqueraient comment, parmi tous les choix possibles, toutes les lignées ont convergé vers ce même système. En attendant ces arguments, l'évolution contingente du système et la descendance à partir d'un ancêtre commun restent les meilleures explications dont on dispose.

Ainsi, c'est sur la base de critères théoriques mais très bien soutenus que l'on accepte que les mécanismes moléculaires fondamentaux du fonctionnement et de l'hérédité de toutes les cellules partagent un ancêtre commun. A partir de là, la description de l'organisme ancestral dont sont issus les lignées modernes devient un enjeu pour comprendre l'évolution et le fonctionnement des organismes.

Le dernier ancêtre commun de tous les organismes : définitions

Les bases fondamentales de la biologie moderne soutiennent que tous les organismes actuels³ partagent un dernier ancêtre commun. Le terme *dernier* se réfère au fait que cet organisme est le plus proche ancêtre de tous les organismes actuels, pour le différencier de tous les autres organismes qui ont pu exister avant lui et qui, tout en étant aussi des ancêtres des formes de vie actuelles, ne sont pas leurs plus proches parents communs. Autrement dit, lorsqu'on parle de dernier ancêtre commun de tous les organismes, nous ne sommes pas en train de parler de l'origine de la vie.

La différence entre les deux est importante car le dernier ancêtre commun contient des machineries comme celle de traduction intégrant une grande quantité d'éléments qui n'ont certainement pas pu se regrouper en un seul événement. Au contraire, il a dû y avoir une évolution entre les premières formes de vie et le dernier ancêtre commun pour permettre la mise en place de ces structures que l'on infère de la comparaison des organismes actuels [Zhaxybayeva et Gogarten 2004 ; Forterre *et al.* 2005 ; Becerra *et al.* 2007a]. Une autre différence importante entre les travaux sur les premières formes de vie et le dernier ancêtre commun est méthodologique : le fait que tout le vivant utilise les mêmes macromolécules (ADN, ARN, protéines) permet de comparer ces macromolécules dans les organismes actuels et d'inférer la présence d'un inventaire des fonctions présentes chez le dernier ancêtre du

³ Les fossiles dont on dispose sont logiquement aussi issus de ce même ancêtre, car ils sont généralement rattachables à des lignées connues, mais ils ne seront pas pris en compte ici car on ne dispose pas sur eux des informations moléculaires nécessaires à la comparaison phylogénomique. Pour cette raison, on parle généralement de dernier ancêtre commun des organismes *actuels*.

vivant actuel par une approche que l'on dit *top-down*. En dehors de quelques rares machineries qui ont été dupliquées auparavant et qui avaient déjà dérivé à l'époque du dernier ancêtre de tous les organismes, il est par définition très difficile d'appliquer la comparaison des macromolécules à des étapes ayant précédé cet organisme. Les conditions qui ont pu permettre l'apparition de la vie sont étudiées par une approche très différente, dite *bottom-up*, qui fait intervenir des méthodes de chimie, géologie, astrophysique, astrochimie, etc. en essayant de décrire les circonstances dans lesquelles des molécules organiques ont pu être synthétisées, accumulées et auto-organisées pour former les premières entités vivantes [Gargaud *et al.* 2006]. Le sujet de l'origine de la vie est très vaste et ne sera pas traité ici. Je ne résumerai plus loin que les informations qui concernent la possible origine des membranes biologiques, à cause de son rapport direct avec le sujet de mon travail. Enfin, un autre point qui rend importante la distinction entre dernier ancêtre commun et premières formes de vie est conceptuel : l'apparition de la vie n'a pas forcément été un phénomène unique alors que la notion de dernier ancêtre commun souligne par définition la monophylie (descendance commune) des organismes actuels.

Malgré ces différences, la confusion entre dernier ancêtre commun de tous les organismes et première entité vivante est assez fréquente. Un exemple symbolique en est donné par Darwin lui-même, qui dans l'épigraphe présentée plus haut fait déjà cet amalgame⁴. Comme nous avons maintenant accès à davantage d'informations sur les périodes anciennes de la vie, cette distinction est devenue plus évidente mais reste parfois mal comprise. Une difficulté majeure réside dans le fait qu'aucune définition consensuelle de la vie n'existe, de manière que certaines hypothèses qui essayent de faire le lien entre les étapes prébiotiques, les origines de la vie et les formes de vie actuelles rapprochent énormément le premier organisme vivant du dernier ancêtre de tous les organismes [Woese et Fox 1977a ; Wächtershäuser 1988 ; Martin et Russell 2003 ; Dagan *et al.* 2010]. Ce rapprochement s'est même retrouvé dans le premier nom qui a été donné au dernier ancêtre, celui de *progénote*, i. e. un organisme très

⁴ Dans la version de 1860 de *L'Origine des espèces*, Darwin a rajouté à la fin de « from some one primordial form, into which life was first breathed » les mots « by the Creator » probablement à cause de la pression autour de ce sujet sensible dans la société victorienne. Dans les versions suivantes la phrase « into which life was first breathed [by the Creator] » a définitivement disparu (il est possible de comparer toutes ces versions grâce à l'outil en ligne <http://darwin-online.org.uk/Variorum/> [Van Wyhe 2002]). La raison de cette omission a probablement été d'éviter de parler de Création, car Darwin se préoccupait assez peu d'expliquer l'origine de la vie. L'analyse de sa correspondance suggère qu'il pensait que la question de l'origine de la vie pourrait être un jour envisagée scientifiquement, mais que les connaissances de son époque ne permettaient pas de le faire. Ainsi, il plaide pour la prudence et ne traite quasiment jamais ce sujet dans ses publications [Peretó *et al.* 2009].

primitif contenant une première et très inefficace machinerie de traduction et chez lequel la distinction entre phénotype et génotype n'était pas encore bien définie [Woese et Fox 1977a]. Depuis, les noms qui ont été donnés ont été plus neutres : cenancêtre, (cen- étant une contraction des mots grecs *kainos*, « récent », et *koinos*, « commun », [Fitch et Upper 1987]), LUCA (Last Universal Common Ancestor, [Kyrpides *et al.* 1999]), MCRA (Most Recent Common Ancestor, [Galtier *et al.* 1999 ; Fournier et Gogarten 2010]), etc. Dans mon travail je parlerai généralement de cenancêtre car il s'agit de la première dénomination neutre qui ait été proposée.

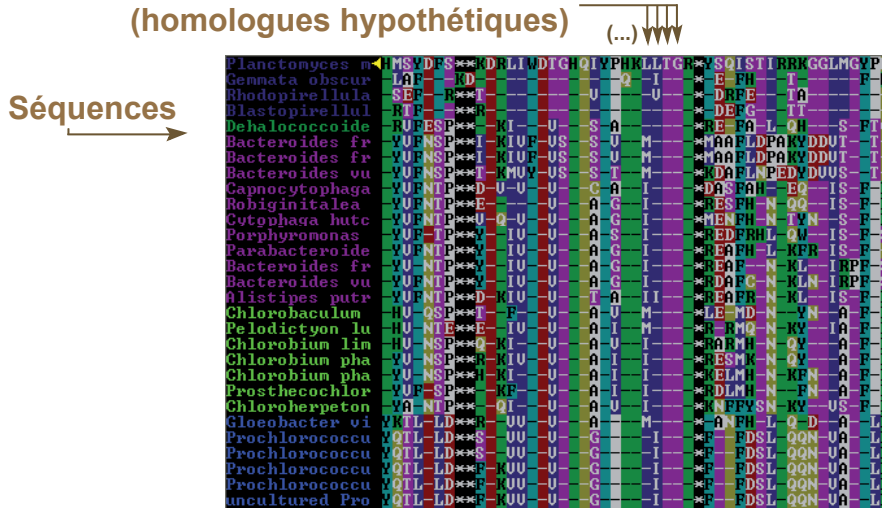
A la rencontre du cenancêtre

Dès leur formulation, l'unicité et l'ascendance commune du vivant ont encouragé les biologistes à essayer de déterminer les relations de parenté qui existent entre les espèces au moyen de ce l'on appelle des *phylogénies*. Il est important de rappeler la différence qui existe entre *généalogie* et *phylogénie*. Une *généalogie* est une figure arborescente qui montre tous les liens génétiques connectant les ancêtres à leurs descendants. Il s'agit d'une nécessité théorique qui découle des fondements de la biologie mais qui reste inaccessible dans la mesure où il nous faudrait détailler les rapports qui ont existé entre toutes les générations depuis le cenancêtre. Une *phylogénie* est une figure arborescente qui montre des degrés d'apparentements relatifs, i. e. qui représente les organismes qui ont hérité des caractères d'un ancêtre commun. Dans ce cas, les ancêtres sont virtuels, comme des portraits robots que l'on peut reconstruire en recoupant les informations obtenues de nombreux organismes. En biologie évolutive on peut reconstruire des *phylogénies* en utilisant des caractères phénotypiques ou moléculaires. Ces *phylogénies* peuvent être ensuite utilisées pour proposer une *phylogénie* des organismes (souvent appelée *arbre du vivant*) et/ou pour comprendre l'évolution des fonctions et structures biologiques.

Traditionnellement c'étaient les caractères phénotypiques qui étaient majoritairement utilisés pour la reconstruction de *phylogénies*. Dès que l'on a compris que tous les organismes fonctionnent sur la base de molécules telles que l'ADN, ARN et protéines, il a été proposé que les séquences de ces molécules puissent être utilisées comme autant de caractères pour étudier l'évolution des processus biologiques et des organismes eux-mêmes [Margoliash 1963 ; Zuckerkandl et Pauling 1965 ; Fitch et Margoliash 1967a]. Ceci a été particulièrement important dans l'étude de l'évolution des microorganismes car leur diversité morphologique

ALIGNEMENT MULTIPLE

Exemples de positions conservées
(homologues hypothétiques)



	C.97	C.98	C.99	C.00	C.01	C.02	J.97	J.98	J.99	J.00	J.01	J.02
C.97	0	0.110	0.116	0.113	0.124	0.145	0.197	0.167	0.183	0.179	0.173	0.168
C.98	0.110	0	0.089	0.094	0.098	0.119	0.182	0.148	0.168	0.167	0.162	0.151
C.99	0.116	0.089	0	0.088	0.092	0.123	0.180	0.151	0.160	0.163	0.160	0.155
C.00	0.113	0.094	0.088	0	0.083	0.118	0.171	0.142	0.155	0.150	0.147	0.146
C.01	0.124	0.098	0.092	0.083	0	0.107	0.166	0.143	0.150	0.150	0.139	0.148
C.02	0.145	0.119	0.123	0.118	0.107	0	0.178	0.150	0.168	0.168	0.161	0.139
J.97	0.197	0.182	0.180	0.171	0.166	0.178	0	0.133	0.123	0.124	0.126	0.175
J.98	0.167	0.148	0.151	0.142	0.143	0.150	0.133	0	0.114	0.118	0.113	0.131
J.99	0.183	0.168	0.160	0.155	0.150	0.168	0.123	0.114	0	0.094	0.096	0.158
J.00	0.179	0.167	0.163	0.150	0.150	0.168	0.124	0.118	0.094	0	0.092	0.152
J.01	0.173	0.162	0.160	0.147	0.139	0.161	0.126	0.113	0.096	0.092	0	0.144
J.02	0.168	0.151	0.155	0.146	0.148	0.139	0.175	0.131	0.158	0.152	0.144	0

Construction d'une matrice de similarité entre les séquences

RECONSTRUCTION PHYLOGENETIQUE

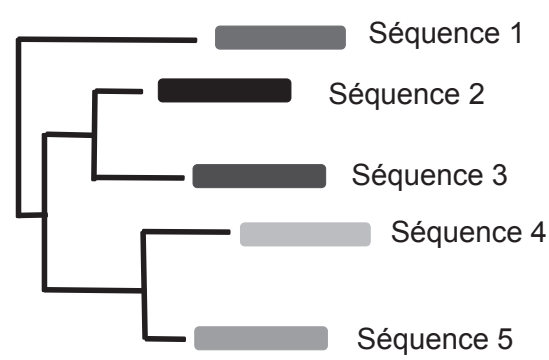


Figure 3. Alignements et phylogénies

est assez limitée alors qu'ils ont largement dominé l'histoire de la Terre et constituent toujours la partie la plus vaste de la diversité génétique [López-García et Moreira 2008].

Pour comparer des structures biologiques, il faut s'assurer qu'elles soient homologues entre elles, i. e. qu'elles partagent un ancêtre commun. Dans la pratique il est impossible de connaître directement l'ancêtre de deux séquences, mais on postule que deux séquences qui se ressemblent sont probablement issues d'un ancêtre commun. Souvent, l'homologie entre séquences est soutenue par le partage d'une même fonction par les molécules qu'elles codent, mais il existe aussi des arguments statistiques qui conduisent à penser que des séquences similaires ont partagé une histoire commune : si l'on fait un calcul très simplificateur, pour une séquence nucléique de n nucléotides, sachant que l'alphabet de l'ADN utilise quatre bases, il existe 4^n séquences différentes possibles ; pour une séquence protéique de m acides aminés, sachant que l'alphabet des protéines utilise généralement 20 acides aminés, il est possible d'imaginer 20^m séquences différentes possibles. Etant donné que la taille moyenne d'une protéine est d'environ 200 acides aminés et que le gène correspondant aura 600 nucléotides, le nombre de séquences protéiques possibles de $m = 200$ sera de l'ordre de 10^{260} et le nombre de séquences nucléiques possibles de $n = 600$ sera de l'ordre de 10^{361} . Ces nombres⁵ sont tellement grands que l'on considère comme plus probable, connaissant les règles d'hérédité, que deux séquences qui se ressemblent en séquence et en fonction aient divergé d'un ancêtre commun au lieu de se ressembler par hasard.

La comparaison de séquences passe par leur alignement, i. e. la construction d'une matrice dont chaque ligne contient une séquence et chaque colonne représente un site (un nucléotide pour les séquences nucléiques ou un acide aminé pour les protéines) que l'on postule être homologue entre les séquences de l'alignement (Figure 3). L'information contenue dans chaque position homologue permettra de construire des matrices de similarité entre les séquences et de construire ainsi des arbres phylogénétiques montrant les degrés d'apparentement relatifs entre les séquences (Figure 3). Ces techniques ont été développées très tôt après la proposition de l'utilisation des informations moléculaires pour reconstruire l'histoire des organismes [Margoliash 1963 ; Fitch et Margoliash 1967a]. Mais à cette époque il existait trop peu de séquences protéiques ou nucléiques capables de servir à la reconstruction des phylogénies, et ce encore moins à l'échelle de l'ensemble du vivant.

⁵ A titre de comparaison, on estime à environ 10^{80} le nombre d'atomes dans l'univers.

Conceptuellement, la caractérisation du cenancêtre semblait donc possible mais sa réalisation devait encore attendre que des nouvelles méthodes produisent des données moléculaires à grande échelle dans toute la diversité du vivant.

Dans ce contexte, Carl Woese a été une figure clé dans le travail moléculaire pour la comparaison universelle des organismes. Je vais présenter maintenant dans le détail une partie du travail de cet auteur et de ses collaborateurs, comme point de départ des discussions actuelles sur la nature du cenancêtre et sur l'évolution dans l'ensemble du vivant. Bien évidemment, ces chercheurs n'ont pas été les seuls à contribuer de façon importante à ce sujet mais ayant été les pionniers des premières études moléculaires à vocation vraiment universelle, ils ont logiquement formulé un grand nombre d'hypothèses en relation avec leurs résultats. Ces hypothèses ont organisé les différents débats qui s'en sont suivis et qui se poursuivent encore.

2. Woese : le bouleversement de l'histoire du vivant

Les trois « royaumes primaires »

Jusque dans les années 1970, la classification phylogénétique des organismes macroscopiques, comme les animaux, les plantes et les champignons, et même de certains protistes (organismes microscopiques avec des ultrastructures identifiables au microscope), était possible grâce principalement à l'utilisation de critères morphologiques. Cependant, pour la plupart des microorganismes les caractéristiques morphologiques (forme, taille, enveloppe et moyen de mobilité) étaient trop peu variables et les propriétés biochimiques apparemment trop chaotiques et peu conservées pour permettre de construire une classification fiable [Stanier et Van Niel 1962 ; Fox *et al.* 1980 ; Woese 1987 ; Sapp 2005]. De ce fait, plusieurs systèmes de classification avaient été proposés [Copeland 1938 ; Whittaker 1969] mais dans tous les cas les microorganismes étaient déconsidérés.

L'un des rares critères morphologiques qui permettait de distinguer les microorganismes les uns des autres était la présence ou l'absence de noyau dans les cellules. Il s'agit, d'une part, des organismes dont les cellules ont un noyau et d'autres compartiments cytoplasmiques, nommés organites (animaux, plantes, champignons, protistes) et que l'on appelle eucaryotes (du grec *eu*, « vrai », et *karuon*, « noyau »). D'autre part, les procaryotes (du grec *pro*, « avant ») sont définis principalement par l'absence du noyau, ainsi que des autres organites intracellulaires visibles chez les eucaryotes. Dans un premier temps on classait parmi les

procaryotes les bactéries et les algues bleues-vertes (*blue-green algae*) que l'on appelle maintenant cyanobactéries car elles appartiennent en réalité au groupe des bactéries. Cette différence, qui dans un premier temps était cytologique [Stanier et Van Niel 1962], avait été acceptée comme une classification phylogénétique [Sapp 2005].

Woese et Fox critiquent cet état de fait car on ne peut pas asseoir l'existence d'un groupe taxonomique sur l'absence d'un caractère partagé, ce qui est fait dans le cas des procaryotes. De plus, le choix du préfixe « pro » pour les procaryotes est tendancieux car il suggère que les procaryotes sont des organismes « antérieurs » aux eucaryotes alors qu'en réalité aucune analyse ne permettait alors de vérifier une telle hypothèse. Afin de construire une classification vraiment phylogénétique, basée sur l'apparentement des organismes, ils proposent d'utiliser un élément universel de la cellule. Le séquençage des protéines et de l'ADN étant laborieux dans les années 1970, Woese et Fox pensent à utiliser l'ARNr pour une étude comparative [Woese et Fox 1977b]. Woese avait déjà participé à des comparaisons similaires entre quelques bactéries [Sogin *et al.* 1971] mais dans leur nouvelle étude la comparaison s'étend à la diversité la plus large possible, comprenant des microorganismes, des cellules animales, végétales et même des ARNr chloroplastiques (les chloroplastes sont les organites responsables de la photosynthèse chez les eucaryotes et ils comportent leurs propres ADN et ribosomes).

Le ribosome semble un bon candidat pour ce travail car il est universel, très abondant dans les cellules, composé de différentes molécules d'ARNr et de protéines. De plus, ces composants sont connus pour changer relativement peu selon les organismes [Moore et McCarthy 1967 ; Sogin *et al.* 1971]. Ce sont donc des bons marqueurs pour travailler sur des événements anciens. Parmi les ARNr qui sont à leur disposition, Woese et Fox décident d'utiliser l'ARNr 16S⁶ ou son équivalent 18S dans certains organismes. Il s'agit d'un ARN de taille moyenne, que l'on pense assez long pour contenir une information discriminante entre taxons sans être trop difficile à manipuler. Les ARNr 16S/18S marqués radioactivement sont indépendamment dégradés par une enzyme (ribonucléase T1) qui coupe spécifiquement après les positions G des ARN simple brin, de manière que chaque ARNr est divisé en morceaux plus courts (oligonucléotides) en fonction de leur séquence. Les oligonucléotides de chaque

⁶ S est pour Svedberg, une mesure du taux de sédimentation qui est utilisé notamment pour la caractérisation des ARN. Les ARNr type des ribosomes de la bactérie *Escherichia coli* ont un taux de sédimentation de respectivement 23S, 5S et 16S ; ceux d'un eucaryote comme l'humain sont de 28S, 5,8S, 5S et 18S.

Tableau 1. Résultats d'après Woese et Fox 1977.

Les chiffres correspondent au coefficient de similarité (S_{AB}) entre les catalogues, comme expliqué dans le texte et la note de bas de page numéro 7. Le coefficient a des valeurs d'autant plus importantes que la similarité est forte.

	1	2	3	4	5	6	7	8	9	10	11	12	13	
"Urkaryotes"	1. <i>Saccharomyces cerevisiae</i> , 18S	-	0,29	0,33	0,05	0,06	0,08	0,09	0,11	0,08	0,11	0,11	0,08	0,08
	2. <i>Lemna minor</i> , 18S	0,29	-	0,36	0,1	0,05	0,06	0,1	0,09	0,11	0,1	0,1	0,13	0,07
	3. Cellule L (humaine)	0,33	0,36	-	0,06	0,06	0,07	0,07	0,09	0,06	0,1	0,1	0,09	0,07
"Eubactéries"	4. <i>Escherichia coli</i>	0,05	0,1	0,06	-	0,24	0,25	0,28	0,26	0,21	0,11	0,12	0,07	0,12
	5. <i>Chlorobium vibrioforme</i>	0,06	0,05	0,06	0,24	-	0,22	0,22	0,2	0,19	0,06	0,07	0,06	0,09
	6. <i>Bacillus firmus</i>	0,08	0,06	0,07	0,25	0,22	-	0,34	0,26	0,2	0,11	0,13	0,06	0,12
	7. <i>Corynebacterium diphtheriae</i>	0,09	0,1	0,07	0,28	0,22	0,34	-	0,23	0,21	0,12	0,12	0,09	0,1
	8. <i>Aphanocapsa</i> 6714	0,11	0,09	0,09	0,26	0,2	0,26	0,23	-	0,31	0,11	0,11	0,1	0,1
	9. Chloroplastes (<i>Lemna</i>)	0,08	0,11	0,06	0,21	0,19	0,2	0,21	0,31	-	0,14	0,12	0,1	0,12
"Archaeobactéries"	10. <i>M. thermoautotrophicum</i>	0,11	0,1	0,1	0,11	0,06	0,11	0,12	0,11	0,14	-	0,51	0,25	0,3
	11. <i>M. ruminantium</i> M-1	0,11	0,1	0,1	0,12	0,07	0,13	0,12	0,11	0,12	0,51	-	0,25	0,24
	12. <i>Methanobacterium</i> sp. JR-1	0,08	0,13	0,09	0,07	0,06	0,06	0,09	0,1	0,1	0,25	0,25	-	0,32
	13. <i>Methanosarcina barkeri</i>	0,08	0,07	0,07	0,12	0,09	0,12	0,1	0,1	0,12	0,3	0,24	0,32	-

ARN sont alors séparés les uns des autres dans une électrophorèse en deux dimensions, ce que la radioactivité permet de révéler avec du papier photographique. Chaque distribution d'oligonucléotides est caractéristique d'un ARNr, i. e. approximativement de chaque espèce. Ces oligonucléotides étant de petite taille, on peut les séquencer plus facilement que les séquences entières. On obtient ainsi un catalogue de séquences nucléotidiques pour chaque espèce. Les oligonucléotides de petite taille peuvent être issus du hasard et ne sont donc pas pris en compte lorsqu'ils ont une taille inférieure à 6 nucléotides (choix arbitraire). Pour les plus grands, chacun d'entre eux est comparé à tous les autres nucléotides des catalogues des autres espèces, de sorte que plus les catalogues ont des oligonucléotides en commun, plus on considère que les ARNr (et par conséquent les espèces porteuses) sont proches parents⁷.

Les résultats ainsi obtenus par Woese et Fox montrent que les séquences forment trois grands groupes (Tableau 1). Sur les trois ensembles, deux correspondent à des organismes sans noyau (procaryotes) et le troisième correspond aux cytoplasmes des eucaryotes. Les séquences des chloroplastes sont regroupées dans un des groupes d'organismes sans noyau. Ils proposent donc de remplacer la séparation en procaryotes et eucaryotes par une classification phylogénétique en trois groupes qu'ils appellent dans un premier temps « royaumes primaires » ou ur-royaumes (certainement en référence à l'ancienneté du royaume mésopotamien d'Ur). Le premier groupe est appelé *eubactéries* et il contient la plupart des bactéries des études classiques et les chloroplastes. Le second comprend les eucaryotes ou *ur-caryotes*, qui correspondent aux cytoplasmes des organismes avec un noyau. Le troisième regroupe quelques organismes méthanogènes sans noyau et sur lesquels on ne savait pas grande chose à l'époque ; à ce moment, on les appellera *archaebactéries*.

Les deux premiers « royaumes primaires » étaient bien caractérisés et leur acceptation ne semblait pas poser de problème particulier, mais les auteurs savaient que la séparation de quelques procaryotes méthanogènes dans un groupe aussi distinct des deux autres « royaumes » pouvait soulever des réserves. Par conséquent, ils essayèrent de trouver d'autres particularités que l'ARNr divergent pour permettre de distinguer les archaebactéries. Ils

⁷ Bien entendu, il faut aussi tenir compte de la taille des catalogues. Le calcul exact du coefficient de similarité entre catalogues (S_{AB}) tient compte du nombre d'oligonucléotides en commun (N_{AB}), et du nombre d'oligonucléotides dans chaque catalogue (N_A , N_B) selon l'expression suivante : $2N_{AB}/(N_A+N_B)$. Le S_{AB} est égal à 1 si on a deux catalogues identiques et égal à 0 si les catalogues sont totalement différents. L'article de 1977 comporte peu d'explications, mais le protocole est plus clairement expliqué dans Sogin *et al.* 1972 et Fox *et al.* 1980.

proposèrent les quelques caractéristiques que l'on savait spécifiques des méthanogènes : des parois cellulaires particulières, une biochimie exotique et une modification particulière des ARNt. L'année suivante ils remarquèrent que des organismes halophiles faisaient aussi partie des archaeobactéries [Magrum *et al.* 1978]. Sachant que les halophiles et certains thermophiles qui pourraient faire partie des archaeobactéries ont des membranes possédant des propriétés particulières, ils ajoutèrent les membranes aux caractéristiques propres des archaeobactéries [Woese *et al.* 1978]. Le nom d'archaeobactéries (du grec *archaios*, « ancien » et *backterion*, « bâton ») a été donné parce que la méthanogenèse semblait caractéristique de ce groupe et que ce type de métabolisme paraissait très bien adapté aux conditions de la Terre primitive, suggérant par là que ces organismes pouvaient conserver un phénotype ancien. Quant au placement des chloroplastes parmi les eubactéries, il semble cohérent avec l'hypothèse selon laquelle cet organite était le résultat de l'endosymbiose d'une cyanobactérie, possibilité déjà envisagée à l'époque et qui est couramment acceptée aujourd'hui [Gray et Doolittle 1982]. Cette position des chloroplastes motivait d'ailleurs dans un premier temps le nom des urcaryotes, signalant qu'il s'agissait là de la cellule eucaryote sans ses endosymbiontes.

Dans l'analyse de Woese et Fox, les trois « royaumes primaires » sont équidistants entre eux. Comme cette distance est mesurable en termes de quantité de mutations mais non pas de temps, il n'est pas possible de déterminer s'il existe un ordre de divergence des trois « royaumes » à partir du progénote. Autrement dit les trois « royaumes » semblent, selon ces résultats, se trouver tous les trois à la même distance phylogénétique du progénote. Ce n'est cependant pas dans cet article mais dans un autre publié la même année que Woese et Fox expriment leur point de vue théorique sur les relations entre les « royaumes primaires » ainsi que sur les relations de ces « royaumes » avec le progénote [Woese et Fox 1977a]. Concernant le rapport entre les eucaryotes et le progénote, contrairement à la vision la plus courante donnée par le nom même des *procaryotes*, ils ne pensent pas que les eucaryotes aient évolué à partir d'un procaryote de type moderne : si les ancêtres des eucaryotes ont été procaryotes dans le passé, c'est uniquement d'un point de vue cytologique, car ils ne devaient pas contenir de noyau, mais non pas d'un point de vue phylogénétique. Autrement dit, la lignée ur-caryote serait indépendante des autres lignées procaryotes et issue directement du progénote, de même que les deux autres « royaumes », comme on verra plus tard.

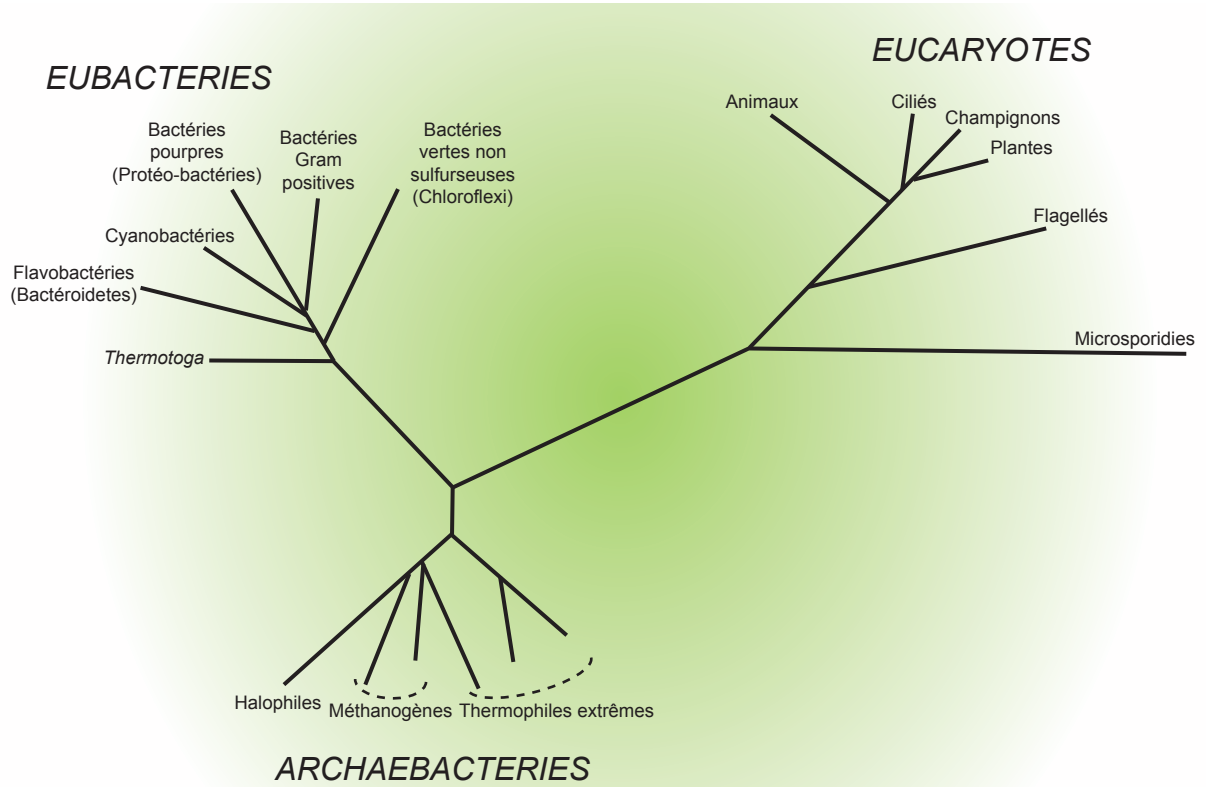


Figure 4. Arbre phylogénétique basé sur les séquences ADN des gènes codant les ARN 16S. D'après Woese 1987.

Relations entre les trois « royaumes primaires »

On remarquera que ces analyses avaient séparé les organismes en trois « royaumes » en utilisant non des phylogénies mais les catalogues d'ARNr. A cette époque il y avait peu de séquences longues disponibles. Les techniques de séquençage inventées dans les années 1970 ont été développées au cours des années 1980 et le séquençage d'ADN est alors devenu un peu plus aisé. Il devenait plus facile de séquencer les ADN codant l'ARNr 16S/18S d'organismes divers et de reconstruire des phylogénies moléculaires avec les séquences. Ces arbres confirment la séparation du vivant en trois « royaumes primaires » et constituent les premières phylogénies moléculaires à grande échelle pour les eubactéries et les archaebactéries [Woese 1987]. Dans ces arbres, bien que les trois groupes soient clairement séparés, la branche qui sépare un « royaume » de l'intersection avec les autres « royaumes » est plus courte pour les archaebactéries que pour les autres « royaumes » (Figure 4). De ce fait, Woese donne du crédit à la possibilité que les archaebactéries ressemblent plus à l'ancêtre de tous les organismes que les autres « royaumes » [Gupta *et al.* 1983].

Par ailleurs, comme une grande partie des archaebactéries connues à l'époque sont (hyper)thermophiles, il en infère que l'ancêtre des archaebactéries l'était aussi⁸. On avait observé que des eubactéries (hyper)thermophiles branchaient aussi à la base de leur « royaume » dans certaines phylogénies. C'était par exemple le cas des bactéries vertes non sulfureuses (aujourd'hui, Chloroflexi), qui sont thermophiles, et des Thermotogales, qui sont hyperthermophiles [Oyaizu *et al.* 1987 ; Woese 1987]. Du branchement basal des séquences hyperthermophiles dans deux « royaumes primaires » du vivant il conclue que le progénote devait être hyperthermophile et que les lignées qui avaient d'abord divergé de lui étaient les archaebactéries, puis les eubactéries quand la température avait un peu baissé [Woese 1987]. Comme on le verra, ce thème a été très débattu par la suite.

Dans son travail de 1987 sur la phylogénie des procaryotes, Woese reconnaît cependant que la longueur des branches n'est pas suffisante pour déterminer s'il existe un ordre de divergence particulier des « royaumes » à partir du cenancêtre, et que pour cela il faudrait *enraciner* l'arbre, i. e. l'orienter en déterminant l'ordre de divergence des lignées modernes depuis le cenancêtre. A cette fin, il suggère d'appliquer une méthode proposée dix ans

⁸ On a l'habitude de diviser les organismes en quatre catégories en fonction de leur température optimale de croissance : pour les psychrophiles, cette température est inférieure à 15°C, pour les mésophiles, elle est entre 15-45°C, pour les thermophiles elle est de 45-80°C et pour les hyperthermophiles, elle est de plus de 80°C.

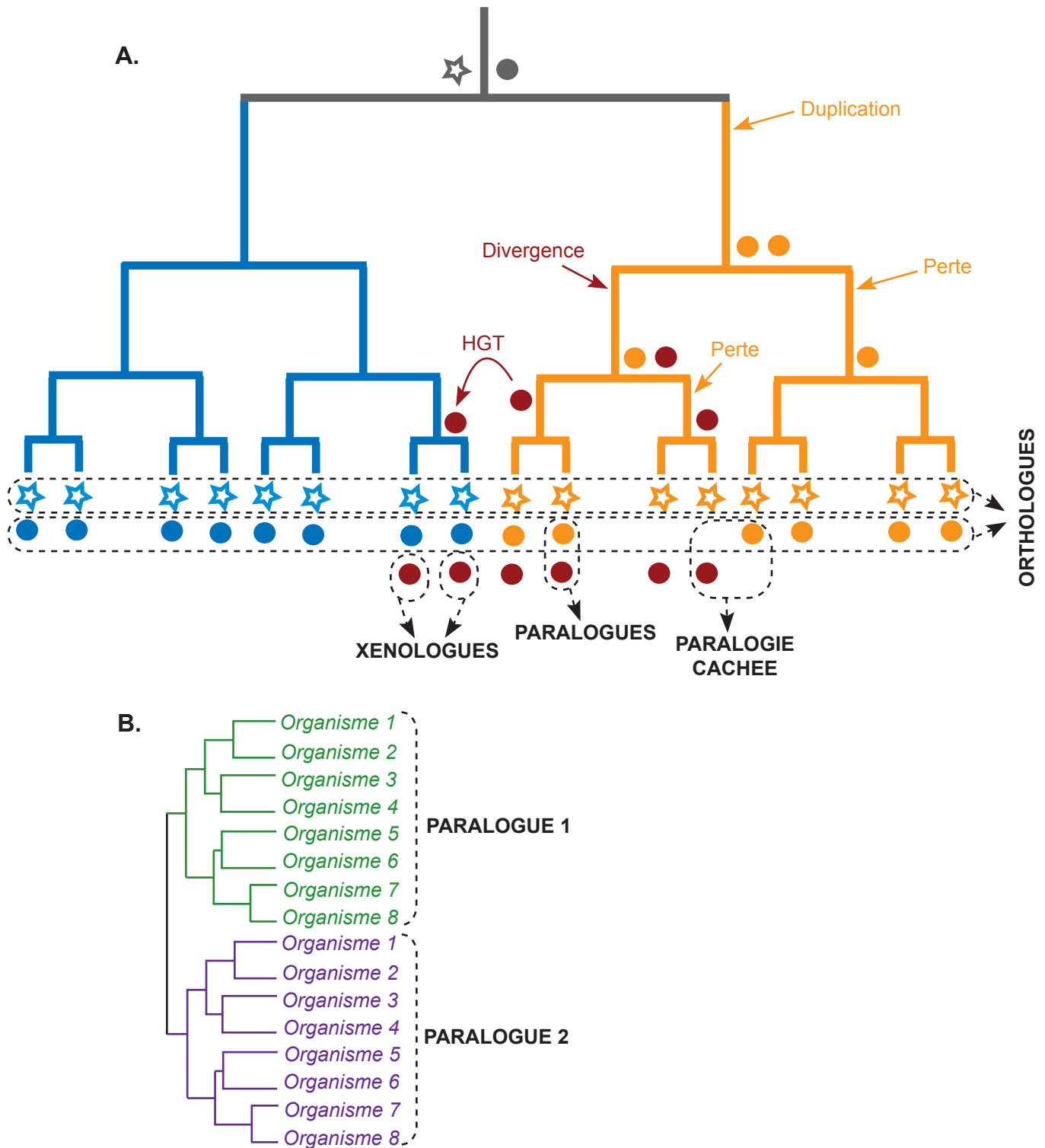


Figure 5. Orthologues, paralogues et xénologues.

A. L'arbre représente une évolution théorique des organismes. L'étoile et le cercle symbolisent des gènes et ils sont représentés sur un nœud à chaque fois qu'il se produit une modification remarquable les concernant. Le gène *étoile* évolue par spéciation dans toutes les lignées, donc les séquences sont toutes orthologues entre elles. Le gène *cercle* subit une duplication et les gènes résultants seront dits paralogues. Les xénologues sont les gènes acquis par transfert horizontal. **B.** L'arbre représente l'évolution théorique de deux gènes paralogues ayant subi leur duplication avant la divergence des organismes 1 à 8. Comme les gènes verts et violets ont évolué par spéciation dans les mêmes organismes, leurs phylogénies sont identiques et permettent de les enraciner réciproquement.

auparavant par Margaret Dayhoff : la reconstruction phylogénétique des gènes *paralogues* universels.

Deux gènes sont paralogues entre eux lorsqu'ils sont issus de la duplication d'un gène ancestral. De ce fait, on peut trouver les deux gènes paralogues dans un même génome, généralement accomplissant des fonctions similaires mais différentes. Les gènes paralogues se distinguent notamment des gènes *orthologues*, qui sont des séquences partageant un ancêtre commun par spéciation et non par duplication, et des gènes *xénologues*, qui ne proviennent pas de la lignée verticale de l'organisme mais sont le résultat de transferts horizontaux de gènes (Figure 5). Un cas remarquable se produit lorsque des gènes paralogues sont alternativement perdus dans différentes lignées, car leurs génomes contiendront des séquences homologues non orthologues et on parle alors de *paralogie cachée* (Figure 5).

Dans le cas présent, l'intérêt d'utiliser des paralogues universels est que, comme ils sont issus d'une duplication génique précédant le cenancêtre (puisque universels), la phylogénie de chaque gène permet d'enraciner réciproquement la phylogénie de l'autre (Figure 5). Cette technique avait été utilisée quelques années auparavant avec des gènes de la ferrédoxine. En effet, chaque gène de ferrédoxine est composé de deux séquences paralogues attachées entre elles, de manière que la phylogénie de chaque moitié de gène peut permettre de raciner la phylogénie de l'autre moitié [Schwartz et Dayhoff 1978]. Cependant, la ferrédoxine est une protéine très courte et seulement une quinzaine de positions conservées avaient été utilisées, ce qui est très insuffisant pour construire une phylogénie des organismes fiable. De plus, seules quelques séquences d'eubactéries et de chloroplaste avaient été utilisées à l'époque, ce qui ne permettait pas d'avoir une idée générale de la phylogénie dans les trois « royaumes ».

Les gènes paralogues qui avaient été utilisés pour enraciner les premiers arbres universels étaient ceux des ATPases et des facteurs d'élongation. Les ATPases sont des complexes moléculaires qui utilisent un gradient de protons pour synthétiser de l'ATP, la « monnaie énergétique » de la cellule. Elles sont composées d'un grand nombre de sous-unités dont deux qui sont des paralogues universels. Les facteurs d'élongation sont des protéines paralogues qui participent à la traduction. Les résultats de ces premières analyses ont révélé que les séquences eubactériennes semblaient dans tous les cas diverger en premier, alors que les archaebactéries et les eucaryotes étaient plus proches entre eux [Gogarten *et al.* 1989 ; Iwabe *et al.* 1989].

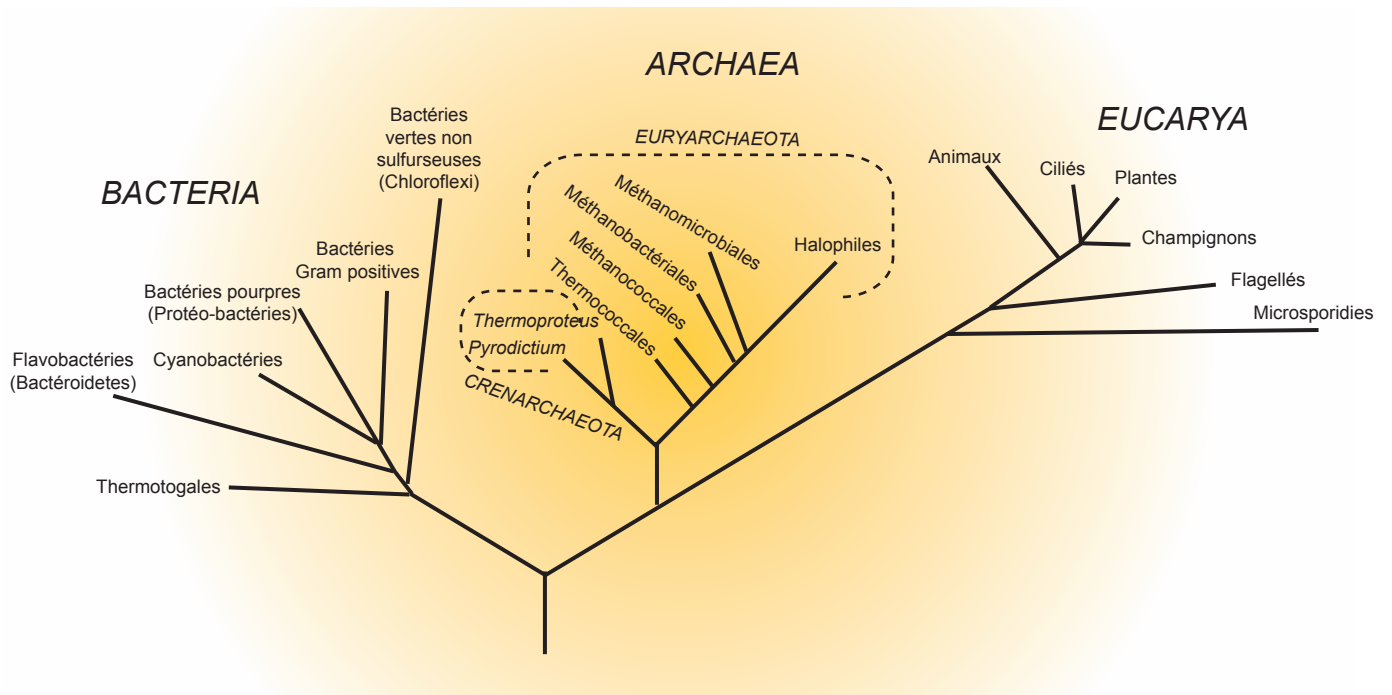


Figure 6. Les trois domaines du vivant.

D'après Woese *et al.* 1990. Cet arbre est identique à celui de 1987 (Figure 4) mais enraciné sur la base des résultats de Gogarten *et al.* 1989 et Iwabe *et al.* 1989.

Woese reprend ce résultat et enracine ainsi son arbre de 1987 basé sur les ARNr en regroupant les archées et les eucaryotes ensemble et en plaçant les bactéries comme premier groupe divergeant du cenancêtre (Figure 6, [Woese *et al.* 1990]). Avec ses collaborateurs, il change aussi une partie de la nomenclature et propose celle qui est actuellement utilisée. Au lieu de parler de « royaumes primaires », ils proposent de parler de *domaines* du vivant pour éviter la confusion avec d'autres groupes taxonomiques que l'on avait déjà l'habitude d'appeler royaumes, comme les animaux et les plantes, et qui se placent à un niveau taxonomique beaucoup plus restreint. Ils conservent le nom *Eucarya* pour les eucaryotes mais proposent de remplacer celui des archaebactéries par *Archaea* et celui des eubactéries par *Bacteria*. Le but est de souligner que les archées sont un domaine à part entière, très différent des bactéries malgré leur cytologie de type procaryote. L'étymologie est d'ailleurs conservée dans le nom des archées (*archaios*, « ancien ») pour rappeler le caractère soi-disant plus primitif de leur phénotype, notamment les hautes températures auxquelles sont adaptées les hyperthermophiles. Enfin, bien qu'ils ne rentrent pas dans le détail de la dénomination des taxons bactériens et eucaryotes, ils proposent des noms très significatifs pour les deux principaux groupes d'archées connus à l'époque : l'un sera appelé les Euryarchaeota (du grec, *euris*, « large »), pour signaler la grande diversité métabolique de ces organismes, et l'autre les Crenarchaeota (du grec, *crinis* « source »), pour signaler que leur phénotype thermophile est probablement celui qu'avaient les premières archées. Comme on le verra, cette classification et le placement de la racine à cet endroit ont suscité un grand nombre de débats qui se poursuivent actuellement.

Une vision particulière du cenancêtre : le progénote

Woese a dès le départ posé la question de la nature de ce qu'il appelle progénote [Woese et Fox 1977a]. Il est important de signaler que la définition originale du progénote est indépendante de celle du cenancêtre, bien que la plupart des travaux de Woese partent du postulat que le cenancêtre *était* un progénote.

Comme on le sait, le cenancêtre est le dernier ancêtre commun des organismes actuels et son existence est inférée par les bases théoriques de la biologie. Le progénote est une entité définie par Woese dans laquelle le phénotype et le génotype se confondent⁹ et dans laquelle la

⁹ En 1977 Woese n'explique pas vraiment ce qu'il entend par non-séparation de phénotype et génotype, mais dans son article de 1987 il fait clairement le rapprochement entre sa conception du progénote de celle du monde ARN postulé dans les hypothèses sur l'origine de la vie.

machinerie de traduction est encore en train d'être mise en place. Du fait du caractère primitif de son système de traduction, le progénote a une fidélité de synthèse des protéines faible. Beaucoup d'erreurs sont introduites et il s'ensuit que la capacité des protéines à réaliser des fonctions était instable, entraînant une faible qualité de la réplication de l'information génétique. Ces contraintes devaient limiter la taille du génome, qui devait alors être composé de gènes isolés. Woese insiste sur le fait que cette entité est très différente des cellules actuelles et ne peut pas être considérée comme un procaryote mais comme une forme beaucoup plus primitive. Au contraire, les domaines modernes du vivant seraient issus directement du progénote. Leurs dissemblances, comme l'organisation du génome et les mécanismes de réplication de l'ADN, seraient dues précisément au fait que le saut qualitatif entre le progénote et les cellules modernes aurait été franchi indépendamment trois fois, correspondant aux trois domaines du vivant.

Le progénote évoque fortement les mondes ARN et ARN-protéines, qui bénéficient d'un grand crédit dans les travaux sur les origines de la vie. L'hypothèse du monde ARN postule qu'il a existé une étape dans l'apparition de la vie pendant laquelle les ARN jouaient à la fois des rôles informationnels et catalytiques. Le monde ARN-protéines est souvent vu comme une étape postérieure dans laquelle des peptides auraient commencé à aider les ARN dans leurs activités catalytiques [Gilbert 1986]. Le rapprochement du progénote de l'origine de la vie n'est pas surprenant car dès 1965, en proposant une hypothèse sur l'*origine* du code génétique, Woese avait décrit un organisme très similaire au progénote mais auquel il ne donnait pas de nom [Woese 1967]. Ces travaux sont très stimulants pour l'étude de l'origine de la vie et l'apparition des *premières formes vivantes* mais ils ne sont pas cohérents avec la sophistication que l'on peut inférer chez le cenancêtre à partir de la comparaison des organismes actuels, notamment l'existence d'un ribosome et d'un contexte dans lequel celui-ci puisse fonctionner. Dans certains travaux, Woese fait explicitement la distinction entre progénote et cenancêtre [Woese 1987] mais conclut malgré tout que pour lui les deux coïncident. Qu'est-ce qui lui fait penser cela ?

Avant Woese, l'idée la plus répandue était que les eucaryotes étaient issus des procaryotes. Woese s'oppose à cette hypothèse car il sait que la diversité des procaryotes est très importante (il a précisément été un des moteurs de cette prise de conscience) et que leur fonctionnement n'est pas aussi trivial que d'autres semblent le penser. Son hypothèse est que les eucaryotes, et notamment leur cytoplasme, ne sont peut être pas si différents des procaryotes si l'on fait abstraction de leurs structures issues d'endosymbioses (mitochondries

et chloroplastes), qui elles-mêmes ont été des procaryotes. Si une cellule eucaryote sans endosymbionte n'est pas très différente d'une cellule procaryote, alors la différence intrinsèque entre ces domaines doit être le résultat d'une divergence à partir d'un organisme antérieur et différent des procaryotes. Autrement dit, pour que l'évolution fasse apparaître les organismes « complexes »¹⁰ modernes très différents, il faut imaginer un ancêtre très « simple » comportant une machinerie encore non mature par rapport à ce qui semble aujourd'hui être universel : la traduction. Et cette entité qui « n'est pas encore un procaryote » est le progénote. L'article de 1987 est probablement celui qui expose le plus clairement ce raisonnement, présent aussi de façon assez cohérente mais moins développée dans d'autres travaux similaires [Woese et Fox 1977a ; Woese 1998 ; Woese 2002]. Pourtant, la ressemblance entre l'origine de la traduction et le cenancêtre est tellement grande dans certains articles que la différence entre origine de la vie et cenancêtre (progénote) devient confuse [Woese 2000].

En résumé, le progénote est un organisme théorique imaginé *a priori* pour expliquer le fonctionnement d'un organisme primitif, et placé au niveau du cenancêtre pour expliquer les différences soi-disant inconciliables entre les domaines du vivant, mais sans essayer de décrire les propriétés du cenancêtre à partir des organismes modernes comme il serait logique de le faire puisque son existence repose sur leur comparaison. De fait il aurait été très difficile à l'époque de comparer les organismes modernes avec assez de détail pour inférer la présence de fonctions cellulaires chez le cenancêtre.

La question de savoir si le cenancêtre était ou non un progénote va nous intéresser pour deux raisons. La première, qui sera développée dans la suite de l'introduction, concerne la nature du cenancêtre et les raisons des très grandes divergences entre les trois domaines, qui sont des sujets à débat dans la littérature. La deuxième sera traitée dans la conclusion car l'hypothèse du progénote de Woese met en jeu des notions de « simplicité » et « complexité » du vivant qui abondent dans la littérature mais qui me semblent à la fois très mal définies et utilisées d'une façon inappropriée.

¹⁰ Comme je le développerai dans la conclusion, le mot « complexe » est très répandu dans la littérature mais rarement défini et souvent ambigu. En ce qui concerne ce texte, je me suis efforcé de remplacer ce mot par d'autres plus appropriés et précis. Lorsque je me réfère au travail d'un auteur il sera toujours entre guillemets pour signaler que j'utilise ce mot avec le sens que l'auteur en question a voulu lui donner.

3. Débats sur le cenancêtre

En dehors de la nature générale du cenancêtre, nous allons voir que les hypothèses de Woese ont structuré les débats actuels en rapport avec cet organisme : ses conditions de vie, la détermination de la racine de l'arbre du vivant, le rapprochement relatif entre les lignées modernes et le cenancêtre, ainsi que l'existence même de l'arbre du vivant.

Le cenancêtre : ni ancêtre totipotent, ni communauté

Quelles sont ces différences irréconciliables entre les trois domaines du vivant invoquées par Woese ? Kandler a été parmi ceux qui ont le plus clairement exposé cette problématique en signalant des caractéristiques partagées par seulement deux domaines sur trois : 1) les bactéries et les archées se ressemblent en ce qu'elles ont des génomes circulaires et des gènes organisés en unités transcriptionnelles (opérons) ; 2) les bactéries partagent avec les eucaryotes des phospholipides membranaires, très différents de ceux des archées comme on le verra plus tard, ainsi que leur phosphofructokinase (une enzyme de la glycolyse) ; 3) les archées et les eucaryotes sont plus proches dans l'arbre raciné du vivant sur la base de leurs ATPases et facteurs d'élongation [Kandler 1995]. On remarquera d'abord que ces différences majeures sont fondées sur les critères très disparates et difficilement comparables entre eux que sont l'organisation générale de la cellule, l'utilisation de certaines enzymes et la distance phylogénétique.

Cependant, Kandler a proposé une explication à cette répartition inégale entre les trois domaines du vivant, explication qui a eu des fortes répercussions sur le débat relatif au cenancêtre. Il postule que toutes ces propriétés existaient à l'époque du cenancêtre à l'échelle d'une *communauté de pré-cellules* mais que l'ensemble de ces caractéristiques n'était présent dans aucun organisme en particulier. A cette époque l'échange direct de gènes (que l'on appelle généralement *transfert horizontal de gènes*, HGT) était très important et permettait par la même occasion l'échange facile de fonctions biologique. Les cellules modernes seraient apparues lorsque la « cellularisation » de ces entités aurait bloqué l'échange d'informations et figé ainsi des sous-ensembles aléatoires des propriétés de la communauté de départ [Kandler 1995]. La communauté d'organismes s'oppose à l'idée que le cenancêtre était un organisme unique.

Woese a considéré d'une façon très favorable cette possibilité et l'a intégrée dans ses hypothèses sur le cenancêtre. En effet, il s'oppose à l'idée que le cenancêtre ait pu posséder

toutes les caractéristiques qui sont partagées deux à deux par les domaines du vivant car cela ferait de lui un organisme totipotent, encore plus « complexe » que les organismes modernes [Olsen et Woese 1996 ; Woese 1998]. A son avis, un tel organisme n'aurait pas eu le temps d'émerger entre le moment de la formation de la Terre et celui correspondant aux premières traces fossiles de marqueurs biologiques, lapse de temps qu'il estime à moins d'un milliard d'années. Il imagine donc une communauté de progénètes, à la fois très similaires au progénète déjà décrit et aux pré-cellules de Kandler. Woese utilise une métaphore pour expliquer ce modèle : le *recuit physique* (*physical annealing*). Le recuit physique se produit lorsqu'un matériau à très haute température, par exemple un métal chauffé, adopte des propriétés particulières au fur et à mesure qu'il perd de la température. Dans son hypothèse, Woese compare l'échange génétique à la température. Au début l'échange génétique aurait été très important au point que la généalogie n'aurait pas encore existé. Ensuite, la « température » de ces échanges se serait réduite progressivement au fur et à mesure du perfectionnement des machineries, établissant un premier palier de sous-ensembles au sein desquels les échanges étaient encore possibles : il s'agirait des ancêtres des trois domaines. Enfin, l'apparition d'autres paliers de « température » à l'intérieur des domaines aurait « figé » les lignées à l'intérieur de chaque domaine [Woese 1998].

Woese n'est pas le seul à avoir été séduit par l'idée que le cenancêtre était une communauté [Kurland *et al.* 2003 ; Wächtershäuser 2003 ; Koonin 2007 ; Glansdorff *et al.* 2008 ; Lake *et al.* 2009 ; Olendzenski et Gogarten 2009] et ce d'autant plus que, comme nous le verrons plus tard, beaucoup d'auteurs pensent que les transferts horizontaux de gènes ne permettent pas de parler d'arbre du vivant. Cependant, son hypothèse est intéressante car elle est la plus explicite sur un certain nombre de points. D'abord, on peut analyser la pertinence de sa métaphore. La diminution de la température physique (qui peut être définie comme une agitation thermique moléculaire ou comme une mesure de l'entropie en thermodynamique) est liée à la diffusion de cette chaleur à un environnement dont la température est plus réduite. Quelle équivalence peut-il exister entre cette métaphore et l'échange de gènes entre organismes ? Si l'on accepte que les échanges génétiques aient été plus importants à l'époque de ces organismes (au point que les organismes n'ont pas de généalogie), quelle a été la raison de la diminution graduelle de ces transferts ?

Woese répond par l'hypothèse de complexité. Les gènes qui codent les parties d'une structure avec plusieurs composantes sont d'autant moins fréquemment transférés entre organismes¹¹ que la structure a beaucoup d'éléments. Cette hypothèse se justifie par le fait qu'un élément d'une grande machinerie trouvera difficilement une fonction dans l'organisme récepteur si le reste de la machinerie ne l'accompagne pas [Jain *et al.* 1999 ; Cohen *et al.* 2010]. Woese suppose donc que les premiers organismes avaient des structures avec peu d'éléments peu intégrés, ce qui aurait permis de nombreux échanges. L'augmentation progressive de l'intégration et du nombre d'éléments qui formaient les structures biologiques aurait réduit les possibilités de transfert efficace de gènes.

Cependant, il semble légitime de se demander comment il est possible que des éléments aient pu former des structures de plus en plus intégrées si la généalogie n'existait pas et si chaque élément était continuellement et indépendamment transféré, presque comme si le lot génétique d'un progénote était tiré au hasard. Woese ne traite pas cette question mais certains auteurs ont proposé des réponses dans d'autres contextes qui pourraient s'appliquer par extrapolation à cette hypothèse. Si l'on accepte que les éléments dont on parle étaient autoréplicatifs, ceux qui se répliquaient le plus vite auraient été plus nombreux que les autres et auraient donc été plus transférés, ce qui a conduit parfois à les comparer à des virus [Koonin et Martin 2005 ; Koonin *et al.* 2006]. Mais si *chaque élément* ne pouvait compter que sur ses propres capacités d'autoréplication, comment est-il possible que des structures avec de nombreux éléments coopératifs aient pu se mettre en place ? Si l'on accepte que ces éléments étaient liés d'une quelconque manière, on pourrait imaginer une co-évolution mais comment assurer cette liaison si chacun était transféré indépendamment ? Or, si l'on invoque le transfert simultané de tous les éléments participant ensemble dans un processus, cela équivaut à une certaine généalogie. Une autre critique qu'il est possible de formuler à l'encontre de cette métaphore est que la comparaison entre température et transfert horizontal ne permet pas de déterminer la relation qui existerait entre l'hypothèse de complexité et les « paliers » représentant les ancêtres des grands taxons.

En dehors du fait qu'elle se base sur une métaphore présentant peu de fondement biologique, l'hypothèse de Woese souffre d'une tendance très répandue consistant à observer

¹¹ Bien évidemment on parle de transferts effectifs, i. e. ceux pour lequel un transfert d'information génétique a eu lieu entre organismes et que l'organisme récepteur a conservé. On ne tient pas compte des cas où du matériel génétique a pu arriver dans une cellule sans être adopté, car cela ne laisse pas de trace génomique.

le cenancêtre *par rapport* aux origines de la vie au lieu de le faire par rapport aux organismes modernes qui permettent de le définir. Ainsi, lorsque Woese exclue la possibilité d'un cenancêtre « complexe », i. e. contenant des unités avec de nombreux composants qui interagissent ensemble sur la base du « manque de temps » (moins d'un milliards d'années) depuis l'origine de la Terre pour générer un tel organisme, il applique abusivement un référentiel sur les origines de la vie qui n'a rien à voir avec l'inférence d'un ancêtre commun à partir des organismes modernes. Cela constitue un exemple de confusion entre origines de la vie et cenancêtre, ou plus exactement entre progénote et cenancêtre. En effet, l'hypothèse des pré-cellules fournit un modèle plus adapté à l'apparition des premières structures vivantes car l'origine de la vie a été vraisemblablement liée à la combinaison de composantes précédemment indépendantes.

Une thématique subsidiaire à celle des grandes différences qui existent entre les trois domaines du vivant est celle de l'explication de grandes branches vides dans les arbres phylogénétiques basés par exemple sur l'ADN codant l'ARNr 16S/18S. Pourquoi n'y a-t-il pas des branches latérales entre le cenancêtre et les derniers ancêtres respectifs des domaines du vivant ? Le scénario des communautés d'organismes justifie cela par le transfert permanent de gènes, qui élimine la généalogie jusqu'à l'apparition des lignées modernes, mais il a déjà été expliqué pourquoi cette hypothèse n'est pas valable. D'autres hypothèses ont été proposées.

L'une postule qu'à l'époque des ancêtres des domaines du vivant (ici, des archées et des bactéries) il y avait une grande diversité d'organismes, dont certains avec des caractéristiques intermédiaires entre les bactéries et les archées, mais qu'un événement brutal aurait éradiqué les autres formes de vie. Aujourd'hui on n'observerait que les descendants des deux lignées survivantes. La cause brutale évoquée pour rendre compte de cette disparition est généralement la chute d'un météorite de très grande taille qui aurait provoqué un réchauffement soudain de la température. Et la raison proposée pour expliquer que les ancêtres des archées et les bactéries aient survécu, mais pas leurs contemporains, serait que les ancêtres bactérien et archéen auraient été des hyperthermophiles vivant dans des sources hydrothermales qui soit les auraient gardés à l'abri, soit auraient constitué une préadaptation pour l'augmentation globale de température [Gogarten-Boekels *et al.* 1995 ; Nisbet et Sleep 2001 ; Boussau *et al.* 2008]. Cette possibilité a été critiquée en partie car dans le cas d'un tel impact toute la Terre aurait été affectée de la même manière, ce qui est très improbable [Miller et Lazcano 1995]. D'autres hypothèses plus fantaisistes ont été proposées pour fonder

un supposé goulet d'étranglement qui aurait marqué la différence entre les domaines du vivant. Selon l'une d'elles, la vie serait d'abord apparue sur Mars et les trois domaines du vivant seraient des échantillons d'un ou plusieurs transports météoritiques entre Mars et la Terre [Nisbet et Sleep 2001 ; Line 2002 ; Boussau *et al.* 2008]. Ces hypothèses rappellent respectivement l'opposition entre gradualisme et catastrophisme [Gould 1989] et la conjecture sur la panspermie [Oparin 1938], mais je ne vais pas détailler ici les arguments qui les invalident car il s'agit de controverses qui vont bien au-delà de mon travail et ces propositions sont rarement considérées dans la communauté qui étudie le cenancêtre.

En revanche, les principes fondamentaux de l'évolution darwinienne permettent de proposer une explication à ces différences entre domaines et aux longues branches basales vides. Pour commencer, le fait que le cenancêtre n'ait pas été une communauté non généalogique d'organismes ne veut pas dire que toutes les informations présentes dans les organismes modernes proviennent de la lignée issue d'un organisme unique. Les structures à plusieurs composants interdépendants dont la présence est inférée chez le cenancêtre impliquent l'existence d'une évolution préalable, et par conséquent la compétition entre les ancêtres du cenancêtre et d'autres organismes. Ainsi, autant le cenancêtre que ses descendants ont dû côtoyer des organismes d'autres lignées pendant des périodes plus ou moins longues avant que les représentants des autres lignées se soient éteints. L'idée que l'extinction soit une composante majeure de la macroévolution (l'évolution à l'échelle des temps géologiques) n'est pas une hypothèse aussi forte que l'on pourrait le penser au premier abord [Philippe 2003].

Premièrement, le registre fossile montre souvent le remplacement d'une large diversité de groupes apparentés par une nouvelle diversité issue d'une autre lignée, et ce sans forcément faire intervenir des événements catastrophiques, mais tout simplement par l'effet de nouvelles adaptations ou de phénomènes stochastiques [Eldredge et Gould 1972]. Deuxièmement, il a été proposé que les longues branches vides puissent être interprétées d'une façon similaire aux arbres de coalescence, qui sont très liés à l'extinction de lignées anciennes. La coalescence est un modèle de génétique des populations (par conséquent plus adapté à des événements évolutifs récents) dont l'objectif est de suivre l'évolution de plusieurs allèles d'un gène jusqu'à leur dernier ancêtre commun. L'un de ses enjeux est de déterminer le nombre de générations qu'il est nécessaire de remonter pour converger vers un seul gène ancestral. Sans entrer dans les détails, des calculs de probabilités montrent que lorsque le nombre d'allèles est important (i. e. la diversité génétique est importante),

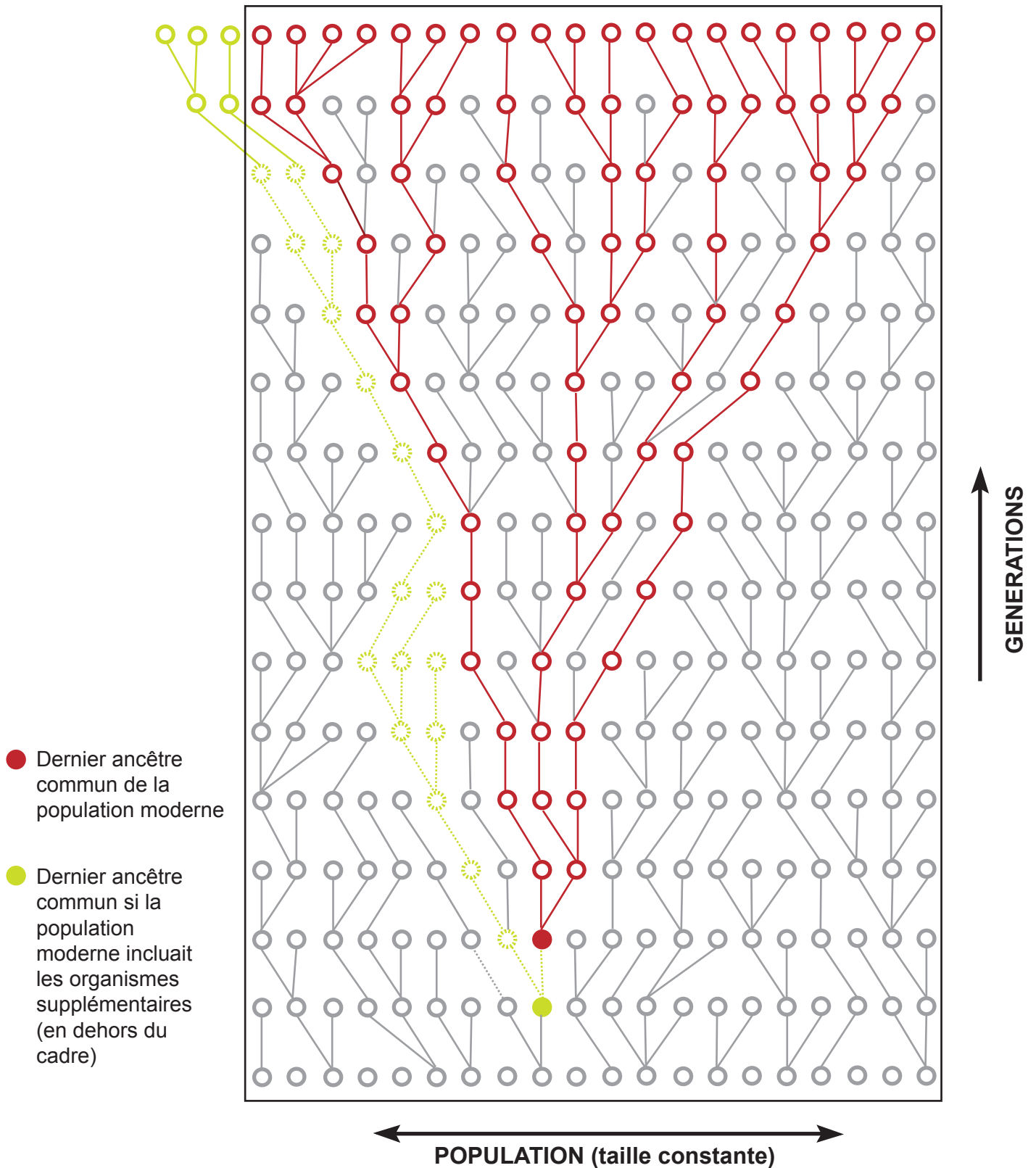


Figure 7. Représentation de la coalescence.

Les lignes correspondent à des générations, les plus anciennes étant en bas et les plus récentes en haut. Les cercles représentent les gènes se trouvant dans des organismes différents d'une même population et les lignes représentent la parenté entre ces gènes. Dans le texte, l'évolution des gènes est extrapolée à l'évolution des organismes.

beaucoup d'allèles convergent dans les premières générations mais qu'ensuite il faut de nombreuses générations (un peu plus de la moitié des générations totales entre les allèles de départ et le dernier ancêtre commun) pour que les dernières coalescent définitivement [Felsenstein 2003]. Ce résultat n'est pas étonnant dans la mesure où, pour une population constante, la quantité d'organismes apparentés dans la génération précédente est de plus en plus faible (Figure 7). Par ailleurs, on observe que, dans une population constante, beaucoup d'organismes anciens n'ont pas laissé de descendants, i. e. leurs lignées se sont éteintes et ont été remplacées par d'autres qui sont arrivées jusqu'au temps final (Figure 7). Bien que l'extrapolation des modèles de génétique de populations à des modèles de macroévolution ne puisse pas être formelle, il a été proposé que les longues branches vides des arbres phylogénétiques construits avec les ADN codant les ARN 16S/18S et les ATPases puissent se justifier, par analogie avec les arbres de coalescence, par le fait que les branches latérales n'aient pas donné de descendants modernes [Zhaxybayeva et Gogarten 2004]. Finalement, l'extinction des lignées qui ont cohabité avec le cenancêtre dans des périodes postérieures est une nécessité théorique, car si elles ne s'étaient pas éteintes et que leurs descendants se trouvaient parmi les organismes actuels, cela ne ferait que reculer à une époque antérieure le cenancêtre, sans rien modifier au raisonnement de base (Figure 7).

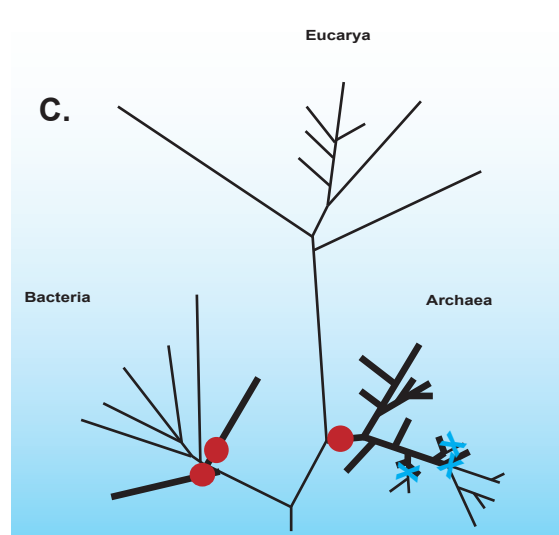
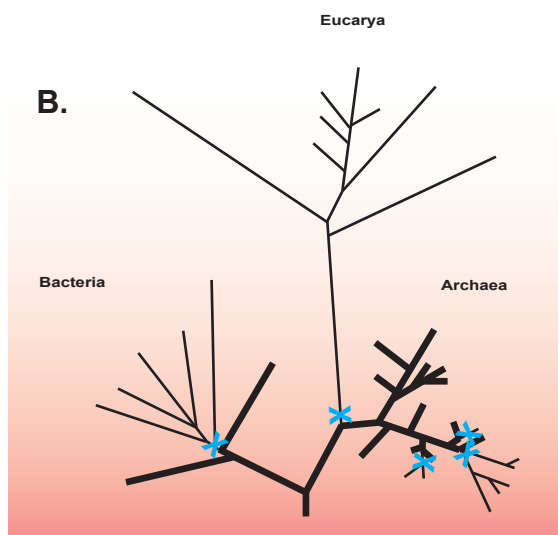
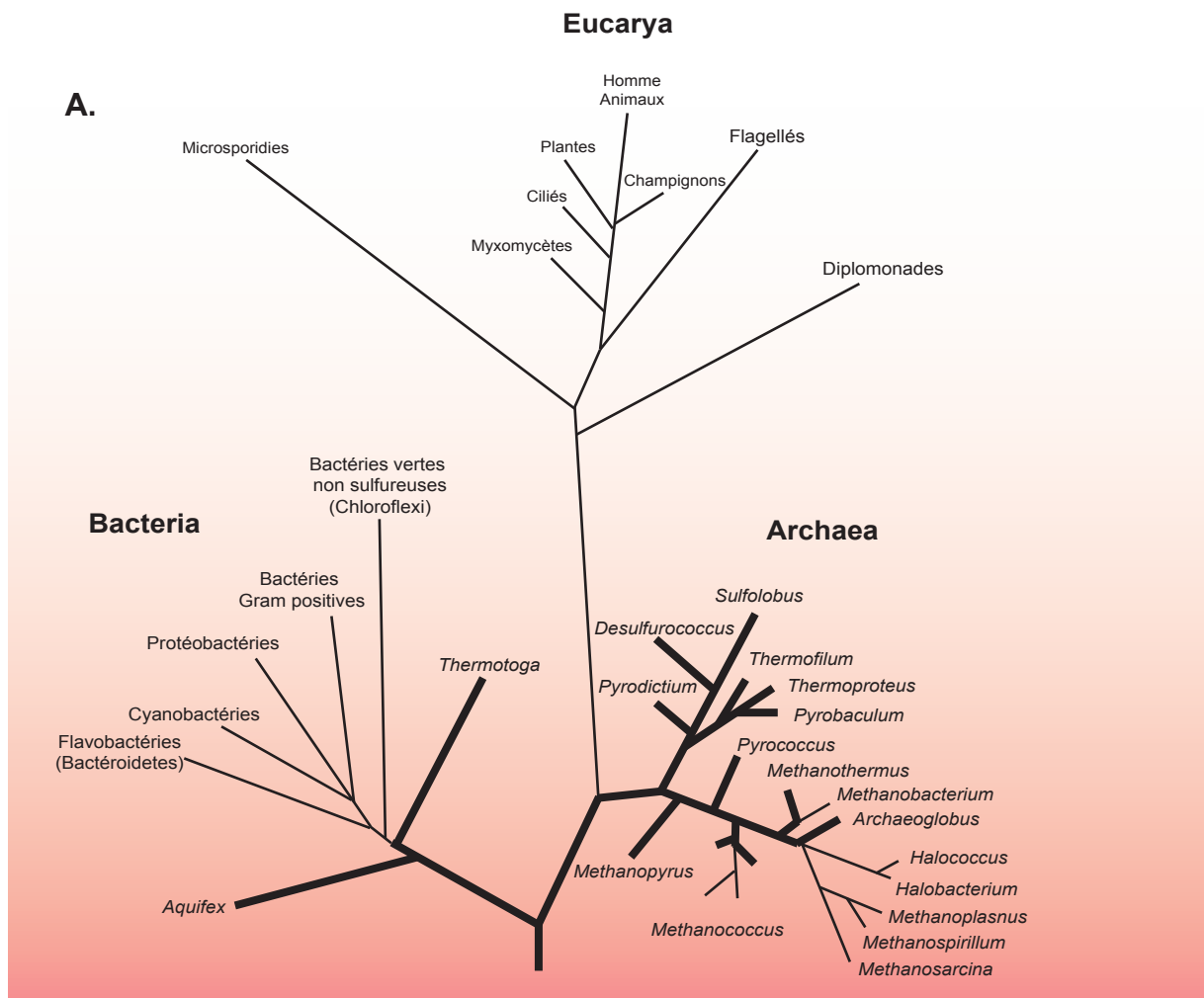
En revanche, le fait que les lignées contemporaines du cenancêtre aient disparu n'empêche pas que certains de leurs gènes se retrouvent actuellement dans les lignées modernes par l'effet de transferts horizontaux de gènes. Ces transferts étant postérieurs au cenancêtre, les séquences devraient se trouver répandues uniquement dans les organismes descendants de l'organisme récepteur, sauf si des transferts très nombreux ont répandu ces caractéristiques dans les lignées modernes. Dans cette dernière hypothèse, la principale difficulté est de construire une phylogénie fiable en tenant compte des transferts horizontaux, comme c'est le cas à chaque fois que ces événements sont fréquents.

Interprétations de l'arbre : cenancêtre et température

Depuis l'obtention des premiers arbres moléculaires du vivant, on a essayé de corrélérer ces résultats avec l'information disponible sur les conditions de vie sur la Terre ancienne. A l'inverse, on a essayé d'inférer la présence de certaines propriétés de l'environnement des organismes anciens à partir de la phylogénie des organismes. Ainsi, les archées portent un nom qui fait référence à l'hypothèse selon laquelle leurs phénotypes sont des restes anciens représentatifs de l'environnement de la Terre primitive. D'abord le nom d'archaebactéries

leur a été donné parce que la méthanogenèse qui caractérisait les organismes connus à l'époque semblait cohérente avec l'environnement supposé de la Terre primitive [Woese et Fox 1977b]. Ensuite, le nom d'Archaea devait marquer le caractère apparemment plus ancien de ces organismes, notamment parce que ce domaine semblait contenir beaucoup d'organismes vivant à haute température, ce qui rappelait l'idée que la vie est apparue dans des milieux très chauds [Oyaizu *et al.* 1987 ; Woese 1987 ; Woese *et al.* 1990 ; Olsen *et al.* 1994]. Des travaux phylogénomiques plus récents signalent qu'en vérité la méthanogénèse est restreinte aux archées et qu'elle n'est pas un caractère ancestral dans ce domaine du vivant mais un caractère dérivé apparu tardivement dans les euryarchées [Gribaldo et Brochier-Armanet 2006 ; Brochier-Armanet *et al.* 2011b]. Entretemps, c'est la température à laquelle vivait le cenancêtre et, par extension, celle à laquelle la vie est apparue, qui est devenue le centre d'un grand nombre de débats.

Il est difficile de déterminer depuis quand existe l'idée que la vie soit apparue à (très) haute température, une hypothèse que l'on appelle parfois « Eden hyperthermophile ». L'information disponible sur la température de la Terre ancienne est faible et a laissé place à beaucoup de spéculations [Miller et Lazcano 1995]. L'une d'entre elles est l'idée selon laquelle, lors de la formation de la Terre, la surface de notre planète était fondue et que son refroidissement aurait été ralenti par l'impact météoritique qui était plus fréquent dans cette période ancienne de l'histoire du système solaire. Selon cette hypothèse, la surface terrestre aurait fini par se refroidir progressivement mais comme les premiers organismes sont apparus très tôt, ils devaient probablement être compatibles avec de très hautes températures [Pace 1991]. Malheureusement, on ne connaît aujourd'hui aucune roche du tout début de la Terre qui pourrait nous permettre de fournir un argument géologique à l'appui de cette hypothèse. Les échantillons les plus anciens qui pourraient nous renseigner sur la température ancienne sont des roches sédimentaires d'environ 3,5 milliards d'années (un milliard d'années environ après la formation de la Terre, [Gargaud *et al.* 2006]). Ces roches montrent un rapport isotopique $^{18}\text{O}/^{16}\text{O}$ qui augmente depuis lors, ce qui a été interprété comme l'évidence que la température de la surface terrestre dans laquelle ces roches se sont formées aurait diminué au cours du temps [Knauth et Lowe 2003 ; Robert et Chaussidon 2006]. Cependant, ces interprétations sont polémiques car d'autres explications, comme le pH, la salinité ou la profondeur des océans ont été proposées pour expliquer ces mêmes ratios [Kasting *et al.* 2006 ; Jaffres *et al.* 2007].



- X Transitions d'un milieu hyperthermophile à un milieu plus froid
- Transitions d'un milieu plus froid à un milieu hyperthermophile

Figure 8. Température environnementale à l'époque du cenancêtre.

A. D'après Stetter 1996. Les traits épais correspondent aux organismes hyperthermophiles. **B.** Phylogénie de Stetter si l'on accepte que le cenancêtre était hyperthermophile. **C.** Phylogénie de Stetter si l'on accepte que le cenancêtre n'était pas hyperthermophile.

Dans ce contexte, le point de savoir si le cenancêtre vivait ou pas à très haute température est devenu un enjeu majeur dans le débat sur la température ancienne de la Terre, notamment au moment de l'apparition de la vie. Woese privilégiait déjà l'idée d'un cenancêtre hyperthermophile dont auraient dérivé d'abord les archées (hyperthermophiles) et ensuite les bactéries, suivant un gradient de température qui serait corrélé à la diminution progressive de la température terrestre [Woese 1987]. Les organismes hyperthermophiles étaient dans ce cas considérés comme primitifs et ayant conservé des niches écologiques similaires à celle du cenancêtre [Pace 1991]. Cette idée semblait soutenue par les premières phylogénies des organismes basées sur les séquences d'ADN codant l'ARN 16S/18S d'organismes des trois domaines du vivant, puisqu'elles plaçaient les espèces hyperthermophiles bactériennes et archéennes à la base de leurs domaines respectifs (Figure 8, [Stetter 1996]).

Avant de revenir sur la question de la température, il semble important de rappeler qu'il est erroné d'inférer les caractéristiques d'un ancêtre à partir des capacités de la première lignée basale à diverger dans les phylogénies. Ces lignées basales constituent ce que l'on entend parfois appeler *fossiles vivants*, en supposant la ressemblance de certaines caractéristiques d'un organisme basal avec celles que l'on peut observer (ou que l'on imagine avoir été présentes) chez un fossile. La notion de fossile vivant est fautive car toutes les lignées ont des caractéristiques propres et les organismes basaux peuvent accumuler autant que les autres de nombreuses modifications par rapport à l'état ancestral, même si elles peuvent se concentrer sur des caractères différents. Autrement dit, si l'on détecte chez un organisme moderne un caractère déjà présent chez un ancêtre, cela ne permet pas de dire que toutes ses autres caractéristiques seront aussi ancestrales, même s'il diverge en premier dans la phylogénie de son groupe taxonomique.

Pour déterminer les caractéristiques d'un ancêtre commun à un groupe il faut déterminer l'homologie des caractères, comparer l'ensemble des organismes qui en descendent et comparer avec un groupe extérieur [Philippe et Laurent 1998 ; Philippe 2003]. Ainsi, si l'on acceptait que la phylogénie de Stetter était correcte (Figure 8), la raison pour postuler l'hyperthermophilie du cenancêtre ne serait pas que les lignées hyperthermophiles sont les premières à diverger dans ces phylogénies à l'intérieur de leurs domaines respectifs, mais plutôt que ce postulat impliquerait moins de changements évolutifs pour cette topologie (Figure 8), i. e. il s'agit de l'hypothèse la plus parcimonieuse. Toutefois, cette hypothèse dépend de la topologie de l'arbre étudié. Certains travaux ont donc débattu de la température

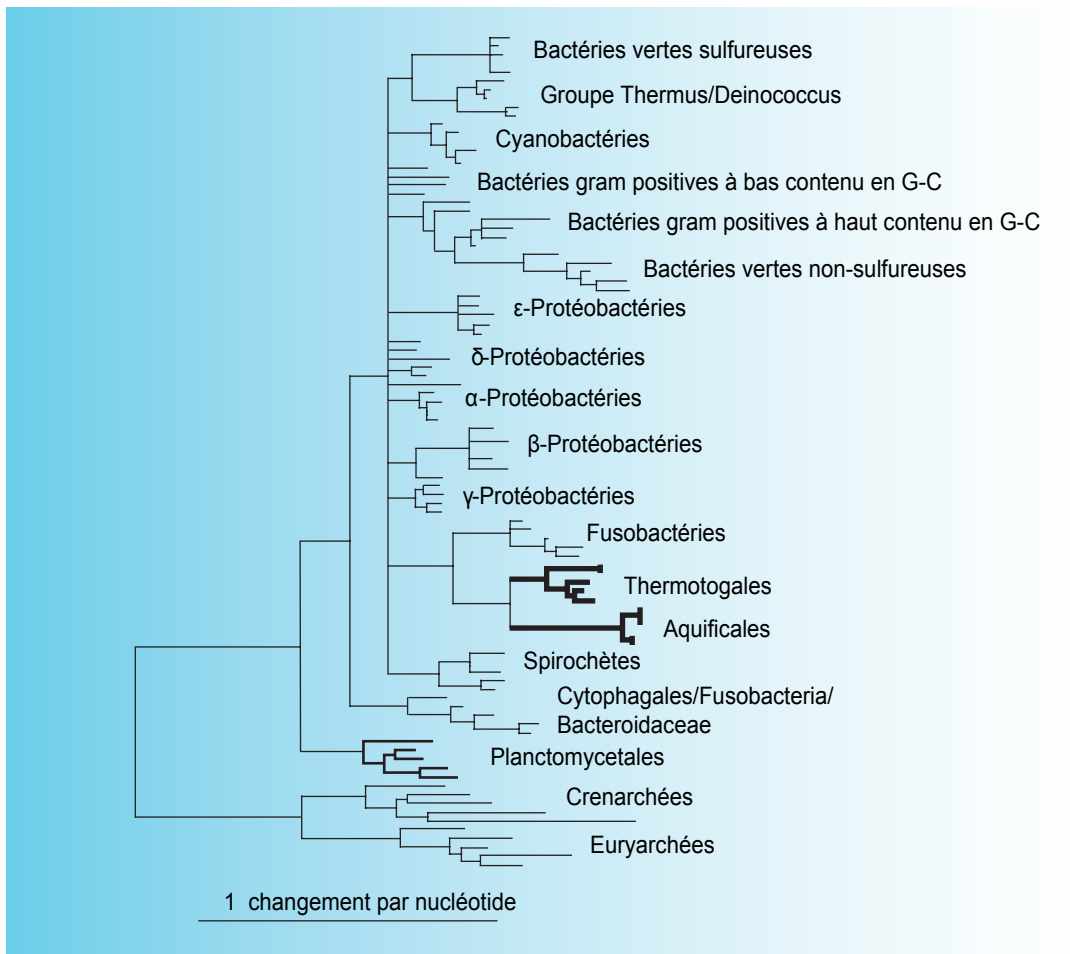


Figure 9. Phylogénie montrant que les bactéries hyperthermophiles ne branchent pas à la base de leur domaine du vivant.

D'après Brochier et Philippe 2002. Cet arbre a été construit à partir des positions évoluant le moins rapidement dans les gènes codant l'ARN 16S. Les branches épaisses correspondent à des organismes hyperthermophiles.

du cenancêtre en s'interrogeant sur la position des hyperthermophiles dans l'arbre. Notamment, l'utilisation de méthodes moins sensibles à l'accumulation de bruit dans les séquences pour reconstruire une phylogénie des gènes codant l'ARNr 16S et des méthodes de superarbres ont mis en cause la position basale des phyla bactériens hyperthermophiles Aquificales et Thermotogales (Figure 9, [Brochier et Philippe 2002 ; Daubin *et al.* 2002]). Ce résultat impliquerait que l'hypothèse la plus parcimonieuse serait l'adoption tardive de l'hyperthermophilie dans ces phyla bactériens. Cependant, la discussion sur ce sujet est difficile car, comme on le verra dans les sections suivantes, la phylogénie des organismes est encore loin de faire l'objet d'un consensus.

Par ailleurs, l'utilisation de l'habitat des organismes comme caractère n'est pas appropriée pour l'analyse phylogénétique car elle ne tient pas compte des mécanismes qui sous-tendent l'adaptation à une niche donnée, i. e. l'homologie des caractères n'est pas avérée et leur similarité pourrait être le résultat d'une convergence récente. C'est ce que l'étude de l'évolution de la reverse gyrase suggère : la reverse gyrase est une protéine qui introduit des supertours positifs, i. e. des surenroulements topologiques formés par la coupure, retournement et fusion, dans des molécules circulaires d'ADN fermées comme le chromosome d'un grand nombre d'archées et de bactéries. La présence de reverse gyrase semblait dans un premier temps spécifique des organismes hyperthermophiles [Forterre 1996 ; Forterre 2002], ce qui suggérait qu'elle était indispensable à la vie dans ces conditions (depuis, on l'a aussi retrouvée dans quelques bactéries thermophiles [Brochier-Armanet et Forterre 2006]). Elle est formée par deux domaines protéiques, une topoisomérase et une hélicase, qui sont fusionnés dans la reverse gyrase mais qui apparaissent indépendamment dans les organismes qui ne sont pas hyperthermophiles. D'abord, le fait que les deux domaines protéiques isolés soient répandus dans les trois domaines du vivant, ce qui n'est pas le cas de la reverse gyrase, a été interprété comme suggérant que le cenancêtre ne devait pas encore avoir de reverse gyrase et, connaissant la dépendance des organismes hyperthermophiles envers cette protéine, on concluait que le cenancêtre ne pouvait pas être hyperthermophile [Forterre 1996]. Ensuite, la phylogénie des reverse gyrases semblait indiquer que les phyla bactériens (hyper)thermophiles avaient acquis cette protéine par des transferts horizontaux indépendants [Forterre *et al.* 2000]. Ce résultat soutenait l'adaptation tardive de ces lignées à l'(hyper)thermophilie mais des nouvelles analyses phylogénétiques donnent des résultats contradictoires et peu soutenus [Brochier-Armanet et Forterre 2006].

Des observations similaires ont été proposées pour d'autres enzymes qui semblent importantes pour la vie à haute température, par exemple les protéine-disulfide oxydoréductases qui aident à la stabilisation de protéines en formant des ponts disulfure. Ces enzymes sont formées de deux domaines issus d'une duplication et fusionnés ensemble qui existent dans d'autres enzymes. Il a été proposé que cette fusion se soit produite dans la lignée crenarchéenne avant d'être transférée horizontalement aux euryarchées et bactéries thermophiles et hyperthermophiles qui les portent [Becerra *et al.* 2007b]. Ces propositions sont intéressantes car elles explorent les mécanismes qui se cachent derrière l'hyperthermophilie des organismes modernes et permettent ainsi de comparer des caractères homologues sur lesquels la reconstruction phylogénétique est possible. Mais ces arguments ne peuvent pas encore être définitifs dans la mesure où il est toujours possible d'imaginer un cenancêtre hyperthermophile sans ces protéines. En fin de compte, la principale information que l'on retire de ces observations est que si le cenancêtre était hyperthermophile, alors ses adaptations à haute température étaient très différentes de celles des organismes modernes.

La question de l'(hyper)thermophilie des premiers êtres vivants a aussi été traitée par d'autres approches. On a émis l'idée que l'ARN était trop sensible à la température pour que le monde ARN ait pu exister à très haute température [Forterre 1995 ; Miller et Lazcano 1995], mais le fait que les organismes modernes hyperthermophiles actuels utilisent des ARN, comme tous les organismes, fragilise cet argument. D'un autre côté, on a émis l'hypothèse que le biais lié à la dégénérescence du code génétique reflète une mise en place à des hautes températures [Di Giulio 2000a] mais ce même biais a été expliqué par d'autres facteurs comme la disponibilité des acides aminés à l'époque de sa mise en place [Fournier et Gogarten 2010].

Des méthodes similaires ont été appliquées pour essayer de déterminer s'il existe dans les séquences nucléiques et protéiques actuelles des traces de l'adaptation à la température ancienne. Dans le cas des acides nucléiques, on sait que les organismes (hyper)thermophiles ont des pourcentages G-C plus riches, ce qui confère plus de stabilité à leur génome malgré leur habitat à haute température. Le même type de biais de composition existe aussi pour les protéines. Ainsi, beaucoup de travaux ont essayé d'inférer la séquence ancestrale de certains gènes ou protéines universels et de déterminer si ces séquences sont plutôt compatibles avec un cenancêtre mésophile, thermophile ou hyperthermophile. Malheureusement, les résultats ne sont pas cohérents entre eux. Certaines analyses montrent que les séquences d'ARNr que l'on peut inférer chez le cenancêtre correspondraient à un organisme mésophile [Galtier *et al.*

1999 ; Galtier 2001] et d'autres qu'il serait (hyper)thermophile [Di Giulio 2000b]. Du côté des protéines, les séquences ancestrales postulent un cenancêtre thermophile mais probablement pas hyperthermophile [Di Giulio 2001 ; Brooks *et al.* 2004 ; Gaucher *et al.* 2008]. La combinaison d'une analyse sur des ARNr et des protéines a donné un cenancêtre mésophile mais des ancêtres respectifs des archées et des bactéries qui seraient thermophiles [Boussau *et al.* 2008]. Ce dernier résultat a été interprété dans des termes similaires à l'hypothèse selon laquelle les ancêtres des organismes modernes seraient les survivants d'un cataclysme provoquant l'augmentation globale de la température, par exemple l'arrivée d'un météorite ou un transfert extraterrestre [Gogarten-Boekels *et al.* 1995 ; Nisbet et Sleep 2001].

Enfin, une autre méthode utilisée a été la « ressuscitation » de protéines, i. e. l'expression artificielle dans des cellules actuelles de protéines avec une séquence « ancestrale » préalablement inférée de la comparaison des séquences modernes. Cette technique permet de tester les propriétés thermiques de ces séquences. Certains de ces travaux se sont concentrés uniquement sur la modification ponctuelle de protéines déjà hyperthermophiles pour introduire des résidus considérés comme importants dans les séquences ancestrales [Miyazaki *et al.* 2001 ; Iwabata *et al.* 2005 ; Shimizu *et al.* 2007]. Mais le fait que les séquences résultantes fonctionnent dans des conditions hyperthermophiles semble peu significatif étant donné que les séquences utilisées étaient déjà hyperthermophiles. La ressuscitation totale de facteurs d'élongation ancestraux soutient, de son côté, la thermophilie de ces séquences [Gaucher *et al.* 2008]. Cependant, les méthodes de reconstruction des séquences ont été mises en cause car elles contiennent des biais intrinsèques qui infèrent des protéines plus stables que la réalité [Williams *et al.* 2006], ce qui ne permet pas de trancher définitivement la question de la température de vie du cenancêtre.

En conclusion, il me semble important de signaler que même si la plupart de ces travaux suivent une démarche parfaitement légitime en comparant les séquences des organismes actuels pour inférer les séquences ancestrales présentes chez le cenancêtre, leur motivation est manifestement de déterminer si la vie est apparue à haute température ou pas (par exemple, [Pace 1991 ; Olsen *et al.* 1994 ; Galtier *et al.* 1999 ; Brooks *et al.* 2004]). Ce rapprochement entre les origines de la vie et le cenancêtre en liaison avec l'étude de la température de la Terre ancienne est contestable et maintient la confusion entre les deux [Dagan *et al.* 2010]. Par ailleurs, même si ces études sont très intéressantes, elles nous renseignent finalement assez peu sur les caractéristiques du cenancêtre car, comme nous l'avons vu, elles s'intéressent au milieu dans lequel cet organisme aurait pu vivre sans s'intéresser à sa

biologie, comme le montre le fait qu'aucun travail ne permette de proposer quelles adaptations auraient permis au cenancêtre de vivre à haute température.

4. Vacillations de l'arbre du vivant

Nous avons vu dans la section précédente que la question de l'habitat du cenancêtre, qui est considérée comme une question majeure par beaucoup de chercheurs, est fondée en grande partie sur la topologie de la phylogénie des organismes telle que proposée par Woese (1990). Cependant, la topologie de l'arbre du vivant a été revue un grand nombre de fois depuis sa formulation et aujourd'hui elle semble très incertaine, notamment sur trois sujets qui vont être traités respectivement dans les trois sections suivantes : la place de la racine, l'importance des transferts horizontaux de gènes dans l'histoire des organismes et l'origine des eucaryotes.

La racine

Retrouver la place de la racine dans l'arbre est indispensable pour l'orienter et déterminer les lignées qui partagent un lien plus fort entre elles. En général, l'enracinement d'un arbre se fait avec un groupe extérieur, mais dans le cas de l'arbre du vivant cela est impossible car aucun groupe extérieur n'existe. A la suite de la suggestion de Woese (1987) d'appliquer la méthode d'enracinement avec des gènes paralogues [Schwartz et Dayhoff 1978], des racines ont été proposées d'abord en utilisant des couples de paralogues de facteurs d'élongation [Iwabe *et al.* 1989], des sous-unités des ATPases [Gogarten *et al.* 1989 ; Iwabe *et al.* 1989], deux couples d'ARNt synthétases [Brown et Doolittle 1995 ; Brown *et al.* 1997], des carbamoylphosphate synthetases [Lawson *et al.* 1996], des enzymes HisA et HisF [Charlebois *et al.* 1997] ou encore de la *Signal Recognition Particle* (SRP) et son récepteur [Gribaldo et Cammarano 1998].

Des critiques pouvaient être faites contre certaines de ces phylogénies, à savoir qu'elles n'étaient pas toujours représentatives de la diversité des organismes, qu'elles contenaient souvent des transferts horizontaux, qu'elles étaient basées sur des alignements trop courts ou avec très peu de signal. Malgré cela, à l'exception de celle de Charlebois *et al.*, elles coïncidaient toutes pour placer les bactéries comme premier groupe à diverger depuis le cenancêtre et les archées et les eucaryotes comme groupe frère, ce que Woese avait reporté sur son arbre des organismes (Figure 6, [Woese *et al.* 1990]). L'arbre des trois domaines a connu dès le départ des critiques, notamment de la part de ceux qui continuent de penser que les caractéristiques morphologiques (ou cytologiques) devraient être les principaux arguments

de la classification des organismes [Margulis 1996 ; Mayr 1998 ; Cavalier-Smith 2002]. Cependant la division en trois domaines du vivant est largement acceptée, en particulier parmi les microbiologistes, à cause du fort pouvoir explicatif de cette classification parmi les organismes unicellulaires, spécialement chez les bactéries et les archées.

Des doutes ont commencé à surgir, cependant, lorsque l'accumulation progressive de données de séquençage a révélé que plus on accumulait de phylogénies de gènes uniques, plus on semblait trouver des incohérences entre les phylogénies de ces gènes, mettant en cause leur véracité et leur pertinence pour comprendre la phylogénie des organismes [Syvanen 1985 ; Brown et Doolittle 1997]. Ces incohérences avaient parfois des explications biologiques, comme les phénomènes de paralogie cachée et de transfert horizontal de gènes, mais elles pouvaient aussi être dues à des artefacts importants dans les méthodes de reconstruction des phylogénies [Philippe et Laurent 1998]. En effet, certains artefacts liés à la *saturation mutationnelle* des séquences¹², à la différence de vitesse d'évolution entre sites et lignées¹³ et aux biais de composition des séquences (par exemple chez les thermophiles : voir section précédente), avaient été signalés beaucoup plus tôt chez des groupes plus restreints d'organismes [Fitch et Margoliash 1967a ; Fitch et Margoliash 1967b ; Fitch 1971 ; Felsenstein 1978 ; Meyer *et al.* 1986] mais ils avaient reçu peu d'attention lors de l'enracinement de l'arbre du vivant.

Lorsque l'accumulation des premiers séquençages de génomes a permis de reconstruire à nouveau les phylogénies de ces mêmes paralogues (sauf pour les HisA/HisF et carbamoyl transférases) avec un échantillonnage plus large, on a observé que les topologies obtenues étaient à la fois beaucoup plus confuses que ne le laissait croire l'arbre classique des trois domaines et contradictoires entre elles [Philippe et Forterre 1999]. La même analyse signalait que la saturation mutationnelle des séquences dans les six alignements était très importante, au point que la majorité du signal phylogénétique qui permettait de reconstruire les arbres était perdue. La question se posait alors de savoir comment il était possible que les premières phylogénies coïncidassent. Tout semblait indiquer que ces phylogénies étaient soumises à des

¹² La saturation mutationnelle des séquences se produit lorsque l'accumulation de mutations sur une même position ne permet plus de différencier les caractères convergents (homoplasiques) des dérivés (synapomorphies), conduisant ainsi à l'effacement du signal phylogénétique pour cette position.

¹³ On entend vitesse d'évolution dans le sens de quantités de mutations fixées dans une séquence par rapport à la séquence ancestrale.

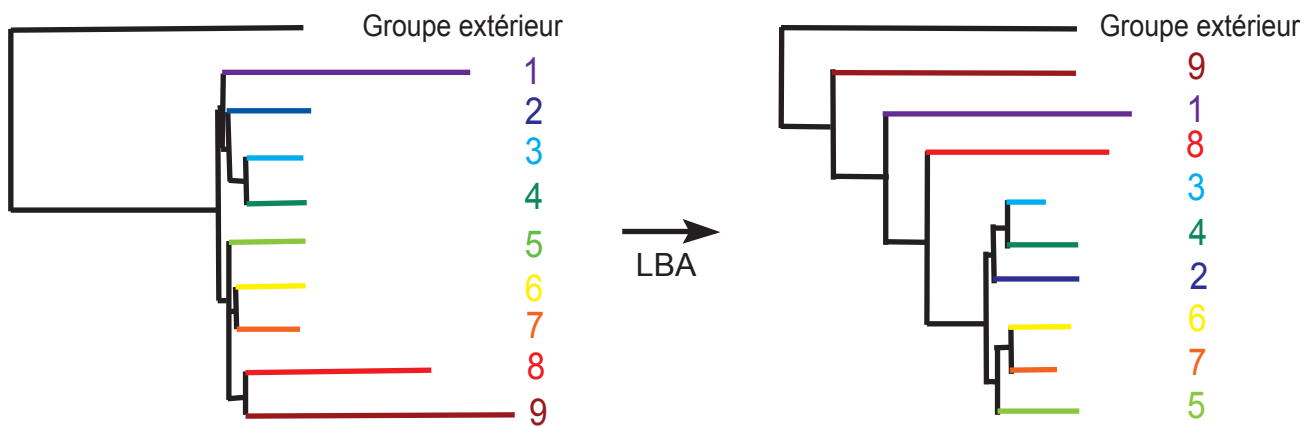


Figure 10. Attraction des longues branches (LBA).

La phylogénie à gauche représente une phylogénie théorique de gènes « vraie ». La phylogénie à droite montre comment l'attraction des branches longues par le groupe extérieur peut produire une topologie erronée.

artefacts d'*attraction des longues branches* [Philippe et Forterre 1999]. L'attraction des longues branches se produit lorsque la vitesse d'évolution entre les séquences est très inégale, quel que soit le rapport phylogénétique entre elles. Dans ce cas, les branches courtes groupent ensemble non pas parce qu'elles sont phylogénétiquement plus proches mais parce qu'elles ont moins divergé entre elles. Ainsi, la branche associée à une vitesse d'évolution rapide se rapproche du groupe extérieur et ce d'autant plus qu'il sera divergent (Figure 10, [Felsenstein 1978]).

Dans le cas des paralogues utilisés, la grande divergence entre les séquences bactériennes d'une part et les archéennes-eucaryotes d'autre part a été proposée comme expliquant l'attraction des séquences bactériennes par le groupe extérieur, en l'occurrence l'autre paraglogue. Lorsque des phylogénies des SRP et des facteurs d'élongation étaient réalisées uniquement en prenant les positions qui évoluent le moins vite (les moins saturées), on retrouvait les trois domaines mais c'étaient cette fois-ci les eucaryotes et non plus les bactéries qui divergeaient en premier du cenancêtre [Philippe et Forterre 1999]. Afin de prendre en compte ces biais, certains travaux ont commencé à ne tenir compte pour la reconstruction de phylogénies anciennes que des positions qui varient peu [Philippe *et al.* 2000 ; Brochier et Philippe 2002]. La méthode des covariations a été développée¹⁴ et appliquée aux facteurs d'élongation et aux SRP [Brinkmann et Philippe 1999 ; López *et al.* 1999]. Enfin, il a été postulé que l'utilisation d'événements rares permettant de déterminer des *synapomorphies* (caractères dérivés partagés) pouvait aider dans la détermination de ces relations [Philippe *et al.* 2000].

Les résultats utilisant les covariations s'accordaient pour montrer d'une façon plus ou moins bien soutenue que les eucaryotes étaient le groupe ayant divergé du cenancêtre en premier, ce qui a conduit à postuler la possibilité que le cenancêtre ressemblait plus à un eucaryote [Forterre et Philippe 1999]. Cette idée avait été formulée beaucoup plus tôt au motif que 1) les eucaryotes possèdent beaucoup d'ADN non codant (non traduit), ce qui était postulé comme une caractéristique ancestrale par rapport aux génomes bactériens beaucoup plus compacts, et 2) qu'ils comportent des introns dans leurs gènes, i. e. des parties d'ADN non codantes s'intercalant à l'intérieur des gènes, ce qui aurait été d'un grand intérêt chez le cenancêtre pour permettre la recombinaison et formation de nouveaux gènes [Reaney 1974 ;

¹⁴ Cette méthode permet de prendre des différences de vitesse d'évolution non seulement entre positions d'un alignement, mais aussi entre différentes lignées (*hétérotachie*).

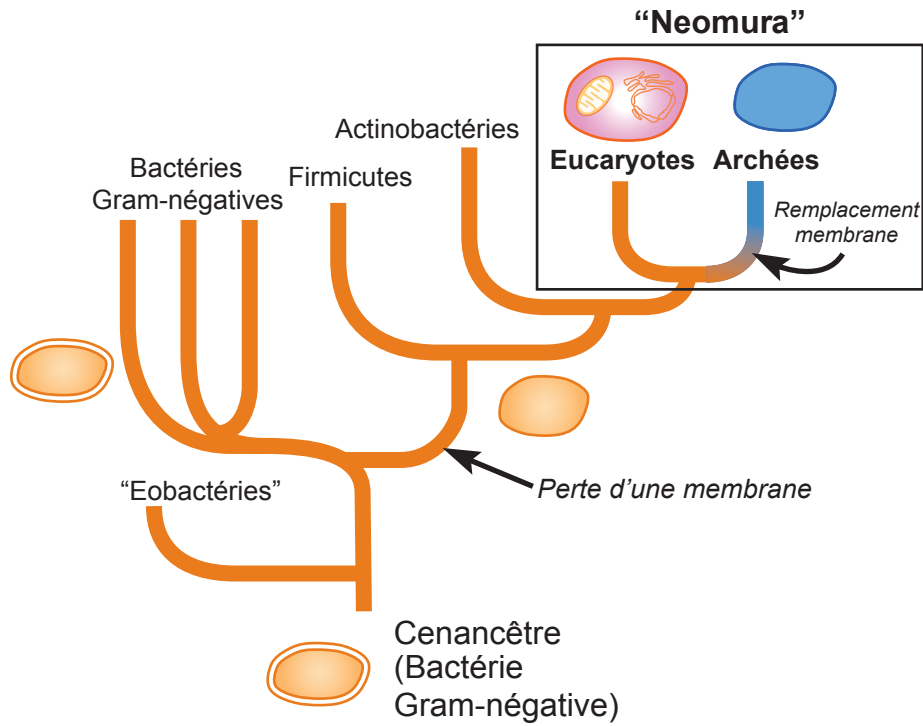


Figure 11. Représentation du modèle de Cavalier-Smith de l'évolution des organismes.

Ce modèle place la nature des membranes (spécialement leur nombre) au centre de l'évolution des organismes. L'enracinement de l'arbre entre les "éobactéries" et les autres bactéries se base sur l'enracinement par des paralogues des gènes de la photosynthèse. Les dessins symbolisent l'évolution supposée par ce modèle du nombre et du type de membranes.

Darnell 1978 ; Doolittle 1978]. Néanmoins, des observations sur l'absence d'opportunités de recombinaison fournies par les introns ont conduit à abandonner cette hypothèse [Stoltzfus *et al.* 1994].

D'autres travaux ont suggéré que les eucaryotes étaient la première branche à dériver du cenancêtre, par exemple en comparant des structures secondaires des ARN [Caetano-Anolles 2002], ou en essayant d'imaginer la machinerie en place dans le monde ARN [Jeffares *et al.* 1998 ; Penny et Poole 1999]. La nature eucaryote du cenancêtre a été discutée (ou plutôt sa nature proto-eucaryote car, comme il est admis que les eucaryotes modernes ont divergé après l'endosymbiose d'une bactérie à l'origine de la mitochondrie, ils doivent avoir évolué plus tard que le cenancêtre). Il a notamment été souligné que les procaryotes ont pu évoluer par perte d'éléments (réduction), voire par thermoréduction, i. e. une réduction liée à l'adaptation à des hautes températures [Forterre 1995 ; Penny et Poole 1999 ; Glansdorff 2000].

La mise en cause de la légitimité des enracinements basés sur des paralogues a conduit d'autres auteurs à considérer comme ancestrales certaines caractéristiques des organismes actuels. Cela leur permet d'orienter les phylogénies sur cette base et d'expliquer ainsi l'histoire des organismes. Le problème majeur de cette approche est qu'elle est arbitraire et soumise à l'avis éclairé de celui qui la formule. Par exemple, Cavalier-Smith a postulé que le principal caractère ancestral des organismes actuels est la présence d'une enveloppe cellulaire formée de deux membranes et d'une paroi de muréine. Parmi les organismes qui ont cette caractéristique, il utilise la phylogénie de deux paralogues non universels intervenant dans la synthèse de chlorophylle (qui n'est pas universelle non plus) pour enraciner l'arbre. Le résultat est un arbre dans lequel la première lignée à diverger du cenancêtre est celle des « éobactéries » (Chloroflexi), i. e. à l'intérieur des bactéries. Ensuite, certaines bactéries auraient perdu leur membrane extérieure, donnant ainsi lieu aux bactéries dotées d'une seule membrane, aux eucaryotes et aux archées (Figure 11, [Cavalier-Smith 2002]). L'inverse a aussi été postulé en considérant que c'étaient les bactéries avec une seule membrane et non pas deux qui devaient enraciner l'arbre [Lake *et al.* 2009]. Des racines à l'intérieur des archées ont aussi été proposées en se basant sur la plus petite divergence entre des ARNt, sur l'accumulation de caractéristiques archéennes considérées comme des adaptations intéressantes pour la vie dans la Terre primitive ou sur l'absence d'opérons et la fragmentation de gènes chez certaines archées très divergentes [Xue *et al.* 2003 ; Di Giulio 2007a ; Wong *et al.* 2007].

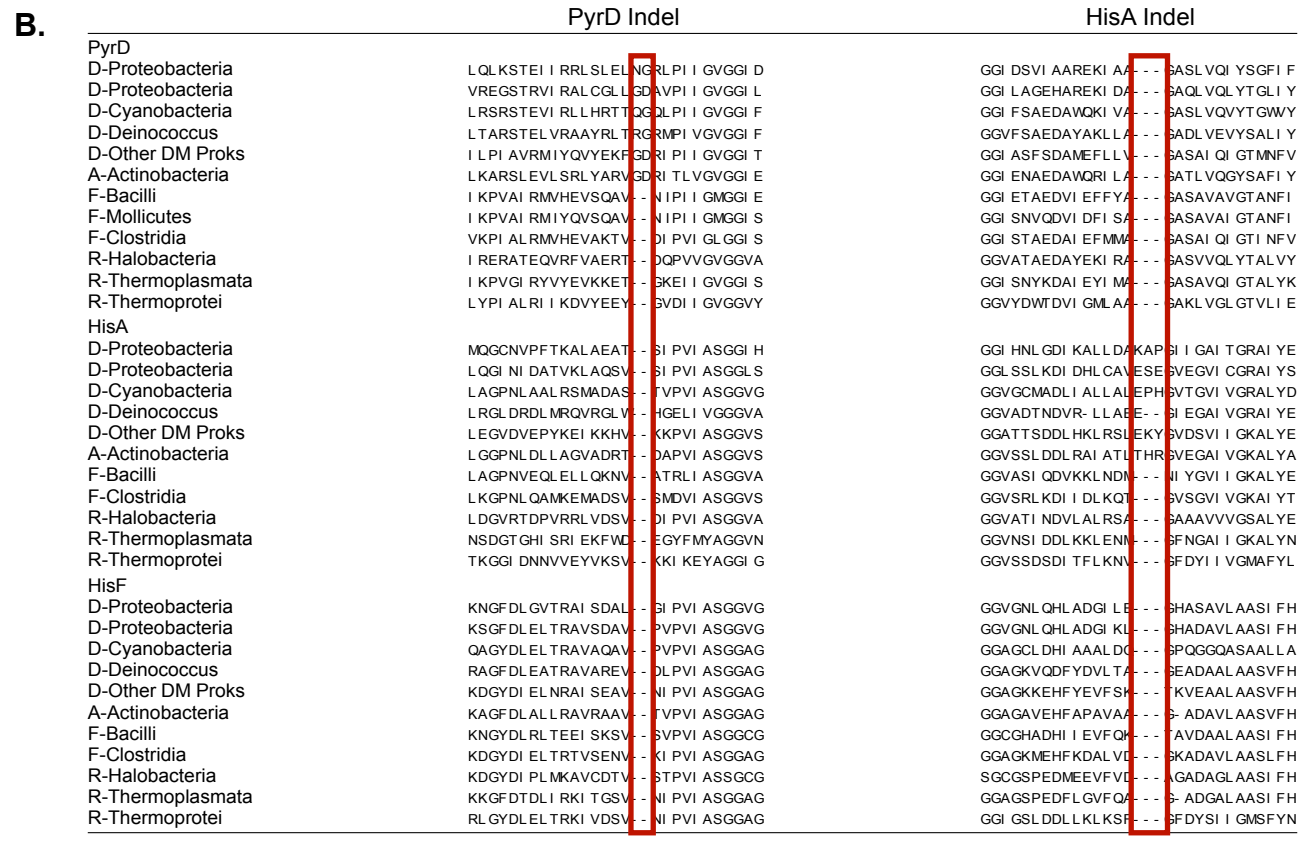
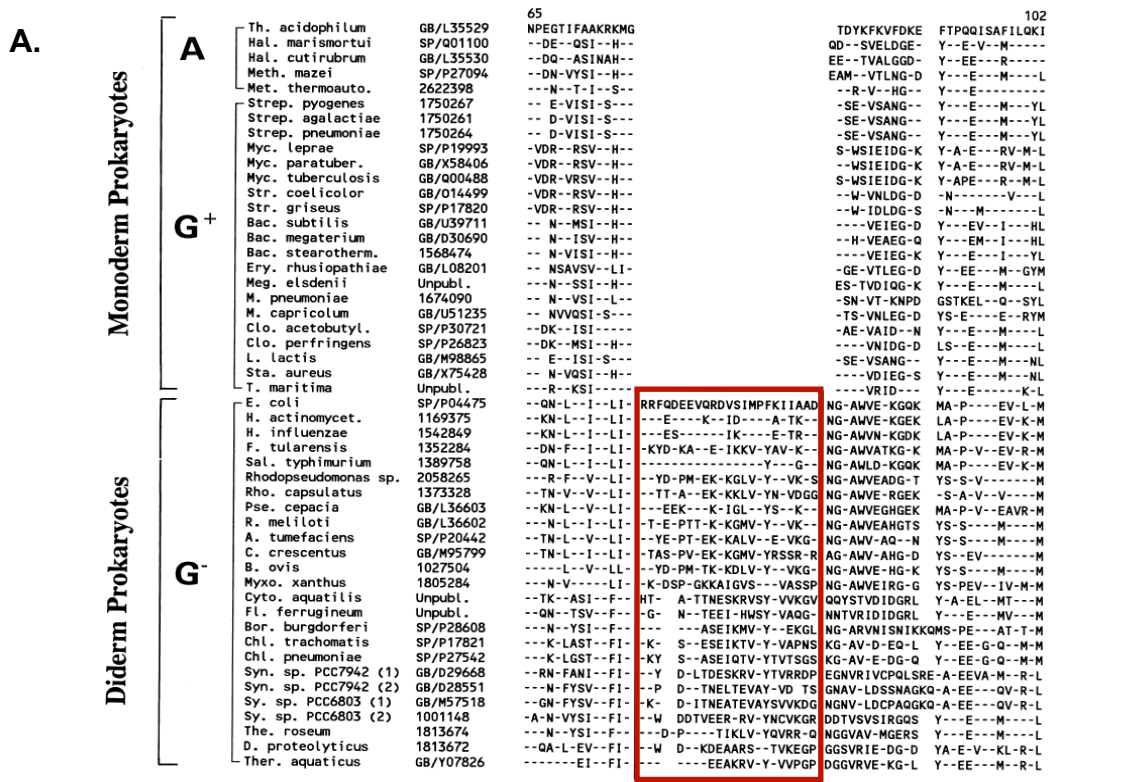


Figure 12. Exemples d'indels qui ont été abusivement utilisés pour déterminer l'emplacement de la racine de l'arbre du vivant.
A. D'après Gupta 1998. Alignement de séquences de la protéine Hsp70. La zone utilisée dans cette étude n'est pas correctement alignée et ne permet pas de visualiser des positions homologues, donc elle ne devrait pas être interprétée évolutivement. **B.** D'après Lake *et al.* 2008. Cette analyse utilise des indels trop courts qui peuvent être facilement le résultat d'homoplasies (convergences).

Ces méthodes sont inaptes à raciner l'arbre non seulement à cause de l'arbitraire qu'elles introduisent mais aussi parce qu'elles considèrent certains organismes comme des fossiles moléculaires vivants. Or la notion de fossile vivant est complètement erronée, comme nous l'avons vu dans la section précédente.

Pour essayer de contourner la difficulté du manque de signal phylogénétique et de l'arbitraire du choix d'un caractère comme ancestral, il peut sembler raisonnable de construire une phylogénie basée sur des événements génomiques rares qui permettraient de définir des groupes monophylétiques partageant une caractéristique remarquable, à l'exception des autres. Cette méthode se base sur l'exclusion de la racine des groupes monophylétiques, jusqu'à ce qu'il ne reste plus qu'un endroit où la racine soit possible. La difficulté de cette méthode réside dans la détection d'un événement génomique rare mais comparable entre tous les organismes. L'événement étudié est généralement un *indel*, i. e. une région génique contenant des insertions ou des délétions.

Malheureusement, cette méthode n'a pas donné non plus de résultats cohérents entre eux. Certains travaux ont proposé d'exclure la racine des archées [Skophammer *et al.* 2006], d'autres ont placé la racine entre les procaryotes qui ont une seule membrane cellulaire (archées et bactéries Gram positives) et ceux qui en ont deux (classiquement, les bactéries Gram négatives, [Gupta 1998b ; Gupta 1998a] et, enfin, d'autres ont utilisé cette méthode pour exclure la racine des archées et des bactéries gram négatives, ce qui placerait la racine dans les bactéries gram positives [Skophammer *et al.* 2006]. Cette méthode a même été utilisée pour essayer d'enraciner un « anneau » du vivant [Lake *et al.* 2008 ; Lake *et al.* 2009], qui représente l'histoire des organismes sous une forme de cercle au lieu d'une forme d'arbre afin de prendre compte de l'existence de transferts horizontaux (cette question sera traitée dans la section sur les transferts horizontaux).

Il existe des problèmes majeurs intrinsèques à l'utilisation d'indels pour déterminer des groupes monophylétiques. La convergence est d'autant plus probable que l'indel est court, et il est très difficile de détecter des transferts horizontaux avec cette méthode [Gribaldo et Philippe 2002 ; Dagan *et al.* 2010]. Or les travaux utilisant les indels pour déterminer la position de la racine ont utilisé des indels extrêmement courts ([Skophammer *et al.* 2006 ; Lake *et al.* 2008 ; Lake *et al.* 2009], Figure 12) ou même des zones visiblement mal alignées, i. e. des parties d'alignements qui ne permettent pas d'assurer l'homologie des positions

concernées et qui sont par conséquent impropres à déterminer des relations phylogénétiques ([Gupta 1998b ; Di Giulio 2007b], Figure 12).

Face à tous ces obstacles, on cherche encore de nouvelles méthodes. Certains ont essayé de détecter un plus grand nombre de paralogues universels pour tenter d'apporter plus d'arguments à l'enracinement mais, bien qu'un enracinement entre bactéries et archées semble celui obtenu le plus fréquemment, ces phylogénies sont soumises aux mêmes artefacts que les phylogénies classiques [Zhaxybayeva *et al.* 2005]. D'autres ont essayé d'enraciner l'arbre du vivant en partant du principe que le code génétique reflète l'ordre d'incorporation des acides aminés aux protéines, ce qui aurait pu se traduire dans un biais de composition en acides aminés des protéines à cette époque [Fournier et Gogarten 2007]. Ainsi, en comparant la composition des protéines ribosomales modernes avec le biais préalablement calculé d'acides aminés supposés être plus ou moins fréquents chez le cenancêtre, il a été montré que le signal des protéines bactériennes ressemble davantage au biais du code, et que ces séquences pourraient par conséquent conserver des caractéristiques plus anciennes, positionnant la racine dans ce domaine [Fournier et Gogarten 2010]. Finalement, on a essayé aussi d'enraciner un réseau du vivant. Pour cela, on a établi d'abord un réseau formé par la phylogénie de nombreux gènes, puis déterminé la racine en calculant le point médian du réseau. Le résultat plaçait la racine entre les archées et les bactéries [Dagan *et al.* 2010], mais les eucaryotes n'étaient pas étudiés dans ce travail.

En résumé, les études qui ont essayé de déterminer la racine du vivant ont donné des résultats insatisfaisants qui ne permettent pas d'établir une vision consensuelle des relations basales de l'arbre du vivant. L'interprétation des différents enracinements a d'ailleurs été trompeuse en de nombreuses occasions, à cause de la tendance erronée de beaucoup d'auteurs à considérer que les caractéristiques du groupe qui diverge en premier du cenancêtre doit conserver davantage de caractéristiques de celui-ci. L'approche qui me semble la plus prudente, et qui sera celle que j'appliquerai lors de mes propres analyses, est d'inférer indépendamment les propriétés de l'ancêtre de chaque domaine du vivant et d'utiliser ces informations pour construire une hypothèse sur le cenancêtre à partir de ces résultats.

Arbre ou forêt du vivant ?

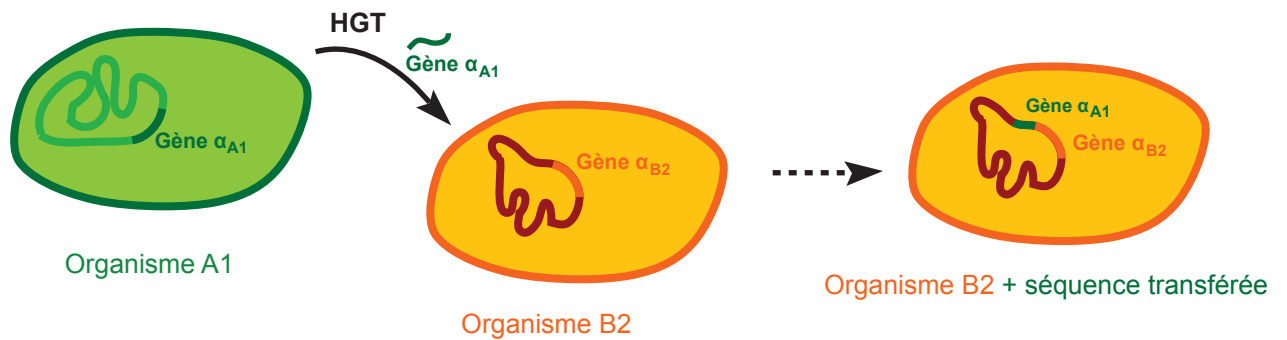
L'application des méthodes moléculaires à la microbiologie, commencée par Woese et développée grâce notamment aux résultats des nouvelles techniques de séquençage et de génomique a totalement changé notre vision de la diversité microbienne. Là où auparavant on

ne pouvait distinguer que quelques caractéristiques morphologiques et physiologiques, on a d'un coup pu découvrir une diversité génétique et métabolique sans précédent chez les organismes unicellulaires (bactéries, archées, mais aussi eucaryotes unicellulaires, Figure 13). Ces méthodes sont loin d'avoir fini de nous apporter des enseignements sur la diversité des microorganismes car elles nous permettent maintenant d'étudier même des organismes que nous sommes incapables de cultiver [López-García et Moreira 2008]. Ces mêmes techniques moléculaires devaient nous permettre de construire une phylogénie des organismes basée sur des caractéristiques universelles, mais on s'est alors retrouvé face à des résultats si contradictoires qu'on est même allé jusqu'à postuler que l'arbre du vivant n'existait pas. Or si la phylogénie des organismes n'existe pas, comment pourrait-on déterminer ce à quoi pouvait ressembler le cenancêtre ?

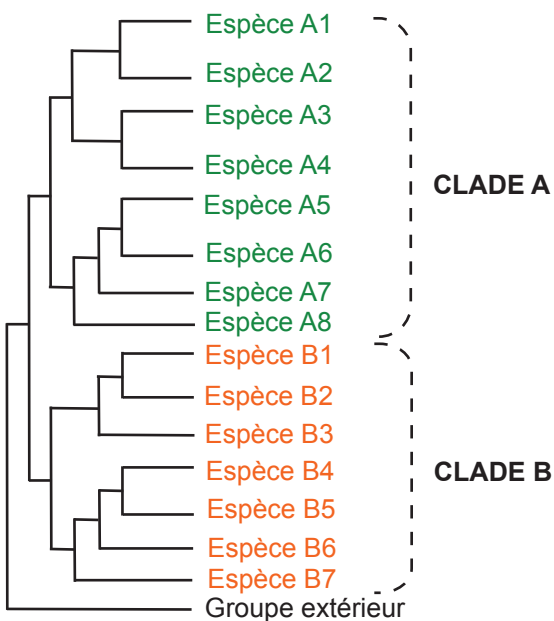
Avec l'accumulation de données génomiques on a pu réaliser des phylogénies d'un nombre de plus en plus grand de gènes et on a remarqué que, contrairement à ce que l'on croyait précédemment [Fox *et al.* 1980], ces phylogénies ne donnaient pas des résultats cohérents entre eux. L'incohérence entre les phylogénies de différents gènes a été assez vite interprétée comme le résultat de transferts horizontaux de gènes (HGT) [Syvanen 1985 ; Brown et Doolittle 1997]. Si ce gène est maintenu (*fixé*) dans les descendants du récepteur, alors la phylogénie du gène devient différente de celle des organismes car les gènes transférés se placeront à côté de leurs plus proches parents, qui sont les gènes des parents de l'organisme donneur (Figure 14). En soi l'existence de HGT n'est pas surprenante car on connaît plusieurs mécanismes qui peuvent véhiculer des tels transferts (Figure 15). Ce qui étonnait dans un premier temps était leur nombre, que ce soit en se basant la détection de transferts sur la comparaison de phylogénies [Doolittle 1999] ou en étudiant les variations de biais de composition dans les génomes [Lawrence et Ochman 1998]¹⁵. Si les gènes sont échangés entre des lignées distantes, comment doit-on comprendre l'évolution des génomes et celle des organismes ?

Très vite on a proposé que la phylogénie des organismes ne soit pas représentable comme un arbre mais plutôt comme un réseau dans lequel on représenterait l'histoire

¹⁵ Les génomes sont caractérisés entre autres par le rapport entre le nombre de bases G-C et celui des bases A-T, qui sont complémentaires. Lorsqu'un gène qui provient d'une lignée arrive dans un récepteur d'une lignée différente, le biais peut-être différent. Bien que les mutations s'accumulent dans le gène transféré aient tendance à effacer ce biais de composition, la détection de gènes avec un biais différent de celui du génome est possible pour détecter des HGT récents.



Phylogénie des organismes



Phylogénie du gène α

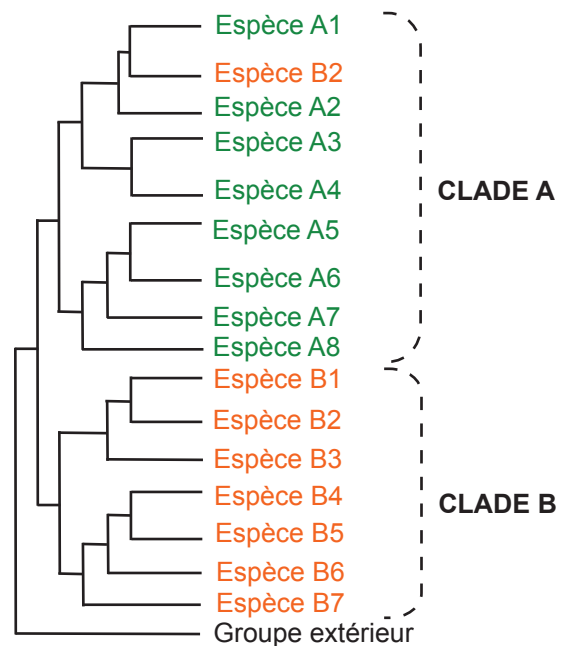


Figure 14. Impact des transferts horizontaux de gènes (HGT) sur la phylogénie des gènes. En bas à gauche est représentée une phylogénie théorique des organismes. La phylogénie de droite montre le transfert d'un gène d'un organisme du clade A vers un organisme du clade B. Du fait de ce transfert, les séquences les plus proches du xénologue qui se trouve dans l'organisme B2 sont les organismes proches de l'organisme donneur.

conflictuelle de tous les gènes [Hilario et Gogarten 1993 ; Kunin *et al.* 2005b]. Le fait que les eucaryotes aient reçu beaucoup de gènes par HGT des procaryotes, comme on le verra dans la section suivante, a même conduit à postuler que les échanges de gènes pourraient permettre de représenter l'évolution du vivant sous forme d'un réseau cyclique fermé au niveau des eucaryotes, un *anneau du vivant* [Rivera et Lake 2004]. Avec le temps, deux écoles de pensée se sont formées, l'une postulant que les HGT n'ont pas été si fréquents et qu'il existe une partie des gènes qui sont très rarement transférés, ce qui permet d'inférer la phylogénie des organismes à partir de celle de ces gènes, et l'autre considérant la quantité de HGT trop importante pour accepter que la phylogénie des organismes puisse être déterminée, ou même pour considérer qu'elle existe. L'une des priorités face à l'incohérence apparente entre les phylogénies des gènes isolés était de déterminer l'étendue des transferts. De nombreux exemples ou méthodes ont été proposés pour estimer la fréquence des HGT dans l'histoire des organismes [Lawrence et Ochman 1998 ; Ochman *et al.* 2000 ; Clarke *et al.* 2002 ; Snel *et al.* 2002 ; Boucher *et al.* 2003 ; Daubin *et al.* 2003 ; Ge *et al.* 2005 ; Kunin *et al.* 2005a ; Lerat *et al.* 2005 ; Pal *et al.* 2005 ; Abby *et al.* 2012] mais, malheureusement, ces méthodes très diverses s'appliquaient à des périodes de l'évolution, des échantillons et des gènes différents, donnant des résultats incohérents entre eux qui n'ont pas permis aux deux écoles de se mettre d'accord.

Chez les défenseurs de l'arbre, il existe l'idée que même si les HGT existent, ils ne doivent pas affecter toutes les fonctions de la même manière. Certains gènes pourraient être préservés et leurs phylogénies seraient alors représentatives de l'histoire des organismes. Généralement, ce sont les machineries de traduction et parfois aussi de transcription (gènes informationnels) qui servent de cœur pour la construction des phylogénies du vivant [Harris *et al.* 2003] car, même si quelques HGT de ces gènes ont été identifiés [Brochier *et al.* 2000], ces événements semblent assez rares. Comme par ailleurs le signal contenu en un seul gène est trop faible pour résoudre les relations taxonomiques à l'échelle de l'arbre entier de façon robuste, on travaille généralement avec des concaténations de ces gènes, que l'on espère congruentes entre elles [Brown *et al.* 2001 ; Delsuc *et al.* 2005 ; Gribaldo et Brochier 2009]. Certaines analyses utilisant uniquement la présence/absence des gènes dans les génomes ont montré que la topologie de l'arbre global des génomes était proche de celle de la machinerie de traduction [Fitz-Gibbon et House 1999 ; Snel *et al.* 1999], et ce même lorsqu'on écartait les gènes impliqués dans ce processus. Cela suggère que le signal est malgré tout assez fort parmi les gènes opérationnels (par opposition aux informationnels) pour reporter une

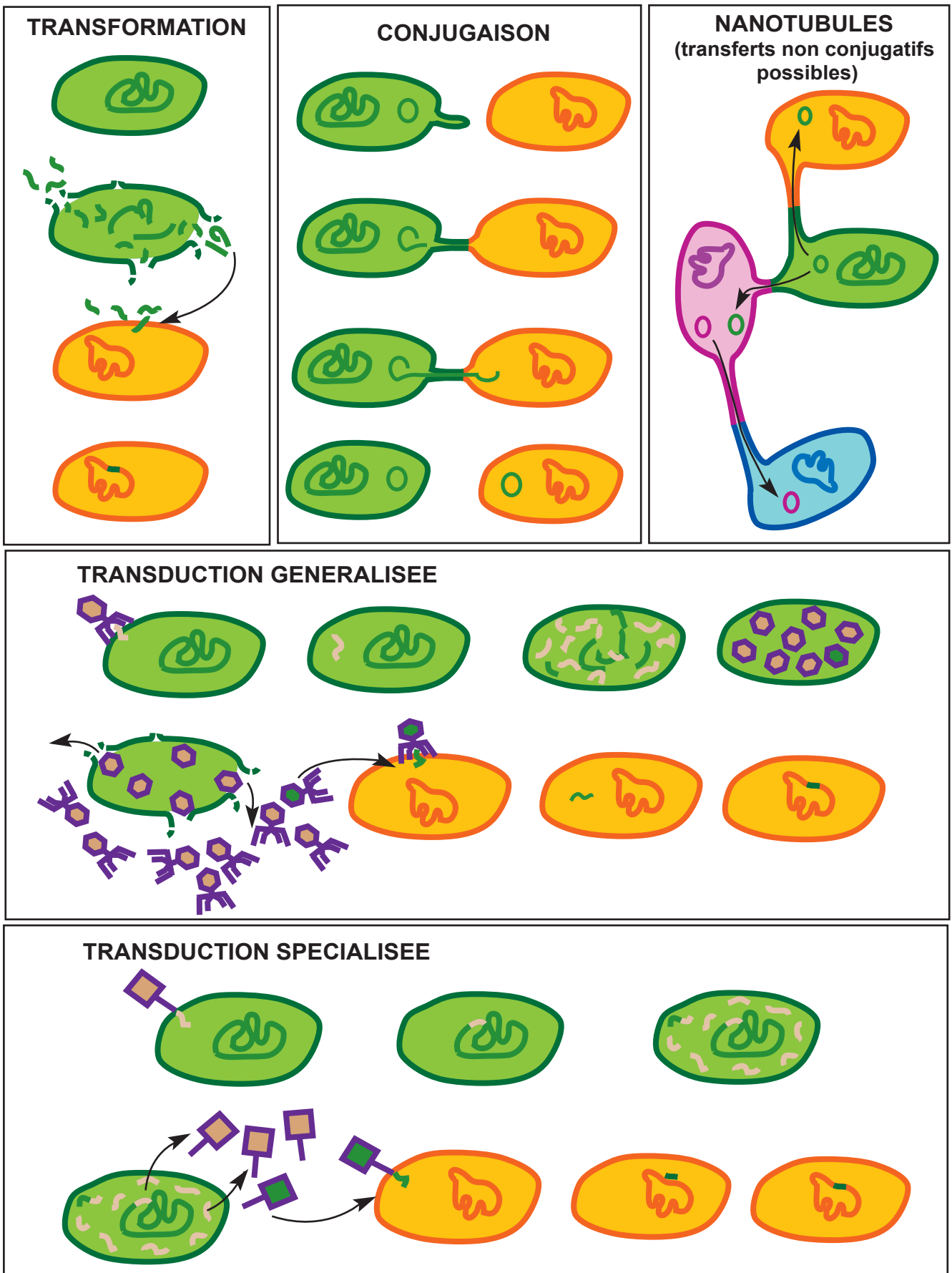


Figure 15. Mécanismes naturels de HGT.

Transformation. L'incorporation de séquences du milieu. **Conjugaison.** Construction de ponts entre les cellules provoquée par un plasmide qui se transmet d'une cellule à l'autre. **Nanotubules** (découverte récente, Dubey et Ben-Yehuda 2011). Formation de communications cellulaires sans faire intervenir un plasmide conjugatif. **Transduction généralisée.** Transfert véhiculé par des phages lytiques dont certains virions peuvent transporter des gènes cellulaires par erreur. **Transduction spécialisée.** Transfert véhiculé par des virus lysogéniques qui s'intègrent dans le génome cellulaire et peuvent emporter des séquences proches de leur site d'insertion.

phylogénie comparable. D'autres études ont intégré une information supplémentaire sur la similarité entre les gènes partagés dans les génomes et ont révélé des phylogénies proches de celles basées sur l'ADN codant l'ARNr 16S [Tekaiia *et al.* 1999 ; Clarke *et al.* 2002 ; Kunin *et al.* 2005a]. Néanmoins, ni la présence/absence des gènes dans les génomes ni la détection des gènes les plus ressemblants ne permettent de prendre en compte les HGT.

Une autre critique formulée par les défenseurs de l'arbre du vivant est que ses adversaires utilisent systématiquement le HGT pour expliquer l'incongruence des arbres. Or d'autres raisons permettent d'expliquer ces désaccords [Glansdorff 2000] tant d'un point de vue biologique (paralogies cachées) que d'un point de vue méthodologique (artéfacts dus à la saturation des séquences et autres). Par exemple, il a été montré que la majorité des incohérences entre phylogénies étaient en réalité statistiquement peu soutenues [Daubin *et al.* 2003] et devaient donc être analysées avec précaution.

Du côté des défenseurs du réseau du vivant, la présence de quelques incohérences entre certains gènes que les partisans de l'arbre du vivant considèrent comme le cœur le plus représentatif de l'histoire des organismes montrerait qu'un tel cœur n'existe pas [Nesbo *et al.* 2001]. D'autres travaux ont essayé d'expliquer la cohérence de certaines topologies en signalant qu'elles pouvaient être corrélées à des niches écologiques, au mode de vie des organismes ou à leur métabolisme plutôt qu'à leur degré de parenté [Gogarten *et al.* 2002 ; Boucher *et al.* 2003 ; Jain *et al.* 2003 ; Pal *et al.* 2005 ; Schliep *et al.* 2010].

Comme dans beaucoup de débats, certains argumentaires sont devenus de plus en plus extrêmes, tant dans le sens de la minimisation de l'importance des HGT [Kurland *et al.* 2003] que dans le sens de l'exagération de leur importance [Woese 2000 ; Doolittle et Baptiste 2007]. Autant on doit reconnaître la réalité des HGT parce qu'on est capable de les détecter et de les expliquer, autant leur apologie révèle des confusions plus profondes sur la nature de l'arbre du vivant. Notamment, Doolittle et Baptiste ont attaqué dans leur article de 2007 la notion même d'arbre du vivant en disant qu'il s'agit d'une idée antérieure à Darwin et que celui-ci l'a reprise à son compte en essayant de la justifier grâce à sa notion de descendance avec modification. Ces auteurs estiment donc que l'arbre est une idée reçue, sans fondement primaire et qui par conséquent doit être abandonnée au profit d'un modèle qui reconnaisse pleinement l'incohérence entre les phylogénies. La réalité à leurs yeux est que la majorité des gènes sont distribués d'une façon chaotique à travers les génomes à cause d'une « histoire

évolutive (non-verticale) et complexe » favorisée par les nombreux moyens d'échange horizontal qui existent dans la nature.

Une telle interprétation du travail de Darwin est incorrecte car inversée. La grande préoccupation de Darwin dans l'*Origine des espèces* est de démontrer qu'il peut exister des mécanismes de sélection naturelle qui favorisent, au cours des générations, l'accumulation de certaines caractéristiques dans les variétés d'organismes, sans se préoccuper de l'origine de ces caractéristiques. Une fois que ce mécanisme a été expliqué, Darwin essaye d'extrapoler son idée à des nombreuses questions (ce que la biologie moderne continue de faire depuis). Il propose notamment la formulation de l'idée que ce que l'on applique aujourd'hui aux variétés devait aussi exister dans le passé, ce qui permettrait d'étudier les organismes par rapport à leur généalogie commune qui, à terme, a été développée sous la forme d'un arbre. Ainsi, l'arbre n'est pas le point de départ, mais l'un des points d'arrivée de Darwin. On ne peut pas l'attaquer comme une idée reçue sans fondement, mais on doit attaquer le raisonnement qui le précède.

C'est d'ailleurs ce que font Doolittle et Baptiste en mettant en cause la descendance avec modification elle-même, ce qui semble extravagant. Il est en effet difficile aujourd'hui de douter de l'existence de la répllication de l'information génétique et de la division cellulaire, et on fait l'expérience tous les jours dans le laboratoire que lorsqu'un gène est présent dans une population c'est parce que les organismes qui la composent sont issus d'un ancêtre commun, et non pas parce que chaque cellule de la population reçoit chaque gène indépendamment par HGT. Ainsi, les HGT peuvent être des événements plus ou moins fréquents et influents dans la génération de nouvelles variantes d'organismes, au même titre que les mutations, réarrangements chromosomiques, duplications ou fusion de gènes, mais leur propagation dans les populations ne peut se faire effectivement que par le mécanisme normal de la vie de la cellule. Par conséquent, ces arguments théoriques pour exclure l'existence de l'arbre du vivant ne sont pas recevables.

En dehors des considérations théoriques, l'une des grandes préoccupations relatives aux HGT concerne le nombre de gènes qui constituent le cœur central libre de HGT, cœur qui permet de construire des phylogénies que l'on peut assimiler à celles des organismes. Dès que la phylogénie moléculaire a commencé à devenir plus accessible dans les années 1990, il apparut l'idée que si la phylogénie construite à partir d'un gène n'était peut être pas assez fiable pour résoudre toute l'histoire des organismes, l'accumulation de beaucoup de gènes,

dont on espérait à l'époque que les phylogénies seraient cohérentes entre elles, permettrait d'avoir plus de signal et par conséquent plus de robustesse [Brown et Doolittle 1997 ; Woese 1998]. Cependant, avec la prise de conscience de l'étendue des HGT on a restreint de plus en plus l'éventail de gènes universels capables de jouer ce rôle jusqu'à arriver à une trentaine de gènes qui sont encore en discussion [Ciccarelli *et al.* 2006].

Les défenseurs du réseau ont considéré que ce petit nombre de gènes était négligeable par rapport à l'ensemble du génome et ils ont appelé l'arbre construit avec cette trentaine de gènes « l'arbre du 1% [du génome] » pour signaler sa faible représentativité [Dagan et Martin 2006]. Ces auteurs ont cependant négligé le fait que lorsqu'on parle d'arbre du vivant, on parle de l'arbre qui représente la généalogie des organismes et que cet arbre est, comme le cenancêtre, une conséquence directe de la l'évolution darwinienne et de la théorie cellulaire qui sont la base de la biologie moderne. Il est en soi possible que les gènes dont on dispose aujourd'hui ne permettent pas de reconstruire l'histoire des organismes, mais cette histoire *doit avoir existé*. Si l'on disposait d'un seul gène capable de nous fournir cette information, on n'aurait pas besoin de plus. Le problème est de déterminer si les marqueurs utilisés sont les bons plutôt que de savoir si les organismes ont une histoire. Si ce 1% est représentatif de l'histoire des organismes, alors ce sera une fenêtre de choix pour comprendre l'histoire de la vie pendant plus de 3 milliards d'années [Gribaldo et Brochier 2009].

Les HGT, comme tous les autres événements évolutifs, peuvent nous renseigner sur l'histoire des organismes et sur leurs caractéristiques mais nous ne devons pas oublier que les organismes sont quelque chose de plus qu'un groupe de gènes [Morange 2011]. La génomique nous permet d'accéder à des informations du passé, mais la question que se pose la biologie évolutive est surtout de savoir comment le fonctionnement des organismes a changé au cours du temps. Pour cette raison, j'ai surtout essayé dans mon travail de proposer les mécanismes métaboliques qui peuvent avoir été importants à l'époque des ancêtres de chaque domaine du vivant et du cenancêtre, en essayant de dépasser la génomique pour arriver à la fonction.

Plusieurs raisons suggèrent que retracer l'histoire des événements aussi anciens semble possible en dépit des HGT. D'abord, bien que les HGT puissent être fréquents à l'échelle du génome, ils le sont beaucoup moins à l'échelle du gène. Ainsi, en dehors des moments où les gènes sont transférés, les phylogénies de gènes peuvent fournir des informations cohérentes avec celle des organismes [Lerat *et al.* 2005]. Les HGT peuvent même être une source

d'information évolutive supplémentaire car ils constituent des événements génomiques rares (voir section précédente) qui peuvent permettre de définir des groupes phylogénétiques, et par conséquent d'apporter plus d'information pour résoudre des relations problématiques de l'arbre du vivant [Daubin *et al.* 2002 ; Abby *et al.* 2012]. Une conséquence importante de ces perspectives est que l'arbre du vivant auquel on peut espérer accéder un jour sera probablement un résultat composite et non pas lié à un seul marqueur, comme on aurait pu le penser dans les années 1980.

Il existe aussi des limites aux HGT. D'abord, pour être fixés, les gènes doivent arriver dans un contexte cellulaire favorable qui peut être très variable [Sorek *et al.* 2007]. Cela explique que les transferts se fassent souvent entre organismes relativement proches soit phylogénétiquement [Andam et Gogarten 2011 ; Popa et Dagan 2011], soit physiquement, en partageant des niches communes [Jain *et al.* 2003 ; Schliep *et al.* 2010]. Ensuite, le fait que certains gènes soient plus faciles à transférer que d'autres est très lié à l'hypothèse de complexité : la fixation d'un gène dans un génome après un transfert sera d'autant plus difficile qu'il entre dans un complexe comprenant beaucoup d'éléments car il aura besoin des autres éléments pour fonctionner [Jain *et al.* 1999 ; Cohen *et al.* 2010]. L'hypothèse de complexité s'applique plus facilement à des structures très intégrées comme celles de la machinerie informationnelle. Néanmoins, les voies de biosynthèse que j'ai étudiées sont aussi des éléments très intégrés dans le réseau métabolique et dont la fonction est souvent essentielle à la vie de l'organisme. Ces contraintes devraient se refléter dans l'histoire évolutive des protéines étudiées, par exemple en distinguant les gènes centraux du métabolisme peu transférés de ceux périphériques, qui seraient plus fréquemment échangés entre lignées lointaines. De ce fait, le détail des informations fonctionnelles disponibles dans la littérature fournira des arguments supplémentaires pour déterminer l'origine orthologue, paralogue ou xénologue des gènes étudiés.

Origine des eucaryotes et relation avec le cenancêtre

L'origine des eucaryotes est l'un des sujets qui soulèvent le plus de questions et de polémiques dans le domaine de la biologie évolutive. Je ne vais pas traiter directement cette question dans mon travail, je vais me contenter de la mentionner brièvement afin d'expliquer comment mon travail a pris en compte l'incertitude actuelle sur les relations de parenté entre les trois domaines du vivant. Comme nous l'avons vu, la séparation originelle des organismes en procaryotes et eucaryotes sous-entend que les premiers ont précédé les seconds. Bien que

A.

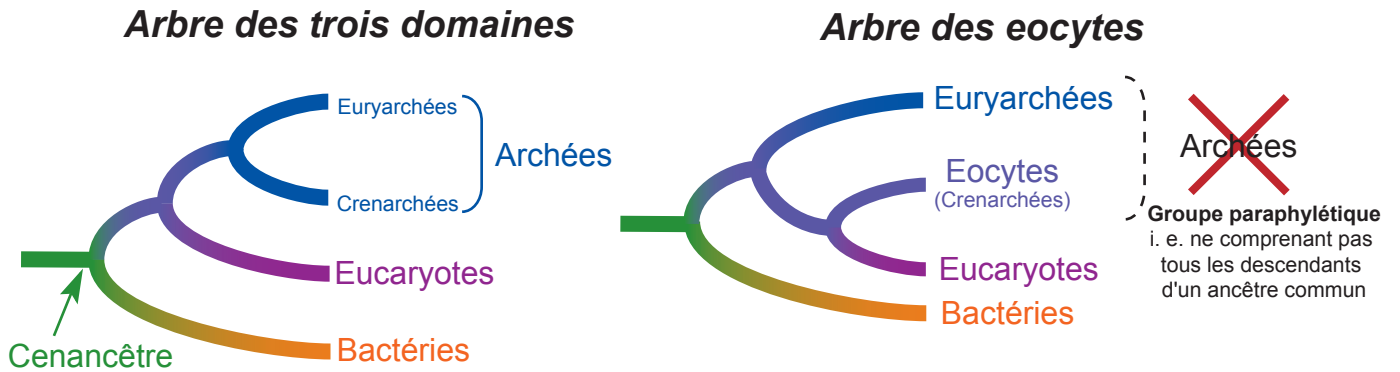


Figure 16. A. Hypothèse des eocytes.

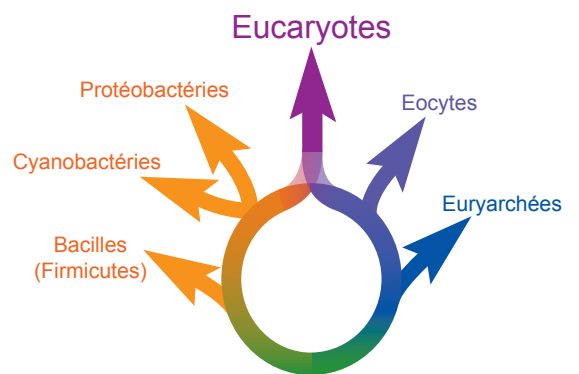
Cette hypothèse se base sur des caractéristiques communes entre les eucaryotes et les crenarchées (ici appelés eocytes) pour proposer que les premiers aient divergé des seconds. De fait, cette hypothèse ne considère plus les archées comme un groupe monophylétique mais paraphylétique.

B. Anneau du vivant.

Lake et ses collaborateurs ont aussi proposé qu'il existe de très nombreux HGT entre les organismes et que les eucaryotes résultent de la combinaison des gènes des eocytes et d'une protéobactérie et/ou cyanobactérie, de sorte que la représentation de l'histoire des organismes devrait se faire par un "anneau du vivant" au lieu d'un arbre.

B.

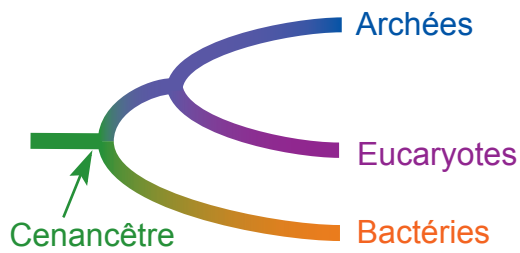
Anneau du vivant



reconnaissant que cette observation était probablement vraie d'un point de vue cytologique, Woese insiste volontairement sur son caractère non phylogénétique lorsqu'il propose la division du vivant en trois domaines [Woese et Fox 1977b]. Ce changement de perspective a même conduit certains auteurs à proposer que le cenancêtre ait pu ressembler davantage à un (proto)eucaryote qu'à un procaryote [Reaney 1974 ; Darnell 1978 ; Doolittle 1978 ; Forterre 1995]. Cependant, il est aujourd'hui admis que le dernier ancêtre commun des eucaryotes (LECA) a dû être un organisme plus récent que les derniers ancêtres communs des autres domaines du vivant, car LECA contenait déjà des mitochondries que l'on sait avoir été le résultat de l'endosymbiose de bactéries de type moderne [Gray et Doolittle 1982]. La lignée de LECA étant plus récente que les lignées procaryotes, cela repose la question du lien direct entre la lignée eucaryote et le cenancêtre [Kurland *et al.* 2006].

L'enracinement de l'arbre de Woese sur la branche bactérienne implique un ancêtre commun partagé entre les archées et les eucaryotes (Figure 6, p.24). Cette hypothèse semble soutenue par un certain nombre de phylogénies [Iwabe *et al.* 1989 ; Brown et Doolittle 1997 ; Gribaldo et Cammarano 1998], par des propriétés enzymatiques partagées des ATPases [Zhaxybayeva *et al.* 2005] et par une machinerie traductionnelle beaucoup plus ressemblante entre les deux domaines [Hartman *et al.* 2006]. Cependant, plusieurs analyses ont montré que les ribosomes des eucaryotes semblent être plus similaires à ceux des crenarchées (eocytes) qu'à ceux des autres archées, ce qui a été interprété comme un indice que les archées ne constituaient pas vraiment un domaine du vivant, mais que les eucaryotes seraient une lignée dérivée des eocytes (Figure 16), [Lake *et al.* 1984 ; Lake 1988 ; Rivera et Lake 1992 ; Baldauf *et al.* 1996 ; Cox *et al.* 2008]. Pourtant, lorsque des homologues des gènes eucaryotes sont cherchés dans les génomes des organismes des autres domaines, on observe qu'il existe une grande quantité de gènes qui n'ont d'équivalent que parmi les bactéries [Doolittle 1998 ; Esser *et al.* 2004 ; Dagan et Martin 2006]. On accepte généralement qu'une partie de ces gènes a été acquise depuis les endosymbiontes à l'origine de la mitochondrie et du chloroplaste. Néanmoins, le nombre de ces gènes semble trop important pour certains et, surtout, leurs phylogénies ne replacent pas toujours les séquences eucaryotes à proximité des alpha-protéobactéries et des cyanobactéries qui sont respectivement les plus proches de la mitochondrie et du chloroplaste. Une hypothèse formulée pour expliquer cette observation est que les ancêtres eucaryotes se nourrissaient par phagocytose et qu'une bonne partie des gènes en cause auraient pu être acquis directement des bactéries digérées par ce mécanisme [Doolittle 1998].

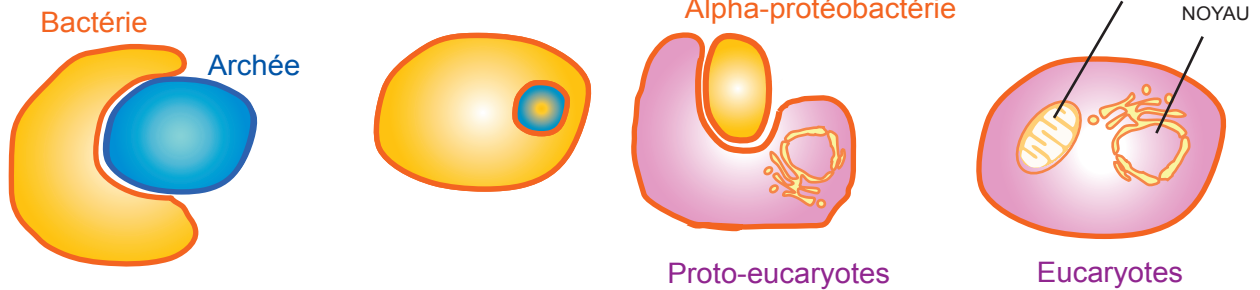
Arbre des trois domaines



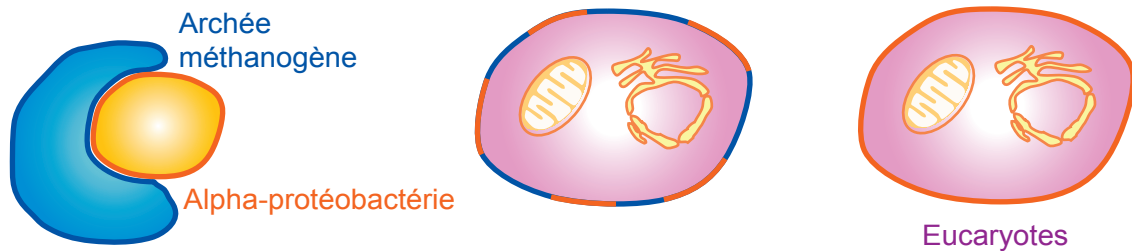
Fusion de deux domaines primordiaux



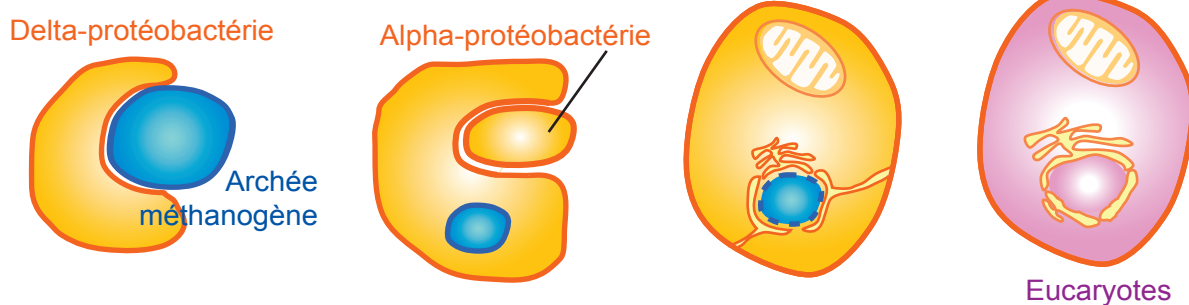
Gupta 1996



Martin & Müller 1998



Moreira & López García 1998



Margulis et al. 2000

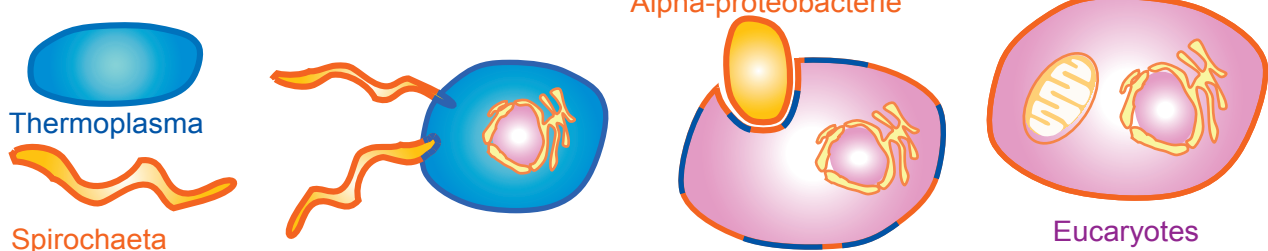


Figure 17. Hypothèses chimériques faisant intervenir des bactéries et des archées pour expliquer l'origine des eucaryotes.

D'autres auteurs ont remarqué que les gènes des processus informationnels des eucaryotes semblaient plus similaires de leurs équivalents archéens alors que les gènes opérationnels ressemblaient plus aux bactéries [Horiike *et al.* 2001]. Cet argument a donné un grand crédit aux hypothèses qui proposaient que les eucaryotes pourraient être le résultat de la fusion entre au moins une bactérie et une archée (Figure 17), la première ayant apporté les gènes métaboliques et la deuxième ayant contribué aux gènes informationnels [Gupta et Golding 1996 ; Martin et Muller 1998 ; Moreira et López-García 1998 ; Margulis *et al.* 2000 ; Rivera et Lake 2004 ; López-García et Moreira 2006]. En revanche, la nature des organismes qui auraient participé à cette fusion reste très discutée. Les archées candidates pour la fusion ont été des eocytes, des euryarchées méthanogènes et des thermoacidophiles, alors que pour les bactéries plusieurs types de protéobactéries et même une spirochète ont été proposées. Finalement, il a été postulé que la fusion aurait pu se produire entre un ou plusieurs de ces procaryotes et une lignée proto-eucaryote issue directement du cenancêtre dont proviendraient les caractéristiques uniques des eucaryotes comme le cytosquelette, ses systèmes de signalisation et autres complexes protéiques [Hartman et Fedorov 2002 ; Hartman *et al.* 2006].

En résumé, on est encore loin d'avoir trouvé un accord sur l'origine des eucaryotes et les relations de ce domaine du vivant avec les deux autres, mais la recherche à ce propos continue [Gribaldo *et al.* 2010]. Je pense notamment que, même s'il est compréhensible d'utiliser les organismes modernes comme approximation, il faut éviter l'assimilation fréquente entre les procaryotes actuels et ceux qui ont pu exister à l'époque de l'apparition des eucaryotes modernes. Les travaux qui ont déjà commencé à décrire la biologie de LECA et des ancêtres bactériens [Eme *et al.* 2009 ; Koonin 2010 ; Eme *et al.* 2011] devraient aboutir à une vision plus complète des mécanismes qui ont conduit à l'apparition des eucaryotes.

En tout cas, la position incertaine des eucaryotes dans la phylogénie des organismes pose un problème lorsqu'il s'agit de déterminer si elle peut s'accorder ou non à une phylogénie de gène. Dans mon travail, j'ai appliqué une interprétation pragmatique (Figure 18). Lorsque la topologie de la phylogénie d'un gène correspond à la séparation des trois domaines (en prenant en compte la possibilité de transferts horizontaux de gènes), j'ai considéré que ce gène devait être présent chez le cenancêtre et hérité dans les trois lignées. Lorsque la monophilie des archées et des bactéries est retrouvée et que les eucaryotes branchent au sein des bactéries avec des indices clairs de réception des gènes notamment par endosymbiose, j'ai aussi considéré que la présence du gène pouvait être inférée chez le

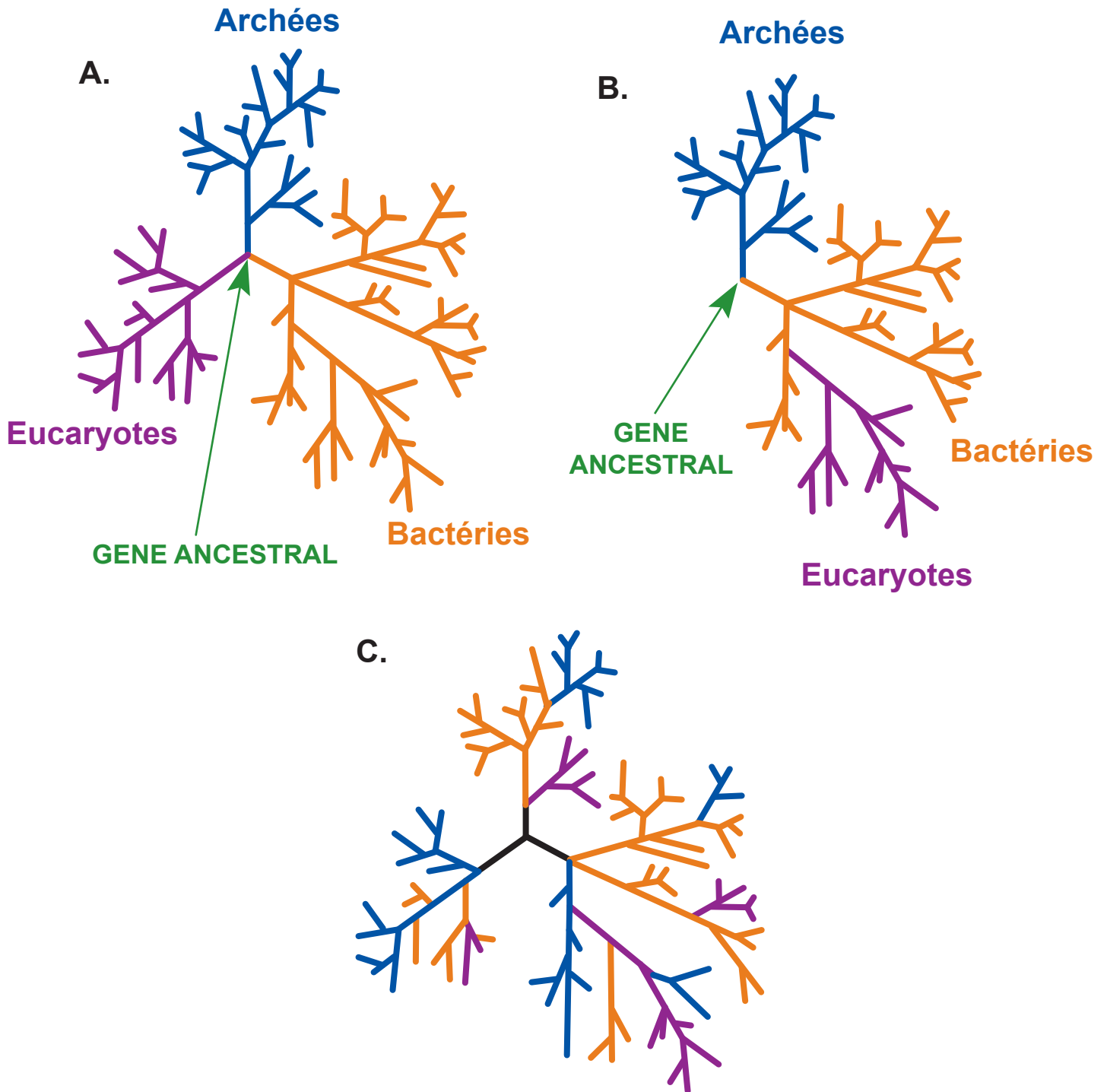


Figure 18. Interprétation de la topologie des phylogénies pour la détermination de la présence d'un gène chez le cenancêtre.

A. La phylogénie du gène est cohérente avec les principaux taxons des trois domaines. **B.** La phylogénie du gène est cohérente avec les principaux taxons des bactéries et des archées et les eucaryotes sont absents ou branchent à l'intérieur d'un groupe procaryote. **C.** La phylogénie du gène est incohérente avec les principaux groupes taxonomiques des trois domaines. Dans les cas A et B, le gène est considéré comme probablement présent chez le cenancêtre. Dans le cas C, l'information contenue dans l'arbre n'est pas suffisante pour résoudre cette question.

cenancêtre, soit en acceptant que la lignée eucaryote a eu un remplacement de son gène ancestral par celui de l'endosymbionte, soit en accord avec les théories de fusion. A l'inverse, lorsque les phylogénies des gènes ne présentent pas une topologie cohérente avec les grands groupes taxonomiques bactériens et archéens j'ai estimé que la phylogénie ne pouvait pas fournir des informations sur le cenancêtre. De ce fait, certaines de ces phylogénies suggèrent une absence de leurs gènes respectifs dans le cenancêtre ou ne sont pas concluantes et doivent donc être interprétées avec beaucoup de précaution (Figure 18).

5. Portrait robot du cenancêtre

A quoi ressemblait le cenancêtre ?

Nous avons vu qu'il existe un très grand nombre d'incertitudes en ce qui concerne l'évolution des organismes anciens. Beaucoup de descriptions du cenancêtre ont été fondées sur la base de caractéristiques qui semblaient plus primitives ou plus compatibles avec l'apparition de la vie, que ce soient les archées à cause de leurs habitats, les bactéries à cause de leur morphologie simple, les eucaryotes à cause de leur machinerie informationnelle utilisant beaucoup l'ARN ou même une population de pré-cellules qui mélangerait toutes ces caractéristiques à la fois. Le placement de la racine a joué aussi un rôle important dans la mesure où les organismes qui divergeaient en premier du cenancêtre étaient interprétés comme primordiaux. Cependant, la détermination *a priori* des caractéristiques primitives est toujours contestable et nous avons vu que pour inférer la présence d'un gène ou d'une fonction chez un ancêtre, quel qu'il soit, il ne suffit pas de caractériser la première lignée divergente, qui d'ailleurs s'avère souvent controversée étant donnée la difficulté à établir une phylogénie des organismes. Au contraire, la comparaison des organismes et la reconstruction phylogénétique restent les meilleures méthodes pour déterminer si un caractère était présent chez l'ancêtre d'un groupe, en l'occurrence le cenancêtre. Quels sont alors les gènes et les fonctions cellulaires dont la présence peut être inférée dans le cenancêtre par comparaison entre les organismes actuels ?

Le code génétique est universel à quelques très rares et récentes exceptions près. Sa présence chez le cenancêtre est donc acceptée et régulièrement étudiée [Fournier et Gogarten 2007]. Dans un premier temps, l'accès à la séquence des génomes a permis de comparer leur contenu en gènes afin de déterminer la partie homologue commune, que l'on postulait avoir été présente chez le cenancêtre [Mushegian et Koonin 1996 ; Kyrpides *et al.* 1999].

Cependant, au fur et à mesure de l'accumulation de génomes on s'est aperçu que le nombre de gènes universels chutait très rapidement. Une estimation relativement récente du cœur génétique universel basée sur l'observation de 3100 COG¹⁶, montre que seulement 80 d'entre eux sont présents dans tous les organismes. Sur ces 80 gènes, une cinquantaine seulement a une phylogénie cohérente avec celle des gènes codant l'ARNr 16S/18S [Harris *et al.* 2003]. Les fonctions associées à ces gènes sont très majoritairement la participation au ribosome et la modification de protéines. On accepte donc qu'un ribosome relativement sophistiqué, comparable aux actuels, existait chez le cenancêtre, organisme qui devait être doué de traduction. Une partie de ces COG constitue le principal cœur fonctionnel de la transcription, donc cette fonction devait exister aussi chez le cenancêtre. En revanche, il y a très peu de protéines intervenant dans la réplication de l'ADN, ce dont on discutera plus bas. Certaines protéines qui interviennent dans l'intégration d'autres protéines dans les membranes sont universelles et ont même servi à enraciner l'arbre du vivant, donc on accepte qu'elles existaient chez le cenancêtre [Gribaldo et Cammarano 1998 ; Philippe et Forterre 1999]. De même, bien que la phylogénie complexe montre de nombreux HGT plus récents, l'existence des gènes paralogues universels, notamment les ATPases, suggère leur présence chez le cenancêtre [Gogarten *et al.* 1989 ; Iwabe *et al.* 1989 ; Mulkidjanian *et al.* 2007]. La présence d'une machinerie d'insertion des protéines dans la membrane et la présence de protéines membranaires (par exemple, les ATPases) suggère que le cenancêtre possédait des membranes, mais aucun mécanisme de synthèse des membranes n'est universel, et ce paradoxe constituera précisément la question principale de ma thèse.

On s'accorde en général pour considérer que ce cœur de protéines universelles est trop réduit pour constituer un cenancêtre fonctionnel. De plus, on sait qu'il y a beaucoup de familles protéiques qui sont présentes dans les trois domaines du vivant sans être universelles, donc on peut concevoir qu'elles aient existé chez le cenancêtre et perdues dans une partie des organismes actuels. Ainsi, on accepte que d'autres protéines aient dû exister chez le cenancêtre, mais l'estimation de la quantité de ces gènes est très difficile à effectuer. Le nombre de gènes que l'on infère chez le cenancêtre dépend de la fréquence que l'on attribue à la perte de gènes par rapport à celle de leur apparition. Si l'on accepte que l'apparition de gènes (et les HGT qui permettent de les échanger entre lignées lointaines) est fréquente, alors

¹⁶ Clusters of Orthologous Genes, un moyen de classer toutes les protéines que l'on connaît dans des grands groupes sur la base de leur homologie [Tatusov *et al.* 1997].

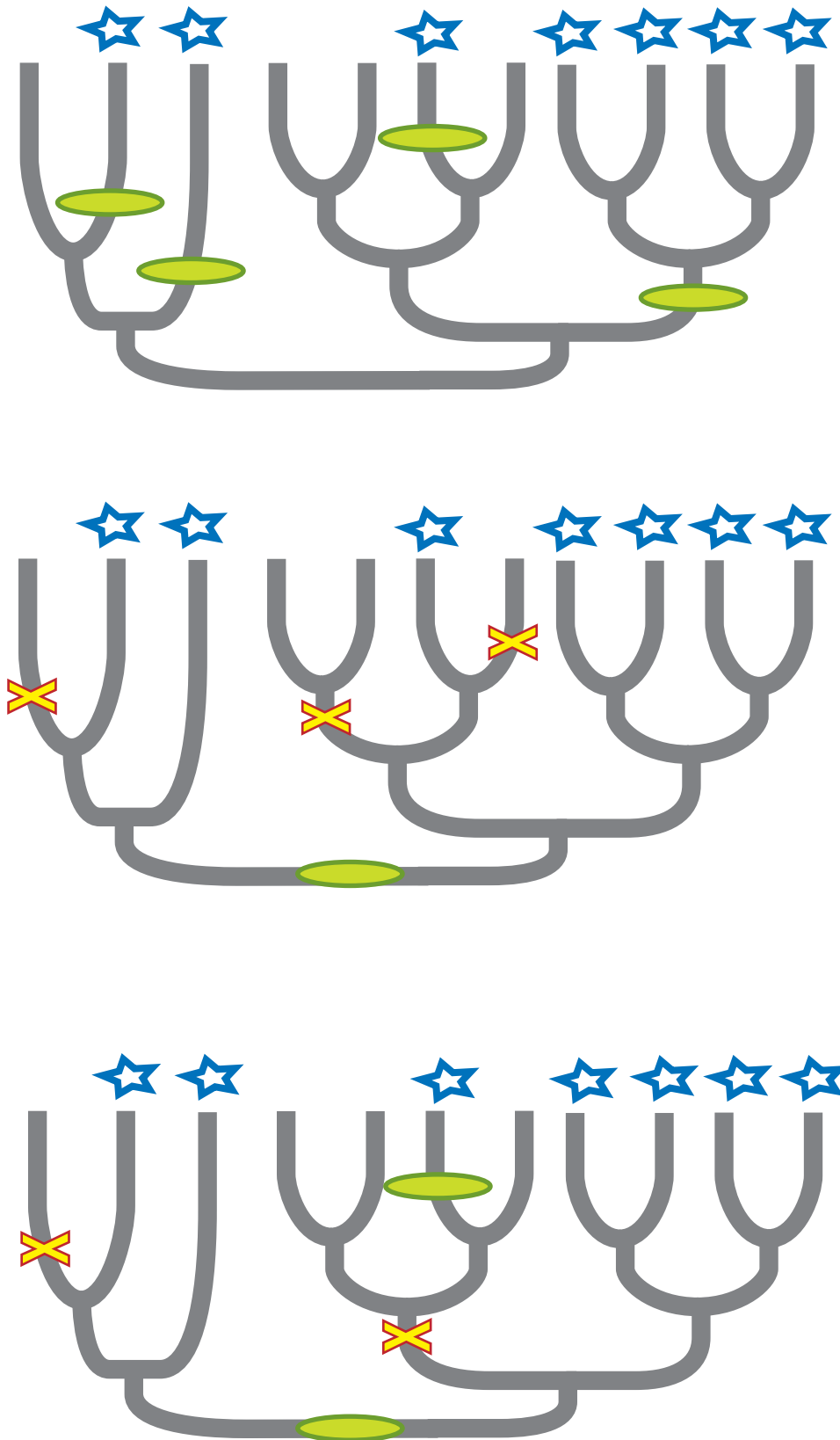


Figure 19. Nombre d'événements que l'on doit formuler pour expliquer la distribution des homologues d'un gène dans une phylogénie des organismes.

Les arbres représentent une phylogénie théorique des organismes, les étoiles représentent des homologues détectées dans les génomes des organismes, les ovals représentent des gains (quelque soit la source du gain) et les croix représentent des pertes. **A.** Hypothèse ne faisant intervenir que des gains. **B.** Hypothèse de faisant intervenir que des pertes. **C.** Hypothèse faisant intervenir une combinaison des deux.

on conclura à un génome ancestral très petit et à une apparition postérieure des gènes modernes. Si, au contraire, on attribue une haute fréquence à la perte de gènes, alors on pourra en déduire la présence chez le cenancêtre de beaucoup de gènes qui sont distribués uniquement dans une partie des organismes actuels, les absences étant expliquées par la fréquence des pertes (Figure 19). Comme on ne connaît pas la valeur relative du nombre de pertes par rapport au nombre de gains au cours de l'évolution, l'équipe de Koonin a mis en place un système capable de déterminer le nombre de gènes chez le cenancêtre pour différents rapports pertes/gains. Cette analyse a montré que si l'on accepte que le nombre de gains et de pertes soit équivalent (1 : 1) alors on peut en inférer la présence de 572 gènes chez le cenancêtre [Mirkin *et al.* 2003]. Selon les auteurs de cette étude, ce nombre constitue une valeur minimale permettant d'assurer les tâches qui semblent indispensables au fonctionnement d'une cellule (métabolisme des acides aminés, des coenzymes, des nucléotides, des glucides, énergétique, chaperones, transport d'ions), mais la question de quelles sont ces tâches essentielles à la vie de la cellule reste encore à l'appréciation du chercheur. En outre, on admet généralement que les pertes soient probablement plus fréquentes que les gains dans la mesure où elles se produisent plus facilement (il est plus aisé de perdre une fonction que de mettre en place un nouveau gène) et où beaucoup de pertes n'ont pas d'impact négatif sur la survie et la valeur sélective des organismes qui portent les gènes [Mirkin *et al.* 2003 ; Pal *et al.* 2005]. Le ratio 1 : 1 est donc indicatif d'une valeur probablement basse de la taille du génome du cenancêtre, en accord avec ceux qui pensent que le cenancêtre avait un génome plutôt petit, encore que non minimal [Koonin 2003]. En postulant une proportion de pertes plus importante, trois méthodes différentes donnent un contenu de gènes dépassant les 1000 gènes, mais le croisement des trois méthodes réduit ce nombre à 669 gènes détectés simultanément par les trois approches [Ouzounis *et al.* 2006], ce qui n'est finalement pas très différent des résultats des travaux de Koonin. Cependant, ces derniers auteurs, tout en décidant de ne discuter que cette limite basse pour des raisons de prudence, plaident pour un cenancêtre qui comporterait déjà la majorité des fonctions cellulaires caractéristiques des cellules modernes.

Génome à ADN ou à ARN

Connaître le contenu en gènes du cenancêtre est important pour se faire une idée générale des fonctions qui pouvaient être présentes dans cet organisme, mais ce contenu ne permet pas de préciser suffisamment son fonctionnement de la même façon qu'il ne suffit pas de connaître le génome d'un organisme moderne pour comprendre la totalité de sa biologie.

En réalité, peu de travaux se sont concentrés sur des fonctions concrètes chez le cenancêtre et le fait que, souvent, seule une partie des gènes intervenant dans un processus puisse être inférée dans cet organisme soulève très facilement le débat.

L'une des questions sur la nature du cenancêtre qui a attiré le plus d'attention est le support de son information génétique. L'ADN constituant le support de l'information génétique dans les organismes des trois domaines du vivant, on pourrait penser qu'il en irait de même pour le cenancêtre. Cependant, comme les machineries de réplication de l'ADN sont très différentes entre les bactéries et les archées/eucaryotes et que celle dont la présence peut être inférée chez le cenancêtre est très incomplète, l'équipe de Koonin a suggéré que le génome ancestral était basé sur l'ARN [Mushegian et Koonin 1996]. Cette proposition a été contestée car la présence même d'une partie de la machinerie de réplication semblait suggérer au moins l'ébauche de cette fonction chez le cenancêtre. De même, les principaux composants de la polymérase qui transcrit l'ADN en ARN font partie du cœur de gènes universels et l'enzyme qui synthétise des déoxynucléotides de l'ADN était également postulée comme ancestrale. Enfin, il a été suggéré qu'un génome à ARN serait peu stable du fait de l'instabilité des grandes molécules d'ARN et qu'il comporterait des très forts niveaux de mutation car on ne connaissait pas de mécanisme de réparation de l'ARN [Becerra *et al.* 1997]. Le fait que les virus à ARN comportent des génomes de très petite taille conforte ces objections.

Ces critiques ont conduit à la comparaison plus détaillée des machineries de réplication et à l'observation qu'une partie de la machinerie était homologue entre bactéries et archées/eucaryotes, une autre partie homologue mais probablement recrutée indépendamment, et une dernière partie sans origine commune [Leipe *et al.* 1999]. Koonin a alors postulé que le cenancêtre pouvait avoir une information génétique mélangée ADN/ARN, dont la réplication ressemblerait beaucoup à la retrotranscription actuelle, i. e. à la copie d'un ADN en ARN suivie de la transcription inverse. Ensuite, les deux machineries de réplication modernes seraient apparues indépendamment dans la lignée bactérienne d'un côté et dans l'archée/eucaryote de l'autre [Leipe *et al.* 1999]. Cette hypothèse est affaiblie par le fait qu'elle suppose l'existence chez le cenancêtre d'une retrotranscriptase dont on ne connaît rien car elle aurait ensuite été remplacée dans les différentes lignées, mais elle reste une possibilité.

Par ailleurs, il a été proposé que le cenancêtre ait possédé une machinerie de réplication de l'ADN mais que celle-ci ait été majoritairement remplacée par une autre provenant d'un virus [Forterre 1999]. Cette hypothèse est séduisante car on connaît plusieurs exemples de protéines informationnelles virales qui semblent avoir remplacé des machineries existantes mais elle souffre aussi d'être très spéculative et de ne pas permettre de déterminer de façon satisfaisante la lignée chez qui le changement a eu lieu. Cette idée a été poussée plus loin en proposant que l'ADN lui-même, et pas seulement sa réplication, pourrait être d'origine virale car des virus auraient pu avoir intérêt à modifier leur information génétique dans un monde ARN pour échapper aux défenses cellulaires [Forterre 2001]. Ensuite, l'ADN et sa machinerie auraient été transférés deux fois aux lignées cellulaires, soit l'une avant le cenancêtre et l'autre après celui-ci, dans une lignée moderne, soit l'une et l'autre dans des lignées après le cenancêtre. Cependant, cette idée est encore plus spéculative et ne permet pas de fournir d'information sur le support génétique du cenancêtre.

D'autres analyses se sont concentrées sur les machineries de réparation, avec des résultats qui restent ouverts. D'une part, il a été montré que la machinerie de réparation de l'ADN pouvait être présente chez le cenancêtre, ce qui soutient l'hypothèse que l'ADN était le support de l'information génétique du cenancêtre [DiRuggiero *et al.* 1999]. D'autre part, une machinerie de réparation de l'ARNm par les principales protéines de la polymérase de la transcription a été identifiée et sa présence inférée chez le cenancêtre [Poole et Logan 2005]. Cette observation a été invoquée pour contrer l'argument selon lequel un génome à ARN comporterait trop d'erreurs car sans réparation, mais elle n'est pas suffisante pour déterminer que l'ARN était le support de l'information génétique chez le cenancêtre car celui-ci contenait aussi de l'ARNm, comme les cellules modernes. Finalement, la discussion a porté sur les ribonucléotides réductases (RNR) qui synthétisent les déoxynucléotides, car les trois classes qui existent sont homologues [Stubbe 2000], mais leurs phylogénies sont marquées par les HGT qui ne permettent ni d'affirmer ni d'infirmer la présence de l'un de ces gènes chez le cenancêtre [Lundin *et al.* 2010].

La question de la nature du support de l'information génétique chez le cenancêtre (ADN ou ARN) reste donc ouverte. Malgré cette incertitude, on voit que l'analyse directe de l'évolution des gènes qui interviennent dans un processus cellulaire permet non seulement de tenter de fournir des descriptions concrètes du fonctionnement du cenancêtre [Koonin 2003] mais surtout de penser le cenancêtre sans faire intervenir des idées *a priori* sur les caractères qui doivent être considérés comme étant ancestraux [Snel *et al.* 2002 ; Schoepp-Cothenet *et*

al. 2012]. C'est ce genre d'approche favorisant l'étude approfondie de l'évolution d'une structure cellulaire que j'ai essayé d'appliquer dans mon travail. Les structures sur lesquelles j'ai concentré mon attention ont été les membranes cellulaires.

Le cenancêtre était-il entouré par des membranes lipidiques ?

Toutes les cellules sont délimitées par des membranes plasmiques et un grand nombre d'organismes possède aussi des membranes internes qui délimitent des organites. Toutes les membranes sont composées d'une matrice lipidique dans laquelle s'insère un grand nombre de protéines. Les lipides majoritaires de ces membranes sont les phospholipides. Bien que les phospholipides existent dans tous les organismes, il en existe deux types, l'un caractéristique des bactéries et des eucaryotes et l'autre caractéristique des archées. Ces différences entre phospholipides sont corrélées avec des différences dans leur synthèse, ce qui a conduit certains auteurs à questionner la présence d'un mécanisme de synthèse des phospholipides chez le cenancêtre et par conséquent aussi à mettre en question la présence même de membranes chez le cenancêtre. Cette supposition a des implications structurelles mais elle affecte aussi directement la conception que l'on a du cenancêtre : le dernier ancêtre commun de toutes les cellules pouvait-il ne pas être une cellule ?

Dans la partie qui suit je vais exposer comment notre conception des membranes a pu changer notre vision de la cellule, que ce soit dans les organismes présents lorsqu'on a essayé de comprendre le fonctionnement de la cellule, ou dans les organismes passés lorsqu'on tenté de comprendre le fonctionnement du cenancêtre. Ensuite, mon travail de phylogénie moléculaire se concentrera sur les mécanismes de synthèse des phospholipides afin de déterminer si une synthèse enzymatique de phospholipides pouvait exister chez le cenancêtre et à quoi elle pouvait ressembler ;

Deuxième partie : les membranes lipidiques

« *Donnez-moi une vésicule organique douée de vitalité, et je vous rendrai tout le monde organisé* »

François-Vincent Raspail,

Nouveau système de physiologie végétale et de botanique, 1837

“If it can be shown that the existence and function of plasma membranes is essential to life of the protoplast, it follows that it is as incorrect to think of protoplasm as a living substance apart from its organization into protoplasts as it would be to think of proteins, or water, or carbohydrates as living unless organized into protoplasm. In that case, the protoplast, rather than protoplasm, becomes ‘the physical basis of life’.¹⁷”

Janet Q. Plowe, *Membranes in the plant cell*, 1931.

¹⁷ « Si l'on démontrait que l'existence et la fonction des membranes plasmiques sont essentielles à la vie du protoplaste [la cellule], il s'ensuivrait qu'il serait faux de considérer le protoplasme [cytoplasme] comme une substance vivante en dehors de son organisation dans les protoplastes comme il est faux de penser que les protéines, ou l'eau, ou les carbohydrates sont vivants à moins de faire partie du protoplasme. Dans ce cas-là, le protoplaste, plutôt que le protoplasme, deviendrait la « base physique du vivant ». » (Ma traduction).

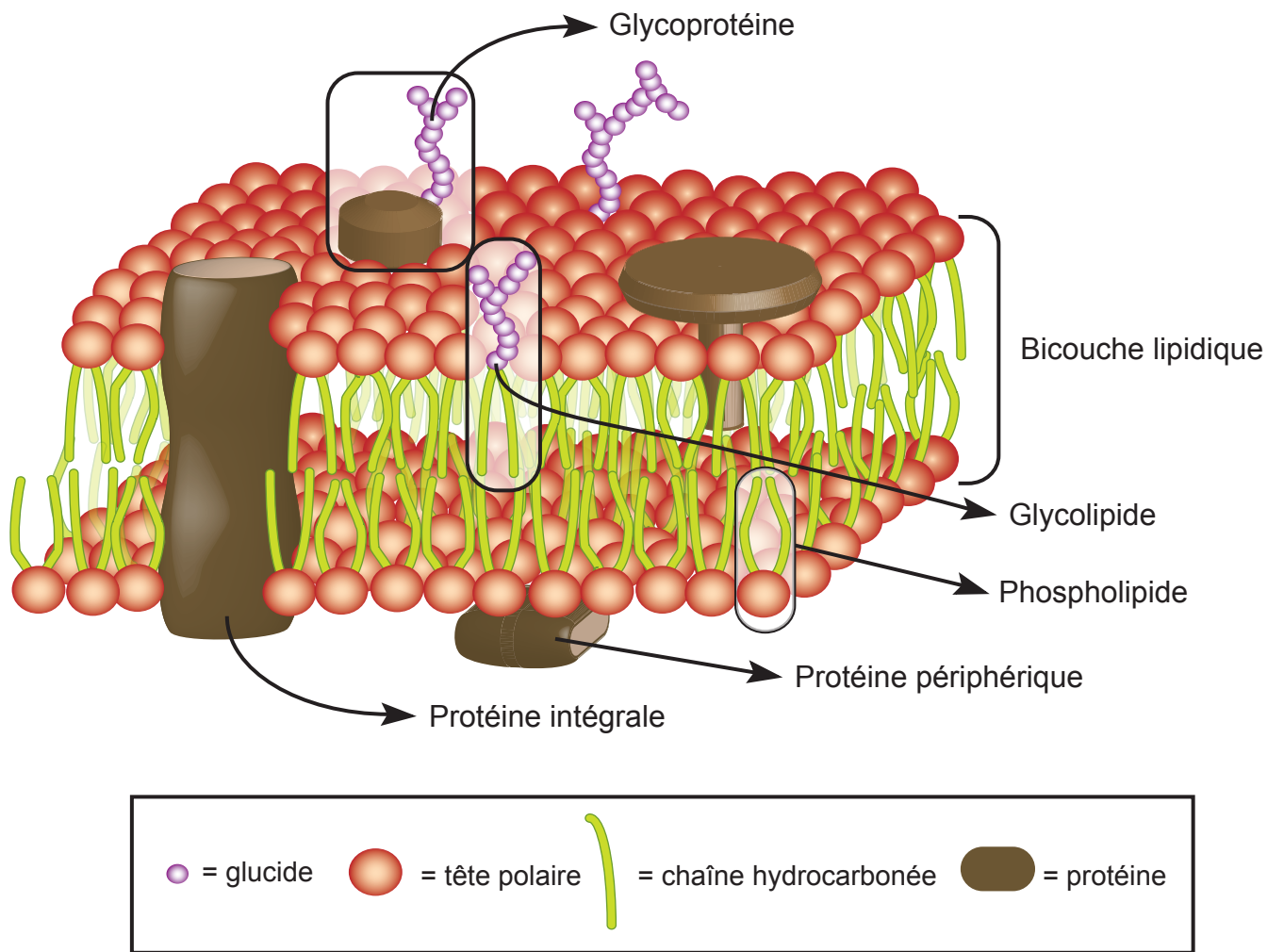


Figure 20. Modèle membranaire de la mosaïque fluide.

1. Importance physiologique et structure des membranes cellulaires

Toutes les cellules sont délimitées par un film que l'on appelle la membrane plasmique. Celle-ci joue un rôle de frontière entre la cellule et son environnement en même temps qu'elle constitue un compartiment cellulaire. Dans le premier rôle, la membrane plasmique régule l'échange de molécules avec l'extérieur, ce qui permet de conserver des conditions osmotiques et de pH, de concentrer les nutriments, d'exclure les déchets et de créer des gradients ioniques transformables en énergie chimique : on dit que les membranes sont sélectives. Elles permettent aussi la transduction de signaux depuis et vers l'extérieur. Dans le second rôle, les membranes agissent comme des compartiments car leurs composants forment un contexte chimique apolaire qui concentre des pigments et des protéines hydrophobes essentielles à de nombreuses voies métaboliques et énergétiques [Campbell et Reece 2004 ; Berg *et al.* 2007b].

Le modèle communément accepté aujourd'hui pour décrire la structure des membranes cellulaires est celui qui a été proposé par Singer et Nicolson en 1972, et que l'on appelle « modèle de la mosaïque fluide » (Figure 20). Les membranes sont composées de lipides, de protéines et, dans une moindre mesure, de glucides. Les lipides les plus abondants sont les phospholipides, qui sont des molécules amphiphiles, i. e. comportant à la fois une partie polaire attirée par l'eau (tête hydrophile) et une partie apolaire qui fuit l'eau (queues hydrophobes). Énergétiquement, la structure la plus stable est celle qui réduit au maximum le contact des parties hydrophobes avec l'eau environnante tout en permettant aux parties polaires d'interagir avec elle [Tanford 1973]. Du fait de ces contraintes thermodynamiques, les phospholipides membranaires forment une bicouche dans laquelle les queues hydrophobes se font face à l'intérieur, pendant que les têtes hydrophiles interagissent avec l'eau de part et d'autre de cette membrane (Figure 20). À cause de ces mêmes contraintes, certaines protéines, dites intégrales ou intrinsèques, s'insèrent dans la matrice bidimensionnelle de phospholipides en établissant des liaisons hydrophobes avec le centre de la membrane, pendant que les parties polaires ressortent au contact de l'eau du côté interne, externe ou des deux. D'autres protéines, dites périphériques ou extrinsèques, peuvent s'associer à la surface des membranes à travers des liaisons faibles transitoires (Figure 20).

Le terme *mosaïque* fait référence au mélange de lipides et protéines intrinsèques qui composent le plan de la membrane. On dit que cette membrane est *fluide* car ses composants

peuvent diffuser au sein d'une même bicouche, ce qui permet à la fois des mouvements stochastiques et la formation par endroits de regroupements spécifiques. Malgré la diffusion latérale des composants membranaires, le passage de phospholipides d'une couche à l'autre est énergétiquement défavorable et nécessite des protéines intégrales spécifiques (flippases) capables de faciliter la traversée de la tête polaire à travers le centre hydrophobe de la membrane [Daleke et Lyles 2000]. Pour les mêmes raisons, les protéines qui ne sont pas attachées à une structure cellulaire (comme par exemple au cytosquelette chez les eucaryotes) peuvent diffuser dans la bicouche mais changeront rarement leur polarité de part et d'autre de la membrane. L'incapacité de ces molécules à changer de couche confère son caractère *asymétrique* à la membrane : la composition des deux couches est différente au niveau à la fois de l'orientation des protéines, de la nature des lipides et de la distribution de glucides qui sont responsables des interactions avec l'extérieur et se trouvent donc majoritairement attachés aux phospholipides et protéines de la couche externe. Des composants lipidiques autres que les phospholipides (tels la chlorophylle ou le cholestérol) jouent des rôles importants comme pigments ou régulateurs de la fluidité de la membrane.

Le modèle de Singer et Nicolson est l'archétype enseigné depuis les années 1970 dans les manuels [Campbell et Reece 2004]. Bien entendu, il a été raffiné depuis sa première formulation mais il continue de fournir le cadre de réflexions, d'interrogations et de méthodes, autrement dit le *paradigme* dans lequel se développent toutes les recherches modernes sur les membranes¹⁸. Il est donc aujourd'hui difficile de concevoir les membranes biologiques autrement qu'à la lumière de ce modèle, mais il est intéressant de souligner que sa naissance ne fut pas immédiate et qu'elle fut liée à un bouleversement radical (une révolution scientifique *sensu* Kuhn 1962) de la compréhension des limites cellulaires¹⁹. Comment a-t-on

¹⁸ La définition de paradigme est souvent ambiguë et peut s'avérer source de polémique [Marris 2008]. On se référera ici au concept avancé par Kuhn en 1962 et défini ici dans le texte.

¹⁹ La théorie sur les révolutions scientifiques de Kuhn développe l'idée que l'histoire des connaissances ne serait pas forcément linéaire dans le sens d'un progrès, mais plutôt un mélange de phases cumulatives et de discontinuités réinterprétant les données disponibles [Kuhn 1962]. Elle a été formulée par induction à partir d'observations de l'histoire des sciences, surtout de la physique et la chimie. Bien que la validité de cette théorie dans les autres domaines scientifiques ait été proposée dès le départ par son auteur, certains travaux ont soulevé les difficultés d'application de cette théorie à la biologie [Wilkins 1996]. Je pense cependant que l'histoire des recherches sur les membranes constitue un exemple remarquable d'application de la théorie de Kuhn aux sciences de la vie. J'invite le lecteur à observer dans le chapitre « Une histoire des membranes plasmiques » les principales phases selon l'analyse de Kuhn, que je résumerais succinctement ainsi : (1) passage d'une étape « pre-scientifique » à une science « mûre » qui constitue le socle de départ commun aux recherches dans un

conçu les bords cellulaires entre le moment de la formulation de la théorie cellulaire en 1839 et l'établissement du modèle de la mosaïque fluide en 1972 ? Avant de m'intéresser aux membranes actuelles et à ce qu'elles peuvent nous apprendre sur les membranes du cenancêtre, je vais traiter brièvement les difficultés historiques qui ont existé dans la formulation du modèle. Ceci nous permettra non seulement de comprendre l'importance du concept de mosaïque fluide mais aussi d'illustrer les difficultés qui existent pour penser une alternative biologique aux membranes lipidiques.

2. Une histoire des membranes plasmiques.

Il est aujourd'hui inconcevable de penser qu'une cellule moderne ne soit pas délimitée par une membrane plasmique comme celle que nous avons décrite plus haut [Tanford 2004]. Pourtant, au XIX^{ème} siècle la nature de la délimitation des cellules était inconnue et, une fois démontrée l'existence de la membrane et la nature chimique de ses composantes, il a encore fallu plusieurs décennies pour établir le paradigme actuel. Je vais résumer ici comment ces découvertes se sont produites.

De l'existence des membranes plasmiques

Le terme de cellule fut utilisé pour la première fois en biologie en 1665 par Robert Hooke, un précurseurs des premiers microscopes [Hooke 1665]. En regardant des coupes très fines de liège, il observa qu'elles étaient composées d'un grand nombre de cavités géométriques séparées par des parois épaisses qu'il compara aux cellules d'une ruche d'abeille²⁰. Les observations de Hooke étaient principalement descriptives et ne portaient pas

domaine : le premier paradigme. Cette transition est compliquée et a été décrite en détail dans des travaux spécifiques [Bachelard 1938] ; (2) science « normale », cumulative, qui répond aux énigmes fournis par le paradigme avec les outils autorisés par celui-ci ; (3) découverte éventuelle d'anomalies et de résultats, qui conduisent au réajustement de la théorie ; (4) l'accumulation d'anomalies ou de versions contradictoires des théories cristallisent dans une mise en question de l'ensemble (crise) ; (5) révolution scientifique conduisant à la proposition d'une nouvelle théorie qui peut devenir un nouveau paradigme. L'augmentation de la précision explicative de la théorie, la formulation de prédictions qui devront être confirmées et même une sensation subjective d'élégance peuvent favoriser l'adoption de la nouvelle théorie ; (6) une fois qu'un nouveau paradigme a été accepté, une nouvelle science « normale » se développe au sein de ce cadre conceptuel.

²⁰ On dit très souvent que Hooke a nommé les cellules de par la ressemblance qu'il aurait trouvée entre ses observations et les cellules des monastères ou des prisons [Campbell et Reece 2004]. Cependant le texte original ne fait pas de référence à des telles comparaisons. Au contraire, il dit : « I took a good clear piece of Cork, and [...] I could exceeding plainly perceive it to be all perforated and porous, much like a Honey-comb, but that the pores of it were not regular; [...] Next, in that these pores, or cells [...] » [Hooke 1665].

sur le rôle, la nature ou l'identité de ces structures. La théorie cellulaire moderne, qui définit la cellule comme l'unité morphologique et fonctionnelle de tous les organismes vivants, est née au XIX^{ème} siècle. Mais la naissance de la théorie cellulaire est un sujet compliqué et très riche qui dépasse la portée de mon travail sur les membranes. Je résumerai donc très brièvement la genèse de la théorie cellulaire en me basant sur l'analyse publiée par Harris en 1999, uniquement afin de donner des repères historiques pour l'étude de l'histoire du concept de membranes biologiques²¹. On peut déterminer une première phase au XVIII^{ème} siècle pendant laquelle ont été accumulées des données physiologiques et microscopiques concernant principalement les végétaux, parfois les animaux et très rarement des « infusoires » (organismes unicellulaires). A cette époque, la diversité des observations et des interprétations a suscité un débat entre les auteurs qui pensaient que la matière biologique était composée de « globules », de « fibres », de « cylindres tortueux » ou d'un mélange de toutes ces structures.

Au tournant du XIX^{ème} siècle, l'accumulation d'observations avait conduit la plupart des botanistes à accepter l'idée que les tissus végétaux étaient formés de cellules. L'amélioration des techniques de fixation du matériel animal a permis d'étendre progressivement ce principe aux animaux et de constituer l'idée que les cellules étaient les structures universelles qui formaient tous les organismes vivants. Autour des années 1820, Dutrochet et d'autres auteurs ont commencé à ne plus considérer les cellules comme des composantes des organismes, mais les organismes comme le résultat de l'activité cellulaire. Cette différence modifie le sujet des études physiologiques et permet la comparaison de cellules et de fonctions appartenant à des organismes et des tissus très différents. Les travaux postérieurs ont été marqués par l'importance grandissante que l'on a attribuée au noyau cellulaire car il ne faut pas oublier que dans ces travaux les organismes utilisés étaient principalement des eucaryotes. En 1839,

²¹ Bien que la théorie cellulaire soit un des sujets fondateurs de la biologie moderne, comparable à la théorie de l'hérédité et à l'évolution darwinienne, relativement peu de travaux d'histoire et de philosophie des sciences ont essayé à ma connaissance de traiter ce sujet en profondeur. La plupart des analyses historiques se trouvent dans des introductions trop succinctes et par conséquent s'arrêtent à des descriptions superficielles qui entretiennent une historiographie fragmentaire qui souvent surévalue l'apport des travaux des figures les plus connues : Schleiden, Schwann et Virchow. Pourtant, la conception moderne de la théorie cellulaire est née d'innombrables contributions expérimentales et théoriques qui ont précédé et continué après ces auteurs. Je pense que le livre de Harris 1999, qui essaye de replacer historiquement la plupart des travaux influents qui ont conduit à la théorie cellulaire moderne, peut être une bonne base pour entamer ce sujet, mais je crois qu'une enquête plus approfondie des textes originaux serait d'un grand intérêt afin de mieux comprendre les implications des observations à leur époque et leur rôle dans la construction de la biologie contemporaine.

Schwann publie une synthèse qui est souvent retenue comme la première formulation de la théorie cellulaire: la cellule est une structure *individuellement vivante* qui constitue l'unité morphologique et fonctionnelle du vivant. Le grand mérite de ce travail a été de fonder un programme de recherche qui a été la base d'une bonne partie des travaux du XIX^{ème} siècle [Duchesneau 1987]. Comme nous le verrons plus tard, ce travail a été très influencé par les travaux de Schleiden qui attribuaient la formation des cellules à une sorte de cristallisation dans un liquide préexistant formant d'abord le noyau cellulaire, puis progressivement le reste de la cellule. Cette conception a ouvert une période de grande confusion car malgré l'avis contraire de Schwann, certains voyaient là une possibilité de génération spontanée et les analyses histologiques étaient à la fois imprécises et parfois en désaccord avec cette hypothèse. Un changement radical s'est produit à partir des années 1850 lorsque l'hypothèse de la division cellulaire a commencé à être plus généralement acceptée et que les premières études sur la mitose se sont généralisées à la fin du XIX^{ème} siècle.

Comment concevait-on les bordures cellulaires lors des premières formulations de la théorie cellulaire ? Le travail d'Henri Dutrochet à ce propos [Dutrochet 1824] est un antécédent intéressant. Non seulement il a étudié l'universalité des cellules mais ses travaux sur l'entrée et la sortie de l'eau dans les cellules font de lui un des pionniers des recherches sur l'osmose [Dutrochet 1828]. Il considérait que les cellules étaient des structures individuelles (ou que l'on pouvait les séparer par des traitements chimiques) et qu'elles étaient toutes entourées par une membrane ou paroi qui leur était propre²². En plus de constituer des unités structurelles, les cellules étaient à l'origine des fonctions physiologiques de l'organisme : la cellule devait acquérir et transformer des substances à l'extérieur, puis les « sécréter » soit à l'intérieur soit à l'extérieur de la cellule (aujourd'hui on parlerait plutôt de métabolisme que de sécrétion). Etant donné que les substances extérieures devaient forcément traverser la paroi avant d'être « sécrétées » de part et d'autre de celle-ci, Dutrochet considérait logiquement que les membranes étaient le siège de ce que l'on appellerait aujourd'hui le métabolisme, ce qui leur conférait une grande importance.

²² On remarquera que pendant longtemps il a existé une certaine confusion entre les termes « membrane » et « paroi ». Aujourd'hui on parle de membrane dans le cas des bicouches lipidiques décrites par le modèle de la mosaïque fluide, alors que la paroi est une structure extracellulaire plus rigide et de composition chimique diverse dans les différents groupes taxonomiques.

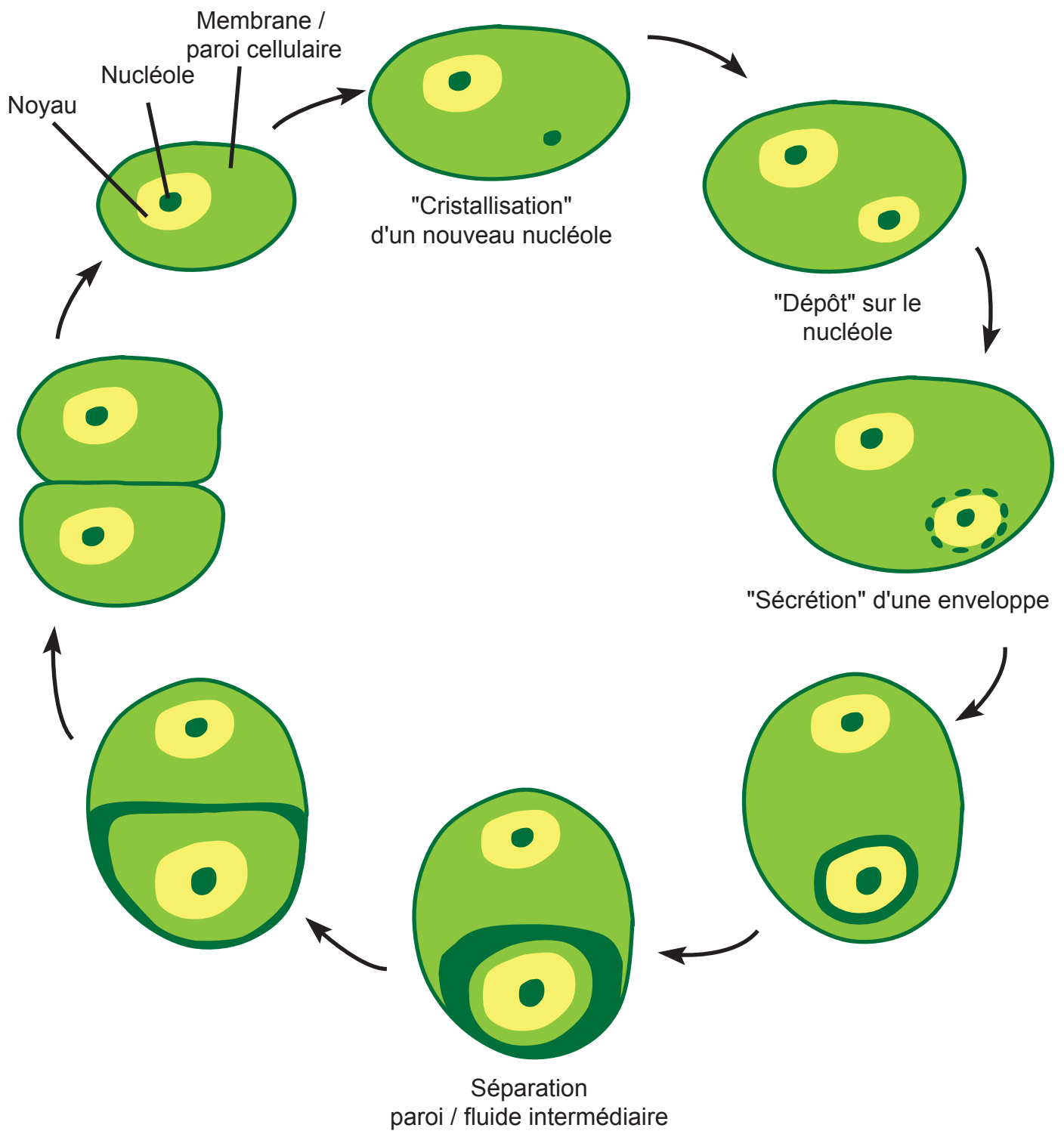


Figure 21. Modèle proposé par Schleiden et Schwann pour expliquer la division cellulaire.

Pour Schleiden et Schwann, les cellules étaient composées d'un fluide entouré d'une paroi responsable de nombreuses réactions chimiques et métaboliques. Mais ces auteurs ont surtout attiré l'attention par leur interprétation du noyau et du nucléole, dont les observations se sont popularisées dans les années 1830. Schleiden avait observé que le noyau était la première structure observable chez de jeunes cellules végétales et il en avait conclu que la cellule se formait à partir de celui-ci [Schleiden 1837]. Lors de l'apparition de nouvelles cellules, les premières choses que l'on voyait étaient des « points » (Schwann les nommera ensuite nucléoles) dans un milieu granuleux. Le dépôt de molécules sur les nucléoles permettait la formation du noyau, puis l'accrétion continuait jusqu'à atteindre une épaisseur permettant la ségrégation d'une paroi et d'un fluide intermédiaire entre le noyau et la paroi (Figure 21). Schwann a étendu l'hypothèse de Schleiden aux animaux [Schwann 1839]. Il a proposé que les nouvelles cellules se formaient à partir du *blastoplasme*, un fluide qui pouvait être soit à l'intérieur des cellules, comme dans le modèle de Schleiden, soit dans l'espace interstitiel. Le blastoplasme fournissait les substances qui cristallisaient d'abord en un nucléole, puis en un noyau et enfin en l'ensemble de la cellule, comme dans le modèle de Schleiden (Figure 21).

Ainsi ces auteurs attribuaient une grande importance aux membranes mais la recherche de leurs contemporains s'est plutôt concentrée sur les polémiques autour de l'importance du noyau pour la formation de nouvelles cellules. En plus de ce décalage de l'intérêt de la communauté des histologistes vers le noyau, la deuxième moitié du XIX^{ème} siècle a connu la mise en cause systématique de l'existence universelle des membranes. D'abord, les membranes cellulaires sont des structures très fines (~8 nm) et par conséquent difficiles à observer avec le matériel biologique le plus souvent utilisé à l'époque. Les cellules végétales sont entourées d'une épaisse paroi externe qui rend difficile la visibilité des membranes plasmiques. Quant aux cellules animales, on observait souvent des tissus compliqués comme ceux des muscles ou du système nerveux dont l'interprétation et la comparaison avec d'autres tissus était difficile. Le contraste entre la facilité à voir les séparations entre les cellules végétales (à cause des parois) et la difficulté à trouver des équivalents chez les animaux et certains organismes unicellulaires a conduit certains auteurs à penser que les « membranes » étaient en réalité des structures rigides qui empêchaient le mouvement et ont donc conclu que les cellules capables de mouvement étaient dépourvues de telles structures (Reichert 1841,

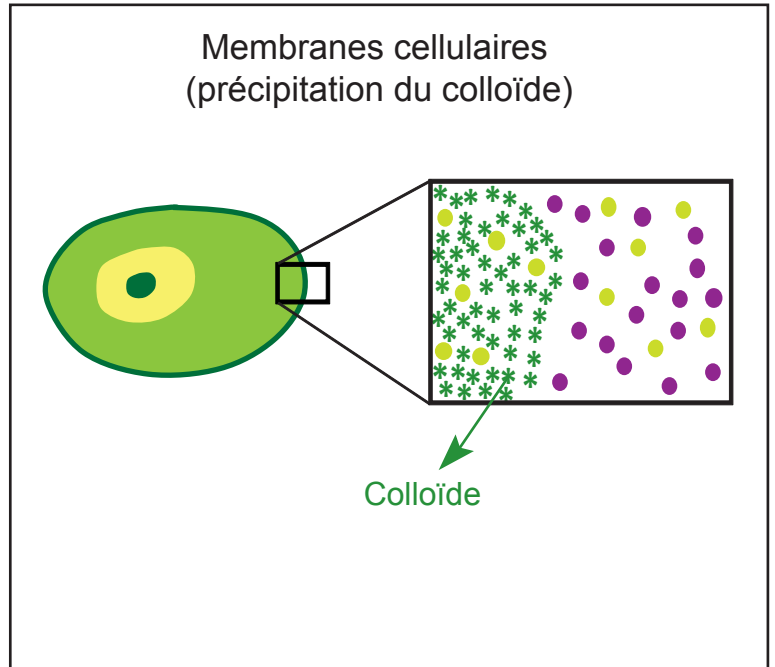
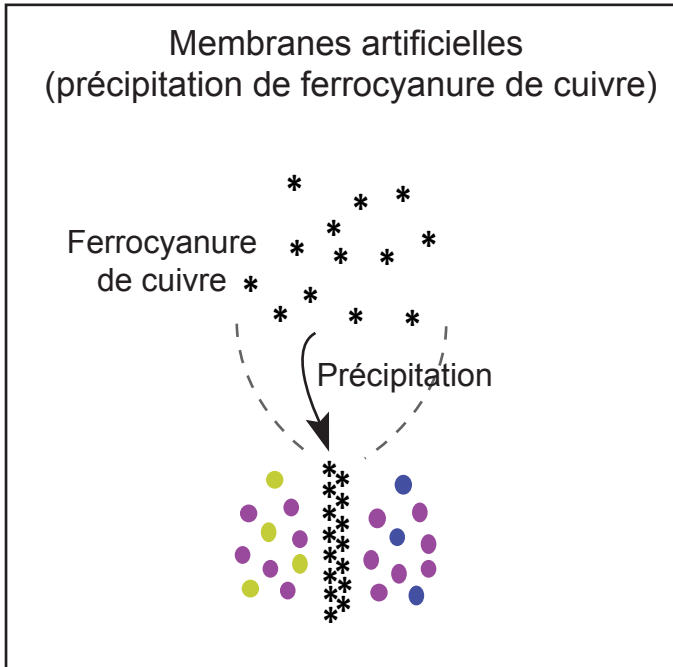


Figure 22. Comparaison entre les membranes artificielles en ferrocyanure de cuivre et les membranes cellulaires à la fin du XIX^{ème} siècle.

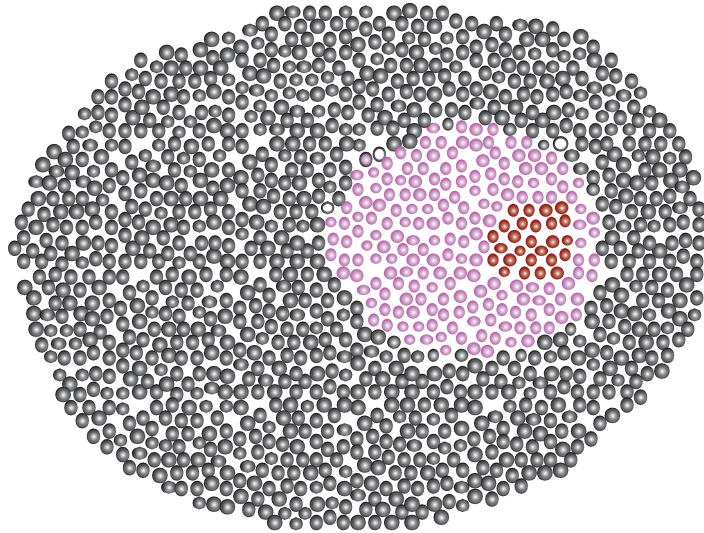
Ecker 1848, de Bary 1864, Haeckel 1866²³). Cette idée a d'ailleurs été soutenue par certains observateurs de protistes qui ont observé plusieurs noyaux à l'intérieur d'organismes que l'on sait aujourd'hui être unicellulaires mais qui pouvaient apparaître à l'époque comme des agglomérations de cellules nucléées mais sans membrane (de Bary 1864). L'idée que les membranes étaient des structures secondaires ou accessoires a alors connu un essor mais le sujet est resté polémique. Remak, qui a été parmi les premiers à étudier la division cellulaire de façon systématique et à militer en faveur de l'universalité de ce mécanisme plutôt que de celui annoncé par Schleiden et Schwann, avait notamment basé son travail sur les premières phases de division de l'œuf fécondé. A cette occasion, il avait traité les œufs avec différents agents chimiques qui lui ont permis d'observer des membranes, raison pour laquelle il a soutenu l'idée que les cellules étaient entourées d'une membrane qui n'était pas forcément rigide (Remak 1855). Le débat ne fut pas clos pour autant, car on a proposé que les structures que l'on voyait apparaître avec la technique chimique de fixation des échantillons de Remak étaient le résultat de la précipitation ou la dégénérescence du matériel cellulaire à cause du traitement chimique. Les membranes pouvaient n'être qu'un artefact dû à l'expérience (Schultze 1860-1863) ou encore une partie morte résultant de l'activité des cellules (Beale 1861). L'alternative avancée par les tenants de ces critiques contestant l'existence de membranes était que la limite cellulaire observable pouvait être juste l'interface entre le fluide cellulaire et le milieu extérieur, sans qu'une structure particulière soit nécessaire pour séparer les deux [Turner 1890].

En dehors des arguments histologiques, les premiers travaux sur la perméabilité des membranes ont aussi encouragé ce débat. La première séparation de membranes et parois végétales fut réalisée grâce à des traitements chimiques (von Mohl 1851)²⁴ mais elle ne permettait pas de trancher sur la nature des membranes (structure biologique ou artéfactuelle). Les avancées sur l'osmose dans les cellules végétales ont alors été importantes : lorsque la cellule perd de l'eau, elle perd aussi du volume et entraîne sa membrane ; celle-ci se sépare de la paroi, plus rigide, qui reste en place (Nägeli 1855). On pouvait donc visualiser les bords cellulaires mais on n'en apprenait pas plus sur leur composition. Cette méthode fut surtout

²³ Ces textes étant écrits en allemand, je me suis basé ici sur les interprétations du livre de Harris 1999. Se référer à ce livre pour avoir les références correspondantes.

²⁴ Cette partie a également été étudiée principalement par des allemands. Dans le texte je cite les documents originaux de ces auteurs, mais je me suis référé ici à certains textes en anglais qui ont fait des analyses de ces études : [Loeb 1906 ; Jacobs 1962 ; Smith 1962].

Absence de membranes différenciées
(limites du colloïde)



Membranes différenciées
du milieu intérieur

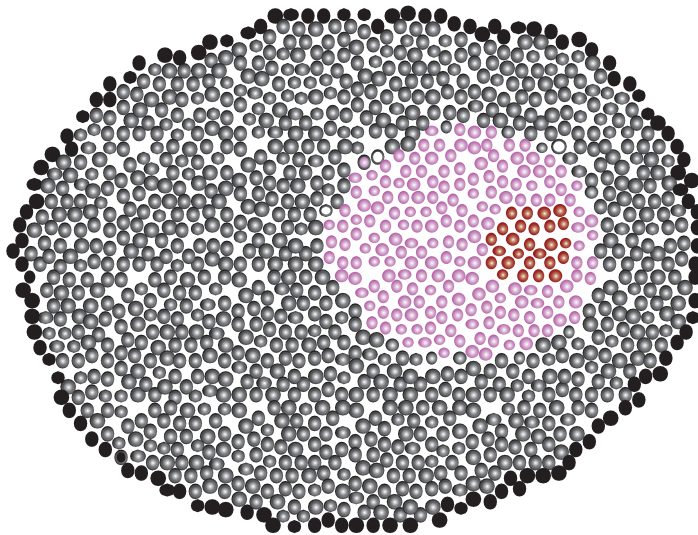


Figure 23. Représentation des visions qui s'affrontaient à la fin du XIX^{ème} siècle à propos de la nature des délimitations cellulaires.

importante pour comprendre que la membrane cellulaire se comportait de la même façon que les barrières semi-perméables artificielles et qu'on pouvait donc voir la cellule comme une cavité osmotique (Nägeli 1855, Traube 1867, de Vries 1884, Pfeffer 1890, Hamburger 1895).

Harris dit que le fait que des travaux sur l'osmose des cellules se soient développés à la fin du XIX^{ème} siècle montre qu'au tournant du siècle il n'existait plus de doute sur l'existence d'une membrane entourant la cellule [Harris 1999]. Toutefois, je crois que l'avis sur cette question à la fin du XIX^{ème} était en réalité plus complexe que cela, en partie pour des raisons méthodologiques. A cette époque le modèle courant de membranes artificielles était celui formé par la précipitation de ferrocyanure de cuivre dans une interface (Figure 22). Il a donc semblé naturel de proposer que l'intérieur de la cellule, formé par une suspension de protéines, glucides, lipides et sels (le colloïde), précipitait au contact du milieu aqueux environnant et que la barrière cellulaire en était le résultat (Nägeli 1855, Traube 1867, Pfeffer 1877). Et là encore deux visions se sont alors affrontées [Jacobs 1962] : l'une considérait que des membranes de précipitation existaient autour de la cellule avec des propriétés comparables aux membranes semi-perméables artificielles (Pfeffer 1877), l'autre soutenait qu'il n'existait pas de membrane autour des cellules et que les propriétés déduites à la surface des cellules étaient les propriétés du cytoplasme lui-même (Fischer 1921, Lepeschkin 1924, Figure 23).

Il me semble plus prudent de penser que la question de l'existence de la membrane cellulaire était déjà au XIX^{ème} siècle un sujet circonscrit à une petite communauté et que l'attention au tournant du XX^{ème} siècle s'étant tournée vers d'autres sujets (notamment avec la redécouverte du mendélisme, les débats évolutionnistes et la description de plus en plus détaillée de la biochimie et de la mitose cellulaire), l'existence d'une membrane autour des cellules a été plus ou moins tacitement acceptée comme une hypothèse de travail opérationnelle. Il a fallu en réalité attendre un quart de siècle, jusqu'à la mise en place de la technique des microinjections par Chambers, pour pouvoir lever cette hésitation.

Cette technique utilise une micropipette qui permet d'appliquer un marqueur de couleur ou une molécule cytolytique (qui dégrade le cytoplasme) de façon très précise. D'une part, Chambers a appliqué des composés du côté extérieur de cellules animales et observé que les composés diffusaient dans le milieu extérieur de la cellule mais n'entraient pas à l'intérieur ; d'autre part, il a microinjecté ces molécules à proximité du côté intérieur des

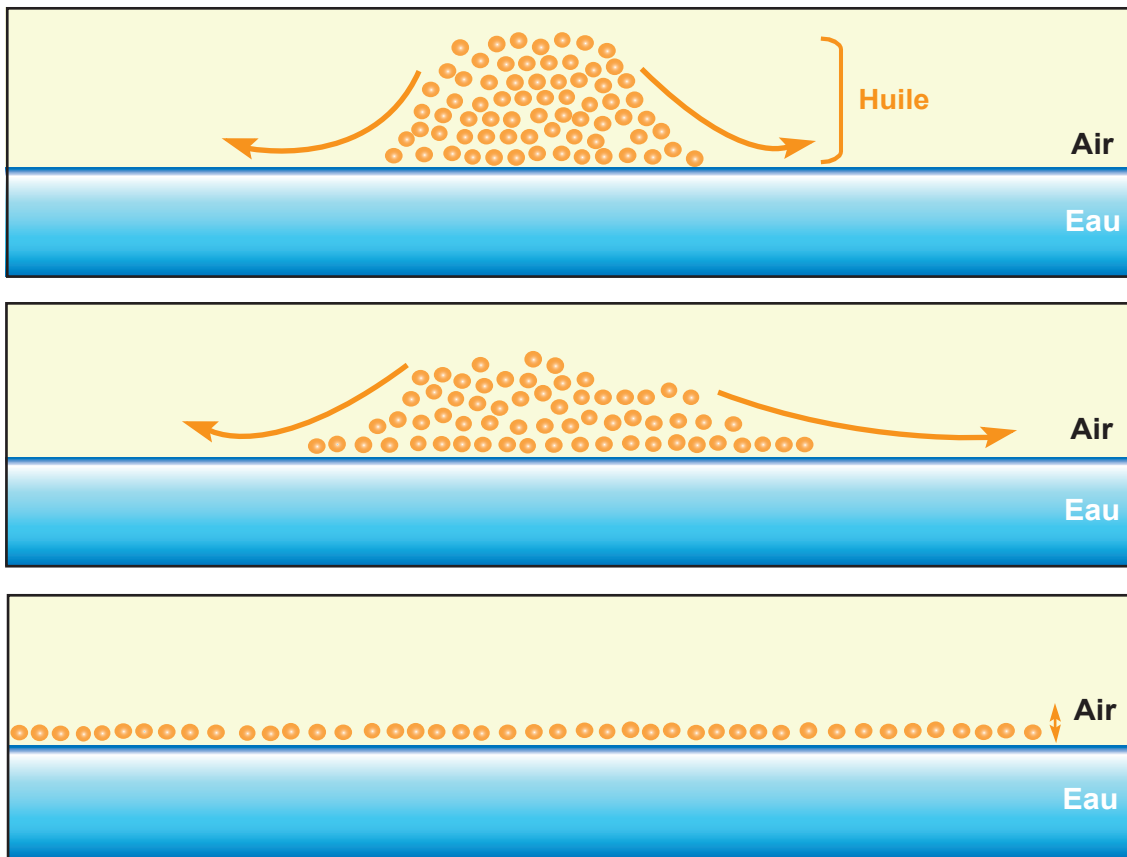


Figure 24. L'huile se répand sur l'eau jusqu'à former une couche de seulement une molécule de hauteur.

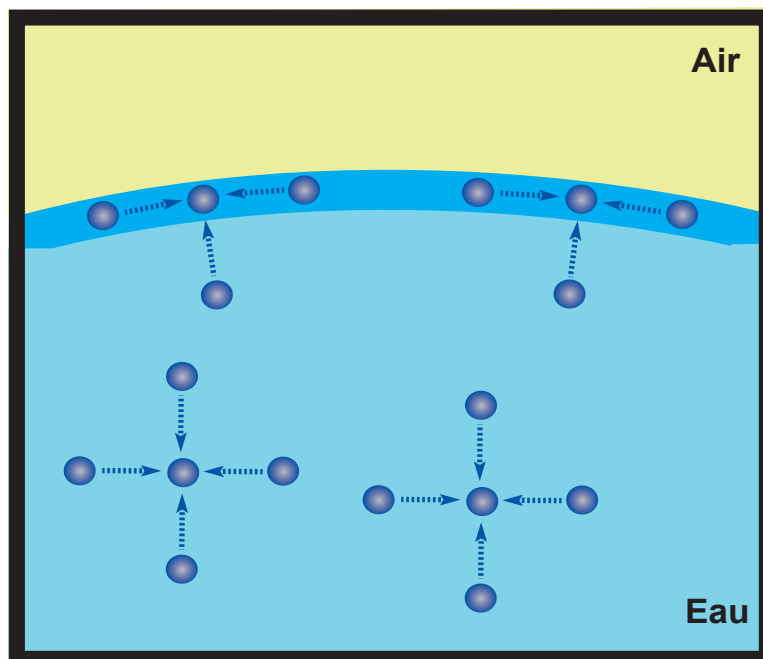


Figure 25. Tension superficielle.

Les molécules d'eau qui se trouvent sur la surface ont des interactions avec leur voisines mais pas avec l'extérieur. Cette cohésion se traduit par une certaine résistance qui n'existe pas à l'intérieur du liquide.

bords cellulaires et montré que les molécules, tout en diffusant à l'intérieur de la cellule, ne sortaient pas de celle-ci et s'arrêtaient à la surface. Les cellules animales étaient donc délimitées par une structure dont les propriétés étaient différentes du reste du cytoplasme [Chambers 1922]. Il faudra encore attendre quelques années pour que soit développée une technique applicable aux cellules végétales, donnant des résultats similaires [Plowe 1931]. Presqu'un siècle s'est donc écoulé entre la formulation de la théorie cellulaire et la démonstration de l'existence des membranes cellulaires par Chambers et Plowe.

Les premiers modèles de structure membranaire

Les propositions initiales sur la composition des membranes peuvent être difficilement considérées autrement que comme des convictions éclairées. Les premiers résultats dont on tiendra compte ici sont ceux d'Ernest Overton qui a publié en 1899 une étude sur la perméabilité de la membrane à plus de 500 composés. Overton observe que les composés qui se solubilisent dans l'éther (apolaires) traversent plus facilement les membranes biologiques que les composés qui se solubilisent dans l'eau (polaires). Il interprète son résultat en postulant que les membranes elles-mêmes sont composées de substances apolaires qui dissolvent et permettent la diffusion d'autres molécules apolaires mais non pas des polaires. Il tente ensuite de proposer une composition des membranes en accord avec ces résultats : il exclut les acides gras libres car ils seraient saponifiés dans les conditions de vie des cellules, et propose que les phospholipides et le cholestérol soient les principaux composants des membranes [Overton 1899]. La suite des études sur la structure des membranes s'est donc construite sur le comportement des lipides dans l'eau.

En dehors de quelques observations plus anciennes [Tanford 2004], on peut considérer que les recherches sur le comportement des lipides dans l'eau ont débuté avec les travaux de Lord Rayleigh à la fin du XIX^{ème} siècle. Cet auteur savait que lorsqu'on met une petite quantité d'huile sur l'eau, l'huile se répand jusqu'à former une très fine couche, qu'il a assimilée à une couche d'une seule molécule de hauteur, comme indiqué sur la Figure 24²⁵. Il a essayé de mesurer la hauteur de ce film afin d'estimer la taille des molécules qui le composaient [Rayleigh 1889]. La technique de Rayleigh consistait au départ à mettre une goutte d'huile sur l'eau, attendre qu'elle prenne ses dimensions maximales et mesurer la surface résultante, mais il était assez difficile de connaître précisément avec cette méthode les

²⁵ Quelques années plus tard, le calcul du nombre d'Avogadro permit de confirmer que les films d'huile forment bien des monocouches au dessus de l'eau [Devaux 1913].

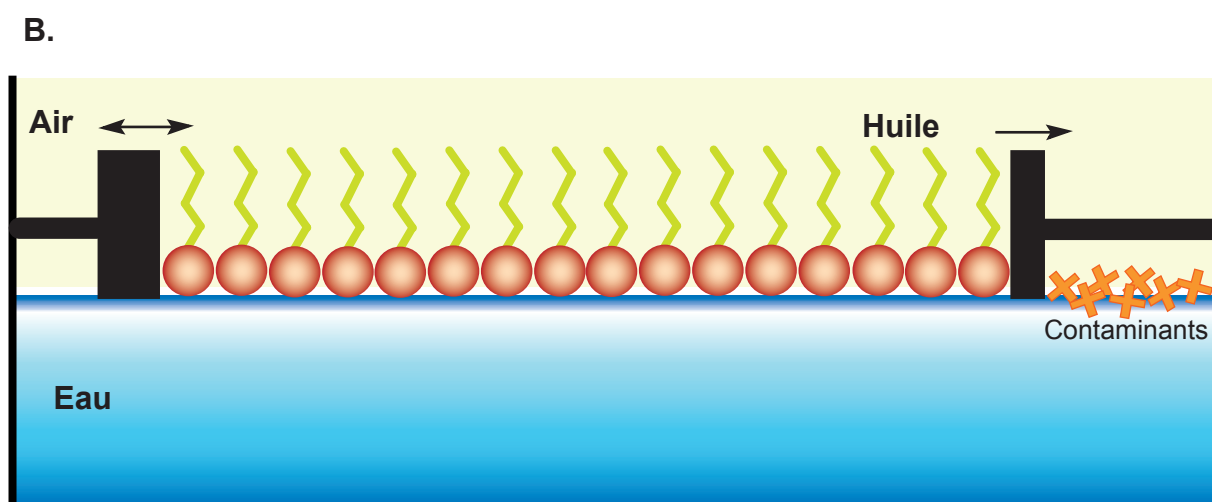
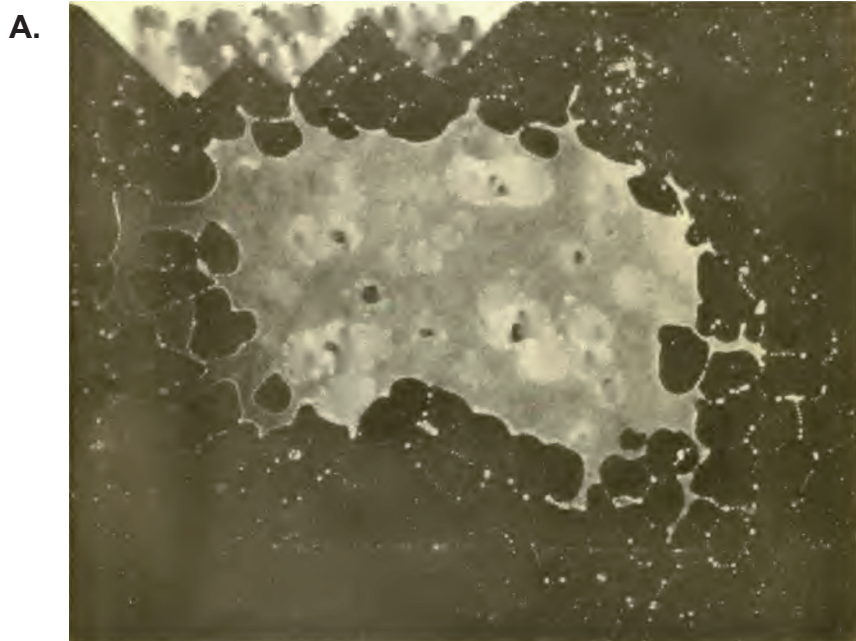


Figure 26. Mesure de la surface occupée par l'huile sur l'eau.

A. D'après Devaux 1913. Il est difficile de mesurer la surface occupée par l'huile sur une surface ouverte, et les contaminants peuvent influencer la tension superficielle. **B. Balance de Langmuir.** Elle est composée d'un récipient rectangulaire contenant de l'eau. La surface de l'expérience est délimitée par deux barrières. Sur le schéma, la barrière de droite permet d'écartier les contaminants de la surface de l'eau et celle de gauche permet de ne libérer que la surface nécessaire pour que les phospholipides forment une seule couche régulière à la surface de l'eau.

surfaces atteintes par l'huile. Il était également difficile de s'assurer de l'absence d'impuretés sur l'eau, ce qui rendait difficile la mesure de la *tension superficielle* de l'interface eau-huile (la tension superficielle est la résistance qu'un liquide offre à une force extérieure de par les interactions qui existent au sein du liquide (Figure 25). Rayleigh a alors été contacté par une dame (Mrs Pockels) qui avait inventé chez elle un dispositif (Figure 26) capable de faciliter la mesure de l'expansion de l'huile ainsi que de la tension de surface [Rayleigh 1891]²⁶. Ce dispositif reçut le nom de balance de Langmuir en l'honneur d'Irving Langmuir qui l'améliora quelques années plus tard.

Les travaux de Langmuir ont joué un rôle assez important dans l'histoire de notre compréhension des membranes. D'une part, lui et d'autres auteurs ont compris que, contrairement aux molécules totalement polaires ou apolaires qui se dissolvent dans les solvants de même nature qu'eux, les acides gras qui composent l'huile forment des monocouches à l'interface eau-air parce que qu'ils sont amphiphiles et que la tête polaire tente de maximiser son contact avec l'eau en même temps que la partie hydrophobe la minimise [Harkins *et al.* 1917 ; Langmuir 1917]. D'autre part, en 1925, Gorter et Grendel ont utilisé la balance de Langmuir dans une expérience historique qui les a conduits à proposer que les cellules étaient entourées d'une bicouche lipidique (Figure 27). Ils ont purifié les membranes des érythrocytes de plusieurs animaux et ils ont mesuré à l'aide de la balance de Langmuir la surface que leurs lipides occupaient à la surface de l'eau. L'utilisation d'érythrocytes était justifiée par l'absence de membranes internes dans ces cellules, ce qui permet de s'assurer que tous les lipides extraits proviennent de la membrane plasmique. Gorter et Grendel ont observé que la surface occupée par les lipides était deux fois plus grande que la surface qu'ils avaient préalablement estimée pour les cellules. Etant donné ce rapport 2 : 1, ils ont conclu que la membrane plasmique était composée d'une bicouche lipidique où les têtes hydrophiles des phospholipides sont tournées vers l'extérieur et les queues hydrophobes vers l'intérieur [Gorter et Grendel 1925].

Ce calcul astucieux est considéré comme la première preuve de l'existence d'une bicouche lipidique autour des cellules. Cependant, deux difficultés de ce travail sont souvent négligées. D'un côté, la technique utilisée par Gorter et Grendel ne permettait pas d'extraire la totalité de lipides des membranes, de manière que la quantité de lipides a été sous-estimée. Ils

²⁶ L'aspect anecdotique de l'initiative spontanée d'Agnes Pockels de contacter Lord Rayleigh a été développé dans un ouvrage de vulgarisation sur l'histoire des sciences. Voir Tanford 2004.

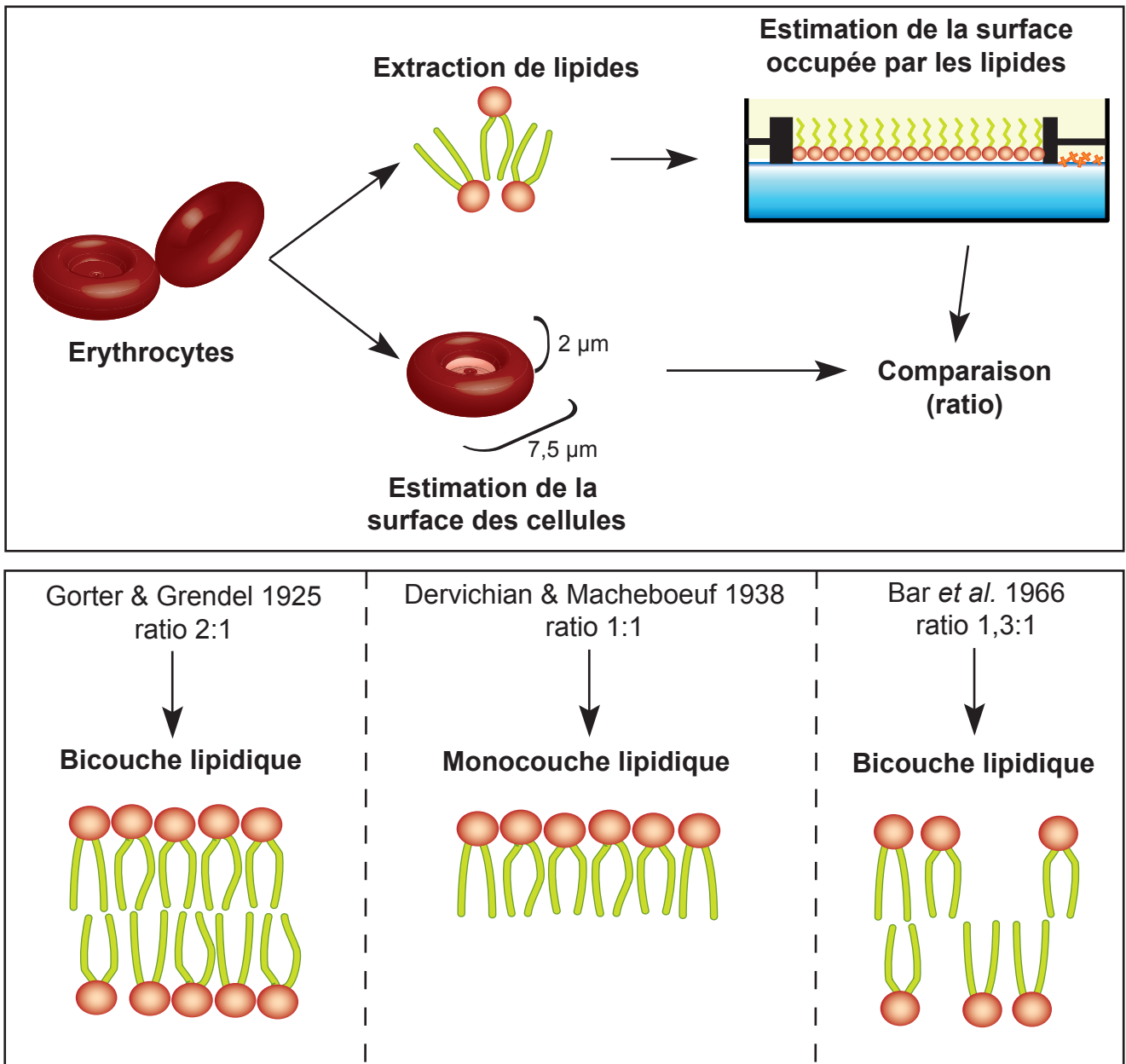


Figure 27. Mesure de la surface occupée par les phospholipides membranaires pour déterminer leur arrangement dans la membrane.

A. Méthode utilisée. **B.** Résultats obtenus par différentes études et interprétations. Voir explications dans le texte.

ont également sous-estimé la surface des cellules, ce qui rend difficilement exploitables leurs résultats car le rapport 2 : 1 semble être en réalité le fruit du hasard [Bar *et al.* 1966]. D'un autre côté, Gorter et Grendel ont mesuré la surface *maximale* que les lipides pouvaient occuper (Figure 27) alors que d'autres auteurs qui ont mesuré la surface *minimale* occupée par la monocouche de lipides membranaires ont obtenu un ratio de 1 : 1 ou de 1,3 : 1 entre la surface occupée par les phospholipides et la surface calculée pour les cellules [Dervichian et Macheboeuf 1938 ; Bar *et al.* 1966]. Le ratio 1 : 1 fut considéré par ses auteurs comme la preuve d'une monocouche de lipides autour de la cellule, alors que le résultat 1,3 : 1 fut expliqué par l'hypothèse selon laquelle les phospholipides membranaires formaient une bicouche peu condensée (Figure 27). Nous savons aujourd'hui que les phospholipides dans les membranes sont plutôt condensés et que la différence qui « reste à couvrir » dans ces résultats est en réalité occupée par des protéines [Singer 2004].

Malgré ces difficultés et les quelques arguments qui s'opposaient à leurs résultats [Fricke 1925 ; Dervichian et Macheboeuf 1938], la bicouche de Gorter et Grendel fut majoritairement acceptée dès qu'elle fut proposée. Mais elle avait des limites. D'abord, même si la composition lipidique des membranes pouvait expliquer le passage de molécules apolaires observé par Overton, on savait que la perméabilité de composés polaires était aussi un phénomène important [Loeb 1906]. De plus, les années 1930 ont connu une profusion d'études sur la tension de surface qui ont montré que des gouttes d'huile dans des œufs (ou extraits d'œufs) de plusieurs téléostéens ou d'oursin avaient des tensions superficielles très inférieures à celles de l'interface eau-huile [Cole 1932 ; Harvey et Shapiro 1934]. On a donc d'abord postulé qu'une molécule était responsable de la diminution de cette tension, et que cette molécule pouvait être une protéine [Danielli et Harvey 1935]. Ensuite, on a étendu cette hypothèse aux membranes cellulaires en proposant qu'elles fussent recouvertes de part et d'autre d'une couche de protéines (Figure 28, [Danielli et Davson 1935]). Malgré le peu d'éléments qui soutenaient le modèle de Danielli et Davson lorsqu'il fut formulé, il devint très vite le paradigme de la structure des membranes. On peut signaler que les observations exposées dans ces travaux semblent être liées au choix du modèle des œufs, qui comporte principalement des triglycérides au lieu de phospholipides [Branton et Park 1968]. Des travaux postérieurs ont montré que la basse tension superficielle était explicable par les phospholipides eux-mêmes [Danielli 1937 ; Harkins 1952 ; Haydon et Taylor 1963].

Les techniques employant la polarisation de la lumière [Schmidt 1936] et la diffraction des rayons X [Schmitt *et al.* 1935] sur des nerfs myélinisés ont détecté des couches

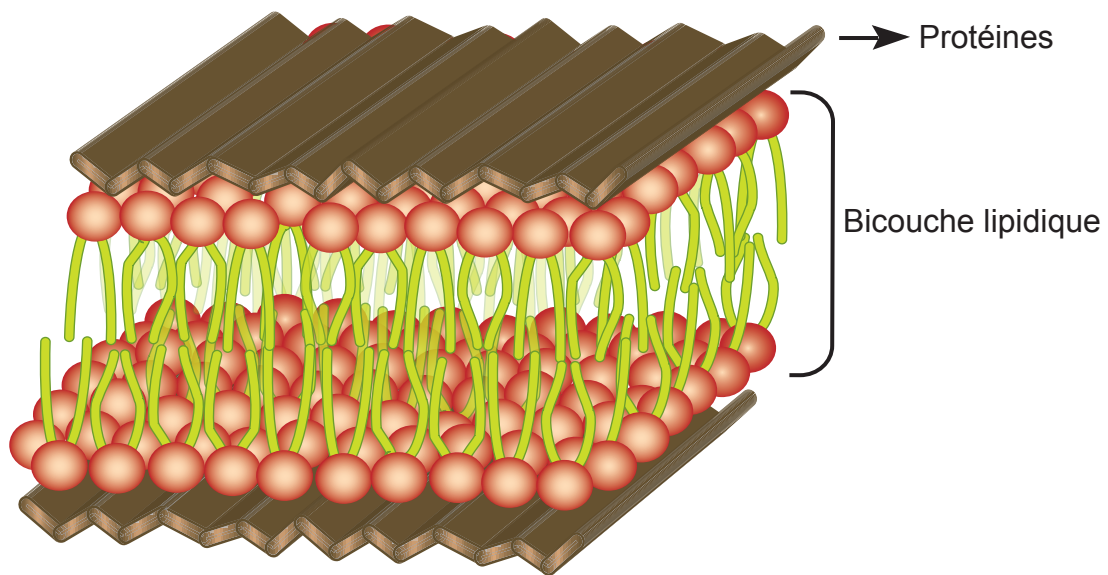
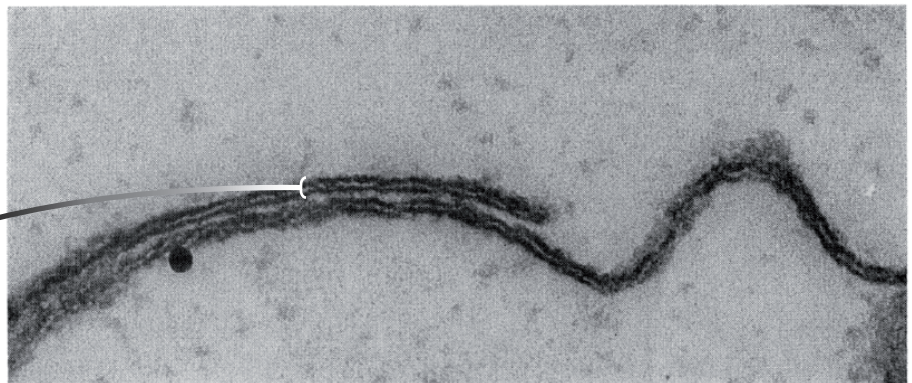
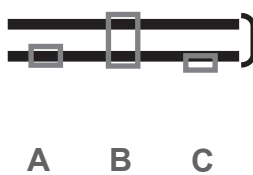


Figure 28. Modèle de Danielli et Davson sur la structure de la membrane plasmique.

La membrane serait composée d'une bicouche lipidique entourée de part et d'autre d'une couche de protéines.

Structure en voie ferrée



Robertson 1957

Figure 29. Interprétation des images de membranes au microscope électronique à transmission.

L'image est extraite de Robertson 1957. Le dessin à gauche schématise la structure en voie ferrée observée dans l'image. En 1959 Robertson se pose explicitement la question d'où se trouve la membrane plasmique dans ces images : s'agirait-il uniquement de la couche foncée interne (A) ? De l'ensemble de la structure (B) ? Ou la voie ferrée correspondrait-elle déjà au milieu extracellulaire (C) ? Il penche pour l'option B. L'unité de membrane proposée par Robertson implique la même structure que celle du modèle de Danielli et Davson (Figure 28).

superposées de protéines et de lipides. Ces observations ont été interprétées comme des arguments en faveur du modèle de Danielli et Davson. Lorsque le développement des microscopes électroniques permit l'observation directe des membranes, on vit des structures dites en *voie ferrée* (Figure 29) : deux couches foncées de part et d'autre d'une partie plus claire [Robertson 1957]. On suggéra que les couches foncées correspondaient aux protéines et têtes polaires des phospholipides alors que l'intérieur était composé des parties hydrophobes de ces derniers. Le développement de la microscopie permit d'observer une grande variété de membranes. Les querelles à propos des interprétations de ces images conduisirent certains auteurs à se demander s'il était vraiment possible de représenter les membranes selon un modèle unique [Branton et Park 1968] ou même laquelle des parties observables dans les images obtenues devait être considérée comme étant la membrane : la surface interne de la couche intérieure, la couche foncée interne ou l'ensemble des trois couches ? (Figure 29, [Robertson 1959]).

En 1959-1960, Robertson fit une récapitulation des observations au microscope électronique et, se basant sur le modèle de Danielli et Davson, il extrapola un socle commun aux clichés observés. Le modèle qui en résulta, très similaire à son prédécesseur, fut connu sous le nom d'*unité de membrane*, un nom choisi à la fois pour souligner l'existence d'une membrane unique comprenant les trois couches observées au microscope électronique et l'universalité de ce type de membrane dans toutes les membranes cellulaires [Robertson 1960]. L'unité de membrane était le paradigme au début des années 1960.

La crise d'un paradigme

En même temps que l'unité de membrane était présentée comme le modèle universel de la structure des membranes, les hésitations autour de la présence de protéines à la surface de la bicouche lipidique ont commencé à grandir au début des années 1960. En premier lieu, la composition des membranes semblait très différente selon l'origine des membranes étudiées, avec des proportions protéines/lipides allant de 1 : 4 à 5 : 1 [Branton et Park 1968]. Il paraissait donc difficile d'admettre que la quantité de protéines suffise toujours pour recouvrir la bicouche lipidique. Deuxièmement, la plupart des observations qui soutenaient l'unité de membrane étaient basées sur le nerf myélinisé, une structure connue pour son caractère isolant et métaboliquement inerte, incompatible avec les nombreuses activités cellulaires attribuables aux membranes [Korn 1966 ; Branton et Park 1968]. Enfin, il était devenu plus facile à l'époque de mesurer la stabilité thermodynamique d'une structure. Dans le domaine des

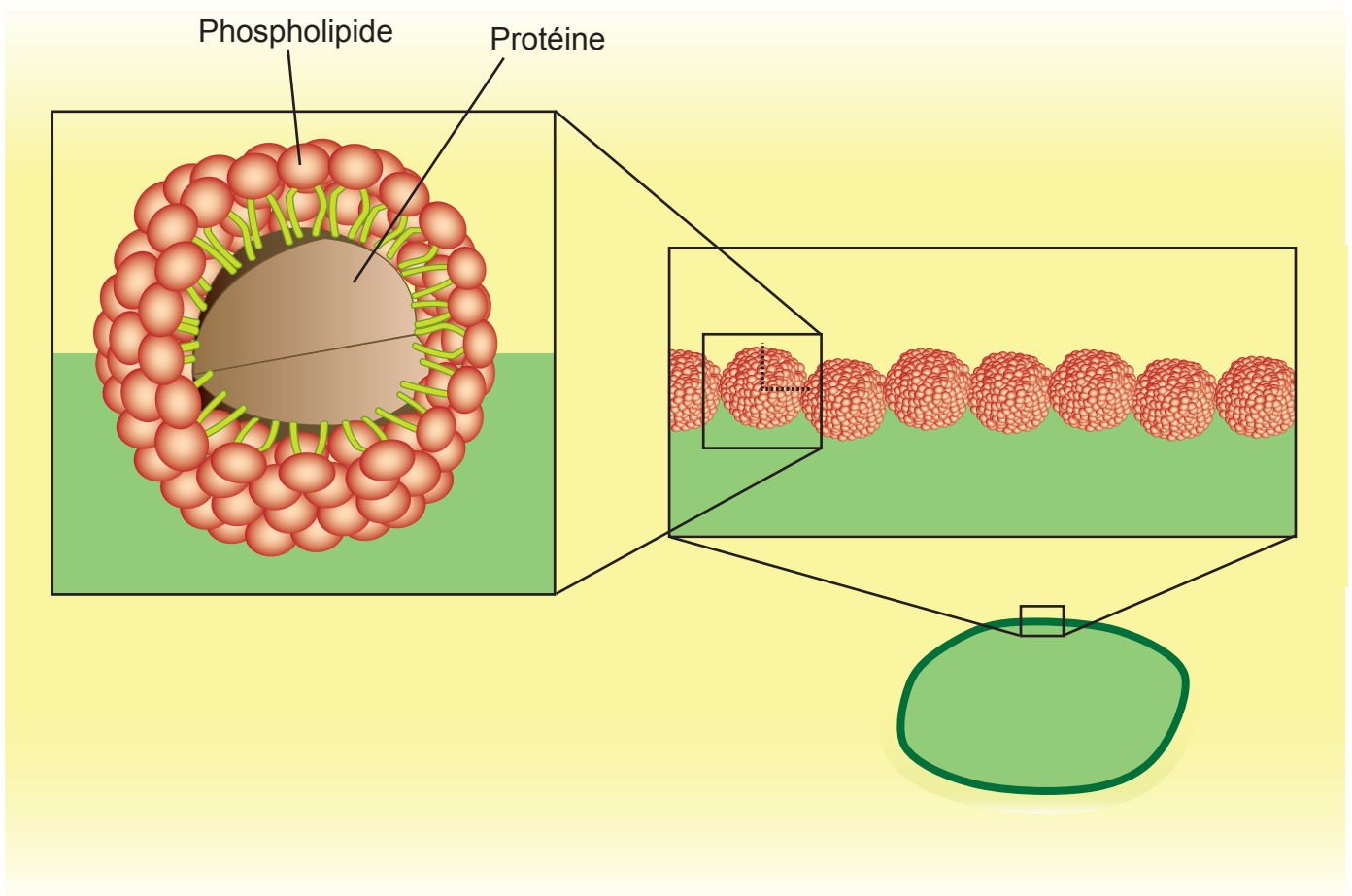


Figure 30. Modèle de Benson 1966 pour la structure de la membrane plasmique.
La membrane serait constituée d'unités composées de protéines entourées de phospholipides.

lipides, l'importance thermodynamique des interactions hydrophobes permit d'expliquer la formation spontanée de bicouches (ou de micelles, en fonction des conditions) dès que l'on met des molécules amphiphiles dans l'eau. Dans le cas des protéines, il apparut que leur interaction avec les lipides était plus stable lorsqu'elle se faisait par leurs parties hydrophobes au lieu de par leurs parties polaires. La superposition d'une couche protéique sur la bicouche lipidique était donc énergétiquement défavorisée [Haydon et Taylor 1963 ; Richardson *et al.* 1963].

L'unité de membrane est donc née au début des années 1960 dans un contexte de fortes critiques [Haydon et Taylor 1963 ; Korn 1966] et on lui a opposé l'hypothèse que la membrane était peut-être composée de sous-unités. En effet, étant donné que les interactions les plus stables entre protéines et lipides se font par leurs parties hydrophobes, plusieurs auteurs commencèrent à penser que les protéines membranaires pouvaient s'entourer de lipides (Figure 30) pour former des lipoprotéines avec un cœur hydrophobe et une surface hydrophile. Ces lipoprotéines composeraient le plan de la membrane [Richardson *et al.* 1963 ; Benson 1966]. Ce modèle fut clairement influencé par la découverte que les capsides qui délimitent certains virus sont composées de protéines adoptant des formes géométriques [Hechter 1965]. L'observation dans les membranes mitochondriales et les thylakoïdes chloroplastiques de structures répétitives qui furent appelées dans le premier cas « particules élémentaires » [Fernandez-Moran 1962 ; Fernandez-Moran *et al.* 1964] et dans le deuxième « quantasomes » [Park et Pon 1961 ; Park 1965] parut confirmer l'hypothèse des membranes composées de sous-unités. Pour l'anecdote, je précise que ces « sous-unités » observées au microscope électronique correspondent probablement à une grande concentration d'ATPases dans les membranes mitochondriales et de photosystèmes dans les thylakoïdes.

Naissance et vie de la mosaïque fluide

Vers la moitié des années 1960, l'hypothèse que la membrane était composée de sous-unités lipoprotéiques concurrençait sérieusement le modèle de l'unité de membrane, mais ce furent finalement d'autres arguments qui conduisirent à rejeter ce paradigme.

a) Pendant cette période la fabrication des premiers *liposomes* (vésicules artificielles composées d'une bicouche de lipides), montra que la formation de ces structures est spontanée sous certaines conditions dès lors que des molécules amphiphiles sont mises dans un milieu aqueux. Les liposomes ont d'ailleurs des propriétés de semi-perméabilité qui furent très importantes pour la compréhension des capacités des membranes biologiques [Bangham

et al. 1965]. Lorsque des bicouches synthétiques composées uniquement de lipides furent observées au microscope électronique, on obtint des clichés en voie ferrée comme ceux des membranes cellulaires [Henn *et al.* 1967], ce qui contredisait la conviction précédente selon laquelle l'absence de protéines dans les membranes changeait l'image obtenue au microscope électronique [Stoeckenius 1962].

b) Les techniques de microscopie électronique avaient été très critiquées en raison de la technique chimique utilisée pour fixer le matériel biologique, que l'on suspectait de modifier les structures naturelles. On développa alors le cryodécapage, une méthode consistant à fixer les échantillons par le froid et à les briser avant l'observation au microscope électronique [Moor et Muhlethaler 1963]. Cette technique découpe les membranes au niveau de la partie hydrophobe de la bicouche, permettant de séparer chaque monocouche lipidique et d'observer l'intérieur de la membrane. On y observa alors des protubérances plus ou moins nombreuses selon les membranes fracturées. Lorsqu'on remarquait une protubérance sur une couche, la couche opposée présentait un creux en miroir, ce qui semblait indiquer que des « globules » se trouvaient à l'intérieur du plan de la bicouche et que la fracture les contraignait à rester attachés à l'une des deux couches [Branton 1966 ; Deamer et Branton 1967 ; Pinto da Silva et Branton 1970].

c) Il fut publié une étude de la fusion de deux cellules (une humaine, une de souris) comportant un suivi des composants membranaires par deux marquages fluorescents différents. Il fut observé qu'après la fusion des deux cellules, les deux fluorescences se trouvaient progressivement mélangées, ce qui suggérait la capacité des composants membranaires à diffuser librement sur le plan de la membrane [Frye et Edidin 1970].

d) Le modèle de l'unité de membrane prédisait que les protéines membranaires devaient se trouver en conformation de feuillet beta par-dessus la bicouche lipidique [Davson 1962]. Or plusieurs travaux montrèrent que les protéines membranaires ne se trouvaient pas dans cette conformation [Maddy et Malcolm 1965 ; Lenard et Singer 1966 ; Wallach et Zahler 1966].

Ces derniers travaux concluaient que l'unité de membrane était incompatible avec leurs résultats et ceux des autres auteurs. Les premières ébauches du modèle de la mosaïque membranaire furent formulées : les protéines sont insérées dans une matrice de phospholipides avec lesquels elles établissent des interactions hydrophobes, ce qui leur permet d'accomplir un grand nombre de fonctions, notamment le passage de molécules de

part et d'autre de la membrane [Lenard et Singer 1966 ; Wallach et Zahler 1966]. L'insuffisance de phospholipides pour composer deux couches complètes [Bar *et al.* 1966], l'observation d'une structure en voie ferrée des bicouches lipidiques sans protéines [Henn *et al.* 1967], les résultats de cryodécapage [Branton 1966] et les indices de fluidité membranaire [Frye et Edidin 1970] furent revus à la lumière de cette nouvelle hypothèse, conduisant à l'actuel modèle de la mosaïque fluide [Singer et Nicolson 1972].

Dès sa formulation, ce modèle prenait en compte le caractère asymétrique de la membrane qui avait déjà été postulé auparavant [Robertson 1960 ; Pinto da Silva et Branton 1970 ; Bretscher 1972] mais dont les premières preuves en dehors de l'asymétrie des oligosaccharides [Nicolson et Singer 1971] arrivèrent un peu plus tard [Gordesky et Marinetti 1973 ; Fisher 1976]. Ce modèle admettait également la possibilité d'un rassemblement de composants à certains points précis de la membrane (microdomaines), sujet qui a beaucoup attiré l'attention ces dernières années [Simons et Ikonen 1997]. La détermination de la structure des membranes cellulaires a par ailleurs permis de comprendre que les protéines membranaires ont besoin d'interagir avec la matrice formée par la bicouche lipidique pour adopter leur conformation active [Phillips *et al.* 2008 ; Zhang et Rock 2008]. La prise en compte du contexte lipidique déboucha en 1978 sur la détermination de la première conformation d'une protéine membranaire, première d'une longue série qui se poursuit encore aujourd'hui [Singer 2004]. Ce paradigme est aussi celui dans lequel on conçoit les futurs défis concernant l'étude des membranes : une meilleure compréhension de leur dynamique et de leurs relations avec le reste de la cellule [Edidin 2003 ; Zhang et Rock 2008].

Perspectives de l'analyse historique des membranes cellulaires pour celle des membranes anciennes

Dans la première partie de cette analyse historique nous avons vu que la démonstration de l'existence d'une membrane plasmique différenciée du reste de la cellule a été difficile et que d'autres hypothèses ont été envisagées. Or, comme le signale Plowe dans le texte cité dans l'épigraphe de cette partie, l'existence ou non de membranes change totalement notre façon de concevoir le fonctionnement des cellules et leur indépendance face à leur environnement. Nous retrouverons cette difficulté lorsqu'on parlera de la compartimentation du cenancêtre.

Dans la deuxième partie, nous avons vu aussi que la détermination du modèle actuel de la mosaïque fluide n'a pas du tout été immédiate mais qu'au contraire de nombreux modèles

ont été proposés puis abandonnés au profit de celui qui semble aujourd'hui expliquer le mieux les observations. Ces études ont permis de décrire la membrane comme un compartiment cellulaire avec des propriétés spécifiques. Comme je me suis principalement concentré sur l'existence ou l'absence de membranes chez le cenancêtre, j'aborderai très peu la question des membranes du cenancêtre en tant que compartiment mais ce sujet sera de grand intérêt comme perspective pour la suite de mon travail.

3. Quand la diversité membranaire actuelle met en cause l'existence de membranes chez le cenancêtre

Le modèle de la mosaïque fluide est aujourd'hui considéré comme étant universel. On accepte ainsi que toutes les membranes soient composées d'une matrice de phospholipides dans laquelle on trouve une diversité d'autres lipides et protéines, glycosylés ou non. Cette structure de base est considérée comme étant la même partout mais il existe en réalité une grande diversité de membranes : certains organismes sont entourés de deux membranes (notamment chez les bactéries) et beaucoup d'autres contiennent des membranes intracellulaires (surtout les eucaryotes mais aussi certains procaryotes). De ce fait, il existe de très nombreuses différences entre les membranes en fonction de leur composition et des proportions de différents types de protéines, phospholipides et autres lipides²⁷. Par ailleurs, comme on l'a vu dans la section précédente, beaucoup de membranes externes sont doublées par une paroi qui peut-être formée d'une grande diversité de molécules (cellulose, chitine, différents peptidoglycanes...). Cependant, les parois sont très différentes entre elles et ne peuvent donc pas être comparées directement dans les trois domaines du vivant. Je parlerai ici seulement des membranes et jamais des parois. Comme les composants fondamentaux des membranes lipidiques sont les phospholipides, c'est sur eux que mon travail s'est concentré.

²⁷ Certains organismes sont même connus pour manquer presque totalement de phospholipides dans des parties importantes de leurs membranes plasmiques, comme c'est par exemple le cas du stramenopile *Ochromonas danica*, dont la quasi-totalité des phospholipides des membranes flagellaires et une partie du reste est remplacée par une famille très rare de chlorosulfolipides [Elovson et Vagelos 1969 ; Bedke *et al.* 2009 ; Bedke et Vanderwal 2011]. En revanche, ces lipides forment des bicouches et sont synthétisés par une voie très similaire à celle des autres organismes, donc leur cas semble être le résultat d'une adaptation relativement récente et ne saurait mettre en question le modèle de la mosaïque fluide.

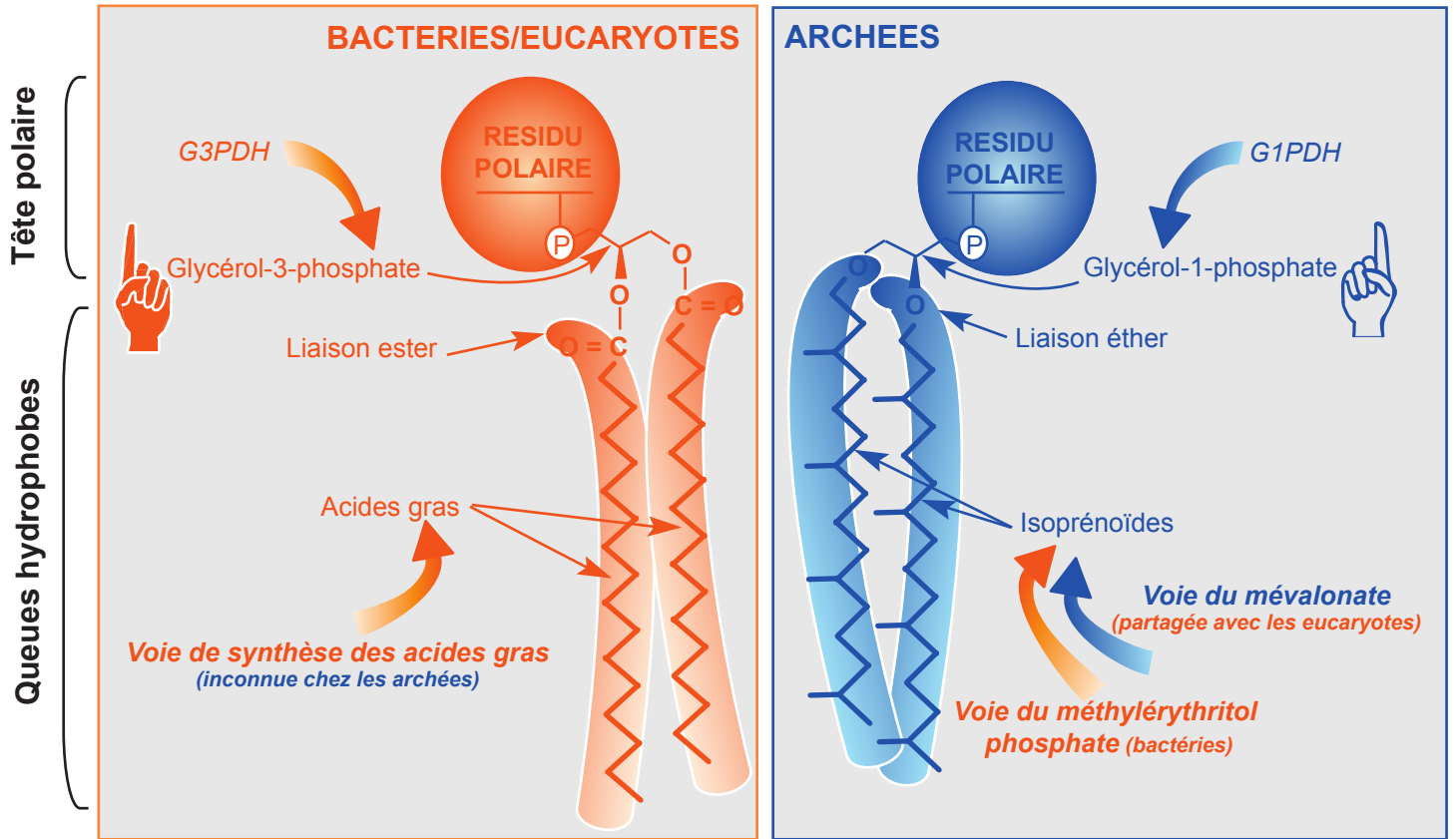


Figure 31. Deux types de phospholipides dans les trois domaines du vivant.

Les cercles colorés représentent les têtes polaires (le résidu polaire pouvant être l'éthanolamine, le glycérol phosphate, le sérine, etc) et les parties allongées colorées représentent les queues hydrophobes. Les flèches fines signalent les principales caractéristiques de chaque type de phospholipide et les flèches épaisses représentent les différentes voies de synthèse des composants des phospholipides.

Des phospholipides exotiques chez les archées

La plupart des travaux historiques sur la composition et la structure des membranes que nous avons vu plus haut se sont concentrés sur les érythrocytes, les cellules du système nerveux des vertébrés, certaines cellules végétales, et, dans une moindre mesure, sur quelques champignons et bactéries [Fricke et Curtis 1934 ; Robinow 1962]. Ainsi, au début des années 1960 tous les phospholipides connus provenaient d'eucaryotes et de bactéries et avaient une composition très similaire. Dans ces phospholipides, le *sn*-glycérol-3-phosphate (G3P) sert d'ossature (Figure 31). Les deux carbones du G3P qui ne sont pas liés au groupement phosphate sont attachés à deux acides gras, i. e. des chaînes carbonées linéaires (aliphatiques) avec un groupement carboxylique terminal leur permettant d'établir une liaison ester avec le G3P. Cette partie du phospholipide constitue sa partie hydrophobe (queues hydrophobes). Le phosphate est attaché à une molécule polaire plus variable constituant la partie hydrophile (tête hydrophile, Figure 31).

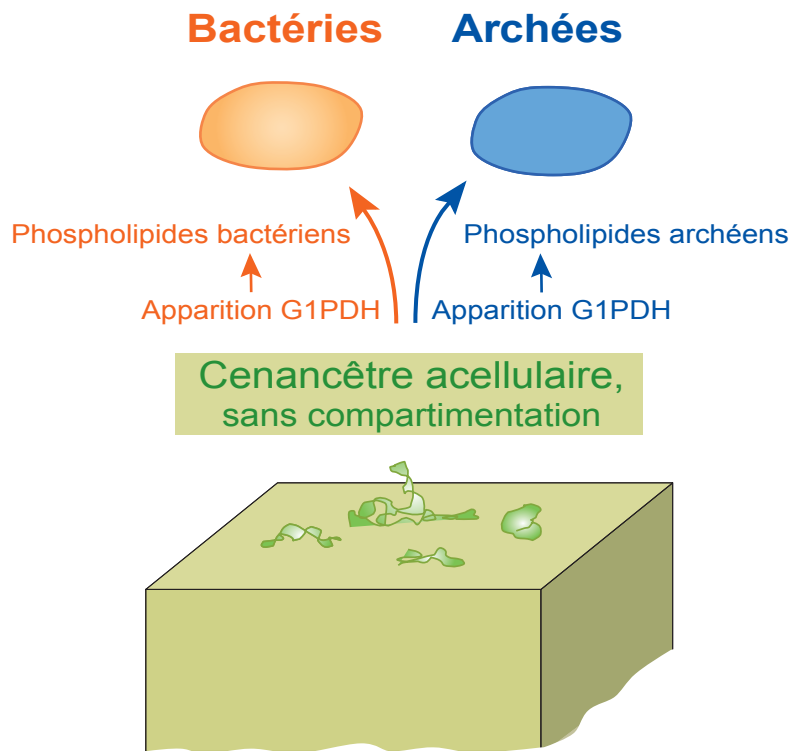
A partir de 1965, l'étude de ce que l'on pensait à l'époque être des bactéries halophiles révéla des phospholipides dont la partie hydrophobe était totalement différente de celle connue auparavant (Figure 31). L'ossature de ces phospholipides n'était pas constituée de G3P mais de l'énantiomère opposé (son image dans un miroir), le *sn*-glycérol-1-phosphate (G1P). En outre, les queues hydrophobes n'étaient pas des acides gras mais des chaînes carbonées branchées que l'on appelle des isoprénoïdes, attachées par des liaisons éther et non pas ester [Kates *et al.* 1965]. Très vite après leur découverte, ces phospholipides atypiques ont été observés chez plusieurs autres « bactéries » vivant dans des milieux exotiques : des halophiles [Kates *et al.* 1966], thermoacidophiles [Langworthy *et al.* 1972 ; Langworthy *et al.* 1974] ou méthanogènes [Tornabene *et al.* 1978 ; Tornabene et Langworthy 1979]. On corréla d'abord ces résultats avec les conditions de vie extraordinaires de ces organismes, en pensant que ces phospholipides atypiques pouvaient s'expliquer par l'adaptation à de tels environnements [Kates *et al.* 1965 ; Langworthy *et al.* 1974].

Parallèlement, la comparaison des ARN ribosomiques (plus tard la comparaison des séquences d'ADN qui les codent) avait permis de comprendre que ces organismes (archées) étaient aussi différents des bactéries qu'ils l'étaient des eucaryotes, divisant le vivant en trois domaines [Woese et Fox 1977b]. Afin de soutenir la validité de cette classification, des caractères supplémentaires spécifiques aux archées furent recherchés mais ces caractères étaient souvent confrontés à des critiques : (1) les archées n'ont pas les parois cellulaires en

peptidoglycane qui sont fréquentes chez les bactéries mais l'on pouvait répondre que l'absence d'un caractère ne constitue pas une preuve d'évolution commune ; (2) des métabolismes jusqu'alors inconnus avaient été détectés chez les archées mais toutes les archées ne partagent pas le même métabolisme et des organismes autres que les archées peuvent aussi avoir des métabolismes exotiques ; (3) on a pensé que les habitats extrêmes des archées pouvaient fournir des renseignements sur leur appartenance à un domaine du vivant différent mais on sait que les niches écologiques sont des mauvais indicateurs des relations de parenté ; (4) enfin, en dehors de la ressemblance de leur ARNr, le principal argument recevable pour accepter que les archées composaient un domaine du vivant différent était la présence d'un caractère dérivé partagé (*synapomorphie*) dans leurs ARNt.

Dans ce contexte, la découverte précédemment décrite de l'existence de phospholipides propres aux archées a joué un rôle important dans l'acceptation de l'idée que les archées sont un domaine du vivant à part [Woese *et al.* 1978]. Plus tard, cette caractéristique a été confortée par la découverte que leurs voies de biosynthèse sont aussi distribuées irrégulièrement entre les trois domaines du vivant (Figure 31). D'une part, les voies de synthèse des acides gras des bactéries et des eucaryotes ont une origine commune [Smith et Sherman 2008] alors que l'existence d'une telle voie n'a pas encore été démontrée chez les archées. D'autre part, bien que les isoprénoïdes soient des molécules présentes chez tous les organismes, les archées et les eucaryotes utilisent la voie du mévalonate pour les synthétiser [Kates *et al.* 1968 ; Boucher *et al.* 2004 ; Berg *et al.* 2007a] alors que les bactéries utilisent le plus souvent une voie alternative non homologue, celle du methylerythritol phosphate [Lange *et al.* 2000 ; Rodríguez-Concepción et Boronat 2002]. Finalement, les archées synthétisent leur G1P grâce à une *sn*-glycerol-1-phosphate dehydrogenase (G1PDH, [Nishihara et Koga 1997]) qui n'est pas homologue de la *sn*-glycerol-3-phosphate dehydrogenase (G3PDH, [Kito et Pizer 1969]) responsable de la synthèse de G3P chez les bactéries et les eucaryotes.

La découverte des différences radicales qui existent entre ces deux types de phospholipides a donc été un fort soutien de la classification taxonomique la plus courante aujourd'hui, qui considère les archées comme étant un domaine du vivant différent des bactéries et des eucaryotes. Cependant, cette distinction a fait surgir une énigme concernant l'origine évolutive des phospholipides et, par conséquent, des membranes plasmiques elles-mêmes [Koga *et al.* 1998]. D'une part, comme toutes les cellules contemporaines sont délimitées par une membrane plasmique respectant le modèle de la mosaïque fluide, il semblerait raisonnable d'accepter que leur dernier ancêtre commun possédait déjà de telles



Surface de pyrite en conditions hydrothermales
(modèle d'apparition de la vie)

Figure 32. Hypothèse de Koga *et al.* 1998 d'un cenancêtre non compartimenté.

Apparition des composants dont la présence peut être inférée chez le cenancêtre directement sur une surface minérale non compartimentée (modèle d'origine de la vie exposé dans Wächtershäuser 1988). L'apparition des déshydrogénases spécifiques capables de synthétiser respectivement un énantiomère de glycérol phosphate aurait conduit à l'apparition des premiers phospholipides et à la divergence des archées et des bactéries.

membranes. D'autre part, étant données les différences radicales qui existent entre les deux types de phospholipides et la distribution différentielle de leurs voies de biosynthèse chez les trois domaines du vivant, quelle pouvait être la nature des phospholipides chez le cenancêtre ?

Un cenancêtre acellulaire

La première réponse donnée à ces contradictions fut que le cenancêtre ne possédait ni un type de phospholipide ni l'autre, que chaque mécanisme de synthèse était apparu indépendamment et avait été à l'origine des lignées modernes d'archées et de bactéries (les eucaryotes n'étant pas pris en compte ici car on admet qu'ils seraient apparus ultérieurement d'une fusion entre une archée et une bactérie, [Koga *et al.* 1998]). Ce scénario se base sur l'idée que la vie aurait pu commencer sous la forme d'un métabolisme local situé sur une surface de pyrite en conditions hydrothermales [Wächtershäuser 1988]. Sur cette surface seraient apparues les structures communes conservées par les bactéries et les archées et qui sont par conséquent déductibles chez le cenancêtre : les coenzymes, le métabolisme central, une partie de la machinerie informationnelle (les ribosomes, entre autres) et même les deux énantiomères du glycérol phosphate, synthétisés de façon non stéréospécifique. Ce cenancêtre n'aurait donc pas été compartimenté mais acellulaire. L'apparition indépendante des G1PDH et G3PDH aurait permis de synthétiser spécifiquement chaque énantiomère de glycérol phosphate, qui serait alors rentré dans la composition des deux types de phospholipides. La biosynthèse de phospholipides aurait permis la cellularisation des composants biologiques déjà présents, donnant naissance séparément à la première bactérie et à la première archée (Figure 32, [Koga *et al.* 1998]).

Ce travail a eu pour mérite d'être le premier à signaler la contradiction qui existait entre le caractère universel des membranes et les différences majeures entre les mécanismes de synthèse des phospholipides. Cependant, le scénario auquel il aboutit suscite des critiques sérieuses du fait des graves problèmes soulevés par la nature acellulaire de son cenancêtre. Bien que l'hypothèse d'un métabolisme catalysé et localisé sur une surface de pyrite puisse être stimulante lorsqu'on parle des origines de la vie, nous pouvons inférer chez le cenancêtre la présence de machineries comme celles de la traduction et de la transcription ainsi que des ATPases dont l'apparition paraît très improbable sans un mécanisme de concentration des composants biologiques que nous ne connaissons que dans des entités compartimentées [Oparin 1976 ; de Duve et Miller 1991]. Par ailleurs, il est difficile de concevoir que les ribosomes ou les ATPases soient apparus sans l'action de la sélection naturelle car il s'agit de

structures composées de très nombreux éléments et dont l'avantage sélectif n'est évident que par rapport au reste des composants cellulaires. Or, conceptuellement, pour que la sélection naturelle puisse agir sur un tel réseau comportant de *nombreux éléments intégrés*, il faut un mécanisme qui garantisse le couplage de l'information génétique (génotype) avec les caractéristiques qui peuvent conférer un avantage sélectif (phénotype). Le seul mécanisme présentant cette garantie que l'on connaisse aujourd'hui est la compartimentation [Oparin 1976 ; Tawfik et Griffiths 1998 ; Szostak *et al.* 2001 ; de Duve 2005]. On peut aussi se demander pourquoi le scénario de Koga et collaborateurs accepte la synthèse des deux énantiomères du glycérol phosphate de façon non stéréospécifique chez le cenancêtre mais n'envisage pas leur utilisation dans des phospholipides. D'ailleurs, comme il sera signalé plus loin, la synthèse abiotique de molécules amphiphiles et la formation de vésicules membranaires lorsque ces molécules sont au contact de l'eau semble plausible dans les conditions prébiotiques de la Terre primitive, ce qui semble indiquer que les membranes sont plus faciles à mettre en place dans ces conditions que des structures biologiques plus sophistiquées, comme le ribosome.

En définitive, ce scénario acellulaire a été important car il fut le premier à susciter un débat qui s'est grandement étendu depuis. Cependant, son caractère non compartimenté est très discutable car, en rapprochant le cenancêtre de l'étape prébiotique, il suppose l'émergence de systèmes biologiques très compliqués à partir de mécanismes chimiques très basiques sans proposer un mécanisme qui expliquerait l'accumulation de telles structures ou le moteur de la sélection qui a conduit à leur apparition. Ceci est d'autant plus paradoxal que la sophistication de la machinerie informationnelle et du métabolisme semble bien plus difficile à expliquer dans les conditions prébiotiques que la formation de vésicules à partir de molécules amphiphiles probablement présentes sur la Terre primitive.

Une compartimentation minérale

Toujours en se concentrant sur les grandes différences qui existent entre les deux types de phospholipides, mais en reconnaissant aussi l'importance de la compartimentation pour permettre d'atteindre la concentration suffisante pour former des structures biologiques à partir de leurs monomères, il a été proposé que le cenancêtre ait été borné non pas par des structures lipidiques mais par un autre type de délimitation : des compartiments minéraux.

On sait que les cheminées hydrothermales sont des sites prometteurs pour l'origine de la vie car elles fournissent des limites physiques qui peuvent ralentir la diffusion des molécules

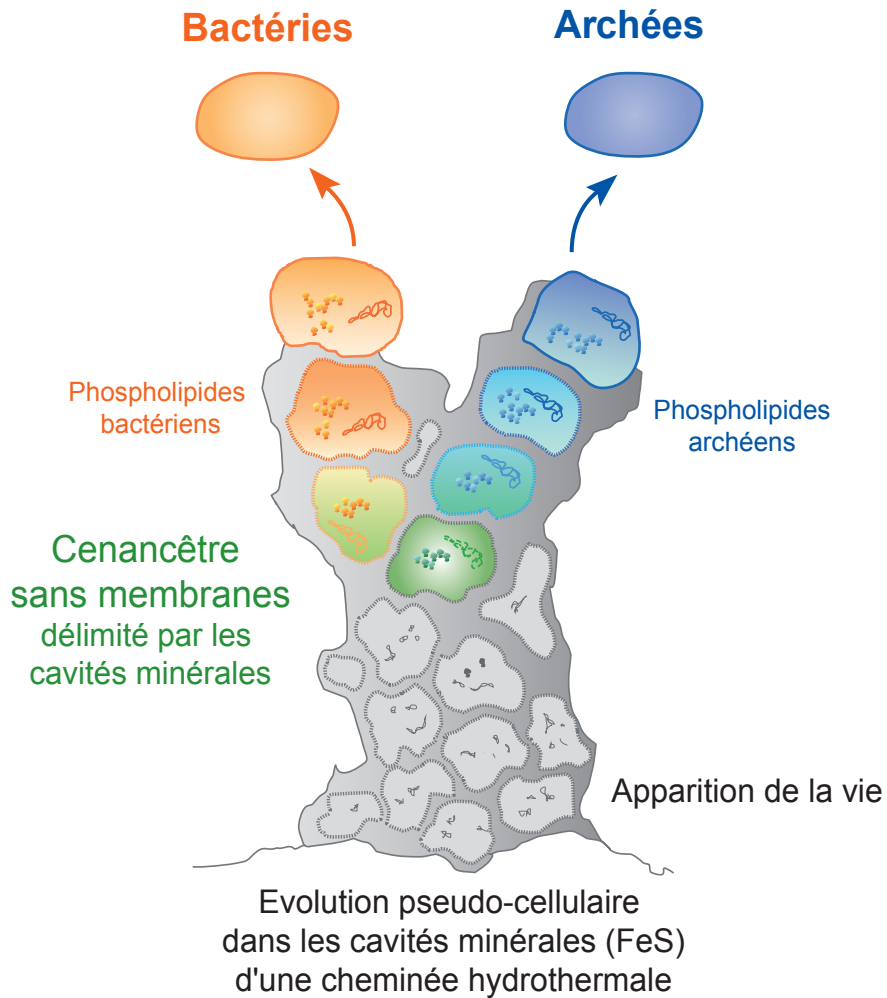


Figure 33. Hypothèse de Martin et Russell 2003 d'un cenancêtre compartimenté dans des cavités minérales.

La vie serait apparue dans une cheminée hydrothermale et aurait évolué jusqu'à l'apparition du cenancêtre, dont les délimitations auraient été minérales. L'apparition des phospholipides à deux endroits différents de la cheminée aurait permis la formation des premières cellules, qui seraient les ancêtres des bactéries et des archées.

organiques vers l'océan, elles contiennent des centres Fe-S et Fe-Ni-S qui auraient pu agir comme catalyseurs primaires et elles sont traversées en permanence par un flux d'énergie redox potentiellement utilisable pour la synthèse de composés biologiques [Russell et Hall 1997]. On sait également que ces conditions hydrothermales conduisent à la précipitation de sulfure de fer (FeS) en forme de bulles. Il a donc été proposé que ces cavités aient constitué les premières délimitations de la vie. Dans ce scénario, les flux énergétiques qui traversent les cavités de sources hydrothermales alcalines auraient formé des molécules organiques de façon abiotique et la présence de la cheminée hydrothermale les aurait accumulées en limitant leur dilution dans l'océan environnant. L'augmentation de la concentration de ces composés organiques aurait conduit progressivement à l'apparition d'un monde ARN, ARN-protéines, à des mécanismes de transcription et traduction et, enfin, au cenancêtre (Figure 33).

Ce cenancêtre, incapable de synthétiser ses phospholipides, aurait été composé d'un réseau de cavités de sulfure de fer contenant des unités autorépliquatives d'ARN ressemblant à des rétrovirus. Ces unités génétiques seraient des objets soumis à variation et sélection naturelle grâce à leur capacité d'autosynthèse, de manière que ceux se répliquant le plus rapidement auraient occupé préférentiellement les compartiments minéraux voisins au fur et à mesure de leur précipitation abiotique. A l'intérieur des cavités minérales il y aurait eu également un réseau capable d'un métabolisme enzymatique primordial, responsable par exemple du métabolisme de l'azote et de la synthèse d'acides nucléiques, de cofacteurs et d'acides aminés. Par la suite, les deux GPDH, les deux machineries de synthèse des phospholipides et d'autres machineries qui sont différentes entre les bactéries et les archées, comme la machinerie de répllication de l'ADN, seraient apparues à des endroits différents de la cheminée hydrothermale. La synthèse de phospholipides aurait permis la formation de vésicules qui se seraient affranchies de la cheminée et auraient emporté avec elles les composants biologiques présents, donnant naissance indépendamment aux bactéries et aux archées (dans cette hypothèse aussi, on considère que les eucaryotes sont apparus plus tard par la fusion d'une bactérie et d'une archée [Martin et Russell 2003 ; Koonin et Martin 2005]).

Ce scénario est devenu très populaire dans la communauté de recherche sur l'origine de la vie, au point qu'on la retrouve aujourd'hui dans certains manuels d'enseignement aux chapitres sur l'évolution précoce de la vie [Madigan *et al.* 2009]. En effet, les cheminées hydrothermales restent aujourd'hui parmi les sites les plus suggestifs en ce qui concerne la période prébiotique et les premières étapes de la vie car elles fournissent des mécanismes de synthèse abiotique de composés organiques et un moyen de les concentrer. Cependant, il ne

faut pas confondre origines de la vie et cenancêtre et l'extension de cette hypothèse au dernier ancêtre commun de tous les organismes se heurte à des problèmes majeurs qui sont malheureusement souvent omis.

Un des principaux problèmes de ce scénario est que, contrairement aux affirmations de ses auteurs [Koonin et Martin 2005], l'évolution darwinienne qui est nécessaire pour expliquer l'ensemble des éléments présents chez le cenancêtre est impossible dans les conditions évoquées. Les éléments dont la présence a déjà été inférée chez le cenancêtre sont les machineries de transcription et traduction, des enzymes comme les ATPases et un métabolisme central capable de synthétiser des cofacteurs, des acides aminés et des acides nucléiques. En général, la fonction de ces composants s'applique à d'autres composants (une ATPase produit de l'ATP dont elle n'est pas le bénéficiaire direct), établissant ainsi un réseau très interconnecté dont chaque élément bénéficie du résultat de l'activité de sa propre fonction uniquement dans la mesure où le reste du réseau fonctionne. Dans ce contexte, la modification d'un composant du réseau peut changer la valeur sélective de l'ensemble par rapport à d'autres réseaux (d'autres organismes) concurrents. Or, étant donné que la plupart de ces composants ne sont pas physiquement liés entre eux, leur *individualisation* doit passer par une délimitation qui les permette de rester ensemble pour bénéficier des avantages du réseau [Szabó *et al.* 2002 ; Ichihashi *et al.* 2010].

Si l'on prend un exemple, un organisme comportant une voie de synthèse des acides nucléiques plus efficace permettra de synthétiser plus rapidement ces molécules, de faciliter éventuellement la réplication de l'information génétique et donc d'accélérer le cycle de vie de l'ensemble de l'organisme. Le résultat sera une augmentation du nombre de ces organismes par rapport à leurs concurrents et donc la propagation de cette voie de synthèse dans la population sans qu'elle ait eu besoin de s'auto-réplicer. Dans le modèle de la cheminée hydrothermale, on suppose que les cavités minérales suffisent à assurer cette individualisation, mais ce n'est pas satisfaisant car les cavités elles-mêmes précipitent de façon abiotique et totalement *indépendante* de leurs contenus biologiques qui ne font que *diffuser* à l'intérieur de ces bulles. En soutenant que la sélection s'applique uniquement à des unités génétiques capables de diffusion dans des structures minérales dont la formation était totalement indépendante, les auteurs oublient que le cenancêtre n'était pas uniquement une structure génétique mais qu'il était déjà constitué d'un réseau de composantes dont la cohésion était indispensable pour former une entité autopoïétique [Fleischaker 1990] et pour que leurs variations puissent conférer un avantage sélectif par rapport à d'autres entités

individuelles [Szostak *et al.* 2001 ; de Duve 2005]. Par rapport à l'hypothèse du cenancêtre acellulaire, ce scénario prend en compte l'importance de la concentration des composants biologiques à l'origine de la vie mais oublie que l'individualisation obtenue grâce à la compartimentation est surtout nécessaire à la sélection et à l'évolution de structures aussi compliquées que les ribosomes qui ont précédé le cenancêtre.

Ce problème n'est pas le seul posé par ce scénario. Dans cette hypothèse, les voies de biosynthèse des phospholipides seraient apparues à différents endroits de la cheminée hydrothermale après la période du cenancêtre et auraient colonisé les cavités en même temps que les machineries de réplication de l'ADN et autres mécanismes cellulaires apparus à l'époque [Koonin et Martin 2005]. Cependant, ce modèle ne propose pas de fonction pour ces phospholipides et, par conséquent, il ne parvient pas à indiquer un avantage sélectif pour encourager la mise en place de telles molécules. On pourrait proposer que l'avantage sélectif de la présence de ces molécules était de réunir les éléments génétiques et métaboliques, mais dans ce cas les cavités minérales qui sont le fondement du scénario deviennent inutiles.

Des arguments indirects ont également été avancés contre cette hypothèse, par exemple l'inférence chez le cenancêtre de la présence d'ATPases et du système SRP : sachant que l'ATPase est un complexe qui utilise le gradient transmembranaire pour synthétiser l'ATP et que le SRP est un système d'adressage de protéines à la membrane, comment de telles structures pouvaient-elles fonctionner et à quoi pouvaient-elles servir en l'absence de membranes ? [Peretó *et al.* 2004 ; Jekely 2006]. Afin de répondre à cette critique, les auteurs du modèle ont complété celui-ci en postulant que des molécules aliphatiques pouvaient être synthétisées de façon abiotique dans la cheminée hydrothermale, formant des plaques lipidiques contenant les ATPases et SRP [Koonin et Martin 2005]. Mais si des membranes portant des ATPases et des SRP existaient, on peut se demander pourquoi elles ne contenaient pas les éléments génétiques et métaboliques permettant l'individualisation de ces structures et rendant inutile le recours à des cavités minérales chez le cenancêtre. D'ailleurs, même si les entités génétiques étaient externes aux membranes, quel était le moteur évolutif pour développer des systèmes d'adressage aux membranes (SRP) si les éléments génétiques n'y étaient pas encore liés ?

En définitive, le scénario des cheminées hydrothermales semble aujourd'hui être un modèle prometteur pour essayer d'en apprendre plus sur les étapes prébiotiques mais, malgré la faveur qui semble lui avoir été accordée, il n'est pas bien adapté à l'explication des

frontières du cenancêtre. La différence principale entre les hypothèses du cenancêtre acellulaire et du cenancêtre délimité par des cavités FeS est que le deuxième fournit un moyen de concentration des substances biologiques. Mais comme la fonction des compartiments cellulaires n'est pas uniquement de concentrer les composants biologiques, les deux scénarios présentés jusqu'ici demeurent sensibles tous les deux à la plupart des critiques qui ont été faites à chacun d'entre eux. De par leurs défauts, ces hypothèses ont permis d'établir les premières exigences auxquelles doit aujourd'hui répondre une hypothèse de compartimentation du cenancêtre : capacité de concentrer les substances biologiques et capacité de définir les entités individuelles sur lesquelles la sélection naturelle puisse agir.

Bilan des hypothèses acellulaires

Considérant que les différences entre bactéries et archées sont trop importantes pour permettre de reconstruire les caractéristiques du cenancêtre, les deux hypothèses qui ont été présentées jusque là (le cenancêtre acellulaire et celui délimité par des compartiments minéraux) ont essayé d'expliquer cet organisme surtout par des approches bottom-up dans lesquelles les mécanismes caractéristiques des organismes modernes n'acquièrent tout leur sens qu'à la suite des divergences des domaines du vivant.

Les critiques à l'encontre de ces hypothèses se concentrent autour de trois axes principaux : (1) l'importance de la compartimentation pour établir des entités regroupant plusieurs composants qui n'auraient pas de cohérence commune autrement ; cette critique a déjà été expliquée en détail dans ce chapitre ; (2) certains axes de recherche sur les origines de la vie indiquent que la formation de structures de type membranaire est relativement facile à obtenir dans des conditions prébiotiques plausibles dans la Terre primitive ; par conséquent, des telles membranes ne sont pas forcément des structures dont la synthèse soit limitante ; elles ont donc pu jouer des rôles très importants dès l'origine de la vie et ce bien avant le cenancêtre ; dans le chapitre suivant je vais résumer les résultats des approches bottom-up relatives aux membranes et sur leur rôle dans les premières étapes de l'apparition de la vie ; (3) la présence de nombreuses protéines membranaires a pu être inférée chez le cenancêtre, ce qui implique indirectement qu'une membrane lipidique similaire aux membranes actuelles devait exister chez cet organisme ; je résumerai les informations disponibles dans la littérature à ce sujet dans le chapitre 5.

4. Avant le cenancêtre : le principe de continuité des membranes

Les approches bottom-up ne sont pas les plus adaptées pour donner des informations précises sur le cenancêtre, spécialement à cause de la définition de celui-ci à partir des organismes modernes. Cependant, il est intéressant de savoir ce que ces approches ont apporté sur la naissance des premières membranes pour démystifier la possibilité que des telles structures aient pu exister très tôt dans l'histoire de la vie. Ainsi, on décrira les membranes du cenancêtre uniquement par rapport aux membranes modernes mais si ces membranes ont existé dans cet organisme alors il sera stimulant de penser qu'une continuité a existé entre les membranes prébiotiques et les membranes modernes.

La grande majorité des travaux sur les origines de la vie ont cristallisé autour de deux grands débats portant l'un sur le substrat d'origine des premières entités vivantes (soupe primordiale et/ou surface catalytique [Pascal *et al.* 2009]) et l'autre sur leur nature (unités répliquatives et/ou métabolisme [Peretó 2005]). Des chercheurs de nombreux domaines (astrophysique, astrochimie, planétologie, géologie, chimie organique, biologie, etc) participent à ce débat en essayant de proposer des mécanismes de synthèse, d'accumulation et d'organisation des molécules qui ont servi de premiers ingrédients aux structures vivantes. Ce sujet est passionnant mais trop vaste. Je concentrerai donc mon attention uniquement sur les travaux qui ont un rapport direct avec des compartiments lipidiques hypothétiques sur la Terre primitive.

Premiers travaux sur les compartiments primordiaux

On attribue généralement à Oparin l'idée que la vie aurait commencé à partir d'un auto-assemblage de matériel organique, sous forme de coacervats, qui sont des agrégats de matériel organique, principalement protéique [Oparin 1924 ; Oparin 1938]. Indépendamment mais presque en même temps, J. B. S. Haldane développait l'hypothèse d'origine de la vie qui faisait des virus des formes intermédiaires entre le vivant et le non vivant. Il est aussi probablement le premier auteur à avoir suggéré qu'une fois que les composants du vivant étaient nés dans la soupe primitive, le passage à la vie actuelle, i. e. à la vie cellulaire, aurait eu lieu par la compartimentation d'un « film huileux » caractéristique des cellules modernes [Haldane 1929]. Cependant le sujet de la compartimentation des organismes a longtemps été en retrait par rapport aux fortes controverses auxquelles j'ai fait référence. Dans un premier temps, tant que les délimitations des cellules modernes n'étaient pas encore bien comprises, le

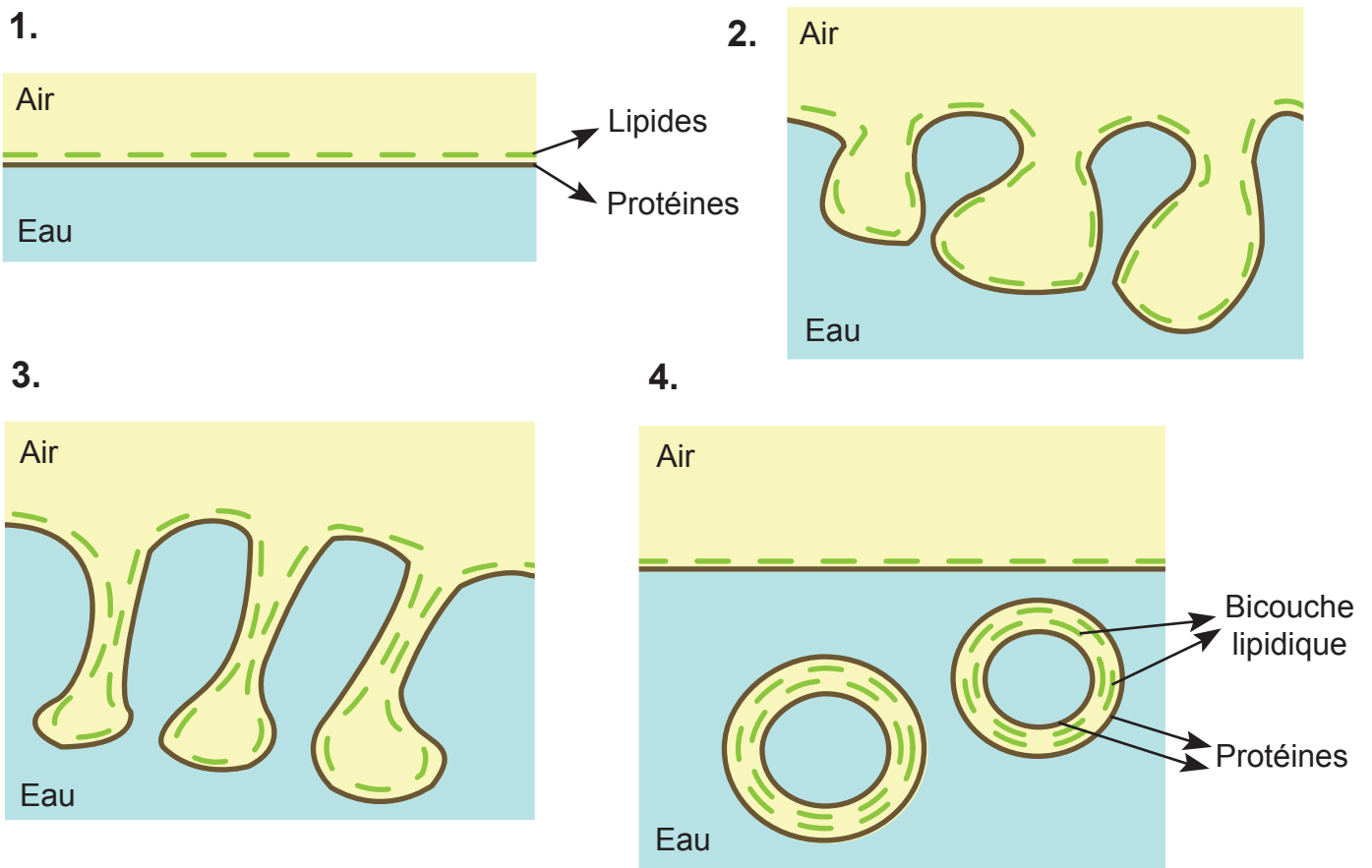


Figure 34. Hypothèse de Goldacre 1958 pour l'apparition des premières cellules ou vésicules se basant sur le modèle membranaire de Danielli et Davson.

Les lipides et les protéines auraient formé des couches superposées à l'interphase entre eau et air. Le mouvement de l'eau aurait conduit la surface à se plier en formant des membranes avec les lipides à l'intérieur et des protéines à la surface, membranes qui auraient permis de délimiter une cavité interne.

modèle de prédilection pour parler de compartimentation était celui des coacervats [Oparin 1938].

Au cours de la trentaine d'années pendant laquelle le modèle de Danielli a été le paradigme, quelques hypothèses ont essayé d'expliquer l'apparition de telles membranes dans les conditions prébiotiques que l'on pensait être plausibles à l'époque. Le meilleur exemple est probablement celui de Goldacre qui a suggéré que les mouvements sur l'interface air-lipide-protéines-eau pouvaient expliquer l'apparition de membranes comme celles du modèle de Danielli ([Goldacre 1958], Figure 34). D'autres auteurs se sont concentrés sur l'importance des protéines dans le modèle de Danielli et ont proposé que le refroidissement brutal d'une solution concentrée d'acides aminés pouvait être à l'origine de la synthèse de vésicules comparables aux membranes cellulaires [Fox *et al.* 1959].

En 1965, on montre que des phospholipides mis dans des conditions aqueuses forment spontanément des vésicules dites *multilamellaires* constituées de plusieurs couches de phospholipides entourant un volume aqueux interne [Bangham *et al.* 1965]. Les propriétés étonnantes des vésicules lipidiques (on observe dès le départ la capacité des structures vésiculaires à former spontanément des gradients ioniques) et l'adoption de la mosaïque fluide comme nouveau paradigme des membranes cellulaires montrent les limites des premières tentatives d'explication de l'apparition des compartiments cellulaires dans la Terre primitive [Hargreaves *et al.* 1977 ; Lazcano 2010].

Dans ce contexte, de nombreuses études on essayé d'approfondir la connaissance du comportement des molécules hydrophobes dans l'eau [Tanford 1973], conduisant à de nouveaux protocoles capables de fabriquer des vésicules *unilamellaires* entourées d'une seule bicouche lipidique, les *liposomes*. Ce n'est qu'à partir de la fin des années 1980 que les études sur ces structures et sur leur formation dans la Terre primitive sont devenues un des axes principaux de recherche dans le domaine des origines de la vie.

Des molécules amphiphiles dans la Terre primitive : synthèse endogène

La première condition qui doit être remplie par toute hypothèse s'intéressant aux premières étapes de la vie est de faire intervenir des molécules dont la présence sur la Terre primitive soit plausible. Etant donné que l'on avait déjà identifié des conditions prometteuses pour la synthèse de peptides et oligonucléotides [Nooner *et al.* 1977 ; Sherwood *et al.* 1977 ; Sherwood et Oro 1977], les premiers travaux consacrés à la synthèse de molécules amphiphiles aptes à former des liposomes dans un contexte de Terre prébiotique ont appliqué

ces conditions à des précurseurs de phospholipides. Ces conditions consistaient en des températures modérément élevées (60°C à 90°C) et en la présence de catalyseurs comme la cyanamide. En fonction des précurseurs on a obtenu du glycérol, du glycérol phosphate [Epps *et al.* 1979 ; Maheen *et al.* 2010], des acylglycérols [Eichberg *et al.* 1977 ; Hargreaves *et al.* 1977] et des phospholipides [Epps *et al.* 1978]. La principale critique à l'encontre de ces travaux est que l'occurrence simultanée des précurseurs utilisés dans ces expériences dans la Terre primitive est très incertaine [Monnard et Deamer 2002]. Beaucoup de travaux se sont alors tournés vers l'étude de molécules amphiphiles plus petites contenant une chaîne carbonée hydrophobe et un groupement hydrophile, comme les acides monocarboxyliques (e. g. les acides gras) qui sont aussi capables de former des vésicules [Hargreaves et Deamer 1978]. Une grande partie des recherches prébiotiques sur les membranes va donc se concentrer sur la synthèse de ces molécules.

Les travaux de Miller ont été pionniers dans l'étude des molécules organiques synthétisables dans des conditions prébiotiques. Il appliqua des décharges électriques dans un mélange de méthane, ammoniac, eau et hydrogène supposé simuler l'atmosphère réductrice que l'on pensait avoir existé dans la Terre ancienne [Miller 1953]. Cette expérience permit de produire un grand nombre de composés organiques mais les acides monocarboxyliques majoritairement obtenus (acide formique, acétique et propanoïque [Miller 1955]) étaient trop courts pour former des vésicules. D'autres études utilisant des décharges électriques comme source d'énergie révélèrent des acides monocarboxyliques contenant jusqu'à 7 carbones, mais ceux-ci étaient branchés et non pas droits [Yuen *et al.* 1981]. Un autre système de synthèse initialement testé fut le mélange de méthane et d'eau soumis à un effet corona. Ce protocole permit la synthèse d'hydrocarbures comprenant jusqu'à 16 carbones et la formation d'acides monocarboxyliques comprenant jusqu'à 12 carbones [Allen et Ponnampereuma 1967]. Cependant, ces chaînes étaient majoritairement branchées et, de plus, la pertinence des conditions de réalisation de ces expériences au regard du contexte plausible de la Terre primitive est contestable.

Aujourd'hui une hypothèse très populaire de fabrication abiotique de chaînes carbonées est celle des synthèses de Fischer-Tropsch (FTS). Les FTS ont d'abord été proposées dans l'industrie comme un moyen de synthétiser des hydrocarbures [Fischer et Tropsch 1926]. Dans leur sens premier, elles consistent en la production à haute température de composés organiques à partir d'un mélange gazeux d'hydrogène et de monoxyde de carbone (ou d'autres molécules capables de former du CO dans ces conditions, comme le CO₂, le CH₃OH

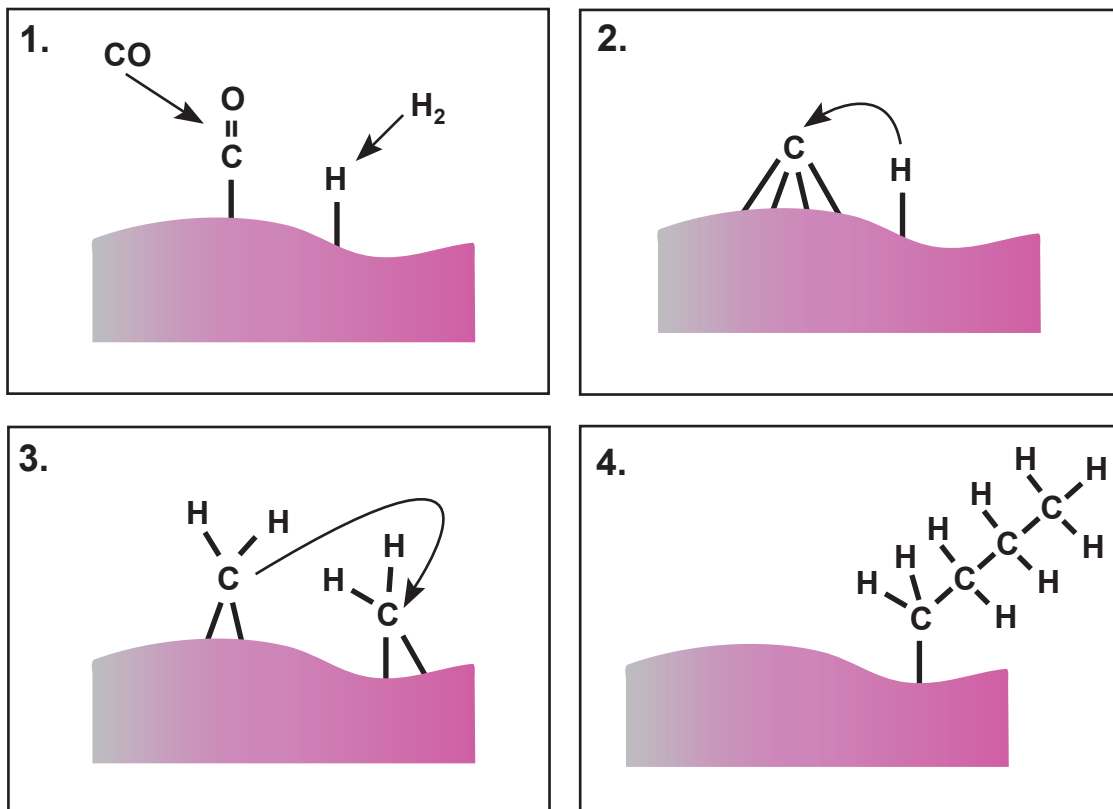


Figure 35. Réactions de Fischer-Tropsch.

Les réactions commencent par l'attachement de monoxyde de carbone et d'hydrogène sur des surfaces catalytiques minérales. L'attaque de l'hydrogène sur le carbone permet la formation d'alcane qui dans un premier temps ne comportent qu'un seul carbone mais dont des nouvelles attaques permettent l'élongation des chaînes jusqu'à l'intégration d'un carbone dont toutes les valences sont occupées. Si des molécules carbonées oxygénées ou azotées s'attachent à la surface, elles peuvent rentrer aussi dans la composition des chaînes.

ou le HCOOH). Bien que ces réactions soient énergétiquement favorables, elles nécessitent des catalyseurs comportant des surfaces capables de fixer les composants pendant l'élongation des chaînes (Figure 35). Les premières applications d'intérêt prébiotique des FTS ont concerné des réactions très proches de celles de la synthèse industrielle, considérant que le mélange gazeux pouvait être rapproché des nuages protostellaires et utilisant des fragments météoritiques riches en fer comme catalyseurs. Ces premières études ont mis en évidence la synthèse d'un grand nombre d'alcane, alcènes, molécules aromatiques, acides gras et même isoprénoides [Studier 1968 ; Nooner *et al.* 1976 ; Nooner et Oro 1979]. Plus récemment, la possibilité que des réactions de FTS puissent être adaptées aux conditions océaniques hydrothermales a commencé à être explorée [McCollom et Seewald 2007]. La reproduction en laboratoire d'un système assimilable aux FTS en conditions hydrothermales a permis de synthétiser à partir d'acide formique ou oxalique des alcane, alcènes, alcools gras, acides gras et alcanones, mais pas d'isoprénoides [McCollom *et al.* 1999 ; Rushdi et Simoneit 2001].

Finalement, la popularité des conditions hydrothermales comme conditions plausibles dans l'Hadéen a conduit à essayer d'autres mécanismes en dehors des FTS. L'une de ces études a mis en évidence la production à partir de petites molécules comme le pyruvate de mélanges de molécules amphiphiles capables de former des vésicules [Hazen et Deamer 2007]. D'autres ont permis l'estérification d'acides gras courts avec du glycérol ou de l'éthylène glycol [Apel et Deamer 2005 ; Rushdi et Simoneit 2006 ; Simoneit *et al.* 2007], ce qui fait penser aux expériences de synthèse des phospholipides dans les années 1970 décrites plus haut. Cependant, la majorité des synthèses organiques hydrothermales qui ne font pas intervenir les réactions FTS n'ont aboutit jusqu'à présent qu'à la production de petites molécules (méthane, méthanol ou acide acétique) ou à l'allongement des molécules préexistantes plus longues [McCollom et Seewald 2007].

Ceci place les réactions FTS comme les plus prometteuses pour la synthèse de molécules amphiphiles dans des conditions hydrothermales mais pointe aussi les efforts considérables qui restent encore à faire pour démontrer la plausibilité de son application à des environnements primitifs, spécialement en ce qui concerne les précurseurs utilisés.

Des molécules amphiphiles dans la Terre primitive : synthèse exogène

L'accrétion de matériel organique d'origine extraterrestre peut être une alternative à la synthèse terrestre des molécules qui ont participé aux premiers pas de la vie. Les météorites [Chamberlin et Chamberlin 1908] puis les comètes [Oro 1961] ont été proposés comme

vecteurs capables de transporter dans l'environnement terrestre des matériaux organiques synthétisés dans l'espace. On admet aujourd'hui que les premiers temps de la Terre connurent un apport important de matière interplanétaire [Chyba et Sagan 1992] : on sait que la chute de ces corps à certains moments de l'Hadéen était massive [Gomes *et al.* 2005] et que, même aujourd'hui, l'accrétion de particules submillimétriques s'élèverait à plusieurs dizaines de milliers de tonnes par an [Love et Brownlee 1993]. Beaucoup de travaux se sont donc intéressés à la synthèse et à la présence de molécules organiques dans l'espace.

Dans un premier temps, il a été proposé que le fractionnement isotopique observé dans certains météorites puisse être expliqué par un mécanisme réactionnel analogue aux FTS en phase gazeuse [Lancet et Anders 1970]. Cependant, d'autres mécanismes de synthèse ont été proposés et le fait que les déséquilibres isotopiques ne soient pas les mêmes pour tous les types moléculaires suggère que la composition observée est probablement issue d'un mélange de différents types réactionnels [Pizzarello et Shock 2010]. Il a par exemple été montré en laboratoire que des analogues de glaces interstellaires et cométaires soumis à des ultraviolets ou à des radiations ioniques forment des molécules organiques [Bernstein *et al.* 1995 ; Gerakines *et al.* 2001]. Les chocs au sein de la poussière interstellaire pourraient également fournir une énergie pour la synthèse de molécules organiques [Henning et Salama 1998]. Ces conditions sont probables à la fois dans les nuages interstellaires et dans l'histoire des systèmes solaires. Leur reconstitution en laboratoire a permis d'identifier des hydrocarbures de type C_nH_m et des acides carboxyliques [Cottin *et al.* 1999], bien qu'ils ne soient pas présents sur les mêmes molécules.

Les indices signalant la possible présence de molécules organiques dans l'espace soutient l'origine exogène de ces molécules dans la période prébiotique. D'abord, une grande diversité moléculaire a été décelée dans les nuages moléculaires. On parle souvent de la détection spectroscopique de molécules polyaromatiques (PAH) et même d'acides aminés comme la glycine [Henning et Salama 1998] mais en raison des difficultés d'identification de ces molécules, ces résultats doivent être pris avec précaution. En réalité, les seules molécules dont la présence dans l'espace est fortement probable sont des petites molécules qui dépassent rarement la dizaine d'atomes [Thaddeus 2006]. En ce qui concerne les hydrocarbures, la détection de molécules de type C_nH_m en laboratoire et la proposition que des voies de synthèse existent dans les nuages stellaires [Henning et Salama 1998] semblent contredites par le fait que les principales chaînes carbonées observées comportent de nombreux groupements alcynes et nitriles à intérêt membranaire contestable [Thaddeus 2006].

Les comètes ont pendant longtemps fait l'objet d'une attention particulière car on considérait qu'elles étaient formées par l'agrégation de poussières interstellaires qui n'auraient pas été accrétées dans les planètes au moment de la formation du système solaire. On a donc supposé que les comètes étaient des témoins de la composition du nuage moléculaire à l'origine de notre système solaire et qu'elles auraient pu ainsi ensemençer la Terre primordiale avec les molécules détectables dans les nuages moléculaires. Cependant, cette hypothèse est souvent rejetée aujourd'hui car on pense que les conditions de formation du système auraient dû détruire ou modifier fortement cette composition originelle et l'on sait que la composition des comètes varie en fonction de leur lieu d'origine et de leur histoire [Cottin *et al.* 1999 ; Ehrenfreund et Cami 2010]. On privilégie donc de nos jours l'idée que la composition organique des corps interplanétaires serait le résultat de synthèses plus récentes dans le système solaire.

Une question très similaire se pose en ce qui concerne la préservation de ces molécules lors de leur entrée dans l'atmosphère terrestre, car la chaleur libérée à ce moment-là devrait détruire une grande partie des molécules organiques portées par les corps interplanétaires. Des calculs ont été conduits pour démontrer que la pression de CO₂ plus grande dans l'atmosphère de l'époque aurait pu freiner les corps interplanétaires plus efficacement qu'aujourd'hui [Chyba *et al.* 1990] mais il est également intéressant de signaler que cette chaleur a aussi été proposée comme une source énergétique pour la synthèse de composés organiques [Chyba et Sagan 1992]. Quoi qu'il en soit, la meilleure preuve de la résistance de ces molécules à l'entrée dans l'atmosphère est probablement leur détection dans les météorites chondritiques tombés récemment sur la Terre.

Les *chondrites* sont des météorites rocheux qui ne montrent pas de traces de métamorphisme ou de modifications dues à la chaleur. Elles sont formées par l'agrégation de poussières et de petits grains qui étaient présents dans le système solaire primordial et qui n'auraient pas acquis des tailles suffisantes pour permettre la différenciation des matériaux caractéristiques des corps planétaires. Elles sont majoritairement composées de silicates et contiennent des *chondrules* (petits grains qui donnent leur nom aux chondrites) entourés par une matrice rocheuse [Pizzarello et Shock 2010]. Elles portent souvent des marques de chocs ainsi que des traces d'eau à l'état liquide. Leur aspect est très similaire à celui des roches terrestres et de ce fait il est difficile de les trouver, excepté lorsqu'on les voit tomber (c'est le cas du célèbre météorite de Murchison) ou lors d'expéditions d'échantillonnage dans des déserts comme ceux de l'Antarctique (où les roches au dessus de la calotte ne peuvent venir

que du ciel). Ces expéditions ont permis la découverte de la majorité des météorites connus aujourd'hui [Bischoff 2001]. Il existe plusieurs types de chondrites, dont les *carbonées* qui, comme leur nom l'indique, ont la particularité de contenir des quantités remarquables de carbone (entre 1,5% et 4%), souvent sous forme de composés organiques [Pizzarello et Shock 2010]. Ces météorites sont aujourd'hui considérés comme des sources d'information très prometteuses sur les molécules organiques de la Terre primitive, bien qu'il faille rester prudent car ils ne constituent aujourd'hui que 2,5% des météorites identifiés sur Terre [Bischoff 2001].

Les chondrites carbonées ont été étudiées depuis très longtemps mais les analyses réalisées au début du XX^{ème} siècle sont suspectées de contaminations biotiques terrestres. Depuis 50 ans, le développement des sciences spatiales a permis de mieux étudier de nouveaux météorites, le plus abondamment étudié étant probablement celui tombé près de Murchison (Australie) en 1969. Les molécules organiques du météorite de Murchison peuvent être séparées en une fraction insoluble majoritaire (70% des composants organiques) et en une fraction soluble [Pizzarello et Shock 2010]. La fraction insoluble est souvent comparée au kérogène car il s'agit d'un matériel macromoléculaire avec une composition complexe dont le détail reste inconnu. Une bonne partie de ce matériel semble être constituée de composés aromatiques reliés entre eux par des chaînes aliphatiques dont des fragments composent aussi la fraction insoluble. La fraction soluble contient de très nombreuses molécules dont des polyalcools et des acides aminés très divers et inconnus dans la biologie terrestre. Bien que sa composition totale soit très compliquée et n'ait pas encore été intégralement étudiée, on sait que cette partie contient des molécules amphiphiles intéressantes. Elle contient notamment des acides monocarboxyliques allant de 2 à 8 carbones pour lesquels la plupart des isomères branchés possibles sont présents et d'autant plus rares qu'ils comportent une grande quantité d'atomes [Yuen et Kvenvolden 1973 ; Lawless et Yuen 1979]. La diversité des isomères trouvés dans le météorite contraste avec la préférence biologique des hydrocarbures linéaires, ce qui est en général considéré comme une preuve de l'origine extraterrestre de ces composés [Yuen et Kvenvolden 1973]. Des études isotopiques ont essayé de déterminer l'origine exogène de ces molécules [Epstein *et al.* 1987] de même que le mécanisme par lequel ces composés ont pu être synthétisés [Lawless et Yuen 1979 ; Yuen *et al.* 1984]. Pendant longtemps les observations faites sur le météorite de Murchison ont été considérées comme généralisables mais les analyses récentes, notamment sur des échantillons antarctiques, montrent une plus grande diversité météoritique non réductible au cas de cet unique météorite.

Par exemple, contrairement à ce qui a été observé sur Murchison, certaines chondrites carbonées ne comportent pas d'acides monocarboxyliques ; d'autres en contiennent mais avec des composants plus grands (jusqu'à 12 carbones) et dans une proportion de chaînes aliphatiques supérieure à celle des composés branchés [Naraoka *et al.* 1999 ; Pizzarello et Shock 2010].

En résumé, les résultats obtenus jusqu'à présent sont prometteurs car ils fournissent des preuves de principe de la possible présence de composants organiques d'intérêt membranaire sur la Terre primitive. Cependant, ces résultats exigent aussi de la prudence car il est difficile aujourd'hui de déterminer dans quelle mesure chaque mode de synthèse et de transport interplanétaire a pu influencer la composition organique de la Terre primitive. Il reste donc encore beaucoup de recherches à faire pour déterminer l'efficacité des synthèses proposées sur Terre, leur application à l'espace et les moyens par lesquels ces composés se sont retrouvés ensemble dans des concentrations nécessaires à l'apparition de la vie. En attendant, beaucoup de travaux se sont concentrés sur le moyen par lequel les molécules amphiphiles décrites jusqu'ici auraient pu former les premières membranes, et les propriétés de celles-ci.

Auto-assemblage spontané de vésicules dans l'environnement prébiotique

En partant du principe que des molécules amphiphiles pouvaient se retrouver dans les milieux aqueux de la Terre primitive, certains travaux se sont concentrés sur la formation spontanée de membranes dans ces conditions. Un des travaux fondateurs de cette discipline est l'expérience réalisée par David Deamer sur des extraits du météorite de Murchison. La partie non polaire de la fraction soluble d'un échantillon de ce météorite fut extraite à l'aide d'un solvant et le mélange obtenu mis au contact de l'eau. Des vésicules multilamellaires se formèrent spontanément et furent observées au microscope. L'introduction d'un marqueur fluorescent qui resta piégé plusieurs minutes à l'intérieur de la vésicule illustre la stabilité de la structure [Deamer 1985 ; Deamer et Pashley 1989]. Depuis, d'autres travaux ont montré des résultats similaires à partir de mélanges différents. Par exemple l'application de rayonnement ultraviolet à un mélange supposé simuler les particules interstellaires (eau, méthanol, ammoniac et monoxyde de carbone) a permis de fabriquer un mélange complexe de composés dont le contact avec l'eau permet la synthèse de structures vésiculaires [Dworkin *et al.* 2001]. Dans ces deux cas, on ne sait pas quelles molécules sont responsables de l'auto-assemblage car la composition organique exacte des mélanges n'est pas connue. Cependant, la richesse du mélange a conduit à proposer que ces molécules ne soient pas seulement

capables de former des bicouches mais aussi de dissoudre des composants aromatiques qui pourraient agir comme pigments photosensibles d'importance énergétique [Morowitz *et al.* 1988 ; Dworkin *et al.* 2001 ; Cape *et al.* 2011].

Néanmoins, l'utilisation de mélanges complexes rend difficile l'étude en laboratoire et la plupart des travaux concernant la formation de vésicules à partir de molécules amphiphiles s'est concentrée sur des modèles de laboratoire monolamellaires contenant moins de types moléculaires différents. Les premiers travaux ont utilisé des phospholipides [Hargreaves *et al.* 1977 ; Hargreaves et Deamer 1978] mais l'idée que la synthèse prébiotique de ces molécules était peu probable a conduit à préférer l'auto-assemblage de molécules amphiphiles plus petites qui, en contrepartie, sont aussi plus sensibles aux conditions environnementales.

En effet, plusieurs paramètres déterminent l'auto-assemblage des molécules amphiphiles en gouttelettes, micelles ou vésicules [Tanford 1973] : (1) le pH détermine l'état d'ionisation de la tête polaire et par conséquent modifie la solubilité des parties hydrophiles [Apel *et al.* 2002] ; (2) les concentrations des molécules amphiphiles en solution doivent être supérieures à certains seuils pour déplacer l'équilibre en faveur de la formation de micelles ou de vésicules [Deamer 1986 ; Monnard et Deamer 2003] ; (3) la taille des chaînes hydrophobes est un paramètre majeur car il détermine la solubilité des molécules dans l'eau et, par conséquent, la stabilité des membranes ; par exemple, bien que cette formation dépende aussi des autres paramètres, les chaînes de moins de 10 carbones ont tendance en règle générale à former des micelles alors que les chaînes de plus de 12 carbones forment plutôt des vésicules [Deamer 1986] ; la longueur des chaînes conditionne aussi la fluidité au sein des membranes et la perméabilité de celles-ci, caractéristiques qui vont de pair avec un autre paramètre, (4) la température qui facilite ou rend plus difficile les mouvements des molécules dans la bicouche [Deamer 1986] ; (5) alors que les phospholipides sont assez résistants à la concentration ionique, les acides gras et les alcools gras qui sont privilégiés dans les études sur les origines de la vie précipitent à des très faibles concentrations d'ions cationiques bivalents, comme le Ca^{2+} , le Mg^{2+} ou le Fe^{2+} ; cette caractéristique est d'autant plus importante que les ARN qui sont fréquemment utilisés comme modèles du matériel génétique primitif ont souvent besoin de ces cations pour réaliser une partie de leurs fonctions [Monnard *et al.* 2002].

Les incertitudes qui existent sur la nature et la concentration des molécules amphiphiles dans la Terre primitive ainsi que sur les conditions ioniques, de pH et de température rendent

difficile la proposition d'un modèle convaincant d'auto-assemblage prébiotique de vésicules basées uniquement sur des acides gras. En revanche, lorsque les acides gras sont complétés avec d'autres molécules, il se produit une stabilisation des membranes qui deviennent ainsi moins sensibles aux paramètres précédemment évoqués. Les principales molécules candidates ont été les alcools gras et les glycérolipides²⁸ dont la synthèse hydrothermale semble plausible [Apel et Deamer 2005 ; Rushdi et Simoneit 2006 ; McCollom et Seewald 2007 ; Simoneit *et al.* 2007]. Cette question pointe l'importance des mélanges pour la formation de vésicules prébiotiques. D'une part, il semble raisonnable de penser que la Terre ancienne devait contenir un mélange de composés, plutôt que des solutions pures [Oparin 1938 ; Deamer 1986]. D'autre part, les vésicules dont les membranes sont composées par des mélanges de molécules ont une résistance aux conditions environnementales beaucoup plus large que les vésicules dont les membranes ne contiennent qu'un seul type de molécule. Plus exactement, les membranes mélangées sont plus stables même lorsque leurs chaînes sont plus courtes ou se trouvent dans des concentrations plus faibles [Apel *et al.* 2002 ; Mansy et Szostak 2008]. Les membranes mélangées résistent mieux aussi à des rangs très différents de pH, à des concentrations cationiques plus fortes et à la température [Apel *et al.* 2002 ; Monnard *et al.* 2002 ; Mansy et Szostak 2008 ; Maurer *et al.* 2009]. Un moyen supplémentaire et non exclusif d'établir des vésicules stables dans des conditions diverses est l'utilisation de vésicules multilamellaires plutôt que monolamellaires [Zhu et Szostak 2009].

Par ailleurs, les vésicules d'acides gras n'ont pas été les seules à avoir fait l'objet d'investigations. Des vésicules prébiotiques d'isoprénoïdes ont aussi été proposées [Ourisson et Nakatani 1994] malgré la mise en question de l'existence de mécanismes abiotiques de synthèse de ces molécules [Monnard et Deamer 2002]. Les polyprényl phosphates sont des chaînes d'isoprénoïdes attachées à des groupements phosphates (partie polaire) qui forment des vésicules dans l'eau [Plobeck *et al.* 1992]. Comme dans le cas des acides gras, la formation de vésicules dépend de la taille des chaînes, du pH et de la concentration en ions qui, dans ce cas, favorise la formation de bicouches au lieu de les faire précipiter [Ariga *et al.* 2005 ; Gotoh *et al.* 2006]. De même, l'utilisation d'un mélange de polyprényl phosphates et d'alcools polyprényls stabilise les vésicules et permet leur constitution dans des conditions moins restrictives que les compositions pures [Streiff *et al.* 2007].

²⁸ Les glycérolipides sont des molécules formées d'une ossature de glycérol et d'une seule chaîne courte d'acides gras dont la synthèse prébiotique.

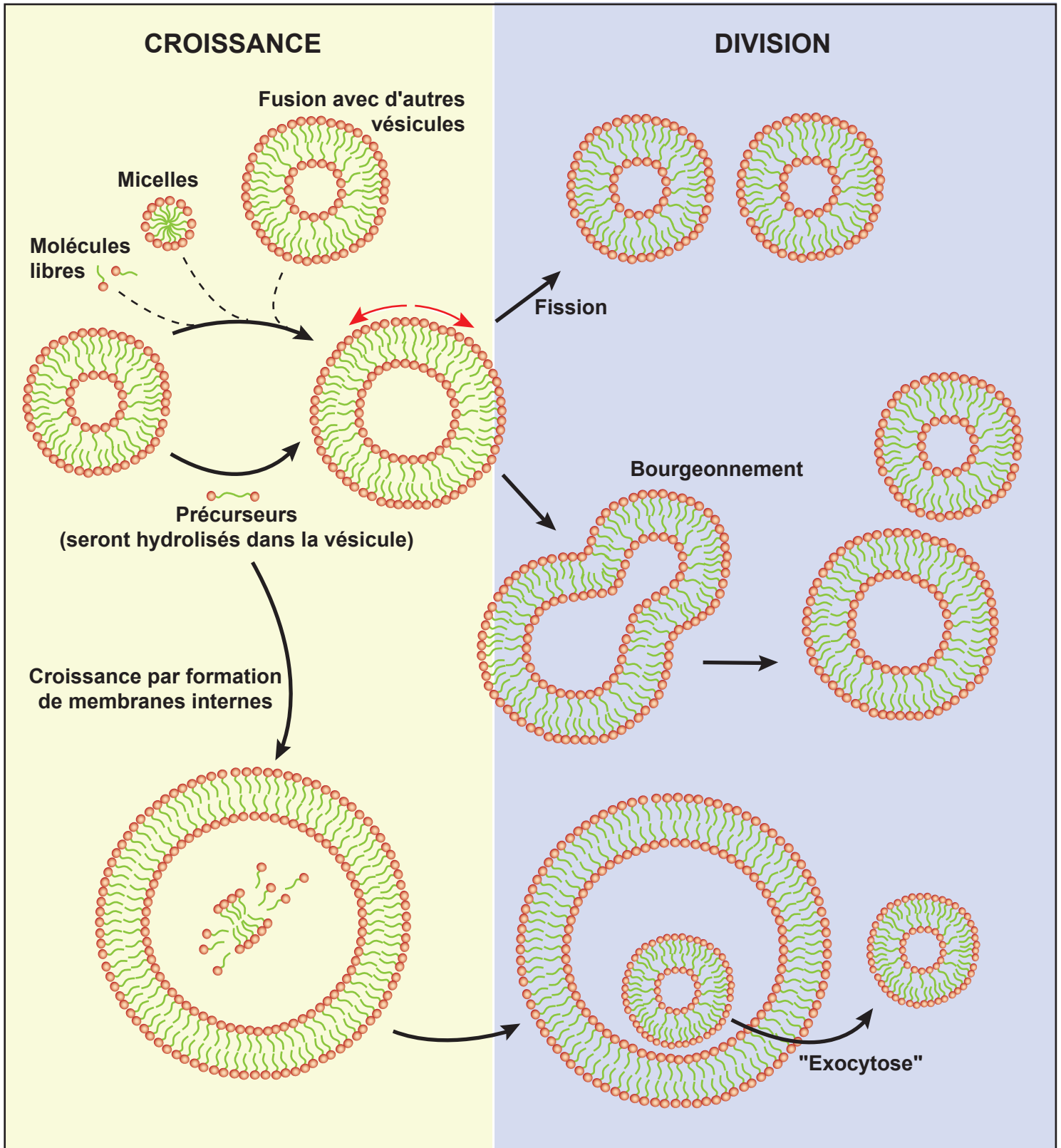


Figure 36. Croissance et division abiotique de vésicules.

Croissance. Les vésicules peuvent s'agrandir en intégrant soit des molécules amphiphiles libres, soit faisant partie de micelles ou en fusionnant avec d'autres vésicules. Elles peuvent aussi incorporer des précurseurs dont les produits après transformation pourront intégrer les membranes. Dans certains cas l'assemblage des bicouches peut se faire à l'intérieur des vésicules. **Division.** Lorsque les surfaces membranaires sont grandes les perturbations du milieu peuvent être à l'origine d'une fission des vésicules. Si l'incorporation des membranes est très asymétrique il peut y avoir aussi bourgeoisement spontané. Enfin, lorsque des vésicules se trouvent à l'intérieur d'autres vésicules, celles à l'intérieur peuvent sortir des plus grandes en fusionnant momentanément leurs membranes.

Ces lignes de recherche sont prometteuses mais elles montrent aussi que la formation prébiotique de vésicules est un processus plus compliqué de ce que l'on pouvait penser au premier abord. Nous allons cependant explorer les principaux mécanismes de croissance et de division vésiculaire qui sont actuellement évoqués dans les scénarios prébiotiques.

Croissance et division de vésicules prébiotiques

Pour que des liposomes puissent être des bons modèles de protocellules, il est souhaitable que de nouvelles vésicules soient issues des précédentes afin d'assurer une continuité (Figure 36). Morowitz a proposé que la formation de nouvelles molécules amphiphiles à partir de précurseurs externes permette d'augmenter l'extension des membranes. A terme, lorsque leur augmentation de taille les rendrait instables, les vésicules se diviseraient spontanément [Morowitz *et al.* 1988]. On a donc commencé par tester l'accroissement de micelles [Bachmann *et al.* 1992] et de vésicules [Walde *et al.* 1994] par l'incorporation de précurseurs hydrolysés par les lipides eux-mêmes. Plus récemment, beaucoup de travaux ont étudié l'effet de l'apport direct de molécules amphiphiles (acides gras à longue chaîne carbonée, généralement) à des vésicules déjà existantes. Lorsque les acides gras sont apportés sous forme de micelles on observe une multiplication du nombre de vésicules [Berclaz *et al.* 2001] mais seule une partie des nouvelles vésicules semble provenir de la division des vésicules précédentes, les autres se formant *de novo* [Hanczyc et Szostak 2004]. De fait, d'autres travaux ont signalé que l'ajout de matériel devait se faire de façon progressive et en petites quantités pour qu'il intègre les vésicules antérieures plutôt que d'en créer de nouvelles [Hanczyc *et al.* 2003 ; Chen et Szostak 2004a], (Figure 37). L'apport d'acides gras d'une vésicule à une autre est également possible, créant ainsi une compétition entre les deux. En effet, lorsque des vésicules avec une forte pression osmotique de sucrose sont mélangées à des vésicules vides, les premières incorporent une partie des acides gras des secondes, probablement à cause de la dilatation provoquée par la pression osmotique, qui détend la membrane et libère ainsi de la place pour de nouvelles molécules [Chen *et al.* 2004].

Sur la division des membranes (Figure 36), la plupart des travaux considère comme plausible le fait que l'accroissement de taille conduise à des structures instables qui finissent par se fractionner spontanément [Berclaz *et al.* 2001]. D'autres proposent des mécanismes physiques simples pour accélérer la division des vésicules, par exemple le passage à travers des pores de petite taille [Hanczyc *et al.* 2003]. Dans les membranes composées de différents types de molécules (phospholipides, sphingomyéline et cholestérol), les molécules ont

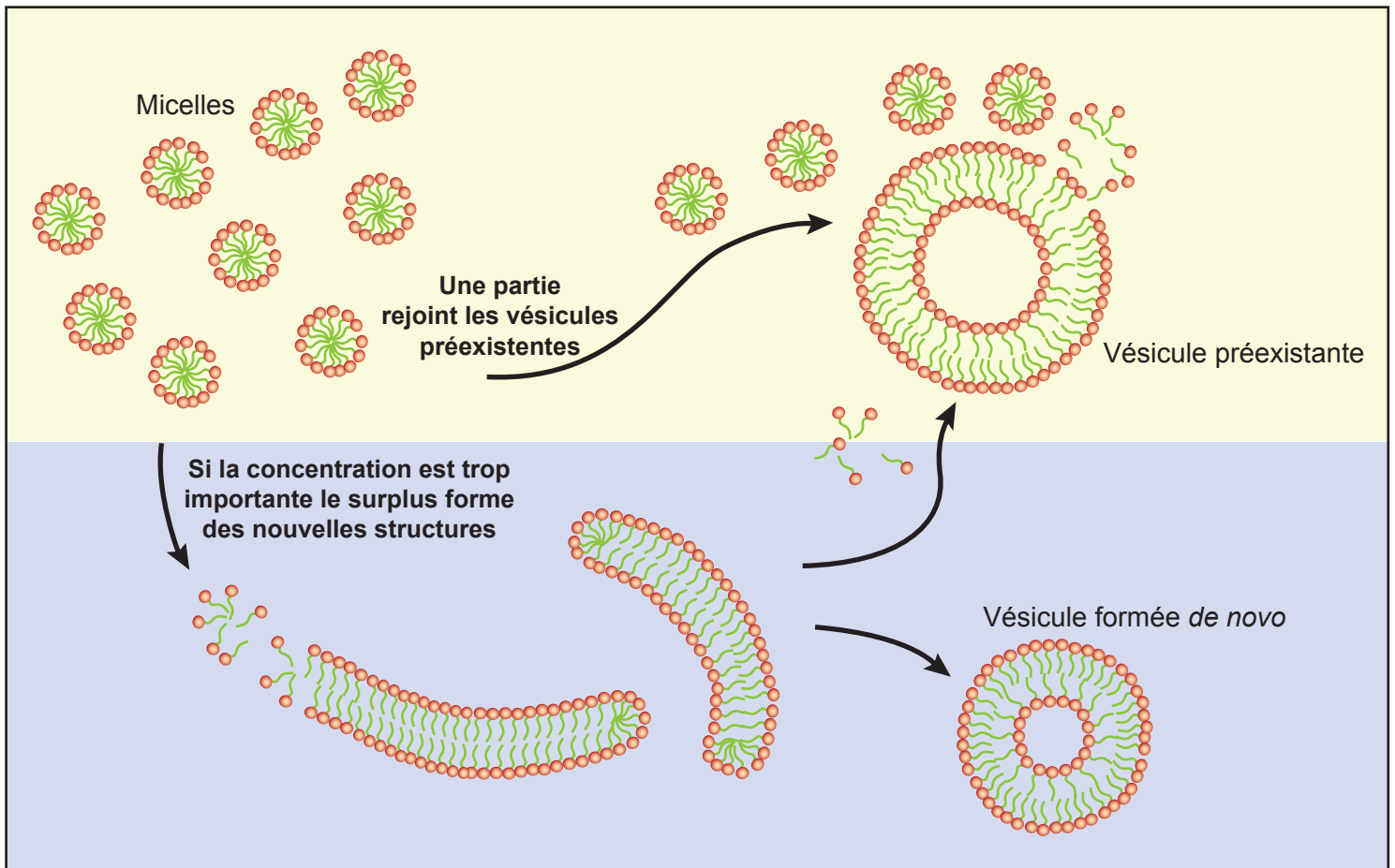


Figure 37. Croissance vésiculaire ou formation *de novo*.

Si la concentration de micelles dans le milieu est faible, les molécules amphiphiles rejoindront très majoritairement les vésicules préexistantes (partie supérieur du schéma). Si la concentration est plus grande, les membranes des vésicules préexistantes ne pourront pas accueillir toutes les molécules amphiphiles à la fois. Ainsi, seulement une partie des molécules des micelles les rejoindra, pendant que le reste forme des structures compliquées. Bien qu'une partie des molécules des nouvelles structures puisse intégrer tardivement les vésicules préexistantes, la plupart formera des nouvelles vésicules.

tendance à ségréger et à adopter des courbures différentes qui peuvent conduire à la séparation définitive en plusieurs vésicules qui ne partageront pas la même composition que la vésicule mère [Baumgart *et al.* 2003]. Cependant, les vésicules sont en général assimilées à des sphères et l'on sait que les rapports volume/surface des sphères changent en fonction de leur rayon. En conséquence, la division en deux d'une vésicule sphérique entraîne la perte d'une partie du volume initial, ce qui est un modèle peu souhaitable dans le contexte biologique. Des mécanismes classiques de bourgeonnement ont été proposés [Takakura et Sugawara 2004], de même qu'une formation de nouvelles vésicules à l'intérieur des vésicules existantes, suivie d'une sorte d'exocytose [Wick *et al.* 1995].

Mais la réponse qui semble aujourd'hui la plus prometteuse est celle utilisant des vésicules multilamellaires au lieu des monolamellaires classiques. Il a été observé que l'ajout d'acides gras à des vésicules multilamellaires ne conduit pas à une augmentation uniforme de la surface vésiculaire mais à la formation de prolongements filamenteux qui sont assez fragiles et se séparent de la cellule mère assez facilement, formant une nouvelle vésicule multilamellaire contenant une partie du volume original de la vésicule mère sans perte de volume global [Zhu et Szostak 2009]. Cette hypothèse est apparue d'autant plus séduisante qu'un mécanisme similaire de division a été mis en évidence dans une souche de la bactérie *Bacillus subtilis* ne comportant ni paroi cellulaire ni mécanisme Fts de division cellulaire [Leaver *et al.* 2009]. Cela suggère que le mécanisme de division des vésicules multilamellaires est propre aux bicouches lipidiques et a pu être conservé entre les premières vésicules et les premières cellules, avant l'apparition de mécanismes cellulaires de division qui caractérisent les organismes modernes [Chen 2009]. Pourtant, les études concernant les propriétés biophysiques des membranes ne devraient pas exclure la possibilité que le matériel contenu dans les vésicules ait influencé les divisions de celles-ci. Par exemple, bien que cela reste encore très hypothétique, il a été proposé que des ARN particuliers avec une affinité spécifique pour les membranes puissent contrôler la forme adoptée par les membranes et ainsi réguler la division vésiculaire [Vlassov *et al.* 2001].

Quoi qu'il en soit, il est intéressant de signaler comment les récentes observations sur la formation, la croissance et la division des vésicules plaident en faveur de l'utilisation de systèmes membranaires plus sophistiqués qu'admis jusqu'à présent. L'utilisation de vésicules monolamellaires avec un seul composant amphiphile a été longtemps justifiée par la difficulté à travailler sur des systèmes avec plus d'éléments mais les propriétés plus intéressantes des vésicules multilamellaires à composition mixte et leur plus grande plausibilité dans les

conditions prébiotiques sont un défi à l'état actuel de notre connaissance des origines des compartiments lipidiques et elles devraient encourager le développement de nouvelles techniques pour l'étude de ces structures [Szostak 2011]. Une autre critique que l'on peut adresser aux études sur la croissance et la division des vésicules est leur utilisation d'acides gras à longue chaîne carbonée au lieu de molécules dont la synthèse abiotique semble plus plausible [Thomas et Rana 2007], ce qui laisse encore la porte ouverte à de nombreuses perspectives dans ce domaine.

De la vésicule à la protocellule

Nous avons résumé les mécanismes de synthèse des molécules amphiphiles et de formation, croissance et division des vésicules mais pour que ces structures aient un intérêt biologique elles doivent aussi contenir du matériel génétique et si possible un métabolisme. Beaucoup de travaux parlent de *protocellules* pour désigner des vésicules membranaires contenant un matériel génétique et la machinerie nécessaire à sa réplication *sans que les mécanismes soient couplés entre eux*. Dans ce cas, il est considéré que la différence entre ces protocellules et les cellules (qui dans ces modèles apparaissent ultérieurement) réside dans l'indépendance du mécanisme de réplication génétique par rapport à la division membranaire [Szostak *et al.* 2001]. Ce scénario n'est pas le seul à avoir été proposé mais il est majoritaire au sein de la communauté travaillant sur les membranes dans les premières étapes de la vie. Pour l'instant je vais donc me limiter à ce modèle et récapituler comment des liposomes peuvent encapsuler des molécules d'intérêt biologique.

Parmi les premiers mécanismes d'encapsulation proposés se trouvent les cycles de déshydratation/réhydratation : lorsqu'une solution contenant des vésicules est soumise à dessiccation, les vésicules s'agrègent et forment des structures multilamellaires. Si cette solution contient aussi d'autres molécules, ces substances se retrouvent alors mélangées aux structures multilamellaires. Une fois réhydratés, les lipides reforment de nouvelles vésicules qui contiendront une partie des molécules qui initialement étaient à l'extérieur. Ceci a notamment été testé sur des macromolécules comme l'ADN et sur des ions [Deamer et Barchfeld 1982]. De tels cycles réalisés à haute température et en présence de ribonucléotides ont d'ailleurs montré que les lipides peuvent à la fois faciliter la synthèse d'oligomères entre 25 et 75 nucléotides pendant les phases sèches du cycle et les retenir à l'intérieur pendant les phases hydratées [Rajamani *et al.* 2008]. L'encapsulation peut aussi se produire lors de la formation *de novo* des liposomes, par exemple grâce à la modulation du pH. Dans cette

technique on prépare une solution contenant les acides gras et les macromolécules que l'on souhaite encapsuler à un pH impropre à la formation de vésicules puis, en ramenant le pH à des valeurs compatibles avec la formation de vésicules, on produit de telles structures contenant une partie des macromolécules [Monnard et Deamer 2003].

Plusieurs travaux se sont intéressés au devenir des molécules à l'intérieur de la membrane : la réplication non enzymatique d'oligonucléotides de cytosine a été mise en évidence dans des vésicules à composition mixte et on a observé que la présence des vésicules facilite la synthèse d'oligopeptides [Tsukahara *et al.* 2002 ; Zepik *et al.* 2007 ; Mansy *et al.* 2008]. Par ailleurs, les macromolécules sont protégées des agents de dégradation extérieurs (protéases, nucléases) et peuvent réaliser leur activité à l'intérieur des vésicules. L'encapsulation de macromolécules, leur protection et leur maintien d'activité ont été observés pour l'ADN [Apel *et al.* 2002], des ribozymes [Chen *et al.* 2005], des ARN-polymérase capables de synthétiser l'ARN soit *de novo* soit à partir d'une matrice ADN [Chakrabarti *et al.* 1994 ; Monnard *et al.* 2007] et des protéines réalisant des activités métaboliques diverses [Apel *et al.* 2002]. Cependant, les travaux qui ont essayé d'étudier les possibilités de croissance des vésicules en parallèle avec des activités enzymatiques internes ont été assez rares jusqu'à présent. L'équipe de Luisi est celle qui s'est probablement le plus concentrée sur ce problème. Elle a par exemple réussi à montrer la croissance de vésicules d'acide oléique à partir de précurseurs externes en même temps que des protéines encapsulées dans ces vésicules (des polynucléotide phosphorylases) synthétisaient des oligomères polyadénylés d'ARN [Walde *et al.* 1994]. Mais ces modèles doivent encore se rapprocher des protocellules en essayant de suivre à la fois l'auto-réplication du matériel génétique et la multiplication des vésicules et ce en utilisant des composants amphiphiles dont la synthèse abiotique semble plus vraisemblable que celle des acides gras à chaîne longue qui ont été utilisés jusqu'à présent [Thomas et Rana 2007].

Propriétés des membranes protocellulaires

Comme nous l'avons vu, les vésicules protègent les macromolécules encapsulées et favorisent certaines réactions mais ce faisant elles ne doivent pas empêcher d'autres molécules plus petites de traverser les membranes afin d'assurer les besoins internes de la protocellule et d'éliminer ses déchets. Cette perméabilité sélective²⁹ constitue l'une des

²⁹ D'ailleurs, la perméabilité des membranes a pu avoir des répercussions historiques très importantes. Par exemple, il a été montré que les membranes composées d'acides gras sont plus perméables au ribose qu'aux

principales caractéristiques des membranes modernes. Néanmoins, comme ces membranes contemporaines, composées de phospholipides, sont très peu perméables naturellement, elles nécessitent la présence d'un grand nombre de protéines spécifiques pour transporter les molécules. Or la présence de ces protéines semble exclue dans les premières structures vivantes. Ce paradoxe est souvent résolu par le fait que les membranes d'acides gras ou de glycérolipides considérées dans les études prébiotiques sont perméables à plus de molécules que les membranes modernes [Monnard et Deamer 2001 ; Mansy 2010]. La difficulté réside en ce que chaque membrane a une perméabilité différente selon les molécules qui la constituent et selon les conditions ioniques, de pH et de température [Apel *et al.* 2002]. Ainsi, beaucoup de travaux ont signalé des perméabilités intéressantes de certaines membranes mais il est plus difficile de déceler des conditions qui conviennent simultanément à tous les processus recherchés.

La règle d'Overton stipule que les grandes molécules traversent plus facilement la membrane quand elles sont apolaires que quand elles sont polaires [Overton 1899]. Pour les petites molécules polaires le transport passif est plus compliqué [Walter et Gutknecht 1986 ; Paula *et al.* 1996]. Elles peuvent profiter de failles temporaires dans les membranes qui semblent plus fréquentes dans les membranes d'acides gras et de d'autant plus que les chaînes carbonées sont courtes. Certaines molécules peuvent occuper des espaces vides dans la membrane et finir par diffuser. Les acides gras qui composent des membranes ont aussi la capacité de changer de couche lipidique (alors que ce mouvement de *flip-flop* est beaucoup plus difficile dans les membranes phospholipidiques) de sorte que des molécules s'attachant à leur tête polaire peuvent profiter du basculement pour traverser la membrane. Les études citées précédemment sur la synthèse interne d'ARN et de protéines [Chakrabarti *et al.* 1994 ; Monnard *et al.* 2007 ; Zepik *et al.* 2007 ; Mansy *et al.* 2008] sont la preuve qu'un certain transport peut exister à travers des membranes d'acides gras au point de permettre le fonctionnement de structures internes. Néanmoins, il a aussi été postulé que les premières membranes aient bénéficié de mécanismes qui auraient spécifiquement facilité le transport de certaines substances. Le plus original de ces mécanismes est certainement celui des canaux formés de polyhydroxyalkanoates et polyphosphates dans certaines cellules modernes. Les premières molécules permettent leur insertion commune dans la membrane alors que les

autres sucres, ce qui a été proposé comme une possible raison de l'utilisation du ribose plutôt que d'une ose plus stable dans le matériel génétique, ARN ou ADN [Sacerdote et Szostak 2005].

deuxièmes forment un canal hydrophile qui peut être emprunté par des petites molécules [Reusch 2000]. Aujourd'hui la synthèse de ces molécules se fait par des voies métaboliques qui peuvent difficilement être évoquées à l'époque mais ce système est un modèle intéressant pour expliquer comment des composants relativement simples peuvent former des structures membranaires spécifiques. Plus classiquement, on a proposé des peptides aléatoires capables de transporter des composés [Pohorille *et al.* 2003] ou encore des ARN particuliers sélectionnés pour faciliter le transport de certains acides aminés [Janas *et al.* 2004 ; Vlassov 2005].

Etant donné que les membranes considérées dans les scénarios protocellulaires sont majoritairement composées d'acides gras mais que les membranes modernes sont principalement constituées de phospholipides, il est implicite dans ces modèles que le remplacement d'un type de membrane par l'autre a eu lieu à un moment donné, et ce avec un impact probablement important sur la perméabilité des membranes. Sans s'intéresser à l'origine des phospholipides, certains travaux récents ont essayé d'expliquer cette transition. Les membranes d'acides gras étant plus stables à haute température que leurs équivalents phospholipidiques, il a été proposé que le passage d'un type de membrane à l'autre ait été dû à l'adaptation à une plus basse température que celle dans laquelle les premières membranes seraient apparues [Furuuchi *et al.* 2005]. D'autres ont observé que lorsque des vésicules composées à la fois de phospholipides et d'acides gras libres étaient mises en présence de vésicules composées uniquement d'acides gras, les premières avaient tendance à dérober une partie de leurs acides gras aux deuxièmes, en s'accroissant par la même occasion. Dans ce contexte il a été proposé que des vésicules ayant acquis un mécanisme de synthèse des phospholipides aient absorbé et progressivement fait disparaître les vésicules d'acides gras [Budin et Szostak 2011].

Quel que soit le mécanisme par lequel le remplacement s'est produit, il semble raisonnable de penser qu'il a dû exister un processus de co-évolution des composants lipidiques et des composants protéiques des membranes, de sorte que plus les protéines prenaient en charge le transport de molécules, plus la barrière lipidique pouvait devenir sélective de ces molécules et réciproquement [Szathmary 2007]. Malheureusement, l'idée de *compartmentation progressive* a été parfois quelque peu surinterprétée comme on le verra dans le chapitre suivant à propos des ATPases. Par ailleurs, certaines approches bottom-up minimisent la difficulté à maintenir un gradient de protons attribuée aux liposomes d'acides gras. Il a notamment été montré que la croissance vésiculaire par acquisition d'acides gras

d'origine externe peut entraîner la formation de gradients de protons. Ceux-ci sont maintenus le temps de l'expérience s'il existe des cations divalents dans le milieu, de sorte que lesdits gradients auraient pu constituer une énergie utilisable pour les premières protocellules [Chen et Szostak 2004b]. Une autre proposition suggestive de bioénergétique ancienne a été formulée à la suite de l'observation de molécules aromatiques dans les mélanges supposés représenter correctement la composition de la Terre primitive car ces molécules auraient pu se dissoudre dans les premières membranes et agir ainsi comme pigments énergétiques [Morowitz *et al.* 1988 ; Dworkin *et al.* 2001]. Bien entendu, ces preuves de principe sont très préliminaires et doivent être interprétées avec prudence mais elles encouragent à explorer une énergie utilisable par les premières structures vivantes et nous rappellent la difficulté à reconstruire la séquence historique d'événements aussi anciens.

Membranes prébiotiques et principe de continuité

Comme toutes les approches bottom-up des origines de la vie, l'étude des premières membranes est soumise à la proposition de preuves de principe de la possibilité de certains processus sans que pour autant on sache encore précisément comment la vie est apparue. Nous avons vu plusieurs preuves de principe relatives à la synthèse et accumulation de molécules amphiphiles sur la Terre primitive, ainsi que des hypothèses sur les mécanismes par lesquels ces molécules ont pu former des vésicules, croître, se diviser et incorporer d'autres molécules indispensables au métabolisme et à l'information génétique.

Pour la communauté des chercheurs travaillant sur les origines des membranes, il semble que le recours à des systèmes comportant de plus en plus de composants (molécules différentes et vésicules multilamellaires) devienne un impératif [Chen et Walde 2010 ; Szostak 2011]. D'autres idées encore plus audacieuses ont été proposées pour souligner la possible importance des membranes lipidiques et des vésicules dans des conditions prébiotiques. Par exemple, les molécules amphiphiles pourraient avoir un potentiel catalytique [Segre *et al.* 2001] et la composition même des vésicules pourraient être un support d'information héréditaire transmise de génération en génération et soumise à la concurrence [Morowitz *et al.* 1991 ; Segre *et al.* 2001]. D'une manière plus générale, étant donné que la vie actuelle est composée simultanément de membranes, de matériel génétique et d'un métabolisme, il ne fait pas de doute que l'explication de son origine nécessite l'intégration en un seul modèle de ces trois composantes, même si pour des raisons méthodologiques on a eu

plutôt tendance jusqu'à maintenant à étudier chaque constituant séparément [King 1990 ; Deamer 1997 ; Szostak *et al.* 2001 ; Moreno et Ruiz-Mirazo 2009 ; Schrum *et al.* 2010].

La similarité entre les lipides probables des premières formes de vie et les lipides actuels encourage à postuler une continuité entre les premières membranes et les membranes actuelles, ce qui semble être la possibilité la plus parcimonieuse [Morowitz *et al.* 1991]. Certes, la précision que l'on peut fournir aujourd'hui sur les premières membranes n'est pas satisfaisante mais les résultats résumés ici nous incitent à penser que, tout autant qu'un support génétique et une forme de métabolisme, les membranes ont été importantes pour l'apparition de la vie qui, rappelons-le, a précédé le cenancêtre. Tout en reconnaissant son imprécision, il importe de prendre en compte cette information lorsqu'on étudie la possible compartimentation du cenancêtre.

5. Des protéines membranaires chez le cenancêtre

Je vais exposer maintenant la dernière grande critique contre les hypothèses qui suggèrent l'absence de membranes lipidiques chez le cenancêtre : les membranes ne servent pas uniquement à délimiter la cellule mais sont aussi des compartiments métaboliques et énergétiques essentiels. De ce fait, elles contiennent de nombreuses protéines membranaires participant à la transformation d'énergie, au transport de molécules et à l'insertion des protéines dans la membrane. Certaines études évolutives ont pris comme sujet de travail ces protéines membranaires et suggéré leur présence chez le cenancêtre. Je vais donc résumer brièvement les implications de ces résultats en ce qui concerne non seulement la présence de membranes chez le cenancêtre mais aussi les fonctions que l'on peut leur attribuer.

Fonctions énergétiques de la membrane

Les ATPases sont le premier complexe protéique membranaire dont la présence chez le cenancêtre a été étudiée car ce sont des complexes protéiques universels composés, entre autres, de deux sous-unités qui sont paralogues entre elles, et qui ont été utilisées pour enraciner l'arbre du vivant [Gogarten *et al.* 1989] comme nous l'avons vu dans la première partie de l'introduction. Ces protéines sont capables d'utiliser un gradient ionique entre les deux versants de la membrane pour synthétiser de l'ATP, qui est la principale « monnaie énergétique » de la cellule, ou d'hydrolyser l'ATP pour le transport de certains solutés [Nakanishi-Matsui et Futai 2006]. Les ATPases sont donc parmi les quelques complexes minimaux dont on peut être assez certains de leurs présence chez le cenancêtre. Or, le fait que

leur fonction dépende d'un gradient transmembranaire soutient fortement la présence d'une enveloppe lipidique dans cet organisme. D'ailleurs, leur lien très fort avec les membranes a conduit récemment à formuler des hypothèses sur les propriétés des membranes anciennes à partir des informations disponibles sur l'évolution des ATPases.

En effet, des travaux sur les ATPases ont été interprétés sur la base de l'hypothèse que les premières membranes ont permis une compartimentation progressive des premières structures vivantes [Mulkidjanian *et al.* 2007 ; Szathmary 2007] dont j'ai brièvement parlé à la fin du chapitre précédent. Cette compartimentation progressive se traduirait entre autres par une réduction graduelle de la perméabilité des membranes qui, au départ, auraient été imperméables aux protéines et acides nucléiques mais perméables aux ions, puis elles seraient devenues imperméables à des molécules de plus en plus petites : le sodium et finalement les protons [Mulkidjanian *et al.* 2007]. Les ATPases fourniraient des arguments en faveur de cette hypothèse : il existe des ATPases capables de fabriquer de l'ATP à partir soit d'un gradient de Na^+ soit de H^+ . Les ATPases utilisant le H^+ semblant majoritaires, on a longtemps cru qu'elles étaient présentes chez le cenancêtre et que celles utilisant le Na^+ étaient apparues plus tard. Cependant, la distribution phylogénétique très large des ATPases à Na^+ a conduit certains auteurs à proposer que les ATPases les plus anciennes utilisaient le Na^+ avant d'utiliser le H^+ [Mulkidjanian *et al.* 2008]. Il a donc été proposé que les membranes du cenancêtre étaient imperméables au Na^+ mais perméables aux H^+ et que les ATPases de l'époque auraient utilisé le gradient de Na^+ . Comme le gradient de H^+ est plus efficace d'un point de vue bioénergétique que le gradient de Na^+ , puisqu'il est couplé à des chaînes de transport des électrons qui le génèrent, les membranes seraient devenues imperméables aux H^+ dans plusieurs lignées différentes postérieures au cenancêtre, permettant aux ATPases d'utiliser le H^+ à la place du Na^+ [Mulkidjanian *et al.* 2007 ; Mulkidjanian *et al.* 2009].

Cette hypothèse est intéressante car elle suggère des propriétés des membranes anciennes à partir de protéines dont la présence a été inférée chez le cenancêtre, mais elle est critiquable dans sa formulation. Même en acceptant que les ATPases du cenancêtre utilisaient du Na^+ (ce qui est une option mais doit encore être pris avec réserve), les membranes n'ont pas pu devenir imperméables aux H^+ uniquement *pour* permettre aux ATPases postérieures d'adopter un mécanisme énergétiquement plus favorable, puisque cela supposerait une « prévoyance » impossible de la part de l'évolution. Autrement dit, les membranes n'ont pas pu devenir plus serrées en prévision de la plus grande efficacité énergétique des ATPases mais au mieux au cours d'un processus de co-évolution des membranes, des chaînes de

transport des électrons et des ATPases. Il faut donc, encore une fois, traiter ce sujet d'une façon plus large en prenant en compte d'autres facteurs que les membranes et les ATPases proprement dites pour pouvoir comprendre leur évolution. Autrement dit, ce problème pose la question subsidiaire de l'existence de chaînes respiratoires dans le cenancêtre.

Dans les cellules modernes, les chaînes respiratoires sont des complexes membranaires qui utilisent des réactions redox pour transférer un électron depuis un donneur initial jusqu'à un accepteur final et utilisent l'énergie libérée pour transférer des protons d'un côté de la membrane à l'autre. Il existe des chaînes respiratoires très diverses dans tous les organismes vivants. Quatre d'entre elles ont été analysées et il a été proposé qu'elles aient été présentes chez le cenancêtre en raison de leur large distribution dans une grande diversité d'archées et de bactéries [Castresana et Moreira 1999]. La présence d'une chaîne respiratoire aérobie a été la plus contestée parce que son utilisation de l'O₂ comme accepteur final semblait contradictoire avec l'idée que l'O₂ n'existait pas dans l'atmosphère primitive et qu'il serait apparu beaucoup plus tard, lorsque les cyanobactéries ont mis en place la photosynthèse oxygénique [Klein *et al.* 1992]. Cependant, des analyses phylogénétiques récentes ont montré qu'au moins l'une des dioxygène réductases (A-O₂Red) [Brochier-Armanet *et al.* 2009 ; Sousa *et al.* 2012] existait déjà avant la divergence entre les trois domaines du vivant, ce qui suggère que même si la proportion d'O₂ était faible à l'échelle de l'atmosphère, une certaine quantité de ce gaz pouvait être disponible à petite échelle, par exemple à cause de la photolyse de l'eau dans l'atmosphère [Kasting 1993]. Des chaînes respiratoires basées sur le nitrate, le sulfate ou le soufre ont aussi été proposées chez le cenancêtre à cause de la détection d'homologues des protéines qui transfèrent les électrons aux accepteurs inorganiques terminaux dans une grande diversité d'archées et de bactéries [Castresana et Moreira 1999]. Cette question a été l'objet de débats pour essayer de déterminer si cette distribution très large dans les deux domaines procaryotes était due à l'héritage vertical depuis le cenancêtre ou à des transferts horizontaux plus récents (voir par exemple le débat sur la respiration du sulfate dans [Wagner *et al.* 1998 ; Larsen *et al.* 1999 ; Klein *et al.* 2001 ; Fukuba *et al.* 2003]).

Sans rentrer dans les détails des accepteurs finaux des chaînes respiratoires, il importe de signaler que certains composants des chaînes de transports des électrons, comme le cytochrome b, les ferredoxines 2Fe-2S et 4Fe-4S, la protéine de Rieske, la rubredoxine et la flavodoxine, sont aussi très répandus dans les génomes des trois domaines du vivant [Castresana et Moreira 1999]. Des phylogénies plus récentes [Vignais *et al.* 2001 ; Baymann *et al.* 2003 ; Lebrun *et al.* 2006] confortent la présence de certains de ces gènes chez le

cenancêtre et différents résultats soutiennent que la très grande diversité bioénergétique des organismes actuels soit due à différentes combinaisons d'une machinerie bioénergétique déjà présente chez le cenancêtre [Baymann *et al.* 2003 ; Schoepp-Cothenet *et al.* 2012]. Ces protéines étant par définition membranaires, leur présence chez le cenancêtre implique d'abord que cet organisme possédait des membranes lipidiques et deuxièmement que ces membranes étaient probablement imperméables aux protons et par conséquent dotées de propriétés physico-chimiques probablement très proches de celles de la plupart des membranes cellulaires modernes.

Adressage, insertion et translocation de protéines

La particule de reconnaissance du signal (SRP) et son récepteur (SR pour *SRP receptor*) interviennent respectivement dans la détection de peptides en cours d'élongation et dans leur attachement aux ribosomes concernés par l'élongation. Ce système permet à des protéines qui n'ont pas encore adopté leur configuration tridimensionnelle définitive soit d'être insérées dans les membranes soit d'être sécrétées à l'extérieur de la cellule. SRP et SR ont attiré l'attention des études évolutives anciennes car, comme les ATPases, ces protéines sont des paralogues réciproques et universels et, de ce fait, nous les avons vus aussi dans la première partie de l'introduction car leur phylogénie a été utilisée pour essayer d'enraciner l'arbre du vivant [Gribaldo et Cammarano 1998]. On peut donc dire qu'un mécanisme au moins d'adressage des protéines à la membrane existait déjà chez le cenancêtre.

Cependant, le système SRP/SR n'est probablement pas le seul à avoir existé chez le cenancêtre car trois autres mécanismes d'interaction entre les protéines et les membranes existent dans une grande diversité d'organismes. Le plus connu est la voie Sec (pour *secretion*) qui est responsable de l'export de certaines protéines à travers les membranes et de l'insertion d'autres protéines dans les membranes de nombreuses bactéries. Il existe des homologues du système Sec chez les archées mais, malgré la grande divergence qui sépare les séquences des deux domaines, le mécanisme utilisé par ce système semble conservé au point d'arriver à se compléter mutuellement [Auer *et al.* 1991]. La comparaison fonctionnelle de ces protéines et leurs phylogénies soutiennent d'ailleurs la présence de plusieurs des composants Sec chez le cenancêtre [Cao et Saier 2003]. Chez les bactéries le système Sec peut collaborer avec un autre système appelé YidC en référence au nom du principal gène de ce mécanisme [Yuan *et al.* 2010]. YidC est une intégrase qui peut fonctionner seule ou avec Sec et dont des comparaisons génomiques ont permis de détecter des homologues dans les

génomés d'une grande diversité d'euryarchées. Bien que la caractérisation biochimique des protéines archéennes n'ait pas encore été réalisée, la phylogénie de YidC est congruente avec la séparation en trois domaines du vivant, suggérant la présence de ces protéines chez le cenancêtre [Yen *et al.* 2001]. Le troisième système connu de translocation protéique est la voie Tat (*twin arginine targeting*). Ce mécanisme a la particularité d'être le seul capable de transporter à travers les membranes des protéines ayant déjà adopté leur conformation active [Palmer et Berks 2012]. C'est un complexe formé d'au moins deux sous-unités, l'une étant un pore transmembranaire et l'autre étant un peptide de reconnaissance. Les deux existent chez les bactéries, les chloroplastes et les archées [Yuan *et al.* 2010]. Bien que chez les archées la plupart des caractérisations se soient concentrées chez les halophiles, des analyses phylogénomiques montrent que les deux composants sont répandus à la fois chez les bactéries et les archées, et sont probablement ancestraux à chaque lignée procaryote, donc probablement aussi présents chez le cenancêtre [Yen *et al.* 2002]. Finalement, et bien que des phylogénies ne soient pas encore disponibles pour ces enzymes, il semblerait que non seulement les mécanismes de translocation mais aussi les protéines de traitement des signaux de reconnaissance soient anciens. Il s'agit notamment des peptidases qui coupent les signaux des protéines précurseurs une fois qu'elles ont été adressées à la membrane, et dont les homologues sont répandus dans les trois domaines du vivant [Eichler 2002 ; Ng *et al.* 2007].

L'ensemble de ces résultats suggère qu'au moins quatre machineries de translocation ou d'insertion des protéines dans les membranes ainsi peut-être que la machinerie de maturation protéique existaient chez le cenancêtre. Cette richesse de mécanismes de traitement des protéines membranaires n'est pas seulement importante à cause de ses implications pour l'utilisation de protéines membranaires ou extracellulaires chez le cenancêtre mais aussi à cause des implications de leur présence chez le cenancêtre pour d'autres composants cellulaires. Notamment, on sait que les mécanismes Sec, YidC et Tat utilisent tous les trois le gradient de protons (ou, rarement, un gradient de sodium, [Kwan *et al.* 2008]) à travers la membrane comme source d'énergie [Brundage *et al.* 1990 ; Mould et Robinson 1991 ; He et Fox 1997 ; Yahr et Wickner 2001]), ce qui représente un argument supplémentaire en faveur de la présence d'un tel gradient chez le cenancêtre. Cette possibilité semble d'autant plus plausible que les machineries YidC et Tat participent activement à l'insertion dans les membranes de certains composants des ATPases et des chaînes respiratoires comme la protéine de Rieske et des cytochromes [Yuan *et al.* 2010 ; Palmer et Berks 2012].

Protéines et fonctions membranaires chez le cenancêtre

La présence chez le cenancêtre d'ATPases et de chaînes de transport des électrons, qui sont des protéines membranaires dont la fonction biologique n'acquiert de sens que dans la mesure où elles créent ou utilisent un gradient ionique transmembranaire, est un argument très fort, bien qu'indirect, en faveur du caractère incontournable de la présence de la membrane lipidique dans n'importe quel modèle représentant le cenancêtre. On ne décompte pas moins de quatre machineries de translocation et d'intégration des protéines dans les membranes, ce qui suggère que les capacités de la membrane du cenancêtre en matière de sécrétion et d'incorporation de protéines devaient être très similaires à celles des cellules modernes. Ainsi, même si la reconstruction phylogénétique d'autres protéines membranaires est nécessaire pour affirmer avec assurance l'existence de protéines supplémentaires chez le cenancêtre, il semble probable que des protéines très répandues dans les organismes actuels, tels les membres de la famille des transporteurs ABC ou des perméases, étaient déjà présentes chez le cenancêtre [Delaye *et al.* 2005].

La forte intégration des protéines membranaires dont la présence peut être inférée chez le cenancêtre (ATPases, chaînes respiratoires, intégrases et translocases) enrichit d'ailleurs l'argument opposé aux compartiments minéraux dans la mesure où tous ces composants ont dû être physiquement liés entre eux pour qu'une telle co-évolution ait été possible. Par conséquent, l'ensemble de ces résultats ne doit plus nous conduire à nous demander s'il existait des membranes chez le cenancêtre mais à quoi elles ressemblaient. Nous devons garder aussi à l'esprit qu'il est difficile de tirer des conclusions sur les propriétés des membranes anciennes à partir d'informations qui concernent uniquement une partie de leurs composants. Les efforts futurs devraient essayer de décrire un maximum d'éléments ayant pu être présents chez le cenancêtre pour permettre de déterminer non seulement les fonctions que cet organisme pouvait probablement réaliser mais aussi comment ces fonctions pouvaient communiquer entre elles à cette époque très ancienne.

6. Quelle composition possible pour des membranes lipidiques chez le cenancêtre ?

Nous avons vu que face aux deux types très différents de phospholipides qui existent d'une part chez les archées et d'autre part chez les bactéries et les eucaryotes, certains auteurs ont postulé que le cenancêtre pouvait ne pas avoir contenu de phospholipides du tout et ne pas

avoir été délimité par des membranes lipidiques. Cependant, ces hypothèses ne permettent pas d'expliquer l'apparition de structures très compliquées qui sont universelles et généralement déduites chez le cenancêtre comme la machinerie de transcription, les ribosomes ou les ATPases. Nous avons vu également que certaines études sur les origines de la vie estiment que les membranes lipidiques ont pu jouer un rôle important dès les premières étapes de la vie, suggérant par la même occasion une continuité entre les premières formes de vie et les formes de vie actuelle en passant par le cenancêtre. Mais si des membranes ont existé chez le cenancêtre, à quoi pouvaient-elles ressembler ?

Des membranes de type bactérien chez le cenancêtre

Certains modèles ont postulé que les membranes du cenancêtre pouvaient être composées d'un des deux types modernes de phospholipide, qui aurait ensuite été remplacé par l'autre type de phospholipide dans la lignée correspondante [Lake *et al.* 2008]. L'exemple le plus connu est celui de l'hypothèse du cenancêtre négibactérien qui érige la membrane en un caractère évolutif majeur dont les changements d'état permettraient la reconstruction de l'évolution ancienne (Figure 11, p. 41, [Cavalier-Smith 2002]). Dans cette hypothèse l'arbre du vivant est raciné au sein des bactéries (exactement entre les eobactéries, qui sont un groupe défini par Cavalier-Smith contenant les Chloroflexi et les Deinococcus-Thermus, et le reste des bactéries), de sorte que le cenancêtre aurait été entouré d'une double membrane bactérienne de type Gram négatif.

Au cours du temps, un groupe de bactéries (appelées posibactéries par Cavalier-Smith et qui recouperaient les bactéries Gram positives) aurait perdu leur membrane externe, remplacée par des parois plus épaisses. Les archées et les eucaryotes formeraient un groupe monophylétique, les neomura (« nouvelles parois »), dont les premiers ancêtres seraient apparus au sein des actinobactéries. Les neomura partageraient avec les bactéries Gram positives la caractéristique d'être entourés d'une seule membrane mais se distingueraient de ces mêmes organismes par la présence de glycoprotéines dans les enveloppes cellulaires. Au sein des neomura, les eucaryotes auraient conservé leur membrane unique dont la plasticité aurait permis la phagocytose et de nombreuses fonctions associées à des membranes internes, alors que les archées auraient remplacé leur membrane unique originelle par celle avec leurs phospholipides caractéristiques. Le moteur de ce changement vers des phospholipides de type archéen aurait été l'adaptation aux conditions hyperthermophiles et acides dans lesquelles l'ancêtre des archées est supposé avoir vécu.

L'originalité principale de ce travail réside probablement dans le fait qu'il ne considère pas les archées, les eucaryotes et les bactéries comme les trois domaines du vivant, mais considère au contraire que les deux premiers groupes branchent en réalité dans le troisième. Cette hypothèse fait des bactéries un groupe paraphylétique car elle considère les archées et les eucaryotes comme des bactéries dérivées. Selon l'auteur, la raison de ce parti pris est sa préférence pour l'*analyse des transitions*³⁰ et les données fossiles au détriment des analyses moléculaires qui ont été surestimées à ses yeux [Cavalier-Smith 2002]. Pourtant, Cavalier-Smith n'hésite pas à utiliser les phylogénies moléculaires de deux gènes paralogues intervenant dans la synthèse de la chlorophylle pour placer la racine de l'arbre du vivant entre les eobactéries et le reste des organismes, alors que les deux gènes en question ne sont pas universels et que la plupart des auteurs pense que la photosynthèse n'est probablement pas une caractéristique présente chez le dernier ancêtre commun [Raymond *et al.* 2002]. Quant à l'utilisation des fossiles, cette tâche s'avère impossible dans la pratique lorsqu'on travaille en grande partie sur des procaryotes ayant vécu il y a plusieurs milliards d'années et pour lesquels l'interprétation des quelques rares traces reste polémique et largement insuffisante pour déterminer le détail de leur biologie.

Par ailleurs, bien que Cavalier-Smith ait raison de rappeler l'importance de la séquence historique pour comprendre l'évolution des organismes, ce qui vaut aussi pour les analyses de données moléculaires, sa méthode *a priori* souffre de la nature arbitraire des caractères qu'il choisit pour construire l'histoire de l'évolution des organismes. Il fait tout reposer sur l'idée que les doubles membranes sont plus anciennes que les autres [Cavalier-Smith 2002]. Or ce caractère morphologique apparaît difficilement discriminant lorsque l'on connaît la diversité membranaire du vivant. Associer ensemble sur la base de leur nombre les membranes archéennes et Gram positives alors que leur composition et même le reste de leur enveloppe sont aussi différents n'est pas un choix trivial : on connaît des archées comportant une double membrane (*Ignicoccus hospitalis* [Junglas *et al.* 2008]), ce qui, même dans le cas d'une convergence avec les bactéries Gram négatives, montre la faiblesse de ce caractère morphologique comme principal critère directeur de classification des organismes.

³⁰ L'analyse de transitions est définie par Cavalier-Smith comme la construction conceptuelle d'une séquence rationnelle d'étapes permettant de transformer un ancêtre en un descendant en tenant compte des mécanismes évolutifs et des avantages évolutifs au cours du temps.

Le travail de Cavalier-Smith accepte que les doubles membranes aient été plus anciennes que les membranes uniques pour des raisons qui sont liées à la bioénergétique qui a précédé le cenancêtre, qui auraient privilégié une membrane double chez les premiers organismes. Pourtant, comme nous l'avons vu auparavant, les processus biologiques à l'origine de la vie restent hautement spéculatifs et l'on ne peut pas exclure, étant donnée la diversité des membranes, qu'une membrane unique puisse accueillir une activité bioénergétique comme c'est le cas dans les organismes Gram positifs modernes. D'ailleurs, la caractéristique inverse (une seule membrane) a été postulée aussi comme une caractéristique ancestrale, précisément parce qu'elle rend plus facile le transport de substances [Lake *et al.* 2009]. Cette hypothèse soulève donc à nouveau la principale critique que l'on peut faire à ce type d'hypothèse, et que j'ai longuement développée dans la première partie de l'introduction : l'arbitraire du choix du caractère utilisé.

Une autre critique que l'on peut adresser à toutes les hypothèses proclamant un changement tardif de type membranaire est qu'un tel événement n'a aucun précédent et semble improbable étant donné le grand nombre d'acteurs qui interviennent dans la synthèse de phospholipides et l'absence de moteur sélectif pour les étapes intermédiaires. Il a d'ailleurs été proposé que les membranes composées d'un mélange des deux types de phospholipides (hétérochirales) soient moins stables que celles composées d'un seul type de phospholipide (homochirales). Si cela s'avérait juste, les étapes de transition entre deux membranes homochirales seraient mises en cause [Wächtershäuser 2003] mais, comme nous allons le voir, la supposée instabilité des membranes hétérochirales reste encore à démontrer.

Des membranes hétérochirales chez le cenancêtre

En partant du principe hypothétique qu'une membrane hétérochirale serait plus instable qu'une membrane homochirale, il a été proposé que la transition tardive d'un type de membrane à l'autre soit impossible car les étapes intermédiaires n'auraient pas été favorablement sélectionnées. Cependant, en utilisant le même raisonnement, il a été proposé que le cenancêtre ait pu être entouré d'une membrane mixte et que l'homochiralité des membranes modernes ait pu être causée par leur plus grande stabilité [Wächtershäuser 2003]. Ce scénario se place dans le contexte d'un cenancêtre qui aurait été composé d'une communauté de pré-cellules [Kandler 1995]. Rappelons que les pré-cellules auraient été soumises à de fréquentes fusions et fissions, de sorte que le matériel génétique aurait été

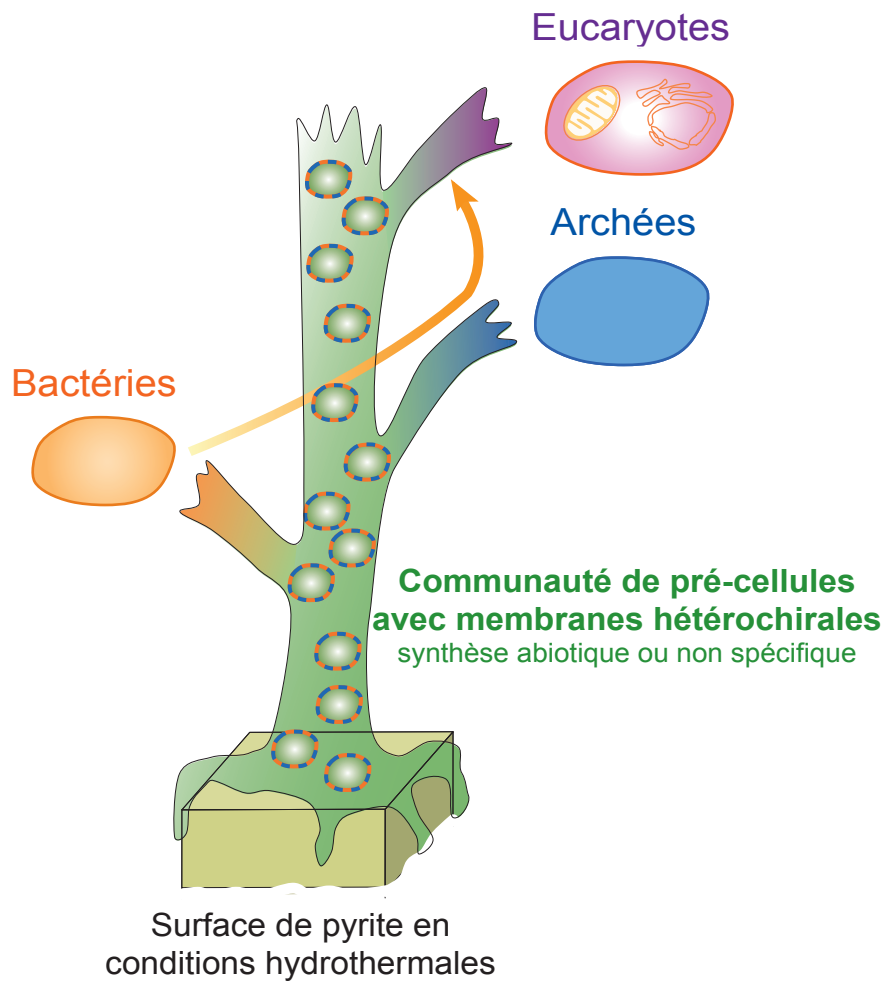


Figure 38. Hypothèse de Wächtershäuser 2003 selon laquelle le cenancêtre était une communauté de pré-cellules entourées de membranes hétérochirales.

Apparition de la vie sur une surface minérale (comme exposé dans Wächtershäuser 1988). Les couleurs des membranes indiquent le type de phospholipides : le type bactérien est en orange, l'archéen en bleu, les pré-cellules ont un mélange des deux et les eucaryotes ont des membranes de type bactérien. Dans ce modèle, les fusions et fissions entre pré-cellules sont supposées être très fréquentes.

fréquemment mélangé. Les pré-cellules sont donc des structures délimitées par une membrane dont Wächtershäuser va proposer qu'elle soit hétérochirale (Figure 38).

Les phospholipides contenant respectivement un énantiomère de glycérol phosphate ou l'autre (G1P ou G3P), ils auraient des propriétés physico-chimiques différentes qui auraient tendance à les séparer et donc à former des membranes homochirales supposées plus stables [Wächtershäuser 2003]. Parallèlement, selon Wächtershäuser, la ségrégation aurait été facilitée par les fréquentes fusions et fissions des pré-cellules, favorisant l'obtention de pré-cellules avec des phospholipides de type archéen et d'autres pré-cellules avec des phospholipides de type bactérien. Ces différences auraient été fixées dans les lignées modernes lorsque des déhydrogénases responsables de la synthèse des énantiomères de glycérol phosphate auraient été recrutées séparément (Figure 38). Plus récemment, il a été proposé que la chaîne latérale ait joué un rôle supplémentaire dans la séparation membranaire car, alors que les phospholipides composés de G1P/G3P et acides gras sont des énantiomères, ceux composés de G1P/G3P et isoprénoïdes comportent deux centres chiraux et sont donc des diastéréoisomères avec des propriétés différentes [Koga 2010].

Bien que très intéressante à cause de son pouvoir explicatif, cette hypothèse est en réalité assez confuse car elle ne propose pas un seul mais trois mécanismes concurrents pour expliquer la ségrégation différentielle des phospholipides : d'abord les phospholipides auraient ségrégré spontanément dans les membranes pré-cellulaires au cours des divisions pré-cellulaires. Par ailleurs, les membranes se seraient séparées aussi à cause de fusions et de fissions fréquentes (ce qui d'ailleurs semble étrange car les fusions auraient dû, au contraire provoquer à nouveau le brassage des lipides ; on suppose peut-être dans ce cas que les fusions entre membranes de type opposé sont plus rares que celles avec des membranes identiques, mais cela reste encore à prouver). Finalement la synthèse enzymatique des glycérol phosphate aurait fixé les deux types de phospholipides dans chaque lignée, ce qui sous-entend qu'en réalité les critères physico-chimiques n'étaient pas suffisants et qu'ils ont été doublés de cet autre mécanisme qui a définitivement empêché tout retour en arrière. Ce scénario ne peut donc pas être compris comme une hypothèse unifiée mais plutôt comme une proposition de l'importance possible de la chiralité des membranes pour l'explication des différences actuelles.

A première vue, l'idée que les membranes hétérochirales seraient moins stables semble plausible car beaucoup de composés avec des propriétés physico-chimiques différentes ont

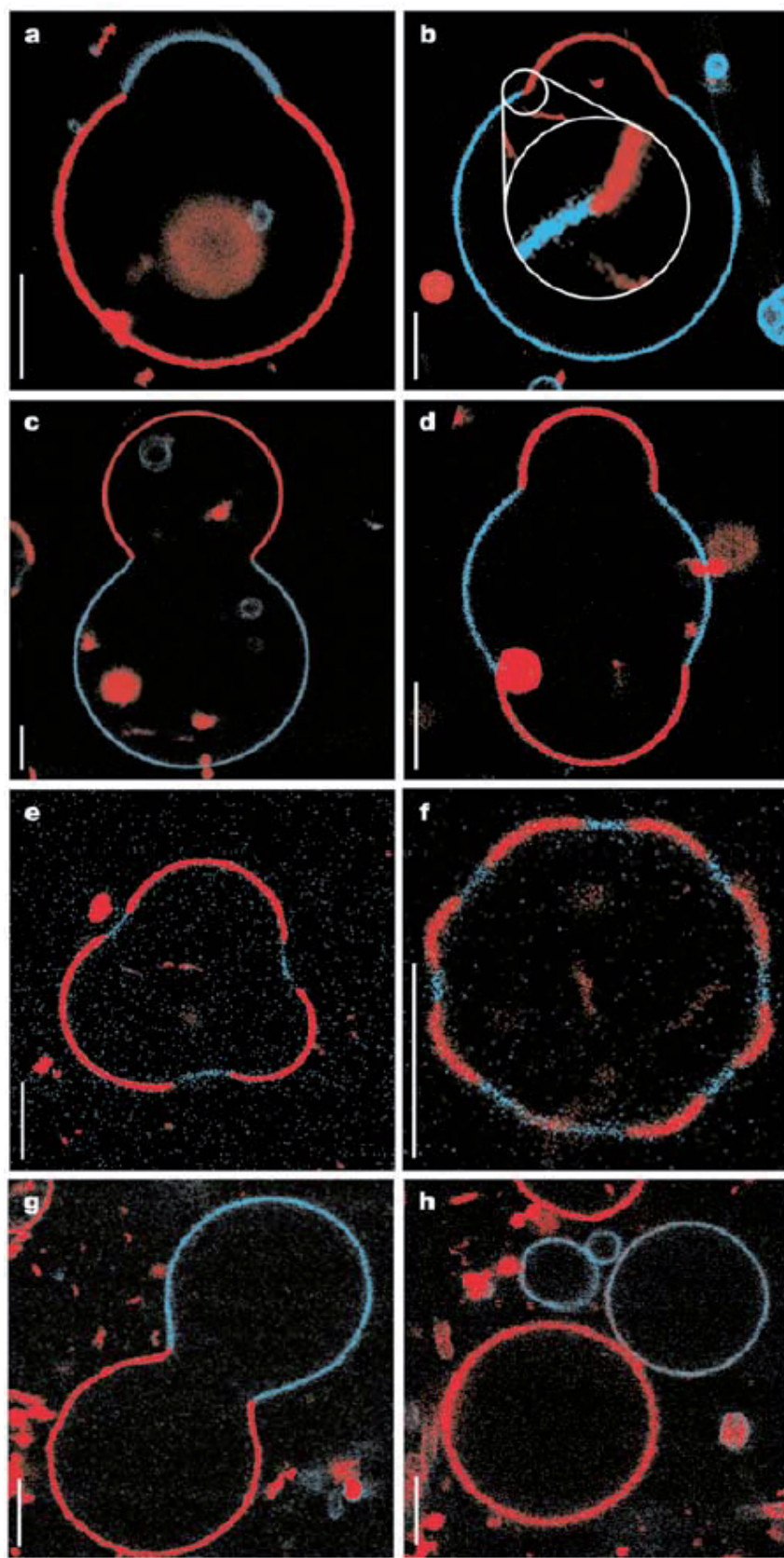


Figure 39. Comportement des membranes liposomiques mélangées. D'après Baumgart et al. 2003.

Les membranes sont composées de plusieurs types de sphingomyéline, cholestérol et dioleylephosphatidylcholine, marqués avec des substances fluorescentes. La séparation spontanée des composées est très rapide mais elle dépend fortement des lipides utilisés.

tendance à ségréger spontanément [Vollhardt *et al.* 1996 ; Uragami *et al.* 2000]. Ceci avait par exemple été observé pour des mélanges racémiques de D- et L- myristoylalanine [Nassoy *et al.* 1995] et pour des mélanges de phospholipides archéens et bactériens [Longo *et al.* 2007]. Le travail sur des liposomes monocouches composés de plusieurs types de lipides (sphingomyéline, cholestérol et dioleylphosphatidylcholine, DOPC) marqués avec des substances fluorescentes permet non seulement de voir la ségrégation spontanée des domaines membranaires mais aussi d'observer que ces domaines adoptent des courbures différentes en fonction de leur composition (Figure 39). Lorsque de faibles agitations du milieu provoquent la cassure des vésicules, les nouveaux liposomes résultants n'auront donc plus la composition des vésicules mères mais un enrichissement en un des types de phospholipide [Baumgart *et al.* 2003]. On peut opposer pourtant à ces expériences montrant la ségrégation spontanée de molécules des expériences signalant que, lorsque des phospholipides de type archéen sont mélangés à ceux des eucaryotes (similaires aux bactériens), on observe une stabilité égale ou même parfois plus grande [Elferink *et al.* 1992 ; Fan *et al.* 1995]. Afin de surmonter cette contradiction, des travaux plus récents ont testé des vésicules composées d'une plus grande diversité de lipides et observé que la stabilité des liposomes³¹ est indépendante de la chiralité des phospholipides et dépend surtout de la longueur des chaînes latérales [Shimada et Yamagishi 2011]. Ces résultats mettent directement en cause le fondement même des hypothèses de stabilité proposées par Wächtershäuser pour expliquer la séparation des deux types de membranes et témoignent du manque d'analyses comparatives fiables. En effet, beaucoup d'analyses ont été faites avec des phospholipides de différentes tailles [Longo *et al.* 2007], des petites molécules [Nassoy *et al.* 1995] ou des molécules trop différentes entre elles [Baumgart *et al.* 2003] pour être représentatives du comportement de membranes hétérochirales plus homogènes.

En dehors de cette remise en cause fondamentale, d'autres critiques ont été formulées à l'encontre de l'hypothèse de Wächtershäuser. Il a par exemple été dit que si le cenancêtre avait possédé des membranes hétérochirales, cela aurait fait de lui un organisme bien plus polyvalent que les organismes actuels, ce qui semblait difficile à croire [Koonin et Martin 2005]. Dans la conclusion, je développerai l'idée que cette interprétation est seulement une

³¹ La mesure de la stabilité est ici considérée comme la inversement proportionnelle à la quantité de fuites enregistrées pour une molécule fluorescente encapsulée dans les vésicules [Shimada et Yamagishi 2011].

interprétation possible parmi d'autres et qu'elle repose sur un préjugé sur le cenancêtre plutôt que sur des observations réelles.

Plus importantes, des expériences comme celle de Baumgart *et al.* montrent que si les conditions sont réunies pour la séparation spontanée des molécules, alors ce processus devrait être très rapide et s'accomplir en très peu de divisions. Cette hypothèse impose donc un nombre de générations absurde petit entre l'apparition de la vie (que l'on imagine hétérochirale dans ce contexte) et la divergence des bactéries et des archées. Les fusions évoquées par Wächtershäuser pourraient expliquer que les mélanges aient été maintenus pendant plus longtemps mais cela contredit l'idée de cet auteur selon laquelle les fusions et les fissions étaient le moteur de la ségrégation plutôt que du brassage membranaire. Ce problème met en évidence une autre lacune de cette hypothèse : elle reste ouverte sur l'origine des phospholipides présents dans les membranes des pré-cellules.

L'auteur propose que ces phospholipides aient pu être catalysés abiotiquement grâce à l'activité de certains métaux de transition, processus qui serait très proche de son modèle d'apparition de la vie sur des surfaces minérales [Wächtershäuser 1988] mais cette hypothèse rapprocherait sérieusement le cenancêtre des premières étapes de la vie. Plus tard il a clairement postulé la possibilité d'une synthèse enzymatique non-stéréospécifique [Wächtershäuser 2006], ce qui peut sembler étrange dans la mesure où les enzymes sont en général considérées comme des structures hautement stéréospécifiques et où la diversité des substrats utilisés par les enzymes généralistes ne se fait en général pas sur des énantiomères différents mais avec des molécules diverses.

Malgré toutes ces critiques, le plus grand mérite de cette hypothèse est d'avoir posé la question de l'existence d'enzymes capables de synthétiser des phospholipides de façon non spécifique (mais stéréospécifique) chez le cenancêtre. C'est sur la recherche de ces enzymes que les travaux postérieurs, y compris les miens, se sont concentrés.

Synthèse enzymatique de glycérol phosphate chez le cenancêtre

La stabilité plus faible des membranes hétérochirales par rapport aux membranes homochirales proposée par Wächtershäuser comme un déclencheur de la transition des premières dans le cenancêtre vers les secondes dans les lignées modernes est mise en cause par l'observation que les membranes mixtes composées de phospholipides comparables ne sont pas moins stables que les homochirales [Shimada et Yamagishi 2011]. Cependant, cette

hypothèse pose surtout la question de l'existence d'un mécanisme enzymatique de synthèse de phospholipides chez le cenancêtre.

Cette question a d'abord été posée à propos de la synthèse de G1P et G3P. Bien que chacune de ces deux molécules constitue l'ossature d'un phospholipide caractéristique d'un domaine procaryote et absent dans l'autre, ces molécules ne sont pas spécifiques d'un seul domaine. Par exemple, le G3P est en réalité un métabolite très répandu dans les trois domaines du vivant et chez certaines archées hétérotrophes il peut être catabolisé comme substrat [Nishihara *et al.* 1999]. Des activités enzymatiques capables de synthétiser du G3P ont été signalées [Nishihara *et al.* 1999] et des gènes codant des enzymes capables de synthétiser le G3P à partir de déhydroxyacétone phosphate ou de glycérol sans être des homologues de la G3PDH ont été détectés dans les génomes d'archées [Peretó *et al.* 2004]. Le G1P synthétisé par une G1PDH de type archéen chez certaines bactéries a aussi été décrit [Guldan *et al.* 2008] mais la phylogénie de ce gène signale qu'il a pu être récemment reçu par HGT par certaines bactéries [Peretó *et al.* 2004]. Bien que marginal, le G1P est pourtant utilisé chez les bactéries à la fois comme composant des acides lipotéichoïques qui forment les parois de peptidoglycane des bactéries Gram positives [Taron *et al.* 1983 ; Rahman *et al.* 2009] et comme tête polaire de certains phospholipides, qui interviennent d'ailleurs parfois dans la synthèse des acides lipotéichoïques [Taron *et al.* 1983 ; Itabashi et Kuksis 1997] et des lipopolysaccharides des bactéries Gram négatives [Perepelov *et al.* 2006]. Le G1P fait aussi parti du BMP, ou acide lysobisphosphatidique³² chez les animaux [Tan *et al.* 2012]. Les deux énantiomères de glycérol phosphate peuvent donc être retrouvés dans des organismes des trois domaines du vivant et sont le résultat de divers mécanismes réactionnels différents mais cela ne suffit pas à permettre de considérer que la molécule ait été présente chez le cenancêtre, en particulier dans les phospholipides.

En ce qui concerne la synthèse des phospholipides, ce sont les réactions des G1PDH et G3PDH que l'on prend en compte prioritairement car ce sont elles qui sont spécialisées dans cette fonction aujourd'hui. Comme nous l'avons vu auparavant, la comparaison des séquences primaires et des structures tertiaires des deux dehydrogenases montre qu'elles ne sont pas homologues entre elles [Koga *et al.* 1998 ; Daiyasu *et al.* 2002]. Cependant, chacune des deux

³² Le BMP est une molécule composée de deux phospholipides attachés par leur groupement phosphate. Il existe chez certaines bactéries et chez les eucaryotes, mais jusque-là, l'utilisation de G1P dans les BMPs n'a été démontrée que chez les mammifères [Tan *et al.* 2012].

appartient à une grande superfamille d'enzymes qui, elles, sont présentes chez les trois domaines du vivant.

La G1PDH appartient à une superfamille contenant l'alcool déshydrogénase (ADH), la 3-déshydroquinase synthase (DHQS) et la glycérol déshydrogénase (GDH). Elles utilisent le NAD^+ comme cofacteur, sont structurellement très similaires et catalysent des réactions similaires [Carpenter *et al.* 1998 ; Ruzheinikov *et al.* 2001]. La phylogénie de cette superfamille montre d'abord que les séquences regroupent en relation directe avec leur fonction. Quelques HGTs sont observables mais n'empêchent pas d'inférer la monophylie de chaque groupe : les ADH et DHQS sont caractéristiques des bactéries et des eucaryotes, les GDH des bactéries et les G1PDH des archées. Etant donné qu'aucune des fonctions n'est universelle on ne peut inférer la présence d'aucune d'entre elles chez le cenencêtre mais la présence dans cet organisme d'au moins un membre de cette superfamille semble très probable. Il a été proposé que chez le cenencêtre ce gène ait codé une déshydrogénase utilisant le NAD^+ dont l'activité ne serait pas spécifique, i. e. qu'elle aurait pu catalyser sa réaction sur plusieurs substrats avant d'être dupliquée et recrutée pour des activités spécialisées dans les différentes lignées. L'enzyme codée par ce gène aurait notamment pu synthétiser du G1P chez le cenencêtre et constituer ainsi un mécanisme enzymatique de synthèse dans cet organisme [Peretó *et al.* 2004].

L'analyse des phylogénies de la superfamille à laquelle appartient la G3PDH donne des résultats très similaires. Cette superfamille, qui en plus de la G3PDH concerne l'UDP-glucose 6-déshydrogénase (UDPGDH) et la 3-hydroxyacyl-CoA déshydrogénase (HACDH), est caractérisée par un mécanisme d'oxydoréduction basé sur l'utilisation d'un cofacteur NAD^+ [Campbell *et al.* 2000 ; Kavanagh *et al.* 2003]. La phylogénie de cette superfamille montre aussi trois grands groupes de séquences correspondant aux trois fonctions et, malgré quelques HGT, la détermination de l'origine de chaque gène n'est pas compromise. L'UDPGDH est universelle. Elle est responsable de la synthèse de différents polymères chez les bactéries (polysaccharides destinés à la sécrétion et participant à la synthèse des parois cellulaires) et les eucaryotes (protéoglycanes, hémicellulose ou pectine) et elle est probablement liée aussi à l'utilisation d'acide D-glucuronique dans certaines parois d'archées [Dalessandro et Northcote 1977 ; Arrecubieta *et al.* 1994 ; Kandler 1995 ; Binari *et al.* 1997]. La HACDH est également universelle et elle fait partie de plusieurs polypeptides catalysant des fonctions très diverses [Youngleson *et al.* 1989 ; Hiltunen et Qin 2000]. La G3PDH est limitée aux bactéries et aux eucaryotes et l'éparpillement des séquences d'archées dans ce groupe soutient fortement

l'idée qu'il s'agisse ici de xénologues dans ces organismes. Cependant, étant donné que la superfamille contient deux sous-familles universelles qui sont capables d'utiliser des substrats très divers, il est possible d'inférer la présence d'au moins deux de ces gènes chez le cenancêtre et de proposer qu'ils aient été capables de réaliser différentes fonctions, dont la synthèse de G3P, avant de se spécialiser dans les lignées postérieures [Peretó *et al.* 2004].

En conclusion, bien que la G1PDH et la G3PDH appartiennent à des superfamilles différentes, on a pu déduire la présence chez le cenancêtre de représentants à fonction non-spécifique dans les deux cas et proposer ainsi que ces molécules aient existé chez le cenancêtre. Le G1P et le G3P auraient alors pu entrer dans la composition des phospholipides du cenancêtre. En proposant que cette observation puisse s'étendre aux autres composants phospholipidiques, il a été postulé que malgré la distribution disparate des mécanismes de synthèse de phospholipides dans les trois domaines du vivant, ils auraient pu être présents chez le cenancêtre sous des formes moins spécifiques que les voies de synthèse actuelles. Cette hypothèse conserve le principe de continuité entre l'origine de la vie et les organismes modernes et propose en plus un mécanisme plausible de synthèse des phospholipides. Ces observations ne permettent pas de déterminer précisément quel type de membrane existait chez le cenancêtre mais il a été proposé qu'il s'agisse de membranes hétérochirales dont la transformation en membranes homochirales modernes aurait résulté du recrutement postérieur de machineries spécifiques dans les différentes lignées [Peretó *et al.* 2004].

Synthèse enzymatique des autres composants phospholipidiques chez le cenancêtre

Les travaux sur la présence chez le cenancêtre de mécanismes enzymatiques de synthèse de G1P et G3P par des méthodes de phylogénomique ont montré que l'étude phylogénétique des protéines participant au métabolisme moderne pouvait nous éclairer sur la nature des membranes du cenancêtre [Peretó *et al.* 2004]. Cependant, une étude détaillée de l'évolution des mécanismes de synthèse des composantes phospholipidiques autres que les énantiomères de glycérol phosphate était nécessaire pour compléter les hypothèses formulées.

J'ai donc focalisé mon travail sur l'évolution des principales voies de synthèse des phospholipides modernes : les voies de synthèse des isoprénoïdes, des acides gras ainsi que les principales enzymes qui construisent les phospholipides à partir de ces composantes. Mes résultats ont permis non seulement de proposer des mécanismes enzymatiques de synthèse de ces composants chez le cenancêtre mais également de révéler l'importance de certains mécanismes qui avaient jusque-là été largement ignorés dans beaucoup d'organismes

OBJECTIFS

Analyse phylogénomique des mécanismes actuels de synthèse des phospholipides

Dans l'introduction, j'ai détaillé l'importance qu'une description du cenancêtre aurait sur notre connaissance de l'évolution des organismes des trois domaines du vivant, spécialement dans les étapes les plus anciennes. Ensuite, j'ai expliqué que l'une des caractéristiques les plus polémiques concernant la nature du cenancêtre est la présence ou non de membranes phospholipidiques autour de cet organisme. Bien que tout semble indiquer que le cenancêtre possédait probablement une membrane, les seules fonctions de la synthèse des phospholipides dont la présence avaient été inférée dans cet organisme avant mon travail étaient celles des deux énantiomères du glycérol phosphate.

L'objectif principal de mon travail a donc été d'étudier l'évolution des mécanismes de synthèse des autres composants des phospholipides et de leur assemblage. Plus exactement, mon travail s'est construit autour trois axes :

1. Etude évolutive des voies de synthèse des isoprénoïdes dans les trois domaines du vivant et implications pour le cenancêtre.
2. Analyse des homologues archéens de la voie bactérienne de synthèse des acides gras et relations phylogénétiques entre les séquences des deux domaines procaryotes du vivant.
3. Analyse évolutive des mécanismes d'assemblage des phospholipides à partir de leurs composants dans les trois domaines du vivant et implications pour le cenancêtre.

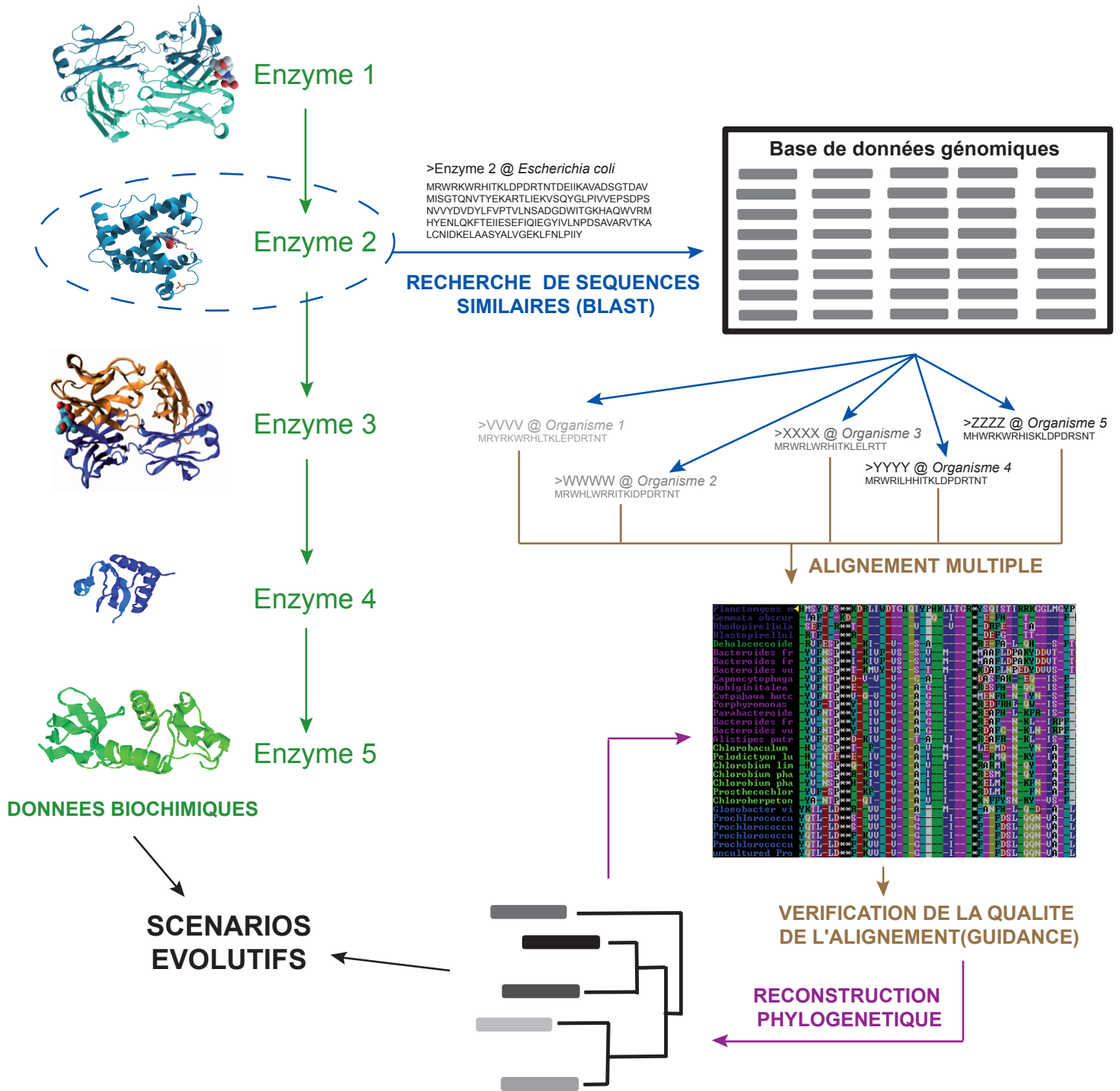


Figure 40. Résumé des principales étapes méthodologiques de mon analyse.

L'analyse se fait d'abord individuellement pour chaque enzyme. Une séquence de départ par enzyme et domaine du vivant est utilisée pour une recherche de similarité par BLAST dans une liste de génomes complets (Annexe 1). Les séquences obtenues sont alignées et la qualité des alignements peut être vérifiée avant la reconstruction phylogénétique (par exemple, avec le logiciel Guidance). L'analyse des alignements et des phylogénies intègre les éventuelles informations biochimiques disponibles. La reconstruction phylogénétique permet la sélection de séquences représentatives qui permettront de réaliser des nouvelles reconstructions avec des méthodes plus fiables mais plus gourmandes en temps de calcul. Les phylogénies finales sont interprétées enzyme par enzyme, puis les informations pour l'ensemble de la voie de synthèse sont intégrées dans une hypothèse évolutive commune.

MATÉRIEL ET MÉTHODES

Dans toutes les analyses que je présenterai dans mes résultats, le cadre général des méthodes utilisées reste le même, bien que des nuances et des modulations aient été nécessaires afin de tenir compte des particularités phylogénomiques des protéines étudiées. Notamment, la distribution taxonomique des protéines est très irrégulière, spécialement à l'échelle des différents domaines du vivant, et la multiplicité des duplications et HGT au cours de l'évolution conduit à la présence de nombreux gènes homologues (paralogues) dans les génomes utilisés. Je vais d'abord présenter un aperçu du sens général de mon analyse, puis je présenterai les programmes qui ont été employés. Lorsque la stratégie générale a dû être adaptée pour tenir compte de la diversité évolutive des protéines étudiées, les affinements seront expliqués directement dans la présentation des résultats protéine par protéine, afin de faciliter à la fois la compréhension des techniques et l'interprétation des résultats.

1. Vue d'ensemble de la stratégie utilisée

Mon travail a été le fruit d'un aller-retour constant entre les données disponibles dans la littérature et les analyses phylogénomiques. Les voies de biosynthèse dont j'ai étudié l'évolution synthétisent des composants indispensables au fonctionnement d'un grand nombre d'organismes. De fait, elles attirent l'attention de la recherche depuis longtemps et leur découverte reste parmi les grands événements marquants de la biologie moderne : les descriptions de la voie du mévalonate, qui permet la synthèse des isoprénoïdes, et de la voie de synthèse des acides gras ont valu le Prix Nobel de Physiologie et Médecine à Konrad Bloch et Feodor Lynen en 1964 [Bloch 1964 ; Lynen 1964]. Depuis, ces voies sont restées au centre de nombreuses recherches fondamentales mais aussi cliniques, ce qui permet d'avoir accès à beaucoup d'informations pour un grand nombre d'enzymes dans un large spectre d'organismes.

Mon analyse a d'abord concerné chaque protéine d'intérêt (Figure 40). En partant d'une séquence dont la fonction avait été caractérisée expérimentalement et qui ne faisait pas de doute, des recherches de similarité ont été faites contre les bases de données génomiques. Cela permet de détecter des séquences similaires que l'on considère statistiquement comme des séquences homologues de la séquence de départ (voir Introduction, partie 1, chapitre 1, section « A la rencontre du cenancêtre »). Les informations contenues dans ces banques ont

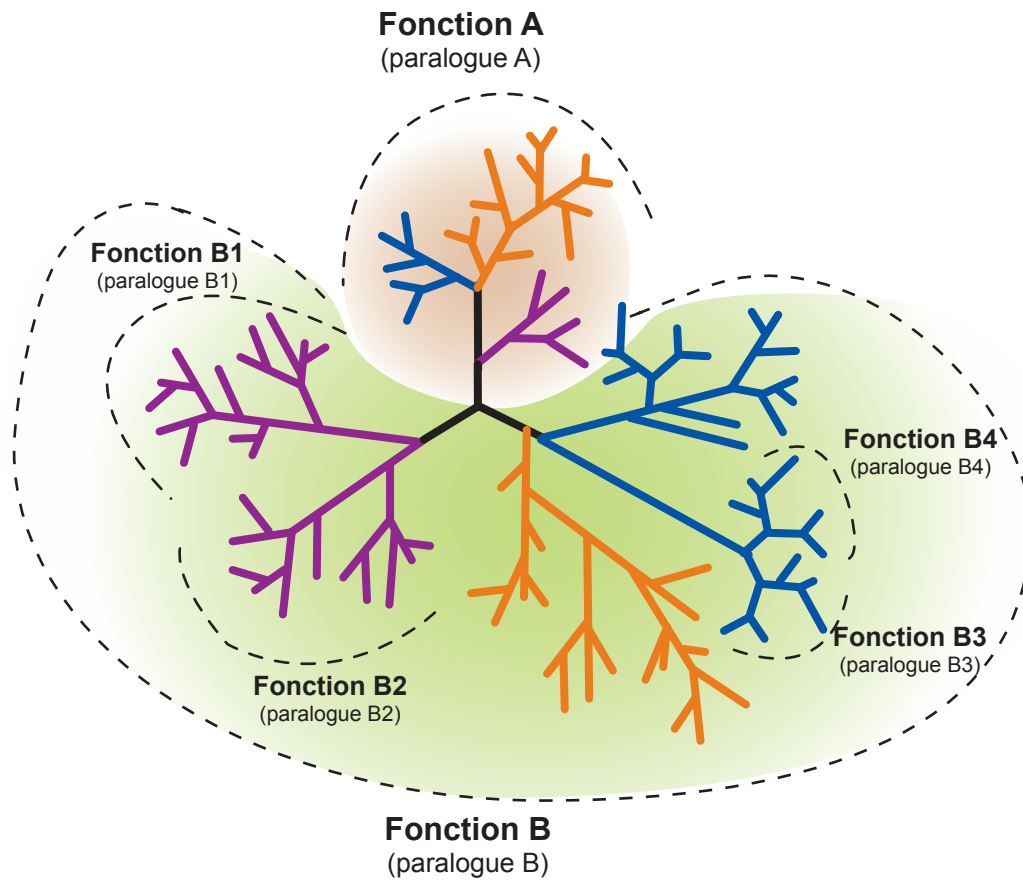


Figure 41. Séparation des paralogues dans les arbres phylogénétiques.

Les gènes issus de duplications divergent d'un point de vue phylogénétique mais généralement aussi d'un point de vue fonctionnel. De fait, dans les arbres phylogénétiques les séquences groupent d'accord non seulement aux groupes taxonomiques mais aussi à la divergence des paralogues/fonctions. Dans cette figure théorique les couleurs des branches représentent les organismes des trois domaines du vivant. Les accolades signalent les fonctions biochimiques indiquées par l'information disponible dans la littérature.

En l'occurrence, on voit deux fonctions générales associées à deux paralogues universels, A et B. Le gène B a subi une duplication avant la divergence de la lignée violette, donnant lieu à deux sous-fonctions dans ces organismes, B1 et B2. La même chose semble avoir eu lieu dans un ancêtre d'une partie des organismes de la lignée bleue, donnant lieu à d'autres sous-fonctions, B3 et B4.

augmenté exponentiellement au cours des dernières années grâce à l'accumulation de génomes complets mais aussi de séquençages partiels (gènes isolés, séquençages massifs de transcriptomes ou de métagénomes). Or, étant donné que mon travail s'intéresse à l'évolution à l'échelle des trois domaines du vivant, le nombre de ces séquences pourrait devenir excessif pour permettre l'application de certaines méthodes trop gourmandes en temps de calcul, comme le maximum de vraisemblance. De ce fait, j'ai généralement travaillé uniquement avec une sélection de génomes complets représentative de la diversité des trois domaines, avec éventuellement l'élargissement à un plus large spectre lorsque le nombre d'homologues détectés était plus faible.

Une fois obtenues, les séquences homologues sont utilisées pour reconstruire un alignement multiple. Cet alignement permet de détecter d'éventuelles séquences non homologues, particulièrement divergentes ou comportant des caractéristiques remarquables, comme des fusions avec d'autres gènes ou des insertions-délétions (indels). Les séquences non homologues doivent être retirées de l'analyse ; les séquences très divergentes peuvent introduire des biais (plus de bruit dans l'alignement, phénomène d'attraction des longues branches), donc en général elles ont été éliminées, sauf si elles comportaient un intérêt particulier de par leur fonction ou distribution taxonomique ; les indels doivent être pris en considération lors de l'interprétation des résultats. Les alignements permettent aussi de déterminer les positions homologues suffisamment conservées pour être utilisables pour la reconstruction de phylogénies.

Pour la plupart des enzymes étudiées, la reconstruction d'une phylogénie préliminaire a été nécessaire pour établir les grandes tendances observables qui guideront la suite de la recherche. D'une part, les enzymes sont assez souvent soumises à des duplications qui génèrent des paralogues, et lorsque les duplications sont anciennes, en général leurs fonctions sont assez différentes. Cette disparité entre paralogues se traduit souvent dans les arbres sous la forme des regroupements associés à la fonction des gènes (Figure 41). Ainsi, il est indispensable de reporter sur la phylogénie l'information fonctionnelle issue de la littérature afin soit de se concentrer uniquement sur les paralogues d'intérêt, soit d'essayer de comprendre l'évolution des fonctions au cours du temps. L'annotation fonctionnelle a été faite à la main en signalant sur l'arbre les séquences caractérisées et leurs fonctions. Ensuite, cette annotation peut être étendue lorsque des synapomorphies permettent de distinguer des groupes fonctionnels, par exemple dans le cas de la conservation de certains acides aminés dans une position donnée, notamment dans le site catalytique qui détermine l'activité

enzymatique. Cette méthode manuelle demande beaucoup de temps et doit être refaite à chaque nouvelle phylogénie mais est nécessaire pour éviter de perdre les informations fonctionnelles qui permettent une mesure qualitative supplémentaire de la fiabilité des phylogénies étudiées.

La phylogénie préliminaire de chaque gène permet aussi de déterminer la position de certains grands groupes taxonomiques par rapport à d'autres, spécialement les séquences eucaryotes. Ainsi, lorsque les protéines eucaryotes se placent parmi les bactéries, notamment à proximité des cyanobactéries dont on connaît le rapport avec les plastes ou des alpha-protéobactéries dont on connaît le rapport avec les mitochondries, l'évolution de ces gènes chez les eucaryotes sera peu informatif pour comprendre la nature du cenancêtre. Cette étude portera moins sur ces séquences et elles seront même parfois éliminées pour réduire la taille des jeux de données et éliminer des éventuels biais dus à la divergence propre à ces séquences eucaryotes d'origine bactérienne.

Les phylogénies préliminaires sont très importantes dans tous les travaux qui seront présentés par la suite, mais les analyses sur la synthèse des isoprénoïdes et sur celle des acides gras sont des cas particuliers de l'application de cette méthode et assez opposés entre eux. Dans le premier cas les phylogénies préliminaires ont permis une analyse fonctionnelle très minutieuse, alors que dans le deuxième cas le nombre de phylogénies préliminaires a dû être multiplié pour extraire l'information d'intérêt à partir d'une très grande quantité de données. Un cas intermédiaire et assez illustratif de l'importance des phylogénies intermédiaires est celui de mon travail sur l'évolution des carboxylases à biotine, dont je ne parlerai pas dans mes résultats car il a peu de rapport avec les membranes du cenancêtre, mais qui est consultable dans l'article correspondant (Annexe 3).

A terme, la phylogénie préliminaire permet de sélectionner des séquences représentatives de la topologie générale pour la reconstruction de phylogénies plus réduites qui conservent les grandes tendances phylogénétiques et sur lesquelles seront reportées à nouveau les informations fonctionnelles. Le but de la diminution du nombre de séquences est de pouvoir appliquer des méthodes de reconstruction plus fiables mais qui, demandant des temps de calcul plus longs, sont difficilement applicables à l'ensemble des données. A cette étape, différentes méthodes de reconstruction ou modèles d'évolution de séquences ont été testées car la comparaison de méthodes peut permettre de révéler des biais éventuels dus aux méthodes et aux modèles de reconstruction utilisés et, sans être une mesure de robustesse

précise, la congruence entre résultats issus de différentes méthodes soutient les relations observées. Dans les figures qui seront montrées, seuls seront fournis les résultats d'une méthode par enzyme, correspondant à celle qui présente les résultats les plus soutenus et qui ont été publiés dans les articles correspondants.

La phylogénie finale obtenue pour chaque enzyme est discutée. Ensuite, les histoires évolutives des enzymes participant dans une même voie peuvent être comparées, en tenant en compte des informations fonctionnelles disponibles pour l'ensemble de la voie. Bien que l'histoire individuelle de chaque enzyme puisse être très compliquée à cause du grand nombre de fonctions et relations phylogénétiques possibles, il est intéressant de signaler que l'histoire ancienne des enzymes qui participent dans une même voie est généralement congruente entre les différentes enzymes. Cette observation est importante car elle suggère que malgré la malléabilité des protéines du métabolisme, il semble rester suffisamment d'information dans les séquences pour nous renseigner, au moins en partie, sur l'évolution ancienne de ces enzymes, et par conséquent sur les fonctions présentes chez les organismes anciens.

2. Construction des jeux de données

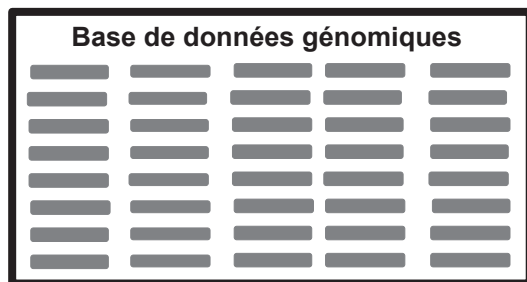
Pour chaque enzyme étudiée, une séquence a été sélectionnée par domaine du vivant dans lequel la fonction est connue. Ces graines ont été obtenues à partir de la base de données KEGG (<http://www.genome.jp/kegg/>) à quelques exceptions près : (1) dans l'étude de la voie du mévalonate de synthèse des isoprénoïdes l'hydroxyméthylglutaryl-CoA réductase (HMGR) de *Pseudomonas mevalonii* a été obtenue dans la Protein Data Bank (<http://www.rcsb.org/pdb/home/home.do>) et l'hydroxyméthylglutaryl-CoA synthase (HMGS) de *Halobacterium salinarum* et l'isopentenyl kinase (IPK) de *Methanocaldococcus jannaschi* de GenBank (<http://www.ncbi.nlm.nih.gov/>) ; (2) dans l'étude de la synthèse des phospholipides à partir de leurs composants de base, la géranylgeranyl-glycérol-phosphate synthase (GGGPS) de *Thermoplasma acidophilum*, la CDP-diacylglycérol-inositol-3-phosphatidyl transférase (PIS) de *Mycobacterium tuberculosis* et une autre CDP-alcool phosphatidyl transférase de *Methanohalobium evestigatum* ont également été obtenues de GenBank.

J'ai réalisé des recherches de similarité par BLASTp [Altschul *et al.* 1990] pour chaque domaine du vivant à partir des enzymes modèles de ce domaine. Lorsqu'aucun homologue d'une enzyme n'était caractérisé dans un domaine du vivant j'ai utilisé les représentants des

>Enzyme X @ *Escherichia coli*

```
MRWRKWRHITKLDPDRNTDEIHKAVADSGTDAV  
MISGTQNVTYEKARTLIEKVSQYGLPIVVEPSDPS  
NVVYDVDYLFVPTVLNSADGDWITGKHAQWVRM  
HYENLQKFTEIIESEFIQIEGYIVLNPDSAVARVTKA  
LCNIDKELAASYALVGEKLFNLPIIY
```

**RECHERCHE DE SEQUENCES
SIMILAIRES (BLAST)**



- Séquences associées à la fonction de l'enzyme X (homologues de l'enzyme X)
- Séquence associée à une fonction très différente (non homologue de l'enzyme X, la première détection n'était pas significative)

```
>VVVV @ Organisme 1  
MRYRKWRHLTKLEPDRNT  
>WWWW @ Organisme 2  
MRWHLWRRITKIDPDRNT  
>XXXX @ Organisme 3  
MRWRLWRHITKLELRIT
```

**BLAST des séquences
dont l'homologie est incertaine**

Figure 42. Vérification par BLAST de l'homologie des séquences détectées.

Suite à une recherche par BLAST on obtient plusieurs séquences similaires. Si l'on hésite sur la pertinence de la détection, on peut utiliser les séquences obtenues pour chercher dans la base de données les séquences les plus proches. Si dans cette deuxième recherche on obtient des séquences dont la fonction est similaire à celle de la protéine de départ, on pourra estimer que la détection n'a pas été aléatoire. Dans le cas inverse la séquence détectée dans la première analyse sera considérée comme non homologue.

autres domaines afin de chercher les séquences similaires. Parfois des doutes pouvaient se présenter quant à l'homologie des séquences détectées (voir par exemple la discussion sur l'évolution des enzymes de synthèse des phospholipides). Dans ces cas j'ai vérifié l'homologie des séquences en utilisant les premières séquences détectées pour faire des BLAST réciproques contre le domaine du vivant de départ dont est issue la protéine étudiée (Figure 42). Si les premiers résultats sont homologues de la séquence de départ, le BLAST réciproque permettra de détecter les séquences réalisant la fonction recherchée dans le domaine original, alors que les non-homologues donneront des résultats différents. Lorsqu'aucun homologue (ou très peu) n'était détecté dans un domaine, des recherches avec PSI-BLAST [Altschul *et al.* 1997] ont été réalisées ; ce programme permet d'utiliser les résultats d'un premier BLAST pour tenir en compte la variabilité qui peut exister pour les différentes positions d'une protéine lors d'une deuxième recherche ; en répétant la procédure sur plusieurs itérations, des protéines homologues mais plus distantes peuvent être détectées.

La liste des génomes dans lesquels les recherches de similarité ont été réalisées a changée au cours de mon travail à cause de l'accumulation de nouveaux génomes séquencés, mais aussi en fonction de la quantité de séquences homologues qui était obtenue dans chaque étude, et qui, notamment dans le cas de la synthèse des acides gras, pouvait facilement saturer les méthodes employées. De fait, la liste des génomes est détaillée pour chaque voie dans une l'Annexe 1. Lorsqu'un domaine ne semblait pas posséder d'homologues, ou en possédait très peu, les recherches étaient complétées avec l'utilisation de banques plus larges. Dans le cas de la voie du mévalonate (synthèse des isoprénoïdes), les recherches de similarité chez les bactéries ont été complétées avec les génomes en cours de séquençage (BLAST assembled genomes de NCBI : http://www.ncbi.nlm.nih.gov/sutils/genom_table.cgi). Dans le cas de certaines enzymes participant dans l'assemblage des phospholipides (partie 3 des résultats) les recherches ont été faites contre l'ensemble de la banque de données non-redondante (nr) de GenBank en restreignant la recherche sur les gènes annotés comme bactériens ou archéens en fonction du domaine dans lequel les homologues des gènes étudiés étaient difficiles à détecter. La recherche des séquences eucaryotes s'est faite contre toute la base de données non-redondante (nr) en restreignant la recherche sur les gènes annotés « eucaryote » dans l'analyse sur la synthèse des isoprénoïdes et contre une liste de génomes complets dans les autres cas (Annexe 1). Pour la voie du méthylerythritol phosphate (synthèse des isoprénoïdes) ces données ont aussi été complétées avec celles d'eucaryotes ne se trouvant pas dans nr. C'était le cas des algues vertes des genres *Micromonas*, *Chlorella* et *Volvox*

(<http://genome.jgi-psf.org/>) et des rhodophytes *Cyanidioschyzon merolae* (<http://merolae.biol.s.u-tokyo.ac.jp/>) et *Galdieria sulphuraria* (<http://genomics.msu.edu/galdieria/>). Les séquences de plusieurs protistes photosynthétiques et non-photosynthétiques ont été recherchées dans les banques de données d'EST (expressed sequence tags) du GenBank et de la Protist EST Database (<http://anabench.bcm.umontreal.ca/anabench/index.jsp>).

3. Analyses phylogénétiques

Pour chaque enzyme, les séquences identifiées par les recherches de similarité au sein des trois domaines ont été regroupées et alignées. Dans le cas des voies de synthèse des précurseurs des isoprénoïdes tous les alignements ont été faits avec Muscle version 3.6 [Edgar 2004]. Pour les premières analyses de la synthèse des acides gras, l'alignement a été fait avec MAFFT [Katoh *et al.* 2005], mais quand le nombre de séquences dans l'alignement a baissé en dessous des 1500 séquences, des nouveaux alignements ont été faits en appliquant indépendamment les deux méthodes, Muscle et MAFFT, à chaque fichier. De même, pour toutes les autres protéines, l'alignement a été fait systématiquement avec les deux méthodes. Souvent, les deux programmes donnaient des alignements comparables, ce qui est à l'avantage de MAFFT, qui est beaucoup plus rapide. Cependant, dans certains cas, spécialement quand les séquences étaient très divergentes ou que les différences de taille entre protéines étaient importantes, Muscle permettait de détecter plus de positions homologues, et avec moins de morcellement. De ce fait, les alignements issus de Muscle ont été privilégiés dans l'ensemble de l'étude.

Les alignements ont été édités grâce aux programmes ED et NET du package MUST [Philippe 1993]. Dans le cas des phylogénies des enzymes des voies de synthèse des acides gras, des analyses avec GUIDANCE [Penn *et al.* 2010] ont permis de raffiner les alignements de façon sensible. Ce programme mesure statistiquement la fiabilité de chaque position de l'alignement ; cela permet de visualiser des sites ou des séquences complètes pour lesquels l'homologie n'est pas bien soutenue, et qui par conséquent ont été éliminés de la suite des analyses.

Les phylogénies préliminaires ont été réalisées par Neighbour Joining [Saitou et Nei 1987] en utilisant l'outil NJ du package MUST [Philippe 1993] dans le cas des analyses sur l'évolution des isoprénoïdes. Dans les autres cas, les analyses préliminaires ont été faites avec

la méthode de maximum de vraisemblance approximée implémentée dans FastTree [Price *et al.* 2001], qui obtient des résultats remarquablement similaires à ceux des phylogénies définitives dans un temps beaucoup plus court. Les analyses phylogénétiques finales ont été faites par maximum de vraisemblance ou avec une inférence bayésienne. Les reconstructions par maximum de vraisemblance ont utilisé le programme TREEFINDER [Jobb *et al.* 2004] avec la matrice de substitution LG et une loi gamma à quatre catégories pour tenir compte de l'hétérogénéité de vitesse d'évolution entre sites (LG + G) [Le et Gascuel 2008]. Ce programme permet aussi d'utiliser d'autres modèles d'évolution, mais bien que certains d'entre eux aient été testés pour certaines protéines, les résultats n'étaient pas sensiblement différents et j'ai opté pour n'analyser en détail que les résultats du modèle LG + G par souci d'uniformisation. Les mesures de robustesse dans l'étude sur la biosynthèse des précurseurs des isoprénoïdes ont été calculées avec 1000 répliquas de bootstrap en utilisant la même méthode. Dans les autres cas, les soutiens des arbres sont les approximations de bootstrap soit de Treefinder soit de FastTree, selon la méthode montrée. L'inférence bayésienne a été possible grâce au programme MrBayes v. 3.0b4 [Ronquist et Huelsenbeck 2003], avec un modèle d'évolution mixte et une distribution gamma à quatre catégories appliqués sur 4 chaînes pendant 1.000.000 de générations. Le consensus a été fait à partir des arbres échantillonnés toutes les 100 générations en ne tenant pas compte des premières 250.000 générations. La stabilisation des paramètres des quatre chaînes ont été vérifiés grâce à l'utilisation du programme TRACER [Rambaut et Drummond 2003].

Pour plus de détails sur les méthodes utilisées dans chaque analyse spécifique, se référer à la partie Résultats et Discussion ou aux articles correspondants (Annexes 2, 5 et 7).

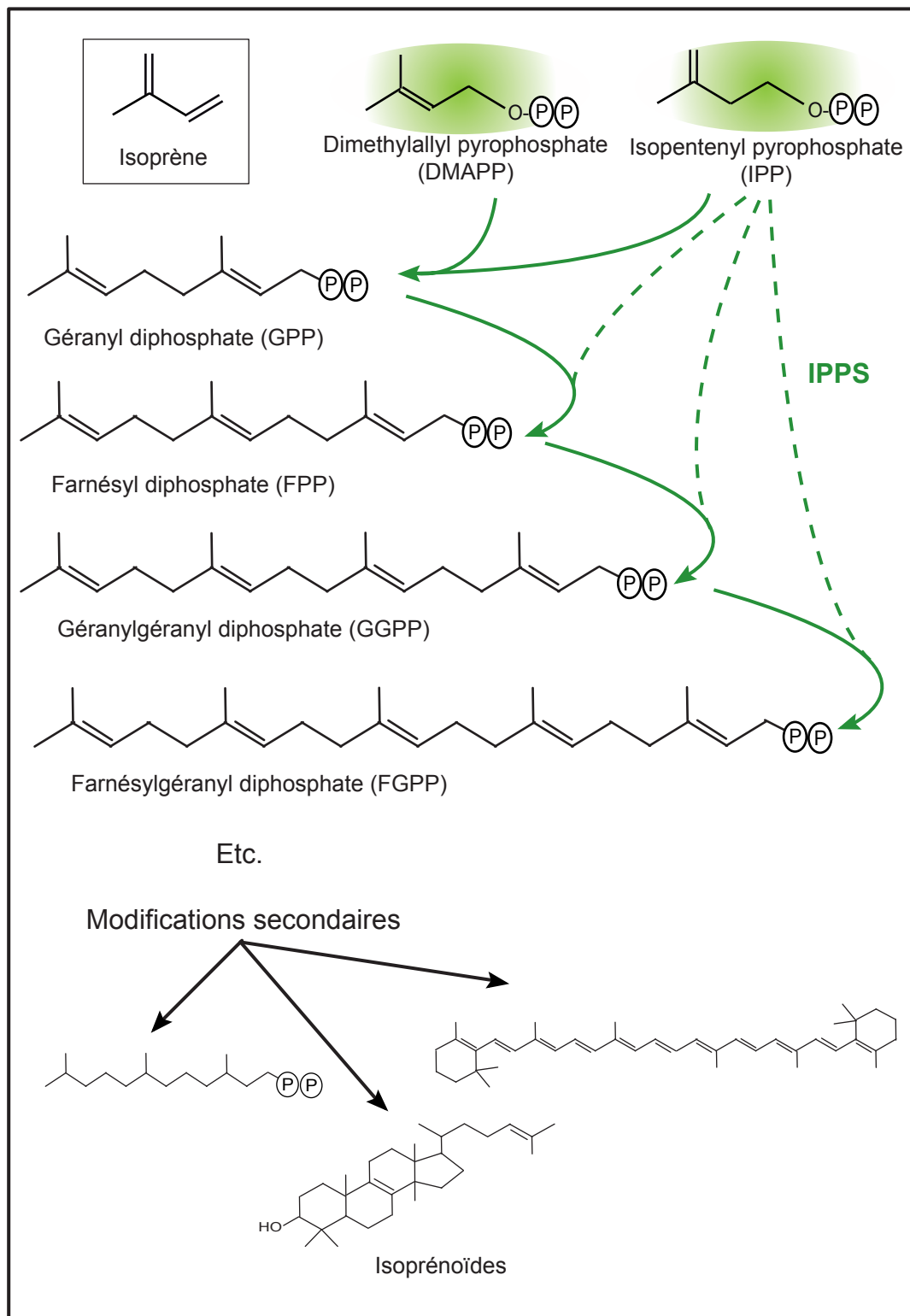


Figure 43. Les isoprénoïdes.

Les isoprénoïdes sont des chaînes composées de formes activées d'isoprène (IPP et DMAPP, molécules avec fond coloré sur la figure) et leurs dérivés. La figure représente la condensation des unités dans des chaînes de plus en plus longues (fonction réalisée par les isoprényl diphosphate synthases, IPPS) puis des modifications secondaires diverses réalisées par une grande diversité d'autres enzymes.

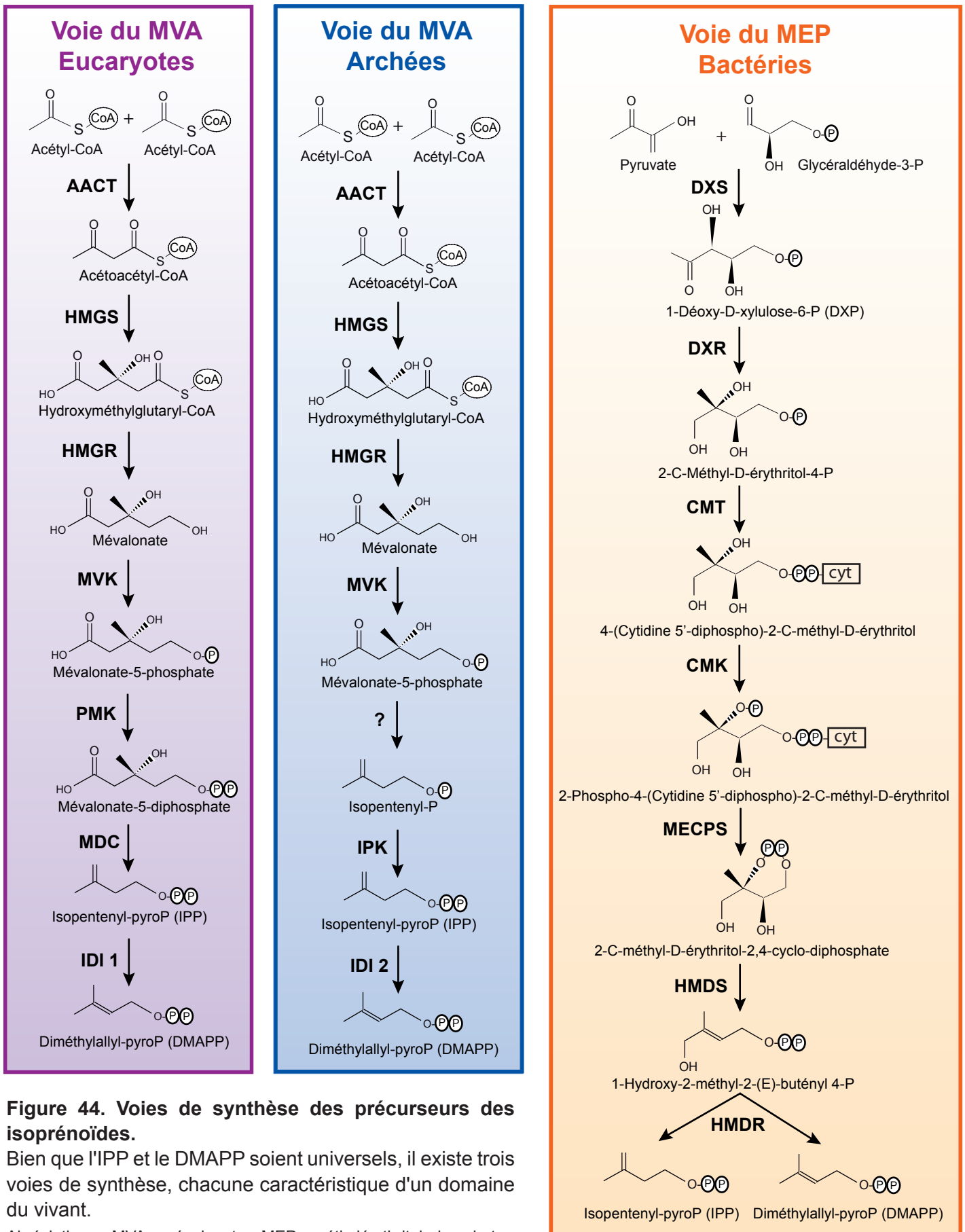
RÉSULTATS ET DISCUSSION

Première partie : Synthèse d'isoprénoïdes dans les trois domaines du vivant.

1. Contexte : Voies de synthèse des isoprénoïdes

Les isoprénoïdes constituent une des familles de molécules les plus riches du vivant, avec plus de 30.000 composés décrits. Ils regroupent par exemple les pigments photosynthétiques, le cholestérol et les hormones stéroïdes, les quinones qui participent aux chaînes de transport des électrons, les chaînes latérales des phospholipides des archées et certaines substances défensives des végétaux [McGarvey et Croteau 1995 ; Lange *et al.* 2000]. Le point commun de toutes ces molécules est qu'elles sont toutes des polymères d'isoprène ou des molécules dérivées de ces chaînes (Figure 43) [Ruzicka 1953]. Cependant, les organismes vivants ne synthétisent pas ces composés directement à partir de l'isoprène mais à partir de formes activées de cette molécule : l'isopentenyl pyrophosphate (IPP) et le diméthylallyl diphosphate (DMAPP, Figure 43). Bien que ces deux unités de base, l'IPP et le DMAPP, soient ubiquitaires chez les organismes des trois domaines du vivant, il existe plusieurs voies de biosynthèse de ces composants (Figure 44).

La première voie à avoir été découverte fut celle du mévalonate (MVA, Figure 44), qui permet de synthétiser les deux unités de base des isoprénoïdes à partir d'acétyl-CoA. Elle fut mise en évidence d'abord chez la levure et les animaux dans les années 1950 (pour une synthèse sur la découverte de ces travaux, voir [Wright 1961 ; Bloch 1964] et on supposa pendant longtemps qu'il s'agissait de la voie de synthèse des isoprénoïdes chez tous les organismes [McGarvey et Croteau 1995]. On a cependant remarqué que les enzymes de cette voie semblaient être absentes chez beaucoup de bactéries [Zhou et White 1991] et qu'elles n'étaient pas responsables dans la production d'au moins une partie des isoprénoïdes chloroplastiques des chlorophytes [Treharne *et al.* 1966 ; Kleinig 1989 ; Rohmer 1999]. En effet, dans les années 1990-2000 une deuxième voie de production d'IPP et de DMAPP fut mise en évidence chez les bactéries et les plantes (pour une synthèse de ces travaux, voir [Rodríguez-Concepción et Boronat 2002 ; Phillips *et al.* 2008]. Cette deuxième voie appelée voie du méthylerythritol phosphate (MEP) produit les unités des isoprénoïdes à partir de pyruvate et de D-glycéraldéhyde-3-phosphate (G3PD, Figure 44).



A la suite de cette découverte, la voie du MVA fut considérée comme la voie majeure de production d'IPP et de DMAPP chez les eucaryotes et les archées [Kuzuyama 2002 ; Boucher *et al.* 2004], et la voie du MEP comme la voie principale de synthèse de ces précurseurs chez les bactéries [Lange *et al.* 2000]. Malgré l'universalité des précurseurs des isoprénoïdes, l'association de la voie du MVA aux archées et eucaryotes et de celle du MEP aux bactéries a conduit à penser que le cenancêtre n'avait peut-être aucune de ces deux voies, qui seraient apparues dans des lignées ultérieures [Lange *et al.* 2000 ; Boucher *et al.* 2004]. Etant donnée l'importance des isoprénoïdes dans les phospholipides des archées, l'incapacité apparente du cenancêtre à produire leurs précurseurs a été l'un des arguments soutenant l'absence de membranes lipidiques chez le cenancêtre et sa compartimentation plutôt dans des cavités minérales [Martin et Russell 2003].

Cependant, la distribution de ces voies de synthèse est plus compliquée et il existe plusieurs exceptions à ces généralités. D'abord, la voie du MVA avait été décrite chez un petit nombre de bactéries [Bochar *et al.* 1999 ; Romanowski *et al.* 2002 ; Voynova *et al.* 2004 ; Steussy *et al.* 2006]. Il était généralement admis que ces bactéries avaient acquis les gènes de cette voie par de multiples HGTs depuis les autres domaines du vivant [Boucher et Doolittle 2000 ; Wilding *et al.* 2000]. Ensuite, on savait que les trois dernières enzymes de la voie classique du MVA, telle qu'elle fut d'abord identifiée chez les eucaryotes, semblaient être absentes de la majorité des génomes d'archées [Smit et Mushegian 2000 ; Boucher *et al.* 2004] et il a été proposé que les rares organismes qui les portent les aient en réalité reçus par HGT [Boucher *et al.* 2004]. Comme nous le verrons plus tard, des enzymes non-homologues qui pourraient catalyser ces réactions ont été proposées [Kaneda *et al.* 2001 ; Barkley *et al.* 2004 ; Grochowski *et al.* 2006 ; Dutoit *et al.* 2008] (Figure 44), laissant ouverte la question de l'évolution de la voie du MVA chez les archées par rapport à celle des eucaryotes. Enfin, la voie du MEP s'ajoute à celle du MVA chez les eucaryotes photosynthétiques [Lange et Ghassemian 2003] et les apicomplexes comme *Plasmodium falciparum* et *Toxoplasma gondii* (qui possèdent des plastes atrophiés non photosynthétiques : les apicoplastes). Le fait que les eucaryotes qui portent la voie du MEP appartiennent à des lignées comportant ou ayant comporté des plastes fait penser à des HGTs des gènes de la voie à partir de l'endosymbionte à l'origine des premiers plastes [Lange *et al.* 2000]. Pourtant, si certaines phylogénies des

enzymes de la voie du MEP soutiennent l'origine cyanobactérienne de ces gènes³³, d'autres pointent plutôt des origines bactériennes différentes des cyanobactéries [Brinkman *et al.* 2002 ; Matsuzaki *et al.* 2008 ; Moustafa *et al.* 2008].

Dans ce contexte, j'ai donc décidé d'étudier précisément l'évolution de chacune des enzymes qui composent les deux voies du MVA et MEP afin de mieux comprendre leur évolution dans les trois domaines du vivant et d'évaluer la possibilité que le cenancêtre ait eu un mécanisme de synthèse des précurseurs des isoprénoïdes ([Lombard et Moreira 2011b], voir Annexe 2).

2. Distribution des enzymes des voies du MVA et du MEP chez les trois domaines du vivant

La voie du mévalonate chez les eucaryotes

La première voie de synthèse des précurseurs des isoprénoïdes à avoir été identifiée dans les années 1950 est celle du mévalonate [Wright 1961]. Le mécanisme de cette voie fait intervenir des réactions classiques (Figure 44). Le réactif de départ est l'acétyl-CoA, dont deux unités sont condensées en acétoacétyl-CoA par une acétoacétyl-CoA thiolase (AACT), puis une troisième est ajoutée par une hydroxyméthylglutatyl-CoA synthase (HMGS). La molécule résultante est alors réduite par une hydroxyméthylglutaryl-CoA réductase (HMGR) qui enlève le coenzyme A et synthétise ainsi le mévalonate qui donne son nom à la voie. Le mévalonate est alors activé par deux phosphorylations consécutives, la première catalysée par une mévalonate kinase (MVK) et la deuxième par une phosphomévalonate kinase (PMK). Le groupement carboxyle terminal du mévalonate-5-diphosphate résultant est alors éliminé par une mévalonate-5-diphosphate carboxylase (MDC) afin de produire le premier précurseur des isoprénoïdes, l'isopentenyl-pyrophosphate (IPP). L'autre précurseur, le diméthylallyl diphosphate (DMAPP), est un isomère du premier qui est produit par le déplacement d'une double liaison par une isopentenyl diphosphate isomérase (IDI).

Cette voie a été mise en évidence principalement chez les eucaryotes et notre recherche des homologues de cette voie dans ce domaine du vivant nous a permis d'identifier leur présence dans un large éventail d'organismes (résultats non montrés). Il est notable que chez

³³ Pour rappel, il est aujourd'hui accepté que l'ancêtre des plastes aurait été une cyanobactérie devenue endosymbiotique dans le cytoplasme d'un ancêtre des plantae [Gray et Doolittle 1982].

les animaux la PMK classique présente chez les autres eucaryotes a été remplacée par une PMK non-homologue [Houten et Waterham 2001], ce qui est corroboré par notre analyse génomique.

La voie du mévalonate chez les archées

L'analyse est plus compliquée chez les autres domaines du vivant. Dans le cas des archées, le principal argument qui soutient la synthèse des isoprénoïdes par la voie du MVA est l'incorporation dans ces organismes de molécules intermédiaires (principalement du ¹⁴C-mévalonate) comportant des marquages isotopiques [Kates *et al.* 1966 ; Kates *et al.* 1968]. Cependant, seules quelques enzymes de cette voie ont été caractérisés biochimiquement chez les archées [Lam et Doolittle 1992 ; Yang *et al.* 2002] et on infère la présence de cette voie dans ce domaine de la détection d'homologues de certains gènes eucaryotes dans les génomes des archées. Dès les premières recherches dans les génomes d'archées il a été signalé que les homologues des quatre premières enzymes de la voie eucaryote du MVA pouvaient être détectés, mais que ce n'était pas le cas des trois dernières étapes [Smit et Mushegian 2000]. Certains auteurs ont donc cherché des gènes candidats pour remplacer ces fonctions chez les archées. La recherche d'une enzyme catalysant les activités de PMK ou de MDC a été infructueuse jusqu'à présent, mais une enzyme capable de synthétiser l'IPP à partir d'isopentenyl phosphate chez *Methanocaldococcus jannaschii* a été mise en évidence, ce qui suppose que les deux étapes (phosphorylation et decarboxylation chez les eucaryotes) soient inversées chez les archées (Figure 44) [Grochowski *et al.* 2006]. L'enzyme archée en question a été appelée isopentenyl phosphate kinase (IPK). L'IDI eucaryote (que l'on appellera ici IDI1) n'ayant pas non plus été détectée chez les archées, une enzyme non homologue (IDI2) capable de catalyser l'isomérisation d'IPP en DMAPP chez les archées a été recherchée et détectée³⁴ [Kaneda *et al.* 2001 ; Barkley *et al.* 2004 ; Dutoit *et al.* 2008]. La recherche des deux IDI chez les eucaryotes montre que ces organismes comportent des IDI1 mais pas d'IDI2 (résultats non montrés).

Dans le cadre de mon analyse chez les archées, *Nanoarchaeum equitans* est la seule archée qui semble totalement dépourvue d'une voie du MVA, ce qui explique la dépendance de cet organisme vis-à-vis de son hôte, *Ignicoccus hospitalis*, pour obtenir des lipides [Jahn *et*

³⁴ Chez les mammifères il existe deux paralogues de l'enzyme IDI1, qui dans certains articles s'appellent IDI1 et IDI2 respectivement [Breitling *et al.* 2003 ; Clizbe *et al.* 2007]. Afin d'éviter toute confusion, j'appellerai ces deux paralogues IDI1a et IDI1b, et IDI2 fera toujours référence à l'enzyme typique des archées.

Tableau 2. Présence ou absence d'homologues dans les génomes complets des archées.

(+) Espèces chez lesquelles ont été détectés des homologues. (II) HMGR de classe II. (?) Séquence détectée mais dont l'orthologie est incertaine.

	HMGS	HMGR	MVK	PMK	MDC	IPK	ID I1	ID I2
Crénarchées	<i>Aeropyrum pernix</i>	+	+	+			+	+
	<i>Desulfurococcus kamchatkensis</i>	+	II	+			+	+
	<i>Staphylothermus marinus</i>	+	II	+			+	+
	<i>Hyperthermus butylicus</i>	+	+	+			+	+
	<i>Ignicoccus hospitalis</i>	+	+	+			+	+
	<i>Metallosphaera sedula</i>	+	+	+	+	+	+	+
	<i>Sulfolobus acidocaldarius</i>	+	+	+	+	+		+
	<i>Sulfolobus solfataricus</i>	+	+	+	+	+	+	+
	<i>Sulfolobus tokodaii</i>	+	+	+	+	+		+
	<i>Pyrobaculum aerophilum</i>	+	+	+			+	+
	<i>Pyrobaculum arsenaticum</i>	+	+	+			+	+
	<i>Pyrobaculum calidifontis</i>	+	+	+			+	+
	<i>Pyrobaculum islandicum</i>	+	+	+			+	+
	<i>Caldivirga maquilingensis</i>	+	+	+			+	+
	<i>Thermofilum pendens</i>	+	II	+			+	+
	<i>Thermoproteus neutrophilus</i>	+	+	+			+	+
	Thaumarchées	<i>Cenarchaeum symbiosum</i>	+	II	+			+
<i>Nitrosopumilus maritimus</i>		+	II	+			+	
Euryarchées	<i>Haloarcula marismortui</i>	+	+	+		+	+	
	<i>Halobacterium salinarum</i>	+	+	+		+	+	+
	<i>Haloquadratum walsbyi</i>	+	+	+		+	+	
	<i>Halorubrum lacusprofundi</i>	+	+	+		+	+	+
	<i>Natronomonas pharaonis</i>	+	+	+		+	+	+
	<i>Natrialba magadii</i>	+	+	+		+	+	+
	<i>Methanoregula boonei</i>	+	+	+			+	+
	<i>Methanospirillum hungatei</i>	+	+	+			+	+
	<i>Methanocorpusculum labreanum</i>	+	+	+			+	+
	<i>Methanoculleus marisnigri</i>	+	+	+			+	+
	<i>Methanococcoides burtonii</i>	+	+	+			+	+
	<i>Methanosaeta thermophila</i>	+	II	+			+	+
	<i>Methanosarcina acetivorans</i>	+	+	+			+	+
	<i>Methanosarcina barkeri</i>	+	+	+			+	+
	<i>Methanosarcina mazei</i>	+	+	+			+	+
	<i>Archaeoglobus fulgidus</i>	+	II	+			+	+
	<i>Ferroplasma acidarmanus</i>	+	II	?			?	+
	<i>Thermoplasma acidophilum</i>	+	II				+	+
	<i>Thermoplasma volcanium</i>	+	II	?			+	+
	<i>Methanocaldococcus jannaschii</i>	+	+	+			+	+
	<i>Methanococcus aeolicus</i>	+	+	+			+	+
	<i>Methanococcus maripaludis</i>	+	+	+			+	+
	<i>Methanococcus vanniellii</i>	+	+	+			+	+
	<i>Methanococcus voltae</i>	+	+	+			+	+
	<i>Methanothermobacter thermautotrophicus</i>	+	+	+			+	+
	<i>Methanobrevibacter smithii</i>	+	+	+			+	+
	<i>Methanosphaera stadtmanae</i>	+	+	+			+	+
	<i>Methanopyrus kandleri</i>	+	+	+			+	+
	<i>Pyrococcus abyssi</i>	+	+	+			+	+
	<i>Pyrococcus furiosus</i>	+	+	+			+	+
	<i>Pyrococcus horikoshii</i>	+	+	+			+	+
	<i>Thermococcus barophilus</i>	+	+	+			+	+
	<i>Thermococcus kodakarensis</i>	+	+	+			+	+
<i>Thermococcus onnurineus</i>	+	+	+			+	+	
<i>Thermococcus</i> sp.	+	+	+			+	+	
<i>Nanoarchaeum equitans</i>								

al. 2004]. Pour les autres archées, nos résultats confirment qu'elles possèdent généralement des homologues des enzymes qui catalysent les premières réactions de la voie du MVA (Figure 44 et Tableau 2) mais pas de celles des trois dernières étapes de la voie eucaryote (PMK, MDC et IDI1). Des homologues d'IPK sont retrouvés dans la grande majorité des archées dont les génomes ont été complètement séquencés (Tableau 2), à l'exception du genre *Sulfolobus*, qui par ailleurs possède les enzymes de type eucaryote PMK et MDC qui manquent en général aux autres archées, et de *N. equitans*, qui ne possède aucune autre enzyme de la voie. Les archées contiennent généralement des IDI2, à l'exception des halobactériales et des thaumarchées, qui possèdent IDI1 en plus -ou à la place- d'IDI2, ce qui fait penser à un remplacement non-homologue d'IDI2 suite à un HGT (Tableau 2). Cette répartition des gènes montre que la voie alternative du MVA (avec IPK et IDI2, Figure 44) semble vraiment caractéristique des archées alors que la version « classique » (avec PMK, MDC et IDI1, voir Figure 44) est typique des eucaryotes. Une première analyse de ces résultats amène à penser que les gènes très répandus dans les génomes d'un domaine étaient probablement présent chez l'ancêtre du domaine respectif et que les gènes limités à quelques groupes taxonomiques sont probablement le résultat de HGTs plus récents, mais ces hypothèses seront réévaluées sur les données des phylogénies.

Les voies du MVA et du MEP chez les bactéries

En ce qui concerne les bactéries, bien que la voie du MVA soit généralement considérée comme absente dans ce domaine, les gènes de celle-ci ont été caractérisés chez certains de ces organismes [Bochar *et al.* 1999 ; Romanowski *et al.* 2002 ; Steussy *et al.* 2006] et semblent être responsables des mêmes fonctions de production des précurseurs des isoprénoïdes [Wilding *et al.* 2000 ; Voynova *et al.* 2004]. Cependant, le faible nombre de bactéries comportant ces gènes a été interprété comme un indice de l'origine xénologue de ces séquences dans ce domaine [Boucher et Doolittle 2000 ; Wilding *et al.* 2000]. Nos recherches des enzymes des voies du MVA dans les génomes bactériens disponibles en 2010 montrent 93 espèces appartenant à 43 genres et à 6 phyla ayant ces enzymes (Tableau 3). Des six phyla, quatre comportent des organismes avec la voie complète de type eucaryote (Firmicutes, Actinobacteria, Proteobacteria, Spirochaeta), le taxon le mieux représenté étant celui des Firmicutes. Il ne manque qu'une enzyme aux Chloroflexi (la PMK), qui sont d'ailleurs parmi les rares bactéries à avoir des homologues de l'enzyme archéenne IPK. Chez les Bacteroidetes seulement une partie des enzymes de la voie du MVA est détectée.

Tableau 3. Présence ou absence d'homologues dans l'ensemble de séquences bactériennes disponibles en 2010.

(+) Espèces chez lesquelles ont été détectés des homologues. (I) HMGR de classe I. (II) HMGR de classe II. (?) Séquence détectée mais orthologie incertaine.

	HMGS	HMGR	MVK	PMK	MDC	IPK
Actinobactéries	<i>Corynebacterium amycolatum</i>	+	I	+	+	
	<i>Gardnerella vaginalis</i>	+	II	+	+	
	<i>Mycobacterium marinum</i>	+	I	+	+	
	<i>Nocardia farcinica</i>	+	I	+	+	
	<i>Streptomyces griseoflavus</i>	+	I	+	+	
Bactéroïdètes	<i>Croceibacter atlanticus</i>		II		+	
	<i>Dokdonia donghaensis</i>		II		+	
	<i>Flavobacteria bacterium</i>	+	II		+	
	<i>Flavobacterium johnsoniae</i>	+	II		+	
	<i>Gramella forsetii</i>		II		+	
	<i>Kordia algicida</i>	+	II		+	
	<i>Leeuwenhoekiella blandensis</i>		II		+	
	<i>Polaribacter irgensii</i>		II		+	
	<i>Psychroflexus torquis</i>		II		+	
	<i>Robiginitalea biformata</i>		II		+	
	<i>Unidentified eubacterium</i>		II		+	
Firmicutes	<i>Bacillus coagulans</i>	+	II	+	+	
	<i>Gemella haemolysans</i>	+	II	+	+	
	<i>Listeria grayi</i>	+	II	+	+	
	<i>Oceanobacillus iheyensis</i>	+	II	+	+	
	<i>Staphylococcus capitis</i>	+	II	+	+	
	<i>Catonella morbi</i>	+	II	+	+	
	<i>Carnobacterium</i> sp.	+	II	+	+	
	<i>Enterococcus faecalis</i>	+	II	+	+	
	<i>Lactobacillus johnsonii</i>	+	II	+	+	
	<i>Lactococcus lactis</i>	+	II	+	+	
	<i>Leuconostoc citreum</i>	+	II	+	+	
	<i>Oenococcus oeni</i>	+	II	+	+	
	<i>Pediococcus pentosaceus</i>	+	II	+	+	
	<i>Streptococcus agalactiae</i>	+	II	+	+	
	<i>Acholeplasma laidlawii</i>	+	II	+	+	
	Chloroflexi	<i>Chloroflexus aggregans</i>	+	II	+	+
<i>Herpetosiphon aurantiacus</i>		+	II	+	+	+
<i>Roseiflexus castenholzii</i>		+	II	+	+	+
Proteobactéries	Rhodobacteraceae bacterium	+	I	+	+	
	<i>Candidatus Liberibacter</i>	+	I	+	+	
Spirochètes	<i>Bdellovibrio bacteriovorus</i>		II	+	+	
	<i>Myxococcus xanthus</i>	+	II	+	+	
	<i>Plesiocystis pacifica</i>	+	II	+	+	
	<i>Sorangium cellulosum</i>	+	II	+	+	
	<i>Stigmatella aurantiaca</i>	+	II	+	+	
	<i>Coxiella burnetii</i>	+	I	+	+	
Spirochètes	<i>Dichelobacter nodosus</i>	+	I	+	+	
	<i>Thiomicrospira crunogena</i>	+	I	+	+	
	<i>Borrelia burgdorferi</i>	+	II	+	?	

La distribution des enzymes IDIs est beaucoup plus compliquée que celles des autres enzymes de la voie du MVA probablement parce que l'isomérisation dont sont responsables les IDIs est aussi une fonction importante dans la voie bactérienne du MEP. La voie du MEP est chimiquement moins directe que celle du MVA (Figure 44). Sa découverte a été possible grâce à une élégante approche multidisciplinaire dont l'historique a été très bien résumé dans [Rodríguez-Concepción et Boronat 2002]. Elle commence avec la décarboxylation d'un pyruvate dont le produit est condensé à une molécule de glycéraldéhyde-3-phosphate pour donner du DXP³⁵. La transcétolase qui catalyse cette réaction est appelée DXP synthétase (DXS). Le DXP doit subir un réarrangement moléculaire suivi d'une réduction pour donner du MEP³⁶, réaction catalysée par la DXP réductoisomérase (DXR). Comme dans le cas de la voie du MVA, le MEP a donné son nom à cette voie car il est considéré comme le premier intermédiaire spécifique de cette route métabolique. Le MEP est doublement activé, d'abord avec le transfert d'un nucléotide triphosphate (le CTP) par une CDP-MEP transférase (CMT)³⁷, puis par la phosphorylation du produit par une CDP-MEP kinase (CDK)³⁸. Ce dernier phosphate ainsi qu'un de ceux du CDP permettent à une MEP-CDP synthase (MCS) de cycliser la molécule³⁹. La fin de la voie du MEP est originale car la résolution du cycle permet la synthèse d'une molécule hydroxylée⁴⁰ qui peut être transformée à la fois en IPP ou DMAPP sans passer par une isomérisation entre les deux molécules, comme dans la voie du MVA. Ceci est possible grâce à deux enzymes (hydroxyméthylbutényl diphosphate synthase et réductase, HMDS et HMDR respectivement), dont la deuxième détermine la quantité produite de chaque isomère. Chez *E. coli*, le ratio est de 5 : 1 en faveur de l'IPP par rapport au DMAPP [Rohdich *et al.* 2002 ; Rohdich *et al.* 2003]. Ce biais explique certainement la présence d'une IDI chez des bactéries qui ne possèdent que la voie du MEP, car elle leur permet ainsi de réguler plus finement leur stock de chaque isomère.

Mon analyse révèle que les deux types d'IDI peuvent être retrouvés chez les bactéries, mais que la distribution des deux est inégale (résultats non montrés). Les actinobactéries,

³⁵ 1-deoxy-D-xylulose-5-phosphate.

³⁶ 2-C-méthyl-D-erythritol-4-phosphate.

³⁷ Le produit de cette réaction est le 4(cytidine 5'-diphospho)-2-C-méthyl-D-erythritol.

³⁸ Le produit de cette réaction est le 2-phospho-4(cytidine 5'-diphospho)-2-C-méthyl-D-erythritol.

³⁹ 2-C-méthyl-D-erythritol-2,4-cyclo-diphosphate.

⁴⁰ 1-hydroxy-2-méthyl-2-(E)-butényl 4-diphosphate.

bactéroidètes et certaines protéobactéries ont IDI1 alors que les firmicutes, cyanobactéries, chlorobi, chloroflexi, *Deinococcus-Thermus*, delta-protéobactéries et certaines spirochètes, actinobactéries et autres protéobactéries ont IDI2. Par conséquent, un plus grand nombre de phyla bactériens possèdent IDI2 qu'IDI1, mais il est difficile de tirer des conclusions en se basant uniquement sur ce point.

La voie du MEP chez les archées et les eucaryotes

Finalement, le même type de recherche a été fait avec les enzymes de la voie du méthylerythritol phosphate chez les archées et les eucaryotes (MEP). Des homologues de chaque enzyme ont été identifiés dans un très grand nombre des bactéries dont les génomes ont été analysés (301 sur 361 génomes ; résultats non montrés). Les séquences homologues chez les archées sont très rares et, en général, très divergentes. Comme attendu, chez les eucaryotes des séquences des gènes de la voie du MEP ont été retrouvées seulement chez les organismes qui ont des plastes ou qui en ont probablement eu dans le passé mais qui les ont perdu (comme les apicomplexes). Etant donné la plus faible quantité de génomes complètement séquencés d'eucaryotes unicellulaires non-photosynthétiques nous avons étendu la recherche d'homologues des gènes de la voie du MEP aux banques d'EST de GenBank et à la Protist EST Database⁴¹, mais aucun résultat positif nouveau n'a pu être identifié.

3. Evolution des enzymes de la voie du mévalonate

Comme nous l'avons vu, le début de la voie du mévalonate est partagé par une majorité d'eucaryotes et d'archées, alors que la fin de la voie diffère entre les deux domaines. Les phylogénies des enzymes de la voie du MVA montrent des topologies qui peuvent être attribuées à deux catégories selon que les protéines soient largement distribuées parmi les archées ou non. De fait, je vais parler d'abord les phylogénies des enzymes de la première partie (commune) de la voie du MVA, puis les phylogénies des gènes terminaux des voies caractéristiques des eucaryotes et des archées.

⁴¹ <http://anabench.bcm.umontreal.ca/anabench/index.jsp>

Enzymes communes aux deux voies

La première enzyme des voies MVA, l'AACT, appartient à la superfamille des thiolases dont les membres sont spécialisés dans l'addition (ou la soustraction) de deux thioesters de type acyl-CoA ou acyl-ACP (Figure 73, p. 163). Cette superfamille contient à la fois l'AACT et HMGS, les deux premières enzymes de la voie du mévalonate, et les céto-acyl synthases (KS) qui catalysent la première étape du cycle d'élongation des acides gras. Le fait que les enzymes responsables de la première réaction des deux voies soient homologues entre elles est évocateur car il suggère que les deux voies aient pu partager une période d'évolution commune dans une époque ancienne. Cependant, l'histoire évolutive des thiolases avait déjà été étudiée dans d'autres travaux et semble être très complexe [Peretó *et al.* 2005 ; Jiang *et al.* 2008]. Je ferai donc uniquement une courte analyse dans la section sur la voie de synthèse des acides gras qui concernera principalement la spécificité de substrat des thiolases. On peut se limiter ici à signaler à propos de la phylogénie de l'AACT que les séquences connues pour catalyser cette réaction groupent ensemble dans les arbres phylogénétiques de la superfamille (Figure 74, p.164). Les AACT, aussi appelées thiolases de type II, sont paralogues d'autres thiolases dites de type I⁴². Les thiolases de type I et de type II existent chez les bactéries et les eucaryotes et sont groupe frère d'un autre groupe de thiolases archéennes qui sont supposées réaliser la fonction AACT chez les archées. Une telle topologie est en général signe que la fonction existait avant la divergence des lignées modernes, mais la phylogénie interne des thiolases archéenne et des thiolases de type I et II montre un très grand nombre de HGTs, paralogies et probablement d'artefacts dus à la divergence de certains de ces gènes. Ainsi, l'étude évolutive à l'échelle des trois domaines du vivant s'avère difficile. La présence d'un gène avec une fonction AACT est probable chez le cenancêtre mais doit être considérée avec précaution.

HMGS, la deuxième enzyme de la voie du MVA, appartient aussi à la superfamille des thiolases. Contrairement aux AACT, la phylogénie de HMGS a été peu étudiée et les séquences bactériennes n'ont pas été prises en compte [Jiang *et al.* 2008]. Pourtant, la

⁴² Il peut apparaître ici une certaine confusion autour du terme thiolase. Le terme *thiolase* fait référence d'abord à des acyltransférases avec la capacité d'ajouter ou de soustraire des thioesters à des chaînes acylées. Certaines thiolases ont des noms spécifiques, comme les KS et HMGS, mais celles qui ont été découvertes sont souvent appelées tout simplement *thiolases*. Enfin, comme les protéines qui réalisent cette fonction sont parentes entre elles, on a l'habitude de les regrouper sous le nom de *superfamille des thiolases*. Il faut donc faire la distinction entre la superfamille des thiolases caractérisée par la fonction de ses membres et les protéines appelées thiolases qui ne constituent qu'une partie de cette superfamille.

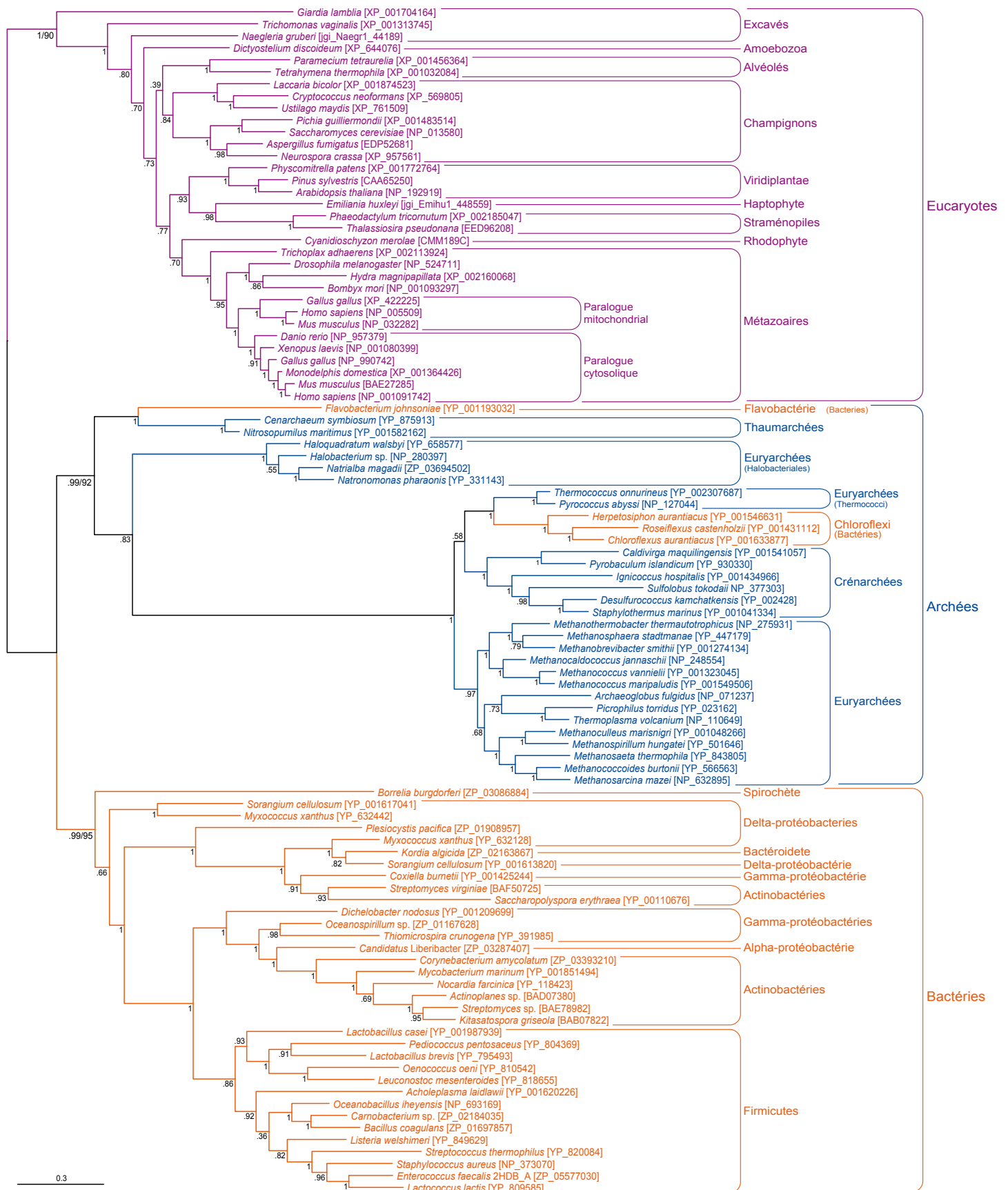


Figure 45. Phylogénie de HMGS.

Reconstruction bayésienne utilisant 99 séquences représentatives et 311 positions conservées. Les nombres à proximité des noeuds indiquent les probabilités postérieures. Les soutiens de bootstrap calculés par des méthodes de maximum de vraisemblance sont aussi indiqués pour les principaux clades. Les couleurs correspondent à l'origine des séquences : les séquences bactériennes apparaissent en orange, les séquences archéennes en bleu et les séquences eucaryotes en violet. Ce code couleur est le même dans toutes les phylogénies présentées dans les résultats de cette thèse.

phylogénie de HMGS est représentative des autres enzymes communes aux deux voies du MVA. Cette phylogénie montre trois grands clades bien soutenus et correspondant aux trois domaines du vivant (Figure 45). A quelques exceptions près, la topologie à l'intérieur des groupes s'accorde bien avec les principaux groupes taxonomiques. Chez les eucaryotes, les exceptions concernent l'existence de deux paralogues chez les animaux [Ayte *et al.* 1990] et la séparation des séquences de champignons et d'animaux. Parmi les archées on observe des séquences bactériennes (des chloroflexi et *Flavobacterium johnsoniae*), qui sont probablement le résultat de transferts depuis ce domaine du vivant. Les archées forment un groupe monophylétique, mais on observe une longue branche qui sépare la plupart d'entre elles d'un sous-groupe formé par les halobactériales, les thaumarchées et la bactérie *F. johnsoniae*. Cette anomalie pourrait s'expliquer par l'acquisition d'une enzyme d'origine bactérienne par HGT ou par l'existence de deux paralogues très distants chez les archées. Nous essayerons plus bas de donner des pistes sur ce point en utilisant la comparaison avec d'autres enzymes de la superfamille des thiolases. Enfin, pour les bactéries nous trouvons aussi un groupe bien soutenu, différencié des autres domaines du vivant et respectant les principaux taxons bactériens, à l'exception de quelques HGTs possibles entre protéobactéries et actinobactéries. Ceci est contradictoire avec l'origine des séquences bactériennes par transfert horizontal (HGT) qui est généralement admise [Wilding *et al.* 2000]. Si cette hypothèse était vraie, on ne s'attendrait pas à trouver un groupe monophylétique mais plutôt des séquences bactériennes dispersées dans l'arbre, se situant à proximité de leurs organismes donneurs respectifs. Une grande divergence des séquences bactériennes après des transferts multiples aurait pu provoquer un artefact d'attraction de longues branches capable de placer ces séquences à la base de l'arbre, mais la bonne conservation de ces séquences et la monophylie de grands taxons bactériens dans cette phylogénie contredisent cette hypothèse. Finalement, la séparation des séquences bactériennes par rapport aux autres domaines n'est pas seulement soutenue par les valeurs de robustesse de l'arbre, mais se retrouve aussi dans les phylogénies de toutes les enzymes de la voie.

Etant donné l'appartenance des HMGS à la superfamille des thiolases, on pourrait mettre en question la fonction des séquences bactériennes (elles pourraient réaliser des fonctions de type thiolase autres que celles des HMGS) ou de grand nombre de séquences archées (étant donné la longue branche à leur base). Dans le cas des bactéries cette possibilité peut être écartée car certaines séquences bactériennes ont été caractérisées biochimiquement [Steussy *et al.* 2006] ; dans le cas des archées, en revanche, aucune caractérisation n'est



Figure 46. Phylogénie de HMGS et de quelques séquences représentatives de KS.

Reconstruction par maximum de vraisemblance utilisant 123 séquences représentatives. Les nombres à proximité des noeuds indiquent les supports de bootstrap.

disponible. Nous avons donc comparé ces séquences à la fois aux HMGS et à d'autres thiolases afin d'approfondir cette question.

Dans la phylogénie de la superfamille, les HMGS forment un groupe clairement distinct des autres thiolases (Figure 74, p.164). Les séquences archéennes qui nous intéressent se placent à la base des HMGS mais sont plus proches de ce groupe que des autres thiolases. Cependant, cette superfamille est très diverse et leurs séquences sont très divergentes, donc l'alignement n'est pas assez fiable pour une analyse détaillée. La reconstruction d'une phylogénie plus fiable contenant uniquement les HMGS et ses paralogues les plus proches, les KS-III, montre des résultats similaires (Figure 46). Afin de compléter cette analyse, nous avons comparé toutes les séquences par rapport aux sites catalytiques de chaque paraglogue. En effet, les acides aminés E79, F185 et S308 du site actif de la séquence HMGS d'*Enterococcus faecalis* [Brown *et al.* 2005 ; Steussy *et al.* 2006 ; Jiang *et al.* 2008] permettent de distinguer les HMGS des KS-III dans les alignements. Les séquences archéennes ont des résidus caractéristiques du site catalytique de HMGS, ce qui soutient l'attribution de cette fonction aux séquences archéennes qui nous intéressent. Bien entendu, la confirmation définitive de la fonction de ces séquences sera soumise à leur caractérisation biochimique. D'un autre côté, la comparaison des Figures 45 et 46 nous montre que la position du groupe basal d'archées n'est pas soutenue. De ce fait, et sachant que les thaumarchées et haloarchées qui forment ce groupe sont souvent soumises à des HGTs [Boucher *et al.* 2003 ; Brochier-Armanet *et al.* 2011b], on peut proposer que ces séquences aient beaucoup divergé après avoir été acquises par HGT depuis un donneur inconnu et que leur position non soutenue soit le résultat d'artefacts de reconstruction.

En résumé, ce que l'on peut retenir de la phylogénie de HMGS est que, malgré l'existence de quelques paralogues et la détection de HGTs, sa topologie est congruente avec les principaux groupes taxonomiques, y compris chez les bactéries, ce qui contredit l'origine par multiples HGTs de ces séquences dans le domaine bactérien.

Le même type d'analyse couplant des reconstructions phylogénétiques et des informations structurelles des sites catalytiques a été réalisé pour chacune des enzymes des voies qui nous intéressent. Des résultats remarquables ont été obtenus avec l'enzyme suivante de la voie du MVA, HMGR. Il existe deux classes différentes mais homologues de cette enzyme [Bochar *et al.* 1999 ; Friesen et Rodwell 2004 ; Hedl *et al.* 2004]. Les enzymes des deux classes forment des homodimères dont chaque monomère apporte à la réaction soit le



Figure 47. Phylogénie de HMGR.

Reconstruction bayésienne utilisant 110 séquences représentatives et 274 positions conservées. Les nombres à proximité des noeuds indiquent les probabilités postérieures. Les soutiens de bootstrap calculés par des méthodes de maximum de vraisemblance sont aussi indiqués pour les principaux clades.

substrat soit le cofacteur. Le site actif des deux classes est très similaire, se situe à l'interface des deux monomères et fait intervenir le même mécanisme de réaction, mais la classe I utilise uniquement le NADPH comme cofacteur tandis que la classe II est aussi capable d'utiliser le NADH. Bien que l'acide aminé catalytique soit dans les deux cas une lysine, la lysine des enzymes de classe I se trouve dans le monomère attaché au substrat contrairement à celle des HMGR de classe II, qui se situe dans le monomère qui apporte le cofacteur. Ces différences structurelles se traduisent aussi par le fait que les HMGR de classe I soient inhibées par des statines⁴³, contrairement aux HMGR de classe II.

Il est intéressant de constater que dans la littérature de biochimie on associe la classe I aux eucaryotes et aux archées, alors que la classe II est caractéristique des bactéries [Bochar *et al.* 1999 ; Friesen et Rodwell 2004]. Quelques exceptions à cette règle existent, car on trouve quelques bactéries comportant une enzyme de classe I et quelques archées avec une classe II ; toutes ces exceptions ont été interprétées comme autant d'indices de nombreux HGTs [Boucher *et al.* 2001 ; Gophna *et al.* 2006]. Nos résultats montrent une phylogénie comportant deux groupes correspondant respectivement aux classes I (eucaryotes, archées, quelques protéobactéries et actinobactéries) et II (bactéries et quelques archées, Figure 47). A l'intérieur de la classe I on observe une séparation des archées et des eucaryotes et on retrouve les principaux groupes taxonomiques de ces domaines, ce qui fait penser à une origine ancestrale chez ces organismes. Ce groupe contient aussi des séquences bactériennes assez divergentes. Contrairement aux résultats précédents qui insistaient sur le grand nombre de HGTs de HMGR de type I chez les bactéries [Gophna *et al.* 2006], nous trouvons un seul clade de bactéries contenant à la fois les protéobactéries et les actinobactéries et ne branchant ni au sein des archées ni au sein des eucaryotes. On peut noter que *Natronomonas pharaonis* a deux paralogues de ce gène dont un qui se place à la base du groupe bactérien. L'origine des séquences bactériennes de la classe I est par conséquent plus énigmatique que ne le laissent penser les travaux précédents, bien que l'origine par HGT semble probable. En ce qui concerne les HMGR de classe II, les bactéries forment à quelques exceptions près des groupes congruents avec les principaux taxons bactériens. Ceci suggère la présence d'une HMGR chez l'ancêtre des bactéries, hypothèse soutenue par les caractéristiques biochimiques de la classe II, qui permettent de considérer les particularités biochimiques de l'HMGR de type II comme

⁴³ Les statines sont des médicaments qui permettent la diminution de synthèse de cholestérol par l'inhibition de la HMGR [Endo 1992].

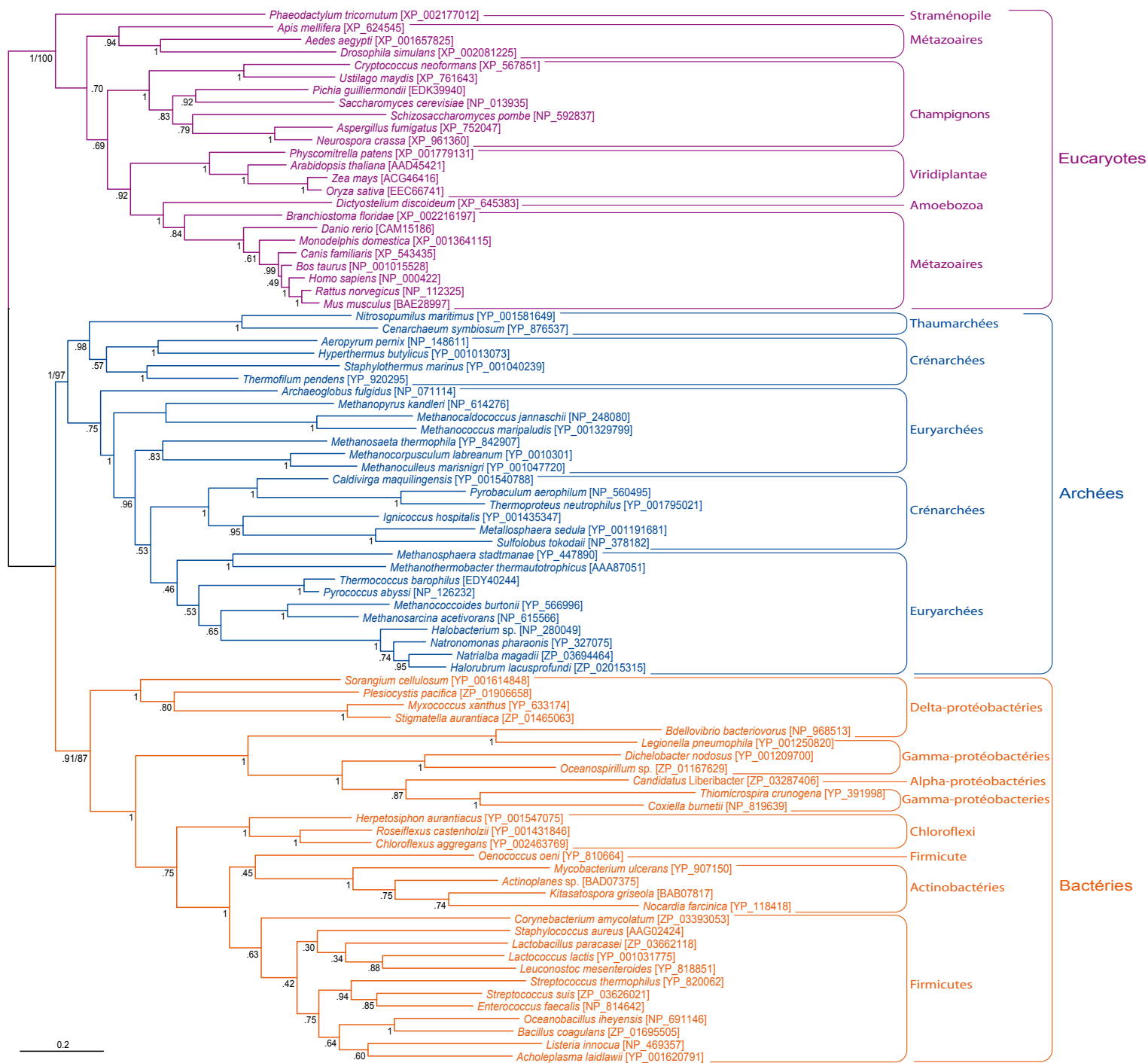


Figure 48. Phylogénie de MVK.

Reconstruction bayésienne utilisant 84 séquences représentatives et 232 positions conservées. Les nombres à proximité des noeuds indiquent les probabilités postérieures. Les soutiens de bootstrap calculés par des méthodes de maximum de vraisemblance sont aussi indiqués pour les principaux clades.

une synapomorphie bactérienne. Les archées qui portent des gènes HMGR de classe II branchent parmi les protéobactéries, suggérant un HGT. Cependant, il est curieux de signaler qu'à l'exception de *Methanosaeta thermophila*, ces séquences branchent ensemble et sont organisées par phylum, suggérant une acquisition relativement ancienne par HGT ou des HGTs multiples entre des archées de groupes différents.

En conclusion, la phylogénie de HMGR montre la même distinction des séquences en trois domaines du vivant que HMGS et les autres enzymes de la voie. En outre, dans le cas de HMGR la séparation des séquences bactériennes par rapport à celles des autres domaines est soutenue par des données biochimiques.

La mévalonate kinase (MVK) est la dernière enzyme commune aux voies du MVA des eucaryotes et des archées (Figure 48). Avec la PMK, la MDC (les deux enzymes qui catalysent les étapes suivantes dans la voie du MVA chez les eucaryotes) et d'autres enzymes participant dans d'autres voies, elle constitue la superfamille des GHMP kinases (Galacto-, Homoserine-, Mévalonate- et Phosphomévalonate-kinases) [Bork *et al.* 1993]. Comme les précédents, l'arbre des MVK montre une division des séquences en trois grands clades correspondant aux trois domaines du vivant (Figure 48). Les séquences bactériennes ne branchent donc pas à l'intérieur des autres groupes à proximité de potentiels donneurs comme on s'y attendrait dans le cas de multiples HGTs récents. Au contraire, elles forment un groupe à part à l'intérieur duquel on observe une topologie en accord avec les principaux groupes taxonomiques bactériens. On retrouve dans les clades eucaryote et archéen les grands taxons acceptés pour ces domaines, bien que cet arbre soit un peu moins bien soutenu que les précédents et que les principaux phyla ne soient pas toujours monophylétiques, sans doute à cause de la moindre conservation des séquences.

Le regroupement des séquences par domaine et la conservation de la phylogénie acceptée des organismes au sein de chaque domaine (notamment des bactéries) sont donc les caractéristiques principales des arbres des premières enzymes des voies du MVA. Les enzymes bactériennes semblent conserver leurs fonctions malgré la grande distance phylogénétique qui les sépare des homologues archéens et eucaryotes. Nous verrons dans la section suivante que les séquences bactériennes de la voie du MVA caractéristique des eucaryotes modernes présentent une distribution et des phylogénies similaires à celles de la première partie de la voie.

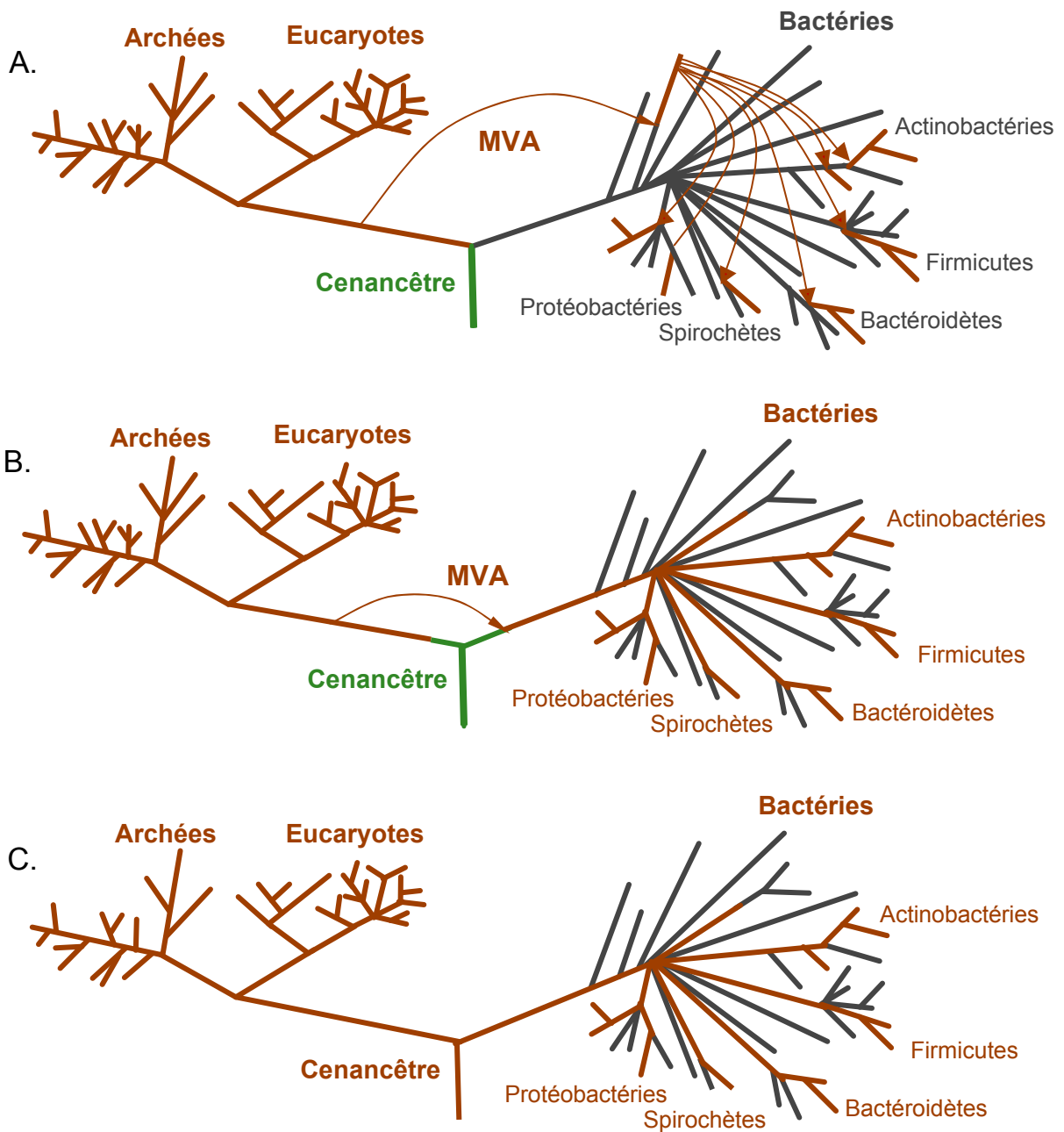


Figure 49. Hypothèses possibles pour expliquer la présence des gènes de la voie du MVA chez les bactéries.

Les schémas représentent une phylogénie des organismes comme celle de Woese *et al.* 1990. Le marron et le gris correspondent respectivement à la présence ou l'absence de la voie du mévalonate au sein des taxons concernés. Les flèches signalent les principaux HGT de l'ensemble de la voie. **A.** Transfert de la voie vers une bactérie et grande divergence suivie de nombreux transferts vers différents taxons bactériens. **B.** Transfert unique vers un ancêtre commun de la lignée bactérienne et transmission verticale ultérieure. Pertes massives de la voie parmi les bactéries. **C.** Héritage vertical à partir du cenancêtre et pertes massives chez les bactéries.

Trois scénarios peuvent expliquer les observations faites sur les séquences bactériennes (Figure 49) : i) il est possible que la voie du MVA ait été transférée vers une bactérie ancienne chez qui il y aurait eu une grande accélération de la vitesse d'évolution pour toutes les enzymes de la voie qui expliquerait leur divergence actuelle par rapport à celles des autres domaines. Ensuite elle aurait transmis horizontalement la voie entière à plusieurs classes de bactéries de différents phyla de façon indépendante. Une autre possibilité est que la voie du MVA soit ancestrale chez les bactéries et proposer ii) soit qu'une bactérie ancestrale vivant avant la diversification des phyla bactériens actuels l'aurait reçue par transfert horizontal ; iii) soit que la voie était déjà présente chez le cenancêtre et héritée verticalement ensuite par les trois domaines du vivant. Ces deux cas ont en leur faveur que la perte de la voie du MVA chez un grand nombre de taxons bactériens pourrait s'expliquer par la présence chez ces mêmes organismes de la voie fonctionnellement redondante du MEP.

Enzymes de la voie eucaryote

Les phylogénies de la PMK et la MDC, les deux premières enzymes de la voie eucaryote du MVA absentes de la voie archéenne (Figure 44), montrent des résultats similaires aux précédents. On détecte la PMK chez les eucaryotes (sauf chez les métazoaires, qui l'ont remplacé par une enzyme non-homologue) [Houten et Waterham 2001 ; Herdendorf et Mizioro 2006], les bactéries et les Sulfolobales (les seules archées à avoir une voie du MVA complète de type eucaryote). Beaucoup de nœuds de l'arbre de PMK ont un support statistique faible et certains groupes, comme les protéobactéries, semblent paraphylétiques (Figure 50), mais la séparation entre les séquences d'eucaryotes et de bactéries et un regroupement des séquences en phyla globalement correct sont encore retrouvés. Dans le cas de MDC on détecte en plus des halobactériales et les thermoplasmatales qui branchent au sein des bactéries (Figure 51). Ceci a été interprété comme un HGT [Boucher *et al.* 2004], et nos résultats confirment cette hypothèse car les séquences d'archées branchent au sein du clade bactérien. En ce qui concerne les séquences de Sulfolobales de PMK et MDC, il a été aussi proposé une origine bactérienne par HGT vers un ancêtre de ces archées [Boucher *et al.* 2004]. Cependant, dans nos arbres de PMK et MDC ces archées se situent à une place intermédiaire entre les eucaryotes et les bactéries avec un fort soutien (Figures 50 et 51).

Cette observation n'est conforme à une origine ni bactérienne ni eucaryote de ces séquences. Au contraire, elle ouvre la possibilité que PMK et MDC aient été ancestrales chez les archées mais que la mise en place de la voie du MVA alternative dans la lignée archéenne

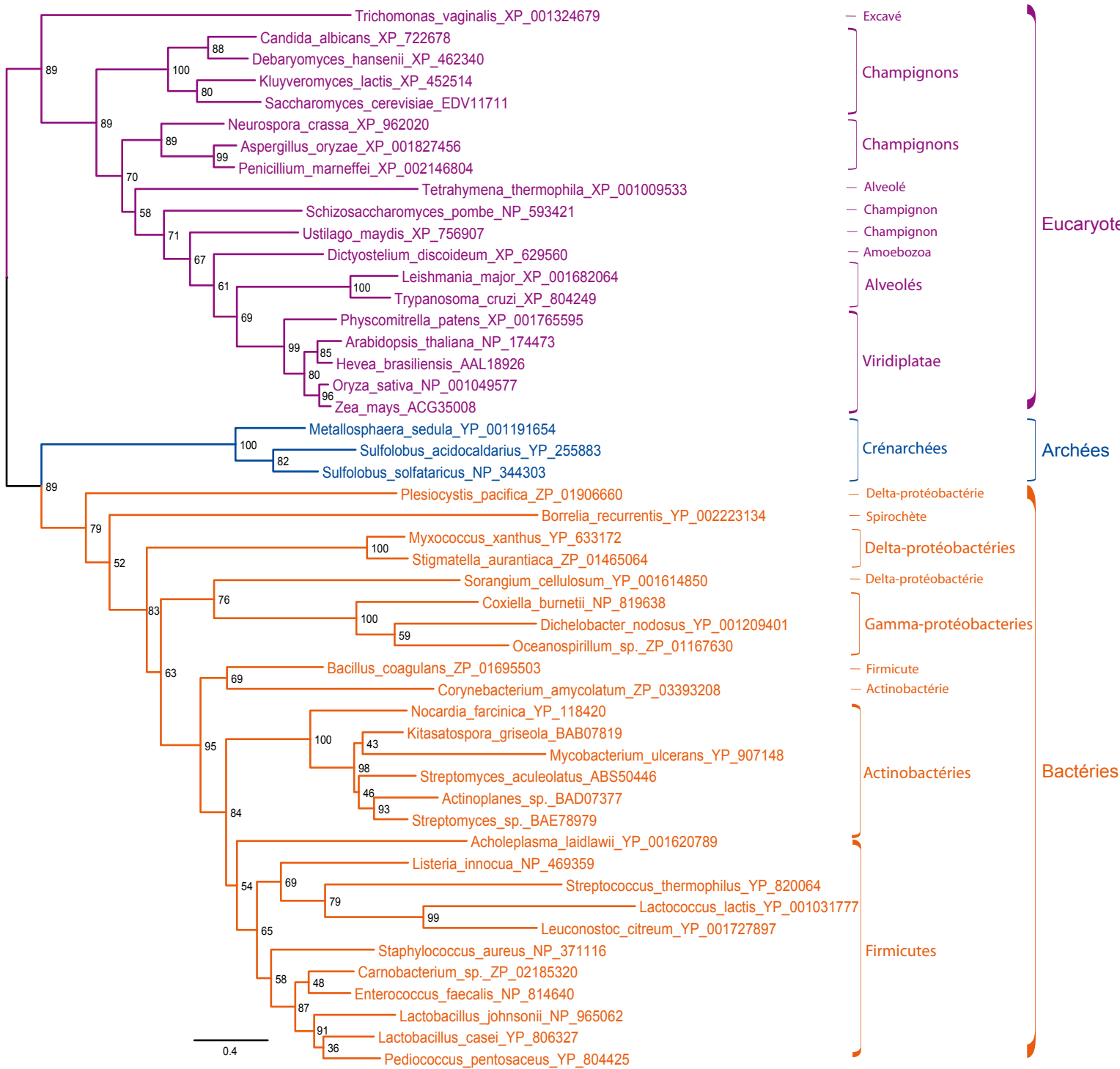


Figure 50. Phylogénie de PMK.

Reconstruction par maximum de vraisemblance utilisant 49 séquences représentatives et 267 positions conservées. Les nombres à proximité des noeuds indiquent les supports de bootstrap.

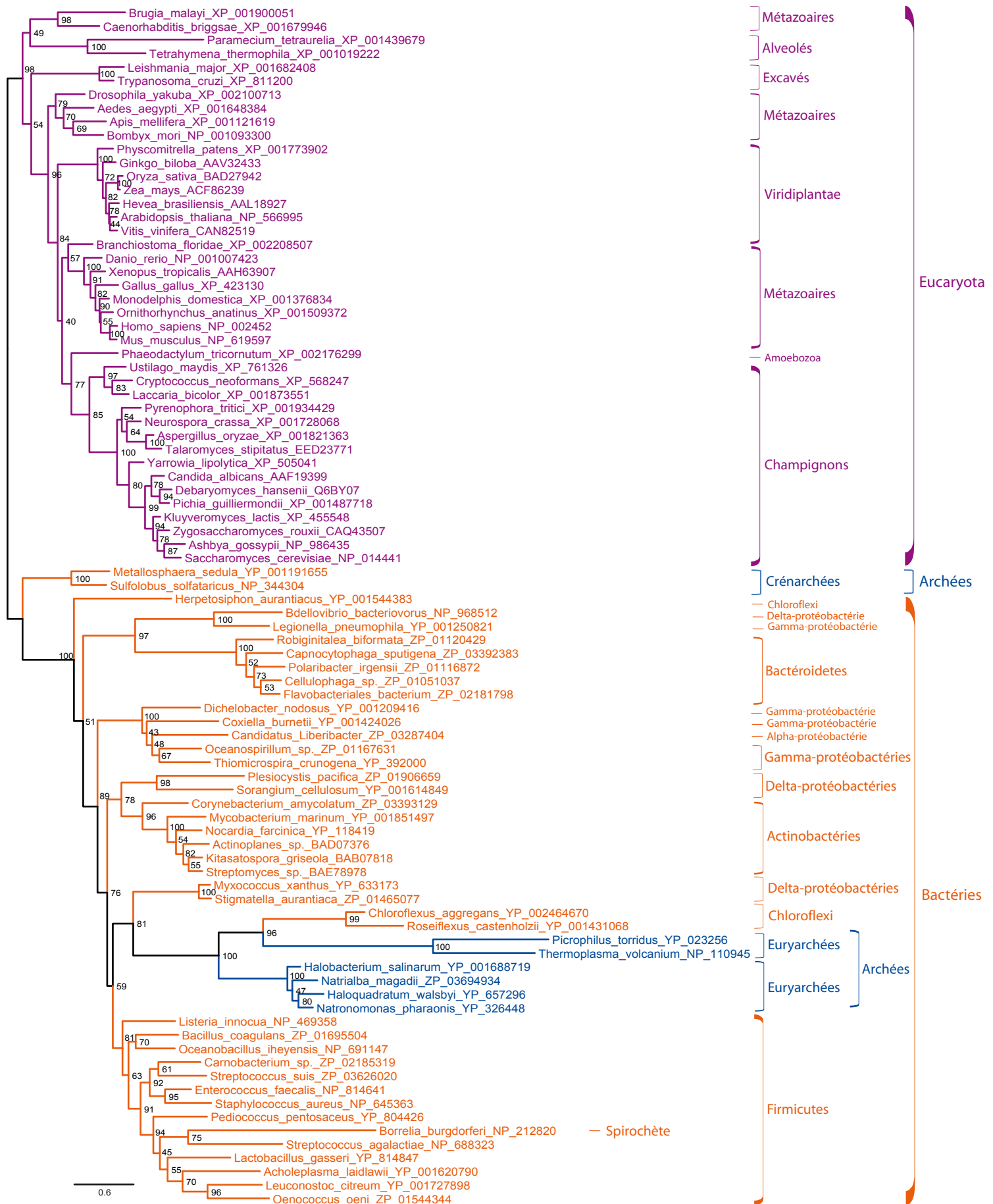


Figure 51. Phylogénie de MDC.

Reconstruction par maximum de vraisemblance utilisant 88 séquences représentatives et 249 positions conservées. Les nombres à proximité des noeuds indiquent les supports de bootstrap.



Figure 52. Phylogénie d'IDI 1.

Reconstruction par maximum de vraisemblance utilisant 99 séquences représentatives et 152 positions conservées. Les nombres à proximité des noeuds indiquent les supports de bootstrap.

aurait remplacé la voie originale dans ce domaine du vivant sauf chez les Sulfolobales, qui auraient gardé les deux. Bien que ces arguments soient affaiblis par la restriction de ces enzymes à la classe des Sulfolobales parmi les archées, PMK et MDC appartiennent à la superfamille des GHMP kinases : cette superfamille est répandue dans les trois domaines du vivant et contient à ce jour huit membres dont la conservation de la structure tertiaire est très forte malgré une conservation faible de la séquence primaire. Ceci se traduit notamment dans un mécanisme réactionnel très conservé malgré l'utilisation de substrats assez différents [Andreassi et Leyh 2004]. Ces indices permettent de proposer que les réactions de PMK et MDC pouvaient être réalisées ancestralement par une ou plusieurs enzymes de cette famille mais que, n'étant pas très spécifiques au départ, elles ont été massivement remplacées dans la lignée archéenne par des enzymes recrutées dans d'autres fonctions.

Dans le cas d'IDI1, il existe à nouveau une séparation claire entre les séquences des eucaryotes et des bactéries (Figure 52). Seules quelques archées comportent des homologues de ces séquences (haloarchées et thaumarchées) et elles branchent parmi les bactéries, ce qui privilégie la thèse d'un HGT. Le support statistique est faible, probablement à cause du petit nombre de positions qui a pu être conservé pour reconstruire les arbres (152 acides aminés). Les eucaryotes et les bactéries semblent monophylétiques, mais la quantité réduite d'information pourrait expliquer les inexactitudes qui existent dans l'arbre, comme la polyphilie des opisthokontes et celle des gamma-protéobactéries. La présence des gènes de la voie eucaryote du MVA chez les bactéries peut s'expliquer avec les mêmes scénarios que pour le reste de la voie du MVA: soit la présence chez l'ancêtre des bactéries et une perte généralisée chez beaucoup de groupes bactériens, soit un transfert ancestral vers une bactérie chez qui la séquence aurait beaucoup divergé suivi de nouveaux transferts vers les ancêtres des différents groupes bactériens.

Enzymes de la voie archéenne

Comme il a été signalé précédemment, le fonctionnement de cette voie n'est pas encore complètement caractérisé. Concrètement, l'enzyme qui catalyserait la première réaction après la MVK n'a jamais été décrite. L'IPK a été identifiée chez *Methanocaldococcus jannaschii* et serait la deuxième enzyme de cette voie, mais il n'existait pas de phylogénie de cette enzyme avant ce travail. La distribution d'IPK est très différente du reste des enzymes des voies du MVA, car elle est présente chez la grande majorité des archées mais chez très peu d'organismes des autres domaines (Tableaux 2 et 3). Les séquences eucaryotes forment un

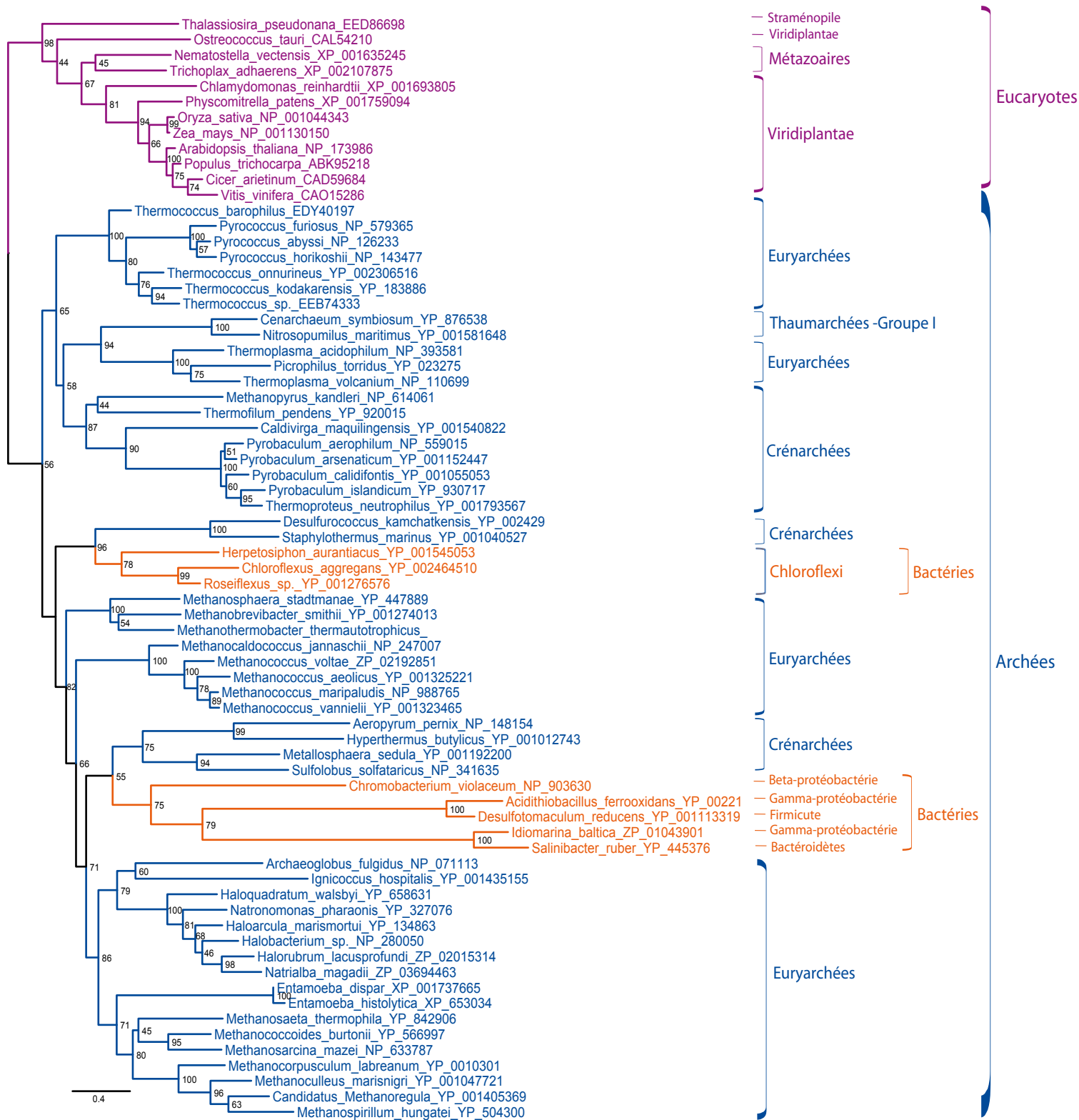


Figure 53. Phylogénie de l'ensemble des séquences d'IPK.

Reconstruction par maximum de vraisemblance utilisant 71 séquences représentatives et 204 positions conservées. Les nombres à proximité des noeuds indiquent les supports de bootstrap.

groupe extérieur aux archées (Figure 53) mais elles sont peu nombreuses et leur fonction est inconnue [Dellas et Noel 2010]. Les séquences bactériennes branchent au sein des archées, mais la plupart d'entre elles ont de très longues branches qui font douter d'une conservation de la fonction IPK. Cette réserve semble d'autant plus pertinente que depuis notre analyse des caractérisations biochimiques des gènes IPK de *Thermoplasma acidophilum* et *Methanothermobacter thermautotrophicus* ont montré leur capacité à phosphoryler aussi des substrats autres que l'isopentenyl phosphate [Chen et Poulter 2010]. Le branchement des séquences bactériennes parmi les archées fait penser à des HGTs. La faible diversité eucaryote suggérerait que ces séquences soient d'origine archéenne, mais qu'après plusieurs transferts il y aurait eu une grande divergence responsable d'un phénomène d'attraction de longues branches ; l'hypothèse alternative serait la présence ancestrale d'un homologue de ces séquences chez les eucaryotes et leur perte dans la plupart des lignées eucaryotes.

Une difficulté des analyses d'IPK est que ses séquences sont très divergentes entre elles. Par conséquent, leur alignement et l'arbre qui en découle ne sont pas très robustes. Nous avons donc inféré un nouvel arbre uniquement avec les séquences d'archées afin de minimiser le bruit (Figure 54). Bien que le nouvel alignement comporte encore beaucoup de sites variables probablement à l'origine d'artefacts de reconstruction phylogénétique dont sont témoins les faibles valeurs de soutien, les groupes archéens obtenus sont représentatifs des classes admises dans ce domaine. On obtient un groupe d'euryarchées et un groupe contenant à la fois des crenarchées, thaumarchées et des euryarchées des classes Thermoplasmatales et Thermococcales. Ces résultats soutiennent l'origine ancestrale d'IPK chez les archées, puis de nombreux HGTs entre archées. Cette hypothèse semble d'autant plus probable que certains de ces groupes, comme les Thermoplasmatales et les thaumarchées, sont particulièrement connus pour avoir reçu de nombreux gènes par HGT ([Ruepp *et al.* 2000 ; Brochier-Armanet *et al.* 2011a]).

Comme signalé précédemment, IDI2 se retrouve chez un grand nombre de phyla bactériens en plus des archées (Figure 55). La topologie de l'arbre semble montrer certains artefacts de reconstruction (en particulier la paraphylie des euryarchées), mais nous observons une séparation des archées et des bactéries. A l'intérieur des phyla, on observe des groupes taxonomiques de rang inférieur, mais beaucoup de séquences ne groupent pas en fonction de leur taxonomie, ce qui est probablement le résultat de HGTs. Notamment, il existe un groupe de firmicutes qui branche parmi les crenarchées et des nombreux transferts semblent avoir existé entre protéobactéries. Malgré ces événements phylogénétiques et quelques possibles

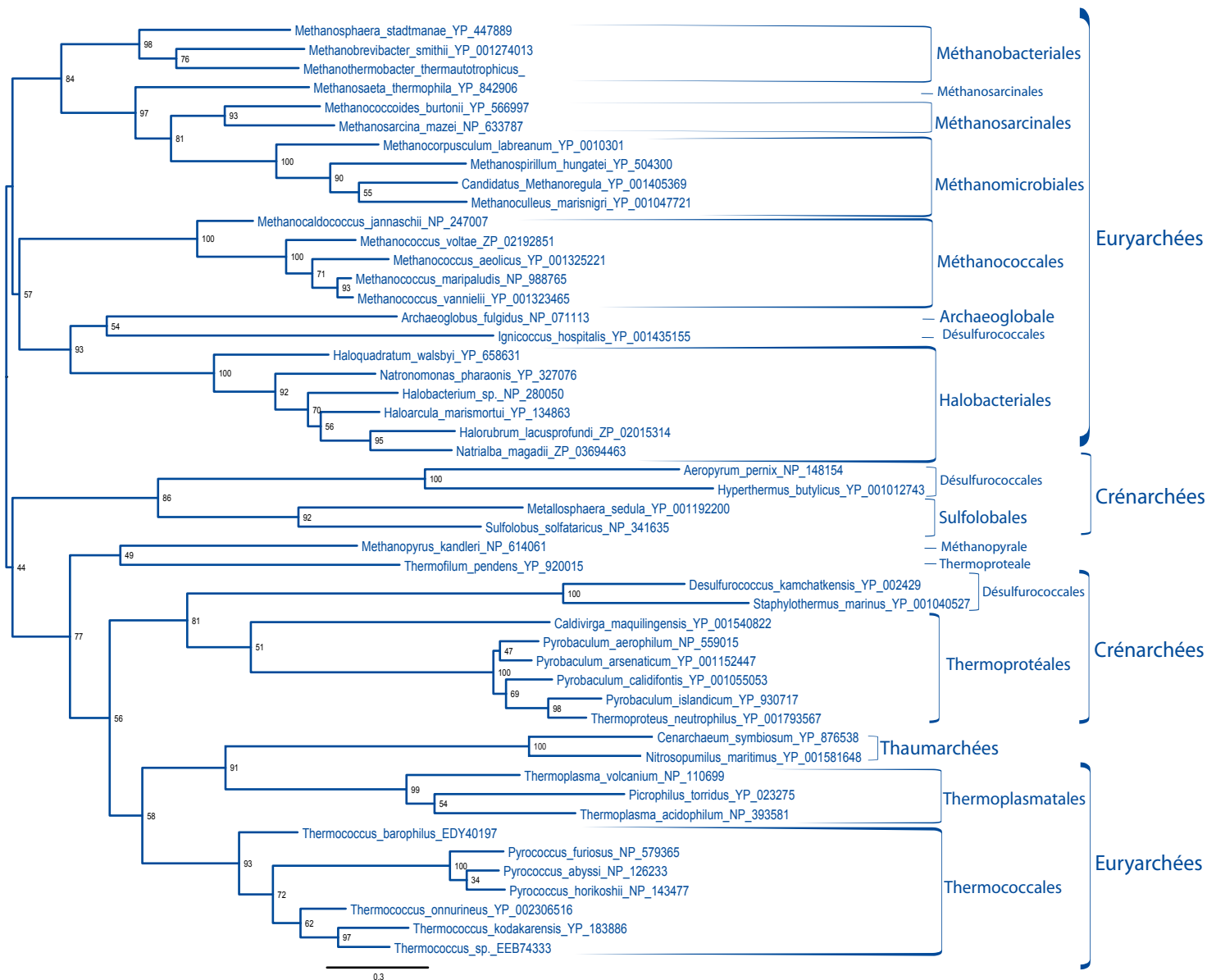


Figure 54. Phylogénie des séquences IPK archéennes.

Reconstruction par maximum de vraisemblance utilisant 49 séquences représentatives et 199 positions conservées. Les nombres à proximité des noeuds indiquent les supports de bootstrap.

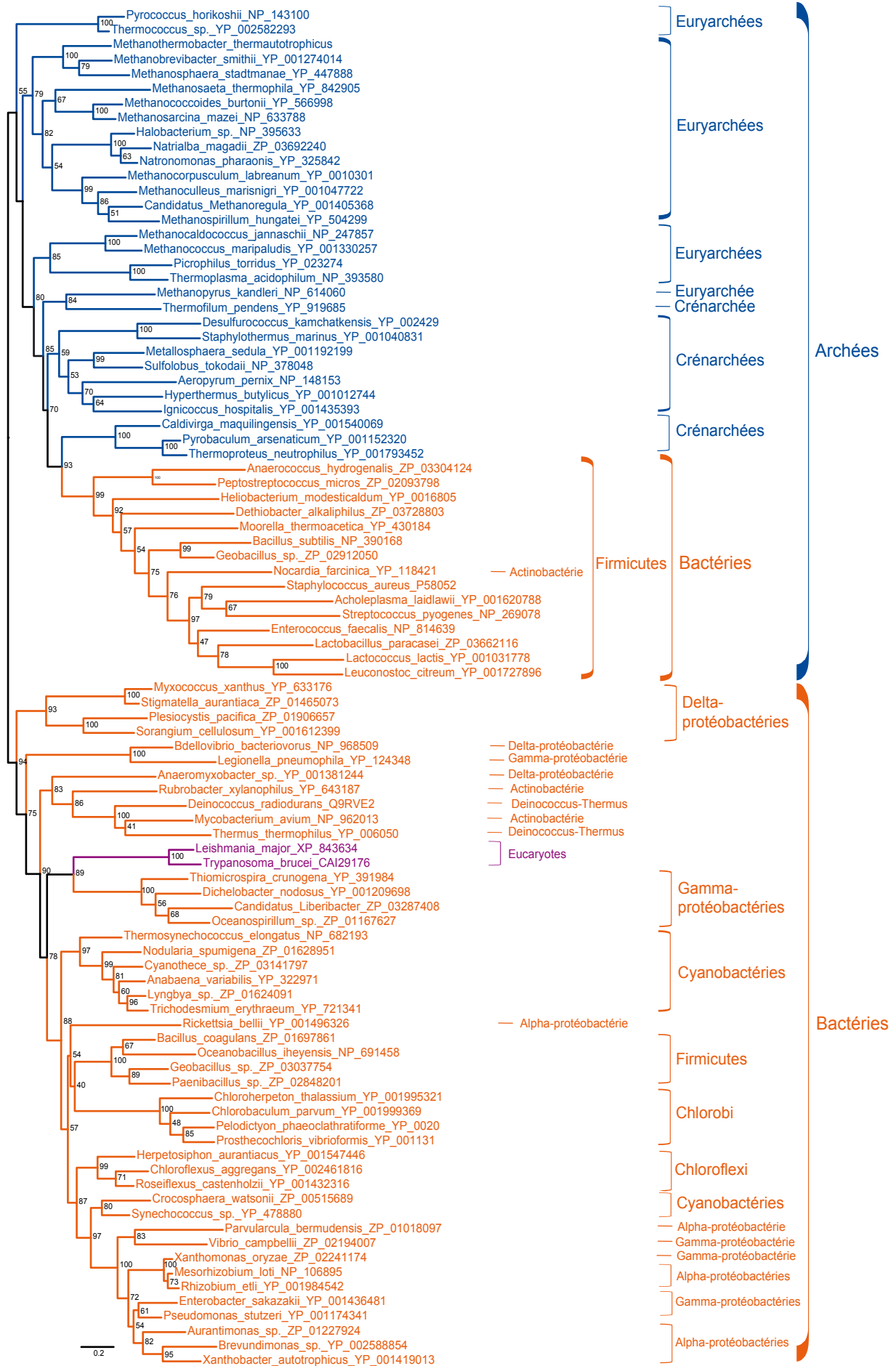


Figure 55. Phylogénie d'IDI 2.

Reconstruction par maximum de vraisemblance utilisant 93 séquences représentatives et 308 positions conservées. Les nombres à proximité des noeuds indiquent les supports de bootstrap.



Figure 56. Phylogénie de DXS.

Reconstruction par maximum de vraisemblance utilisant 124 séquences représentatives et 508 positions conservées. Les nombres à proximité des noeuds indiquent les supports de bootstrap.

erreurs de reconstruction, ces résultats suggèrent l'ancestralité d'IDI2 chez les archées. Etant donné le grand nombre de phyla bactériens qui possèdent IDI2 et le regroupement taxonomique de beaucoup de ces séquences, il semblerait aussi possible qu'IDI2 ait été ancestrale chez ces organismes, puis massivement perdue. Les raisons des pertes peuvent être diverses : la capacité de la voie du MEP présente chez les bactéries à produire directement l'IPP et le DMAPP, la présence redondante d'IDI1 et IDI2, voire leur remplacement réciproque.

Ces résultats sont en accord avec l'existence d'une voie alternative du MVA chez les archées [Grochowski *et al.* 2006] et, même si une phosphomévalonate decarboxylase reste encore à identifier, ils soutiennent que cette voie est répandue, caractéristique et probablement ancestrale aux archées.

4. Evolution des enzymes de la voie du méthylerythritol phosphate

La voie du MEP est la principale voie de production d'IPP et de DMAPP chez les bactéries. Selon la littérature, tous les phyla bactériens examinés possèdent cette voie et très peu de séquences non-bactériennes ont pu être identifiées pour ces enzymes, à l'exception des eucaryotes possédant ou ayant pu posséder dans le passé des plastes [Matsuzaki *et al.* 2008]. Cette distribution a fait penser à un transfert des gènes de type cyanobactérien depuis l'endosymbionte à l'origine du plaste vers le génome nucléaire de ces eucaryotes [Lange *et al.* 2000]. Cependant, certains travaux ayant reconstruit les phylogénies indépendantes des gènes de la voie du MEP ont montré que pour certains gènes les eucaryotes étaient plus proches de groupes autres que les cyanobactéries, ce qui a été source d'un fort débat à propos de l'origine de ces séquences chez les eucaryotes. Bien que, comme nous le verrons plus tard, dans différents arbres ont trouvé les eucaryotes associés à des phyla bactériens différents, la relation qui a le plus attiré l'attention était celle entre les eucaryotes et les chlamydiales, un groupe de bactéries parasites intracellulaires. Certains auteurs ont proposé un lien évolutif étroit entre les chlamydiales et les cyanobactéries avant l'endosymbiose plastidiale pour expliquer le rapprochement entre certains gènes d'eucaryotes photosynthétiques et les chlamydiales [Brinkman *et al.* 2002]. D'autres auteurs ont suggéré des modèles dans lesquels la présence d'une chlamydiale dans le cytoplasme de l'hôte eucaryote primitif aurait facilité le maintien de la cyanobactérie nouvellement endocytée [Moustafa *et al.* 2008]. Une autre hypothèse explique l'écart entre les cyanobactéries et les eucaryotes par des HGTs entre les

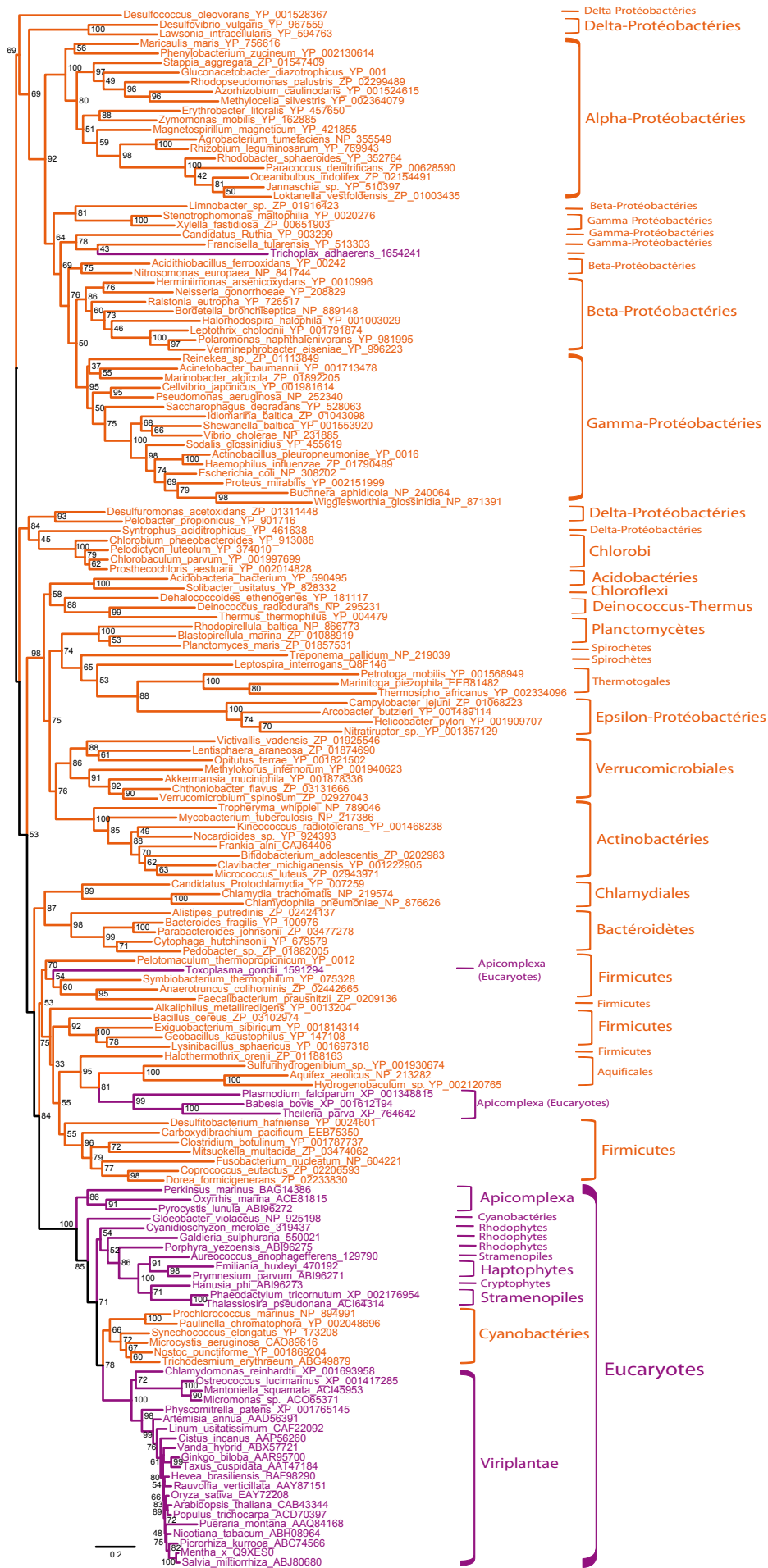


Figure 57. Phylogénie de DXR.

Reconstruction par maximum de vraisemblance utilisant 163 séquences représentatives et 336 positions conservées. Les nombres à proximité des noeuds indiquent les supports de bootstrap.

bactéries postérieurs à l'endosymbiose plastidiale, ce qui occulterait la vraie phylogénie des enzymes [Lange *et al.* 2000]. Pourtant, en ce qui concerne les gènes de cette voie, toutes ces hypothèses sont confrontées à la présence chez les eucaryotes de gènes proches de groupes bactériens autres que les chlamydiales. Je vais donc analyser brièvement les phylogénies des gènes de la voie du MEP afin d'essayer de résoudre cette question.

Origine des gènes de la voie du MEP chez les eucaryotes

Bien que la phylogénie des bactéries soit difficile à résoudre, le fait que les phylogénies des enzymes de la voie du MEP respectent les grands groupes taxonomiques des bactéries fait penser que cette voie était déjà présente chez les bactéries. Je n'ai pas traité le problème de la phylogénie des séquences bactériennes et je me suis plutôt concentré sur la position des séquences eucaryotes par rapport aux séquences bactériennes. Comme pour les voies du MVA, les phylogénies des enzymes de la voie du MEP montrent des caractéristiques communes, donc nous les détaillerons pas toutes ici, et on se focalisera sur les spécificités de chaque reconstruction.

La première enzyme de la voie du MEP est la DXS. Son arbre phylogénétique montre le regroupement des bactéries en phyla et superphyla avec très peu d'exceptions, comme par exemple les spirochètes (Figure 56). Au sein des superphyla, la topologie est bien soutenue et cohérente avec la phylogénie admise des organismes, et ceci s'avère aussi exact pour les séquences d'eucaryotes. Dans cet arbre, les eucaryotes se placent de façon très soutenue comme groupe frère des alpha-protéobactéries et loin des cyanobactéries, au contraire de ce à quoi on se pouvait s'attendre pour expliquer la présence de la voie chez les eucaryotes à partir d'un endosymbionte plastidial. La proximité des DXS eucaryotes aux DXS d'alpha-protéobactéries soutient plutôt leur origine mitochondriale (l'ancêtre de cet organelle serait une alpha-protéobactérie) et, par conséquent, on peut proposer le remplacement de la DXS plastidiale par la DXS mitochondriale chez les eucaryotes comportant des plastes. De leur côté, les cyanobactéries branchent parmi les firmicutes, ce qui reflète un possible HGT des secondes vers les premières comme proposé dans [Lange *et al.* 2000].

La phylogénie de DXS donne le ton général des phylogénies des enzymes de la voie du MEP : des phyla bactériens souvent monophylétiques et bien soutenus, mais une mauvaise résolution entre les différents phyla. C'est notamment ce que l'on trouve dans les phylogénies de DXR, CMK et HMDR. La phylogénie de DXR montre les eucaryotes associés aux cyanobactéries (Figure 57). Celles-ci se placent entre la lignée rouge (algues rouges et autres

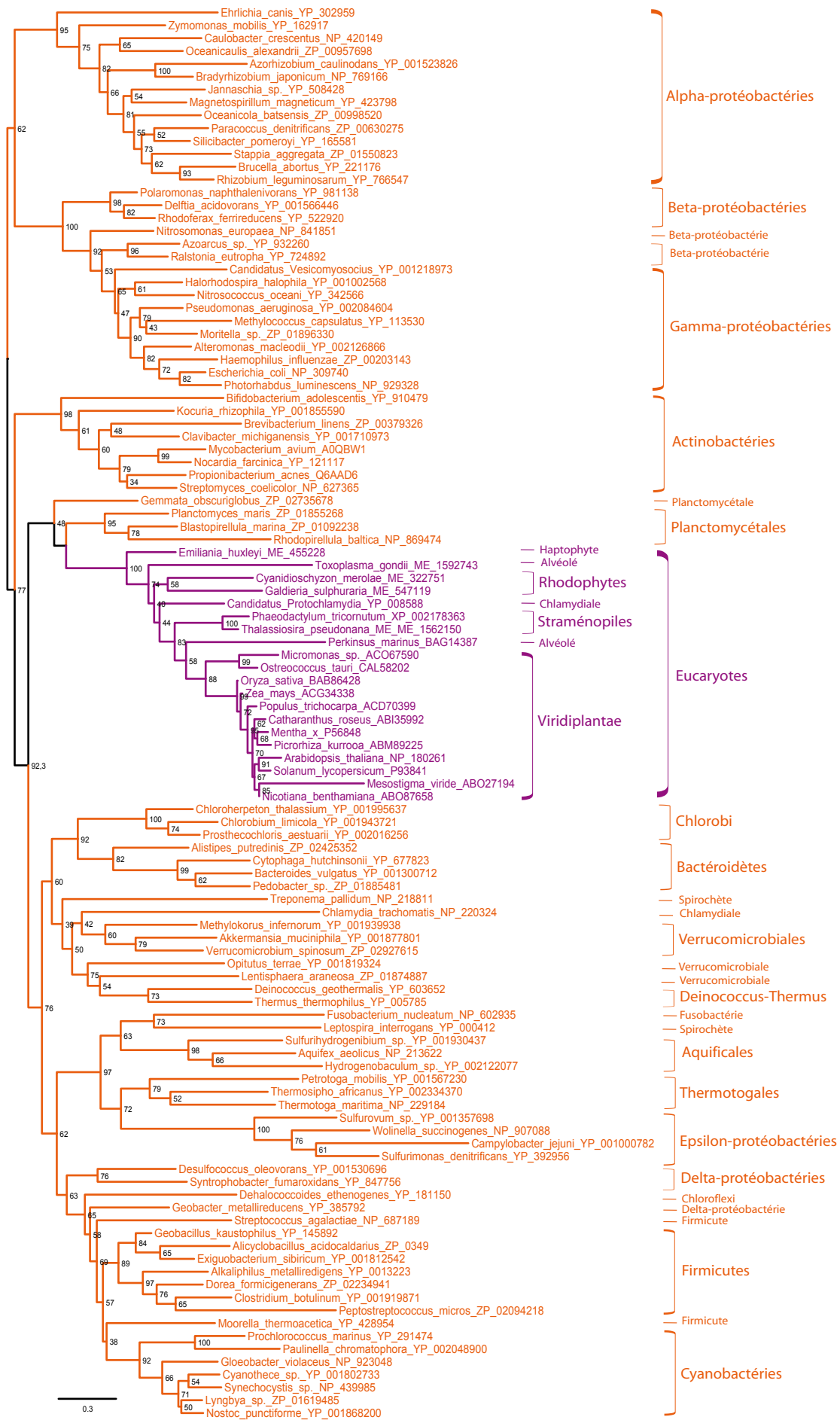


Figure 58. Phylogénie de CMK.

Reconstruction par maximum de vraisemblance utilisant 110 séquences représentatives et 170 positions conservées. Les nombres à proximité des noeuds indiquent les supports de bootstrap.

eucaryotes qui ont acquis un plaste par endosymbiose d'une algue rouge) et la lignée verte (algues et plantes vertes) probablement à cause d'un artefact lié à la grande divergence des séquences de la lignée rouge. Dans le cas de CMK les principaux groupes taxonomiques bactériens sont respectés malgré quelques imprécisions concernant par exemple la place des delta-protéobactéries, des spirochètes et des cyanobactéries (Figure 58). Dans cette phylogénie les eucaryotes branchent de façon peu soutenue parmi les planctomycétales. Pour HMDR les cyanobactéries branchent parmi les eucaryotes au lieu de l'inverse, ce qui est probablement le résultat d'un artefact comme dans le cas de DXR (Figure 59). Malgré cet artefact il semble raisonnable de penser que les cyanobactéries sont le groupe bactérien donneur de ces séquences, plutôt que les verrucomicrobiales ou les bactéroïdètes qui sont groupe frère des eucaryotes. CMT et MECPS sont des petites enzymes, ce qui ne permet d'utiliser que peu de positions pour reconstruire leurs phylogénies (147 et 136 acides aminés respectivement). De plus, ces deux enzymes se trouvent côte à côte dans les génomes de beaucoup de bactéries et chez certains groupes (les alpha- et les epsilon-protéobactéries, quelques firmicutes et des actinobactéries) et elles ont subi des événements indépendants de fusion [Gabrielsen *et al.* 2004]. Cette convergence de caractères, qui a pu entraîner des nouvelles contraintes structurales, s'ajouterait à leur petite taille pour expliquer la faible cohérence des arbres de CMT et de MECPS avec la taxonomie des organismes et leur faible robustesse (Figures 60 et 61). Ces phylogénies doivent donc être interprétées avec prudence mais on peut signaler que la phylogénie de CMT montre que le groupe le plus proche des eucaryotes est celui des chlamydiales, tandis que pour MECPS les méthodes de reconstruction par maximum de vraisemblance et bayésiennes ne donnent pas les mêmes résultats et placent les planctomycétales ou les aquificales à la base des eucaryotes. Etonnamment, dans la phylogénie des MECPS les apicomplexes sont associées aux chlamydiales, loin des autres eucaryotes. Les informations sur le groupe bactérien le plus proche des eucaryotes pour chaque enzyme sont résumées dans le Tableau 4.

Je vais me concentrer sur l'HMDS car la phylogénie de cette enzyme pourrait aider à éclaircir les histoires évolutives des autres enzymes de la voie. L'évolution de cette enzyme chez les chromalvéolés (l'ensemble eucaryote composé par les alvéolés, les straménopiles, les haptophytes et les cryptophytes) est extrêmement atypique, comme il avait été observé dans des études précédentes [Frommolt *et al.* 2008]. Comme pour tous les autres arbres de la voie, celui des HMDS montre un bon regroupement des espèces en fonction de leurs phyla mais une pauvre résolution des relations entre les phyla (Figure 62). Chez les eucaryotes, on trouve

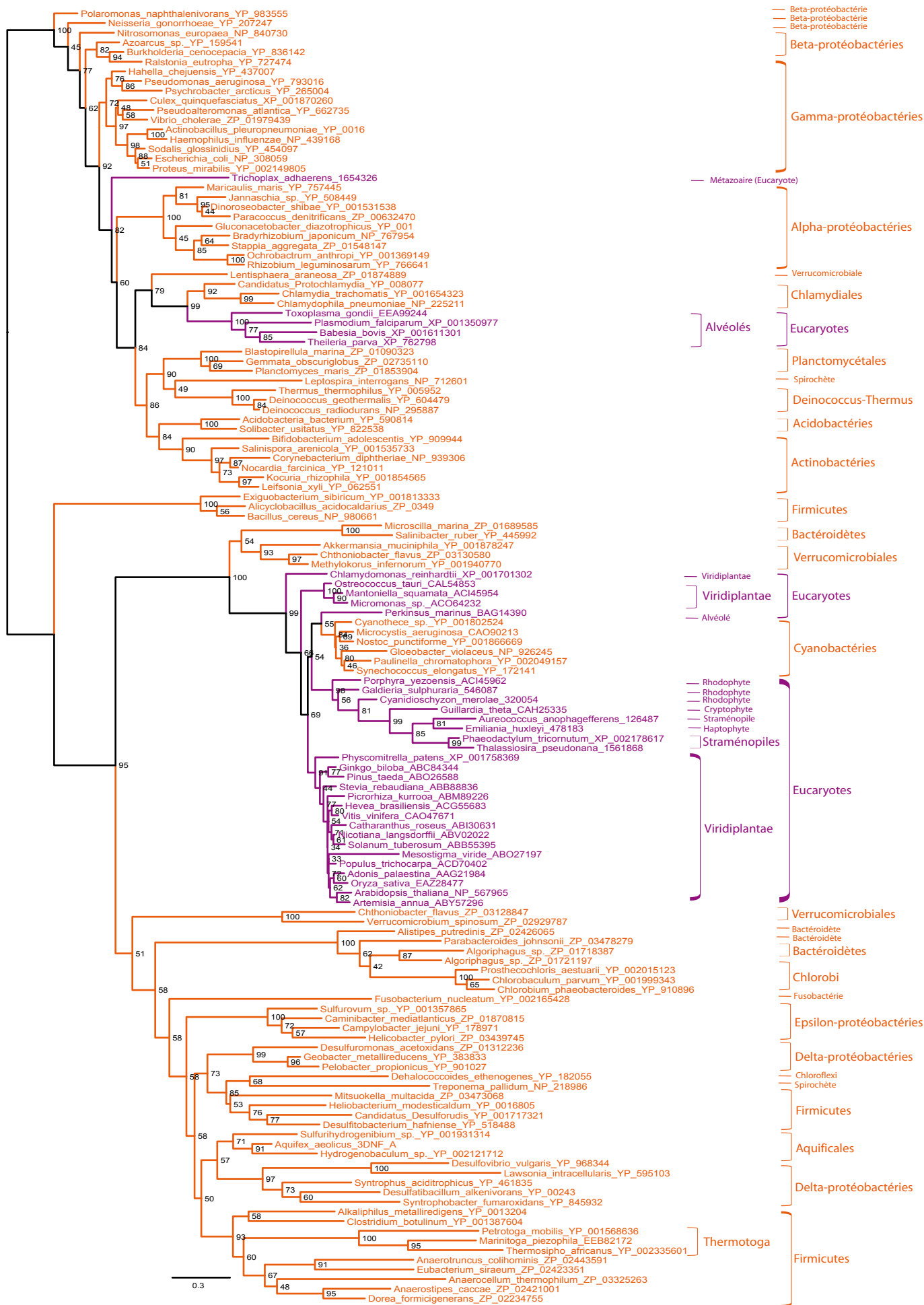


Figure 59. Phylogénie de HMDR.
 Reconstruction par maximum de vraisemblance utilisant 134 séquences représentatives et 192 positions conservées. Les nombres à proximité des noeuds indiquent les supports de bootstrap.

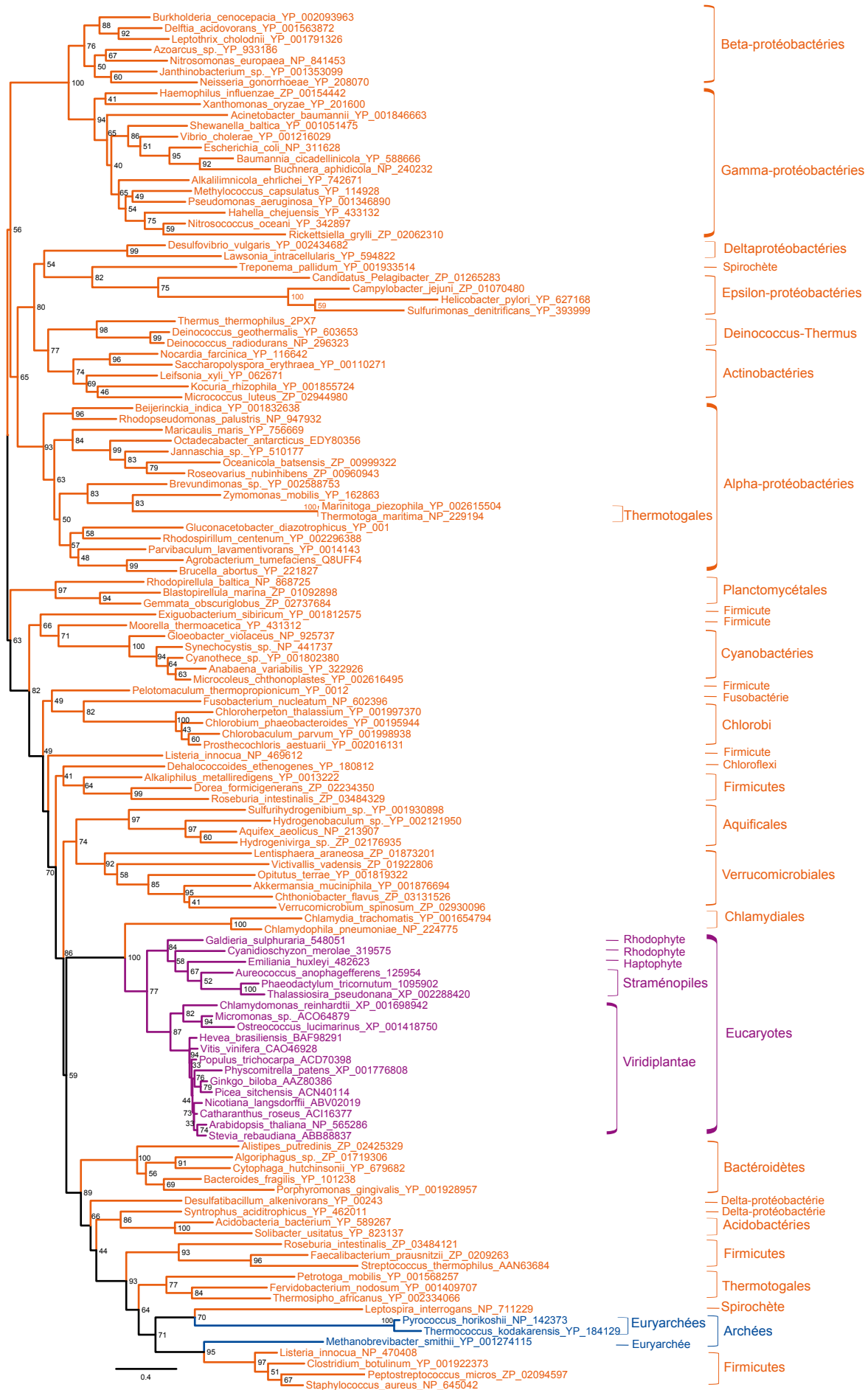


Figure 60. Phylogénie de CMT.

Reconstruction par maximum de vraisemblance utilisant 127 séquences représentatives et 147 positions conservées. Les nombres à proximité des noeuds indiquent les supports de bootstrap.



Figure 61. Phylogénie de MECPS.

Reconstruction par maximum de vraisemblance utilisant 146 séquences représentatives et 136 positions conservées. Les nombres à proximité des noeuds indiquent les supports de bootstrap.

Tableau 4. Récapitulatif des groupes bactériens les plus proches parents des séquences eucaryotes dans les arbres de chaque enzyme de la voie du MEP.

(*) Taxon bactérien proche parent des séquences eucaryotes de chaque enzyme par deux méthodes de reconstruction phylogénétique (maximum de vraisemblance et méthode bayésienne). (?) Taxon bactérien proche parent des séquences eucaryotes par une seule des méthodes de reconstruction phylogénétique.

	DXS	DXR	CMT	CMK	MECPS	HMDS	HMDR
Alpha-protéobactéries	*						
Cyanobactéries		*				*	*
Chlamydiales			*			*	
Planctomycétales				*	?		
Aquificales					?		
Bactéroidètes						*	
Chloroflexi						*	

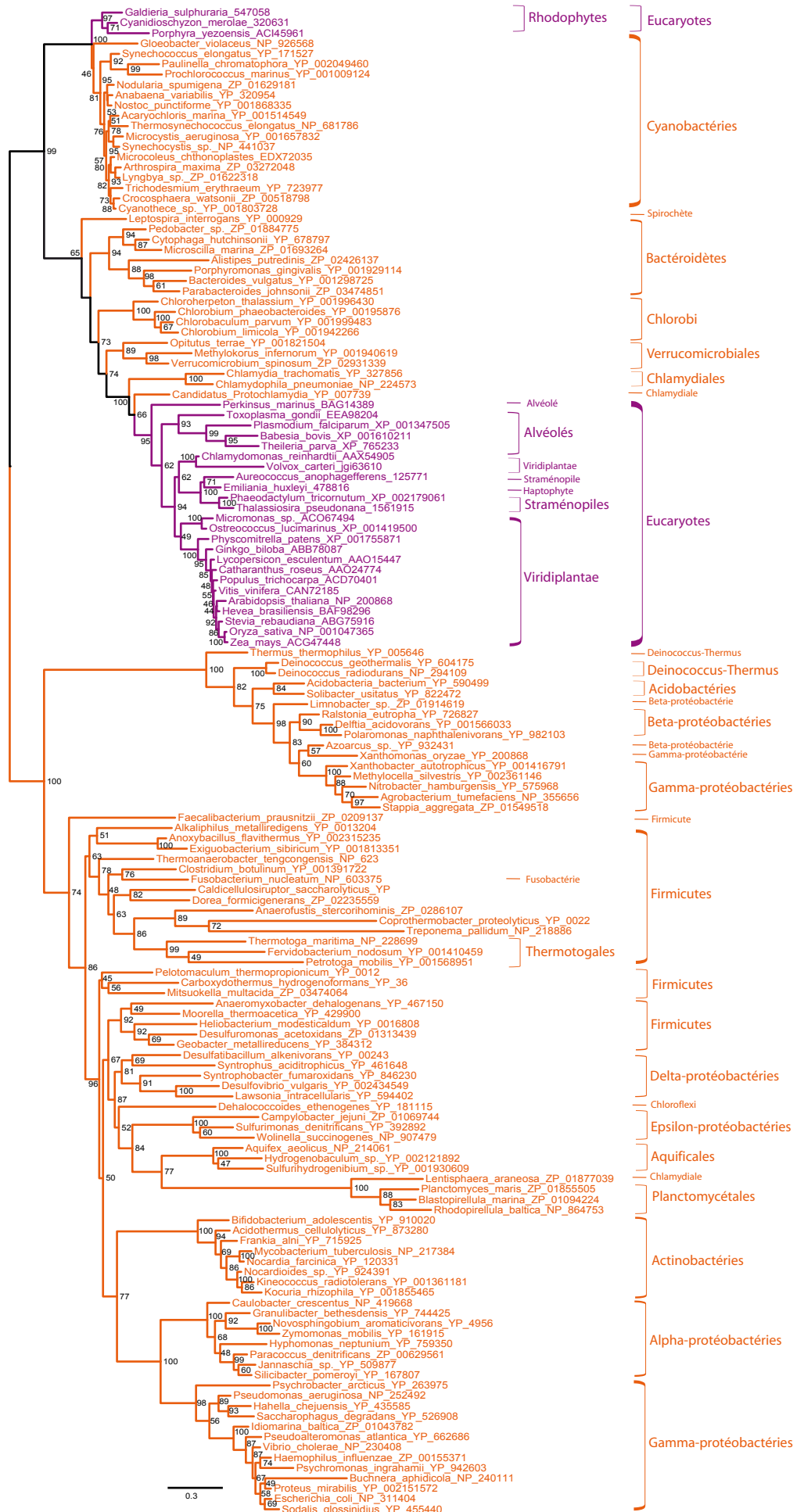


Figure 62. Phylogénie complète de HMDS.

Reconstruction par maximum de vraisemblance utilisant 146 séquences représentatives et 327 positions conservées. Les nombres à proximité des noeuds indiquent les supports de bootstrap.

les trois lignées d'eucaryotes photosynthétiques primaires (i. e. issus de l'endosymbiose primaire d'une cyanobactérie à l'intérieur d'un hôte eucaryote) : les glaucophytes, les viridiplantae (plantes et algues vertes), et les rhodophytes (algues rouges). On trouve aussi des séquences de chromalvéolés, que l'on sait avoir réalisé des endosymbioses secondaires avec des rhodophytes. L'ensemble de ces séquences HMDS eucaryotes émerge dans un grand groupe très bien soutenu qui contient les cyanobactéries, les chlamydiales-verrucomicrobiales, les chlorobi et des bactéroïdètes. Nous avons reconstruit un arbre phylogénétique plus ciblé sur ce groupe (Figure 63). Il montre une claire séparation entre un groupe formé par les cyanobactéries, les glaucophytes et les rhodophytes et le reste (algues et plantes vertes, chlamydiales-verrucomicrobiales, chlorobi, bactéroïdètes et chromalvéolés), que l'on appellera sous-groupe vert.

La séparation entre les deux sous-groupes est aussi soutenue par la présence d'une insertion d'environ 250 acides aminés toujours placée à la même position des séquences du sous-groupe vert mais n'existant ni dans le groupe cyanobactéries-rhodophytes-glaucophytes ni chez les autres bactéries qui constituent le groupe extérieur de l'arbre. Ceci est surprenant car les chromalvéolés, supposés être issus d'endosymbioses secondaires avec des algues rouges, se placent à l'intérieur du sous-groupe vert et possèdent l'insertion qui le caractérise. Les équipes qui se sont intéressées à la phylogénie de HMDS ont essayé d'expliquer comment un gène de la lignée verte se retrouve dans des organismes issus d'endosymbioses avec des algues rouges, ce qui donnerait des pistes sur l'évolution des chromalvéolés et des arguments dans le débat sur leur monophylie [Frommolt *et al.* 2008 ; Keeling 2009]. Etant partagé par la majorité des bactéries, l'état ancestral était très probablement l'absence de l'insertion. Les cyanobactéries, qui seraient à l'origine du HMDS des glaucophytes et des rhodophytes, n'ont d'ailleurs pas cette insertion. Cela semble donc être apparu dans le sous-groupe vert et transmis aux chromalvéolés par un mécanisme inconnu. Pour expliquer la présence de l'insertion chez certaines bactéries deux possibilités se présentent à nous :

1) l'insertion était un caractère ancestral de ce petit groupe de bactéries (chlamydiales, verrucomicrobiales, chlorobi et bactéroïdètes), qui l'auraient transféré uniquement vers la lignée verte des eucaryotes. Cependant, même si les chlamydiales et les verrucomicrobiales semblent faire partie du même super-phylum bactérien, les autres bactéries ayant l'insertion (chlorobi et bactéroïdètes) ne sont pas des proches parents des chlamydiales [Wagner et Horn 2006].

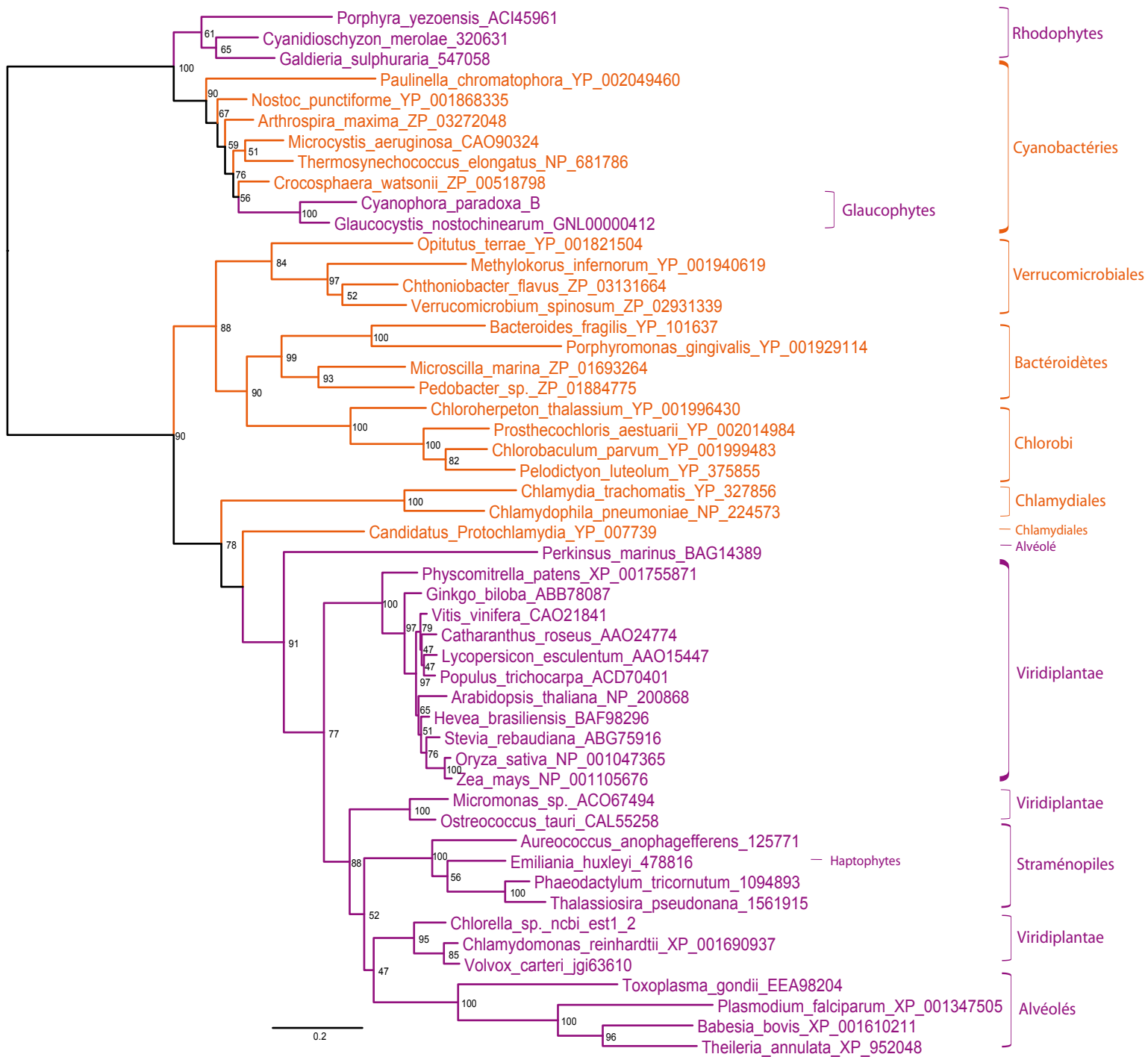


Figure 63. Phylogénie des HMDs eucaryotes et leurs homologues bactériens les plus proches.

Reconstruction par maximum de vraisemblance utilisant 50 séquences représentatives et 383 positions conservées. Les nombres à proximité des noeuds indiquent les supports de bootstrap.

2) le caractère est apparu et a été sélectionné dans la lignée verte, puis transmis vers les bactéries de la même façon que vers les eucaryotes issus d'endosymbioses secondaires rouges (les chromalvéolés). La position des séquences bactériennes à la base et non à l'intérieur de la lignée verte eucaryote peut s'expliquer par un artefact d'attraction de longues branches car ces séquences bactériennes sont très divergentes par rapport à la très forte conservation des séquences eucaryotes. Ceci est soutenu par le fait que le support statistique pour l'émergence basale de ces séquences bactériennes est plus faible dans la phylogénie reconstruite avec un sous-échantillon de séquences proches (Figure 63) que dans la phylogénie globale avec l'ensemble de groupes bactériens (Figure 62). En effet, il est bien connu que l'attraction de longues branches s'accroît quand on utilise un groupe extérieur plus éloigné [Felsenstein 1978]. Des transferts horizontaux de gènes participant dans d'autres processus cellulaires à partir des eucaryotes vers les bactéries [Wolf *et al.* 1999], et même des eucaryotes vers les chlamydiales [Stephens *et al.* 1998], ont aussi été proposés pour expliquer des observations similaires.

Le transfert horizontal de HMDS des eucaryotes vers les chlamydiales-verrucomicrobiales, bactériodètes et chlorobi semble donc être le scénario évolutif le plus plausible. Nous pouvons donc nous demander si la position de certains groupes de bactéries (notamment les chlamydiales) comme groupe frère des eucaryotes pour d'autres enzymes (Tableau 4) n'est pas tout simplement le résultat d'autres HGTs des eucaryotes vers ces bactéries, accompagnés d'une divergence très importante chez celles-ci.

Bien que l'interprétation de l'origine des séquences bactériennes chez les eucaryotes reste encore ouverte au débat, il semble à la fois cohérent avec ce que l'on sait de l'histoire évolutive de ces organismes et bien soutenu par une partie des phylogénies, que les cyanobactéries soient à l'origine de la voie du MEP chez les eucaryotes qui la possèdent. D'autres bactéries comme les chlamydiales ont pu adopter des enzymes de la voie des eucaryotes en remplacement de leurs propres enzymes. Il reste possible que d'autres bactéries aient aussi pu contribuer à la voie MEP des eucaryotes, comme c'est le cas notamment pour la DXS d'origine alpha-protéobactérienne à travers l'endosymbiose mitochondriale.

Origine de la voie du MEP

Nous avons vu que la présence des gènes de la voie du MEP chez les eucaryotes est le résultat de l'origine bactérienne de ces séquences, comme semble l'indiquer la distribution de cette voie principalement dans des organismes comportant des plastes. La présence d'une telle

voie ne peut donc pas être inférée chez le dernier ancêtre eucaryote et des homologues de ses gènes n'ont été détectés qu'à de très rares et partielles occasions chez les archées. Parallèlement, les gènes de cette voie sont très répandus dans les principaux taxons bactériens et les phylogénies sont cohérentes avec ces groupes. Par conséquent, tout indique que la voie du MEP est apparue dans la lignée bactérienne et qu'elle était déjà présente chez le dernier ancêtre commun de ce domaine, mais pas dans les ancêtres partagés avec les autres domaines.

5. Evolution de la synthèse d'isoprénoïdes à partir de leurs précurseurs

Une fois que les précurseurs des isoprénoïdes, l'IPP et le DMAPP, sont synthétisés par l'une des voies de synthèse décrites jusque-là, ces unités sont utilisées comme briques pour la synthèse d'une chaîne qui sera l'isoprénoïde à proprement parler. La première partie de cette synthèse est commune à tous les organismes et à toutes les synthèses d'isoprénoïdes et consiste à la condensation d'une chaîne linéaire à partir des unités de base. Ensuite, ces chaînes peuvent servir chez les archées comme composants des phospholipides membranaires, ou, d'une façon plus large chez des organismes de tous les domaines, comme des intermédiaires dont la modification permet la synthèse de molécules comme le cholestérol et des pigments (Figure 43, p. 123). Comme la première partie de cette synthèse est indispensable à la formation des molécules qui font partie des phospholipides archéens et que cette fonction existe dans des organismes des trois domaines du vivant, je vais étudier l'évolution des enzymes qui en sont responsables.

L'évolution des isoprényl diphosphate synthases dans la littérature

L'unité qui sert à amorcer la synthèse des isoprénoïdes est le DMAPP (5 carbones), sur lequel l'addition progressive d'unités d'IPP donnera du géranyl diphosphate (GPP, 10 carbones), du farnésyl diphosphate (FPP, 15 carbones), du géranylgéranyl diphosphate (GGPP, 20 carbones ; cet isoprénoïde est le plus fréquent dans les phospholipides archéens), du farnésylgéranyl diphosphate (FGPP, 25 carbones), etc [Kellogg et Poulter 1997]. Cette fonction est réalisée par des prényltransférases de la famille des isoprényl diphosphate synthases (IPPS). Bien que les IPPS se différencient par rapport à leur substrat préférentiel de départ (DMAPP, GPP, FPP) et la stéréochimie des doubles liaisons produites, le classement le plus fréquent des IPPS se fait par rapport à la taille du produit synthétisé. On sépare ainsi les IPPS qui produisent des molécules à chaîne courte (jusqu'à 20-25 carbones) de celles qui produisent des chaînes plus longues. Toutes les IPPS sont homologues et partagent leur

mécanisme réactionnel ; les enzymes qui synthétisent des chaînes courtes sont des homodimères dont les deux monomères portent des sites actifs qui travaillent indépendamment l'un de l'autre. Ces enzymes diffèrent par la profondeur de la poche hydrophobe qui attache le substrat (les poches les plus petites n'étant capables de produire que des chaînes plus courtes et vice-versa) [Kellogg et Poulter 1997] mais il a été montré que des mutations ponctuelles peuvent modifier grandement la taille de la poche hydrophobe, ce qui se répercute dans la taille des chaînes produites [Ohnuma *et al.* 1996 ; Tarshis *et al.* 1996 ; Ohnuma *et al.* 1997 ; Ohnuma *et al.* 1998]. Il a été montré que cette propriété a été exploitée dans la nature par des organismes [Tachibana *et al.* 2000], ce qui rend difficile de déterminer précisément le produit d'une enzyme à chaîne courte en connaissant uniquement sa place dans un arbre phylogénétique [Chen *et al.* 1994 ; Tachibana *et al.* 2000].

Les premiers arbres phylogénétiques des IPPS ne concernaient qu'une douzaine de séquences d'enzymes à chaîne courte [Chen *et al.* 1994]. Elles montraient une séparation entre les séquences procaryotes et eucaryotes et il a été proposé que les séquences archéennes soient plus anciennes étant donné qu'elles sont capables de produire une gamme un peu plus large d'isoprénoïdes. Des analyses plus récentes contenant des enzymes à chaîne longue en plus de celles à chaîne courte permettent de voir que les séquences forment deux groupes, l'un correspondant aux enzymes à chaîne courte et l'autre correspondant aux enzymes à chaîne longue [Tachibana *et al.* 2000 ; Boucher *et al.* 2004]. Cette différence est probablement due au fait que les enzymes à chaîne plus longue ont souvent besoin de protéines supplémentaires pour stabiliser le substrat hydrophobe [Kellogg et Poulter 1997], ce qui doit se traduire par une plus grande divergence entre les deux types d'enzymes qu'à l'intérieur des groupes. La phylogénie publiée dans Tachibana *et al.* 2000 est intéressante car elle montre qu'au sein du groupe d'enzymes à chaîne courte on peut observer trois clades indépendants de la fonction des enzymes et correspondant aux trois domaines du vivant (à l'exception des plantes, qui ont aussi une copie de type bactérien probablement d'origine plastidiale). Ceci suggère la présence d'une enzyme de ce type chez les ancêtres communs respectifs des trois domaines, et par conséquent, aussi chez le cenancêtre. En outre, cette phylogénie ne place pas les archées à la base des enzymes à chaîne courte, mais les fonctions étant mélangées dans l'arbre, cela semble indiquer la capacité du gène ancestral à changer facilement de fonction par des mutations ponctuelles comme celles décrites dans la littérature. En ce qui concerne les enzymes à chaîne longue, cette analyse n'avait pas détecté d'homologues chez les archées, mais d'autres plus récentes rapportent de telles séquences dans ce domaine aussi [Boucher *et*

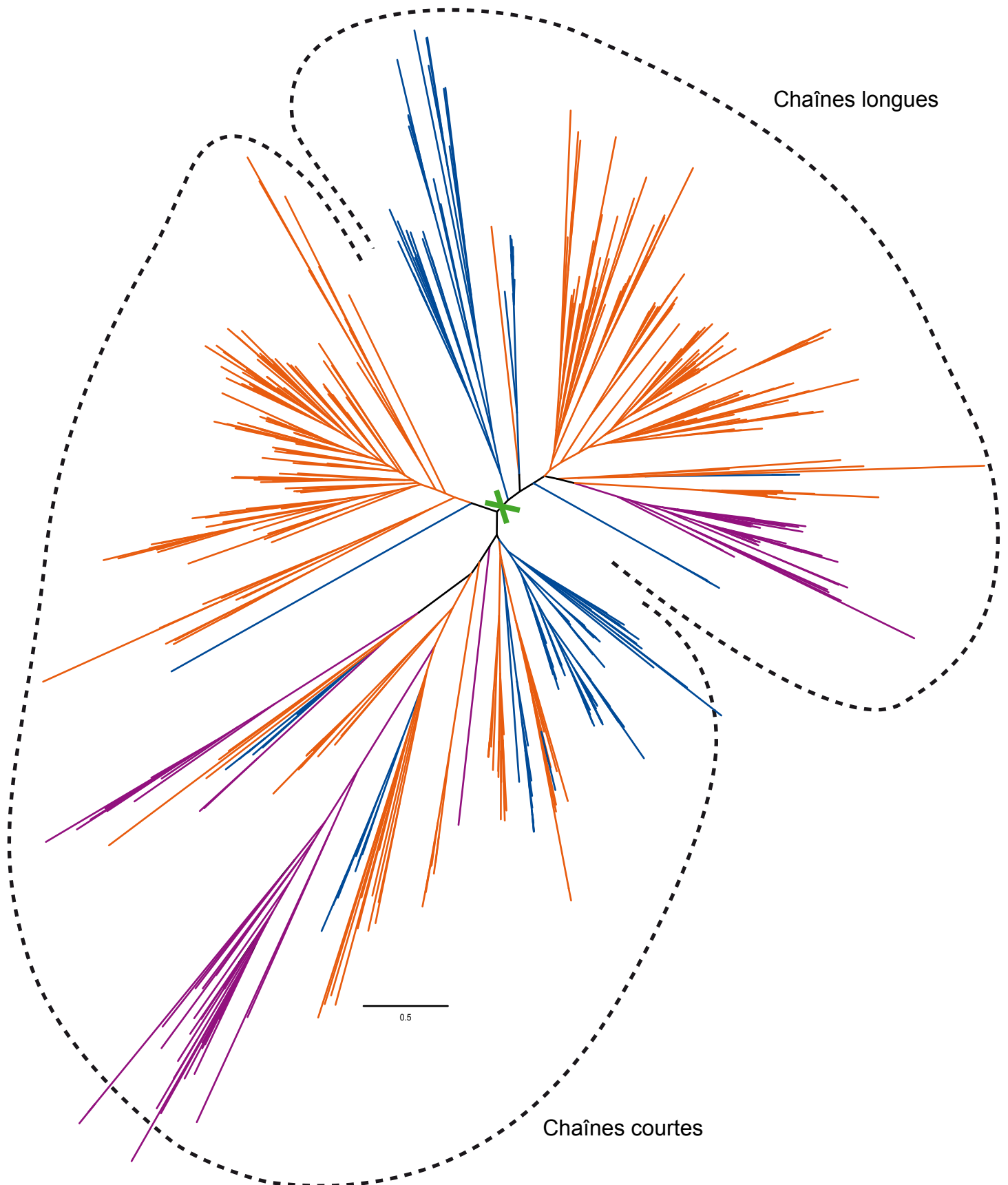


Figure 64. Phylogénie globale des IPPS.

Reconstruction avec FastTree (approximation de maximum de vraisemblance) utilisant 537 séquences représentatives et 216 positions conservées. Les séquences bactériennes apparaissent en orange, les archéennes en bleu et les eucaryotes en violet. La croix verte indique la séparation entre les protéines majoritairement associées à l'utilisation de substrats à chaîne courte et celles utilisant principalement les substrats à chaîne longue.

al. 2004]. Dans cette autre étude les séquences archéennes des enzymes à chaîne longue (ou au moins supposées comme telles à cause de la comparaison de leur site catalytique) ne branchent pas parmi les autres domaines mais sont paraphylétiques et probablement soumises à un effet d'attraction des longues branches, ce qui rend difficile l'établissement de leur origine.

Mise à jour de la phylogénie des isoprenyl diphosphate synthases

Aujourd'hui nous avons accès à beaucoup plus de séquences de génomes complets, ce qui permet la mise à jour de ces analyses avec un échantillon plus large et plus représentatif de la diversité des organismes. La reconstruction phylogénétique des séquences homologues des IPPS confirme la séparation des séquences en deux groupes à peu près équivalents qui correspondent respectivement aux enzymes à courte et à longue chaîne (Figure 64). Les deux types d'enzymes semblent très répandus dans les trois domaines du vivant et les séquences se regroupent selon les grands groupes taxonomiques dans cet arbre malgré quelques séquences très divergentes et certains HGTs très probables. Parmi les enzymes à chaîne courte on observe un groupe représentatif de la diversité archéenne, un grand groupe de bactéries et un groupe de séquences plus divergentes parmi lesquelles branchent les séquences eucaryotes. Dans le groupe des enzymes à chaîne longue on observe plusieurs groupes archéens à la base et un grand clade bactérien au sein duquel branchent aussi les eucaryotes. La séparation nette entre les séquences archéennes et bactériennes pour les deux types d'enzymes et l'observation des principaux groupes taxonomiques dans les deux cas soutient fortement la présence d'un gène de chaque type dans leurs ancêtres respectifs, et donc probablement aussi chez le cenancêtre. Cependant, ces résultats posent aussi la question de l'enracinement des eucaryotes (au sein des bactéries ou comme un clade indépendant) et, dans le cas des enzymes à produit long, questionnent la monophylie et donc l'origine des séquences archéennes. Afin d'approfondir ces questions, des reconstructions phylogénétiques de chaque type d'enzyme ont été réalisées ([Lombard *et al.* 2012b], voir Annexe 7).

La phylogénie des enzymes à chaîne courte montre un clade très robuste d'archées avec une phylogénie en accord avec les principaux taxons archéens, ce qui soutient la présence d'une de ces enzymes chez l'ancêtre de ce domaine (Figure 65). A la base du groupe de bactéries on trouve un groupe de séquences de bacteroidètes-chlorobi, de chloroflexi, de thermoprotéales (crenarchées), et quelques autres séquences divergentes. D'une part on pourrait penser que ce groupe est uniquement le résultat d'un artéfact d'attraction de longues

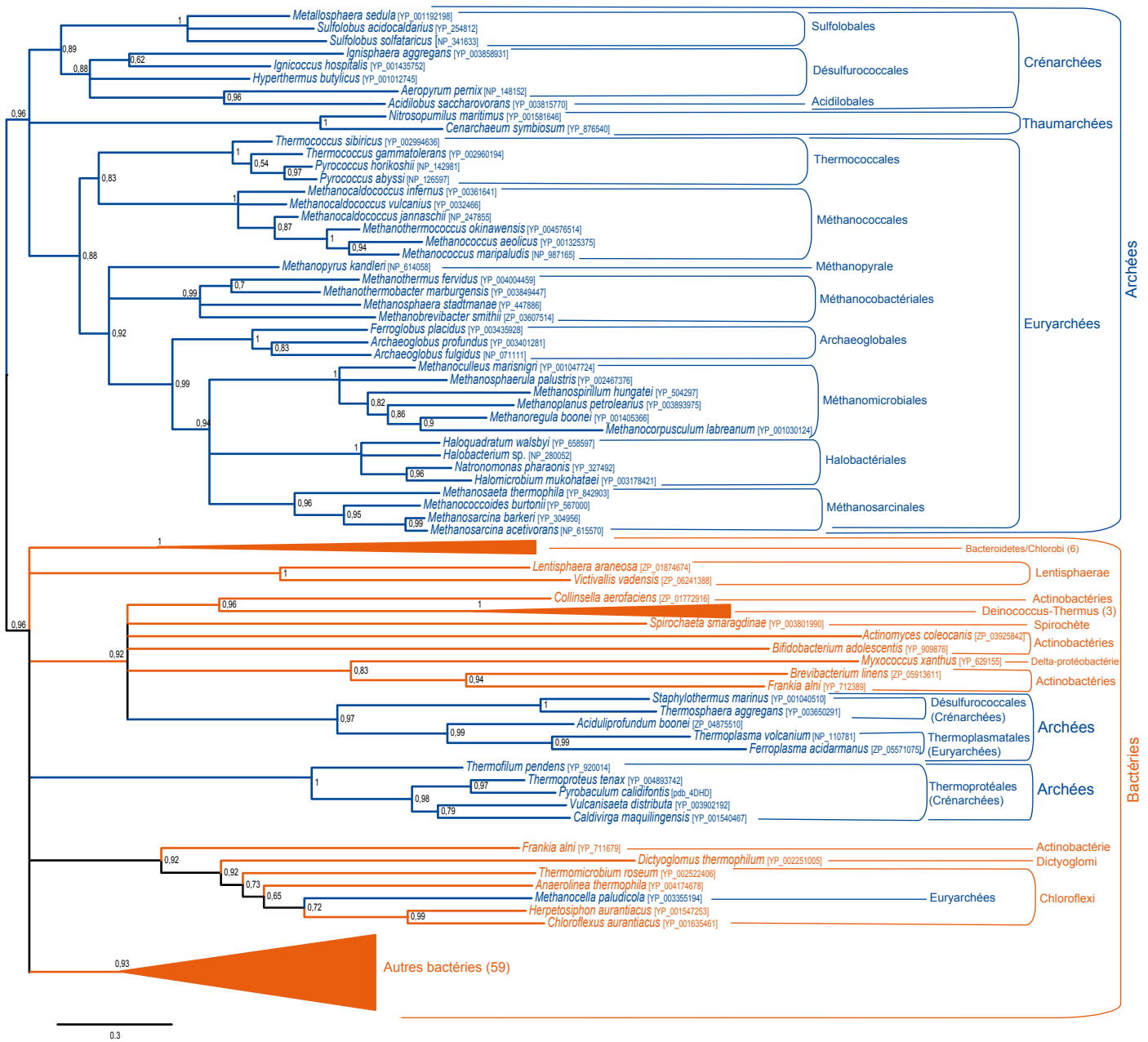


Figure 65. Phylogénie des IPPS à chaîne courte.

Reconstruction avec FastTree utilisant 136 séquences représentatives et 244 positions conservées. Les nombres à proximité des noeuds indiquent les valeurs de robustesse calculées par FastTree.

branches, mais ceci semble peu probable car il existe dans l'arbre des branches plus longues (notamment à proximité des eucaryotes). D'un autre côté, les groupes de thermoprotéales, bactéroïdètes-chlorobi et chloroflexi sont représentatifs de la diversité de ces groupes, dont on ne trouve pas d'autres séquences ailleurs dans l'arbre. On peut donc proposer que dans ce cas les thermoprotéales aient reçu par HGT une copie d'origine bactérienne de cette enzyme et que cette enzyme ait remplacé la version originale archéenne dans cet ordre archéen. Bien que l'arbre soit conforme à la plupart des grands taxons bactériens, ce qui soutient l'origine ancestrale de cette enzyme chez les bactéries, d'autres HGTs ponctuels sont possibles. La position des séquences eucaryotes est plus intrigante. On détecte des groupes remarquables d'eucaryotes à trois endroits différents dans l'arbre (ce qui d'ailleurs est cohérent avec l'idée que ces organismes possèdent un grand nombre de copies d'IPPS) [Kellogg et Poulter 1997]. L'un d'entre eux ne contient que des organismes de la lignée verte et groupe avec les cyanobactéries, ce qui suggère l'origine plastidiale de ces séquences. Un autre groupe contient des séquences un peu plus diverses et associées à des séquences très divergentes comme celles des spirochètes (parasites) et des copies probablement xénologues et très divergentes d'haloarchées. Cette topologie ressemble à un artefact dû à la divergence des séquences et ne permet donc pas de déterminer leur origine. Le troisième groupe eucaryote est beaucoup plus représentatif de la diversité de ce groupe et tout semble indiquer que cette séquence ait été héritée de l'ancêtre des eucaryotes. En revanche, le contexte de ce groupe est étonnant car on y trouve des séquences divergentes de bactéries (des thermotogales, un planctomycètes et une delta-protéobactérie) et d'archées (thermoplasmatales, desulfurococcales et *Aciduliprofundum boonei*). Les HGTs endosymbiotiques classiques ne peuvent donc pas être invoqués mais ce groupe n'apparaît pas non plus comme basal, aussi l'origine de ces séquences chez les eucaryotes reste non résolue malgré l'inférence possible de la présence de ce gène chez le dernier ancêtre eucaryote.

La phylogénie des enzymes à chaîne longue révèle un groupe bien soutenu d'archées dont les principaux groupes taxonomiques sont respectés mais les euryarchées apparaissent comme paraphylétiques (Figure 66). Le fort soutien du groupe justifie cependant de l'ancestralité de ce gène chez les archées et plaide pour un artefact de reconstruction pour expliquer la paraphylie des euryarchées. Ces séquences sont très répandues parmi les bactéries et elles forment des groupes qui respectent la taxonomie bactérienne, ce qui soutient la présence ancestrale de ce gène dans ce domaine. La situation des eucaryotes est dans ce cas à nouveau étonnante : alors que les algues vertes et rouges branchent parmi les cyanobactéries,



Figure 66. Phylogénie des IPPS à chaîne longue.

Reconstruction avec FastTree utilisant 218 séquences représentatives et 241 positions conservées. Les nombres à proximité des noeuds indiquent les valeurs de robustesse calculées par FastTree.

suggérant une origine plastidiale de ces séquences, le groupe principal et représentatif de la diversité des eucaryotes (y compris des eucaryotes qui n'ont pas de plastes) apparaît comme groupe frère des cyanobactéries de façon assez soutenue. Ceci suggère que la copie ancestrale de ce gène chez les eucaryotes pourrait être d'origine cyanobactérienne mais *a fortiori* non plastidiale, ou alors que l'emplacement de ces séquences est le produit d'un artefact.

6. Conclusions

Une voie de synthèse des précurseurs caractéristique pour chaque domaine du vivant

Comme nous avons pu le constater, l'histoire évolutive des voies de biosynthèse des isoprénoïdes est plus compliquée de ce que l'on pourrait penser au premier abord. Nos analyses montrent que, contrairement à l'idée généralement admise [Lange *et al.* 2000 ; Kuzuyama 2002], il existe trois voies de synthèse des précurseurs des isoprénoïdes très répandues à l'échelle de l'arbre du vivant. C'est le cas de la voie du MEP principalement retrouvé chez les bactéries, de la voie du MVA traditionnelle chez les eucaryotes ou de la voie alternative du MVA chez les archées. Les deux voies du MVA partagent un certain nombre d'enzymes mais d'autres les différencient (IPK et IDI2 dans la voie archéenne et PMK, MDC et IDI 1 dans la voie eucaryote). La voie du MEP est très répandue chez les bactéries et les phylogénies de ses enzymes sont cohérentes avec leur présence chez le dernier ancêtre commun des bactéries. La même chose peut être dite pour les voies respectives du MVA des eucaryotes et des archées. Nous passons ainsi d'une vision à deux voies pour trois domaines vers une vision à trois voies pour trois domaines, ce qui complique au premier regard les tentatives de déduction des caractéristiques métaboliques du cenancêtre.

Il faut également signaler que ces gènes se retrouvent aussi en dehors du domaine dont ils sont caractéristiques à la suite de HGT. Le cas le plus remarquable est la présence de la voie du MEP de type bactérien chez un grand nombre d'eucaryotes. Celle-ci serait issue soit d'enzymes apportées par l'endosymbionte cyanobactérien (et éventuellement par l'alpha-protéobactérie à l'origine de la mitochondrie), soit à partir d'une mosaïque de bactéries, la première option semblant être la plus plausible. Le transfert d'une bonne partie de la voie eucaryote du MVA depuis des bactéries vers les halobactériales et les thaumarchées semble probable, ainsi qu'un transfert vers les Sulfolobales bien qu'il soit moins soutenu.

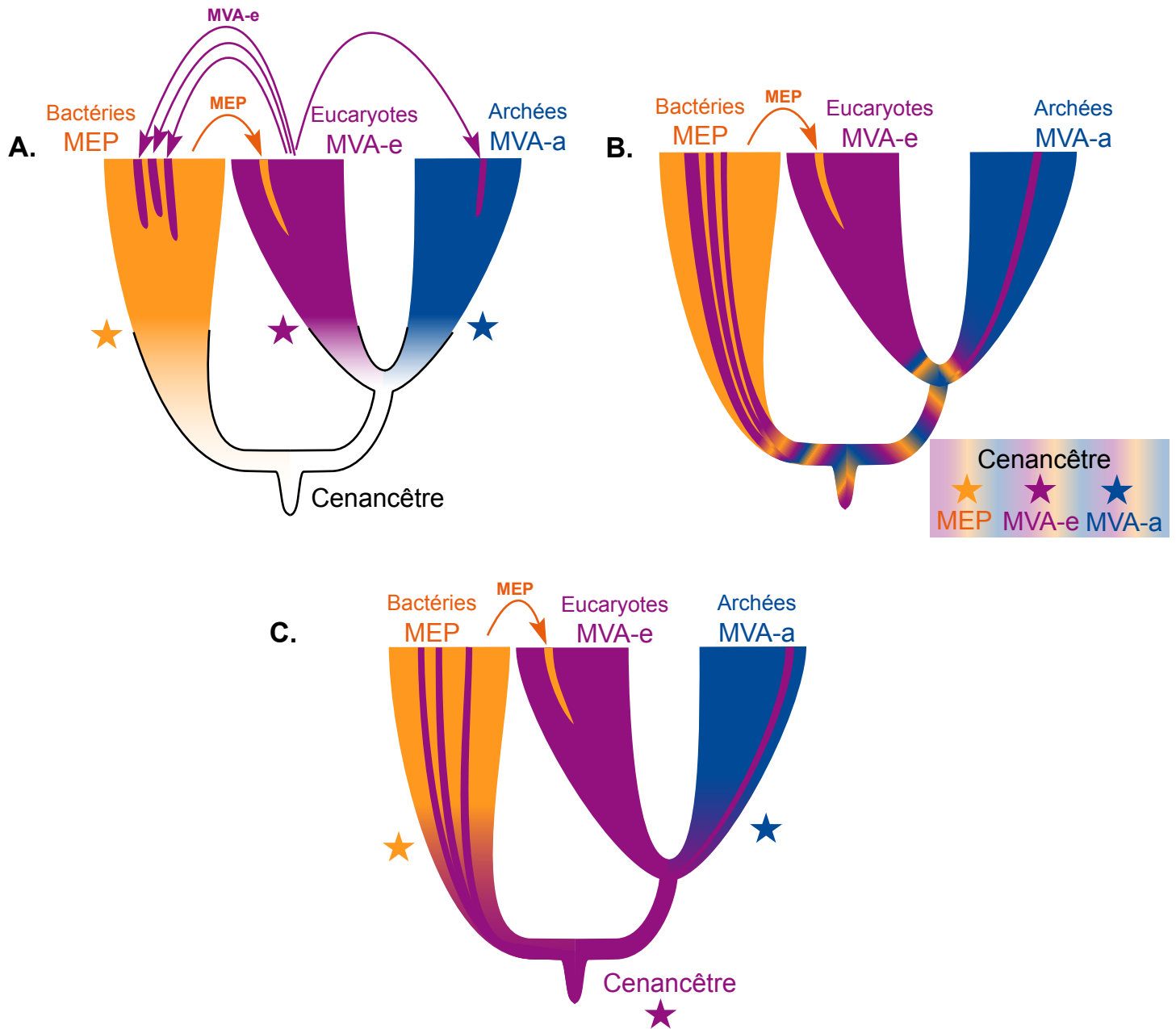


Figure 67. Scénarios évolutifs sur l'origine de la biosynthèse des précurseurs des isoprénoïdes.

Les arbres schématisent les relations entre les trois domaines du vivant et le cenancêtre. L'orange représente la voie du MEP, le violet représente la voie du MVA caractéristique des eucaryotes et le bleu représente la voie du MVA caractéristique des archées. La voie du MEP chez les eucaryotes est toujours représentée par des HGT depuis les bactéries car tous les indices s'accordent sur cette hypothèse comme étant la plus vraisemblable.

A. Le cenancêtre ne possédait aucune synthèse des précurseurs des isoprénoïdes et chaque voie est apparue indépendamment dans le domaine respectif dont elle est caractéristique. Des transferts ont eu lieu plus tard. **B.** Les trois voies étaient présentes chez le cenancêtre et chaque domaine en a conservé une. Les autres voies auraient été perdues à l'exception de quelques rares taxons bactériens et archéens qui auraient conservé la voie du MVA de type eucaryote. **C.** Une voie ancienne comparable à celle du MVA de type eucaryote était présente chez le cenancêtre. Cette voie aurait été remplacée presque complètement chez les bactéries et les archées.

Possibilité d'une voie de synthèse des précurseurs des isoprénoïdes chez le cenancêtre

L'attribution d'une voie principale de synthèse des précurseurs des isoprénoïdes à chaque domaine du vivant permet d'inférer sa présence chez le dernier ancêtre commun respectif de chaque domaine sans nous éclairer sur le métabolisme du cenancêtre. Les deux possibilités que l'on peut proposer au premier abord sont : i) l'absence de métabolisme des isoprénoïdes chez le cenancêtre ; ou ii) la présence de toutes les voies chez le cenancêtre et une spécialisation de chaque voie dans chaque domaine du vivant (Figure 67). La première possibilité rejoindrait les thèses d'une absence de membrane chez le cenancêtre basées sur les manques de voies de synthèses des composantes [Koga *et al.* 1998 ; Martin et Russell 2003], mais nous avons vu les critiques qui peuvent être faites contre ces modèles. D'un autre côté, la deuxième possibilité n'est pas soutenue par les données, notamment lorsqu'on observe la distribution taxonomique des voies du MEP et du MVA chez les archées.

Une troisième possibilité peut être envisagée à partir de nos résultats. Comme nous l'avons vu, une voie MVA de type eucaryote est présente chez un nombre relativement large de bactéries, qui forment un groupe monophylétique séparé des eucaryotes dans les analyses phylogénétiques. Bien que la possibilité d'un transfert très ancien à partir d'un donneur eucaryote, suivi d'une grande divergence chez les bactéries, ne soit pas complètement à écarter, ces phylogénies sont en accord avec l'hypothèse de la présence de la voie eucaryote du MVA chez l'ancêtre des bactéries ou même chez un ancêtre antérieur. Par ailleurs, il existe des archées (des sulfobactéries) qui possèdent la totalité de la voie eucaryote du MVA à l'exception d'IDI1, et dont les séquences ne se placent ni au sein des eucaryotes ni au sein des bactéries. Même s'il s'agit d'une distribution taxonomique très réduite, elle pourrait soutenir le caractère ancestral de la voie eucaryote du MVA chez les archées. Tous ces arguments sont en faveur de la présence de la voie eucaryote du MVA chez les ancêtres des trois domaines du vivant, et donc peut-être présente chez le cenancêtre. Dans ce scénario, les autres voies seraient apparues plus tard dans l'évolution, et auraient massivement remplacé la voie du MVA originelle. Afin d'expliquer la divergence de la voie archéenne, on peut proposer que les GHMP kinases qui participent dans la fin de la voie du MVA chez les eucaryotes n'avaient pas totalement acquis leur spécificité de substrat à l'époque du cenancêtre, laissant la porte ouverte au recrutement d'enzymes non-homologues, ce qui fut le cas chez les archées. Cette possibilité est confortée, bien qu'indirectement, par le remplacement récent avéré d'une des ces enzymes chez certains eucaryotes [Houten et Waterham 2001]. Dans le cas des bactéries, la présence simultanée chez leur dernier ancêtre d'une voie du MVA et une voie du

MEP rend plausible la perte massive de l'une des deux voies redondantes, en l'occurrence celle du MVA.

En résumé, il est possible d'attribuer une voie de synthèse des précurseurs des isoprénoïdes aux ancêtres de chaque domaine du vivant et de tracer l'histoire de certains transferts horizontaux de façon relativement fiable ([Lombard *et al.* 2011b], voir Annexe 2). La distribution taxonomique des gènes de la voie eucaryote du MVA, les phylogénies et l'existence ponctuelle de synapomorphies caractéristiques des bactéries indiquent que, contrairement aux hypothèses précédentes, la voie de synthèse des précurseurs des isoprénoïdes qui existe aujourd'hui chez les eucaryotes semble avoir été présente chez le cenancêtre. La voie présente chez le cenancêtre comprenait probablement certaines enzymes moins spécifiques que les enzymes modernes, mais elle fournit un mécanisme enzymatique de synthèse des isoprénoïdes probable chez ce dernier qui s'ajouterait à celui des énantiomères du glycérol phosphate décrit dans l'introduction [Peretó *et al.* 2004].

Des isoprényl diphosphate synthases chez le cenancêtre

Les analyses phylogénomiques montrent que les deux sous-familles des IPPS (à courte et à longue chaîne) sont répandues dans les trois domaines du vivant et que l'évolution de ces gènes est cohérente avec la présence d'une copie de chaque type d'IPPS chez les ancêtres respectifs des trois domaines du vivant. Il semble aujourd'hui difficile de déterminer l'origine des séquences eucaryotes, qui pourrait être bactérienne ou le résultat d'un héritage directement du cenancêtre si l'on accepte que leur position dans les arbres soit le résultat d'artéfacts. Quoiqu'il en soit, la présence de ces gènes chez les ancêtres respectifs des archées et des bactéries soutient l'existence chez le cenancêtre de deux enzymes de cette famille, probablement déjà avec une affinité respectivement pour les chaînes courtes ou longues ([Lombard *et al.* 2012b], voir Annexe 7). Ce résultat est important non seulement parce qu'il suggère directement la présence de chaînes d'isoprénoïdes chez le cenancêtre, mais aussi parce qu'il soutient indirectement la présence d'IPP et DMAPP dans cet organisme.

	Phospholipides	Acides gras	Alcools gras	Synthèse endogène	Synthase	Intéraction protéines (acylation?)
Halobacterales	Gattinger 2002	Kates 1968		Corcelli 1996	Kates 1968 Pugh 1971 Pugh 1994	Pugh 1994 Colella 1998 Kolbe 2000
Methanosarcinales	Gattinger 2002	Tornabene 1979				
Methanomicrobiales		Tornabene 1979				
Methanococcales	Gattinger 2002	Tornabene 1979				
Methanobacterales	Gattinger 2002	Tornabene 1978 Tornabene 1979 Jones 1991			Pugh 1994	Pugh 1994
Methanopyrales	Gattinger 2002					
Thermococcales		Carballeira 1997	Nishihara 2000	Carballeira 1997 Nishihara 2000		
Sulfolobales		Langworthy 1974				

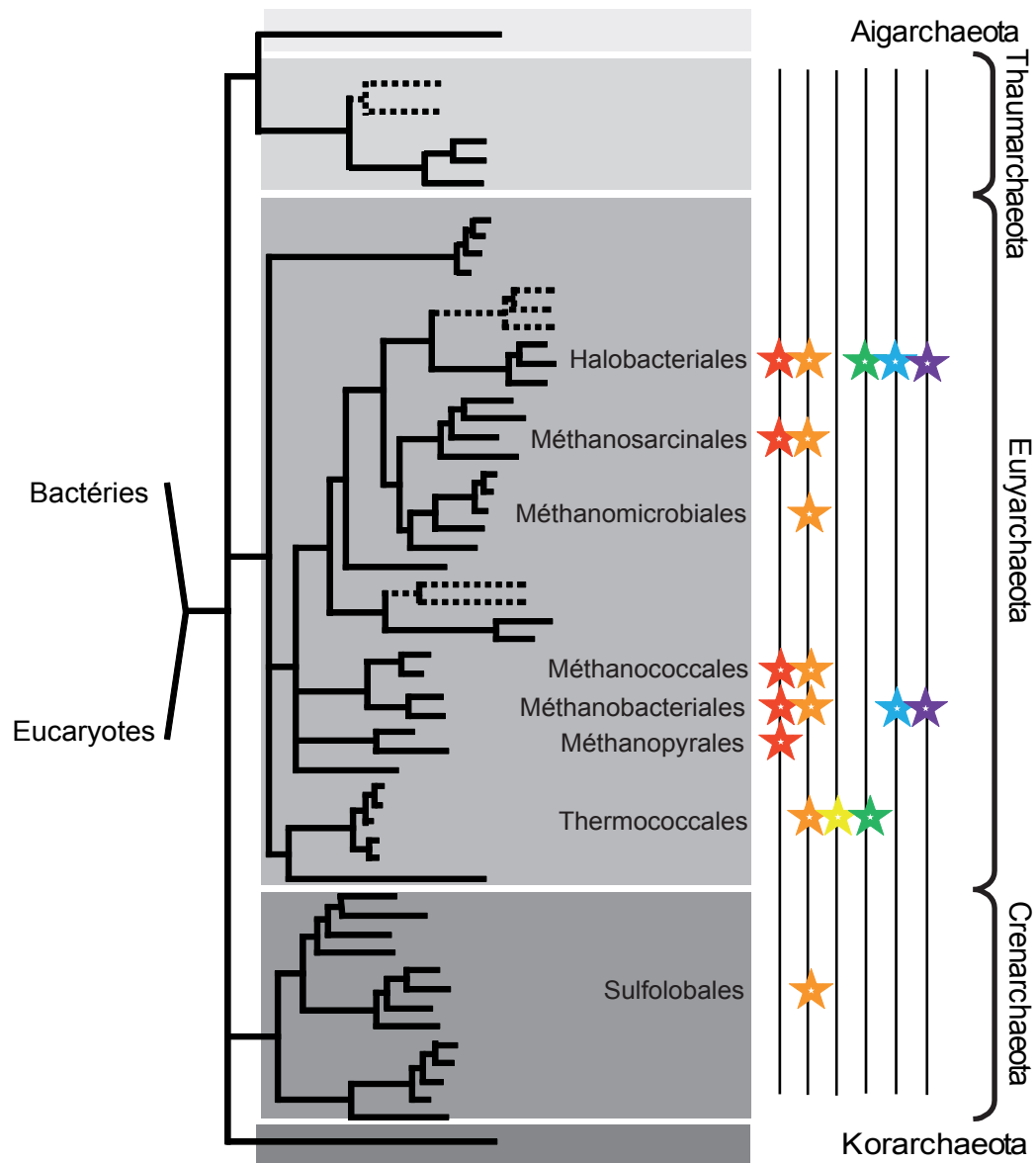


Figure 68. Descriptions dans la bibliographie d'éléments liés aux acides gras chez les archées.

Les colonnes du tableau présentent quelques molécules liées aux acides gras ou fonctions faisant intervenir des acides gras. Les lignes du tableau correspondent à des taxons archéens dont certaines espèces possèdent les caractéristiques des colonnes et les cases du tableau signalent les travaux dans lesquelles ces observations ont été faites. L'arbre schématise une phylogénie récente des archées (adaptée de Brochier *et al.* 2011) et les étoiles correspondent aux informations présentes dans le tableau.

Deuxième partie : Synthèse d'acides gras chez les archées et implications pour le cenancêtre.

1. Contexte

Des acides gras chez les archées

Malgré la grande diversité de phospholipides qui existe chez les bactéries et les eucaryotes, la découverte chez les archées de phospholipides contenant des isoprénoïdes à la place des acides gras provoqua beaucoup d'étonnement parce qu'elle faisait intervenir un type de phospholipide jamais vu auparavant [Kates *et al.* 1965]. De ce fait, on comprend que les principaux travaux sur les lipides des archées aient été focalisés sur les phospholipides, et que la détection de petites quantités d'acides gras dans les membranes de certaines archées [Kates *et al.* 1968 ; Langworthy *et al.* 1974 ; Tornabene *et al.* 1978 ; Tornabene et Langworthy 1979] (Figure 68) ait été éclipsée. Cette focalisation fait qu'actuellement c'est plutôt la découverte d'acides gras chez les archées qui étonne [Jones et Holzer 1991 ; Carballeira *et al.* 1997 ; Iverson *et al.* 2012]. Pourtant, les archées contiennent non seulement des acides gras mais aussi leurs dérivés, comme par exemple les alcools gras dont on sait maintenant qu'ils constituent la moitié des lipides neutres (sans partie hydrophile) de la membrane de *Pyrococcus furiosus* [Nishihara *et al.* 2000]. Des quantités variables d'acides gras ont même été détectées dans des phospholipides⁴⁴ d'une assez large diversité d'euryarchées [Gattinger *et al.* 2002]. En outre, il ne faut pas oublier que chez les bactéries et les eucaryotes les acides gras ne sont pas utilisés uniquement dans les phospholipides, mais qu'ils jouent aussi d'autres rôles importants par exemple dans le stockage d'énergie ou la modification post-traductionnelle de protéines [Schweizer et Hofmann 2004 ; Berg *et al.* 2007b]. Chez les archées, il a été proposé que les acides gras pouvaient servir à l'acylation de protéines [Pugh et Kates 1994 ; Eichler et Adams 2005] et on sait que certaines protéines comme l'halorhodopsine ont besoin d'interagir avec des acides gras libres pour adopter leur conformation active [Corcelli *et al.* 1996 ; Colella *et al.* 1998 ; Kolbe *et al.* 2000]. Bien qu'au départ il y ait eu hésitation sur l'origine de ces acides gras, des travaux ont confirmé leur

⁴⁴ Il s'agit ici principalement de phospholipides dont le glycérol phosphate ne constituerait pas l'ossature, par exemple des aminolipides comme les sphingolipides.

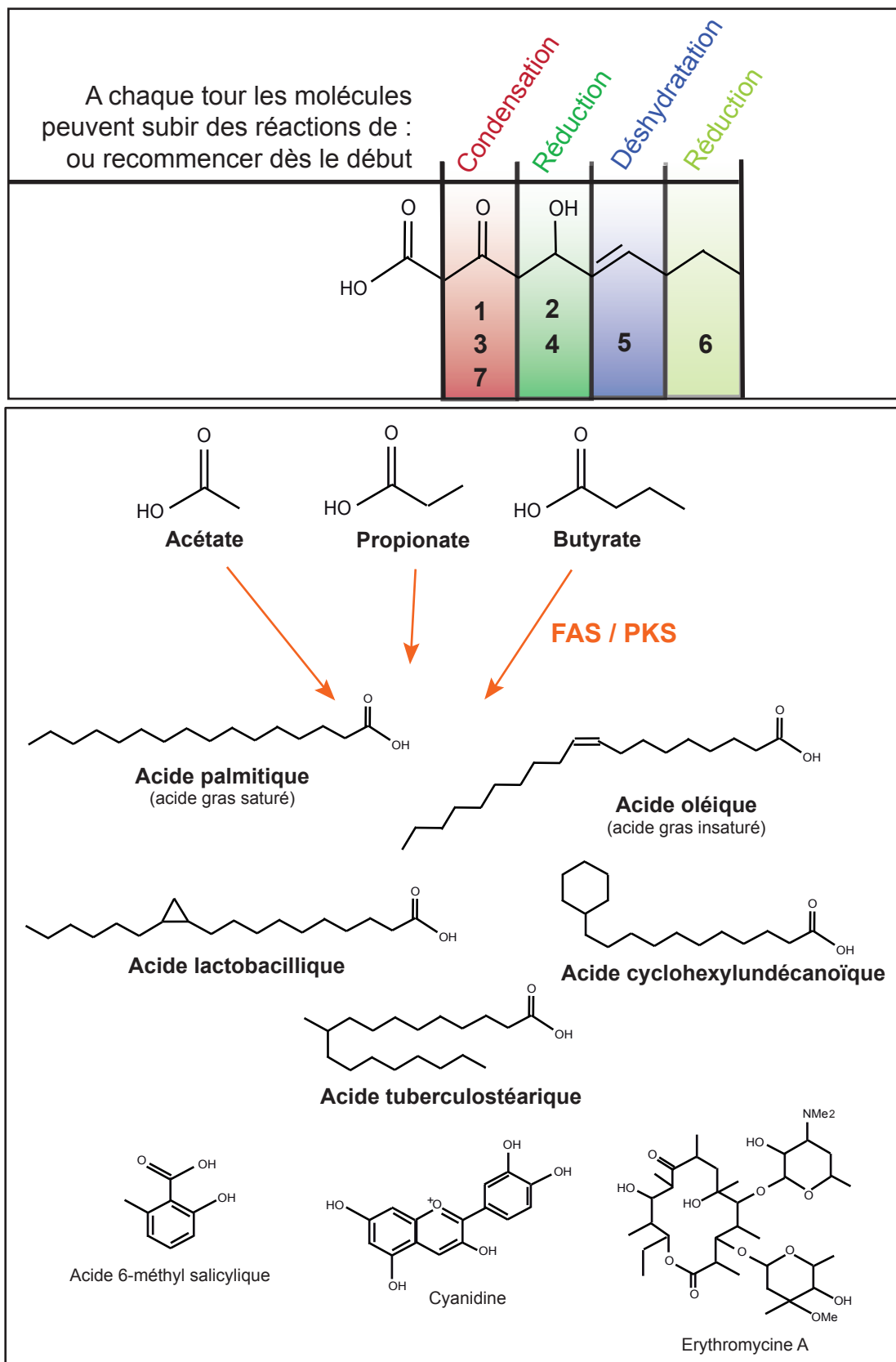


Figure 69. Les polycétides.

Les polycétides sont des composés synthétisés à partir d'unités carboxyliques comme l'acétate, le propionate ou le butyrate. Les réactions de condensation peuvent être intercalées avec des réactions de réduction et déshydratation successives, dans un ordre indiqué dans la partie supérieure de la figure. Les chiffres en dessous de la molécule proposent une série hypothétique de 7 étapes conduisant à la synthèse d'un polycétide. Ces réactions sont réalisées par des machineries appelées PKS (synthèses de polycétides) ou FAS (synthèses d'acides gras). La partie inférieure de la figure donne quelques exemples de cette famille, notamment des acides gras et quelques molécules plus grandes.

synthèse endogène [Kates *et al.* 1968 ; Corcelli *et al.* 1996 ; Carballeira *et al.* 1997 ; Nishihara *et al.* 2000] et certains travaux ont même essayé de caractériser biochimiquement ce qui pourrait être leur machinerie de synthèse [Kates *et al.* 1968 ; Pugh *et al.* 1971 ; Pugh et Kates 1994].

Malgré ces observations, la plupart des travaux ont tendance à minimiser l'importance des acides gras chez les archées [De Rosa *et al.* 1986 ; Martin et Russell 2003 ; Boucher 2007 ; Dibrova *et al.* 2011]. Je pense que la raison à cela est double. D'une part, historiquement, la découverte d'un type de phospholipide différent chez les archées a été très importante pour l'acceptation des archées comme un domaine du vivant différent des autres dans les années 1970, laissant de côté les traces d'acides gras détectables. D'autre part, conceptuellement, il semblerait que l'absence apparente d'acides gras dans les phospholipides ait imperceptiblement glissé vers l'idée que les acides gras eux-mêmes étaient absents de ces organismes [Martin et Russell 2003]. Pourtant, les acides gras existent chez les archées, il est donc important de les étudier afin de mieux comprendre leurs rôles biologiques dans ces organismes. Dans le contexte de mon travail cette question était primordiale à cause de ses implications évolutives sur les lipides ancestraux.

A ce jour, aucun mécanisme de synthèse des acides gras n'est connu chez les archées, mais des recherches génomiques ont détecté des homologues des gènes des voies bactériennes de synthèse et de dégradation des acides gras dans les génomes d'archées [Peretó *et al.* 2004 ; Jenke-Kodama *et al.* 2005 ; Iverson *et al.* 2012]. J'ai donc décidé d'étudier l'histoire évolutive des gènes de la voie bactérienne de synthèse des acides gras en m'intéressant tout particulièrement aux archées afin de comprendre si une telle voie peut exister chez ces organismes et de déterminer l'origine de cette voie.

Les acides gras, représentants élémentaires d'une famille moléculaire très riche

Les polycétides forment une famille de composées très diverse qui inclue un grand nombre de métabolites secondaires (messagers, antibiotiques, flavonoïdes, etc). Leur point commun est la synthèse par addition successive d'unités carboxyliques comme l'acétate, le propionate ou le butyrate intercalées avec des réactions très variables, principalement des réductions et déshydratations (parfois isomérisations et même cyclisations) qui modifient les chaînes carbonées en cours d'élongation (Figure 69) [Hopwood et Sherman 1990]. Les acides gras sont des chaînes aliphatiques avec un groupement carboxylique terminal pouvant comporter, ou pas, des insaturations et parfois des branchements (Figure 69). Le mécanisme

de synthèse des acides gras fonctionne sur le même principe que celle des polycétides, et peuvent par conséquent être classés parmi eux [Hopwood et Sherman 1990]. Les machineries responsables de la synthèse des acides gras et des polycétides secondaires sont d'ailleurs homologues entre elles et des nombreux échanges fonctionnels et structurels existent entre ces mécanismes [Summers *et al.* 1995 ; Gokhale *et al.* 2007 ; Wesener *et al.* 2011], au point que les découvertes sur la synthèse des acides gras ont permis de mieux comprendre la synthèse des polycétides secondaires, et réciproquement (pour des révisions historiques sur la découverte des voies de synthèse des acides gras et autres polycétides, voir [Lynen 1964 ; Summers *et al.* 1995 ; Smith et Tsai 2007]). Etant donné que la plus grande diversité des polycétides se trouve parmi les métabolites secondaires et que, contrairement à celle permettant la synthèse des acides gras, leur machinerie de synthèse est très variable, il est implicitement admis que la synthèse des PKS a évolué par duplication et néofonctionalisation des mécanismes de synthèse des acides gras [Hopwood et Sherman 1990 ; Summers *et al.* 1995 ; Kroken *et al.* 2003 ; Schweizer et Hofmann 2004 ; Castoe *et al.* 2007]. Cependant, cette relation est compliquée, comme nous le verrons plus tard, à cause des très nombreuses communications fonctionnelles et structurelles qui existent entre toutes ces machineries [Summers *et al.* 1995 ; Gokhale *et al.* 2007 ; Wesener *et al.* 2011]. En résumé, les acides gras peuvent être chimiquement classés parmi la famille plus large des polycétides, mais on accepte généralement que les machineries de synthèse des polycétides secondaires soient dérivées des machineries de synthèse des acides gras.

Il peut être révélateur ici d'établir un parallèle théorique entre les polycétides et les isoprénoïdes. Dans les deux cas, ce sont des familles moléculaires très étendues dont les représentants les plus élémentaires constituent les chaînes hydrophobes des phospholipides dans les différents domaines du vivant, alors que les molécules dérivées sont beaucoup plus diverses et réalisent des fonctions incomparables entre les différents organismes. Cependant, une différence remarquable entre les isoprénoïdes et les polycétides est que la synthèse *de novo* des isoprénoïdes passe par la synthèse de chaînes d'unités d'isoprène qui sont ultérieurement modifiées, alors que les polycétides sont modifiés au fur et à mesure de leur élongation par un petit nombre de réactions (condensation, réduction, déshydratation). Ceci se traduit par le fait que l'on a pu étudier l'évolution de la machinerie de synthèse des chaînes élémentaires d'isoprénoïdes sans prendre en compte la machinerie secondaire, alors que, dans le cas des polycétides, il existe un dialogue constant entre synthèse des acides gras et des

A chaque tour les molécules peuvent subir des réactions de :

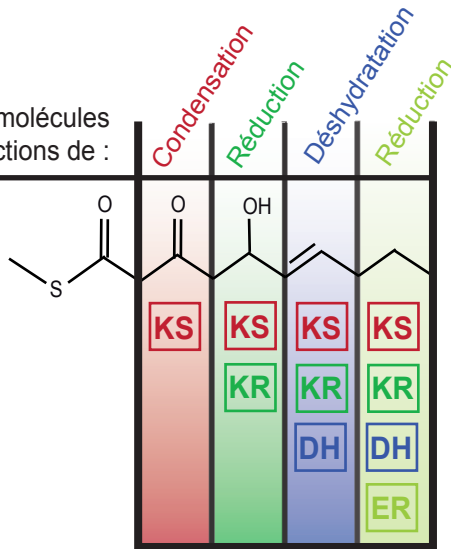


Figure 70. Architectures des synthèses de polycétides (au sens large).

La partie supérieure de la figure présente les réactions que peut subir une chaîne à chaque addition d'une nouvelle unité et les domaines protéiques associés. La partie inférieure présente les principales architectures qui ont été observées à partir de ces domaines (information adaptée de Jenke-Kodama *et al.* 2005). Les machineries dissociées sont soulignées en jaune.

Abréviations : KS : cétoacyl-ACP synthase ; KR : cétoacyl-ACP réductase ; DH : déhydratase ; ER : enoyl réductase ; ACP : acyl carrier protein ; AcT : acétyltransférase ; MCAT : malonyl-CoA:ACP transacylase ; PPT : phosphopantéthéinyl transférase ; TE : thioestérase.

FAS II bactérien



FAS I bactérien (actinobactéries)



FAS I fongique



FAS I animal



PKS II itératif (bactéries)



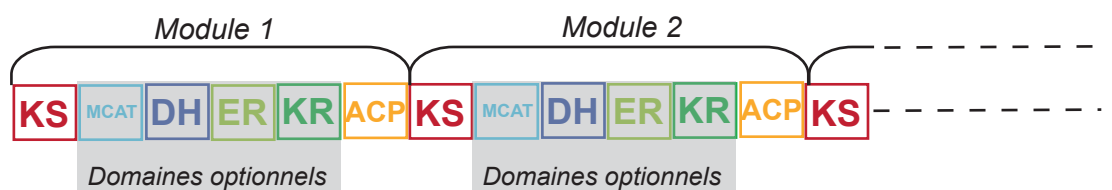
PKS I itératif (bactéries)



PKS I itératif (bactéries/champignons)



PKS I modulaire (bactéries)



polycétides secondaires par le biais de nombreuses duplications, HGTs et recrutements qui introduisent beaucoup de désordre dans l'évolution de ces machineries de synthèse.

Mécanismes de synthèse des acides gras et des polycétides secondaires

Une des caractéristiques principales de la synthèse des polycétides est que chaque activité enzymatique est associée à un domaine protéique dont l'arrangement avec les autres génère un grand nombre d'architectures différentes ([Jenke-Kodama *et al.* 2005], Figure 70). Les complexes enzymatiques qui synthétisent les acides gras sont appelés FAS (*Fatty Acid Synthases*). On les classe en général en deux grands groupes ; les FAS-I ont été les premières à être identifiées et sont composées de grands⁴⁵ polypeptides qui contiennent plusieurs domaines de synthèse des acides gras. Plusieurs de ces polypeptides interagissent ensemble et forment des grandes structures macromoléculaires qui permettent au substrat en cours d'élongation d'être modifié successivement par chaque domaine sans quitter la structure. Les FAS-I existent chez certaines bactéries et sont surtout connues dans le cytoplasme de plusieurs eucaryotes, notamment des animaux et des champignons [Maier *et al.* 2006 ; Jenni *et al.* 2007]. Dans les FAS-II chaque domaine se trouve dans un peptide différent et les chaînes en cours d'élongation doivent par conséquent transiter entre des peptides différents. Les FAS-II sont présentes chez la grande majorité des bactéries et dans les plastes [White *et al.* 2005] et ont été récemment identifiées dans les mitochondries [Schneider *et al.* 1997 ; Stuible *et al.* 1998 ; Miinalainen *et al.* 2003 ; Zhang *et al.* 2003a ; Hoja *et al.* 2004 ; Olsen *et al.* 2004 ; Christensen *et al.* 2007 ; Autio *et al.* 2008 ; Hiltunen *et al.* 2010]. La découverte récente d'une FAS-II dans les mitochondries a beaucoup attiré l'attention car les premières recherches historiques s'étaient efforcées de montrer que la voie mitochondriale de dégradation (β -oxydation) des acides gras était réversible et pouvait produire les précurseurs à partir des produits finaux [Wakil 1961 ; Lynen 1964 ; Whereat *et al.* 1967]. Un tel mécanisme s'était avéré impossible et la synthèse mitochondriale est en réalité comparable au FAS-II bactérien plutôt qu'à l'hélice de Lynen [Hiltunen *et al.* 2010].

Les archétypes des deux types de FAS, FAS-I et FAS-II, utilisent un cofacteur particulier, l'*acyl carrier protein* (ACP). L'ACP est attachée aux intermédiaires en cours d'élongation et participe au cheminement du substrat d'une enzyme à la suivante [Zhang *et al.* 2003b]. Les FAS sont itératives : les mêmes étapes enzymatiques se répètent à chaque fois

⁴⁵ La taille de ces polypeptides est très variable, mais les plus petits comptent autour des 2000 acides aminés.

qu'une nouvelle unité est ajoutée à la chaîne en cours d'élongation [White *et al.* 2005 ; Cronan et Thomas 2009]. Il y a quatre étapes : la condensation de deux unités acylées par une β -cétoacyl synthase (KS), la réduction par une β -cétoacyl réductase (KR), une déshydratation par une 3-hydroxyacyl déshydratase (DH) et une nouvelle réduction par une enoyl réductase (ER). Etant donné que la répétition de ces réactions se fait avec des chaînes de plus en plus longues, qui ne sont donc pas les mêmes à chaque tour, il semble approprié de parler de spirale de synthèse plutôt que de cycle [Lynen 1964]. La fin de la synthèse est variable et fait généralement intervenir, soit le transfert de l'acide gras directement sur une molécule réceptrice (phospholipide ou cofacteur), soit la libération de la chaîne par la simple séparation de l'ACP ou son remplacement par une CoA [Schweizer et Hofmann 2004].

Les complexes capables de synthétiser les autres polycétides sont nommés en référence à la classification la plus courante des FAS [Staunton et Weissman 2001]. On appelle ces complexes PKS (*PolyKetide Synthases*) et on les divise généralement en PKS-I lorsqu'ils sont composés de polypeptides comportant plusieurs domaines protéiques attachés et PKS-II quand les enzymes se trouvent dans des peptides dissociés (Figure 70, [Jenke-Kodama *et al.* 2005]). Chez les plantes, les mécanismes de synthèse des chalcones, qui sont des polycétides intermédiaires importants par exemple pour la synthèse de pigments, sont parfois appelés PKS-III. Les raisons de cette classification seraient liées au fait que ces enzymes n'utilisent pas des thioesters d'ACP mais des thioesters de CoA [Staunton et Weissman 2001 ; Moss *et al.* 2004] et à l'apparente limitation de ces fonctions aux plantes. Cependant, nous verrons que la capacité des domaines de synthèse des polycétides à utiliser les thioesters de CoA semble en réalité être beaucoup plus répandue que et la fonction chalcone synthase a maintenant été décrite chez plusieurs bactéries [Moore et Hopke 2001]. De fait, et parce que la rigidité d'une telle classification s'oppose à l'énorme diversité fonctionnelle et structurelle des PKS, cette classification a été déconseillée [Muller 2004]. Nous n'utiliserons donc ici la notation PKS-I/PKS-II que pour se référer aux PKS polypeptidiques et à ceux composés par des enzymes distinctes respectivement, sans sous-entendre de fonction, structure ou origine commune.

Les domaines protéiques des PKS sont homologues à ceux des FAS, ce qui soutient l'origine commune des PKS et des FAS. La différence principale entre les PKS et les FAS est que dans les premiers l'addition d'une unité acylée n'est pas systématiquement suivie de toutes les modifications répétitives des FAS (réductions et déshydratation), ce qui fait que les produits obtenus sont bien plus divers [Hopwood et Sherman 1990]. Bien que les classifications soient rapidement dépassées lorsqu'elles concernent les PKS [Muller 2004], on

distingue souvent les PKS-I dites *itératives*, dont les polypeptides ne contiennent pas tous les domaines caractéristiques des FAS mais qui sont utilisés de façon répétée comme les FAS-I, des PKS-I dites *modulaires*, qui sont d'immenses polypeptides multifonctionnels (les mieux connus, ceux des mycobactéries, accumulent presque 17000 acides aminés) composés de plusieurs *modules* regroupant chacun des domaines protéiques (KS, KR, ER, etc), de sorte que chaque domaine n'est utilisé qu'une seule fois et que toute (ou grande partie) de la voie de synthèse du produit se trouve dans un seul polypeptide [Moss *et al.* 2004] (Figure 70). La similarité entre les domaines qui composent ces machineries permet des recrutements réciproques entre les PKS et les FAS [Summers *et al.* 1995 ; Gokhale *et al.* 2007 ; Wesener *et al.* 2011].

Evolution des machineries de synthèse des polycétides

Etant donné l'énorme variabilité chimique et fonctionnelle des polycétides et de leur synthèse [Smith et Tsai 2007], il serait impossible de suivre dans le détail l'évolution de toutes les FAS et PKS dans les trois domaines du vivant [Fischbach *et al.* 2008]. Les principales difficultés sont que les polypeptides sont souvent soumis à des recombinaisons⁴⁶ [Jenke-Kodama *et al.* 2005 ; Castoe *et al.* 2007], que beaucoup de composés peuvent être à la fois utilisés par des composants des PKS et des FAS [Summers *et al.* 1995 ; Gokhale *et al.* 2007 ; Wesener *et al.* 2011], que nous connaissons peu des fonctions de tous les paralogues de ces domaines protéiques dans un nombre suffisant d'espèces [Moss *et al.* 2004 ; Smith et Tsai 2007] et que le nombre de paralogies cachées et de HGTs, surtout parmi les PKS, semble très important [Kroken *et al.* 2003 ; Jenke-Kodama *et al.* 2005 ; Castoe *et al.* 2007]. Par conséquent, il n'existe pas de scénario évolutif complet pour expliquer l'origine de toutes ces machineries de synthèse, mais on peut résumer l'avis général présent dans la littérature comme suit :

1. Evolution des machineries dissociées, FAS-II et PKS-II : Le caractère essentiel des acides gras par rapport aux polycétides secondaires fait penser que les PKS sont apparus par la duplication et neofonctionalisation de machineries FAS préalables [Hopwood et Sherman 1990 ; Summers *et al.* 1995 ; Kroken *et al.* 2003 ; Schweizer et Hofmann 2004 ; Castoe *et al.* 2007 ; Cronan et Thomas 2009 ; Massengo-Tiasse et Cronan 2009]. Cependant, à ce jour

⁴⁶ On a signalé même des recombinaisons entre les PKS et des systèmes moléculaires totalement différents, comme les NRPS (polypeptides responsables de la synthèse non ribosomale de peptides) [Smith et Sherman 2008].

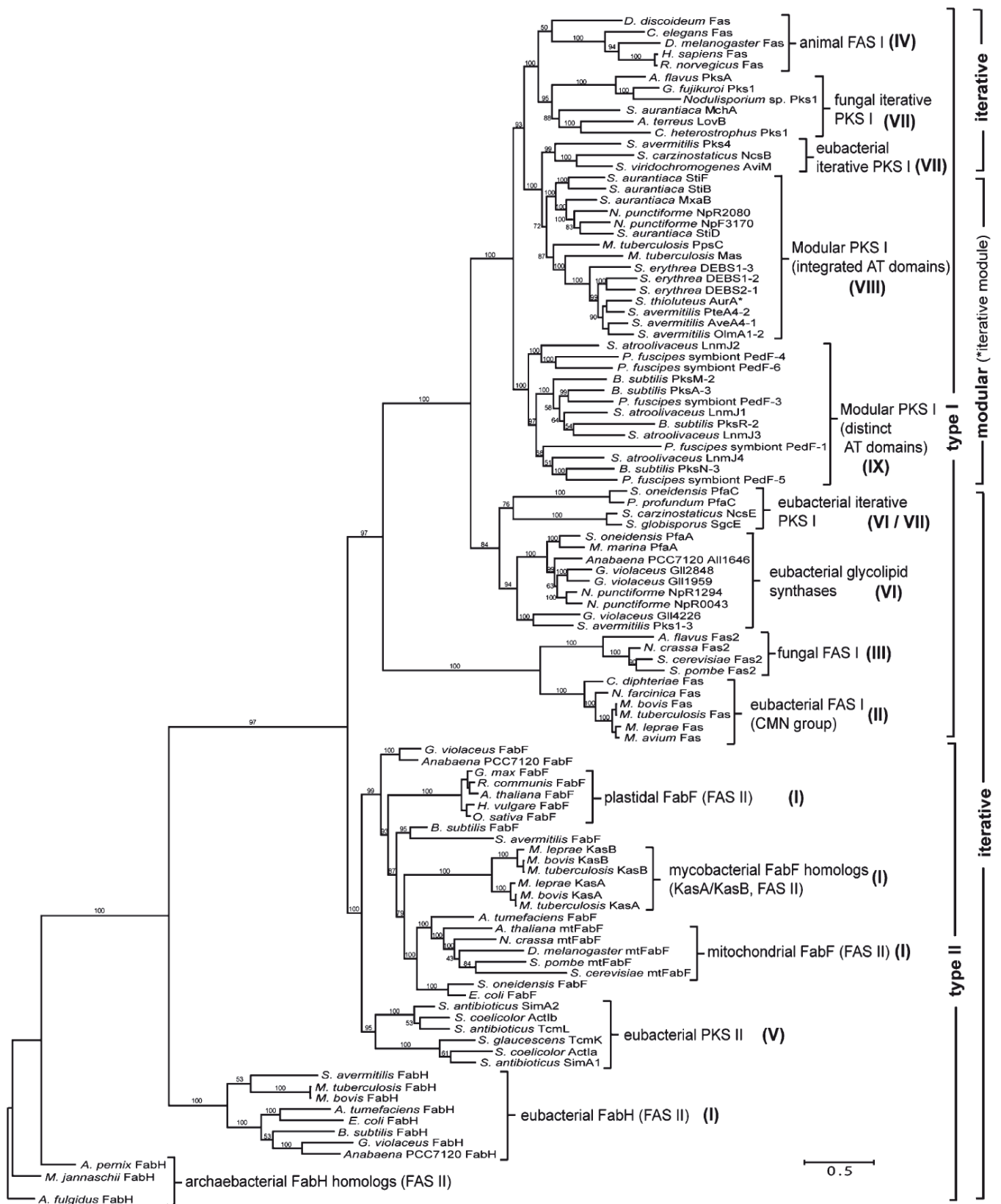


Figure 71. Arbre phylogénétique du domaine KS de différentes FAS et PKS (tiré de Jenke Kodama *et al.* 2005).

Se référer au texte pour discussion.

aucun travail ne s'est vraiment focalisé sur les relations phylogénétiques entre les FAS-II et les PKS-II. Il est admis que les FAS-II sont répandues parmi les bactéries, les mitochondries et les plastes ; avec le support des quelques phylogénies du domaine KS, on accepte que les FAS-II soient anciennes chez les bactéries et que les FAS-II eucaryotes soient d'origine endosymbiotique [Schneider *et al.* 1997 ; Jenke-Kodama *et al.* 2005]. Les phylogénies du domaine KS montrent que les séquences de PKS-II sont plus proches de leurs homologues de FAS-II que de ceux des PKS-I [Jenke-Kodama *et al.* 2005]. Cependant, pour comprendre l'origine des PKS-II, il serait indispensable d'étudier la phylogénie des autres domaines et de le faire avec un échantillon de machineries dissociées beaucoup plus important.

2. Evolution des polypeptides de synthèse des acides gras, FAS-I: Les FAS-I sont décrites chez peu de bactéries (principalement des actinobactéries). Elles regroupent la machinerie déjà présente dans les FAS-II en un seul polypeptide, donc il est généralement admis que les FAS-I sont issus de la fusion des FAS-II [Jenke-Kodama *et al.* 2005 ; Leibundgut *et al.* 2008 ; Smith et Sherman 2008]. Néanmoins, la question de l'origine des FAS-I n'est pas triviale car les FAS-I des champignons et des métazoaires, qui sont les deux archétypes de FAS-I connus chez les eucaryotes, ne se ressemblent pas entre elles [Leibundgut *et al.* 2008] : les domaines du FAS-I des animaux sont dans un ordre qui est caractéristique des PKS-I, alors que les FAS-I des champignons ont une synténie similaire aux FAS-I des actinobactéries⁴⁷ [Jenni *et al.* 2007 ; Maier *et al.* 2008] (Figure 70). Les phylogénies des domaines KS de différents FAS et PKS soutiennent que les FAS-I animales sont groupe frère de certaines PKS-I fongiques [Kroken *et al.* 2003] et que l'ensemble de ces deux clades branche parmi d'autres PKS-I (Figure 71) [Jenke-Kodama *et al.* 2005 ; Castoe *et al.* 2007]. On peut donc admettre que la FAS-I animale soit le résultat d'une néofonctionnalisation d'une PKS-I. Une de ces phylogénies [Jenke-Kodama *et al.* 2005] montre aussi que les FAS-I des actinobactéries et celles des champignons forment un groupe à la fois très distant des ceux des FAS-II et des PKS-I, sans que cela ne permette de déterminer si l'origine des FAS-I de ce type est une fusion indépendante à partir de domaines isolés de type FAS-II ou d'une recombinaison d'un peptide de type PKS-I. En dehors des animaux et

⁴⁷ Il est vrai que l'archétype des FAS-I de champignon est composé de deux polypeptides qui ne forment qu'un seul dans les FAS-I d'actinobactéries (Figure 70). Cependant, le fait qu'il existe des lignées de champignons avec un seul polypeptide comme celui des actinobactéries et le fait que dans les systèmes à deux polypeptides la séparation se fasse à l'intérieur d'un domaine protéique font penser que la machinerie à un polypeptide est plus ancienne et que le système à deux polypeptides résulte de la cassure récente de ce système ancestral [Jenni *et al.* 2007 ; Leibundgut *et al.* 2008].

des champignons, il existe très peu d'information sur les FAS-I et PKS-I des autres eucaryotes. Les recherches dans les génomes des protistes révèlent une distribution désordonnée des FAS-I de type animal et des FAS-I de type fongique [Leibundgut *et al.* 2008], sans que l'on ne puisse tirer de conclusions sur l'évolution des FAS-I à l'échelle de l'ensemble des eucaryotes.

3. Evolution des polypeptides de synthèse des polycétides secondaires, PKS-I: A cause de leur nature polypeptidique et secondaire, pour les PKS-I plus encore que pour les PKS-II, il est généralement accepté que ces machineries sont des dérivés évolutifs des machineries FAS [Hopwood et Sherman 1990 ; Summers *et al.* 1995 ; Kroken *et al.* 2003 ; Schweizer et Hofmann 2004 ; Castoe *et al.* 2007 ; Cronan et Thomas 2009 ; Massengo-Tiasse et Cronan 2009]. Les PKS-I peuvent avoir des architectures très différentes entre eux et sont caractérisés par des nombreuses interactions avec d'autres composants pour synthétiser une grande variété de produits [Smith et Tsai 2007], ce qui rend difficile leur étude évolutive. Cependant, la synténie des domaines dans la plupart des polypeptides PKS-I est généralement conservée et les exceptions sont souvent explicables par des pertes ou gains ponctuels de domaines [Kroken *et al.* 2003]. Je ne rentrerai pas ici dans le détail de ces modifications ponctuelles récentes, mais nous allons analyser l'origine de ces séquences.

Comme vu précédemment, les FAS-I des métazoaires partagent la même synténie que les PKS-I dont ils sont probablement issus [Jenke-Kodama *et al.* 2005 ; Castoe *et al.* 2007] mais diffèrent dans l'ordre des domaines des FAS-I bactériens et fongiques. Des phylogénies des domaines KS des PKS et FAS ont montré que les PKS-I forment un seul groupe subdivisé dans les différents types de PKS-I existants (Figure 71). Il a été proposé qu'un groupe de PKS-I itératif ait été l'ancêtre de ces séquences, parmi lesquelles seraient apparues les PKS-I modulaires. En effet, les PKS-I des champignons et les FAS-I des animaux branchent parmi les PKS-I modulaires même s'ils sont itératifs, suggérant une acquisition secondaire de l'activité itérative dans ces séquences eucaryotes [Jenke-Kodama *et al.* 2005]. Le groupe des PKS-I (au sens large, car cela inclus aussi les FAS-I des animaux qui en dérivent) est frère de celui des FAS-I des actinobactéries et des champignons. On pourrait penser qu'une fusion primaire des domaines protéiques pour former un premier polypeptide aurait pu avoir lieu, suivi de plusieurs recombinaisons qui auraient donné lieu à deux architectures différentes, de type FAS-I et PKS-I. Il est important de souligner que dans cette phylogénie les clades correspondant aux FAS-II, aux FAS-I de type actinobactérien et fongique et aux PKS-I sont tous séparés par des très longues branches qui sont probablement provoquées par les fortes

divergences enregistrées dans les séquences lorsque les domaines protéiques ont été soumis à de nouvelles contraintes structurelles pour former les polypeptides. Malheureusement, ces longues branches rendent incertaine l'origine commune des FAS-I (type actinobactérien/fongique) et PKS-I, car il pourrait être proposé que les deux fusions aient été indépendantes et que ces groupes se retrouvent à proximité dans l'arbre à cause d'un phénomène d'attraction des longues branches. Par ailleurs, la grande divergence entre les séquences des polypeptides et celles des machineries dissociées est telle qu'il est pour l'instant impossible de déterminer de façon fiable les organismes ancestraux chez qui se sont produits la ou les fusions primaires des PKS-I et FAS-I (type actinobactérien/fongique) ou de déterminer la fonction précise de ces gènes avant et après la ou les fusions (synthèse d'acides gras et/ou de polycétides secondaires). Comme pour les FAS-I, il existe peu d'information sur les PKS-I des eucaryotes en dehors des opisthokontes, mais des analyses récentes ont permis de détecter des séquences classées comme PKS-I chez certains protistes [John *et al.* 2008 ; Leibundgut *et al.* 2008]. Néanmoins, ces travaux doivent encore être approfondis pour s'assurer de leur fonction et déterminer si ces séquences ont été héritées de l'ancêtre commun des eucaryotes ou par des HGT multiples.

4. Conclusion : on accepte généralement que les polypeptides de FAS-I et PKS-I sont le résultat de fusions des domaines indépendants de type FAS-II ou PKS-II. Les FAS-I et PKS-I modernes forment des groupes issus respectivement d'un seul ancêtre commun, mais il n'est pas clair, pour l'instant, si les deux sont issus d'une seule fusion primaire ou de fusions indépendantes. Ni l'identité de l'ancêtre chez qui ces fusions ont pu avoir eu lieu, ni les fonctions des gènes qui ont donné lieu aux premiers polypeptides n'ont été déterminées. Ces questions pourront certainement être traitées plus en détail lorsque l'on connaîtra mieux les machineries FAS et PKS d'autres eucaryotes en dehors des opisthokontes et que l'on disposera d'une caractérisation fonctionnelle plus large de tous ces composants dans une plus grande diversité d'organismes.

On remarquera que, dans la littérature, l'attention s'est concentrée sur l'évolution des polypeptides et surtout des PKS-I, alors que les machineries dissociées ont été moins étudiées. Pourtant, la question de la présence d'une machinerie de synthèse des acides gras chez le cenancêtre est subordonnée à une étude plus approfondie de l'évolution ancienne des voies de synthèse actuelles, et surtout à la détermination d'un tel mécanisme chez les archées. J'ai donc focalisé mon travail sur l'évolution des machineries dissociées (principalement les FAS-II) pour deux raisons principales : d'abord, si l'on accepte que les polypeptides sont issus de

la fusion d'une machinerie dissocié, ce sera surtout l'étude de cette machinerie dissociée qui nous donnera des informations sur l'évolution ancienne de la synthèse des polycétides ; deuxièmement parce que la littérature et mes vérifications préliminaires ont signalé dans les génomes des archées des homologues prometteurs des machineries dissociées [Peretó *et al.* 2004 ; Jenke-Kodama *et al.* 2005 ; Iverson *et al.* 2012], mais pas de polypeptides [Jenke-Kodama *et al.* 2005 ; Castoe *et al.* 2007].

2. Evolution des gènes de synthèse des acides gras FAS-II chez les archées

Bien que la synthèse de polycétides utilise un nombre très réduit de domaines protéiques, les mécanismes sont trop divers pour permettre une étude évolutive complète. Par conséquent, j'ai limité mon travail à une parcelle abordable en fixant trois axes restrictifs de départ : (1) étant donné que les polypeptides n'ont pas été détectés dans les génomes archéens et que les polypeptides bactériens et eucaryotes introduisaient beaucoup de bruit dans les alignements, je ne les ai pas inclus dans mes analyses ; (ii) bien que les recherches de similarité dévoilent de très nombreuses séquences de PKS-II, leur caractérisation biochimique et fonctionnelle fait encore défaut. Je me suis donc concentré sur les séquences bactériennes les plus proches des représentants avérés de FAS-II ; (iii) d'après ce que nous avons vu, les mécanismes de synthèse de polycétides chez les eucaryotes font intervenir soit des polypeptides qui n'ont pas été conservés dans mon analyse, soit des machineries dissociées considérées comme étant d'origine endosymbiotique et, par conséquent, ne fournissant pas d'information ni sur les archées ni sur le cenancêtre ; l'origine des séquences eucaryotes ne sera donc pas discutée dans mon analyse.

Etapes préalables à la synthèse des acides gras par la FAS-II bactérienne

En 1958, alors que la voie de synthèse des acides gras n'était pas encore déterminée, une fraction de foie de pigeon capable de fabriquer de l'acide palmitique (acide gras saturé de 16 carbones) à partir d'acétyl-CoA fut identifiée. Pour que la synthèse soit possible, il fut déterminé qu'il lui fallait, en dehors de l'acétyl-CoA, de l'ATP, du NADPH, des ions Mn^{2+} , du bicarbonate et de la biotine [Gibson *et al.* 1958]. De ces composants, celui qui étonna le plus était le bicarbonate, surtout parce que l'on savait qu'un radiomarquage du bicarbonate n'était pas retrouvé dans les acide gras produits [Brady et Gurin 1950 ; Klein 1957] : le bicarbonate était donc nécessaire mais ne rentrait pas dans la composition de l'acide gras synthétisé. Pendant la Conférence Gordon sur le métabolisme lipidique de 1958, le besoin à la

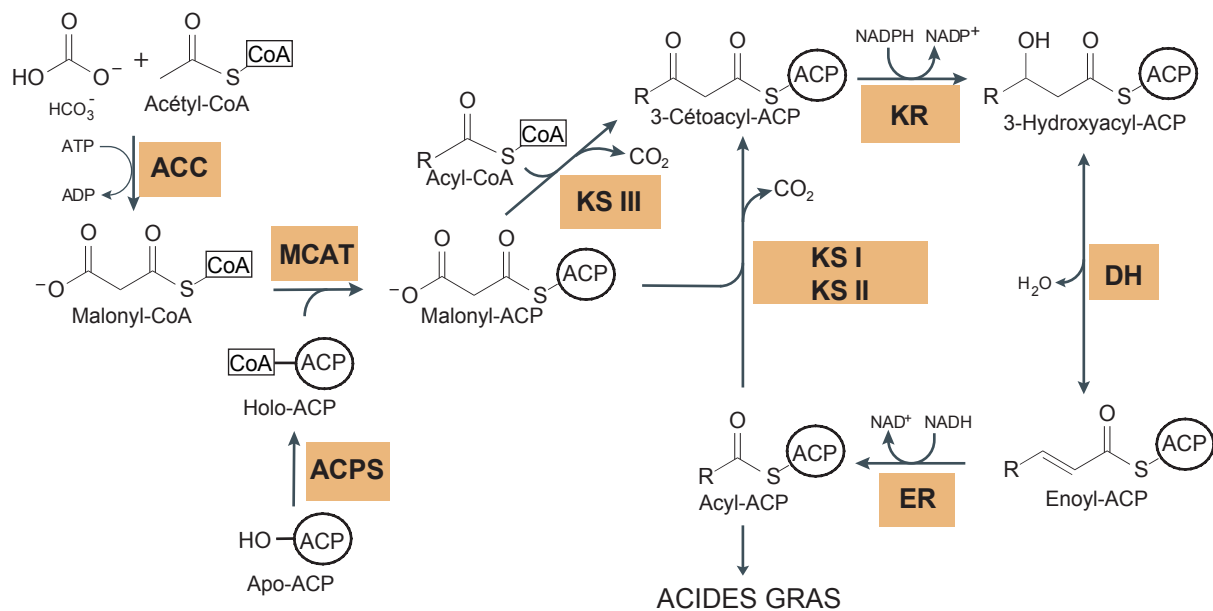


Figure 72. Réactions caractéristiques des FAS-II bactériennes.

Abréviations. CoA : coenzyme A ; ACP : acyl carrier protein ; ACC : acétyl-CoA carboxylase ; MCAT : malonyl-CoA:ACP transacylase ; ACPS : ACP synthase ; KS : cétoacyl-ACP synthase ; KR : cétoacyl-ACP réductase ; DH : déshydratase ; ER : enoyl réductase.

fois d'ATP et de bicarbonate a conduit à penser que la brique condensée dans la synthèse des acides gras pouvait être le malonyl-CoA au lieu de l'acétyl-CoA ; cette molécule subirait une décarboxylation au moment de la condensation, ce qui ferait que seuls les deux carbones de l'acétyl-CoA seraient intégrés (Figure 72), expliquant la contradiction entre le besoin de bicarbonate et l'absence de marquage radioactif dans les produits finaux [Lynen 1964 ; Smith et Tsai 2007]. Très rapidement plusieurs laboratoires ont mis en évidence une carboxylase dépendante de la biotine et responsable de la carboxylation d'acétyl-CoA en malonyl-CoA (acetyl-CoA carboxylase, ACC), expliquant par la même occasion pourquoi les fractions qui synthétisaient les acides gras étaient riches en cette vitamine [Brady 1958 ; Wakil 1958]. Très souvent les travaux concernant les voies de synthèse des acides gras incluent l'acétyl-CoA carboxylase dans la voie de synthèse [Campbell et Cronan 2001 ; White *et al.* 2005]. Ceci se comprend d'un point de vue historique puisque la découverte de cette enzyme a été une des clés de la découverte du mécanisme de synthèse des acides gras. De plus, chez certains organismes modèles comme *Escherichia coli* on ne connaît pas d'autre utilisation du malonyl-CoA en dehors de celle-ci [White *et al.* 2005]. Cependant, le malonyl-CoA est une molécule ubiquitaire et centrale dans des métabolismes très divers et participe à de nombreux processus métaboliques dans d'autres organismes. De plus l'ACC appartient à la famille des carboxylases à biotine, dont les autres membres participent aussi dans des fonctions diverses. J'ai étudié l'évolution de cette superfamille dans un contexte différent de celui de la synthèse de membranes (celui de la fixation du carbone). Je ne discuterai pas ici des résultats car ils sortiraient du sujet de la synthèse des membranes mais nous pouvons retenir que des homologues de l'ACC ont été détectés et caractérisés biochimiquement chez les archées [Menendez *et al.* 1999 ; Chuakrut *et al.* 2003 ; Hugler *et al.* 2003], que sa fonction est relativement répandue dans ce domaine et que sa présence chez le dernier ancêtre des archées ne peut pas être exclue ([Lombard et Moreira 2011a], voir Annexe 3).

La synthèse des acides gras utilise un cofacteur protéique très important, l'acyl carrier protein (ACP). Ce cofacteur est très abondant dans la cellule car, si on prend l'exemple d'*Escherichia coli*, il constitue la quatrième molécule avec le plus d'interactions protéiques dans la cellule (après les polymérases à ADN et à ARN et les protéines ribosomales) et ses fonctions ne se limitent pas à la synthèse des acides gras, mais s'étendent à la fabrication de vitamines et lipopolysaccharides, l'activation de toxines, la participation dans le *quorum sensing* et certains phénomènes de bioluminescence [Chan et Vogel 2010]. L'ACP est une protéine assez courte (environ 80 acides aminés) dont la séquence primaire est relativement

peu conservée à l'exception d'un résidu de sérine qui permet la liaison d'un groupement 4'-phosphopantéthéine provenant d'une molécule de CoA (Figure 72). Ce groupement permet à la protéine de former un pont thioester avec les intermédiaires de la synthèse des acides gras, et de faciliter ainsi leur acheminement d'une enzyme à la suivante dans la spirale de synthèse [Zhang *et al.* 2003a ; Zhang *et al.* 2003b] (Figure 72). En effet, l'ACP peut adopter deux conformations principales, l'une qui accueille la chaîne hydrophobe en cours d'élongation pour l'écarter du milieu aqueux environnant, et l'autre qui met en évidence l'intermédiaire de réaction pour le mettre à disposition des enzymes de la FAS [Chan et Vogel 2010]. L'activation de l'ACP par la 4'-phosphopantéthéine est possible grâce à l'ACP synthase (ACPS). Ensuite, une malonyl-CoA:ACP transacylase (MCAT) transfère le malonyl-CoA à l'holo-ACP, donnant lieu au malonyl-ACP (Figure 72) utilisé par la voie de synthèse des acides gras [Chan et Vogel 2010]. Il est intéressant de signaler que certaines ACP sont capables d'activer et/ou de transférer le malonyl-CoA à d'autres ACP sans faire intervenir ACPS et MCAT ([Misra *et al.* 2009]).

La recherche de gènes similaires à ceux de la machinerie de traitement de l'ACP (l'ACP lui-même, l'ACPS et la MCAT) n'a révélé que quelques très rares homologues chez les archées, disséminés dans ce domaine du vivant. Concrètement, des homologues de l'ACP n'ont été détectés que chez *Methanospirillum hungatei*, *Candidatus Methanoregula boonei*, *Haloterrigena turkmenica* et *Halogeometricum borinquense*, les ACPS dans les mêmes methanomicrobiales, *H. borinquense* et *Nitrosopumilus maritimus* et les MCAT dans les deux halobactériales.

Cette distribution éparpillée et le fait que les halobactériales et *N. maritimus* soient connus pour avoir reçu de nombreux HGTs depuis des bactéries [Boucher *et al.* 2003] font penser que les rares gènes liés à l'ACP détectables dans les génomes des archées ont probablement été hérités par des HGTs indépendants. Ces résultats épars sont congruents avec les recensements génomiques précédents, qui n'avaient pas été capables de détecter des homologues de la machinerie ACP chez les archées [Peretó *et al.* 2004]. En revanche, ces résultats contredisent des rapports biochimiques selon lesquels des ACP et même une MCAT étaient détectés dans les archées *Halobacterium salinarium* et *Methanothermobacter thermoautotrophicum* [Pugh *et al.* 1971 ; Pugh et Kates 1994]. Pour expliquer ces observations, il existe deux possibilités : soit l'ACP et sa machinerie existent chez les archées mais sont trop divergentes pour être détectées, soit elles sont tout simplement absentes dans ce domaine. Dans le cas de l'ACP, sa petite taille et la divergence reconnue de ses séquences

primaires pourraient appuyer la première option. Les ACPS sont aussi assez courtes (autour de 150 acides aminés), mais les MCAT sont plus grandes (plus de 300 acides aminés), donc il est plus difficile d'invoquer la divergence des séquences pour expliquer notre incapacité à les détecter. Plus important, dans le cas des ACPS et les MCAT on s'attendrait à ce que la conservation des fonctions enzymatiques respectives modère la divergence des séquences. Par ailleurs, le fait que des homologues du reste de la machinerie FAS soit détectable chez les archées [Peretó *et al.* 2004] privilégie, par opposition, l'hypothèse que la machinerie de l'ACP soit absente de ce domaine du vivant. Néanmoins, il est difficile de trancher cette question avec les éléments dont on dispose, car l'hypothèse d'une absence de ces gènes chez les archées se base sur un argument négatif. Nous reviendrons donc sur cette interprétation ultérieurement.

Le manque de la machinerie ACP peut surprendre au premier abord étant donné l'importance des ACP dans les FAS bactériennes et eucaryotes. Cependant, cette absence n'implique pas forcément une incapacité à fabriquer des acides gras, car des enzymes qui peuvent utiliser des thioesters de CoA au lieu des thioesters d'ACP comme substrats ont été décrites à plusieurs reprises parmi les mécanismes de synthèse de polycétides secondaires ou d'acides gras exotiques. En effet, la machinerie de synthèse des chalcones synthases est souvent donnée en exemple comme une machinerie n'utilisant pas l'ACP mais le CoA [Hopwood et Sherman 1990 ; Moss *et al.* 2004]. Chez les animaux on connaît des mécanismes FAS-II qui utilisent des thioesters de CoA pour produire des acides gras très longs (jusqu'à une trentaine de carbones) utilisés dans la myéline et les phospholipides des photorécepteurs, par exemple [Cook et McMaster 2002] ; on sait d'ailleurs que des enzymes des systèmes FAS-II mitochondriales utilisent préférentiellement des thioesters de CoA [Das *et al.* 2000 ; Miinalainen *et al.* 2003] et que certaines enzymes, tout en préférant les substrats liés à l'ACP, peuvent aussi utiliser ceux attachés au CoA [Zhang *et al.* 2003a].

La superfamille des thiolases

Une fois que les briques de base pour la synthèse des acides gras sont disponibles, la première étape de la spirale de synthèse est la condensation de deux molécules acylées [Alberts *et al.* 1965] (Figure 72). Chez *E. coli*, qui est l'organisme modèle en ce qui concerne les FAS-II, il existe trois β -cétolase synthase (KS) homologues capables de catalyser cette réaction [White *et al.* 2005]. La condensation d'initiation, celle du malonyl-ACP avec un acyl-CoA (en principe l'acétyl-CoA, mais certaines espèces utilisent des chaînes plus longues,

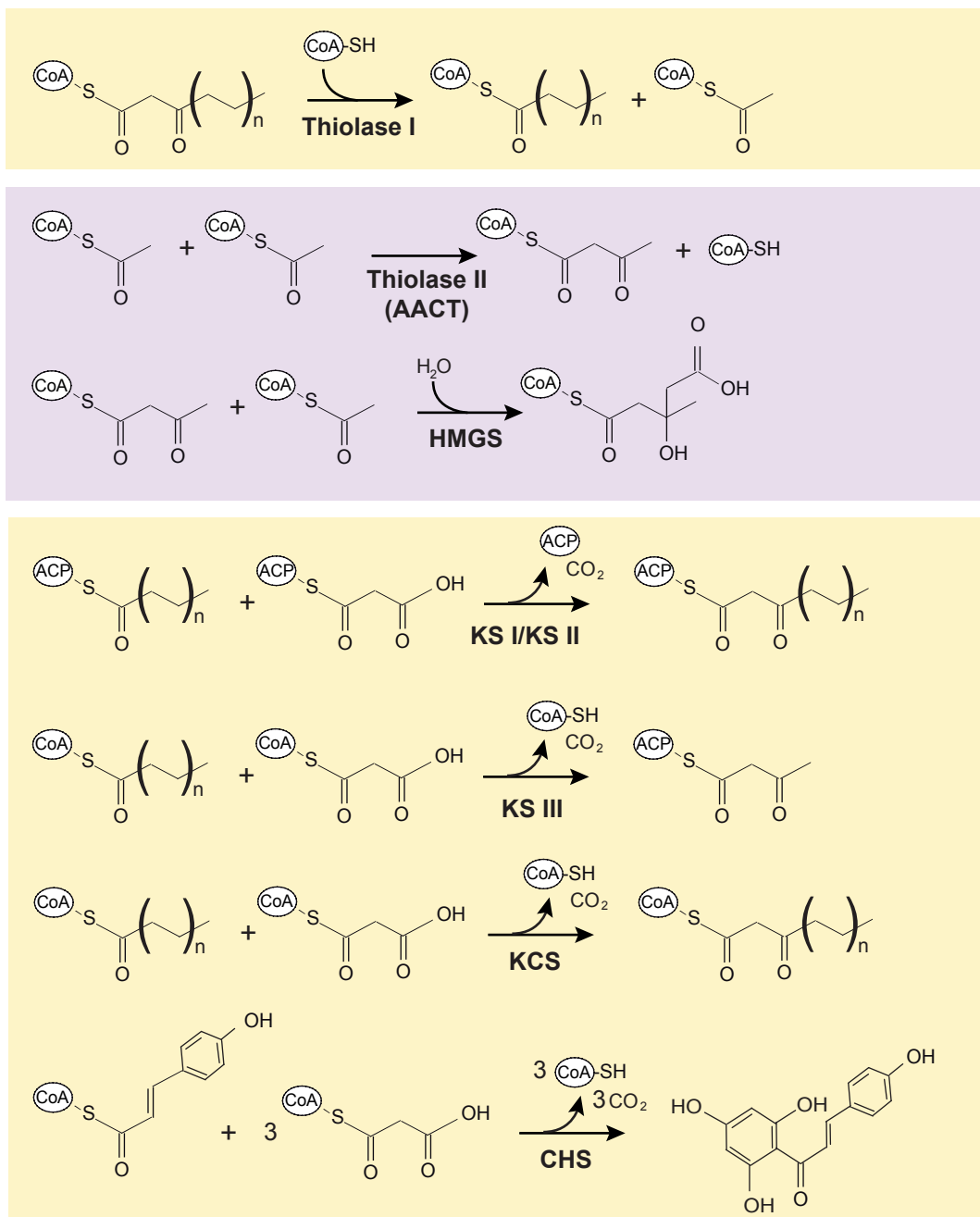


Figure 73. Réactions catalysées par les membres de la superfamille des thiolases.

Toutes les réactions font intervenir l'addition ou la soustraction d'une acyl thioester, attaché soit au CoA soit à l'ACP. AACT et HMGS participent principalement dans la voie du mévalonate de synthèse des isoprénoïdes, alors que toutes les autres enzymes interviennent dans la synthèse ou dégradation de polycétides, qui est la grande famille qui inclut les acides gras.

Abréviations. CoA : Coenzyme A ; ACP : acyl carrier protein ; AACT : acétoacétyl-CoA thiolase ; HMGS : hydroxyméthylglutaryl synthase ; KS : cétoacyl-ACP synthase ; KCS : cétoacyl-CoA synthase ; CHS : chalcone synthase.

[Zhang et Rock 2008], est réalisée par la KS-III, aussi appelée FabH. Son activité précoce limitante joue une fonction régulatrice très importante pour contrôler le flux de l'ensemble de la voie [Jackowski et Rock 1987]. Une fois l'acétoacétyl-ACP résultant totalement réduit, les condensations d'élongation sont catalysées par la KS-II (FabF) à chaque tour de spirale. La KS-I (FabB) intervient dans l'élongation des acides gras qui comportent des insaturations (l'origine des insaturations sera expliquée plus tard, lorsque nous parlerons de la superfamille hot-dog). Bien que ces trois KS constituent l'archétype des FAS-II bactériennes, elles sont toutes les trois homologues entre elles et conservent des mécanismes et des activités très proches, ce qui a permis de montrer qu'une seule de ces enzymes pouvait suffire à assurer les fonctions KS chez certaines bactéries et dans les mitochondries [Morgan-Kiss et Cronan 2008 ; Hiltunen *et al.* 2010].

Les trois KS appartiennent à la superfamille des thiolases (Figure 73) [Jiang *et al.* 2008], qui, en dehors des KS des FAS et PKS, inclut: (i) des thiolases dites de type I capables d'éliminer une chaîne d'acétyl-CoA d'un acyl-CoA dans la voie catalytique de β -oxydation des acides gras ; (ii) des acétoacétyl-CoA thiolases (AACT), aussi appelées thiolases de type II chez les bactéries et les eucaryotes, capables de synthétiser cette molécule à partir de deux unités d'acétyl-CoA. L'acétoacétyl-CoA peut participer dans des nombreuses voies, mais on remarquera particulièrement que cette réaction est la première de la voie de synthèse du mévalonate, et que sa présence a donc été prédite chez les archées, bien que les séquences archéennes soient très différentes des autres thiolases [Liu *et al.* 2002], et, par conséquent, inclassables parmi les thiolases de type I ou II. Chez les eucaryotes plusieurs isoformes des deux types de thiolases existent dans différents compartiments et des phylogénies ont montré l'origine protéobactérienne de ces séquences [Peretó *et al.* 2005] ; (iii) la condensation d'une nouvelle molécule d'acétyl-CoA avec l'acétoacétyl-CoA par l'hydroxymethylglutatyl-CoA synthase (HMGS) permet la synthèse de cette même molécule, qui est le précurseur directe du mévalonate. L'analyse détaillée de l'évolution de cette enzyme dans les trois domaines du vivant a été expliquée dans la partie de cette thèse traitant de l'évolution des isoprénoïdes.

Toutes ces thiolases changent donc la longueur des chaînes acylées en rajoutant ou en enlevant une unité acétyl-CoA. Il existe une différence biochimique importante entre différentes thiolases de cette superfamille [Jiang *et al.* 2008]. En effet, nous avons vu que les KS utilisent du malonyl-ACP pour allonger les chaînes acylées libérant un groupement carboxylique, celui là même qui avait été rajouté par l'acétyl-CoA carboxylase dont il a été question plus haut. On dit des KS que ce sont des enzymes *decarboxylatives*, et cette propriété

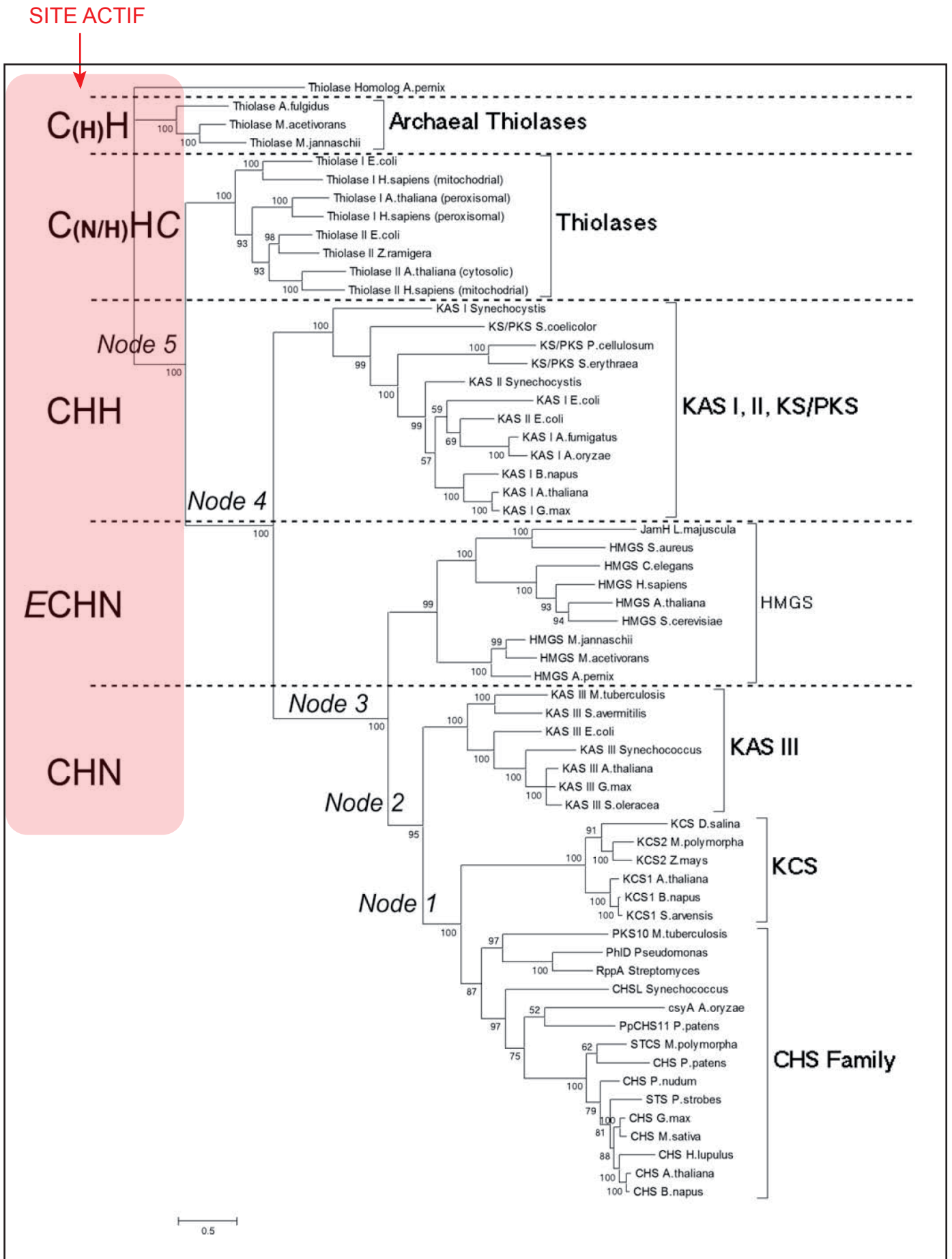


Figure 74. Arbre phylogénétique des thiolases (tiré de Jiang *et al.* 2008).

Se référer au texte pour discussion.

biochimique est partagée par toutes les KS, qu'elles participent dans les FAS-II, FAS-I, ou dans les différents PKS, incluant les chalcones synthases (CHS). En revanche, ce n'est pas le cas de thiolases de type I/II, des thiolases archéennes ni d'HMGS, qui toutes utilisent l'acétyl-CoA directement, sans avoir besoin, par conséquent, de le décarboxyler. Une autre composante biochimique permet de diviser les enzymes en plusieurs groupes, à savoir celle des principaux résidus du site actif à rentrer dans le mécanisme réactionnel de l'enzyme [Jiang *et al.* 2008].

Une étude évolutive de cette superfamille a été publiée récemment [Jiang *et al.* 2008]. Dans ce travail, une reconstruction phylogénétique divisait les thiolases de cette superfamille en cinq grands groupes : un groupe de thiolases archéennes, un groupe de thiolases bactériennes et eucaryotes de type I/II, un groupe contenant les KS-I et KS-II bactériennes, un groupe de KS-III et de chalcones synthases et un groupe d'HMGS eucaryotes et archéens. Les auteurs de ce travail ont reporté l'architecture du site actif sur la phylogénie reconstruite pour ces séquences en essayant de déterminer une séquence de modifications du site actif au cours de l'évolution. Cependant, cette analyse est critiquable dans la mesure où une seule séquence très divergente d'*Aeropyrum pernix* est utilisée pour enraciner l'arbre afin de discuter la série de modifications du site actif au cours de l'évolution, en commençant par les thiolases d'archées, les autres thiolases, les KS-I/KS-II et KS et PKS, les HMGS et enfin les KS-III et CHS (Figure 74). Pourtant, la séquence d'*A. pernix* est la seule à ne pas contenir de résidus catalytiques connus alors que le reste de la superfamille est caractérisée par la grande conservation structurelle des sites actifs. Ainsi, il semble beaucoup plus probable que la séquence d'*A. pernix* soit en réalité une séquence récente très dérivée plutôt qu'un représentant d'une enzyme ancestrale et non fonctionnelle. En dehors de fausser l'interprétation possible de l'évolution des centres actifs dans la superfamille des thiolases, cet enracinement erroné amène des conclusions aberrantes, comme par exemple que les KS-III responsables de l'initiation de la synthèse des acides gras sont apparues après les KS-I/II responsables de l'élongation postérieure des chaînes acylées. Une autre réserve qui peut-être émise sur de ce travail est que les séquences bactériennes d'HMGS n'ont pas été incluses, ce qui est très important notamment pour positionner les HMGS archéens par rapport au reste de la superfamille [Lombard et Moreira 2011b].

Par conséquent, j'ai renouvelé cette analyse sur une sélection restreinte de séquences de la superfamille des thiolases (Figure 75). Une difficulté de cette analyse est que, bien que ces enzymes soient très conservées au niveau structurel, la conservation de leur séquence primaire

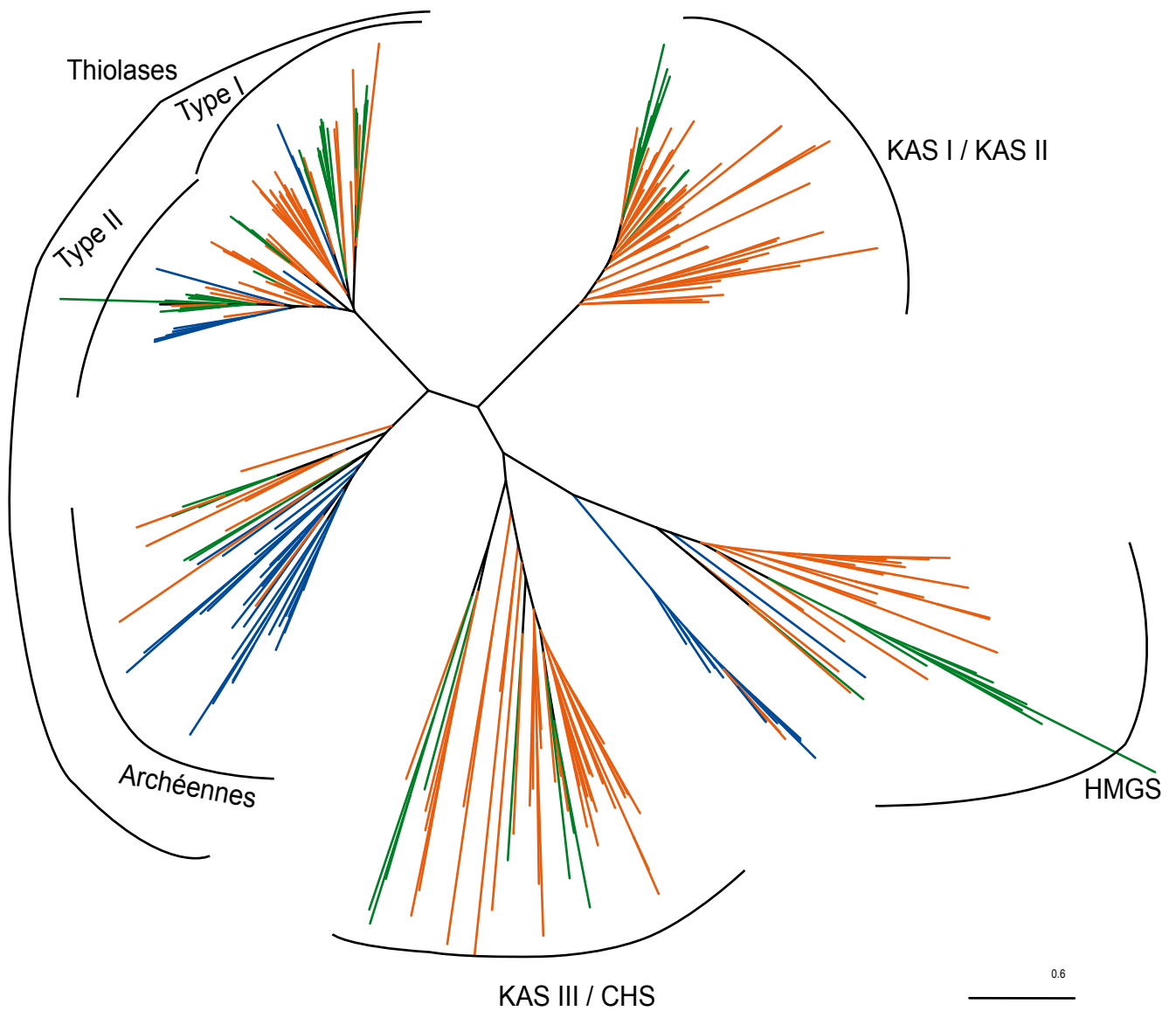


Figure 75. Phylogénie de la superfamille des thiolases.

Reconstruction par maximum de vraisemblance utilisant 273 séquences représentatives et 243 positions conservées. Les séquences bactériennes apparaissent en orange, les archéennes en bleu et les eucaryotes en violet.

est très faible, ce qui se traduit par une faible fiabilité de l'alignement multiple, peu propice à la réalisation d'analyses phylogénétiques précises. Malgré cela, les similarités à l'intérieur de chaque type d'enzyme sont suffisantes pour obtenir quatre groupes clairement identifiés : des thiolases distribuées dans les trois domaines du vivant, des KS-I et KS-II caractéristiques des bactéries et des eucaryotes qui les ont adoptées, les KS-III et CHS avec une répartition taxonomique similaire et les HMGS distribuées dans les trois domaines du vivant.

A l'exception de quelques séquences divergentes, on observe dans le cas de HMGS une division en trois groupes qui correspondent aux domaines du vivant. Ceci suggère qu'une séquence de ce type était déjà présente chez l'ancêtre de chaque domaine du vivant. Ce résultat est cohérent avec les phylogénies d'HMGS rapportées précédemment (Figures 45 et 46, p. 131 et 132) et soutient l'appartenance des séquences archéennes au groupe fonctionnel HMGS. Les thiolases se divisent en deux grands groupes, l'un correspondant aux archées et l'autre correspondant aux thiolases bactériennes et eucaryotes de type I et II. Au sein de ce deuxième groupe, les thiolases de type I forment un clade alors que les thiolases de type II apparaissent paraphylétiques. Ce résultat, de même que la phylogénie confuse du clade archéen sont probablement imputables à l'accumulation de bruit dans l'alignement. Cependant, ces observations montrent que la thiolase décrite chez les archées [Liu *et al.* 2002] est très répandue dans ce domaine. Des analyses précédentes n'utilisant que des thiolases (et donc avec un alignement de meilleure qualité) avaient montré la monophylie des thiolases I/II chez les bactéries et des transferts de ces organismes vers les eucaryotes [Peretó *et al.* 2005]. Ainsi, ces résultats suggèrent la présence d'une enzyme à fonction thiolase chez les derniers ancêtres respectifs des archées et au moins des bactéries.

Ces observations contrastent avec l'absence apparente d'homologues des enzymes KS-I/II/III chez les archées. Les gènes KS semblent spécifiques des bactéries et des eucaryotes ayant obtenu secondairement leurs gènes. Ces observations sont compatibles avec la présence de la machinerie KS chez l'ancêtre des bactéries, mais pas dans l'ancêtre des archées. Que pouvons-nous dire sur l'origine des KS bactériennes ? Bien que faibles, trois arguments permettent de proposer que les séquences KS-I et KS-II aient une origine plus proche des protéines dites thiolases alors que les KS-III semblent plus proches des HMGS : (1) les sites actifs tels qu'analysés dans Jiang *et al.* 2008 ; (2) le rapprochement des KS-III aux HMGS universelles dans les phylogénies (Figure 75) ; (3) la comparaison deux à deux des séquences (données non montrées). On peut donc proposer que les KS bactériennes soient issues des enzymes universelles par des duplications très anciennes dans ce domaine du

vivant. La neofonctionnalisation pourrait expliquer la forte divergence de ces séquences par rapport à leurs séquences mères, et cela d'autant plus que les KS sont des enzymes décarboxylatives qui utilisent principalement du malonyl-ACP alors que les HMGS et thiolases sont non décarboxylatives et utilisent l'acétyl-CoA comme brique fonctionnelle.

En résumé, ces résultats montrent que les archées modernes en général et leur dernier ancêtre commun sont (ou étaient) très probablement capables de (de)acétyler des acyl-CoA sans utiliser du malonyl-CoA ni la machinerie ACP, mais l'acétyl-CoA de façon non decarboxylative. Les machineries KS bactériennes ont probablement leur origine parmi ces séquences ancestrales, mais leurs séquences ont beaucoup divergé pour s'adapter à leur actuel fonctionnement par les machineries FAS que l'on connaît.

Les β -cétolacyl réductases

Dans les FAS-II, l'étape de condensation des chaînes acylées est suivie d'une première réduction, catalysée par une β -cétolacyl réductase, appelée KR ou FabG (Figure 72), [Price *et al.* 2001 ; Marrakchi *et al.* 2002b ; Parish *et al.* 2007], qui remplace la double liaison C=O par un groupement hydroxylé -OH. Cette enzyme est homologue de l'autre réductase qui intervient dans la voie de synthèse des acides gras, l'énoyl réductase (ER, FabI). En effet, il existe un grand nombre de dehydrogenases et réductases homologues qui ont été classées en trois groupes en fonction des motifs enzymatiques intervenant dans leurs réactions, de leur mécanisme réactionnel, de la taille des protéines et des comparaisons structurales [Kavanagh *et al.* 2008] : SDR, MDR et LDR (pour *short- / medium- / long-chain dehydrogenases/reductases*, respectivement). Ces superfamilles sont universelles et comportent une grande quantité d'enzymes chez tous les organismes, ce qui fait penser qu'elles sont très anciennes [Jornvall *et al.* 1999]. KR et ER appartiennent à la superfamille des SDR, qui regroupe des enzymes d'environ 250 acides aminés capables de réduire ou d'oxyder des substrats à l'aide d'un cofacteur NAD(P)(H) fixé à la protéine par un motif protéique d'interaction avec des nucléotides, appelé motif de Rossmann [Kallberg *et al.* 2002]. La partie N-terminale est celle qui fixe les cofacteurs et la partie C-terminale s'attache aux substrats, ce qui rend logiquement cette deuxième partie plus variable. Jusqu'à présent, aucune phylogénie concluante n'a été réalisée pour la superfamille SDR et la plupart des travaux se sont concentrés sur une classification des SDR en cinq groupes par rapport à des motifs protéiques conservés [Kallberg *et al.* 2002]. Le principal obstacle à l'étude phylogénétique exhaustive de cette famille est qu'elle contient beaucoup trop de représentants pour être faisable : en 2008 le

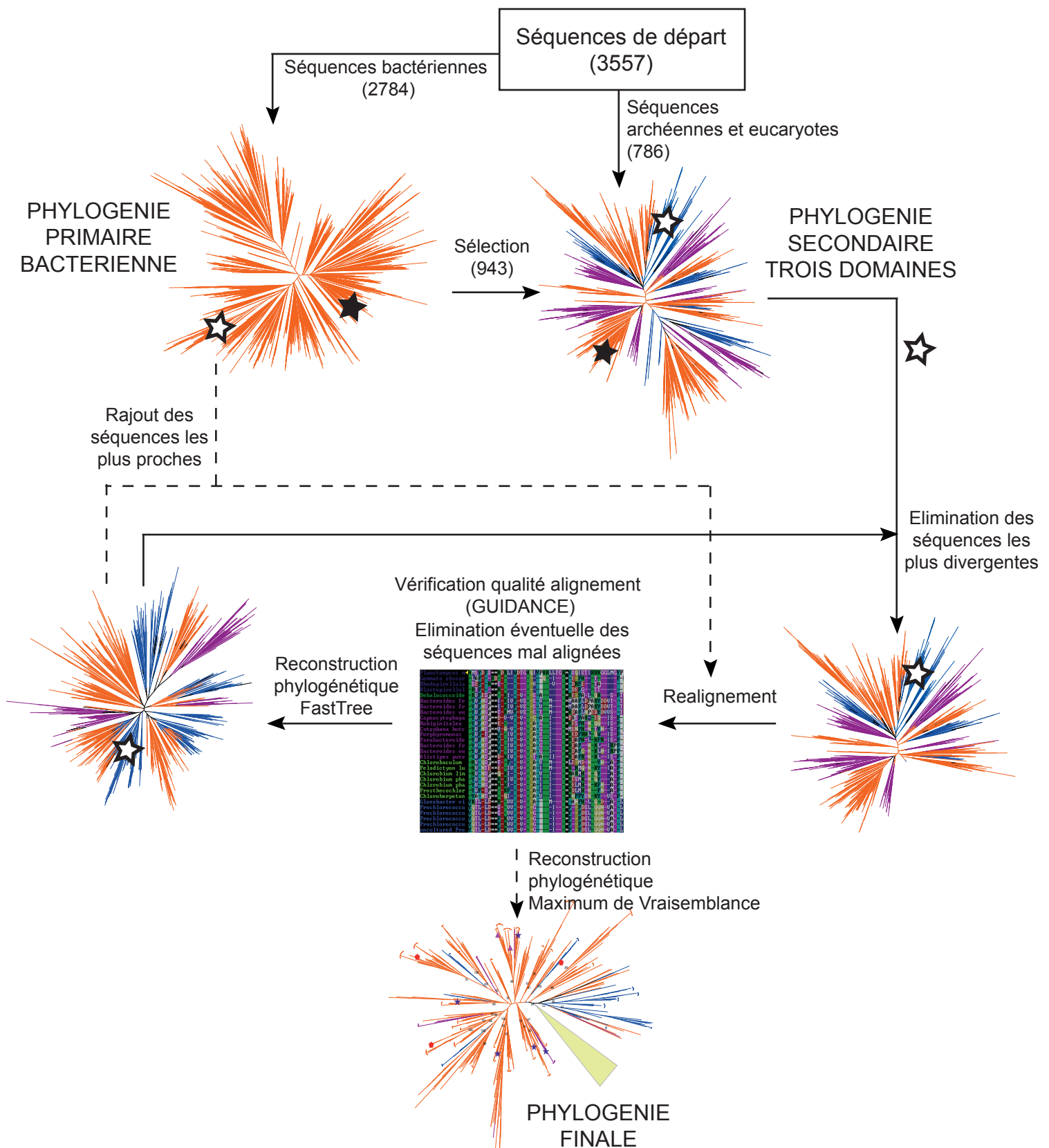


Figure 76. Procédure pour l'analyse des enzymes KR et ER de la superfamille des SDR.

Etapes : (1) Construction d'une phylogénie à partir de toutes les séquences bactériennes. Les étoiles représentent les endroits de l'arbre où groupent les séquences KR et ER. (2) Sélection aléatoire des séquences représentatives de l'ensemble de la topologie bactérienne et fusion de ces séquences avec l'ensemble des données des archées et des eucaryotes. Début des étapes itératives : Pour chaque enzyme séparément, (3) élimination des séquences les plus divergentes. (4) Réalignement et vérification de la qualité de l'alignement avec GUIDANCE. (5) Nouvelle reconstruction phylogénétique. Répétition des étapes 3-5 jusqu'à l'obtention d'uniquement le groupe fonctionnel d'intérêt et un groupe extérieur proche. (6) Rajout des séquences les plus proches des séquences bactériennes de départ. (7-8) Comme étapes 4-5, mais reconstruction phylogénétique finale par maximum de vraisemblance.

nombre d'enzymes différentes dans cette famille a été estimé à 140 et le nombre de séquences présentes dans les bases de données de l'époque à plus de 20000 [Kavanagh *et al.* 2008].

Une première phylogénie contenant toutes les séquences obtenues (3557) ne permettait pas d'identifier des groupes clairs ni sur un critère taxonomique ni sur un critère fonctionnel (données non montrées). J'ai donc réalisé une phylogénie contenant uniquement des bactéries (2784 séquences, Figure 76) qui permettait de voir que les séquences des gènes connus pour réaliser les fonctions d'intérêt (KR et ER) formaient deux groupes distincts dans l'arbre. Cela signifie que l'information dans l'alignement semble suffisante pour séparer les séquences par rapport à leur fonction, bien que les relations phylogénétiques entre les séquences ne soient pas fiables. J'ai donc sélectionné aléatoirement des séquences représentatives de cet arbre primaire bactérien (Figure 76), que j'ai regroupées avec les séquences des autres domaines du vivant (1729 séquences dont 943 bactériennes). Dans cet arbre secondaire complet on pouvait également observer que les séquences de même fonction selon la littérature groupaient ensemble. A partir de cet arbre j'ai éliminé les séquences les plus distantes de chaque fonction enzymatique d'intérêt (dans cette section, KR), ainsi que les séquences dont l'alignement était ambigu selon le logiciel GUIDANCE [Penn *et al.* 2010]. La reconstruction d'un nouvel arbre permet de déterminer à nouveau quelles sont les séquences les plus distantes de mon groupe d'intérêt, et la répétition de la procédure permet au fur à mesure de ne sélectionner que les séquences le plus proches de KR et un petit groupe extérieur. Une fois les séquences divergentes élaguées, j'ai enrichi le sous-ensemble obtenu avec les séquences les plus proches de l'arbre primaire bactérien. Une dernière analyse avec GUIDANCE a permis d'enlever toutes les séquences et les positions dont l'alignement était ambigu. Le résultat est un alignement de 569 séquences avec 187 positions conservées, qui a servi à reconstruire la phylogénie finale des séquences KR (Figure 77). Les arbres ER ont été construits de la même manière mais seront traités dans le paragraphe correspondant.

Dans la phylogénie des séquences KR, on peut observer qu'une grande partie des séquences bactériennes groupent entre elles suivant les principaux groupes taxonomiques bactériens, alors que d'autres séquences se placent de façon désorganisée. La présence de grands groupes bactériens semble indiquer qu'au moins un gène représentatif de cette fonction est ancestral à ce domaine du vivant ; les séquences qui ne s'accommodent pas aux groupes correspondants aux principaux taxons bactériens sont probablement le résultat de multiples paralopies et HGT plus récents. La présence de paralogues issus de duplications relativement récentes de ce gène n'est pas surprenant dans la mesure que les KR peuvent

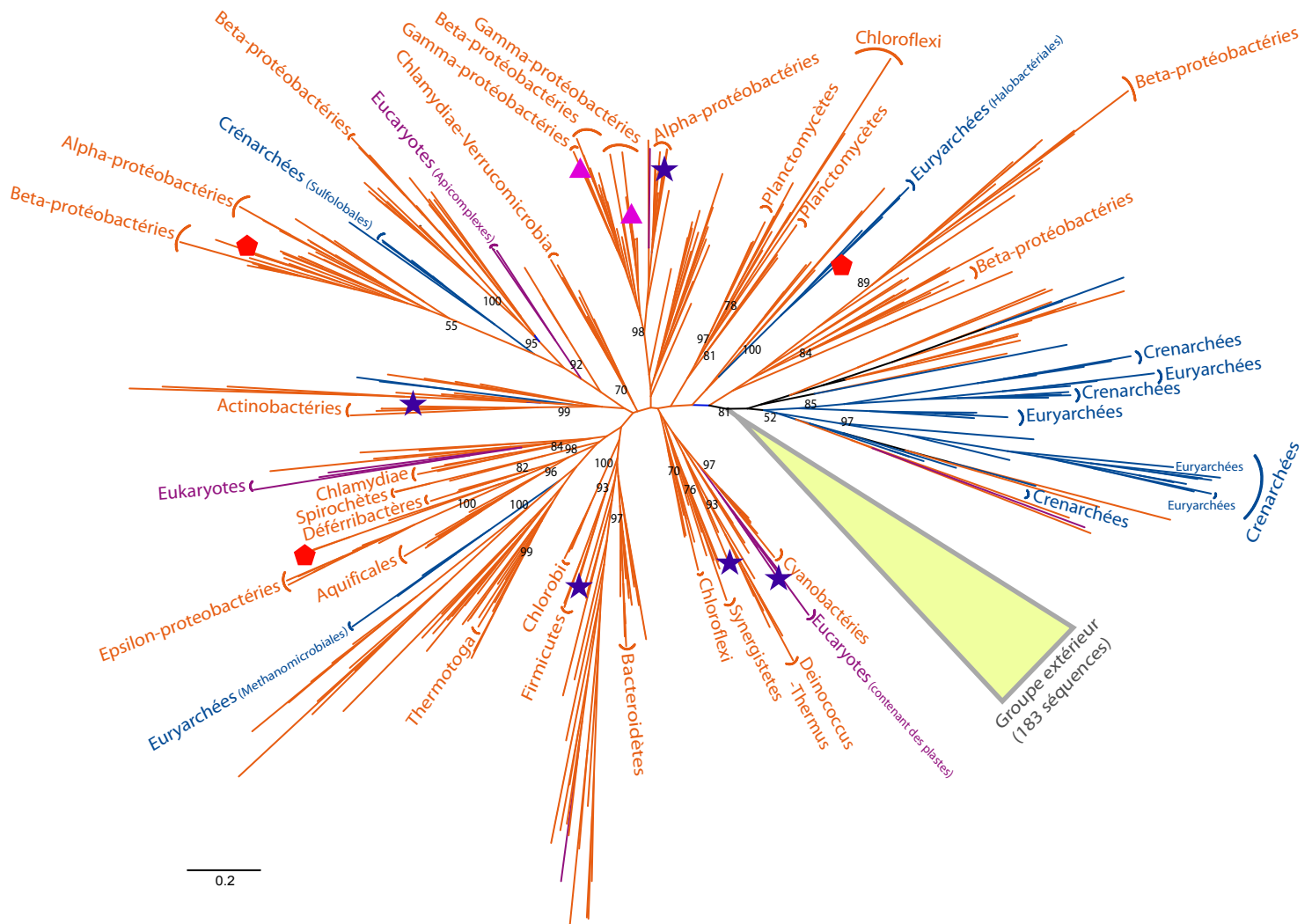


Figure 77. Phylogénie de KR.

Reconstruction par maximum de vraisemblance utilisant 569 séquences représentatives et 187 positions conservées. Les nombres sont les valeurs de robustesse calculées par TreeFinder. Les étoiles signalent les protéines caractérisées pour leur activité KR, les pentagones les protéines caractérisées pour leur activité PhaB et les triangles les protéines capables de catalyser les deux réactions.

participer à de très nombreuses fonctions cellulaires et métaboliques indépendantes de la synthèse des acides gras, comme la fabrication de glycolipides, rhamnolipides, acides mycoliques, acyl-homoserine lactone, etc. [Campos-García *et al.* 1998 ; Hoang *et al.* 2002 ; Lai et Cronan 2004].

Une fonction remarquable des homologues proches de KR est la fabrication de polyhydroxyalkanoates (PHA), qui sont des polymères d'hydroxyacyl-CoA et autres molécules similaires qui servent au stockage d'énergie et de carbone chez les bactéries et les haloarchées [Rehm et Steinbuchel 1999 ; Taroncher-Oldenburg *et al.* 2000 ; Han *et al.* 2009]. Les enzymes homologues de KR qui réalisent cette fonction sont les acetoacetyl-CoA reductases (PhaB). Les PhaB peuvent être couplées à la dégradation des acides gras et font donc intervenir des thioesters de CoA et non pas d'ACP (Figure 78) [Taroncher-Oldenburg *et al.* 2000 ; Han *et al.* 2009], bien que certaines séquences soient reconnues capables d'utiliser les deux cofacteurs [Taguchi *et al.* 1999 ; Ren *et al.* 2000 ; Nomura *et al.* 2008]. En effet, certaines enzymes sont capables de participer à la fois à la synthèse des acides gras et des PHA, et la détermination du cofacteur favori semble dépendre de la mutation ponctuelle de très peu de positions [Zhang *et al.* 2003b ; Wang et Cronan 2004b], ce qui a certainement permis un changement fréquent de spécificité du thioester utilisé par ces enzymes [López-Lara et Geiger 2001 ; Pillai *et al.* 2003]. Cette facilité à passer d'un thioester à l'autre est observable dans les phylogénies, car les séquences PhaB se trouvent éparpillées parmi les KR bactériens ACP-dépendants de la synthèse des acides gras, indépendamment de la taxonomie des organismes (Figure 77). Dans leur ensemble, ces résultats soutiennent que les KR sont ancestrales aux bactéries et que leur spécificité pour un thioester particulier (ACP ou CoA) est encore soumise à changement.

En ce qui concerne les archées, seules quelques PhaB et enzymes à fonction inconnue ont été étudiées [Yamamura *et al.* 2008 ; Han *et al.* 2009]. Certaines séquences d'halobactériales et de sulfobactériales branchent parmi les KR-PhaB bactériens, ce qui suggère pour ces dernières des cas de HGT depuis des bactéries (Figure 77). En revanche, un ensemble contenant une plus grande diversité de séquences archéennes se situe à la base du groupe KR-PhaB dans toutes les phylogénies, donc on peut postuler que ces séquences catalysent des réactions similaires à leurs homologues bactériens. La reconstruction phylogénétique à partir de cet ensemble de séquences montre qu'il existe deux paralogues de ces séquences dont la phylogénie respecte les principaux taxons de crenarchées (Figure 79). Alors que certaines séquences d'euryarchées et de bactéries qui branchent parmi les

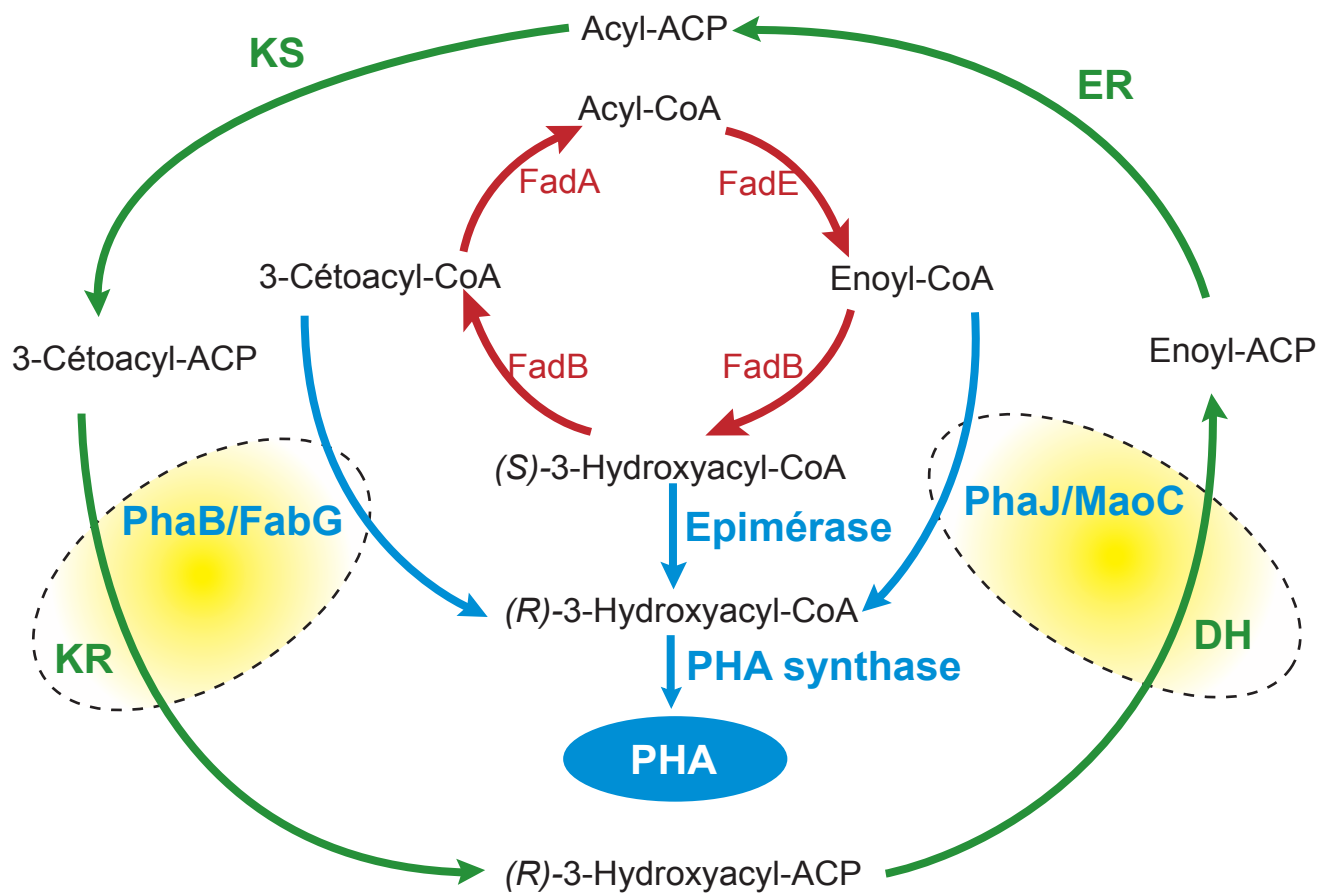


Figure 78. Similarités entre les voies de synthèse et de dégradation des acides gras et la synthèse de PHA.

Les textes en noir correspondent aux molécules intermédiaires et les textes en couleur indiquent les enzymes responsables des réactions. La dégradation des acides gras apparaît au centre, en rouge. La synthèse des acides gras est à l'extérieur, en vert. La synthèse de PHA est entre les deux, en bleu. Les ovals jaunes signalent des homologies remarquables entre les gènes de la synthèse des acides gras et ceux de la synthèse de PHA. La voie de synthèse des PHA partage avec la dégradation des acides gras les thioesters acyl-CoA mais utilise des gènes homologues de la machinerie de synthèse des acides gras.

crenarchées sont probablement le résultat d'HGT, l'existence d'un groupe basal d'euryarchées peut être un indice de la présence d'un de ces gènes non seulement chez le dernier ancêtre commun des crenarchées, mais aussi dans celui des archées en général.

En conclusion, la présence d'au moins un de ces gènes chez le dernier ancêtre des bactéries est très probable. L'utilisation de substrats liés soit à du CoA soit à de l'ACP dépend de mutations ponctuelles fréquentes dans la nature, ce qui plaide pour une versatilité importante de ces enzymes et fait penser que des métabolismes basés sur l'utilisation de thioesters de CoA est possible. L'existence d'un groupe de séquences archéennes inconnues très proches des KR-PhaB bactériennes suggère que ces séquences archéennes pourraient catalyser aussi la réduction de chaînes β -cétoacylées, bien que le type de thioester ne puisse pas être déterminé à cause du manque de caractérisation biochimique pour ces enzymes. La présence de deux paralogues des β -cétoacyl réductases hypothétiques chez les crenarchées semble indiscutable, et la présence dans l'ancêtre de toutes les archées ne peut pas être exclue.

Les hydroxyacyl déshydratases

Dans la FAS-II, la fabrication de 3-hydroxyacyl-ACP est suivie de la perte du groupe hydroxyl (-OH) catalysée par une déshydratase (DH) [Kass et Bloch 1967], amenant la production d'une molécule avec une double liaison entre carbones, l'enoil-ACP (Figure 72). Les DH appartiennent à une très grande famille de protéines, incluant des déshydratases de nombreux composés, de thioestérases qui hydrolysent le CoA ou l'ACP des chaînes acylées, des enzymes de dégradation moléculaire ou même des facteurs de transcription [Dillon et Bateman 2004 ; Pidugu *et al.* 2009]. Le point commun de toutes ces enzymes est qu'elles comportent un domaine protéique dit *Hot dog* : une hélice alpha (la « saucisse ») entourée à deux tiers par six feuillets beta antiparallèles (le « pain ») [Kimber *et al.* 2004]. Les domaines hot dog n'agissent jamais tous seuls ; la construction la plus simple contenant ce domaine exige au moins deux peptides identiques de hot dog qui se font face soit du côté « saucisse » soit du côté « pain », mais les tetramères, hexamères et les associations en polypeptides avec des domaines chargés de fonctions très différentes sont fréquentes [Pidugu *et al.* 2009]. De ce fait, alors que la structure tridimensionnelle est assez bien conservée dans tous les membres de la famille, la séquence primaire varie de façon importante selon la structure quaternaire de chaque protéine. La forte divergence entre ces enzymes est d'autant plus un obstacle pour l'étude comparative que le domaine est très court (environ 140 acides aminés).

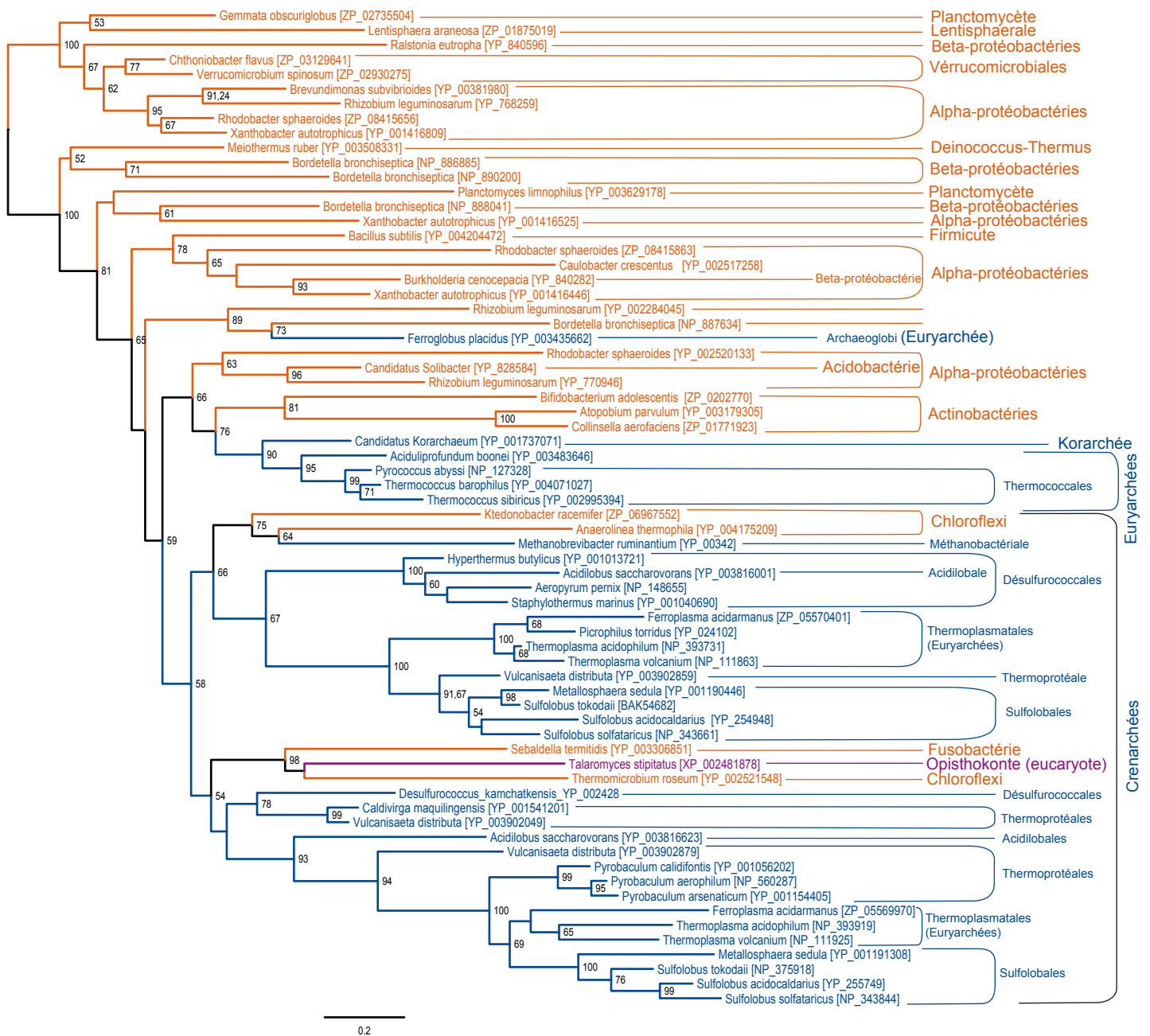


Figure 79. Phylogénie des séquences archéennes du groupe KR-PhaB et de leur plus homologues les plus proches.

Reconstruction par maximum de vraisemblance utilisant 68 séquences représentatives et 166 positions conservées. Les nombres correspondent aux valeurs de robustesse calculés par Tree-Finder.

La forte similarité tridimensionnelle a encouragé certains travaux à essayer d'établir une classification des protéines hot dog soit en fonction des distances deux à deux entre la séquence primaire de chaque protéine [Dillon et Bateman 2004], soit en réalisant des arbres phénétiques basés sur la structure générale des enzymes dont les structures cristallines sont disponibles [Pidugu *et al.* 2009]. Ces travaux permettent d'établir des groupes de similarité mais il est difficile d'établir la relation évolutive entre les différents groupes. En fait, pour détecter, à partir d'une enzyme donnée, les protéines des autres groupes de similarité il faut utiliser un grand nombre d'itérations de PSI-BLAST (données non montrées). L'alignement de séquences aussi divergentes et avec aussi peu de positions donne des alignements multiples inexploitable (non montrés). Par conséquent, la reconstruction d'une phylogénie représentative de l'ensemble de la superfamille hot dog ne semble pas réalisable avec les méthodes actuelles et j'ai concentré mon travail sur les séquences les plus proches de celles responsables de la fonction DH de la FAS-II bactérienne.

La première DH découverte chez *E. coli*, appelée FabA, a la particularité de ne pas être seulement responsable de la déshydratation du 3-hydroxyacyl-ACP, mais de jouer aussi un rôle d'isomérase [Kass et Bloch 1967]. En effet, la double liaison introduite par l'activité réductase de l'enzyme est, dans la plupart des cas, dans la configuration *trans* ; cependant, lorsque la chaîne acquiert une certaine taille, FabA isomérisé la double liaison et produit une configuration *cis*. Ceci est important car l'énoyl-ACP réductase qui suit normalement l'activité de la DH est spécifique des liaisons *trans* mais ne peut pas modifier les liaisons *cis*. De fait, l'élongation de la chaîne se poursuit (grâce à FabB, KS-I) malgré la double liaison et cela aboutit à l'incorporation d'une insaturation (double liaison) dans l'acide gras (Figure 80) [Leesong *et al.* 1996]. Plus récemment, une deuxième DH de la FAS-II, homologue de FabA, a été mise en évidence chez *E. coli* : FabZ [Mohan *et al.* 1994]. Contrairement à FabA, FabZ n'a pas d'activité isomérase, mais elle peut continuer d'allonger les chaînes acylées insaturées, chose que FabA est incapable de faire [Heath et Rock 1996]. Alors que certains travaux avaient proposé dans un premier temps que la différence d'activité isomérase était due à une mutation ponctuelle [Leesong *et al.* 1996], c'est en réalité l'angle formé par deux des feuilletés beta du « pain » qui laisse la place ou pas aux substrats d'adopter les deux conformations, *cis* ou *trans* [Kimber *et al.* 2004]. Il est donc impossible de déterminer en connaissant uniquement la séquence primaire d'une de ces enzymes si elles peuvent agir ou non comme des isomérases [Lu *et al.* 2005], et cette difficulté à déterminer la fonction d'une enzyme à partir de sa séquence primaire est générale aux protéines hot dog.

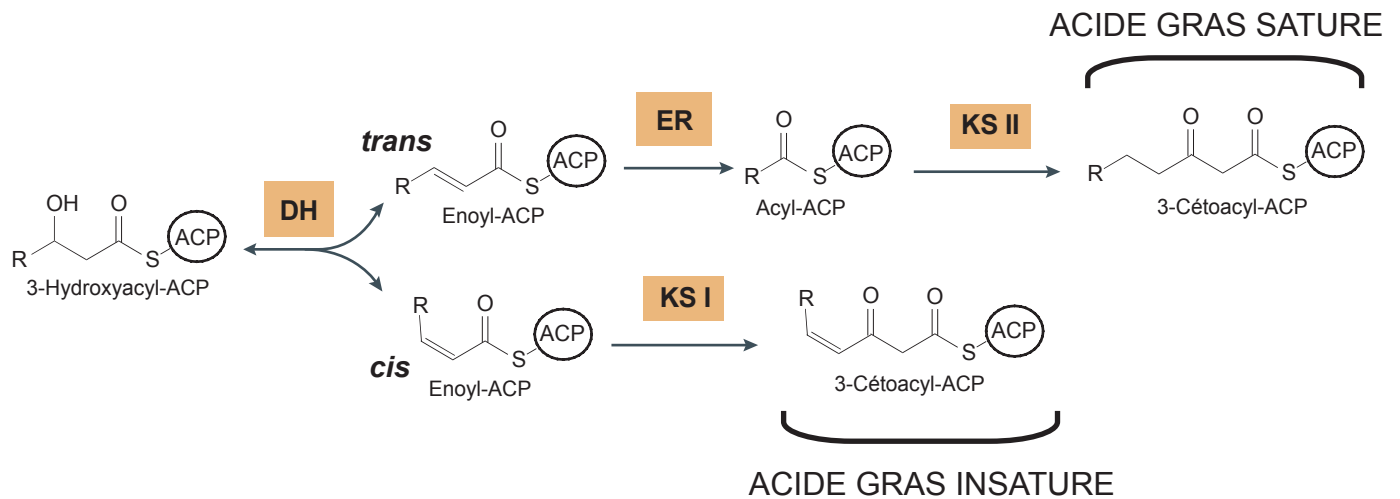


Figure 80. Synthèse d'acides gras saturés ou insaturés chez certaines bactéries.

Certaines DH ont, en plus de leur activité déshydratase, une activité isomérase qui peut changer la double liaison introduite par la déshydratation de la configuration *trans* à la configuration *cis*. A partir d'une certaine taille de l'acide gras en cours d'élongation, l'activité isomérase sera en concurrence avec l'enzyme ER, dont l'issue détermine le maintien ou pas de la double liaison dans la chaîne.

L'accumulation de caractérisations d'enzymes DH chez les bactéries a montré que le mécanisme FabA/FabZ semble caractéristique de certaines protéobactéries et que d'autres méthodes de synthèse des acides gras insaturés existent dans d'autres taxons [Zhu *et al.* 2009]. Il peut s'agir de la duplication et recrutement d'enzymes FabZ qui miment l'angle des feuilletts beta d'une FabA⁴⁸ [Wang et Cronan 2004a], du remplacement par des enzymes non homologues [Marrakchi *et al.* 2002a] ou l'introduction des insaturations non pas dans la synthèse *de novo*, mais dans des acides gras saturés déjà synthétisés [Altabe *et al.* 2003]. Cependant, j'ai étudié uniquement l'évolution de FabZ et ses homologues les plus proches, sans tenir compte des enzymes non homologues qui peuvent jouer un rôle dans cette étape.

La recherche d'homologues de FabA et FabZ dans les génomes d'archées ne révèle pas de séquences par une première recherche simple par BLAST et nécessite de faire appel à deux itérations de PSI-BLAST. Etant donnée la diversité des domaines hot dog citée précédemment, on pourrait s'inquiéter de la conservation fonctionnelle des séquences très divergentes obtenues par PSI-BLAST. Plusieurs arguments nous permettent de penser que cette analyse n'a détecté que les séquences les plus proches des FabZ et non pas d'enzymes très lointaines de cette superfamille dont la conservation de fonction pourrait être mise en cause : d'abord, les séquences obtenues, annotées généralement comme des MaoC ou PhaJ, s'alignent bien deux à deux avec des séquences représentatives des FabZ et ont des structures très proches [Hisano *et al.* 2003]. Ensuite, alors que peu d'itérations de PSI-BLAST sont nécessaires à détecter les MaoC-PhaJ depuis les FabZ et vice-versa, il faut beaucoup plus d'itérations pour détecter les autres protéines de la superfamille hot dog (données non montrées). Enfin, les activités enzymatiques des MaoC-PhaJ sont très similaires à celles de FabZ, les remplaçant dans certains cas [Hiltunen *et al.* 2010] et étant parfois capables de les compléter [Autio *et al.* 2008 ; Gurvitz *et al.* 2008 ; Sacco *et al.* 2010].

Les MaoC ont d'abord été décrites chez les bactéries dans un opéron d' amino oxidasés [Sugino *et al.* 1992] et il fut ensuite compris qu'elles participent à des nombreuses fonctions

⁴⁸ On peut se demander quelle est la différence entre une FabA et une FabZ avec une activité isomérase (qui a été appelée FabN dans *Enterococcus faecalis*). Dans la littérature, les DH sont classées comme FabA ou FabZ en fonction de la séquence qui leur est le plus similaire [Lu *et al.* 2005], mais il existe aussi des différences structurelles, car les FabA sont des dimères [Heath et Rock 1996 ; Swarnamukhi *et al.* 2006], alors que les FabZ sont des hexamères [Lu *et al.* 2005]. Quoiqu'il en soit, à ma connaissance aucune phylogénie n'avait pas été réalisée pour déterminer l'origine de FabN par rapport à celui de FabA, mais mes phylogénies préliminaires montrent que différentes protéines à activité isomérase branchent parmi les FabZ à différents endroits de l'arbre (données non montrées), soutenant l'origine différent des deux enzymes.

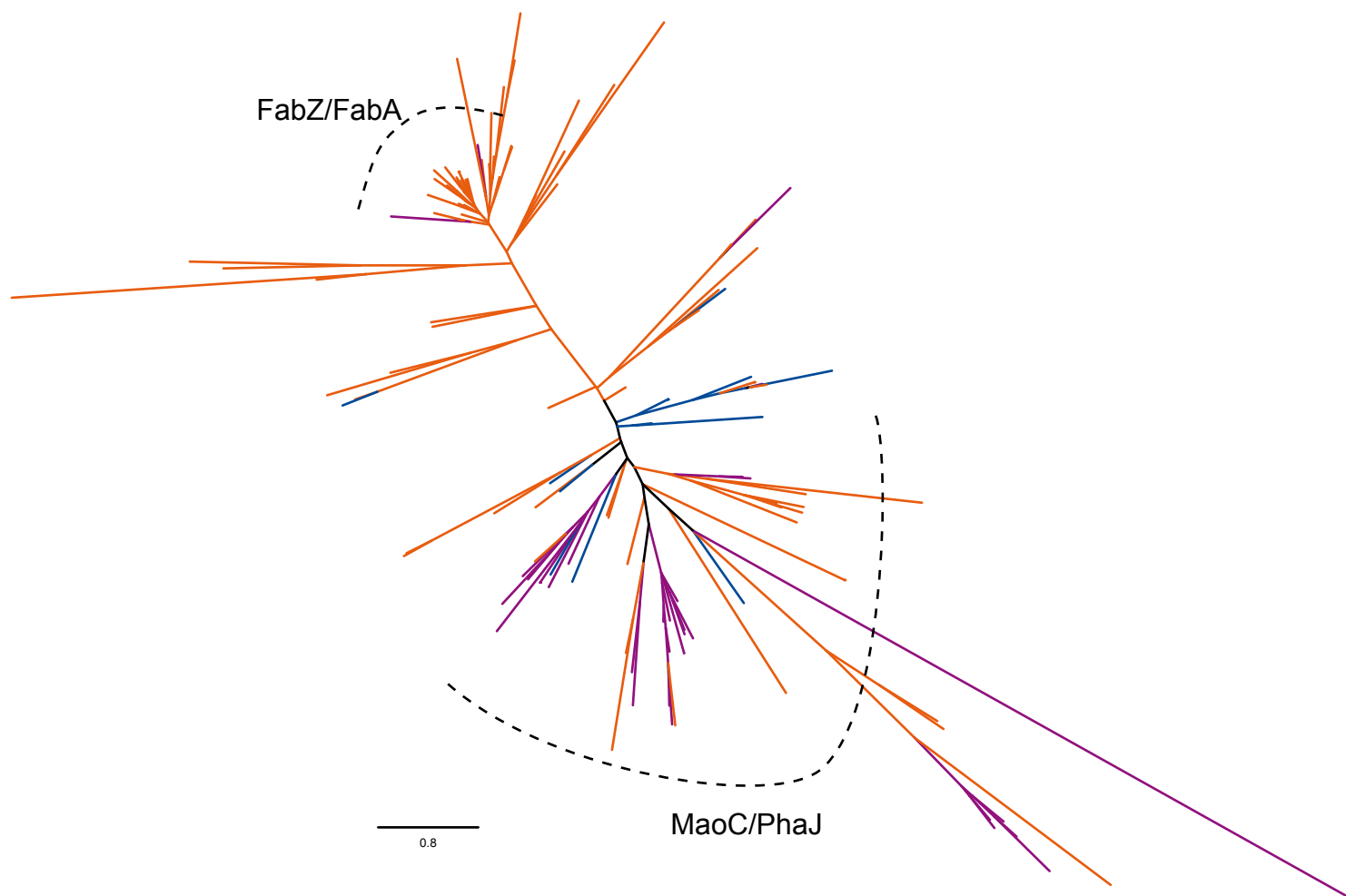


Figure 81. Phylogénie préliminaire contenant des séquences de type FabA/FabZ et MaoC/PhaJ.

Reconstruction avec FastTree utilisant 173 séquences représentatives et 73 positions conservées. Cette phylogénie révèle la grande divergence entre les deux groupes d'homologues.

[Tsuge *et al.* 2003a], dont la mieux connue est la synthèse des unités de base des PHA, pour laquelle elles reçoivent le nom PhaJ [Reiser *et al.* 2000 ; Tsuge *et al.* 2003b ; Wang *et al.* 2010]. PhaJ a un lien particulier avec les acides gras car elle utilise des intermédiaires de leur dégradation (enoyl-CoA) pour produire des (*R*)-3-hydroxyacyl-CoA nécessaires à la synthèse de PHA. Pour comparaison, la FAS-II fait intervenir des (*R*)-3-hydroxyacyl-ACP et enoyl-ACP et la β -oxydation passe par des intermédiaires (*S*)-3-hydroxyacyl-CoA (Figure 78). Autrement dit, FabZ et PhaJ utilisent des substrats très similaires dont la principale différence est la nature du thioester [Park et Lee 2003]. Il a d'ailleurs été montré que certaines enzymes sont capables d'utiliser à la fois des substrats liées à des CoA ou à de l'ACP [Moskowitz et Merrick 1969 ; Gurvitz *et al.* 2009 ; Sacco *et al.* 2010].

Les MaoC présentent une grande diversité d'architectures possibles [Johansson *et al.* 2006], représentative des nombreuses structures des protéines hot dog. Notamment, certaines protéines peuvent contenir deux domaines hot dog d'affilé (dont généralement un des deux est très dérivé et ne conserve plus la fonction d'origine, mais a une fonction stabilisatrice, [Castell *et al.* 2005 ; Sacco *et al.* 2007b ; Sacco *et al.* 2010]). D'autres ont subi une scission à partir d'un tel domaine double [Sacco *et al.* 2007a], ou peuvent encore se trouver dans des polypeptides à fonctions diverses [Qin *et al.* 2000 ; Koski *et al.* 2004 ; Koski *et al.* 2005]. Malgré ces structures diverses, les fonctions sont souvent conservées, comme le montre la restauration d'une mutation de DH de FAS-II par une protéine de type MaoC-PhaJ contenant un double domaine hot dog [Autio *et al.* 2008].

L'analyse phylogénétique de ces séquences est très difficile du fait de leur divergence et de leur courte taille. Les FabZ sont répandus chez les bactéries mais une longue branche les sépare du reste de MaoC-PhaJ (Figure 81). Ceci est probablement dû aux contraintes de la machinerie FAS-II, qui est très bien conservée, et probablement ancestrale, chez les bactéries. Cependant, aucune archée ne branche parmi les FabZ, donc nous avons éliminé ce groupe de nos analyses. Ensuite, nous avons enrichi progressivement l'alignement résultant avec d'autres séquences de MaoC-PhaJ tout en éliminant les séquences les plus divergentes qui dégradaient fortement la qualité de l'alignement. Une fois que toutes les séquences MaoC ont été intégrées, on a éliminé les groupes correspondant aux doubles domaines (il est intéressant de signaler que plusieurs origines différentes semblent avoir existé, ce qui confirme les hypothèses proposées dans la littérature) et aux polypeptides, que l'on suspectait d'introduire encore trop de bruit dans l'alignement. L'arbre résultant montre que les séquences archéennes qui peuvent être classées dans ce groupe sont assez nombreuses (Figure 82). Etant données

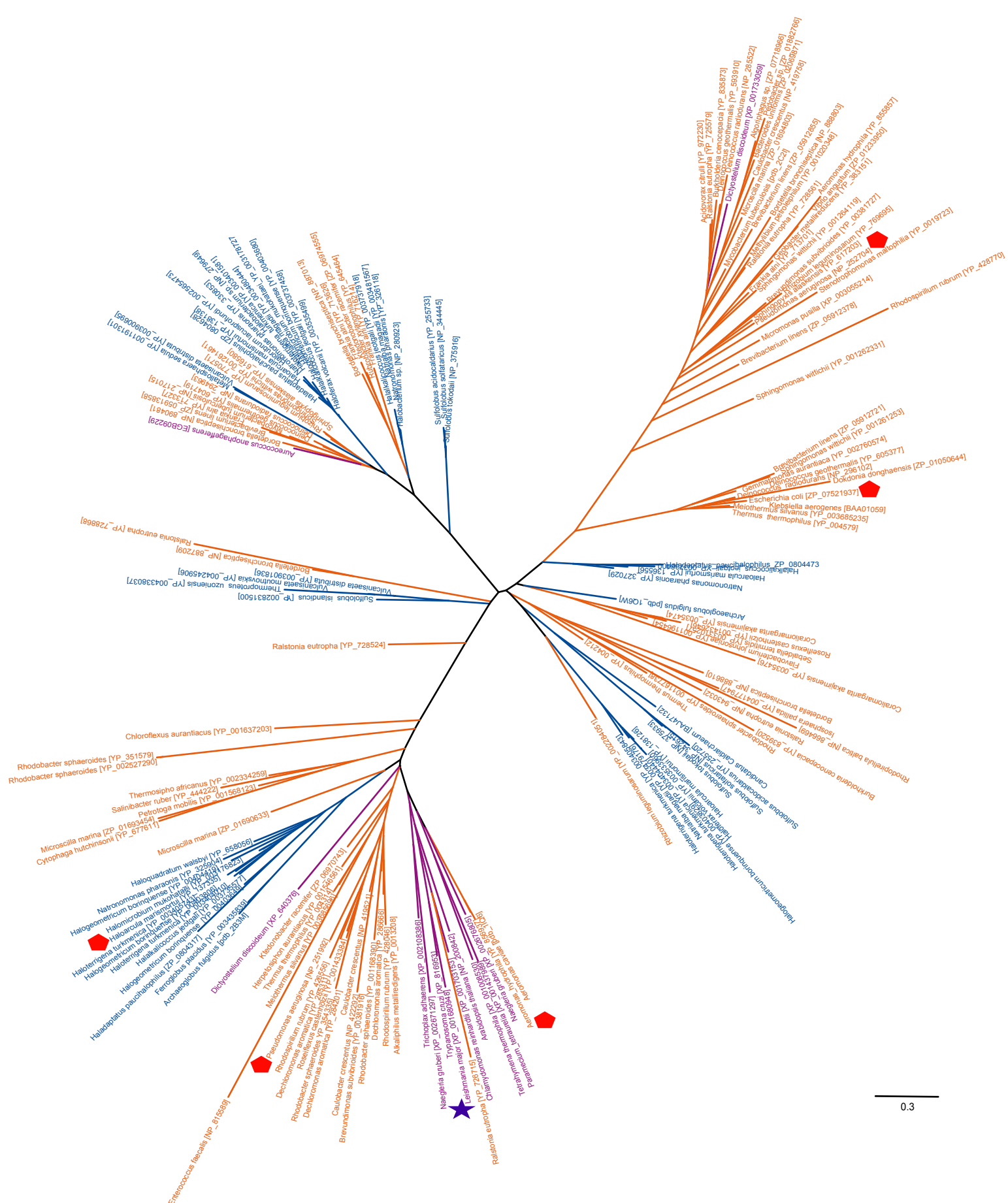


Figure 82. Phylogénie de MaoC.

Reconstruction par maximum de vraisemblance utilisant 163 séquences représentatives et 112 positions conservées. Les pentagones signalent des protéines qui utilisent des thioesters de CoA et l'étoile une protéine capable d'utiliser des thioesters d'ACP.

leurs positions, il peut être proposé qu'elles puissent réaliser une déshydratation similaire à celle des FAS-II bactériennes, mais la caractérisation biochimique est indispensable pour confirmer cela. Néanmoins les relations phylogénétiques entre les différents taxons ne semblent pas fiables à cause de la forte divergence des séquences et le petit nombre de positions conservées dans l'alignement. Pour l'instant il est donc impossible de déterminer si ces séquences sont ancestrales aux archées ou le résultat de nombreux HGTs.

Les enoyl reductases

La dernière étape de chaque tour de la FAS-II (sauf si une insaturation en conformation *cis* est présente), est la réduction de la double liaison *trans* de l'enoyle-ACP en acyl-ACP (Figure 72) par une enoyl-réductase, ER [Massengo-Tiasse et Cronan 2009]. Il existe une grande diversité d'ER, homologues ou non homologues entre elles, chez les bactéries et les eucaryotes. Comme vu précédemment, les KR et une grande majorité des ER appartiennent à la superfamille des SDR : (1) FabI est l'ER dont l'activité a été décrite chez une plus grande diversité d'organismes [Heath et Rock 1995 ; Rozwarski *et al.* 1998 ; Heath *et al.* 2000a ; Hoffmeister *et al.* 2005] ; (2) FabL est une SDR détectée chez *Bacillus subtilis*, assez divergente par rapport à FabI et avec certaines caractéristiques structurales qui la rapprochent de FabG (KR), mais comportant une activité enoyl-ACP reductase [Heath *et al.* 2000b ; Kim *et al.* 2010] ; (3) FabV est aussi une ER de la superfamille SDR, très divergente à la fois de FabI et FabL, et décrite uniquement dans le genre *Vibrio* dont elle semble caractéristique [Massengo-Tiasse et Cronan 2008]. En dehors de la famille SDR, les ER des mitochondries des opisthokontes appartiennent à la superfamille MDR [Airenne *et al.* 2003 ; Miinalainen *et al.* 2003] et il existe chez des firmicutes une ER appelée FabK contenant un domaine protéique dit tonneau TIM (*TIM barrel*) sans lien de parenté avec les autres ER, mais capable de compléter leur fonction [Marrakchi *et al.* 2003]. Finalement, les FAS-I de champignons utilisent une ER de type FabK [Jenni *et al.* 2007], alors que les FAS-I des animaux utilisent un autre domaine MDR sans lien direct avec l'ER utilisé dans la mitochondrie [Maier *et al.* 2008]. Il a été proposé que la grande diversité des ER, par contraste avec la relative conservation des autres enzymes des FAS, soit liée à l'existence d'antibiotiques naturels qui ciblent l'une ou l'autre de ces ER [Massengo-Tiasse et Cronan 2009].

Etant donnée le foisonnement d'ER différentes, nous n'avons étudié que les homologues les plus proches de FabI. Ce choix est dû au fait que ce gène est l'archétype

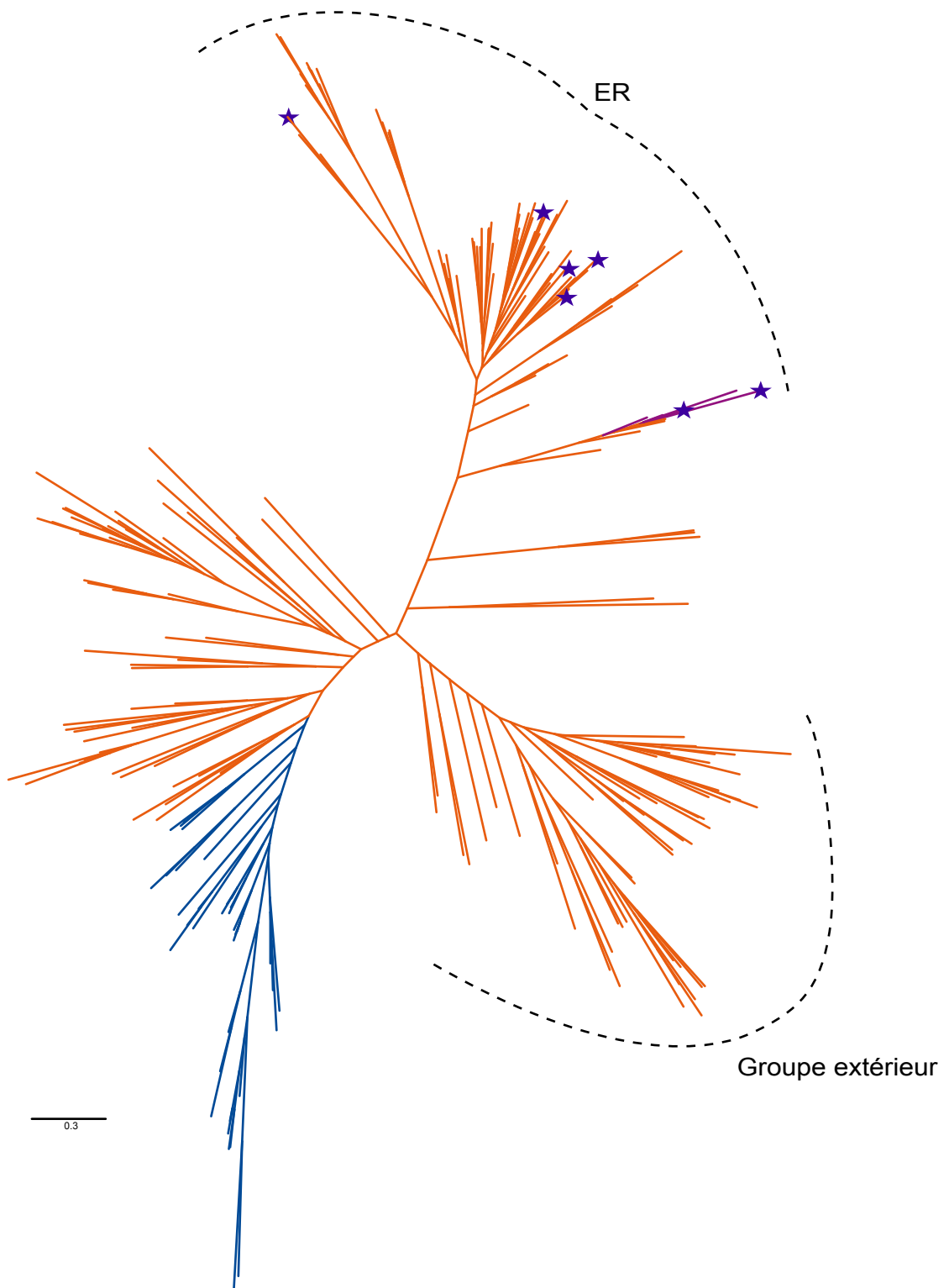


Figure 83. Phylogénie d'ER.

Reconstruction par maximum de vraisemblance utilisant 222 séquences représentatives et 173 positions conservées. Les étoiles signalent des protéines capable d'utiliser des thioesters d'ACP.

classique des FAS-II bactériennes, à sa plus grande distribution taxonomique parmi les bactéries et au fait que les phylogénies intermédiaires (arbre secondaire complet, voir la partie sur les KR) signalaient un groupe de séquences archéennes à la base des FabI bactériennes, suggérant par conséquent une conservation fonctionnelle. La construction de la phylogénie a été faite selon la même procédure que celle détaillée dans la section sur l'évolution des KR.

Après application de cette méthode d'affinage, la phylogénie finale de FabI montre la monophilie des séquences bactériennes connues pour réaliser cette activité et le respect des grands groupes bactériens au sein de ce groupe (Figure 83). Il existe aussi un groupe de séquences bactériennes et archéennes non caractérisées et distinctes à la fois des FabI bactériennes et des séquences représentatives du groupe extérieur. Lorsqu'on étudie en détail la phylogénie des séquences archéennes (Figure 84) on s'aperçoit d'une topologie très similaire à celle des KR (Figure 79). On observe deux groupes frères représentatifs de la diversité des crenarchées parmi lesquels branchent quelques séquences euryarchées. A la base des deux groupes crenarchées on observe un autre groupe d'euryarchées, mais leur recouvrement de la diversité connue de ce phylum est moins important que pour celui des crenarchées. L'interprétation de ces résultats doit donc être similaire à celle des KR : (1) la proximité des séquences euryarchéennes aux séquences FabI fait penser que leur fonction pourrait être similaire ; (2) il existe deux paralogues chez les crenarchées qui sont probablement le résultat d'une duplication avant l'ancêtre de ce phylum ; (3) les euryarchées comme les thermoplasmatales qui branchent parmi les crenarchées sont connues pour avoir reçu des nombreux HGT [Ruepp *et al.* 2000], ce qui semble avoir été le cas également pour ces séquences ; (4) néanmoins, la présence du groupe basal de séquences euryarchées suggère que ces organismes n'ont pas reçu leurs séquences par HGT, et par conséquent on ne peut pas exclure la possibilité que le dernier ancêtre des archées ait également possédé au moins une copie de ce gène, qui aurait ensuite été perdu dans une grande partie des euryarchées actuelles.

Finalement, nous avons vu pour les autres enzymes de la FAS-II qu'elles étaient souvent capables d'utiliser des thioesters de CoA autant que d'ACP. En ce qui concerne les ER, il a été montré que plusieurs de ces enzymes sont capables d'utiliser des thioesters de CoA [Das *et al.* 2000 ; Miinalainen *et al.* 2003 ; Hoffmeister *et al.* 2005 ; Tucci et Martin 2007], mais à ma connaissance ceci n'a été déterminé à ce jour pour aucune FabI. Cependant, cette possibilité ne peut pas être écartée car la plupart des ER utilisant des substrats liés au CoA appartiennent, comme les FabI, à la superfamille des SDR.

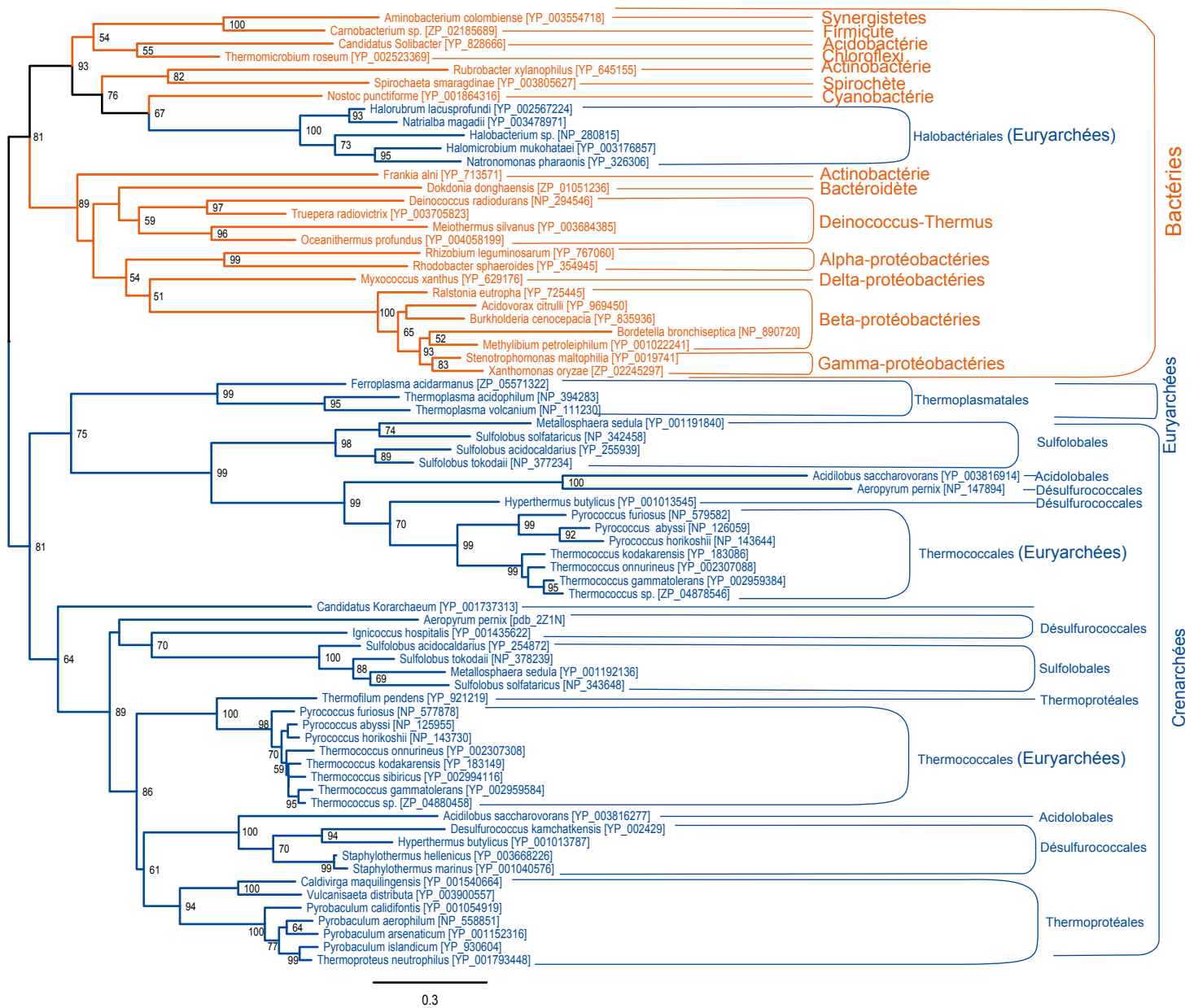


Figure 84. Phylogénie des séquences archéennes du groupe ER et de leur plus homologues les plus proches.

Reconstruction par maximum de vraisemblance utilisant 73 séquences représentatives et 192 positions conservées. Les nombres robustes dépendent aux valeurs de robustesse calculés par Tree-Finder.

3. Une synthèse d'acides gras indépendante de l'ACP ?

Synthèse des acides gras chez les archées

L'analyse phylogénomique des gènes de la FAS-II bactérienne chez les archées montre que la plupart de ces gènes ont des homologues à fonction inconnue dans une diversité remarquable d'archées. Plus spécifiquement, l'absence d'homologues de la machinerie ACP dans les génomes archéens contraste avec la présence des homologues du reste de la voie. Comme dit précédemment, notre incapacité à détecter les gènes de la machinerie ACP pourrait être due à la forte divergence de ces séquences, d'autant plus que certains travaux ont cru pouvoir déterminer la présence d'ACP dans certaines archées [Pugh *et al.* 1971 ; Pugh et Kates 1994]. Cependant, le fait que les homologues archéens des KS, KR et DH soient étroitement liés à des protéines bactériennes ou eucaryotes capables d'utiliser des thioesters de CoA au lieu de ceux d'ACP soutient préférentiellement la possibilité que, si une telle voie de synthèse des acides gras existe chez les archées, elle pourrait être indépendante de l'ACP. Comme chez les bactéries et les eucaryotes, cette voie serait une spirale avec quatre étapes (Figure 85) : (1) la condensation de deux unités d'acyl-CoA soit par une HMGS, soit par une thiolase. Si l'enzyme responsable est la HMGS, l'utilisation de substrats de tailles très différentes à chaque tour d'élongation des acides gras pourrait être à l'origine d'une grande divergence de ces séquences qui expliquerait la longue branche qui sépare les séquences archéennes des autres organismes dans la phylogénie de HMGS (Figure 45, p. 131) ; (2) la réduction du groupement cétone par des séquences similaires aux KR-PhaB bactériennes ; (3) la déshydratation du groupement hydroxyle par une enzyme du groupe MaoC-PhaJ ; (4) la réduction de l'insaturation par une ER similaire aux FasI bactériennes.

A ce jour, aucune FAS basée sur des substrats liés au CoA n'est connue, mais un tel mécanisme a déjà été proposé pour certaines bactéries qui semblent être dépourvues de la FAS-II classique [Tucci et Martin 2007]. Dans ce cas, il a été proposé que la synthèse des acides gras soit très similaire de celle des précurseurs des PHA, ce qui pourrait être aussi le cas chez les archées étant donné le rapprochement des séquences archéennes aux séquences liées à la synthèse des PHA dans certaines de nos phylogénies (KR, DH). Le travail de Tucci et Martin proposait également que la brique de base de la synthèse des acides gras sans ACP pouvait être l'acetyl-CoA au lieu du malonyl-ACP, comme dans la synthèse des PHA. Ceci semble également possible dans le cas de l'hypothétique FAS archéenne, car la HMGS et les thiolases sont des enzymes non-décarboxylatives qui utilisent l'acetyl-CoA au lieu du

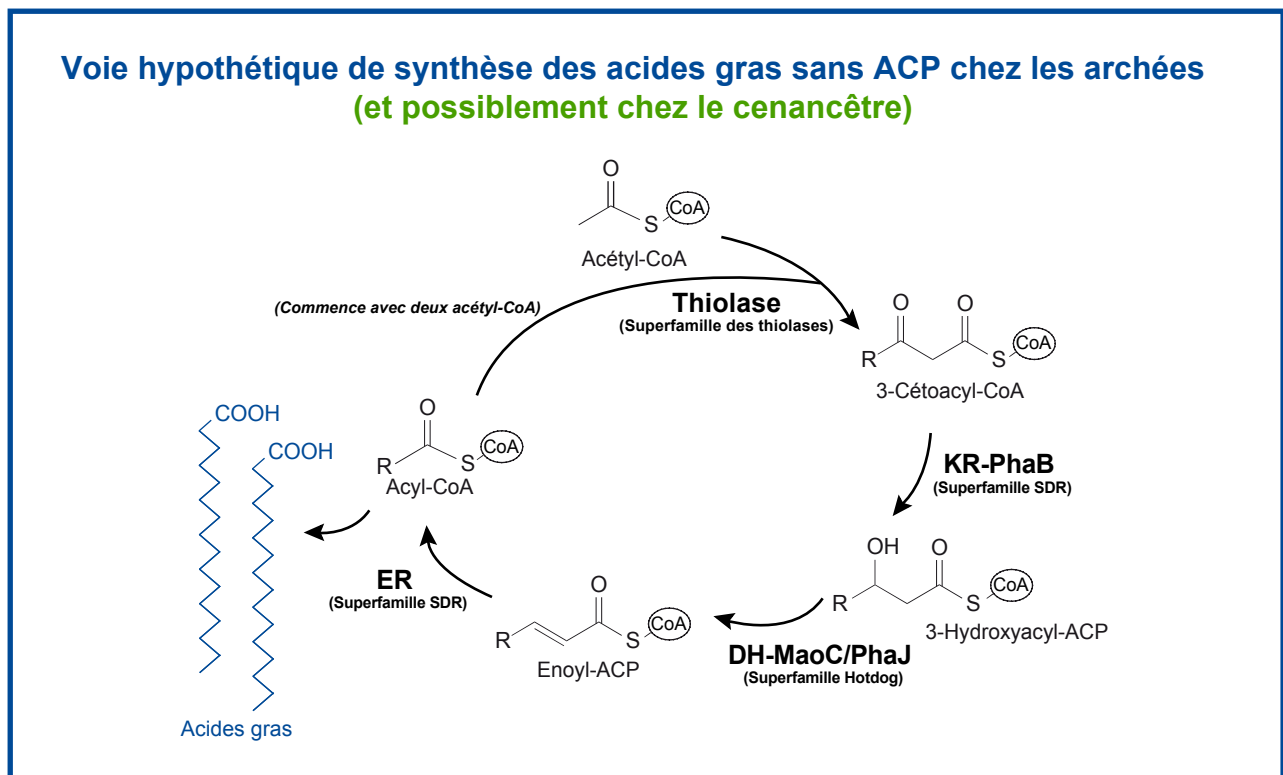
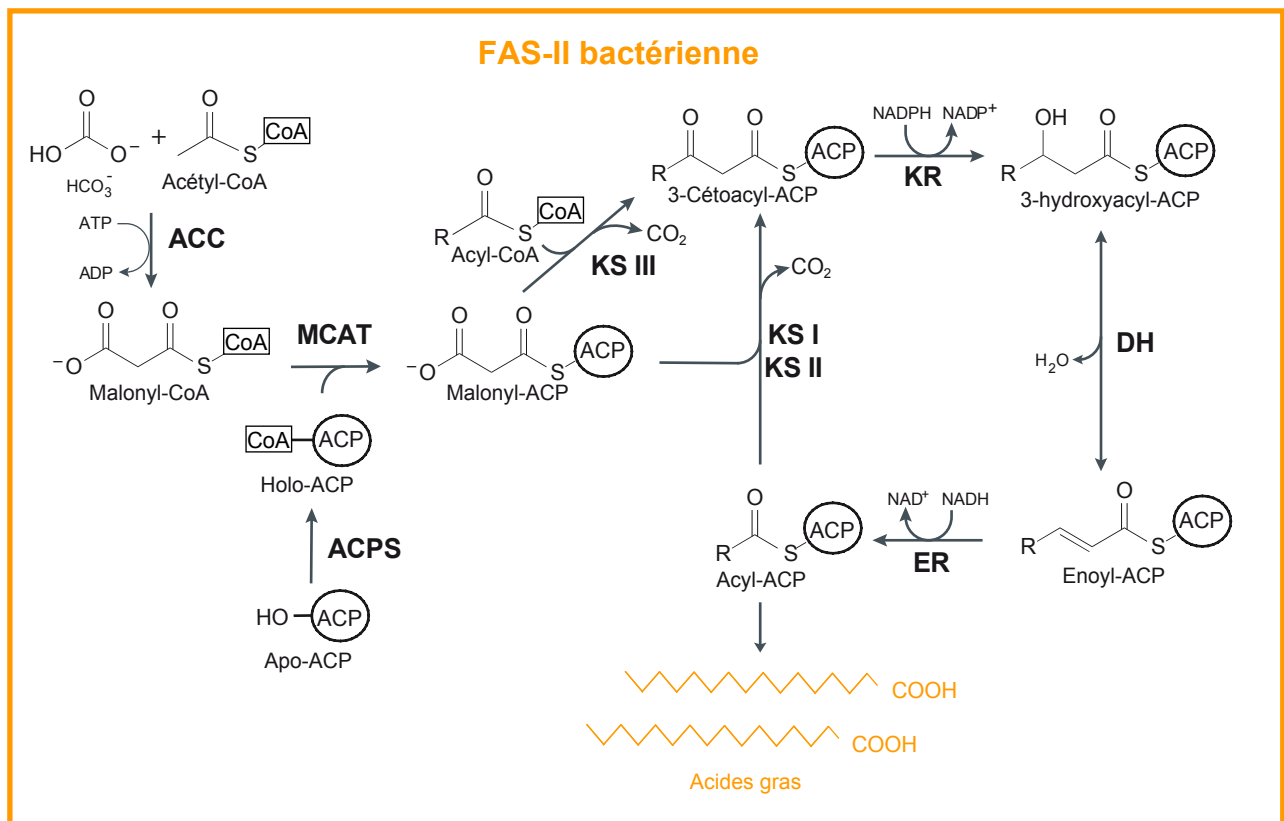


Figure 85. Comparaison entre la FAS-II bactérienne classique et la voie archéenne proposée sur la base de l'analyse phylogénomique.

malonyl-CoA. La vérification ou réfutation biochimique de toutes ces hypothèses sera donc très importante non seulement pour comprendre la biologie des archées et l'origine de leurs acides gras mais aussi pour nous renseigner sur des voies de synthèse plus larges présentes dans des organismes des autres domaines du vivant.

La description d'une FAS archéenne dépendante du CoA au lieu de l'ACP pourrait également apporter des informations différentes sur les règles qui régissent la reconnaissance entre les intermédiaires d'élongation des acides gras et les enzymes de la spirale de synthèse. En effet, pendant longtemps il y a eu un débat sur le rôle de l'ACP dans cette interaction : soit l'ACP se fixe au hasard aux enzymes de la FAS-II, soit elle reconnaît chaque enzyme spécifiquement [Zhang *et al.* 2003b]. Des résultats récents ont montré que l'ACP adopte des conformations différentes en fonction de l'état de réduction du substrat qu'elle porte, et qui lui permettent de donner son substrat uniquement à l'enzyme attendue pour chaque étape de la spirale de synthèse [Ploskon *et al.* 2010]. Cependant, il semble raisonnable de penser qu'une machinerie qui ne comporterait pas d'ACP serait obligée d'utiliser les interactions au hasard entre les métabolites intermédiaires et les enzymes, ce que pourrait se traduire dans une plus faible efficacité d'un mécanisme sans ACP par rapport à la FAS-II. Bien que cela reste hypothétique, la grande efficacité des FAS utilisant l'ACP pourrait expliquer la plus grande importance des acides gras dans les membranes des bactéries et des eucaryotes, alors que les archées, qui ont privilégié l'usage d'isoprénoïdes comme chaînes latérales de leurs phospholipides, ont une voie plus adaptée aux plus faibles quantités d'acides gras nécessaires pour les autres tâches de ces molécules ([Lombard *et al.* 2012a], voir annexe 5).

Implications pour le cenancêtre

Une voie de synthèse des acides gras basée sur des thioesters de CoA peut donc être proposée pour les archées. Les gènes qui ont été postulés dans cette voie sont présents dans une assez large diversité d'archées. La présence d'HMGS et d'une thiolase chez le dernier ancêtre des archées semble très probable, et celle de KR et ER semble également possible. L'existence d'une DH chez l'ancêtre des archées est plus difficile à établir à cause de l'alignement peu fiable de cette enzyme, mais elle ne peut pas être exclue non plus. De fait, on peut proposer une FAS basée sur le CoA non seulement dans les archées actuelles mais aussi dans leur ancêtre (Figure 85). En ce qui concerne la FAS-II bactérienne, il est généralement admis qu'elle a été héritée au moins depuis le dernier ancêtre des bactéries et nos résultats confirment cette possibilité malgré la difficulté à interpréter dans le détail la

phylogénie de chaque gène. Il est d'ailleurs important de noter que les séquences bactériennes et celles des archées forment systématiquement des groupes frères, suggérant que les domaines procaryotes ont hérité ces gènes du cenancêtre, et permettant de ce fait de proposer que cet organisme pouvait être capable de synthétiser des acides gras.

Etant donné qu'une machinerie similaire à celle des archées existe chez les bactéries (la synthèse des PHA, si ce n'est une synthèse d'acides gras) mais que l'ACP est absente chez les archées, il semble raisonnable de proposer que la synthèse d'acides gras chez le cenancêtre se faisait par une FAS sans ACP, et que cette dernière est apparue dans la lignée bactérienne. Ce résultat est intéressant parce qu'il donne aussi une possible explication de la primauté des acides gras dans les membranes bactériennes : alors que la présence d'isoprénoïdes dans les membranes d'archées est souvent justifiée comme une adaptation à la haute température des ancêtres de la lignée archéenne, l'apparition de la machinerie ACP pourrait expliquer pourquoi les bactéries utilisent très majoritairement les acides gras dans leurs phospholipides plutôt que d'avoir conservé une membrane mixte dont on peut supposer la présence chez le cenancêtre.

Enfin, il est intéressant de signaler que les premières étapes de synthèse des isoprénoïdes, qui sont les étapes les mieux conservées de la voie du mévalonate dans tous les organismes, sont également très proches des étapes d'élongation des polycétides. Non seulement les étapes sont identiques, mais elles font intervenir la même superfamille, celle des thiolases. Il semble donc probable que les deux voies aient partagé une période d'évolution commune dans un passé très lointain, puis qu'elles aient divergé quelque temps avant l'émergence du cenancêtre.

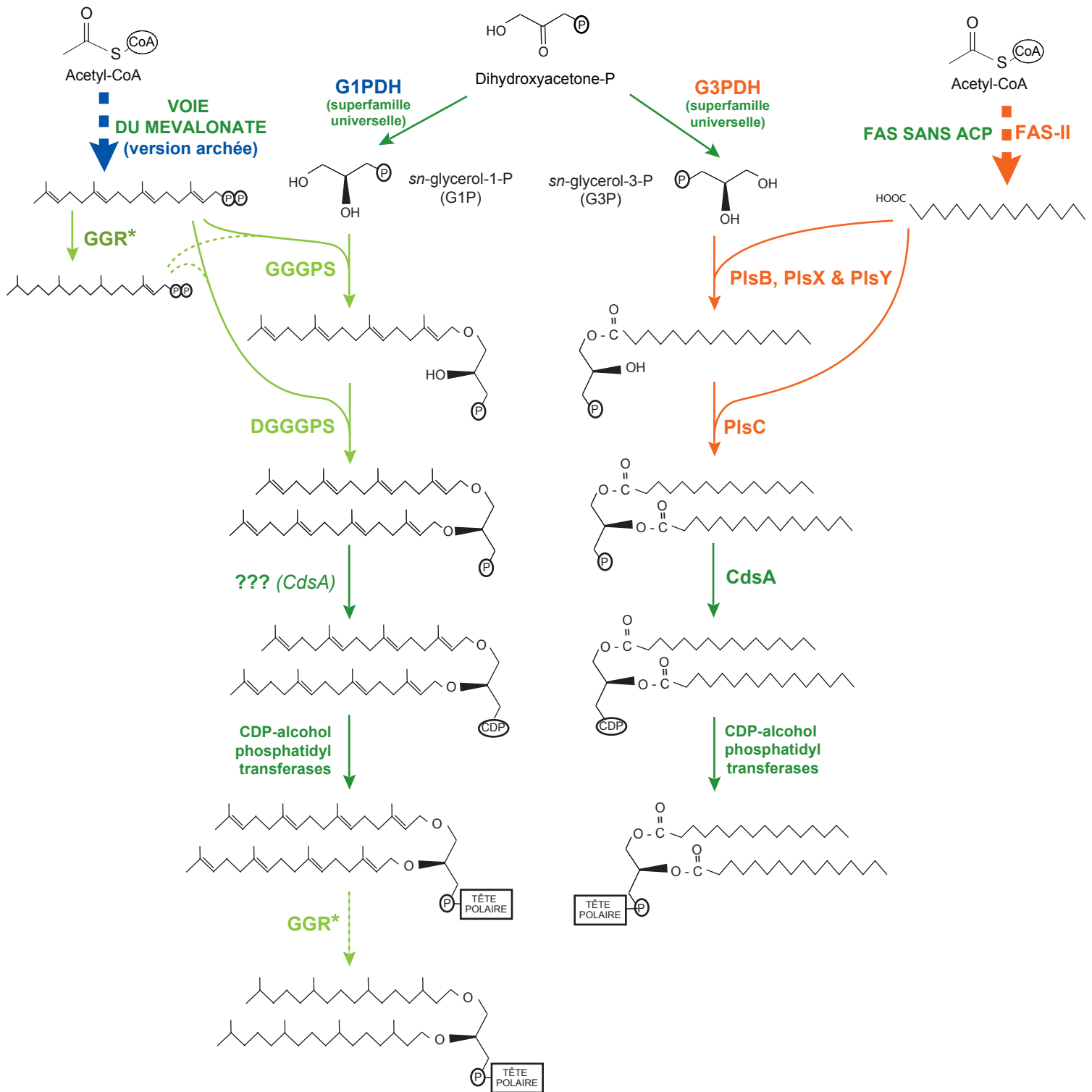


Figure 86. Voies de synthèse des phospholipides chez les archées et les bactéries.

La voie de synthèse archéenne est à gauche et les caractéristiques qui lui sont spécifiques sont marquées en bleu. La voie bactérienne est à droite et ses caractéristiques spécifiques sont en orange. Le vert foncé signale des enzymes ou des voies enzymatiques dont la présence peut être inférée chez le cenancêtre. Le vert clair pointe les cas où la présence de gènes chez le cenancêtre ne peut pas être exclue malgré le fait que les indices de la présence de l'activité enzymatique soient faibles. Les points d'interrogation signalent une réaction pour laquelle l'enzyme responsable n'a pas été décrite chez les archées mais dont il existe des homologues des protéines qui catalysent cette réaction chez les bactéries. L'étoile signale une activité qui peut avoir lieu à deux moments différents.

Abréviations. **Voie archéenne** : G1PDH : *sn*-glycérol-1-phosphate déshydrogénase, GGR : géranylgéranyl réductase ; GGGPS : géranylgéranylglycérol phosphate synthase ; DGGGPS : digéranylgéranylglycérol phosphate synthase ; CdsA : CDP-diacylglycérol synthase. **Voie bactérienne** : G3PDH : *sn*-glycérol-3-phosphate déshydrogénase ; FAS : synthèse d'acides gras ; PlsB, C, X et Y : acyltransférases.

Troisième partie : Synthèse des phospholipides dans les trois domaines du vivant

1. Contexte

Les résultats précédents nous ont permis de proposer que le cenancêtre pouvait synthétiser, au moins de façon non spécifique, les deux énantiomères de glycérol-phosphate [Peretó *et al.* 2004], les isoprénoïdes [Lombard et Moreira 2011b ; Lombard *et al.* 2012b] et les acides gras [Lombard *et al.* 2012a] par des mécanismes enzymatiques similaires à ceux existant actuellement. Dans ce tableau il manque les molécules qui agissent comme têtes polaires, qui sont très diverses mais parmi lesquelles on peut citer l'éthanolamine, le *myo*-inositol, le glycérol, la choline ou des acides aminés comme la sérine ou l'ornithine [Sohlenkamp *et al.* 2003 ; Koga et Morii 2007]. Ces molécules sont des métabolites divers qui jouent des rôles importants dans des très nombreuses fonctions cellulaires que l'on peut proposer chez le cenancêtre, comme c'est le cas des acides aminés dans la composition des protéines. Etant donné que des très nombreuses molécules polaires existaient certainement chez le cenancêtre mais qu'il est difficile de déterminer lesquelles d'entre elles rentrent dans la composition des phospholipides du cenancêtre en étudiant leur synthèse, il faut plutôt se concentrer sur leur incorporation dans les phospholipides.

Maintenant que la présence des mécanismes enzymatiques de synthèse des composants des phospholipides a été inférée chez le cenancêtre, il reste encore à déterminer si la machinerie d'assemblage des phospholipides existait dans cet organisme. Je vais donc traiter maintenant l'évolution des mécanismes de synthèse de phospholipides chez les procaryotes et les implications de ces résultats pour les membranes que l'on peut proposer pour le cenancêtre.

Liaisons ester, liaisons éther

Une fois que les acides gras ou les isoprénoïdes ont été synthétisés, ils doivent être attachés à l'énantiomère correspondant de glycérol phosphate (Figure 86). Traditionnellement, on considère que chez les bactéries et les eucaryotes les liaisons entre acides gras et G3P sont de type ester, alors que chez les archées les liaisons entre les isoprénoïdes et le G1P sont de type éther [De Rosa *et al.* 1986]. Cependant, il existe des nombreuses exceptions à cette règle car des liaisons éther ont été identifiées chez certains eucaryotes [Mangold et Paltauf 1983] et une large diversité de bactéries [Langworthy *et al.*

1983 ; Huber *et al.* 1992 ; Jung *et al.* 1994 ; Rutters *et al.* 2001 ; Sinninghe Damsté *et al.* 2002 ; Weijers *et al.* 2006 ; Sinninghe Damsté *et al.* 2007]. A ma connaissance les liaisons ester n'ont pas été identifiées à ce jour dans les phospholipides des archées.

Chez les bactéries, le premier mécanisme découvert capable de transférer une chaîne acylée vers la première position du G3P fut la G3P acyltransférase PlsB d'*E. coli* (Figure 86) [Lightner *et al.* 1980 ; Green *et al.* 1981]. Cependant, PlsB n'a été détecté que dans les génomes des protéobactéries, notamment les gamma-protéobactéries. Cette observation a conduit à la recherche d'un système d'acyltransférase alternatif chez les autres phyla bactériens et à l'identification de PlsX et PlsY [Lu *et al.* 2006]. Le système PlsX/PlsY est ubiquitaire chez les bactéries à l'exception de l'ordre Xanthomonadales de gamma-protéobactéries, qui utilisent PlsB. Alors que PlsB peut utiliser des thioesters CoA ou ACP pour son activité, le système PlsX/PlsY transforme d'abord l'acyl-ACP en acyl-phosphate, puis transfère celui-ci sur le G3P [Paoletti *et al.* 2007]. Quelque soit la première acyltransférase utilisée (PlsB ou PlsX/PlsY), la deuxième position du 1-acyl-G3P est acylée par une autre acyltransférase ubiquitaire chez les bactéries et dépendante de l'acyl-ACP, PlsC [Coleman 1990]. PlsB, PlsC, PlsX et PlsY sont toutes des enzymes membranaires.

Chez les archées, les enzymes qui transfèrent respectivement la première et la deuxième chaîne d'isoprénoïdes sur le G1P (Figure 86) sont la (*S*)-3-*O*-geranylgeranylglyceryl phosphate synthase (GGGPS) [Chen *et al.* 1993], cytosolique, et la (*S*)-2,3-Di-*O*-geranylgeranylglyceryl phosphate synthase (DGGGPS) [Hemmi *et al.* 2004], membranaire. On considère généralement que les deux enzymes sont caractéristiques des archées [Boucher *et al.* 2004 ; Hemmi *et al.* 2004], et on présume que les quelques bactéries qui possèdent des homologues [Doud *et al.* 2011 ; Guldán *et al.* 2011] les ont acquises par HGT [Boucher *et al.* 2004]. La corrélation de ces enzymes avec les génomes archéens et leur rôle primordial dans la fabrication de phospholipides ont conduit certains auteurs à proposer qu'elles auraient pu être déterminantes dans l'émergence des archées comme un domaine du vivant à part [Hemmi *et al.* 2004 ; Payandeh *et al.* 2006]. Cependant, selon la littérature, les deux enzymes appartiennent à des superfamilles de prenyltransférases présentes dans les trois domaines : GGGPS aurait un mécanisme réactionnel proche de celui d'autres enzymes qui transfèrent des chaînes d'isoprénoïdes sur une grande diversité de molécules [Soderberg *et al.* 2001] et DGGGPS appartiendrait à la superfamille UbiA de prenyltransférases qui synthétisent des nombreuses quinones, hèmes et pigments [Hemmi *et al.* 2004].

Têtes polaires

L'ajout de la tête polaire sur les acides archaetidique ou phosphatidique se fait au moins en deux étapes : la première est le remplacement du phosphate attaché au glycérol par un cytidyl diphosphate (CDP), donnant lieu à des molécules activées, et la deuxième est le remplacement du CDP par la tête polaire (Figure 86) [Sohlenkamp *et al.* 2003 ; Koga et Morii 2007 ; Zhang et Rock 2008]. La mise en place de certaines têtes polaires ou de glycolipides nécessite aussi parfois de modifications ultérieures de la tête polaire primaire, mais nous ne rentrerons pas dans ces modifications secondaires. Les enzymes responsables de ces fonctions sont localisées dans les membranes des cellules actuelles [Chang et Kennedy 1967 ; Sparrow et Raetz 1985 ; Morii *et al.* 2000 ; Daiyasu *et al.* 2005].

La première réaction est bien décrite chez les bactéries [Icho *et al.* 1985 ; Sparrow et Raetz 1985 ; Dowhan 1997] et a été biochimiquement caractérisée chez les archées [Morii *et al.* 2000], mais la protéine qui catalyse cette activité chez les archées reste inconnue, donc aucune comparaison n'a été possible jusqu'à présent. La protéine qui réalise cette fonction chez les bactéries est appelée CdsA pour CDP-diacylglycérol synthase. Il est intéressant de signaler que les protéines avec cette fonction sont capables d'utiliser tous les phospholipides, qu'ils comportent un énantiomère de glycérol phosphate ou l'autre et quelque soit leur chaîne latérale [Morii *et al.* 2000]. En ce qui concerne la deuxième réaction, de nombreux mécanismes sont connus permettant d'attacher la grande variété de têtes polaires des eucaryotes et des bactéries [Sohlenkamp *et al.* 2003 ; Zhang et Rock 2008]. Chez les archées, la diversité des têtes polaires est comparable [Koga et Morii 2005] mais jusqu'à présent seules des enzymes participant respectivement au transfert de sérine [Morii et Koga 2003] et de *myo*-inositol-phosphate [Morii *et al.* 2009] ont été décrites. Comme CdsA, ces protéines ne sont pas spécifiques ni de l'énantiomère de glycérol phosphate qui constitue l'ossature des phospholipides, ni des chaînes hydrophobes qui y sont attachées [Morii et Koga 2003]. Ces enzymes appartiennent à une famille de CDP-alcool phosphatidyltransférases très répandue chez les bactéries et dont les membres transfèrent chacun différentes molécules polaires sur le CDP-archaetol ou CDP-diacylglycérol [Daiyasu *et al.* 2005]. Ce dernier travail a apporté une phylogénie de cette superfamille, qui sera discutée par la suite en relation avec ma propre analyse phylogénétique.

Modifications des chaînes hydrophobes

Une fois que les chaînes hydrophobes sont liées aux phospholipides, elles peuvent subir des modifications importantes de leur degré d'insaturation, ce qui change de façon très importante la fluidité et la stabilité des membranes [Dannenmuller *et al.* 2000 ; Zhang et Rock 2008]. Comme vu précédemment, certaines bactéries sont capables de synthétiser *de novo* des acides gras insaturés grâce aux déshydratases de la FAS-II, alors que d'autres intègrent les insaturations dans les chaînes d'acyl-ACP après leur synthèse [Altabe *et al.* 2003]. Les enzymes responsables de cette désaturation tardive chez les eucaryotes et les bactéries sont appelées acyl-ACP désaturases et elles sont membranaires. Il existe plusieurs désaturases qui se ressemblent peu entre elles, mais qui partagent le même domaine protéique caractéristique de leur fonction. Certaines désaturases (DesA) sont capables d'agir sur les acides gras attachés aux phospholipides de la membrane, alors que d'autres (DesB) modifient les acyl-CoA d'origine exogène avant que ceux-ci soient intégrés dans les phospholipides (par exemple chez les bactéries capables d'incorporer des acides gras extérieurs, comme certaines gamma-protéobactéries) [Zhu *et al.* 2006].

De leur côté, les chaînes d'isoprénoïdes associées aux phospholipides archéens ne conservent généralement pas les insaturations présentes lors de leur synthèse, et les doubles liaisons doivent par conséquent être éliminées. L'enzyme qui se charge de cette réaction chez les archées, dite géranylgeranyl reductase (GGR), est membranaire. Elle a été décrite chez les euryarchées *Thermoplasma acidophilum*, *Archaeoglobus fulgidus* et la crenarchée *Sulfolobus acidocaldarius* [Nishimura et Eguchi 2006 ; Murakami *et al.* 2007 ; Sato *et al.* 2008] et son activité emploie majoritairement des isoprénoïdes déjà attachés à des phospholipides, bien qu'elle puisse également utiliser des chaînes isolées comme substrat [Murakami *et al.* 2007]. Les GGR archéennes ont la particularité de pouvoir éliminer toutes les doubles liaisons des isoprénoïdes lorsqu'ils sont déjà attachés aux phospholipides, mais de laisser intacte la double liaison du carbone 2 dans les isoprénoïdes libres, car cette dernière est nécessaire à l'intégration des chaînes isoprénoïdes dans les phospholipides [Sato *et al.* 2008]. De ce fait, la réduction des isoprénoïdes peut intervenir à différents moments de la synthèse des phospholipides des archées (Figure 86). Des nombreux paralogues des GGR ont été détectés chez les archées, mais jusqu'à présent leurs fonctions restent inconnues [Matsumi *et al.* 2011]. Des homologues des GGR ont été décrits aussi chez les bactéries et les eucaryotes photosynthétiques, et participent dans la synthèse de chlorophylle et de pigments photorésistants [Addlesee *et al.* 1996 ; Keller *et al.* 1998 ; Tanaka *et al.* 1999].

2. Evolution des enzymes de synthèse des phospholipides

Les acyltransférases

La recherche d'homologues des glycérol-phosphate acyltransférases bactériennes PlsX, PlsY, PlsB et PlsC dans les génomes d'archées n'a révélé aucun homologue dans ce domaine du vivant à l'exception d'une séquence de *Methanobacterium* sp. qui pourrait être homologue des PlsX bactériens. Cependant, même si l'homologie de cette séquence était confirmée biochimiquement, le scénario le plus parcimonieux impliquerait un transfert horizontal récent pour expliquer cette observation. Ce résultat s'accorde d'ailleurs avec l'absence apparente de liaisons esters chez les archées. Des recherches avec PSI-BLAST [Altschul *et al.* 1997] ont été entreprises pour essayer de détecter des homologues plus distants de ces acyltransférases et voir si ces homologues pouvaient nous renseigner sur une éventuelle relation évolutive avec les archées.

Dans le cas de PlsX les séquences détectées sont annotées et comportent des domaines protéiques très divers (dont la vérification a été faite avec la base de données de PFAM) très divers, au point que le caractère homologue d'une bonne partie de ces séquences peut être mis en cause. Il est cependant intéressant de signaler la détection parmi ces séquences de quelques phosphate acyltransférases qui participent au transfert d'acétyl ou de butyryl sur des groupements phosphate, une réaction qui est très proche de celle proposée pour PlsX. Quoiqu'il en soit, et malgré la détection de très nombreuses fonctions dans ces analyses, très peu de séquences archéennes sont détectées au bout de quatre itérations. La recherche d'homologues lointains de PlsY détecte principalement des acyltransférases dans les premières itérations, puis des facteurs d'élongation de la transcription. Dans ce cas, des séquences archéennes sont détectées mais elles semblent réaliser des fonctions de facteur d'élongation, éliminant tout intérêt pour la détection d'acyltransférases. La recherche d'homologues de PlsC donne des protéines qui transfèrent des groupements acylés vers des molécules très diverses dans les premières itérations de PSI-BLAST, puis donnant des canaux protéiques et des protéines membranaires à partir de la quatrième itération, sans qu'il y ait pour autant une détection de séquences archéennes. Bien que normalement présente uniquement chez les protéobactéries, j'ai également fait des PSI-BLAST à partir des PlsB, ce qui permet de détecter d'autres acyltransférases bactériennes mais pas de séquences archéennes. Tous ces résultats soutiennent que les différentes acyltransférases ont évolué dans le domaine bactérien à partir d'ancêtres qui étaient déjà restreints aux bactéries et pour

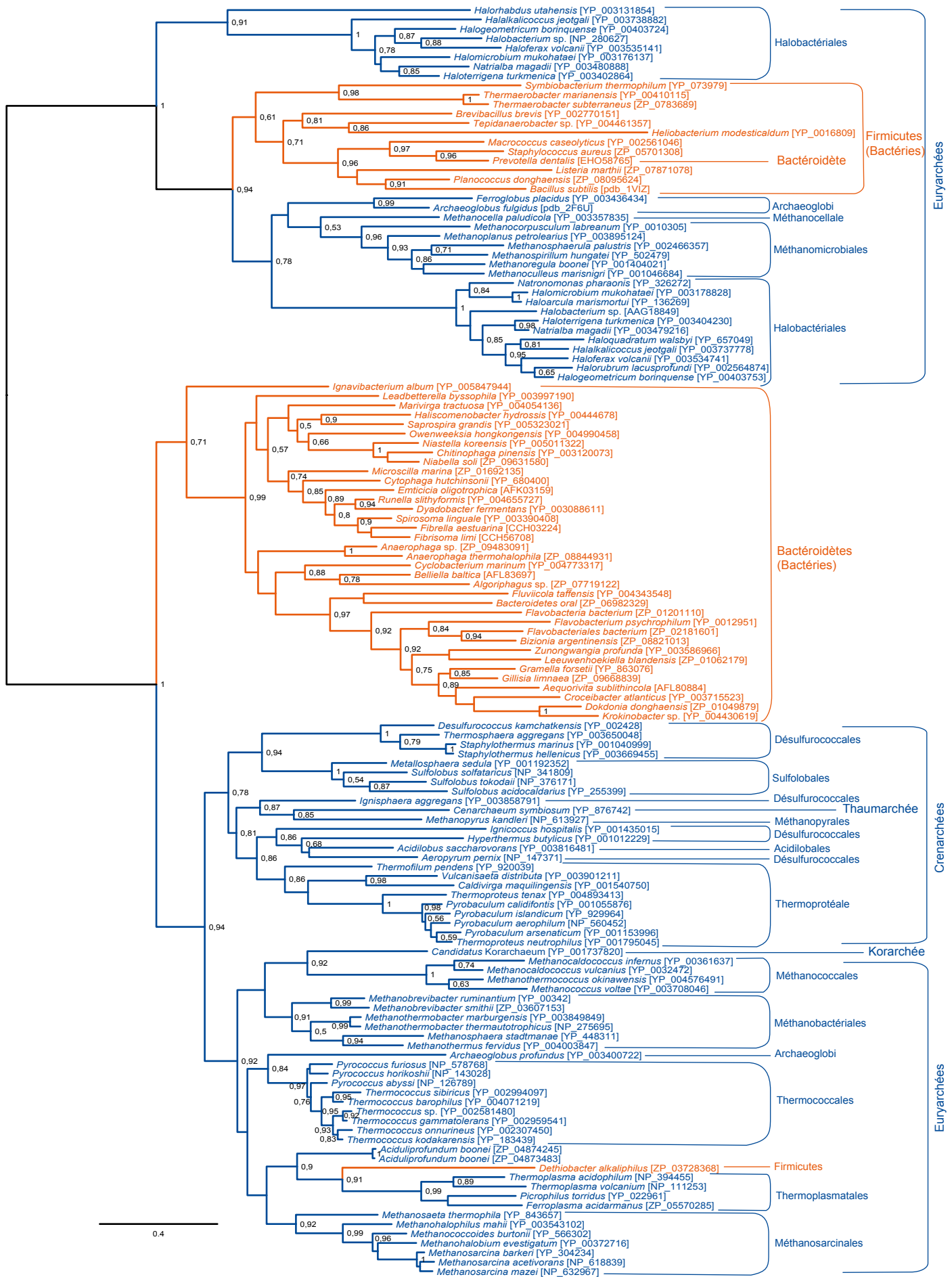


Figure 87. Phylogénie des GGGPS.

Reconstruction avec FastTree utilisant 135 séquences représentatives et 176 positions conservées. Les nombres à proximité des noeuds indiquent les valeurs de robustesse calculées par FastTree.

lesquels aucun ancêtre archéen de fonction comparable n'est connu. Etant donné que l'étude de ces enzymes n'aurait pas pu nous renseigner sur l'existence de liaisons ester dans le cenancêtre, je n'ai pas étudié plus en détail l'évolution des acyltransférases.

Les prenyltransférases

Deux types de GGGPS, homologues mais assez divergentes entre elles, ont été décrites chez les archées [Chen *et al.* 1993 ; Nemoto *et al.* 2003]. Boucher *et al.* (2004) ont remarqué que le type de GGGPS décrit dans Nemoto *et al.* (2003) était le plus courant chez les archées, alors que la copie décrite dans Chen *et al.* (1993), que l'on appellera ici copie divergente de GGGPS, était caractéristique des génomes d'haloarchées et d'*Archaeoglobus fulgidus*. Malgré les différences qui existent entre les deux types d'enzymes, les séquences divergentes semblent avoir conservé leur fonction GGGPS [Payandeh *et al.* 2006]. Des homologues de GGGPS ont également été détectés chez les Bacillales (ordre des Firmicutes) et le bacteroidetes *Cytophaga hutchinsonii*, et appelés PcrB. La capacité des gènes bactériens PcrB à synthétiser des lipides de type archéen a été récemment démontrée, bien que la fonction cellulaire de ces enzymes reste énigmatique [Guldan *et al.* 2011].

L'évolution des GGGPS a été étudiée en utilisant une version actualisée (août 2012) des bases de données génomiques. Ceci a permis d'observer que la copie très divergente de GGGPS est répandue non seulement parmi les archées halophiles mais aussi dans les classes des Archaeoglobi et des Méthanomicrobia (Figure 87). Ces trois classes forment un groupe monophylétique avec les Methanosarcinales [Brochier-Armanet *et al.* 2011b], qui n'ont pas cette version très divergente. Ce résultat pourrait-être interprété comme le produit d'une duplication et grande divergence dans un ancêtre de ce groupe d'archées, suivie de sa perte dans les Methanosarcinales. En dehors de l'existence de cette version très divergente, qui forme un groupe extérieur au reste des archées, la phylogénie des GGGPS archéennes est globalement congruente avec celle des organismes, soutenant la présence de ce gène chez l'ancêtre des archées. Nous observons aussi que les GGGPS sont répandues parmi les Bacillales mais aussi chez grand nombre de Bacteroidetes. Sauf rares exceptions, les séquences de Bacillales et de Bacteroidetes forment respectivement deux groupes dont le premier branche parmi les GGGPS divergentes archéennes et le deuxième à la base du groupe contenant toutes les autres archées. Ceci indique que les séquences bactériennes de chaque groupe ont été reçues par au moins deux événements indépendants de HGT dans les lignées

spécifiques de chacun de ces groupes bactériens. Les rares séquences bactériennes qui ne suivent pas cette distribution sont certainement le résultat d'autres HGTs plus récents.

Malgré la limitation des GGGPS aux archées et à des bactéries qui les ont acquis par HGT, il a été proposé que cette enzyme appartienne à une famille universelle de prenyltransférases [Soderberg *et al.* 2001] incluant des enzymes capables de transférer des groupements isoprénoïdes sur des molécules très diverses comme des protéines [Harris et Poulter 2000], l'ARNt [Moore et Poulter 1997], ou une grande diversité de molécules aromatiques [Kuzuyama *et al.* 2005]. Compte tenu de l'apparente restriction des GGGPS aux archées, je me suis intéressé aux homologues les plus proches de ces séquences afin de mieux comprendre leur origine. Etonnamment, l'hypothèse selon laquelle les GGGPS appartiennent à une famille très large de prenyltransférases se base principalement sur une comparaison des mécanismes réactionnels et des modalités d'interaction avec les substrats respectifs des prenyltransférases [Soderberg *et al.* 2001]. Ces points ne permettent pas de déterminer univoquement la relation évolutive entre ces enzymes, telle leur éventuelle descendance depuis séquence ancestrale unique. Par ailleurs, les GGGPS et les autres prenyltransférases citées dans le travail de Soderberg *et al.* ne contiennent pas les mêmes domaines protéiques. D'une part, les GGGPS contiennent un motif de type *TIM barrel* [Payandeh *et al.* 2006 ; Doud *et al.* 2011], qui reçoit son nom de la triosephosphate isomérase (première enzyme dans laquelle ce motif a été décrit). Depuis, ce motif a été identifié comme l'un des plus répandus dans les protéines des trois domaines du vivant. Il est parfois appelé par le nom plus générique de motif $(\beta/\alpha)_8$ car il est composé de 8 feuillets beta liés entre eux par un réseau de liaisons hydrogène, ce qui forme une structure en forme de tonneau. Chaque feuillet est généralement au contact d'une seule hélice alpha qui se situe à l'extérieur de la structure [Banner *et al.* 1975], mais des nombreuses variantes existent et se recombinent à d'autres domaines, ce qui explique la grande diversité fonctionnelle de ces protéines. D'autre part, les autres prenyltransférases comportent des motifs différents, comme par exemple celui des terpénoïde synthases, qui est composé d'une dizaine d'hélices alpha [Liang *et al.* 2002].

La recherche d'homologues par BLAST et PSI-BLAST sur un échantillon réduit de génomes des trois domaines du vivant permet de détecter, en dehors des gènes GGGPS et PcrB, des nombreuses protéines ou domaines TIM barrel : HisF (qui intervient dans la synthèse de l'histidine), la sous-unité alpha de la tryptophane synthase (synthèse du tryptophane), des domaines IMPDH/GMPR (métabolisme de la purine), l'orotidine 5-phosphate decarboxylase et la dihydroorotate dehydrogenase (biosynthèse de la pyrimidine),

la deoxyribose-phosphate aldolase (métabolisme des nucléotides), la dihydrouridine synthase (modification posttranscriptionnelle des ARNt) ou la TIM. Au vue de la faible similarité des séquences primaires de protéines TIM barrel, une recherche topologique utilisant le serveur DALI a également été menée. Ce serveur permet de détecter les protéines dont la structure tridimensionnelle tel que décrite dans la Protein Data Bank ressemble à une structure cristallisée d'intérêt [Holm et Sander 1996 ; Holm et Sander 1998]. En utilisant l'information de la cristallisation de GGGPS d'*A. fulgidus* [Payandeh *et al.* 2006] il est ainsi possible de détecter les protéines dont la structure tridimensionnelle ressemble le plus aux GGGPS. Les protéines les plus proches des GGGPS sont MoeO5 (synthèse de l'antibiotique moenomycine), une thiamine phosphate phosphorylase (synthèse de la thiamine), la protéine régulatrice TenI (régulatrice les protéases, sans qu'aucune activité enzymatique lui soit connue), la chaîne alpha de la tryptophane synthase, l'orotidine 5'-phosphate decarboxylase, HisF, TIM et la ribulose-phosphate 3-épimérase (cycle de Calvin). Ces résultats montrent que la recherche à partir de GGGPS ne permet de détecter aucune prenyltransférase et, qu'au contraire, toutes les séquences détectées sont des protéines TIM barrel. La recherche avec DALI de structures homologues des prenyltransférases proposées dans le travail de Soderberg *et al.* montre que seules les recherches utilisant les DMAPP-tRNA transférases et les DMAPP-tryptophane synthases se détectent réciproquement, alors que les autres prenyltransférases en sont incapables. Ce résultat contredit l'hypothèse de Soderberg *et al.* (2001) selon laquelle les GGGPS font partie d'une famille très large de prenyltransférases et place plutôt les GGGPS dans la très grande famille des protéines TIM barrel, parmi lesquelles je vais essayer de replacer les GGGPS.

Les protéines TIM barrel ont des séquences primaires très divergentes qui ne permettent pas la construction de phylogénies moléculaires fiables [Reardon et Farber 1995 ; Copley et Bork 2000]. Initialement, la faible similarité de séquence a conduit les premiers travaux sur l'évolution de ces enzymes à se demander si le motif $(\beta/\alpha)_8$ était issu de la divergence à partir d'un seul domaine très ancien ou si sa présence dans des très nombreuses protéines était le résultat de très multiples convergences dont le motif résultant serait toujours le même parce qu'il supposerait un minimum énergétique très stable [Reardon et Farber 1995]. Plusieurs travaux ont essayé de répondre à cette question en comparant un nombre grandissant de substrats, sites actifs, sites de fixation de phosphate ou de co-facteurs, structures tridimensionnelles ou en utilisant des PSI-BLAST et IMPALA pour établir un réseau des protéines TIM barrel se rapprochant le plus les unes des autres [Copley et Bork 2000 ;

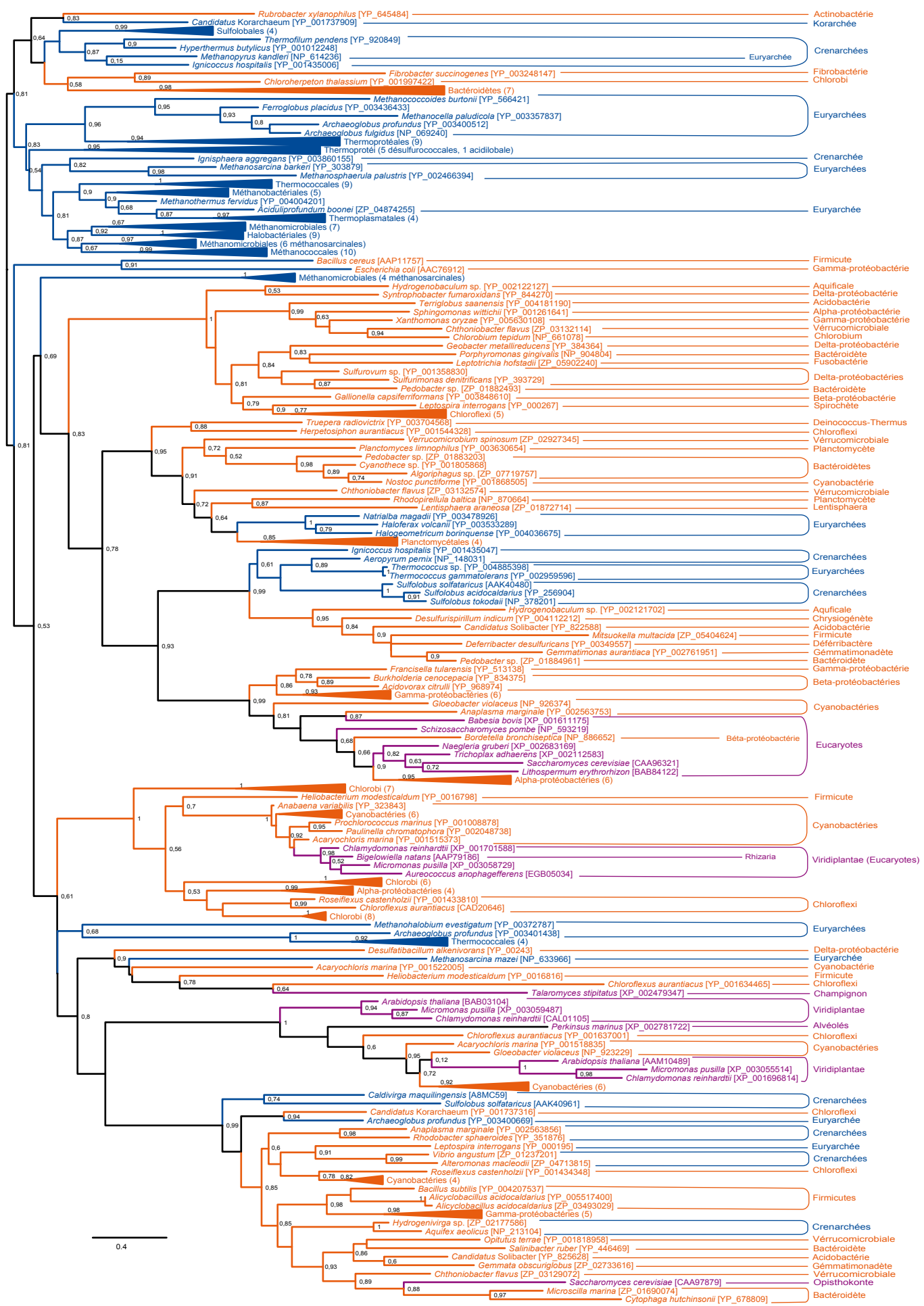


Figure 88. Phylogénie des DGGGS.
 Reconstruction avec FastTree utilisant 278 séquences représentatives et 114 positions conservées. Les nombres à proximité des noeuds indiquent les valeurs de robustesse calculées par FastTree.

Nagano *et al.* 2002]. Ces travaux ont permis d'établir des familles d'homologues fiables, superfamilles d'homologues très probable et des relations possibles entre ces superfamilles, ce qui semble privilégier une divergence à partir de séquences anciennes comme origine de la plupart de ces protéines. Les protéines TIM barrel seraient ainsi l'une des plus anciennes familles de protéines et auraient largement précédé le cenancêtre. Ces protéines étant particulièrement liées aux activités du métabolisme central, il a été proposé que, dans leurs versions les plus anciennes, la forme en tonneau ait permis d'accélérer les réactions de façon très peu spécifique. La spécificité serait devenue plus importante lors de la duplication et de la sousfonctionnalisation de ces enzymes au cours du temps [Anantharaman *et al.* 2003]. GGGPS est notamment proche de TIM, de la ribulose-phosphate 3-épimérase, de la chaîne alpha des tryptophane synthases, des thiamine phosphate synthases, et de l'orotidine 5'-phosphate decarboxylase. Or, toutes ces enzymes appartiennent à la superfamille de protéines TIM barrel appelée FMOP pour *FMN-dependent oxidoreductase and phosphatase binding enzymes* [Reardon et Farber 1995 ; Copley et Bork 2000 ; Nagano *et al.* 2002]. Les protéines FMOP ont la particularité de partager un site de fixation au phosphate qui leur est caractéristique [Bork *et al.* 1995 ; Nagano *et al.* 2002] et il a été proposé que le dernier ancêtre commun de toutes ces protéines ait utilisé le glycéraldéhyde 3-phosphate comme substrat [Copley et Bork 2000]. Ceci est stimulant dans notre cas tant la ressemblance est forte entre cette molécule et le glycérol phosphate dont est dépendant l'activité de la GGGPS. Un argument favorable supplémentaire réside dans le fait que la structure tridimensionnelle la plus ressemblante aux GGGPS est celle de MoeO5, dont la relation aux autres protéines TIM barrel est inconnue mais dont la fonction est le transfert d'une chaîne de FPP sur un acide 3-phosphoglycérique [Doud *et al.* 2011], une autre molécule très proche du glycérol phosphate. En résumé, l'ensemble de ces résultats suggère que les GGGPS sont apparues dans la lignée archéenne à partir d'un ancêtre appartenant à la superfamille FMOP des TIM barrel. L'activité de la protéine ancestrale était probablement peu spécifique et liée à l'utilisation de trioses phosphate dans des nombreuses fonctions métaboliques.

En ce qui concerne la DGGGPS, Hemmi *et al.* avaient réalisé en 2004 une analyse phylogénétique de la superfamille des prenyltransférases UbiA qui contient les DGGGPS. Leur analyse, qui concernait uniquement 17 séquences, les divisait en 6 groupes. Nous avons actualisé ces résultats avec une base de données mise à jour (août 2012), et nous avons obtenu des résultats similaires bien que plus divers (Figure 88). Les séquences archéennes groupent ensemble dans cet arbre, et leur phylogénie est congruente avec les grands groupes d'archées

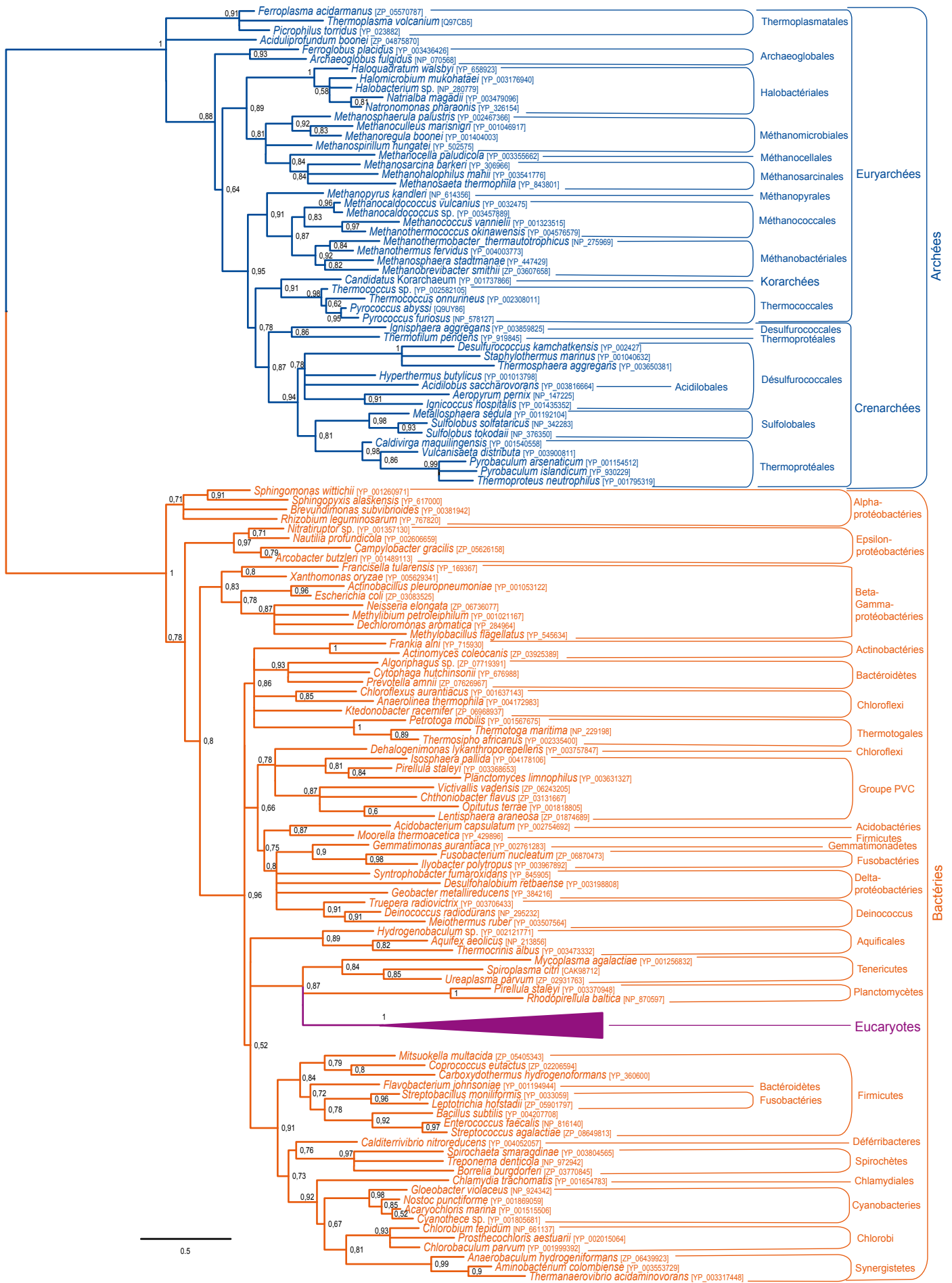


Figure 89. Phylogénie de CdsA

Reconstruction avec FastTree utilisant 133 séquences représentatives et 87 positions conservées. Les nombres à proximité des noeuds indiquent les valeurs de robustesse calculées par FastTree.

bien que les crenarchées apparaissent comme un groupe paraphylétique probablement à cause d'artéfacts de reconstruction. De ce fait, il est très probable que le dernier ancêtre des archées ait déjà possédé un de ces gènes. Des bactéroïdètes branchent parmi les crenarchées, ce qui rappelle la distribution taxonomique de GGGPS parmi les bactéries et met en évidence un HGT vers un ancêtre des bactéroïdètes. Le reste de l'arbre est principalement composé par des chlorobi, cyanobactéries, chloroflexi, planctomycétales et quelques protéobactéries qui comportent des enzymes de la superfamille UbiA de prenyltransférases. On observe ainsi plusieurs paralogues dont les séquences regroupent en grande partie conformément aux activités supposées dans Hemmi *et al.* (2004), i. e. la synthèse de quinones des chaînes respiratoires, d'hèmes, de chlorophylle ou de vitamine E. Du fait de l'empreinte fonctionnelle très marquée dans cet arbre et de la distribution partielle des séquences DGGGPS parmi les bactéries, il semble probable que de nombreuses duplications, neofonctionnalisations et HGTs aient eu lieu dans l'histoire de cette famille, rendant difficile la détermination de la présence chez l'ancêtre des bactéries de la séquence à l'origine de cette prodigalité, ou de l'apparition de ces séquences dans le domaine bactérien suite à des HGT depuis les archées vers les ancêtres de certains phyla bactériens.

La CDP-diacylglycérol synthase

L'enzyme qui utilise un CTP pour remplacer le phosphate de l'acide archaetidique par un CDP chez les archées a été caractérisée biochimiquement mais sa séquence n'a pas été établie [Morii *et al.* 2000]. On se concentrera donc sur la CDP-diacylglycérol synthase (CdsA) qui catalyse la réaction équivalente chez les bactéries [Icho *et al.* 1985 ; Sparrow et Raetz 1985]. En partant de la protéine bactérienne, nous avons pris l'homologue le plus proche parmi les archées, que nous avons utilisé comme graine pour une recherche détaillée dans les génomes archéens. Le résultat de cette recherche nous a permis d'observer que les homologues de cette enzyme sont très répandus à la fois chez les bactéries et chez les archées. Plus remarquable, la phylogénie de ces séquences montre deux clades bien différenciés qui correspondent aux archées et aux bactéries, et les groupes à l'intérieur de chaque domaine sont congruents avec la taxonomie de ces organismes (Figure 89). Ce résultat soutient l'héritage vertical de ces séquences dans les deux domaines procaryotes du vivant et, par conséquent, sa présence chez le cenancêtre. Cependant, à ma connaissance aucune de ces séquences archéennes n'a encore été caractérisée, donc ce résultat est subordonné à la vérification que ces séquences sont bien celles responsables du transfert de CDP sur l'acide archaetidique décrit par Morii *et al.* (2000).

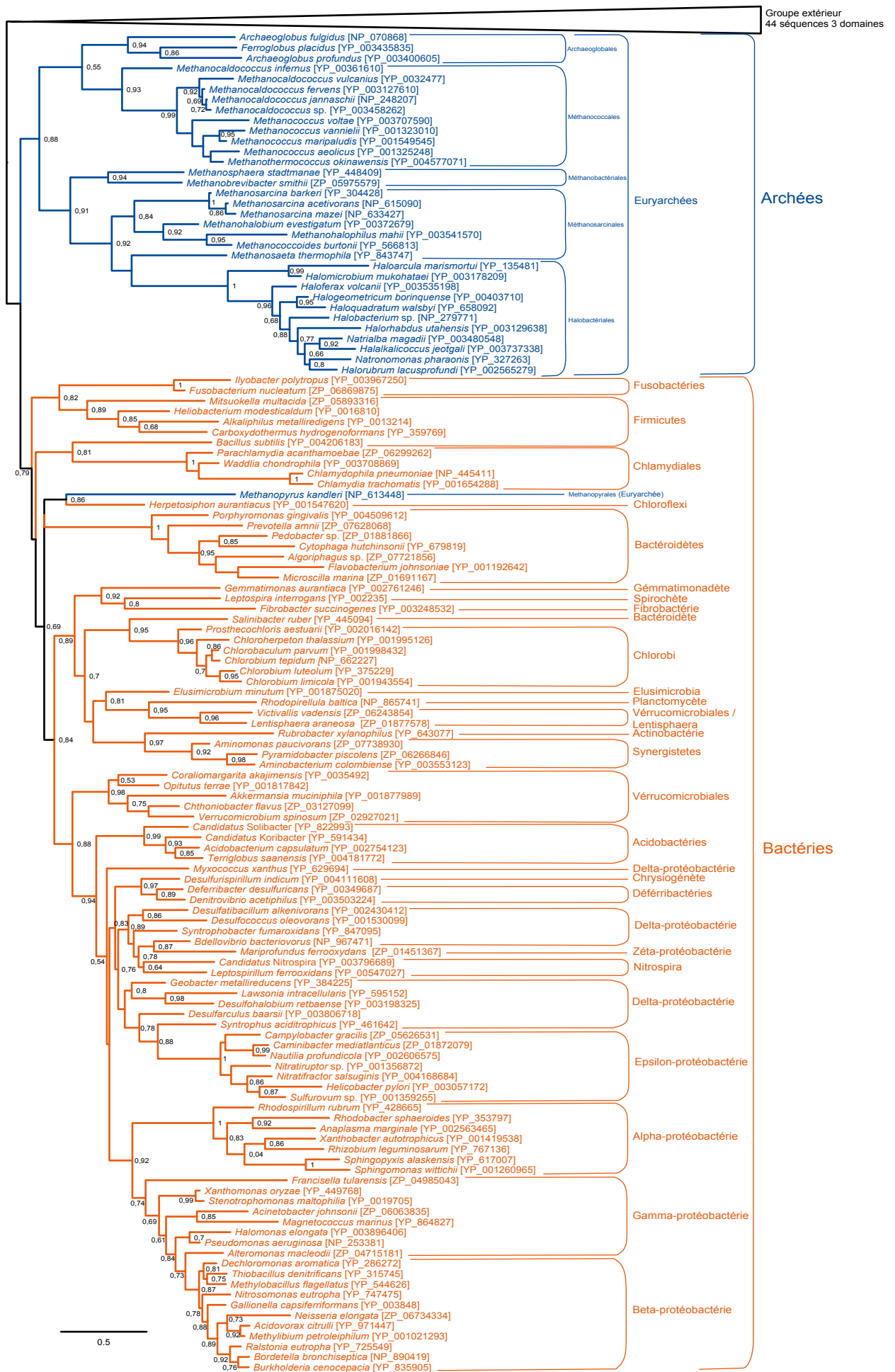


Figure 90. Phylogénie des CDP alcool phosphatidyl transférases probablement liées au transfert de sérine.

Reconstruction avec FastTree utilisant 173 séquences représentatives et 132 positions conservées. Les nombres à proximité des noeuds indiquent les valeurs de robustesse calculées par FastTree.

Les CDP-alcool phosphatidyltransférases

Les premiers travaux phylogénétiques sur la principale superfamille d'enzymes responsables du transfert de têtes polaires sur le phospholipide en cours de synthèse montraient que les séquences groupaient ensemble en accord avec leurs substrats prédits [Katayama *et al.* 2004 ; Daiyasu *et al.* 2005]. Le travail de Daiyasu *et al.* (2005) signalait que la phylogénie obtenue pour les séquences bactériennes respectait les principaux groupes taxonomiques bactériens et que des gènes homologues pouvaient être détectés dans les génomes des archées, ce qui a conduit à proposer que ces gènes aient pu être présents chez le cenancêtre [Koga et Morii 2007]. Notre analyse avec un échantillon taxonomique plus large confirme que les gènes de cette superfamille sont répandus dans les archées. Notre première phylogénie montre que les séquences se divisent en trois catégories (résultats non montrés), l'une associée à l'utilisation de sérine comme tête polaire [Morii et Koga 2003], une autre liée au transfert de *myo*-inositol-phosphate et peut-être aussi de glycérol [Daiyasu *et al.* 2005 ; Morii *et al.* 2009] et la dernière utilisant le glycérol-phosphate. Les séquences bactériennes sont présentes dans les trois catégories, mais les archées n'apparaissent que dans les deux premières. Afin de réduire les artefacts de reconstruction dus à la divergence des séquences, une nouvelle phylogénie a été construite pour chaque catégorie séparément.

Dans l'arbre des séquences associées au transfert de la sérine [Morii et Koga 2003], les séquences d'archées se limitent aux euryarchées (Figure 90). Dans cette phylogénie il existe un groupe de séquences divergentes très mélangées, contenant à la fois des bactéries, des eucaryotes un groupe d'haloarchées et *Methanococoides burtonii* qui ont probablement échangé leurs séquences par des HGT multiples. En dehors de ce groupe très divergent, on observe que les autres séquences (ce qui inclut des bactéries et des euryarchées) forment des groupes congruents avec la phylogénie attendue des organismes à quelques exceptions près. Ces résultats soutiennent qu'au moins le dernier ancêtre des euryarchées, et probablement celui des bactéries possédaient un de ces gènes.

Dans les séquences de la deuxième catégorie, on trouve une plus grande diversité d'archées (Figure 91), que Daiyasu *et al.* (2005) ont prédit comme responsables du transfert de glycérol et dont la capacité à utiliser du *myo*-inositol-phosphate a été caractérisée récemment [Morii *et al.* 2009]. Cette phylogénie soutient la monophylie des crenarchées et une distribution inégale de plusieurs paralogues parmi les euryarchées. Ceci rend difficile la détermination de l'évolution de ce gène dans ce phylum sans bien connaître les fonctions

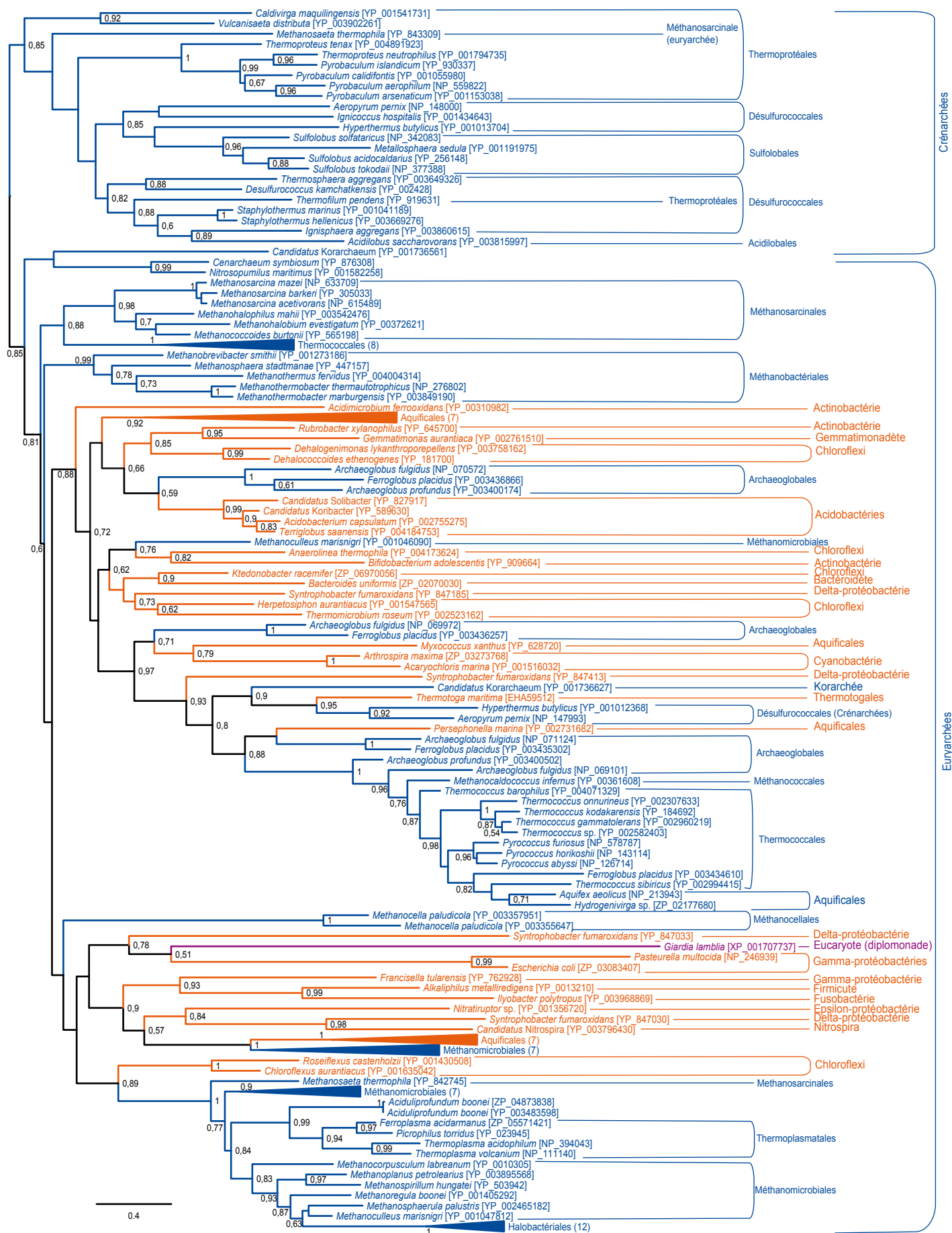


Figure 91. Phylogénie des CDP alcool phosphatidyl transférases peut être liées au transfert de *myo*-inositol.

Reconstruction avec FastTree utilisant 160 séquences représentatives et 159 positions conservées. Les nombres à proximité des noeuds indiquent les valeurs de robustesse calculées par FastTree.

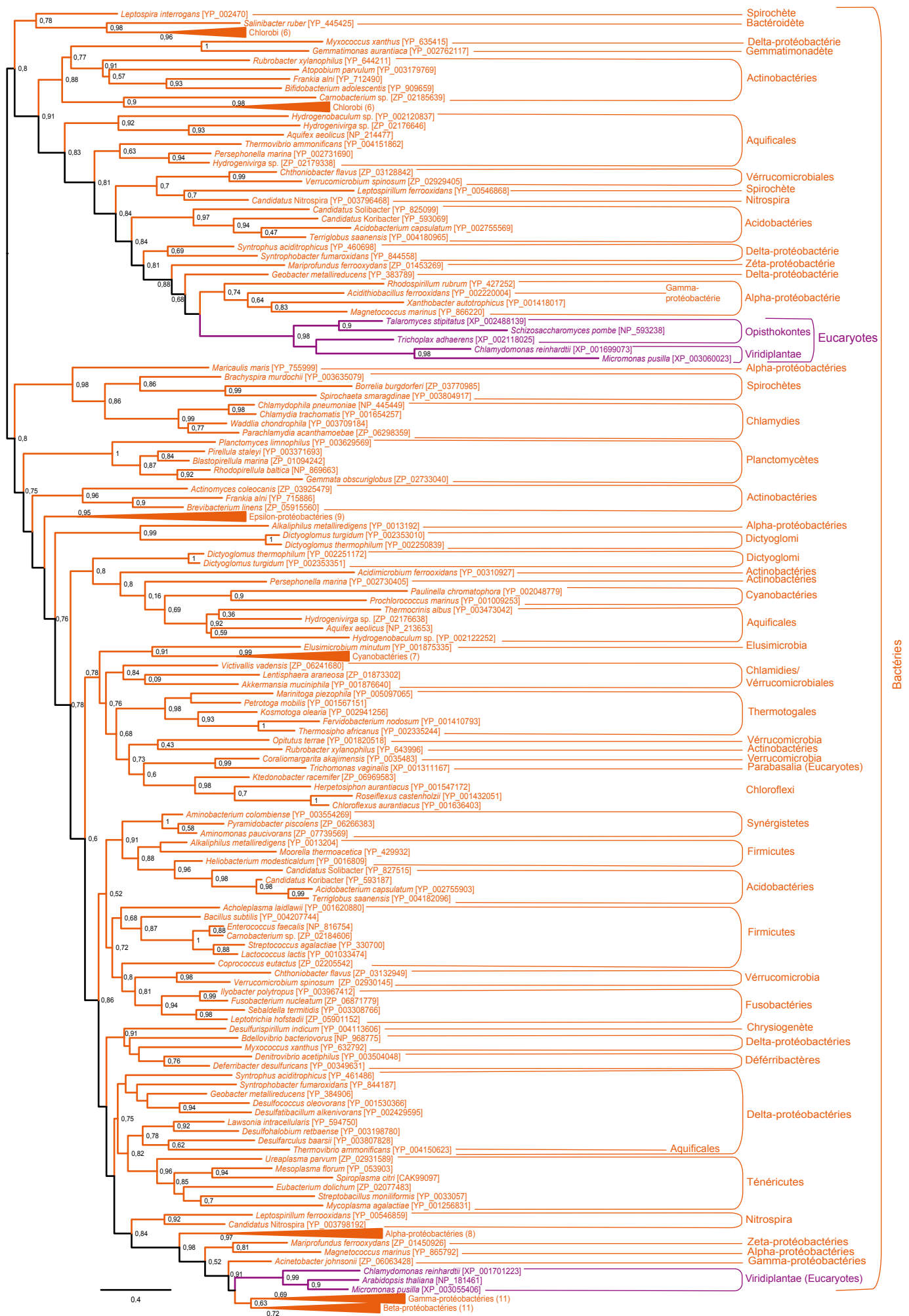


Figure 92. Phylogénie des CDP alcool phosphatidyl transférases du groupe bactérien.
 Reconstruction avec FastTree utilisant 191 séquences représentatives et 131 positions conservées.
 Les nombres à proximité des noeuds indiquent les valeurs de robustesse calculées par FastTree.

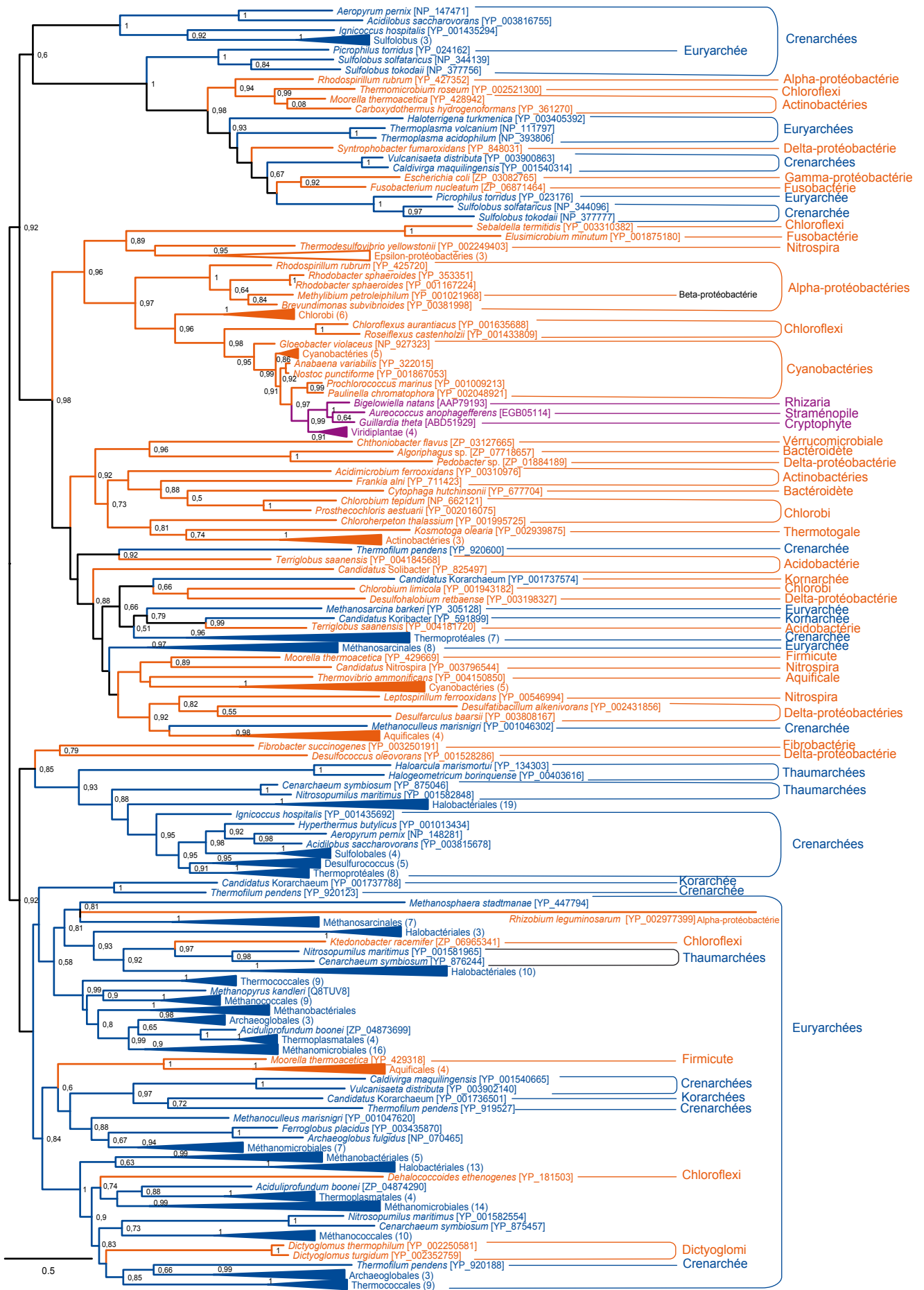


Figure 93. Phylogénie des GGR.
 Reconstruction avec FastTree utilisant 317 séquences représentatives et 266 positions conservées.
 Les nombres à proximité des noeuds indiquent les valeurs de robustesse calculées par FastTree.

biochimiques de ces séquences. Malgré cette difficulté, la monophylie des crenarchées et la grande diversité de ces séquences chez les euryarchées suggèrent qu'au moins l'une de ces séquences devait être présente chez le dernier ancêtre des archées. Ce gène aurait été conservé chez les crenarchées mais aurait subi des nombreuses duplications et neofonctionnalisations chez les euryarchées. Peu de bactéries ont été détectées dans cette catégorie ; leur distribution très irrégulière parmi les séquences d'archées amène à penser qu'elles pourraient résulter de nombreux HGT depuis les archées.

La phylogénie du troisième groupe, probablement liée au transfert de glycérol-phosphate [Chang et Kennedy 1967 ; Dowhan 1992] et ne contenant que des bactéries et quelques eucaryotes, montre que ce gène est répandu dans une très large diversité de bactéries. Les séquences forment dans l'ensemble des groupes conformes aux principaux taxons bactériens, bien que certains écarts par rapport à cette classification suggèrent aussi des HGT plus récents (Figure 92). Ces résultats s'accordent avec la présence d'un de ces gènes chez le dernier ancêtre commun des bactéries. Les séquences eucaryotes sont peu nombreuses et éparées dans l'arbre, suggérant des origines par HGT depuis les bactéries.

Modifications des chaînes hydrophobes.

Une première recherche avec BLAST, dans les génomes d'archées, de séquences homologues des desaturases responsables de l'introduction de doubles liaisons dans les acides gras bactériens et eucaryotes n'a révélé aucune séquence homologue. Les recherches avec PSI-BLAST d'homologues plus lointains de DesA et DesB révèlent quelques facteurs de transcription dans les premières itérations et quelques protéines membranaires et transporteurs à partir de la troisième itération. Toutefois, la multiplicité de fonctions de ces séquences met en question le caractère homologue de ces séquences. De plus, quasiment aucune séquence d'archée n'est détectable, ce qui indique que ces protéines sont principalement bactériennes dans le cas de DesA et eucaryotes dans le cas de DesB. Ainsi l'analyse phylogénétique n'aurait pas apporté d'information sur le cenancêtre et cette piste n'a pas été poursuivie.

Par ailleurs, la phylogénie des géranylgeranyl réductases (GGR) chez les archées confirme que ces gènes sont très largement distribués parmi les crenarchées et qu'il existe au moins deux paralogues très répandus dans la majorité des euryarchées (Figure 93). De nombreux groupes d'archées possèdent d'autres copies additionnelles de ces gènes. Ce résultat s'accorde à l'éventualité qu'au moins un de ces gènes était présent chez l'ancêtre des archées, qu'il ait subi une probable duplication dans la lignée des euryarchées ainsi que de

nombreuses autres duplications et HGT plus récents. Les séquences bactériennes de GGR ne branchent pas parmi les archées, mais les fréquents HGT et le rôle de ces protéines chez les bactéries (très liées aux fonctions de photosynthèse) rendent difficile la détermination de la présence de ce gène chez l'ancêtre des bactéries. Parmi les eucaryotes, il est intéressant de signaler que des homologues de ces séquences se trouvent uniquement parmi des organismes comportant des plastes et que ces séquences branchent parmi les cyanobactéries, soutenant fortement l'origine plastidiale de ces séquences chez les eucaryotes.

3. Synthèse enzymatique de phospholipides chez le cenancêtre

Enzymes de la synthèse des phospholipides présentes chez le cenancêtre

Les résultats obtenus sur l'évolution des enzymes qui participent à la synthèse des phospholipides dans les cellules actuelles montrent que la possibilité d'inférence de la présence de ces protéines chez le cenancêtre est inégale ([Lombard *et al.* 2012b], voir Annexe 7). D'abord, les acyltransférases qui lient les acides gras au G3P et les desaturases qui introduisent des doubles liaisons dans les acides gras bactériens n'ont pas d'homologues capables de réaliser des fonctions comparables chez les archées, ce qui indique que ces protéines sont apparues indépendamment dans la lignée bactérienne. De leur côté, la prenyltransférase DGGGPS et la réductase GGR sont ancestrales aux archées et leur présence chez le dernier ancêtre commun des bactéries ne peut être ni exclue ni confirmée. En effet, leur distribution parmi les bactéries n'est pas très large, comporte des nombreux paralogues et la phylogénie doit être analysée précautionneusement à cause de l'implication fonctionnelle très forte de ces protéines pour la photosynthèse. L'autre prenyltransférase, GGGPS, appartient à une superfamille universelle de protéines contenant le motif TIM barrel et dont le dernier ancêtre commun à la superfamille semble avoir été capable d'utiliser comme substrat ou produit une certaine diversité de trioses phosphate. Ces arguments permettent de proposer qu'une protéine capable de réaliser la fonction GGGPS de façon peu spécifique ait pu exister chez le cenancêtre et que la fonction moderne serait apparue dans la lignée archéenne grâce à la duplication et la sousfonctionnalisation de cette protéine ancestrale. Cependant, la capacité de l'enzyme ancestrale à réaliser cette fonction n'est pas pour autant démontrée. Enfin, les phylogénies soutiennent fortement que le cenancêtre possédait des gènes homologues à ceux qui permettent actuellement l'addition des têtes polaires aux phospholipides. Bien qu'il reste encore à démontrer la fonction biochimique des séquences archéennes, la phylogénie des homologues de CdsA soutient l'ancestralité de ce gène au moins dans les domaines

procaryotes et, par conséquent, sa présence dans le cenancêtre. En ce qui concerne les CDP-alcool phosphatidyltransférases, le gène lié au transfert de sérine semble ancestral chez les bactéries et au moins chez les euryarchées. Le gène qui se charge du transfert de *myo*-inositol-phosphate est ancestral chez les archées et celui responsable de l'addition de glycérol-phosphate est ancestrale chez les bactéries. Une telle distribution permet de postuler qu'au moins un gène, et probablement deux, de la superfamille des CDP-alcool phosphatidyltransférases existaient chez le cenancêtre.

Implications sur la capacité du cenancêtre à fabriquer des membranes lipidiques

Une clé importante dans l'interprétation de ces résultats réside dans la localisation membranaire de toutes les protéines de synthèse des phospholipides à partir de leurs composants, à l'exception de GGGPS. Dans les cellules modernes, cette localisation cellulaire semble très adaptée car elle permet à ces protéines de synthétiser les phospholipides dans le compartiment même auquel ils sont destinés [Zhang et Rock 2008]. Comme nous l'avons vu dans l'introduction, l'inférence de la présence de protéines membranaires chez le cenancêtre est un argument indirect mais important pour soutenir que ce dernier était délimité par des membranes lipidiques. Ceci semble d'autant plus pertinent que ces protéines membranaires sont celles chargées de produire la composante lipidique des membranes.

Bien que des protéines susceptibles de catalyser les premières étapes de synthèse des phospholipides chez le cenancêtre n'aient pas été détectées, l'inférence de la présence des enzymes catalysant les dernières étapes de la voie dans ce dernier implique qu'un moyen d'attacher les chaînes latérales au glycérol phosphate devait y exister. Sans cela, il faudrait invoquer pour la CdsA et les CDP-alcool phosphatidyltransférases ancestrales des fonctions alternatives. Ceci semble difficile dans la mesure où ces enzymes catalysent le transfert de molécules polaires mais qu'elles sont localisées dans les membranes et que peu d'intermédiaires autres que les têtes polaires des phospholipides sont conservés dans les membranes des trois domaines du vivant. Afin d'expliquer la différence de conservation entre les enzymes des premières étapes de synthèse des phospholipides et celles de la fin, on peut rappeler que les premières sont très spécifiques des chaînes d'hydrocarbures et du glycérol phosphate [Soderberg *et al.* 2001 ; Hemmi *et al.* 2004 ; Lu *et al.* 2006], alors que les dernières peuvent utiliser les deux types de phospholipides, quelque soit la stéréochimie du glycérol-phosphate et les chaînes latérales utilisées [Morii *et al.* 2000 ; Morii et Koga 2003].

Cette observation doit être corrélée avec les idées développées dans les parties précédentes de ce travail selon lesquelles le cenancêtre était probablement capable de synthétiser de façon enzymatique tous les composants des phospholipides alors que l'utilisation de chaque molécule dans les lignées modernes semble être le résultat d'un ajustement des molécules à des fonctions précises. On peut donc proposer que, chez le cenancêtre, des mécanismes enzymatiques moins spécifiques permettaient d'abord la synthèse des composants des phospholipides et ensuite la synthèse des phospholipides eux-mêmes. Au cours de l'évolution, des mécanismes plus spécifiques et efficaces de synthèse des composants des phospholipides ont évolué dans les différentes lignées, ce qui s'est répercuté dans les variations de composition entre les phospholipides des archées et celle des bactéries. Concernant les chaînes latérales, le développement de systèmes spécifiques d'attachement au glycérol phosphate est corrélé à une perte de diversité moléculaire dans les phospholipides. Les têtes polaires n'ayant pas beaucoup changé, leur mécanisme d'attachement s'est diversifié sans changer fondamentalement de fonctionnement.

En résumé, il n'y a pas encore d'indices sur les méthodes employées par le cenancêtre pour attacher les chaînes d'hydrocarbures au glycérol phosphate. Il n'est donc pas possible de déterminer complètement comment s'est produit le changement vers les mécanismes des lignées modernes. Il est néanmoins envisageable de proposer avec tous les indices qui ont été rapportés ici que le cenancêtre avait probablement la capacité de synthétiser de façon enzymatique tous les éléments des phospholipides et possédait également une machinerie membranaire pour les assembler.

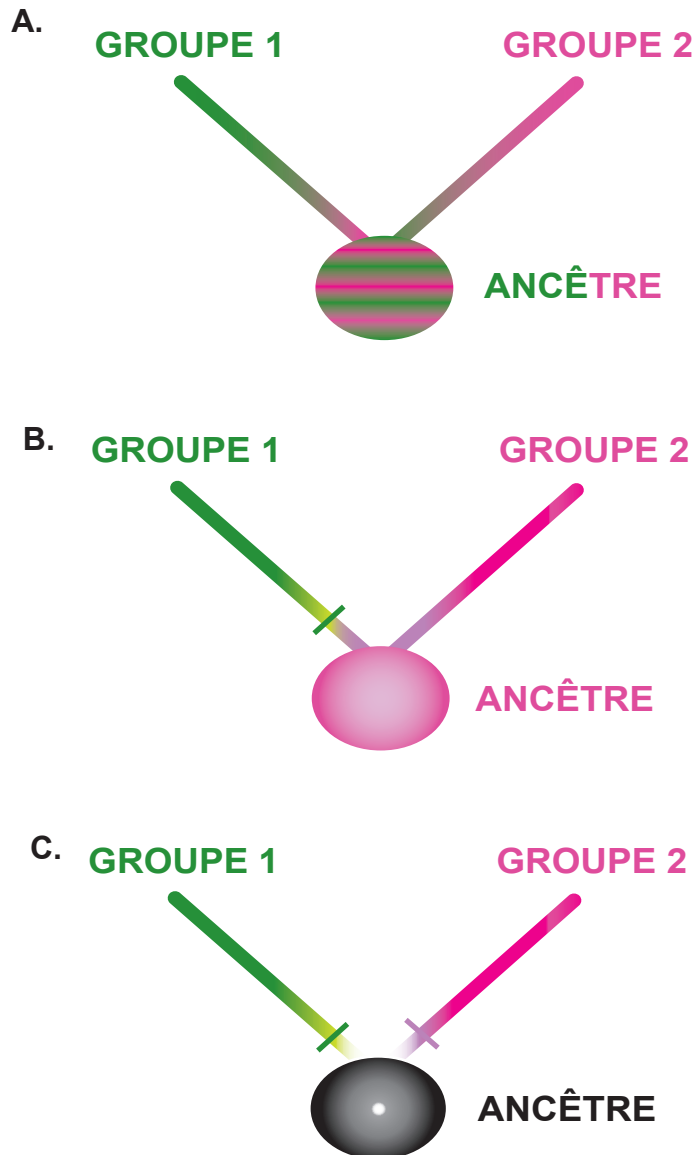


Figure 94. Hypothèses possibles de l'origine de deux structures non-homologues dans des groupes issus d'un ancêtre commun.

A. L'ancêtre avait les deux structures dont une a été perdue dans chaque groupe de descendants.

B. L'ancêtre avait l'une des deux structures et l'autre lignée a subi un remplacement par une nouvelle structure.

C. L'ancêtre n'avait aucune structure et chacune est apparue séparément dans les groupes de descendants.

CONCLUSION ET PERSPECTIVES

1. Sur les descriptions générales du cenancêtre

« *La tâche du sceptique est moins d'expliquer son doute que de combattre les croyances de ceux qui ne doutent pas* »

Victor Brochard (à propos de Sextus, dit Empiricus), *Les sceptiques grecs*, 1887

Bien que le postulat de l'existence du cenancêtre soit une conséquence directe de toutes les bases sur lesquelles repose la biologie actuelle, notre connaissance de cet organisme reste finalement très incomplète. La raison principale de cela est la difficulté à extraire des informations fiables sur des événements aussi anciens à partir de traces qui ont subi autant de changements depuis l'époque du cenancêtre. Cette difficulté explique certainement le petit nombre de tentatives visant à donner un aperçu global du cenancêtre. Nous avons vu, d'une part, que des descriptions rapprochent tout simplement ce à quoi devait ressembler le cenancêtre d'un organisme moderne [Cavalier-Smith 2002 ; Di Giulio 2007a ; Wong *et al.* 2007]. Cependant, ces explications sont inadaptées car leurs choix de caractères ancestraux restent, pour l'instant, arbitraires. D'autre part, le cenancêtre a été décrit comme un progénote [Woese et Fox 1977a], comme une communauté de pré-cellules [Kandler 1995 ; Woese 1998] ou comme un milieu moléculaire contenu dans des cavités minérales [Martin et Russell 2003]. On remarquera que ces hypothèses, bien que fondées sur des observations relativement diverses, ont toutes été engendrées par la difficulté à expliquer l'apparition, à partir d'un ancêtre commun, de systèmes apparemment irréconciliables entre les trois domaines du vivant (ou, dans le cas des progénotes, la grande divergence entre les ribosomes des trois domaines).

Pourtant, lorsque deux structures analogues (avec une même fonction, mais non issues d'un ancêtre commun) s'opposent entre deux groupes, il existe trois interprétations possibles sur le statut de la fonction réalisée par ces structures chez leur dernier ancêtre commun (Figure 94) : (1) l'ancêtre possédait les deux structures et celles-ci ont été alternativement

perdues chez ses descendants ; (2) l'ancêtre possédait l'une des deux structures mais l'un des deux groupes descendants l'a remplacée dans sa lignée par une autre structure assurant la même fonction ; (3) le cenancêtre ne possédait aucune des deux structures et chacune est apparue indépendamment, ce qui fait que l'existence d'une fonction commune est le résultat d'une convergence.

Lorsque l'on parle de cenancêtre, la possibilité que cet organisme ait pu posséder deux systèmes liés à une seule fonction (par exemple des phospholipides ou la machinerie de réplication de l'ADN) est souvent écartée en invoquant le fait que cela ferait du cenancêtre l'organisme ayant le plus de fonctionnalités qui ait jamais existé. La « totipotence » du cenancêtre est explicitement utilisée pour exclure cette possibilité dans certains cas [Woese 1998 ; Koonin et Martin 2005]. Mais ladite possibilité est plus généralement discutée en termes de « complexité » : un cenancêtre avec deux systèmes pour chaque fonction serait beaucoup plus « complexe » que les organismes actuels, ce qui impliquerait une évolution majoritairement basée sur la perte de structures originelles. Cette option est considérée comme très peu probable par de nombreux auteurs [Woese 1987 ; Kandler 1995 ; Martin et Russell 2003 ; Cavalier-Smith 2010]. Face à cette conclusion, certaines hypothèses privilégient un changement tardif dans une lignée [Cavalier-Smith 2002] mais le problème intrinsèque est qu'il est quasiment impossible de déterminer avec certitude et non arbitrairement lequel des deux systèmes modernes serait le plus ancien. Par conséquent, la plupart des auteurs qui proposent des hypothèses sur la nature du cenancêtre préfèrent l'option d'une entité dépourvue de la fonction et une apparition plus tardive des systèmes. Or comme en général les sujets controversés concernent des caractéristiques fondamentales des organismes vivants, tels les ribosomes ou les membranes, la discussion est ramenée à une explication de l'apparition *de novo* de structures primordiales, provoquant irrémédiablement le rapprochement entre origines de la vie et cenancêtre que j'ai déjà critiqué dans l'introduction.

Au vu de mes résultats sur l'évolution de deux systèmes des domaines du vivant soi-disant irréconciliables (les deux systèmes de synthèse de phospholipides), je pense qu'en réalité des nuances sont possibles dans ce débat et rendraient plus souples les frontières entre les trois possibilités sus évoquées. Je traiterai ce thème dans la deuxième partie de cette conclusion, mais je voudrais auparavant m'arrêter brièvement sur l'argument de la « complexité ».

IDEE D'UNE ECHELLE
DES ETRES NATURELS.

L'HOMME.
Orang-Outang.
Singe.
QUADRUPEDS.
Ecureuil volant.
Clouveours.
Autruche.
OISEAUX.
Oiseaux aquatiques.
Oiseaux amphiles.
Poissous volans.
POISSONS.
Poissous rampans.
Anguilles.
Serpens d'eau.
SERPENS.
Limaces.
Limicois.
COQUILLAGS.
Vers crayon.
Tignes.
INSECTS.
Gallinées.
Tens, ou Solitaires.
Polypes.
Orne de Mer.
Sensitive.
PLANTES.
Lichens.
Mouffures.
Champignons, Agarics.
Truffes.
Coraux & Coralloides.
Lithophytes.
Amiante.
Talcs, Gyps, Sélénites.
Ardoiés.
PIERRES.
Pietres figurées.
Crysalisations.
SELS.
Vitriols.
METAUX.
DEMI-METAUX.
SOUFRÉS.
Bicumes.
TERRES.
Terre pure.
EAU.
AIR.
FEU.
Matières plus subtiles.

Figure 95. Exemple d'Echelle des Etres (extrait de Bonnet 1745)

En effet, le fait que la pensée dominante (dans toute sa diversité) privilégie systématiquement un cenancêtre avec le moins de capacités possibles me semble quelque peu suspect, surtout lorsqu'on compare cet avis presque consensuel avec la grande diversité de points de vue qui existe sur d'autres aspects de l'évolution ancienne. Pourquoi la présence de très nombreux éléments chez le cenancêtre serait-elle moins probable que la présence de systèmes essentiels à la vie telle qu'on la connaît ? Je pense que cette opinion est en réalité le produit de quelques préjugés inconscients qui subsistent parfois dans l'étude de l'origine de la vie aux dépens des arguments issus de la comparaison des organismes modernes. Je vais essayer de donner quelques pistes sur les possibles sources de ces préjugés dans le domaine de l'origine de la vie et nous verrons ensuite comment ces interprétations ont été extrapolées au cenancêtre à cause de l'amalgame entre ces deux sujets.

L'Echelle des Etres et la génération spontanée

L'Echelle des Etres est une conception de l'univers qui place tous les éléments existants dans un ordre hiérarchique en fonction de leur degré supposé de perfection (Figure 95). Cette Echelle est très anthropocentrique : l'Homme se situe en haut, au-dessous de Dieu et d'entités spirituelles mais au-dessus de toutes les autres formes vivantes ou non vivantes, lesquelles sont considérées comme d'autant plus perfectionnées qu'elles sont proches de lui. Ce système a été très influent dans l'observation de la nature en Occident pendant deux millénaires mais son histoire et ses déclinaisons sont très riches et outrepassent mes compétences. Je vais donc en parler très brièvement en me basant principalement sur le travail d'Arthur O. Lovejoy [Lovejoy 1936].

Lovejoy observe quatre grandes composantes dans l'organisation de la pensée sur l'Echelle des Etres : la première est l'idée platonicienne de *plénitude* qui suppose que l'univers soit composé d'une diversité d'éléments (reflets des idées) épuisant complètement les possibilités du monde ou, exprimé autrement, que tous les éléments concevables sont présents dans l'Univers. La raison en est que le Bien suprême, précisément à cause de sa suprématie universelle, implique l'existence d'un monde contenant toutes les idées qui le composent. Tout en nuanciant l'idée de plénitude, Aristote a proposé qu'il fût toujours possible de trouver pour chaque élément composant l'univers un autre élément qui ne s'en différencie que par une seule caractéristique : c'est l'idée de *continuité* entre les éléments qui composent le monde. La position d'Aristote est quelque peu ambiguë car, tout en reconnaissant la difficulté d'établir un système pertinent pour classer les éléments qui

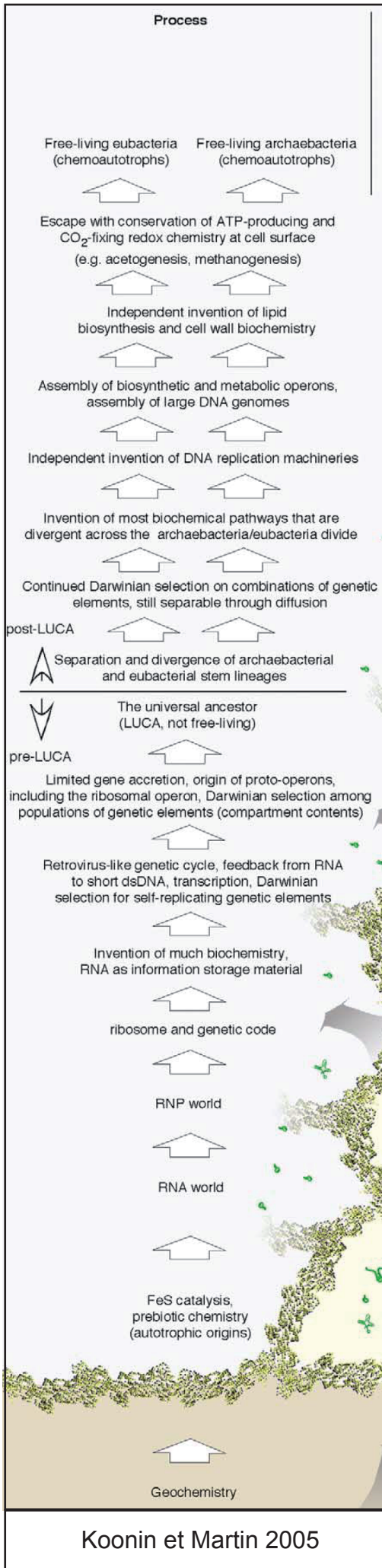


Figure 96. Exemples d'hypothèses sur l'origine de la vie dont une certaine similarité avec l'Echelle des Etres peut être envisagée.

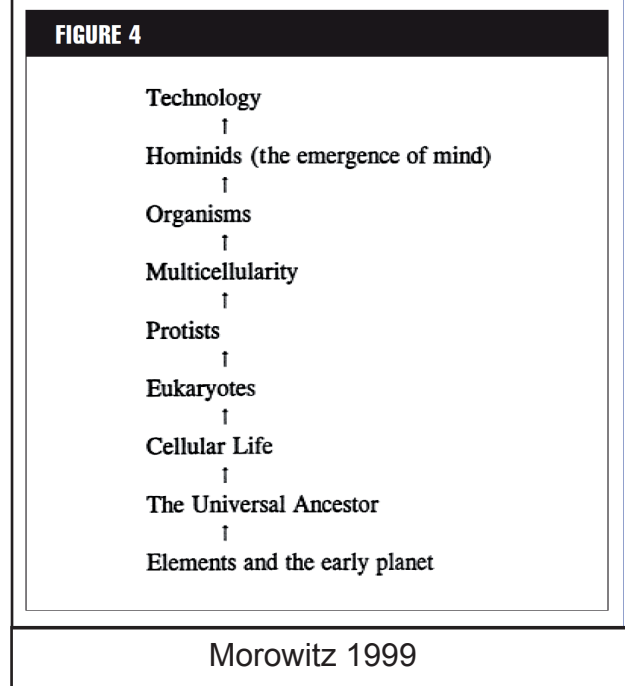
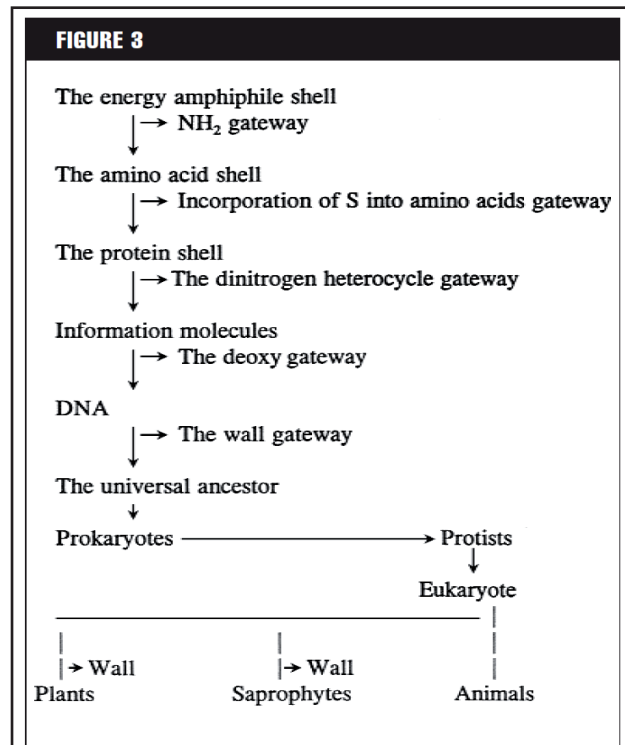


Table 1.2. The major transitions

Replicating molecules	→	Populations of molecules in compartments
Independent replicators	→	Chromosomes
RNA as gene and enzyme	→	DNA + protein (genetic code)
Prokaryotes	→	Eukaryotes
Asexual clones	→	Sexual populations
Protists	→	Animals, plants, fungi (cell differentiation)
Solitary individuals	→	Colonies (non-reproductive castes)
Primate societies	→	Human societies (language)

Maynard-Smith et Szathmáry 1995

composent l'univers, il a été l'initiateur de l'idée selon laquelle, au moins pour les animaux, il était possible de fixer une *gradation* des organismes sur une échelle de perfection. Cette vision, appelée *scala naturae*, a été étendue à l'ensemble des formes animées et inanimées du monde et est devenue le principal système d'explication du monde jusqu'à la formulation de la théorie darwinienne. Le système de l'Echelle des Etres a cependant changé au cours du temps, notamment aux XVIII^{ème} et XIX^{ème} siècles, quand il a été proposé que tous les éléments de l'Echelle puissent ne pas exister simultanément mais se transformer les uns dans les autres, processus que l'on appelle la *temporalisation*. Il se produit à la même époque un changement radical dans le sens de l'Echelle. Alors que pendant longtemps on l'avait considérée comme une conséquence « descendante » de l'existence du Bien (ou de Dieu), l'Echelle s'est alors transformée en un *devenir*, la gradation manifestant une tendance naturelle au perfectionnement. Cette vision explique la popularité de ce système spécialement au XIX^{ème} siècle car elle faisait écho aux idéaux de progrès de cette époque [Lovejoy 1936 ; Ruse 1996].

La théorie darwinienne de l'évolution, en expliquant l'apparition des organismes sur la base de la descendance avec modification et de la sélection naturelle (ou autres moteurs évolutifs), s'oppose directement au système de l'Echelle des Etres. Selon l'évolution darwinienne, les organismes ne représentent qu'une petite diversité parmi toutes les potentialités qui existent, ils ne constituent pas des éléments éternels mais dynamiques, leurs relations sont généalogiques et il n'y a aucune gradation ni perfectionnement linéaire. Au contraire, ce que l'on observe ce sont des adaptations temporaires et buissonnantes à des niches écologiques changeantes. Depuis, l'Echelle a été exclue des réflexions biologiques mais ce système est resté présent dans la société sous une forme similaire à celle du XIX^{ème} siècle, basée sur la notion de progrès et sur l'idée de la supériorité de certains organismes par rapport à d'autres [Gould 1989 ; Ruse 1996].

Je pense que, malgré l'exclusion de l'Echelle des Etres de la biologie, un schéma très similaire a été transposé dans beaucoup d'hypothèses sur l'origine de la vie [Wächtershäuser 1988 ; Morowitz 1992 ; Maynard-Smith et Szathmary 1995 ; Morowitz 1999 ; Koonin et Martin 2005 ; Koonin 2007] (Figure 96). Cette transposition est peut-être explicable par la parenté lointaine qui existe entre l'étude des origines de la vie et les théories anciennes de la génération spontanée, celles-ci ayant toujours eu un rapport avec l'Echelle des Etres.

En effet, la continuité proposée par Aristote a été pendant longtemps l'un des principaux arguments en faveur de la transition possible entre des formes non vivantes et vivantes. Pendant très longtemps on a cru voir dans les petits animaux que l'on pensait « inférieurs » (vers, mouches, etc.) des formes de vie générées nouvellement à partir de la matière morte. Lorsqu'on a découvert que ces animaux étaient plus « complexes » que ce que l'on pensait précédemment et qu'ils avaient besoin de parents pour exister, l'idée de génération spontanée a été abandonnée en ce qui concernait les animaux. Cette idée a alors été transposée à des organismes encore plus petits (encore plus « inférieurs »), les microorganismes. Mais, d'une part, Pasteur a démontré que les microbes aussi sont issus d'autres microbes et, d'autre part, on s'est aperçu que les cellules de ces microorganismes étaient également très « complexes ». De fait, l'idée que la vie se génère *de novo* en permanence a été abandonnée [Oparin 1938 ; Rostand 1943] et le problème de l'apparition de la vie s'est alors posé uniquement à propos des origines anciennes de la vie actuelle [Oparin 1938].

Quels éléments peuvent faire penser que certaines hypothèses sur les origines de la vie ont pu conserver des postulats similaires à ceux fondant l'Echelle des Etres ? D'abord, ces théories conservent l'idée que les premiers organismes devaient forcément être très « simples » et que la « complexité » des organismes suivants a augmenté progressivement puisque les composants primordiaux devaient, logiquement, être plus rudimentaires que la machinerie sophistiquée actuelle, celle-ci étant issue d'une longue évolution. Ceci se traduit dans une certaine gradation, puisque chaque étape est plus perfectionnée (plus « complexe ») que l'étape précédente, et comporte aussi une composante temporelle forte, puisque chaque nouvelle étape annonce l'étape suivante. Ces hypothèses respectent aussi l'idée de continuité puisque chaque étape contenant un certain fonctionnement moléculaire est généralement séparée de la suivante par le rajout d'un nouvel élément. La notion de plénitude est plus subtile car elle a perdu son caractère abstrait et est devenue une limitation logique : la chimie des organismes sert de modèle pour déterminer le champ des possibles rempli par l'hypothèse, donc toutes ces théories n'expliquent qu'une séquence unique et nécessaire. Même lorsqu'elles acceptent des embranchements, il est clair qu'il est très difficile de parler d'événements qui n'ont pas conduit à l'émergence de la vie que l'on connaît.

Ce rapprochement entre les hypothèses sur l'origine de la vie et les idées de génération spontanée ou d'Echelle des Etres peut sembler embarrassant pour les biologistes actuels car on sait que ces idées ont pesé lourdement sur notre compréhension de l'histoire naturelle et

elles sont maintenant connotées d'une façon très négative. Ce parallèle est gênant aussi parce que, pendant longtemps, les doctrines de la génération spontanée et de l'Echelle des Etres ont servi de fondement à l'explication religieuse du monde. Mais certains auteurs ont essayé de dépasser ces critiques épistémologiques et ont proposé que, sans faire intervenir une force supérieure, l'on puisse détecter dans l'histoire de la vie une tendance linéaire du « simple » au « complexe » : si la vie a commencé d'une façon très « simple », alors en variant dans toutes les directions certaines lignées ont forcément augmenté leur « complexité » [Cavalier-Smith 2010] a utilisé cet argument pour l'origine de la vie, mais cette idée a aussi été utilisée plus généralement pour l'apparition des organismes multicellulaires [Maynard Smith 1970] cité dans par McShea en 1991.

Cependant, la question reste ouverte car, notre connaissance de l'origine de la vie étant très parcellaire, comment pouvons-nous être sûrs que la vie a commencé aussi « simplement » ? On peut considérer que les hypothèses sur les origines de la vie critiquées ici ne sont pas comparables à l'Echelle des Etres car elles n'exigent pas l'intervention d'un « esprit organisateur ». Néanmoins, la recherche d'une tendance inévitable de la chimie organique à suivre des étapes linéaires imaginées *ad hoc* est-elle si différente de l'invocation d'un « esprit organisateur » ? Quand on connaît l'influence de l'idée d'échelle de progrès dans le passé et dans la société actuelle, il paraît pour le moins suspect que la majorité des théories sur l'origine de la vie retiennent uniquement l'hypothèse d'une apparition des fonctions biologiques par paliers de « complexité », se contentant de discuter de la nature et de l'ordre des paliers. Etant donné la nature de la question posée, disposons-nous des informations nécessaires pour y répondre ?

Biais expérimental

Alexandr I. Oparin est le principal fondateur des études modernes sur l'origine de la vie⁴⁹. Dans un livre publié en 1938, il analyse l'histoire du concept de génération spontanée et conclut que la prise de conscience de la « complexité » de *tous* les organismes actuels, même ceux qui étaient considérés comme « inférieurs », a été l'une des principales causes de l'abandon de cette doctrine : les organismes actuels sont trop « complexes » pour que l'on

⁴⁹ On parle souvent aussi de J. B. S. Haldane comme un des pionniers du domaine à cause de son article de 1929. Cependant, le sujet des origines de la vie n'était pas du tout le principal sujet d'intérêt pour Haldane et son analyse est beaucoup plus courte, incomplète et ne fournit pas le même type de programme de recherche que les hypothèses d'Oparin.

puisse imaginer leur apparition immédiate à partir de molécules désorganisées. Dans son exposé, Oparin explique que la disparition de l'idée de la génération spontanée comme hypothèse explicative a provoqué deux réactions : d'une part, le développement de théories vitalistes soutenant que la matière vivante est d'une nature différente de celle de la matière non vivante, ce qui n'a jamais reçu d'argument positif et est contredit par la composition chimique banale des organismes vivants ; d'autre part, l'expansion de l'idée de panspermie qui implique que la vie sur Terre soit d'origine extraterrestre mais qui ne fournit pas non plus une explication de la formation des premiers organismes. Ensuite, Oparin présente sa propre théorie de l'origine de la vie. Pour lui, le plus probable est que les premières formes de vie soient apparues à partir d'une combinaison et longue évolution de molécules organiques présentes sur la Terre primitive. Ainsi, il expose les points principaux qui vont constituer le programme de recherches sur l'origine de la vie : l'origine des molécules préalables ; la multiplicité de réactions possibles entre elles conduisant à l'apparition de molécules plus grandes avec des propriétés diverses ; la concentration dans le milieu et la formation d'un colloïde ; le mélange de ces molécules jusqu'à former des coacervats ; et l'individualisation d'entités, qui rentreront en compétition. Les exemples utilisés par Oparin peuvent, aujourd'hui, sembler un peu datés, dans la mesure où à cette époque on ne connaissait ni l'ADN ni la nature de la membrane, mais sa réflexion a été la structure pour la construction des recherches postérieures.

Son apport essentiel a été d'affirmer clairement qu'il était possible d'explorer l'origine de la vie en étudiant les propriétés physico-chimiques de la matière. Cette idée a été importante car elle a permis de découper la question de l'apparition de la vie en une infinité de petites questions auxquelles on est capable de répondre expérimentalement, apportant ainsi des éclaircissements certains sur la nature des molécules et des forces en jeu. Mais, bien que les avancées dans le domaine depuis les années 1930 soient claires, on peut s'inquiéter de l'aspect très mécanistique de ces développements. L'auteur qui a formulé cette nécessité mécanistique de la façon la plus claire est probablement Morowitz dans un livre publié en 1992. Morowitz commence son livre en disant qu'il existe trois manières d'envisager l'origine de la vie : la première est de considérer que la vie est le résultat d'une activité divine et que, par conséquent, son origine échappe au domaine de la science ; la deuxième est de considérer que l'origine de la vie est explicable par les lois physico-chimiques mais qu'elle n'est pas déductible, i. e. que la vie serait apparue suite à un certain enchaînement d'événements historiques et que, par conséquent, cette apparition ne serait pas reproductible ; enfin,

l'origine de la vie pourrait être le résultat déterministe des lois de la physique (l'aspect historique serait apparu plus tard), ce qui permettrait une certaine reproductibilité.

La question du déterminisme et du hasard est une question philosophique fondamentale qui n'a pas lieu d'être traitée ici. Mais je voudrais attirer l'attention sur le fait que Morowitz conclut que l'explication « théologique » et l'explication « historique » ne sont pas étudiables par la science expérimentale et que, par conséquent, seule l'étude déterministe des fonctions physico-chimiques peut être explorée⁵⁰. Comme le nombre de paramètres physico-chimiques qui peuvent entrer en jeu dans ce genre d'étude est très important, pour avoir une réponse claire il faut généralement s'astreindre à poser les questions une à une, ce qui se traduit par un réductionnisme très important.

Il est difficile de savoir si l'avis de Morowitz sur l'interprétation des trois manières d'envisager les origines de la vie est répandu ou non parmi les scientifiques qui travaillent dans ce domaine. Cependant, il est facilement observable que beaucoup d'hypothèses dans ce domaine ont une teinte déterministe très marquée [Wächtershäuser 1988 ; Morowitz 1992 ; Maynard-Smith et Szathmary 1995 ; Morowitz 1999 ; Koonin et Martin 2005 ; Koonin 2007]. La partie de mon introduction sur l'origine prébiotique des membranes permet d'illustrer le fait que, pour l'instant, la plupart des travaux sur les origines de la vie concernent des preuves de principe réductionnistes. Or ce réductionnisme se traduit fatalement par une certaine déconnexion entre les résultats et de grands débats surgissent alors pour essayer de trouver un ordre à tous ces résultats. Qui de l'information génétique ou du métabolisme est apparu en premier ? Les premiers organismes étaient-ils hétérotrophes ou autotrophes ? Quelle partie du code génétique et de la machinerie de traduction est apparue en premier ? A quel moment les premières membranes sont-elles entrées en jeu ? Le plus souvent, les hypothèses sur l'origine de la vie commencent en considérant l'un de ces éléments, qui est censé être le plus élémentaire de tous les composants du vivant, auquel s'ajoutent progressivement les autres éléments grâce à l'apparition de nouvelles structures supposées plus tardives ou remplaçant des structures antérieures. En bref, ces hypothèses considèrent généralement les premières structures comme très « simples » et imaginent comment des ajouts progressifs ont permis

⁵⁰ Cependant cette idée est fautive : en soit, une étude historique est scientifique tant qu'elle reste réfutable. Dans le cas qui nous concerne ici, il est vrai qu'aujourd'hui nous n'avons pas accès à des registres historiques de cette étape, mais la disqualification de l'approche historique est conceptuellement gênante.

l'apparition de structures de plus en plus « complexes » jusqu'à l'apparition des organismes modernes (qui sont souvent confondus avec le cenancêtre)

En soi, la possibilité que les premières formes de vie soient apparues par « l'accrétion » progressive de composants est crédible mais elle n'est pas la seule. Ce qui est étonnant est que quasiment toutes les théories sur les origines de la vie appliquent un schéma similaire, basé sur une observation ou sur la prédiction d'une expérience suivie d'une liste d'autres expériences qui pourraient s'ajouter progressivement à la première. Mais le fait que l'on soit méthodologiquement obligés d'appliquer des méthodes réductionnistes n'implique pas que les premiers organismes vivants soient apparus au terme d'un processus aussi simpliste. Au contraire, Oparin lui-même, en 1938, tout en reconnaissant les limitations méthodologiques, invitait à l'étude des propriétés émergentes de la « chimie sale » :

“Le chimiste, et spécialement le physico-chimiste, préfère toujours travailler avec des substances pures chimiquement isolées, car c'est le seul moyen dont il dispose pour étudier leur nature et propriétés. Cela explique l'attitude traditionnellement méprisante des chimistes pour les mélanges ou “substances sales”. Jusqu'il y a très récemment, on considérait que le plus important pour un chimiste était d'étudier des composés individuels, et de considérer les propriétés des mélanges comme la somme des propriétés de leurs composants. Aujourd'hui cette position est intenable. [...] [Il cite Willstätter :] Les mélanges peuvent en réalité présenter la nature de nouveaux composés chimiques”⁵¹.

Heureusement, comme on l'a vu dans la partie sur les membranes, certains travaux récents commencent déjà à essayer d'imaginer comment des mélanges « sales », qui sont plus plausibles dans la Terre primitive, ont pu conduire à des réactions d'intérêt biologique [Segre *et al.* 2001 ; Szostak 2011].

Définition du « complexe »

A mon sens, l'une des difficultés majeures que l'on rencontre dans ce contexte est que lorsqu'on parle d'une « augmentation de la complexité » on croit comprendre très facilement ce à quoi on fait référence alors qu'en réalité la définition rigoureuse du mot « complexe » est très difficile. En l'occurrence, cette définition n'est probablement pas vraiment une source de biais (c'est plutôt une conséquence) mais elle présente les limites d'utilisation d'un concept

⁵¹ Ma traduction.

polyvalent que l'on utilise généralement sans spécifier le sens que l'on veut lui donner. Les difficultés pour définir la « complexité » sont bien connues et ont été très explorées depuis une vingtaine d'années, lorsqu'une discipline a essayé de se constituer autour de ce concept [Waldrop 2008]. Malgré ces efforts, aucun consensus n'a été trouvé sur cette définition et un travail a même listé 42 définitions différentes du concept de « complexité » dans différents domaines [Lloyd 2001]. Cependant, étant donné que la plupart des personnes qui ont proposé des scénarios pour l'apparition de la vie ne sont pas expertes du sujet de la « complexité » je pense qu'il convient de se référer aux deux sens les plus habituels de ce terme tels qu'ils apparaissent dans un dictionnaire commun [Robert 1984] :

Complexe (*Adjectif*)

1. [Grammaire, Mathématiques] Qui contient, qui réunit plusieurs éléments différents.
2. [Courant⁵², abusivement⁵³] Difficile à cause de sa complication.

En général, lorsque le premier sens du mot est utilisé dans la littérature, les articles correspondants explicitent les éléments dont ils parlent. Dans le cas du cenancêtre, les éléments dont on parle sont généralement des gènes donc lorsqu'on dit que l'on étudie la « complexité » du cenancêtre, on se réfère au nombre de gènes qu'il comporte [Koonin 2003]. Dans d'autres cas, même si le mot utilisé n'est pas explicitement défini, il semble assez clairement faire référence à un nombre de gènes [Ouzounis *et al.* 2006 ; Glansdorff *et al.* 2008]. Etant donné la confusion qui a existé autour du *paradoxe de la valeur C*, qui s'étonnait du fait que la taille des génomes et le nombre de gènes ne se corrélait pas avec la « complexité » apparente des organismes [Gregory 2001], on peut se demander quel est l'intérêt d'employer le terme général « complexité » à la place de la formule plus neutre « nombre de gènes ». Cependant, comme la définition est relativement claire, cette acception ne semble pas vraiment problématique.

La difficulté apparaît lorsque le mot est employé dans son autre acception [Woese et Fox 1977a ; Wächtershäuser 1988 ; Koonin et Martin 2005 ; Cavalier-Smith 2010]. Certes, dans certains cas, le terme « complexe » est aisément utilisable dans le sens de « difficile à cause de sa complication ». Par exemple, en disant qu'une cellule est « complexe », on exprime l'idée qu'il est difficile d'expliquer son fonctionnement à cause de sa complication.

⁵² L'emploi *courant* d'un mot correspond à une utilisation commune, par opposition à un terme savant, qui en ce cas serait la définition grammaticale ou mathématique.

⁵³ *Abusivement* indique que l'emploi du mot est critiquable, qu'il peut comporter des faux sens.

En revanche, il est beaucoup plus difficile d'utiliser cette acception pour une comparaison car pour comparer deux choses on devrait être capable de les décrire en détail alors qu'on utilise cette acception de « complexe » précisément lorsqu'on n'arrive pas à décrire quelque chose. Cette difficulté me semble insurmontable lorsqu'il s'agit de comparer deux entités dont on connaît si peu de choses : les premières formes de vie⁵⁴ et le cenancêtre. Comment peut-on affirmer qu'il y a eu une augmentation de la « complexité » de ces premières formes jusqu'au cenancêtre si on ne connaît ni ces premières formes ni le cenancêtre ? Ou, pour revenir au sujet de départ de cette conclusion : comment peut-on exclure la possibilité que deux structures analogues aient coexisté chez le cenancêtre au motif que cela ferait de lui un organisme trop « complexe » si l'on n'a pas une définition précise (ni aucune preuve) de cette « complexité » ?

Tabula rasa

J'ai commencé cette conclusion en expliquant comment l'exclusion de la possibilité que deux structures cellulaires fondamentales analogues se trouvent chez le cenancêtre conduit certains auteurs à penser que ces structures fondamentales sont apparues après le cenancêtre, rapprochant ainsi le cenancêtre et les origines de la vie. Comme les hypothèses qui en résultent décrivent le cenancêtre comme une étape de plus dans l'apparition de la vie, le cenancêtre se retrouve alors dans un contexte aussi incertain que les autres étapes de la vie primordiale. Je pense notamment que ces hypothèses intègrent une variante moderne de l'Echelle des Etres dans laquelle la progression vers la perfection a été remplacée par une échelle de « complexité ». Je pense aussi que les biais expérimentaux ont aidé à considérer l'apparition de la vie comme un processus allant du plus « simple » (que l'on arrive à reproduire en laboratoire) au plus « complexe » (que l'on n'arrive plus à reproduire). Le cenancêtre, considéré comme un chaînon de plus dans l'apparition de la vie, devient alors une étape plus « complexe » que les précédentes mais moins que les suivantes, conformément au postulat d'origine de ce type de scénarios. Je pense que ce type de raisonnement n'est pas approprié car il se base sur un terme non défini et des préjugés.

⁵⁴ Il me semble important de rappeler que le terme même d' « origine » de la vie est très ambigu, car dans les événements historiques, et même avec une grande dose de déterminisme, l'apparition de la vie a forcément été un événement historique : on pourra donc toujours identifier de très nombreuses origines plus ou moins directes [Morange 2011].

Si l'on parle de « complexité » en référence au nombre de composants, l'idée que leur nombre augmente au cours du temps dans les premiers précurseurs du vivant est très plausible. Il semble aussi raisonnable d'imaginer que les éléments qui interagissaient ensemble dans les premiers systèmes vivants étaient des molécules plus petites que les macromolécules modernes, que leurs interactions étaient moins spécifiques et que l'apparition de la vie a nécessité l'intégration de plus en plus forte des molécules dans un réseau qui a progressivement incorporé des nouvelles molécules mais aussi écarté des anciennes. Evidemment, il y a eu une continuité généalogique entre les premières formes de vie et le cenancêtre. Si le cenancêtre n'avait que 500 gènes, alors on pourra dire qu'il avait moins de gènes que la plupart des organismes actuels (certains diront qu'il est moins « complexe »). Mais cette affirmation doit être la *conclusion* d'un travail complet sur la nature du cenancêtre et non pas le *point de départ* de l'analyse.

L'alternative à l'acceptation d'emblée que le cenancêtre était « simple » est une étude approfondie des relations évolutives entre les organismes modernes comme celle que j'ai essayé de réaliser en étudiant la synthèse des phospholipides membranaires. Ce chemin est plus long et conduira certainement à de nouvelles impasses mais je pense qu'il s'agit de la démarche la plus justifiée étant donnée que c'est à partir des organismes modernes que l'on définit le cenancêtre.

2. Le statut des membranes chez le cenancêtre et l'évolution dans les trois domaines du vivant

« [...] la nature, tout en satisfaisant si largement à la loi de la diversité des organismes, n'a pas eu recours à toutes les combinaisons physiologiques qui auraient été possibles. Elle se montre, au contraire, toujours sobre d'innovations. On dirait qu'avant de recourir à des ressources nouvelles elle a voulu épuiser, en quelque sorte, chacun des procédés qu'elle avait mis en jeu ; et autant elle est prodigue de variété dans ses créations, autant elle paraît économe dans les moyens qu'elle emploie pour diversifier ses œuvres. »

Henri Milne Edwards, *Introduction à la zoologie générale*, 1851

« As Milne Edwards has well expressed it, nature is prodigal in variety, but niggard in innovation. Why, on the theory of Creation, should this be so? Why should all the parts and organs of many independent beings, each supposed to have been separately created for its proper place in nature, be so invariably linked together by graduated steps? Why should not Nature have taken a leap from structure to structure? On the theory of natural selection, we can clearly understand why she should not; for natural selection can act only by taking advantage of slight successive variations; she can never take a leap, but must advance by the shortest and slowest steps⁵⁵ ».

Charles Darwin, *On the Origins of Species*, 1859

⁵⁵ « Pour employer les expressions de Milne Edwards, la nature est prodigue des variétés, mais avare d'innovations. Pourquoi, dans l'hypothèse de la création, y aurait-il autant de variétés et si peu de nouveautés réelles ? Pourquoi toutes les parties, tous les organes de tant d'organismes indépendants, créés, suppose-t-on, séparément pour occuper une place séparée dans la nature, seraient-ils si ordinairement reliés les uns aux autres par une série de gradations ? Pourquoi la nature n'aurait-elle pas passé soudainement d'une conformation à une autre ? La théorie de la sélection naturelle nous fait comprendre clairement pourquoi il n'en est point ainsi ; la sélection naturelle, en effet, n'agit qu'en profitant de légères variations successives, elle ne peut donc jamais faire de sauts brusques et considérables, elle ne peut avancer que par des degrés insignifiants, lents et sûrs ». [Darwin 1859].

Le cenancêtre avait un mécanisme enzymatique de synthèse des phospholipides

Dans l'introduction on a vu qu'un grand nombre d'indices indirects suggèrent que le cenancêtre devait être délimité par une membrane. En effet, il est difficile d'expliquer l'intégration de complexes moléculaires avec des nombreux composants, comme les ribosomes et les ATPases, si ces composants ne faisaient pas déjà partie, avant le cenancêtre, d'entités individualisées. Or aucune alternative de cellularisation n'est convaincante et le fait que la présence d'un grand nombre de protéines membranaires, liées notamment aux gradients transmembranaires, aient pu être inférée chez le cenancêtre suggère très fortement que cet organisme devait avoir une membrane avec des propriétés similaires aux membranes actuelles. Cependant, l'hypothèse que le cenancêtre était délimité par des membranes est sans conséquence sur les hypothèses relatives au mécanisme particulier de leur synthèse. Celui-ci aurait pu être abiotique, en mettant en œuvre l'un des moyens prébiotiques que l'on a vu dans l'introduction, ou enzymatique, comme les mécanismes actuels des bactéries/eucaryotes et/ou des archées.

Dans mon travail j'ai exploré les informations contenues dans les génomes afin de déterminer s'il existe des indices d'une machinerie enzymatique présente chez le cenancêtre. Mes résultats ont montré que la synthèse de chaînes d'isoprénoïdes semble très probable grâce à une voie du mévalonate capable de synthétiser les précurseurs et à des isoprényl diphosphate synthases synthétisant les chaînes à partir de ces précurseurs. La voie du mévalonate présente dans le cenancêtre était probablement similaire à celle qui est aujourd'hui caractéristique des eucaryotes, même si un doute persiste sur la nature de la fin de la voie car il est difficile de déterminer si les trois réactions qui permettent la transformation de mévalonate en isopentenyl pyrophosphate étaient réalisées par une seule protéine de la superfamille des GHMP kinases ou si chaque protéine avait déjà divergé à l'époque du cenancêtre. La synthèse des acides gras par une machinerie sans ACP similaire à celle que nous avons proposé chez les archées semble également probable chez le cenancêtre.

La synthèse enzymatique par des représentants de deux superfamilles de protéines capables d'utiliser plusieurs substrats des deux énantiomères de glycérol phosphate avait déjà été proposée avant mon travail [Peretó *et al.* 2004]. J'ai donc concentré la suite de mon analyse sur les enzymes importantes pour la synthèse des phospholipides à partir de leurs composants. Les enzymes qui lient les acides gras et le glycérol phosphate dans les bactéries n'ont pas d'homologues archéens et leur présence ne peut pas être inférée chez le cenancêtre.

Des deux protéines archéennes qui lient les isoprénoïdes au glycérol phosphate une seule semble avoir eu un ancêtre chez le cenancêtre mais sa fonction est très incertaine. En revanche, les enzymes qui réalisent l'addition ultérieure de la tête polaire au glycérol phosphate lié aux chaînes latérales semblent avoir été présentes chez le cenancêtre.

En résumé, mes résultats suggèrent la présence chez le cenancêtre d'un mécanisme enzymatique pour toutes les réactions qui semblent nécessaires à la synthèse des composants des phospholipides et à leur assemblage postérieur, à l'exception de la liaison des chaînes carbonées au glycérol phosphate qui est incertaine. Malgré ce point d'interrogation, étant donné le besoin théorique d'un mécanisme de synthèse de phospholipides chez le cenancêtre, ces résultats suggèrent que ces voies pouvaient participer à la production de membranes dans ledit organisme. Pour la liaison des chaînes carbonées au glycérol phosphate il est intéressant de signaler que le mécanisme de synthèse des liaisons éther dans les bactéries est actuellement inconnu mais qu'il pourrait fournir des informations lorsqu'il sera décrit. Sinon, le plus probable est que l'un des mécanismes actuels soit ancestral et qu'il ait été remplacé dans une des lignées par un système plus adapté (on peut par exemple proposer que les acyltransférases soient plus récentes puisqu'elles font en partie intervenir l'ACP, qui semble avoir été mis en place plus tard dans la lignée bactérienne). Cependant, comme dans toutes les hypothèses de remplacement, il est très difficile de déterminer quelle transférase de chaîne carbonée est ancestrale et ce point reste donc en suspens.

Le bricolage de l'évolution

Ces résultats ne permettent pas seulement de proposer un mécanisme enzymatique plausible pour la synthèse de phospholipides chez le cenancêtre, mais ils changent aussi notre vision de la distribution de ces molécules dans les trois domaines du vivant. Lorsque j'ai commencé mon travail, les distributions des deux types de phospholipides et de leurs composants semblaient irréconciliables entre les archées d'une part et les bactéries et eucaryotes d'autre part.

Mes résultats soulignent l'universalité des molécules dont j'ai discuté bien que leurs fonctions soient très variables. Les acides gras, qui étaient largement ignorés chez les archées, semblent avoir un rôle important et méconnu dans la physiologie de celles-ci, autant pour la formation de phospholipides que dans l'acylation de protéines et probablement la synthèse de molécules de réserve énergétique et de carbone. Les isoprénoïdes jouent des rôles très importants dans tous les organismes même si chaque domaine semble avoir modifié ou

remplacé la voie ancestrale. La réduction des isoprénoïdes avec des enzymes homologues est d'ailleurs très répandue même si les molécules produites ont des fonctions différentes dans chaque taxon. Les molécules qui constituent les têtes polaires sont des métabolites très courants dans les cellules. Par conséquent, la proposition de l'existence de mécanismes de synthèse de l'ensemble de ces molécules dans le cenancêtre dépasse très largement la question de la présence des phospholipides dans cet organisme et ouvre la possibilité que ces molécules aient été utilisées de façon généraliste dans une pléiade d'autres fonctions.

Ainsi, il ne semble pas nécessaire d'invoquer une apparition tardive des voies métaboliques originales dans chaque domaine du vivant, comme dans les théories proposant l'absence de membranes chez le cenancêtre ou la synthèse abiotique de ces membranes [Koga *et al.* 1998 ; Martin et Russell 2003 ; Wächtershäuser 2003]. Au contraire, ces résultats suggèrent que le cenancêtre possédait une machinerie, peut être moins efficace et moins spécifique que celles des organismes actuels, mais qui permettait la synthèse d'une petite diversité de molécules disponibles pour assurer un grand nombre de fonctions. Ils suggèrent ensuite que l'évolution postérieure dans les trois domaines a principalement consisté en un réassortiment de ces activités initiales, en leur spécialisation et leur augmentation d'efficacité. Cette observation n'est pas la première dans ce sens. Par exemple, dans la partie de l'introduction concernant les protéines membranaires du cenancêtre, nous avons vu qu'un jeu de gènes impliqués dans les chaînes respiratoires semble avoir été présent chez le cenancêtre. La grande diversité de ces chaînes, qui existe actuellement dans les trois domaines du vivant, est probablement le résultat d'une combinatoire et adaptation à partir de quelques protéines ancestrales [Schoepp-Cothenet *et al.* 2012].

L'idée que c'est le remaniement permanent plutôt que l'apparition de nouvelles structures qui permet l'apparition de nouvelles fonctions et de nouveaux organismes est souvent exprimée par la métaphore du *bricolage* développée par François Jacob [Jacob 1981]. Cette métaphore s'oppose à celle de l'*ingénieur* invoquée par les créationnistes : un ingénieur construit en connaissance de cause, à partir d'un plan prédéfini, alors qu'un bricoleur utilisera ce qui sera à sa portée pour répondre à ses besoins et le résultat sera souvent imprévu et incertain. Le recours à la métaphore est toujours discutable en science et doit être utilisé avec beaucoup de précaution mais dans ce cas elle ne fait que traduire en langage courant une des implications majeures de l'évolution darwinienne : la sélection naturelle (et les autres moteurs de l'évolution) s'applique à la diversité des formes qui existent et de nouvelles structures ne peuvent apparaître qu'à partir de structures préexistantes. Comme Darwin lui-même l'a

signalé, son mécanisme d'évolution permet ainsi d'expliquer l'apparent paradoxe bien connu des naturalistes, qui oppose la grande diversité de formes du vivant et le petit nombre de structures originales. Nous voyons dans mes résultats un excellent exemple du fait que cette observation qui au XIX^{ème} siècle était morphologique, se retrouve ici au niveau moléculaire : les trois domaines sont peut être très divers mais ils n'ont fait qu'utiliser différemment les éléments métaboliques qui étaient à leur disposition afin de répondre à leurs besoins.

Ce « bricolage » ne semble pas concerner uniquement l'utilisation des métabolites dans de nombreuses fonctions mais également les enzymes elles-mêmes. Les synthèses de phospholipides montrent de nombreux exemples de recrutement d'enzymes à partir de superfamilles plus larges : les glycérol phosphate dehydrogénases, les différentes thiolases, les GHMP kinases, toutes les enzymes de la voie de synthèse des acides gras, etc. L'histoire de ces enzymes est intéressante car elle évoque l'évolution de la spécificité enzymatique. Traditionnellement, les caractérisations enzymatiques ont pour objectif de déterminer la réaction d'une protéine et tiennent peu compte de la possibilité que les enzymes puissent utiliser des substrats relativement variés en fonction de l'état physiologique de la cellule. Pourtant, certains travaux ont montré que le nombre de protéines à fonction versatile⁵⁶ chez *E. coli* s'élèverait à 37% des enzymes et que ces enzymes peuvent être responsables de jusqu'à 65% des fonctions métaboliques dans cet organisme [Nam *et al.* 2012]. Les enzymes spécifiques sont plus souvent essentielles et participent généralement à des fonctions dont le flux métabolique est très fort mais la majorité des réactions ne répond pas à cette description et n'a donc pas besoin de protéines affectées en permanence à une seule fonction.

La versatilité des protéines est importante car depuis longtemps il a été proposé que les enzymes qui aujourd'hui sont très spécialisées doivent logiquement être issues d'enzymes anciennement versatiles mais qui auraient été recrutées soit directement en les détournant de leur activité initiale, soit après une duplication [Jensen 1976]. Le passage inverse de protéines spécialisées vers des fonctions plus générales a aussi été décrit puisque certaines mutations permettent l'élargissement de l'activité à d'autres substrats sans changer l'efficacité initiale de la protéine [Aharoni *et al.* 2005]. Dans tous les cas, il semble que la majorité des enzymes spécialisées proviennent d'enzymes plus versatiles, soit primairement soit secondairement.

⁵⁶ Le terme anglais « promiscuous » est probablement celui qui décrit le mieux la notion que l'on veut décrire ici, i. e. la capacité d'une enzyme à utiliser plusieurs substrats ou à réaliser un petit groupe de réactions. Ce terme est difficile à traduire en français, donc j'utiliserai le terme de versatilité.

Les homologues des enzymes de la synthèse des acides gras et des carboxylases à biotine semblent en être de bons exemples ([Lombard *et al.* 2011a]).

Certains auteurs ont cru voir une différence entre l'évolution ancienne des enzymes et leur évolution plus récente. D'une part, ils parlent de protéines généralistes pour les protéines anciennes car elles auraient été capables de nombreuses réactions. D'autre part ils considèrent que les enzymes se seraient spécialisées très vite et que les réactions que les protéines versatiles modernes peuvent catalyser en dehors de leur principale activité sont négligeables dans la physiologie normale de la cellule et ne rentrent en jeu que lorsque de nouvelles pressions apparaissent [Copley 2012]. Cette vision des protéines versatiles comme ayant des activités secondaires négligeables ne semble pas confirmée par mes observations qui montrent que de nombreuses protéines peuvent être prédites comme capables d'utiliser des substrats assez divers, malgré leur préférence pour un substrat ou un autre. Cependant, mes analyses n'avaient pas pour objectif de répondre à cette question et ne peuvent qu'encourager la recherche aux fins de déterminer si en effet la versatilité des enzymes est différente entre les enzymes anciennes et les actuelles.

Finalement, on peut signaler que le bricolage n'explique pas l'origine primaire des fonctions de la même manière que l'évolution darwinienne ne pose pas directement la question de l'origine de la vie [Peretó *et al.* 2009]. L'apparition de gènes à partir de séquences aléatoires a certainement existé à une époque [David et Alm 2011] et la possibilité que cela existe encore aujourd'hui est séduisante [Carvunis *et al.* 2012], mais les données disponibles issues de mon travail et de celui d'autres auteurs privilégient clairement le réarrangement de structures préalables comme moteur de l'évolution du métabolisme depuis le cenancêtre.

Quel type de membrane dans le cenancêtre ?

Tout semble indiquer que le cenancêtre avait une machinerie qui lui permettait de synthétiser des phospholipides et, logiquement, on peut supposer que ces phospholipides formaient des membranes dans cet organisme. Cependant, comme toutes les machineries semblent avoir été présentes dans le cenancêtre la composition phospholipidique de ses membranes est une question qui reste ouverte. On pourrait proposer que seul l'un des types de phospholipides possibles était présent chez le cenancêtre et que le reste avait d'autres fonctions mais on n'a aucun indice pour proposer un type de phospholipide plutôt qu'un autre. Une possibilité intéressante est que les membranes du cenancêtre aient été composées d'un mélange de phospholipides, et pas forcément de deux types seulement car on pourrait

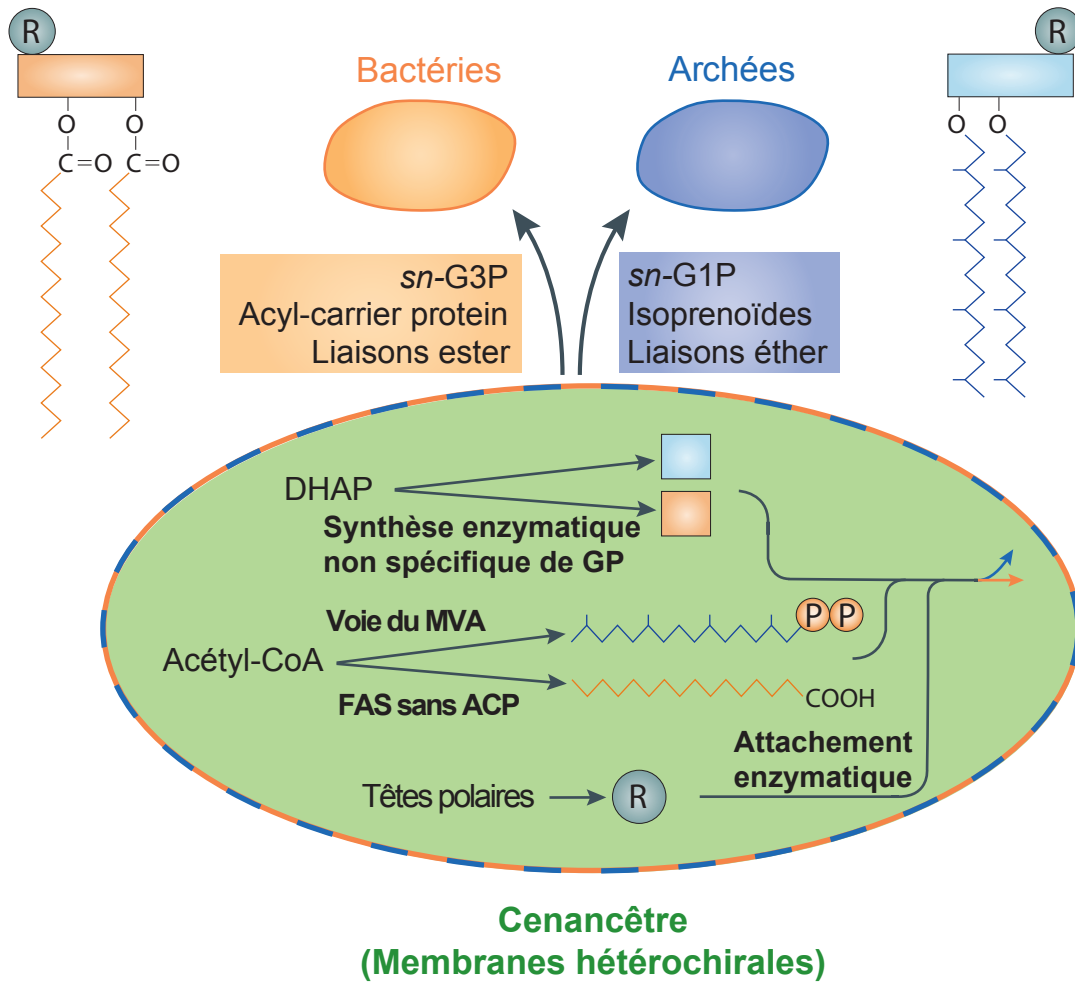


Figure 97. Hypothèse selon laquelle une membrane hétérochirale ait existé autour du cenacêtre sur la base de l'inférence de la présence d'une machinerie enzymatique chez cet organisme (d'après Lombard *et al.* 2012).

imaginer facilement quatre types de phospholipides compte tenu des composants dont on a inféré la présence dans le cenancêtre (Figure 97).

Le premier argument en faveur des membranes mixtes est la difficulté à imaginer un changement complet d'un type de phospholipide à l'autre. D'une part, il a été proposé que le remplacement d'un type de membrane par un autre impliquerait une étape hétérochirale intermédiaire et que cette étape serait plus instable, donc contre-sélectionnée [Wächtershäuser 2003]. D'autre part, j'ai résumé dans l'introduction le fait que l'hypothèse d'une plus grande instabilité des membranes hétérochirales n'est pas soutenue par l'ensemble des observations mais plutôt contredite. En revanche, même si l'étape intermédiaire n'est pas plus instable et dans le cas où il aurait existé un seul type de phospholipide ancestral, on peut se demander comment une machinerie alternative aurait pu être mise en place *de novo*, puis, quelle était la pression sélective qui l'a conservé. En effet, lorsqu'on regarde les organismes actuels, les deux types de membranes semblent s'être adaptés à tous les milieux que l'on connaisse, ce qui suggère que chaque type de membrane est suffisamment plastique pour s'adapter à toutes les situations. Ceci n'est pas surprenant car, comme l'illustre la métaphore du « bricolage », la sélection naturelle privilégie certaines variantes mais ces variantes consistent par définition en de petites modifications d'un système et non en un recrutement spontané d'un système complet non homologue.

En dehors de la difficulté structurelle à imaginer le remplacement total d'un système par l'autre, on a vu que plusieurs enzymes responsables de la synthèse des phospholipides semblent avoir été moins spécifiques que les enzymes actuelles et que leurs produits étaient très probablement utilisés dans plusieurs mécanismes cellulaires différents. Il semble par conséquent probable que les membranes du cenancêtre aient été un mélange hétéroclite de composés et que les membranes modernes comportant un type de phospholipide très majoritaire soient apparues lorsque des fonctions précises ont été assignées à chaque molécule. Finalement, le fait que les membranes du cenancêtre aient été composées d'un mélange de lipides, y compris de phospholipides, semble en continuité à la fois avec les membranes mélangées modernes et avec les membranes comportant de nombreuses molécules différentes que l'on peut imaginer entre les premières membranes et le cenancêtre.

Implications des membranes mixtes du cenancêtre pour les trois domaines

Si le cenancêtre avait des membranes mixtes, comment s'est produit l'évolution vers des systèmes très différenciés comme ceux des domaines du vivant ? La première réponse qui

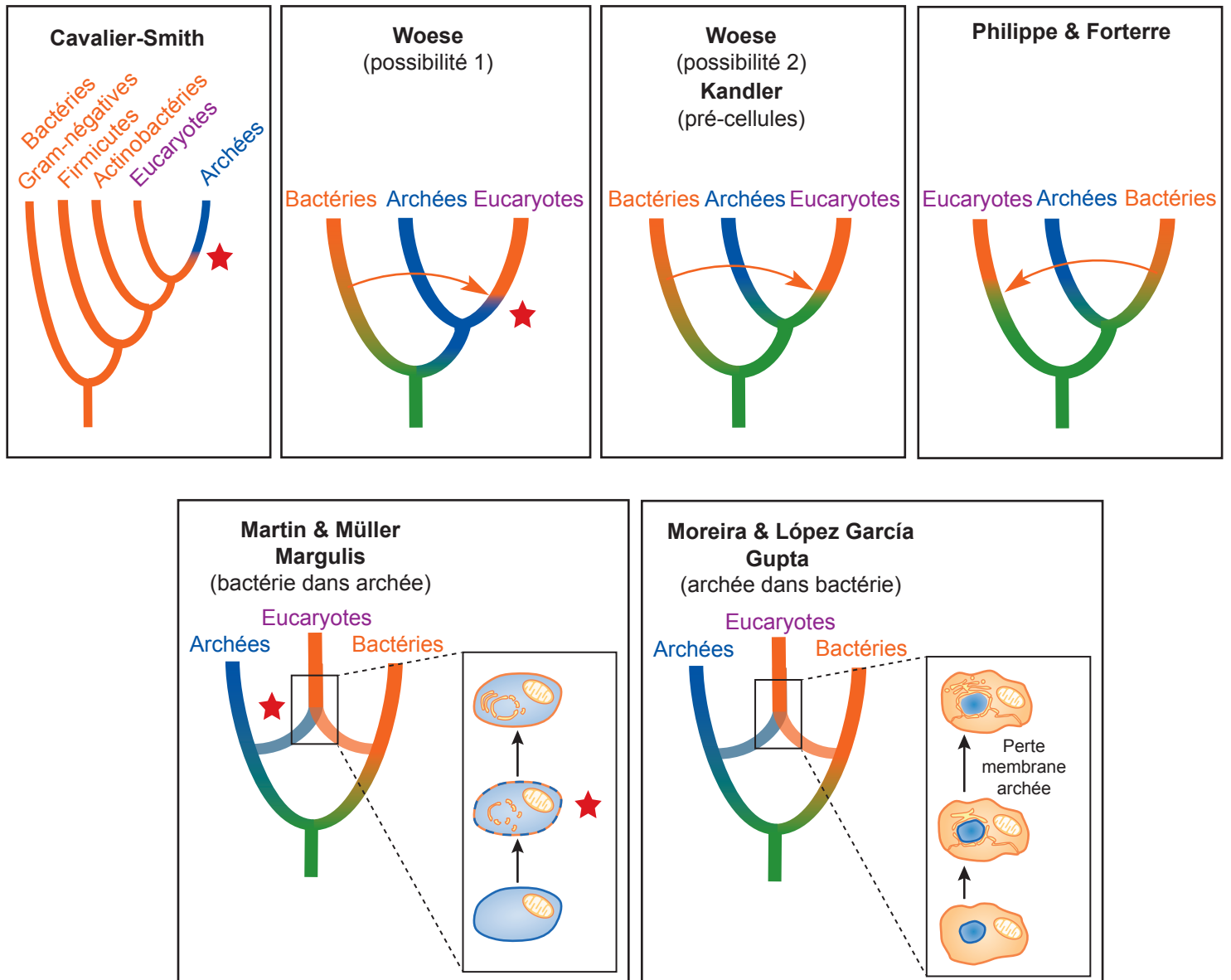


Figure 98. Impact de l'hypothèse selon laquelle le cenancêtre était capable de synthétiser enzymatiquement des membranes hétérochirales sur l'évolution des organismes.

Différentes topologies d'arbres ou hypothèses évolutives sont considérées. Les couleurs des arbres (mais non du texte) correspondent aux types des phospholipides présents au cours de l'évolution : les membranes composées de phospholipides de type bactérien sont en orange et celles avec des phospholipides de type archéen en bleu. Les membranes hétérochirales (anciennes) du cenancêtre apparaissent en vert, alors que les membranes hétérochirales tardives apparaissent comme un mélange de bleu et d'orange. Les étoiles rouges signalent des transitions d'un type de phospholipide vers l'autre. Les flèches indiquent des transferts des voies de synthèse des phospholipides bactériens vers les eucaryotes. L'hypothèse de Cavalier-Smith ne peut pas s'accorder avec un cenancêtre à membrane hétérochirale mais elle est présentée comme un exemple de modèle qui choisit un type de phospholipide comme étant ancestral (ce qui implique forcément une transition).

ait été donnée à cette question est l'instabilité supposée des membranes mixtes mais la réalité de cette instabilité n'est pas prouvée. D'ailleurs, cette hypothèse aurait davantage de sens si l'on invoquait une synthèse abiotique des phospholipides mais comme le cenancêtre semble avoir eu des mécanismes enzymatiques de synthèse, chaque cellule fille aurait continué à synthétiser les différents types de phospholipide et la ségrégation physico-chimique n'aurait pas suffi. On va brièvement explorer quelques événements qui peuvent expliquer les différences entre les domaines modernes (qui ont été exposées d'une façon assez similaire dans l'article Lombard *et al.* 2012c (voir annexes 4 et 6).

Les acides gras des eucaryotes sont synthétisés par des mécanismes différents qui ont été hérités plusieurs fois des bactéries. Par conséquent, la synthèse chez les eucaryotes telle qu'on la connaît aujourd'hui est obligatoirement postérieure à la divergence des bactéries. On va donc se concentrer d'abord sur les phospholipides des deux autres domaines, les archées et les bactéries. Le fait que l'ancêtre des archées ait probablement été hyperthermophile [Boussau *et al.* 2008] a conduit à proposer que les membranes actuelles de ce domaine soient le résultat d'une ancienne adaptation à de hautes températures. Dans le cas des bactéries, la grande efficacité de la machinerie ACP caractéristique de cette lignée pourrait expliquer l'abandon progressif des autres phospholipides. Dans les deux cas, les autres composants ancestraux auraient été conservés mais relégués à d'autres fonctions et une fois qu'un système efficace aurait été mis en place, il aurait été maintenu à cause de l'hypothèse de complexité. Une autre possibilité est que les deux phénomènes aient été en principe le produit du hasard mais qu'une fois chaque machinerie de synthèse mise en place, le changement de type de phospholipides aurait aussi été contre-sélectionné par l'hypothèse de complexité (*frozen accident*). Cette possibilité semble particulièrement tentante dans le cas de la sélectivité de l'énantiomère du glycérol phosphate car il n'y a aucun paramètre adaptatif évident de cette caractéristique.

L'interprétation est plus difficile dans le cas des eucaryotes car ils semblent avoir obtenu leur machinerie actuelle des bactéries mais, en réalité, les scénarios dépendent de la topologie de l'arbre du vivant que l'on utilise (Figure 98). En principe, dans une topologie comme celle de Woese (1990), où les bactéries sont groupe frère d'un regroupement des archées et des eucaryotes, on pourrait imaginer que l'ancêtre archéen-eucaryote avait des membranes de type soit archéen, soit mixte, puisque le type eucaryote actuel est un dérivé du bactérien. Le passage d'une membrane archéenne vers une membrane bactérienne dans la lignée eucaryote semble difficile dans la mesure où aucune autre transition de ce type n'a

jamais été observée. Le changement à partir d'un type mixte semble plus probable car il n'implique pas un changement à proprement parler mais plutôt la promotion d'un des types déjà existants. La lignée eucaryote aurait pu acquérir ses membranes de l'alpha-protéobactérie ancêtre des mitochondries qui comportait déjà la machinerie très efficace de l'ACP. Un scénario similaire devrait être invoqué dans le cas où une ligne proto-eucaryote aurait été la première à diverger du cenancêtre, les bactéries et les archées étant alors groupes frères.

Mais il existe aussi la possibilité que les eucaryotes soient le résultat d'une fusion entre une bactérie et une archée. Dans ce cas, il y a deux possibilités : soit la bactérie est entrée en endosymbiose à l'intérieur d'une archée (comme dans le cas de l'hypothèse de l'hydrogène, [Martin et Muller 1998]) soit c'est l'archée qui est entrée en endosymbiose à l'intérieur d'une bactérie (comme dans le cas de l'hypothèse de syntrophie, [Moreira et López-García 1998 ; López-García et Moreira 2006]). Dans le premier cas il faudrait que l'archée remplace sa membrane par une membrane de type bactérien, ce qui semble improbable comme on l'a dit, à cause de l'hypothèse de complexité et du fait que l'on ne connaisse aucun autre exemple de ce type de changement. En revanche, dans le deuxième cas la membrane externe aurait déjà été celle de la bactérie et les membranes internes en seraient le résultat par invagination, ce qui semble plus vraisemblable car cela n'exige que la perte des membranes archéennes d'origine (Figure 98).

Perspectives

Mon analyse et les résultats présentés ici précisent sensiblement la vision que l'on pouvait avoir jusqu'à présent du cenancêtre : cet organisme avait très vraisemblablement des versions anciennes des voies métaboliques qui permettent les synthèses actuelles des phospholipides. Ces voies métaboliques étaient probablement responsables de la synthèse enzymatique de molécules similaires aux molécules modernes et entraient dans la composition de ses membranes. Ces membranes pouvaient ainsi agir à la fois comme des frontières cellulaires et comme des compartiments très importants, d'une façon similaire à celle rencontrée dans les organismes modernes. Cependant, j'ai montré dans les sections précédentes que la présence de membranes chez le cenancêtre laisse la porte ouverte à de nombreuses questions, notamment sur la nature exacte de ces membranes et sur les mécanismes spécifiques qui ont conduit à la séparation des deux types de phospholipides qui sont aujourd'hui majoritaires dans les différents domaines.

Ce travail pose aussi la question du fonctionnement de cette membrane et des activités qui y avaient lieu. Jusque-là la question des interactions entre protéines membranaires et les différents types de phospholipides a été peu étudiée mais l'on sait qu'elle est importante. L'information contenue dans les protéines nous éclairera peut-être sur le type de phospholipide du cenancêtre et sur la question de la mixité de ses membranes. En outre, les protéines membranaires sont très importantes à la fois en nombre et en fonctions cellulaires fondamentales. Une étude globale les concernant serait donc d'un grand intérêt pour comprendre les fonctions liées aux membranes du cenancêtre.

Mon travail souligne en effet l'intérêt, dans le cas de travaux sur des événements anciens, d'étudier simultanément l'évolution d'un grand nombre d'éléments participant ensemble à une fonction déterminée. Ceci permet, d'une part, de réfléchir à des questions concernant des points sur lesquels on a assez peu d'information en recoupant les résultats et, d'autre part, d'aider à comprendre des fonctions dans leur ensemble. En ce qui concerne les membranes du cenancêtre, il serait intéressant d'étudier l'évolution ancienne des mécanismes de division cellulaire qui sont très différents selon les différents domaines du vivant [Cann 2008 ; Lindas *et al.* 2008]. Cette perspective est d'autant plus prometteuse que des études récentes ont révélé que non seulement les protéines membranaires mais aussi la composition lipidique des membranes sont indispensables dans la régulation de la réplication de l'ADN et la détermination de la division cellulaire [Crooke 2001 ; Mileykovskaya et Dowhan 2005]. On sait encore peu de choses sur le lien qui existe entre les membranes et la machinerie informationnelle des cellules mais la coordination entre les deux semble essentielle. Les découvertes dans ce domaine seront donc d'un grand intérêt pour comprendre aussi l'intégration de ces éléments dans les organismes anciens et même dans le cenancêtre. Enfin, on sait aussi que des vésicules capables de quitter la cellule existent chez des organismes des trois domaines du vivant [Gaudin 2012] et lorsqu'on connaîtra les mécanismes qui régissent ce processus dans les trois domaines, il sera intéressant de les comparer pour déterminer l'ancienneté de cette caractéristique.

D'une façon plus générale, mon étude montre aussi que, contrairement à une idée très répandue, l'étude évolutive d'événements anciens est possible non seulement pour la machinerie informationnelle mais aussi pour le métabolisme. Ceci semble d'autant plus prometteur pour l'étude futur d'autres voies métaboliques clés chez les organismes modernes mais pour lesquelles on dispose de peu d'informations relatives à l'évolution ancienne dans les trois domaines, comme par exemple les voies liées au flux de carbone, à la bioénergétique

et à la synthèse d'acides aminés, de nucléotides ou de co-facteurs. L'étude de l'histoire évolutive de ces protéines reste compliquée mais la prise en compte des voies métaboliques entières et reliées entre elles permet d'avoir une vision d'ensemble indispensable pour comprendre la biologie des organismes, qu'ils soient modernes ou anciens. En outre, l'analyse de l'histoire évolutive des acides gras illustre le fait que ce genre de travail peut permettre de proposer des voies métaboliques qui sont autant de nouvelles pistes pour la recherche en biochimie.

Le cenancêtre, un organisme ancien

L'ensemble des structures et caractéristiques dont la présence a été inférée chez le cenancêtre montre que cet organisme ressemblait beaucoup plus aux organismes modernes qu'à ce que l'on peut imaginer de l'origine de la vie. On a vu qu'il nous reste encore beaucoup à étudier sur les propriétés de cet organisme et j'espère que ces pistes seront explorées prochainement.

Mais en plus d'encourager la recherche sur l'évolution ancienne de composants majeurs des cellules modernes, je pense que mon travail montre aussi que les questions relatives au cenancêtre peuvent être posées autrement. Mon travail est parti du principe que le cenancêtre était un organisme que l'on peut décrire en soi, sans forcément chercher à le comparer à un autre organisme soi-disant primitif, qu'il soit moderne ou encore plus ancien que le cenancêtre. Le cenancêtre n'était ni une bactérie, ni une archée, ni un eucaryote, ni la première cellule, et par conséquent il ne peut être défini par rapport à aucun d'entre eux. Le cenancêtre est un organisme ancien, l'ancêtre des organismes modernes. Ses composants sont par définition plus anciens que ceux de leurs descendants modernes et son fonctionnement était donc forcément différent de celui des organismes actuels. Sa biologie doit être comprise comme celle d'un organisme hypothétique et non pas comme une copie d'un organisme que l'on a préalablement défini comme étant primitif. Cette remarque vaut en réalité pour tout organisme ancestral, qu'il soit le cenancêtre, l'ancêtre d'un domaine, d'un phyla ou de n'importe quel autre taxon.

Concernant la grande diversité qui existe entre les organismes modernes, ce travail rappelle, comme bien d'autres, que la puissance de l'évolution darwinienne est d'expliquer le moteur permettant aux organismes de générations successives d'être de plus en plus divergents et adaptées à des niches qui changent à leur tour, jusqu'à des points surprenants. Si les organismes actuels sont très divers, ce n'est probablement pas parce que le cenancêtre

manquait de tout, mais certainement parce que chaque lignée a utilisé différemment les systèmes qu'elle avait à sa disposition, comme les organismes actuels continuent de faire encore aujourd'hui.

BIBLIOGRAPHIE

- Abby, S. S., E. Tannier, M. Gouy et V. Daubin. 2012. Lateral gene transfer as a support for the tree of life. *Proceedings of the National Academy of Sciences of the United States of America* **109**:4962-7.
- Addlesee, H. A., L. C. Gibson, P. E. Jensen et C. N. Hunter. 1996. Cloning, sequencing and functional assignment of the chlorophyll biosynthesis gene, chlP, of *Synechocystis* sp. PCC 6803. *FEBS letters* **389**:126-30.
- Aharoni, A., L. Gaidukov, O. Khersonsky, Q. G. S. Mc, C. Roodveldt et D. S. Tawfik. 2005. The 'evolvability' of promiscuous protein functions. *Nat Genet* **37**:73-6.
- Airenne, T. T., J. M. Torkko, S. Van den plas, R. T. Sormunen, A. J. Kastaniotis, R. K. Wierenga et J. K. Hiltunen. 2003. Structure-function analysis of enoyl thioester reductase involved in mitochondrial maintenance. *Journal of molecular biology* **327**:47-59.
- Alberts, A. W., P. W. Majerus et P. R. Vagelos. 1965. Acyl carrier protein. VI. Purification and properties of the beta-keto acyl carrier protein synthetase. *Biochemistry* **4**:2265-74.
- Allen, W. V. et C. Ponnampereuma. 1967. A possible prebiotic synthesis of monocarboxylic acids. *Curr Mod Biol* **1**:24-8.
- Altabe, S. G., P. Aguilar, G. M. Caballero et D. de Mendoza. 2003. The *Bacillus subtilis* acyl lipid desaturase is a delta5 desaturase. *Journal of bacteriology* **185**:3228-31.
- Altschul, S. F., W. Gish, W. Miller, E. W. Myers et D. J. Lipman. 1990. Basic local alignment search tool. *Journal of molecular biology* **215**:403-10.
- Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller et D. J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**:3389-402.
- Anantharaman, V., L. Aravind et E. V. Koonin. 2003. Emergence of diverse biochemical activities in evolutionarily conserved structural scaffolds of proteins. *Current opinion in chemical biology* **7**:12-20.
- Andam, C. P. et J. P. Gogarten. 2011. Biased gene transfer in microbial evolution. *Nat Rev Microbiol* **9**:543-55.
- Andreassi, J. L., 2nd et T. S. Leyh. 2004. Molecular functions of conserved aspects of the GHMP kinase family. *Biochemistry* **43**:14594-601.
- Apel, C. L. et D. W. Deamer. 2005. The formation of glycerol monodecanoate by a dehydration condensation reaction: increasing the chemical complexity of amphiphiles on the early Earth. *Orig Life Evol Biosph* **35**:323-32.
- Apel, C. L., D. W. Deamer et M. N. Mautner. 2002. Self-assembled vesicles of monocarboxylic acids and alcohols: conditions for stability and for the encapsulation of biopolymers. *Biochim Biophys Acta* **1559**:1-9.
- Ariga, K., H. Yuki, J. Kikuchi, O. Dannemuller, A. M. Albrecht-Gary, Y. Nakatani et G. Ourisson. 2005. Monolayer studies of single-chain polyprenyl phosphates. *Langmuir* **21**:4578-83.
- Arrecubieta, C., R. López et E. García. 1994. Molecular characterization of cap3A, a gene from the operon required for the synthesis of the capsule of *Streptococcus pneumoniae* type 3: sequencing of mutations responsible for the unencapsulated phenotype and localization of the capsular cluster on the pneumococcal chromosome. *Journal of bacteriology* **176**:6375-83.

- Auer, J., G. Spicker et A. Bock. 1991. Presence of a gene in the archaeobacterium *Methanococcus vannielii* homologous to secY of eubacteria. *Biochimie* **73**:683-8.
- Autio, K. J., J. L. Guler, A. J. Kastaniotis, P. T. Englund et J. K. Hiltunen. 2008. The 3-hydroxyacyl-ACP dehydratase of mitochondrial fatty acid synthesis in *Trypanosoma brucei*. *FEBS letters* **582**:729-33.
- Ayte, J., G. Gil-Gomez, D. Haro, P. F. Marrero et F. G. Hegardt. 1990. Rat mitochondrial and cytosolic 3-hydroxy-3-methylglutaryl-CoA synthases are encoded by two different genes. *Proceedings of the National Academy of Sciences of the United States of America* **87**:3874-8.
- Bachelard. 1938. La formation de l'esprit scientifique. Vrin, Paris.
- Bachmann, P. A., P. L. Luisi et J. Lang. 1992. Autocatalytic self-replicating micelles as models for prebiotic structures. *Nature* **357**:57-59.
- Baldauf, S. L., J. D. Palmer et W. F. Doolittle. 1996. The root of the universal tree and the origin of eukaryotes based on elongation factor phylogeny. *Proceedings of the National Academy of Sciences of the United States of America* **93**:7749-54.
- Bangham, A. D., M. M. Standish et J. C. Watkins. 1965. Diffusion of univalent ions across the lamellae of swollen phospholipids. *J Mol Biol* **13**:238-52.
- Banner, D. W., A. C. Bloomer, G. A. Petsko *et al.* 1975. Structure of chicken muscle triose phosphate isomerase determined crystallographically at 2.5 angstrom resolution using amino acid sequence data. *Nature* **255**:609-14.
- Bar, R. S., D. W. Deamer et D. G. Cornwell. 1966. Surface area of human erythrocyte lipids: reinvestigation of experiments on plasma membrane. *Science (New York, N.Y)* **153**:1010-2.
- Barkley, S. J., R. M. Cornish et C. D. Poulter. 2004. Identification of an Archaeal type II isopentenyl diphosphate isomerase in methanothermobacter thermautotrophicus. *Journal of bacteriology* **186**:1811-7.
- Baumgart, T., S. T. Hess et W. W. Webb. 2003. Imaging coexisting fluid domains in biomembrane models coupling curvature and line tension. *Nature* **425**:821-4.
- Baymann, F., E. Lebrun, M. Brugna, B. Schoepp-Cothenet, M. T. Giudici-Orticoni et W. Nitschke. 2003. The redox protein construction kit: pre-last universal common ancestor evolution of energy-conserving enzymes. *Philosophical transactions of the Royal Society of London* **358**:267-74.
- Becerra, A., L. Delaye, S. Islas et A. Lazcano. 2007a. The very early stages and the nature of the last common ancestor of the three major cell domains. *Annu. Rev. Ecol. Evol. Syst.* **38**:361-79.
- Becerra, A., L. Delaye, A. Lazcano et L. E. Orgel. 2007b. Protein disulfide oxidoreductases and the evolution of thermophily: was the last common ancestor a heat-loving microbe? *J Mol Evol* **65**:296-303.
- Becerra, A., S. Islas, J. I. Leguina, E. Silva et A. Lazcano. 1997. Polyphyletic gene losses can bias backtrack characterizations of the cenancestor. *J Mol Evol* **45**:115-8.
- Bedke, D. K., G. M. Shibuya, A. Pereira, W. H. Gerwick, T. H. Haines et C. D. Vanderwal. 2009. Relative stereochemistry determination and synthesis of the major chlorosulfolipid from *Ochromonas danica*. *Journal of the American Chemical Society* **131**:7570-2.
- Bedke, D. K. et C. D. Vanderwal. 2011. Chlorosulfolipids: structure, synthesis, and biological relevance. *Natural product reports* **28**:15-25.
- Benson, A. A. 1966. On the orientation of lipids in chloroplast and cell membranes. *J Am Oil Chem Soc* **43**:265-70.

- Berclaz, N., M. Müller, P. Walde et P. L. Luisi. 2001. Growth and transformation of vesicles studied by efritin labelling and cryotransmission. *Journal of physical chemistry* **105**:1056-64.
- Berg, J. M., J. L. Tymoczko et L. Stryer. 2007a. Biochemistry. W. H. Freeman and Co., New York.
- Berg, J. M., J. L. Tymoczko et L. Stryer. 2007b. Biochemistry. W. H. Freeman and Company, New York.
- Bernstein, M., S. Sandford, L. Allamandola, S. Chang et M. A. Scharberg. 1995. Organic compounds produced by photolysis of realistic interstellar and cometary ice analogs containing methanol. *The Astrophysical Journal* **454**:327-44.
- Binari, R. C., B. E. Staveley, W. A. Johnson, R. Godavarti, R. Sasisekharan et A. S. Manoukian. 1997. Genetic evidence that heparin-like glycosaminoglycans are involved in wingless signaling. *Development (Cambridge, England)* **124**:2623-32.
- Bischoff, A. 2001. Meteorite classification and the definition of new chondrite classes as a result of successful meteorite search in hot and cold desert. *Planetary and Space Science* **49**:769-76.
- Bloch, K. 1964. The biological synthesis of cholesterol. *Nobel Lectures, Physiology or Medicine 1963-1970*.
- Bochar, D. A., C. V. Stauffacher et V. W. Rodwell. 1999. Sequence comparisons reveal two classes of 3-hydroxy-3-methylglutaryl coenzyme A reductase. *Molecular genetics and metabolism* **66**:122-7.
- Bork, P., J. Gellerich, H. Groth, R. Hooft et F. Martin. 1995. Divergent evolution of a beta/alpha-barrel subclass: detection of numerous phosphate-binding sites by motif search. *Protein Sci* **4**:268-74.
- Bork, P., C. Sander et A. Valencia. 1993. Convergent evolution of similar enzymatic function on different protein folds: the hexokinase, ribokinase, and galactokinase families of sugar kinases. *Protein Sci* **2**:31-40.
- Boucher, Y. 2007. Lipids: Biosynthesis, function, and evolution. Pp. 341-53 in R. Cavicchioli, ed. *Archaea, molecular and cellular biology*. ASM Press, Washington DC.
- Boucher, Y. et W. F. Doolittle. 2000. The role of lateral gene transfer in the evolution of isoprenoid biosynthesis pathways. *Molecular microbiology* **37**:703-16.
- Boucher, Y., C. J. Douady, R. T. Papke, D. A. Walsh, M. E. Boudreau, C. L. Nesbo, R. J. Case et W. F. Doolittle. 2003. Lateral gene transfer and the origins of prokaryotic groups. *Annual review of genetics* **37**:283-328.
- Boucher, Y., H. Huber, S. L'Haridon, K. O. Stetter et W. F. Doolittle. 2001. Bacterial origin for the isoprenoid biosynthesis enzyme HMG-CoA reductase of the archaeal orders Thermoplasmatales and Archaeoglobales. *Molecular biology and evolution* **18**:1378-88.
- Boucher, Y., M. Kamekura et W. F. Doolittle. 2004. Origins and evolution of isoprenoid lipid biosynthesis in archaea. *Molecular microbiology* **52**:515-27.
- Boussau, B., S. Blanquart, A. Necșulea, N. Lartillot et M. Gouy. 2008. Parallel adaptations to high temperatures in the Archaean eon. *Nature* **456**:942-5.
- Brady, R. O. 1958. The Enzymatic Synthesis of Fatty Acids by Aldol Condensation. *Proceedings of the National Academy of Sciences of the United States of America* **44**:993-8.
- Brady, R. O. et S. Gurin. 1950. The biosynthesis of radioactive fatty acids and cholesterol. *J. Biol. Chem.* **186**:461-69.
- Branton, D. 1966. Fracture faces of frozen membranes. *Proceedings of the National Academy of Sciences of the United States of America* **55**:1048-56.

- Branton, D. et R. B. Park. 1968. Introduction *in* D. Branton et R. B. Park, eds. Papers on biological membrane structure. Little, Brown and company, Boston.
- Breitling, R., D. Laubner, D. Clizbe, J. Adamski et S. K. Krisans. 2003. Isopentenyl-diphosphate isomerases in human and mouse: evolutionary analysis of a mammalian gene duplication. *Journal of molecular evolution* **57**:282-91.
- Bretscher, M. S. 1972. Asymmetrical lipid bilayer structure for biological membranes. *Nat New Biol* **236**:11-2.
- Brinkman, F. S., J. L. Blanchard, A. Cherkasov *et al.* 2002. Evidence that plant-like genes in Chlamydia species reflect an ancestral relationship between Chlamydiaceae, cyanobacteria, and the chloroplast. *Genome research* **12**:1159-67.
- Brinkmann, H. et H. Philippe. 1999. Archaea sister group of Bacteria? Indications from tree reconstruction artifacts in ancient phylogenies. *Molecular biology and evolution* **16**:817-25.
- Brocard, V. 1887, Les sceptiques grecs, La librairie française.
- Brochier-Armanet, C., P. Deschamps, P. López-García, Y. Zivanovic, F. Rodriguez-Valera et D. Moreira. 2011a. Complete-fosmid and fosmid-end sequences reveal frequent horizontal gene transfers in marine uncultured planktonic archaea. *The ISME journal* **5**:1291-302.
- Brochier-Armanet, C. et P. Forterre. 2006. Widespread distribution of archaeal reverse gyrase in thermophilic bacteria suggests a complex history of vertical inheritance and lateral gene transfers. *Archaea* **2**:83-93.
- Brochier-Armanet, C., P. Forterre et S. Gribaldo. 2011b. Phylogeny and evolution of the Archaea: one hundred genomes later. *Current opinion in microbiology* **14**:274-81.
- Brochier-Armanet, C., E. Talla et S. Gribaldo. 2009. The multiple evolutionary histories of dioxygen reductases: Implications for the origin and evolution of aerobic respiration. *Molecular biology and evolution* **26**:285-97.
- Brochier, C. et H. Philippe. 2002. Phylogeny: a non-hyperthermophilic ancestor for bacteria. *Nature* **417**:244.
- Brochier, C., H. Philippe et D. Moreira. 2000. The evolutionary history of ribosomal protein RpS14: horizontal gene transfer at the heart of the ribosome. *Trends Genet* **16**:529-33.
- Brooks, D. J., J. R. Fresco et M. Singh. 2004. A novel method for estimating ancestral amino acid composition and its application to proteins of the Last Universal Ancestor. *Bioinformatics* **20**:2251-7.
- Brown, A. K., S. Sridharan, L. Kremer, S. Lindenberg, L. G. Dover, J. C. Sacchettini et G. S. Besra. 2005. Probing the mechanism of the Mycobacterium tuberculosis beta-ketoacyl-acyl carrier protein synthase III mtFabH: factors influencing catalysis and substrate specificity. *The Journal of biological chemistry* **280**:32539-47.
- Brown, J. R. et W. F. Doolittle. 1995. Root of the universal tree of life based on ancient aminoacyl-tRNA synthetase gene duplications. *Proceedings of the National Academy of Sciences of the United States of America* **92**:2441-5.
- Brown, J. R. et W. F. Doolittle. 1997. Archaea and the prokaryote-to-eukaryote transition. *Microbiol Mol Biol Rev* **61**:456-502.
- Brown, J. R., C. J. Douady, M. J. Italia, W. E. Marshall et M. J. Stanhope. 2001. Universal trees based on large combined protein sequence data sets. *Nat Genet* **28**:281-5.
- Brown, J. R., F. T. Robb, R. Weiss et W. F. Doolittle. 1997. Evidence for the early divergence of tryptophanyl- and tyrosyl-tRNA synthetases. *J Mol Evol* **45**:9-16.

- Brundage, L., J. P. Hendrick, E. Schiebel, A. J. Driessen et W. Wickner. 1990. The purified *E. coli* integral membrane protein SecY/E is sufficient for reconstitution of SecA-dependent precursor protein translocation. *Cell* **62**:649-57.
- Budin, I. et J. W. Szostak. 2011. Physical effects underlying the transition from primitive to modern cell membranes. *Proceedings of the National Academy of Sciences of the United States of America* **108**:5249-54.
- Caetano-Anolles, G. 2002. Evolved RNA secondary structure and the rooting of the universal tree of life. *J Mol Evol* **54**:333-45.
- Campbell, J. W. et J. E. Cronan, Jr. 2001. Bacterial fatty acid biosynthesis: targets for antibacterial drug discovery. *Annual review of microbiology* **55**:305-32.
- Campbell, N. A. et J. B. Reece. 2004. Biologie. de Boeck, Bruxelles.
- Campbell, R. E., S. C. Mosimann, I. van De Rijn, M. E. Tanner et N. C. Strynadka. 2000. The first structure of UDP-glucose dehydrogenase reveals the catalytic residues necessary for the two-fold oxidation. *Biochemistry* **39**:7012-23.
- Campos-García, J., A. D. Caro, R. Najera, R. M. Miller-Maier, R. A. Al-Tahhan et G. Soberon-Chavez. 1998. The *Pseudomonas aeruginosa* rhlG gene encodes an NADPH-dependent beta-ketoacyl reductase which is specifically involved in rhamnolipid synthesis. *Journal of bacteriology* **180**:4442-51.
- Cann, I. K. 2008. Cell sorting protein homologs reveal an unusual diversity in archaeal cell division. *Proceedings of the National Academy of Sciences of the United States of America* **105**:18653-4.
- Cao, T. B. et M. H. Saier, Jr. 2003. The general protein secretory pathway: phylogenetic analyses leading to evolutionary conclusions. *Biochimica et biophysica acta* **1609**:115-25.
- Cape, J. L., P. A. Monnard et J. A. Boncella. 2011. Prebiotically relevant mixed fatty acid vesicles support anionic solute encapsulation and photochemically catalyzed trans-membrane charge transport. *Chemical science* **2**:661-71.
- Carballeira, N. M., M. Reyes, A. Sostre, H. Huang, M. F. Verhagen et M. W. Adams. 1997. Unusual fatty acid compositions of the hyperthermophilic archaeon *Pyrococcus furiosus* and the bacterium *Thermotoga maritima*. *Journal of bacteriology* **179**:2766-8.
- Carpenter, E. P., A. R. Hawkins, J. W. Frost et K. A. Brown. 1998. Structure of dehydroquinase reveals an active site capable of multistep catalysis. *Nature* **394**:299-302.
- Carvunis, A. R., T. Rolland, I. Wapinski *et al.* 2012. Proto-genes and de novo gene birth. *Nature* **487**:370-4.
- Castell, A., P. Johansson, T. Unge, T. A. Jones et K. Backbro. 2005. Rv0216, a conserved hypothetical protein from *Mycobacterium tuberculosis* that is essential for bacterial survival during infection, has a double hotdog fold. *Protein Sci* **14**:1850-62.
- Castoe, T. A., T. Stephens, B. P. Noonan et C. Calestani. 2007. A novel group of type I polyketide synthases (PKS) in animals and the complex phylogenomics of PKSs. *Gene* **392**:47-58.
- Castresana, J. et D. Moreira. 1999. Respiratory chains in the last common ancestor of living organisms. *Journal of molecular evolution* **49**:453-60.
- Cavalier-Smith, T. 2002. The neomuran origin of archaebacteria, the negibacterial root of the universal tree and bacterial megaclassification. *International journal of systematic and evolutionary microbiology* **52**:7-76.
- Cavalier-Smith, T. 2010. Deep phylogeny, ancestral groups and the four ages of life. *Philosophical transactions of the Royal Society of London* **365**:111-32.

- Chakrabarti, A. C., R. R. Breaker, G. F. Joyce et D. W. Deamer. 1994. Production of RNA by a polymerase protein encapsulated within phospholipid vesicles. *J Mol Evol* **39**:555-9.
- Chamberlin, T. C. et R. T. Chamberlin. 1908. Early Terrestrial Conditions That May Have Favored Organic Synthesis. *Science (New York, N.Y)* **28**:897-911.
- Chambers, R. 1922. A Micro Injection Study on the Permeability of the Starfish Egg. *J Gen Physiol* **5**:189-93.
- Chan, D. I. et H. J. Vogel. 2010. Current understanding of fatty acid biosynthesis and the acyl carrier protein. *The Biochemical journal* **430**:1-19.
- Chang, Y. Y. et E. P. Kennedy. 1967. Biosynthesis of phosphatidyl glycerophosphate in *Escherichia coli*. *Journal of lipid research* **8**:447-55.
- Charlebois, R. L., C. W. Sensen, W. F. Doolittle et J. R. Brown. 1997. Evolutionary analysis of the hisCGABdFDEHI gene cluster from the archaeon *Sulfolobus solfataricus* P2. *J Bacteriol* **179**:4429-32.
- Chen, A., P. A. Kroon et C. D. Poulter. 1994. Isoprenyl diphosphate synthases: protein sequence comparisons, a phylogenetic tree, and predictions of secondary structure. *Protein Sci* **3**:600-7.
- Chen, A., D. Zhang et C. D. Poulter. 1993. (S)-geranylgeranyl glyceryl phosphate synthase. Purification and characterization of the first pathway-specific enzyme in archaeobacterial membrane lipid biosynthesis. *The Journal of biological chemistry* **268**:21701-5.
- Chen, I. A. 2009. Cell division: breaking up is easy to do. *Curr Biol* **19**:R327-8.
- Chen, I. A., R. W. Roberts et J. W. Szostak. 2004. The emergence of competition between model protocells. *Science (New York, N.Y)* **305**:1474-6.
- Chen, I. A., K. Salehi-Ashtiani et J. W. Szostak. 2005. RNA catalysis in model protocell vesicles. *J Am Chem Soc* **127**:13213-9.
- Chen, I. A. et J. W. Szostak. 2004a. A kinetic study of the growth of fatty acid vesicles. *Biophys J* **87**:988-98.
- Chen, I. A. et J. W. Szostak. 2004b. Membrane growth can generate a transmembrane pH gradient in fatty acid vesicles. *Proceedings of the National Academy of Sciences of the United States of America* **101**:7965-70.
- Chen, I. A. et P. Walde. 2010. From self-assembled vesicles to protocells. *Cold Spring Harbor perspectives in biology* **2**:a002170.
- Chen, M. et C. D. Poulter. 2010. Characterization of thermophilic archaeal isopentenyl phosphate kinases. *Biochemistry* **49**:207-17.
- Christensen, C. E., B. B. Kragelund, P. von Wettstein-Knowles et A. Henriksen. 2007. Structure of the human beta-ketoacyl [ACP] synthase from the mitochondrial type II fatty acid synthase. *Protein Sci* **16**:261-72.
- Chuakrut, S., H. Arai, M. Ishii et Y. Igarashi. 2003. Characterization of a bifunctional archaeal acyl coenzyme A carboxylase. *Journal of bacteriology* **185**:938-47.
- Chyba, C. et C. Sagan. 1992. Endogenous production, exogenous delivery and impact-shock synthesis of organic molecules: an inventory for the origins of life. *Nature* **355**:125-32.
- Chyba, C. F., P. J. Thomas, L. Brookshaw et C. Sagan. 1990. Cometary delivery of organic molecules to the early Earth. *Science (New York, N.Y)* **249**:366-73.
- Ciccarelli, F. D., T. Doerks, C. von Mering, C. J. Creevey, B. Snel et P. Bork. 2006. Toward automatic reconstruction of a highly resolved tree of life. *Science* **311**:1283-7.

- Clarke, G. D., R. G. Beiko, M. A. Ragan et R. L. Charlebois. 2002. Inferring genome trees by using a filter to eliminate phylogenetically discordant sequences and a distance matrix based on mean normalized BLASTP scores. *J Bacteriol* **184**:2072-80.
- Clizbe, D. B., M. L. Owens, K. R. Masuda, J. E. Shackelford et S. K. Krisans. 2007. IDI2, a second isopentenyl diphosphate isomerase in mammals. *The Journal of biological chemistry* **282**:6668-76.
- Cohen, O., U. Gophna et T. Pupko. 2010. The complexity hypothesis revisited: connectivity rather than function constitutes a barrier to horizontal gene transfer. *Molecular biology and evolution* **28**:1481-9.
- Cole, K. S. 1932. Surface forces of the arbacia egg. *Journal of Cellular and Comparative Physiology* **1**:1-9.
- Colella, M., S. Lobasso, F. Babudri et A. Corcelli. 1998. Palmitic acid is associated with halorhodopsin as a free fatty acid. Radiolabeling of halorhodopsin with 3H-palmitic acid and chemical analysis of the reaction products of purified halorhodopsin with thiols and NaBH₄. *Biochimica et biophysica acta* **1370**:273-9.
- Coleman, J. 1990. Characterization of Escherichia coli cells deficient in 1-acyl-sn-glycerol-3-phosphate acyltransferase activity. *The Journal of biological chemistry* **265**:17215-21.
- Cook, H. et C. McMaster. 2002. Fatty acid desaturation and chain elongation in eukaryotes. Pp. 181-204 in D. A. Vance et J. F. Vance, eds. *Biochemistry of lipids, lipoproteins and membranes*. Elsevier Science, Amsterdam.
- Copeland, H. 1938. The kingdoms of organisms. *Quarterly Review of Biology* **13**:383-420.
- Copley, R. R. et P. Bork. 2000. Homology among (betaalpha)(8) barrels: implications for the evolution of metabolic pathways. *Journal of molecular biology* **303**:627-41.
- Copley, S. D. 2012. Toward a systems biology perspective on enzyme evolution. *J Biol Chem* **287**:3-10.
- Corcelli, A., S. Lobasso, M. Colella, M. Trotta, A. Guerrieri et F. Palmisano. 1996. Role of palmitic acid on the isolation and properties of halorhodopsin. *Biochimica et biophysica acta* **1281**:173-81.
- Cottin, H., M. C. Gazeau et F. Raulin. 1999. Cometary organic chemistry: a review from observations, numerical and experimental simulations. *Planetary and Space Science* **47**:1141-62.
- Cox, C. J., P. G. Foster, R. P. Hirt, S. R. Harris et T. M. Embley. 2008. The archaeobacterial origin of eukaryotes. *Proceedings of the National Academy of Sciences of the United States of America* **105**:20356-61.
- Cronan, J. E. et J. Thomas. 2009. Bacterial fatty acid synthesis and its relationships with polyketide synthetic pathways. *Methods in enzymology* **459**:395-433.
- Crooke, E. 2001. Escherichia coli DnaA protein--phospholipid interactions: in vitro and in vivo. *Biochimie* **83**:19-23.
- Dagan, T. et W. Martin. 2006. The tree of one percent. *Genome biology* **7**:118.
- Dagan, T., M. Roettger, D. Bryant et W. Martin. 2010. Genome networks root the tree of life between prokaryotic domains. *Genome Biol Evol* **2**:379-92.
- Daiyasu, H., T. Hiroike, Y. Koga et H. Toh. 2002. Analysis of membrane stereochemistry with homology modeling of sn-glycerol-1-phosphate dehydrogenase. *Protein engineering* **15**:987-95.

- Daiyasu, H., K. Kuma, T. Yokoi, H. Morii, Y. Koga et H. Toh. 2005. A study of archaeal enzymes involved in polar lipid synthesis linking amino acid sequence information, genomic contexts and lipid composition. *Archaea (Vancouver, B.C)* **1**:399-410.
- Daleke, D. L. et J. V. Lyles. 2000. Identification and purification of aminophospholipid flippases. *Biochimica et biophysica acta* **1486**:108-27.
- Dalessandro, G. et D. H. Northcote. 1977. Changes in enzymic activities of nucleoside diphosphate sugar interconversions during differentiation of cambium to xylem in pine and fir. *The Biochemical journal* **162**:281-8.
- Danielli, J. F. 1937. The relations between surface pH, ion concentration and interfacial tension. *Proceedings of the Royal Society (London) B* **122**:155-74.
- Danielli, J. F. et H. Davson. 1935. A contribution to the theory of permeability of thin films. *Journal of Cellular and Comparative Physiology* **5**:495-508.
- Danielli, J. F. et E. N. Harvey. 1935. The tension at the surface of mackerel egg oil, with remarks on the nature of the cell surface. *Journal of Cellular and Comparative Physiology* **5**:483-94.
- Dannenmuller, O., K. Arakawa, T. Eguchi, K. Kakinuma, S. Blanc, A. M. Albrecht, M. Schmutz, Y. Nakatani et G. Ourisson. 2000. Membrane properties of archaeal macrocyclic diether phospholipids. *Chemistry (Weinheim an der Bergstrasse, Germany)* **6**:645-54.
- Darnell, J. E., Jr. 1978. Implications of RNA-RNA splicing in evolution of eukaryotic cells. *Science* **202**:1257-60.
- Darwin, C. 1859. L'origine des espèces au moyen de la sélection naturelle ou la préservation des races favorisées dans la lutte pour la vie. GF Flammarion.
- Das, A. K., M. D. Uhler et A. K. Hajra. 2000. Molecular cloning and expression of mammalian peroxisomal trans-2-enoyl-coenzyme A reductase cDNAs. *The Journal of biological chemistry* **275**:24333-40.
- Daubin, V., M. Gouy et G. Perriere. 2002. A phylogenomic approach to bacterial phylogeny: evidence of a core of genes sharing a common history. *Genome Res* **12**:1080-90.
- Daubin, V., N. A. Moran et H. Ochman. 2003. Phylogenetics and the cohesion of bacterial genomes. *Science* **301**:829-32.
- David, L. A. et E. J. Alm. 2011. Rapid evolutionary innovation during an Archaeal genetic expansion. *Nature* **469**:93-6.
- Davson, H. 1962. Growth of the concept of the paucimolecular membrane. *Circulation* **26**:1022-37.
- de Duve, C. 2005. The onset of selection. *Nature* **433**:581-82.
- de Duve, C. et S. L. Miller. 1991. Two-dimensional life? *Proceedings of the National Academy of Sciences* **88**:10014-17.
- De Rosa, M., A. Gambacorta et A. Gliozzi. 1986. Structure, biosynthesis, and physicochemical properties of archaebacterial lipids. *Microbiological reviews* **50**:70-80.
- Deamer, D. 1985. Boundary structures are formed by organic components of the Murchison carbonaceous chondrite. *Nature* **317**:792-94.
- Deamer, D. W. 1986. Role of amphiphilic compounds in the evolution of membrane structure on the early earth. *Orig Life Evol Biosph* **17**:3-25.
- Deamer, D. W. 1997. The first living systems: a bioenergetic perspective. *Microbiol Mol Biol Rev* **61**:239-61.

- Deamer, D. W. et G. L. Barchfeld. 1982. Encapsulation of macromolecules by lipid vesicles under simulated prebiotic conditions. *J Mol Evol* **18**:203-6.
- Deamer, D. W. et D. Branton. 1967. Fracture planes in an ice-bilayer model membrane system. *Science* **158**:655-7.
- Deamer, D. W. et R. M. Pashley. 1989. Amphiphilic components of the Murchison carbonaceous chondrite: surface properties and membrane formation. *Orig Life Evol Biosph* **19**:21-38.
- Delage, L., A. Becerra et A. Lazcano. 2005. The last common ancestor: what's in a name? *Orig Life Evol Biosph* **35**:537-54.
- Dellas, N. et J. P. Noel. 2010. Mutation of archaeal isopentenyl phosphate kinase highlights mechanism and guides phosphorylation of additional isoprenoid monophosphates. *ACS chemical biology* **5**:589-601.
- Delsuc, F., H. Brinkmann et H. Philippe. 2005. Phylogenomics and the reconstruction of the tree of life. *Nat Rev Genet* **6**:361-75.
- Dervichian, D. et M. Macheboeuf. 1938. Sur l'existence d'une couche monomoléculaire de substances lipoidiques à la surface des globules rouges du sang. *Comptes rendus hebdomadaires des séances de l'Académie des sciences*:1511-14.
- Devaux, H. 1913. Oil films on water and on mercury in C. D. Walcott, ed. Annual report of the board of regents of the Smithsonian institution, Washington.
- Di Giulio, M. 2007a. The tree of life might be rooted in the branch leading to Nanoarchaeota. *Gene* **401**:108-13.
- Di Giulio, M. 2000a. The late stage of genetic code structuring took place at a high temperature. *Gene* **261**:189-95.
- Di Giulio, M. 2001. The universal ancestor was a thermophile or a hyperthermophile. *Gene* **281**:11-7.
- Di Giulio, M. 2000b. The universal ancestor lived in a thermophilic or hyperthermophilic environment. *J Theor Biol* **203**:203-13.
- Di Giulio, M. 2007b. The evidence that the tree of life is not rooted within the Archaea is unreliable: a reply to Skophammer et al. *Gene* **394**:105-6.
- Dibrova, D., K. Makarova, M. Galperin, E. V. Koonin et A. Mulkidjanian. 2011. Comparative analysis of lipid biosynthesis in Archaea and Bacteria: What was the structure of first membrane lipids? Moscow Conference on Computational Molecular Biology, Moscow.
- Dillon, S. C. et A. Bateman. 2004. The Hotdog fold: wrapping up a superfamily of thioesterases and dehydratases. *BMC bioinformatics* **5**:109.
- DiRuggiero, J., J. R. Brown, A. P. Bogert et F. T. Robb. 1999. DNA repair systems in archaea: mementos from the last universal common ancestor? *J Mol Evol* **49**:474-84.
- Doolittle, W. F. 1999. Phylogenetic classification and the universal tree. *Science* **284**:2124-9.
- Doolittle, W. F. 1978. Genes in pieces: were they ever together? *Nature* **272**:581.
- Doolittle, W. F. 1998. You are what you eat: a gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes. *Trends Genet* **14**:307-11.
- Doolittle, W. F. et E. Bapteste. 2007. Pattern pluralism and the Tree of Life hypothesis. *Proceedings of the National Academy of Sciences of the United States of America* **104**:2043-9.
- Doud, E. H., D. L. Perlstein, M. Wolpert, D. E. Cane et S. Walker. 2011. Two distinct mechanisms for TIM barrel prenyltransferases in bacteria. *Journal of the American Chemical Society* **133**:1270-3.

- Dowhan, W. 1997. CDP-diacylglycerol synthase of microorganisms. *Biochimica et biophysica acta* **1348**:157-65.
- Dowhan, W. 1992. Phosphatidylserine synthase from *Escherichia coli*. *Methods in enzymology* **209**:287-98.
- Dubey, P. G. et S. Ben-Yehuda. 2011. Intercellular nanotubes mediate bacterial communication. *Cell* **144**:590-600
- Duchesneau, F. 1987. Genèse de la théorie cellulaire. Bellarmin/Vrin, Montréal/Paris.
- Dutoit, R., J. de Ruyck, V. Durisotti, C. Legrain, E. Jacobs et J. Wouters. 2008. Overexpression, physicochemical characterization, and modeling of a hyperthermophilic pyrococcus furiosus type 2 IPP isomerase. *Proteins* **71**:1699-707.
- Dutrochet, M. H. 1824. Recherches anatomiques et physiologiques sur la structure intime des animaux et des végétaux. B. Baillière, Paris.
- Dutrochet, M. H. 1828. Nouvelles recherches sur l'endosmose et l'exosmose. J-B. Baillière, Paris.
- Dworkin, J., D. Deamer, S. Sandford et L. Allamandola. 2001. Self-assembling amphiphilic molecules: Synthesis in simulated interstellar/precometary ices. *Proceedings of the National Academy of Sciences of the United States of America* **98**:815-9.
- Edgar, R. C. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC bioinformatics* **5**:113.
- Edidin, M. 2003. Lipids on the frontier: a century of cell-membrane bilayers. *Nat Rev Mol Cell Biol* **4**:414-8.
- Ehrenfreund, P. et J. Cami. 2010. Cosmic carbon chemistry: from the interstellar medium to the early Earth. *Cold Spring Harbor perspectives in biology* **2**:a002097.
- Eichberg, J., E. Sherwood, D. E. Epps et J. Oro. 1977. Cyanamide mediated syntheses under plausible primitive earth conditions. IV. The synthesis of acylglycerols. *J Mol Evol* **10**:221-30.
- Eichler, J. 2002. Archaeal signal peptidases from the genus *Thermoplasma*: structural and mechanistic hybrids of the bacterial and eukaryal enzymes. *Journal of molecular evolution* **54**:411-5.
- Eichler, J. et M. W. Adams. 2005. Posttranslational protein modification in Archaea. *Microbiol Mol Biol Rev* **69**:393-425.
- Eldredge, N. et S. J. Gould. 1972. Models in Paleobiology. Freeman, Cooper and Company, San Francisco.
- Elferink, M. G., J. G. de Wit, R. Demel, A. J. Driessen et W. N. Konings. 1992. Functional reconstitution of membrane proteins in monolayer liposomes from bipolar lipids of *Sulfolobus acidocaldarius*. *The Journal of biological chemistry* **267**:1375-81.
- Elias, M. et D. S. Tawfik. 2012. Divergence and convergence in enzyme evolution: parallel evolution of paraoxonases from quorum-quenching lactonases. *J Biol Chem* **287**:11-20.
- Elovson, J. et P. R. Vagelos. 1969. A new class of lipids: chlorosulfolipids. *Proceedings of the National Academy of Sciences of the United States of America* **62**:957-63.
- Eme, L., D. Moreira, E. Talla et C. Brochier-Armanet. 2009. A complex cell division machinery was present in the last common ancestor of eukaryotes. *PloS one* **4**:e5021.
- Eme, L., A. Trilles, D. Moreira et C. Brochier-Armanet. 2011. The phylogenomic analysis of the anaphase promoting complex and its targets points to complex and modern-like control of the cell cycle in the last common ancestor of eukaryotes. *BMC evolutionary biology* **11**:265.

- Endo, A. 1992. The discovery and development of HMG-CoA reductase inhibitors. *Journal of lipid research* **33**:1569-82.
- Epps, D. E., D. W. Nooner, J. Eichberg, E. Sherwood et J. Oro. 1979. Cyanamide mediated synthesis under plausible primitive earth conditions. VI. The synthesis of glycerol and glycerophosphates. *J Mol Evol* **14**:235-41.
- Epps, D. E., E. Sherwood, J. Eichberg et J. Oro. 1978. Cyanamide mediated syntheses under plausible primitive earth conditions. V. The synthesis of phosphatidic acids. *J Mol Evol* **11**:279-92.
- Epstein, S., R. V. Krishnamurthy, J. R. Cronin, S. Pizzarello et G. U. Yuen. 1987. Unusual stable isotope ratios in amino acid and carboxylic acid extracts from the Murchison meteorite. *Nature* **326**:477-9.
- Esser, C., N. Ahmadinejad, C. Wiegand *et al.* 2004. A genome phylogeny for mitochondria among alpha-proteobacteria and a predominantly eubacterial ancestry of yeast nuclear genes. *Molecular biology and evolution* **21**:1643-60.
- Fan, Q., A. Relini, D. Cassinadri, A. Gambacorta et A. Gliozzi. 1995. Stability against temperature and external agents of vesicles composed of archaeal bolaform lipids and egg PC. *Biochimica et biophysica acta* **1240**:83-8.
- Felsenstein, J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. *Systematic Zoology* **27**:401-10.
- Felsenstein, J. 2003. Inferring phylogenies. Sinauer associates.
- Fernandez-Moran, H. 1962. Cell-membrane ultrastructure. Low-temperature electron microscopy and x-ray diffraction studies of lipoprotein components in lamellar systems. *Circulation* **26**:1039-65.
- Fernandez-Moran, H., T. Oda, P. V. Blair et D. E. Green. 1964. A macromolecular repeating unit of mitochondrial structure and function. *Journal of cellular biology* **22**:63-100.
- Fischbach, M. A., C. T. Walsh et J. Clardy. 2008. The evolution of gene collectives: How natural selection drives chemical innovation. *Proceedings of the National Academy of Sciences of the United States of America* **105**:4601-8.
- Fischer, F. et H. Tropsch. 1926. The synthesis of petroleum at atmospheric pressures from the gasification products of coal. *Brenst-Chem* **7**:97.
- Fisher, K. A. 1976. Analysis of membrane halves: cholesterol. *Proceedings of the National Academy of Sciences of the United States of America* **73**:173-7.
- Fitch, W. M. 1971. The nonidentity of invariable positions in the cytochromes c of different species. *Biochem Genet* **5**:231-41.
- Fitch, W. M. et E. Margoliash. 1967a. Construction of phylogenetic trees. *Science* **155**:279-84.
- Fitch, W. M. et E. Margoliash. 1967b. A method for estimating the number of invariant amino acid coding positions in a gene using cytochrome c as a model case. *Biochem Genet* **1**:65-71.
- Fitch, W. M. et K. Upper. 1987. The phylogeny of tRNA sequences provides evidence for ambiguity reduction in the origin of the genetic code. *Cold Spring Harb Symp Quant Biol* **52**:759-67.
- Fitz-Gibbon, S. T. et C. H. House. 1999. Whole genome-based phylogenetic analysis of free-living microorganisms. *Nucleic Acids Res* **27**:4218-22.
- Fleischaker, G. R. 1990. Origins of life: An operational definition. *Origins of life and evolution of the biosphere* **20**:127-37.
- Forterre, P. 1996. A hot topic: the origin of hyperthermophiles. *Cell* **85**:789-92.

- Forterre, P. 1995. Thermoreduction, a hypothesis for the origin of prokaryotes. *C R Acad Sci III* **318**:415-22.
- Forterre, P. 2002. A hot story from comparative genomics: reverse gyrase is the only hyperthermophile-specific protein. *Trends Genet* **18**:236-7.
- Forterre, P. 1999. Displacement of cellular proteins by functional analogues from plasmids or viruses could explain puzzling phylogenies of many DNA informational proteins. *Molecular microbiology* **33**:457-65.
- Forterre, P. 2001. Genomics and early cellular evolution. The origin of the DNA world. *C R Acad Sci III* **324**:1067-76.
- Forterre, P., C. Bouthier De La Tour, H. Philippe et M. Duguet. 2000. Reverse gyrase from hyperthermophiles: probable transfer of a thermoadaptation trait from archaea to bacteria. *Trends Genet* **16**:152-4.
- Forterre, P., S. Gribaldo et C. Brochier. 2005. Luca : à la recherche du plus proche ancêtre commun universel. *Médecine/Sciences* **21**:860-65.
- Forterre, P. et H. Philippe. 1999. Where is the root of the universal tree of life? *Bioessays* **21**:871-9.
- Fournier, G. P. et J. P. Gogarten. 2007. Signature of a primitive genetic code in ancient protein lineages. *J Mol Evol* **65**:425-36.
- Fournier, G. P. et J. P. Gogarten. 2010. Rooting the ribosomal tree of life. *Molecular biology and evolution* **27**:1792-801.
- Fox, G. E., E. Stackebrandt, R. B. Hespell *et al.* 1980. The phylogeny of prokaryotes. *Science* **209**:457-63.
- Fox, S. W., K. Harada et J. Kendrick. 1959. Production of Spherules from Synthetic Proteinoid and Hot Water. *Science (New York, N.Y)* **129**:1221-3.
- Fricke, H. 1925. The Electric Capacity of Suspensions with Special Reference to Blood. *J Gen Physiol* **9**:137-52.
- Fricke, H. et H. J. Curtis. 1934. Electric Impedance of Suspensions of Yeast Cells. *Nature* **134**:102-03.
- Friesen, J. A. et V. W. Rodwell. 2004. The 3-hydroxy-3-methylglutaryl coenzyme-A (HMG-CoA) reductases. *Genome biology* **5**:248.
- Frommolt, R., S. Werner, H. Paulsen, R. Goss, C. Wilhelm, S. Zauner, U. G. Maier, A. R. Grossman, D. Bhattacharya et M. Lohr. 2008. Ancient recruitment by chromists of green algal genes encoding enzymes for carotenoid biosynthesis. *Molecular biology and evolution* **25**:2653-67.
- Frye, L. D. et M. Edidin. 1970. The rapid intermixing of cell surface antigens after formation of mouse-human heterokaryons. *J Cell Sci* **7**:319-35.
- Fukuba, T., M. Ogawa, T. Fujii et T. Naganuma. 2003. Phylogenetic diversity of dissimilatory sulfite reductase genes from deep-sea cold seep sediment. *Marine biotechnology (New York, N.Y)* **5**:458-68.
- Furuuchi, R., E. Imai, H. Honda, K. Hatori et K. Matsuno. 2005. Evolving lipid vesicles in prebiotic hydrothermal environments. *Orig Life Evol Biosph* **35**:333-43.
- Gabrielsen, M., C. S. Bond, I. Hallyburton, S. Hecht, A. Bacher, W. Eisenreich, F. Rohdich et W. N. Hunter. 2004. Hexameric assembly of the bifunctional methylerythritol 2,4-cyclodiphosphate synthase and protein-protein associations in the deoxy-xylulose-dependent pathway of isoprenoid precursor biosynthesis. *The Journal of biological chemistry* **279**:52753-61.
- Galperin, M. Y. et E. V. Koonin. 2012. Divergence and convergence in enzyme evolution. *J Biol Chem* **287**:21-8.

- Galtier, N. 2001. Maximum-likelihood phylogenetic analysis under a covarion-like model. *Molecular biology and evolution* **18**:866-73.
- Galtier, N., N. Tourasse et M. Gouy. 1999. A nonhyperthermophilic common ancestor to extant life forms. *Science* **283**:220-1.
- Gargaud, M., P. Claeys, P. López-García *et al.* 2006. From suns to life. A chronological approach to the history of life on earth. Springer.
- Gattinger, A., M. Schloter et J. C. Munch. 2002. Phospholipid etherlipid and phospholipid fatty acid fingerprints in selected euryarchaeotal monocultures for taxonomic profiling. *FEMS microbiology letters* **213**:133-9.
- Gaucher, E. A., S. Govindarajan et O. K. Ganesh. 2008. Palaeotemperature trend for Precambrian life inferred from resurrected proteins. *Nature* **451**:704-7.
- Gaudin, M. 2012. Etude des vésicules membranaires produites par les Archées hyperthermophiles marines de l'ordre des Thermococcales. Pp. 101. Université Paris Sud, Orsay.
- Ge, F., L. S. Wang et J. Kim. 2005. The cobweb of life revealed by genome-scale estimates of horizontal gene transfer. *PLoS Biol* **3**:e316.
- Gerakines, P. A., M. H. Moore et R. L. Hudson. 2001. Energetic processing of laboratory ice analogs: UV photolysis versus ion bombardment. *Journal of Geophysical research* **106**:33.381-33.85.
- Gibson, D. M., E. B. Titchener et S. J. Wakil. 1958. Requirement for bicarbonate in fatty acid synthesis. *Journal of the American Chemical Society* **80**:2908.
- Gilbert, W. 1986. The RNA World. *Nature* **319**:618.
- Glansdorff, N. 2000. About the last common ancestor, the universal life-tree and lateral gene transfer: a reappraisal. *Molecular microbiology* **38**:177-85.
- Glansdorff, N., Y. Xu et B. Labedan. 2008. The last universal common ancestor: emergence, constitution and genetic legacy of an elusive forerunner. *Biol Direct* **3**:29.
- Gogarten-Boekels, M., E. Hilario et J. P. Gogarten. 1995. The effects of heavy meteorite bombardment on the early evolution--the emergence of the three domains of life. *Orig Life Evol Biosph* **25**:251-64.
- Gogarten, J. P., W. F. Doolittle et J. G. Lawrence. 2002. Prokaryotic evolution in light of gene transfer. *Molecular biology and evolution* **19**:2226-38.
- Gogarten, J. P., H. Kibak, P. Dittrich *et al.* 1989. Evolution of the vacuolar H⁺-ATPase: implications for the origin of eukaryotes. *Proceedings of the National Academy of Sciences of the United States of America* **86**:6661-5.
- Gokhale, R. S., P. Saxena, T. Chopra et D. Mohanty. 2007. Versatile polyketide enzymatic machinery for the biosynthesis of complex mycobacterial lipids. *Natural product reports* **24**:267-77.
- Goldacre, R. J. 1958. Surface films, their collapse on compression, the shapes and sizes of cells, and the origin of life. in J. F. Danielli, K. G. A. Pankhurst et A. C. Riddiford, eds. *Surface Phenomena in Biology and Chemistry*. Pergamon press, New York.
- Gomes, R., H. F. Levison, K. Tsiganis et A. Morbidelli. 2005. Origin of the cataclysmic Late Heavy Bombardment period of the terrestrial planets. *Nature* **435**:466-9.
- Gophna, U., J. R. Thompson, Y. Boucher et W. F. Doolittle. 2006. Complex histories of genes encoding 3-hydroxy-3-methylglutaryl-Coenzyme A reductase. *Molecular biology and evolution* **23**:168-78.
- Gordesky, S. E. et G. V. Marinetti. 1973. The asymmetric arrangement of phospholipids in the human erythrocyte membrane. *Biochem Biophys Res Commun* **50**:1027-31.

- Gorter, E. et F. Grendel. 1925. On bimolecular layers of lipoids on the chromocytes of the blood. *The Journal of Experimental Medicine* **47**:439-43.
- Gotoh, M., A. Miki, H. Nagano, N. Ribeiro, M. Elhabiri, E. Gumienna-Kontecka, A. M. Albrecht-Gary, M. Schmutz, G. Ourisson et Y. Nakatani. 2006. Membrane properties of branched polyprenyl phosphates, postulated as primitive membrane constituents. *Chem Biodivers* **3**:434-55.
- Gould, S. J. 1989. La vie est belle. Les surprises de l'évolution. Editions du Seuil.
- Gray, M. W. et W. F. Doolittle. 1982. Has the endosymbiont hypothesis been proven? *Microbiol Rev* **46**:1-42.
- Green, P. R., A. H. Merrill, Jr. et R. M. Bell. 1981. Membrane phospholipid synthesis in *Escherichia coli*. Purification, reconstitution, and characterization of sn-glycerol-3-phosphate acyltransferase. *The Journal of biological chemistry* **256**:11151-9.
- Gregory, T. R. 2001. Coincidence, coevolution, or causation? DNA content, cell size, and the C-value enigma. *Biological reviews of the Cambridge Philosophical Society* **76**:65-101.
- Gribaldo, S. et C. Brochier-Armanet. 2006. The origin and evolution of Archaea: a state of the art. *Philosophical transactions of the Royal Society of London* **361**:1007-22.
- Gribaldo, S. et C. Brochier. 2009. Phylogeny of prokaryotes: does it exist and why should we care? *Research in microbiology* **160**:513-21.
- Gribaldo, S. et P. Cammarano. 1998. The root of the universal tree of life inferred from anciently duplicated genes encoding components of the protein-targeting machinery. *Journal of molecular evolution* **47**:508-16.
- Gribaldo, S. et H. Philippe. 2002. Ancient phylogenetic relationships. *Theor Popul Biol* **61**:391-408.
- Gribaldo, S., A. M. Poole, V. Daubin, P. Forterre et C. Brochier-Armanet. 2010. The origin of eukaryotes and their relationship with the Archaea: are we at a phylogenomic impasse? *Nat Rev Microbiol* **8**:743-52.
- Grochowski, L. L., H. Xu et R. H. White. 2006. Methanocaldococcus jannaschii uses a modified mevalonate pathway for biosynthesis of isopentenyl diphosphate. *Journal of bacteriology* **188**:3192-8.
- Guldan, H., F. M. Matysik, M. Bocola, R. Sterner et P. Babinger. 2011. Functional Assignment of an Enzyme that Catalyzes the Synthesis of an Archaea-Type Ether Lipid in Bacteria. *Angewandte Chemie (International ed)* **50**:8188-91.
- Guldan, H., R. Sterner et P. Babinger. 2008. Identification and characterization of a bacterial glycerol-1-phosphate dehydrogenase: Ni(2+)-dependent AraM from *Bacillus subtilis*. *Biochemistry* **47**:7376-84.
- Gupta, R., J. M. Lanter et C. R. Woese. 1983. Sequence of the 16S Ribosomal RNA from *Halobacterium volcanii*, an Archaeobacterium. *Science* **221**:656-9.
- Gupta, R. S. 1998a. Protein phylogenies and signature sequences: A reappraisal of evolutionary relationships among archaeobacteria, eubacteria, and eukaryotes. *Microbiol Mol Biol Rev* **62**:1435-91.
- Gupta, R. S. 1998b. What are archaeobacteria: life's third domain or monoderm prokaryotes related to gram-positive bacteria? A new proposal for the classification of prokaryotic organisms. *Molecular microbiology* **29**:695-707.
- Gupta, R. S. et G. B. Golding. 1996. The origin of the eukaryotic cell. *Trends Biochem Sci* **21**:166-71.
- Gurvitz, A., J. K. Hiltunen et A. J. Kastaniotis. 2009. Heterologous expression of mycobacterial proteins in *Saccharomyces cerevisiae* reveals two physiologically functional 3-hydroxyacyl-

- thioester dehydratases, HtdX and HtdY, in addition to HadABC and HtdZ. *Journal of bacteriology* **191**:2683-90.
- Gurvitz, A., J. K. Hiltunen et A. J. Kastaniotis. 2008. Identification of a novel mycobacterial 3-hydroxyacyl-thioester dehydratase, HtdZ (Rv0130), by functional complementation in yeast. *Journal of bacteriology* **190**:4088-90.
- Haldane, J. B. S. 1929. The origin of life in J. D. Bernal, ed. The Origin of Life. The Weidenfeld and Nicolson Natural History. Weidenfeld and Nicolson, London.
- Han, J., Q. Lu, L. Zhou, H. Liu et H. Xiang. 2009. Identification of the polyhydroxyalkanoate (PHA)-specific acetoacetyl coenzyme A reductase among multiple FabG paralogs in *Haloarcula hispanica* and reconstruction of the PHA biosynthetic pathway in *Haloferax volcanii*. *Applied and environmental microbiology* **75**:6168-75.
- Hanczyc, M. M., S. M. Fujikawa et J. W. Szostak. 2003. Experimental models of primitive cellular compartments: encapsulation, growth, and division. *Science (New York, N.Y)* **302**:618-22.
- Hanczyc, M. M. et J. W. Szostak. 2004. Replicating vesicles as models of primitive cell growth and division. *Curr Opin Chem Biol* **8**:660-4.
- Hargreaves, W. R. et D. W. Deamer. 1978. Liposomes from ionic, single-chain amphiphiles. *Biochemistry* **17**:3759-68.
- Hargreaves, W. R., S. J. Mulvihill et D. W. Deamer. 1977. Synthesis of phospholipids and membranes in prebiotic conditions. *Nature* **266**:78-80.
- Harkins, W. D. 1952. The physical chemistry of surface films, New York.
- Harkins, W. D., E. C. H. Davies et G. L. Clark. 1917. The orientation of molecules in the surfaces of liquids, the energy relations at surfaces, solubility, adsorption, emulsification, molecular association, and the effect of acids and bases on interfacial tension. *Journal of the American Chemical Society* **39**:541-96.
- Harris, C. M. et C. D. Poulter. 2000. Recent studies of the mechanism of protein prenylation. *Natural product reports* **17**:137-44.
- Harris, H. 1999. The birth of the cell. Yale university press, New Haven and London.
- Harris, J. K., S. T. Kelley, G. B. Spiegelman et N. R. Pace. 2003. The genetic core of the universal ancestor. *Genome Res* **13**:407-12.
- Hartman, H., P. Favaretto et T. F. Smith. 2006. The archaeal origins of the eukaryotic translational system. *Archaea* **2**:1-9.
- Hartman, H. et A. Fedorov. 2002. The origin of the eukaryotic cell: a genomic investigation. *Proceedings of the National Academy of Sciences of the United States of America* **99**:1420-5.
- Harvey, E. N. et H. Shapiro. 1934. The interfacial tension between oil and protoplasm within living cells. *Journal of Cellular and Comparative Physiology* **5**:255-67.
- Haydon, D. A. et J. Taylor. 1963. The stability and properties of bimolecular lipid leaflets in aqueous solutions. *Journal of Theoretical Biology* **4**:281-96.
- Hazen, R. M. et D. W. Deamer. 2007. Hydrothermal reactions of pyruvic acid: synthesis, selection, and self-assembly of amphiphilic molecules. *Orig Life Evol Biosph* **37**:143-52.
- He, S. et T. D. Fox. 1997. Membrane translocation of mitochondrially coded Cox2p: distinct requirements for export of N and C termini and dependence on the conserved protein Oxa1p. *Molecular biology of the cell* **8**:1449-60.

- Heath, R. J., J. Li, G. E. Roland et C. O. Rock. 2000a. Inhibition of the *Staphylococcus aureus* NADPH-dependent enoyl-acyl carrier protein reductase by triclosan and hexachlorophene. *The Journal of biological chemistry* **275**:4654-9.
- Heath, R. J. et C. O. Rock. 1995. Enoyl-acyl carrier protein reductase (fabI) plays a determinant role in completing cycles of fatty acid elongation in *Escherichia coli*. *The Journal of biological chemistry* **270**:26538-42.
- Heath, R. J. et C. O. Rock. 1996. Roles of the FabA and FabZ beta-hydroxyacyl-acyl carrier protein dehydratases in *Escherichia coli* fatty acid biosynthesis. *The Journal of biological chemistry* **271**:27795-801.
- Heath, R. J., N. Su, C. K. Murphy et C. O. Rock. 2000b. The enoyl-[acyl-carrier-protein] reductases FabI and FabL from *Bacillus subtilis*. *The Journal of biological chemistry* **275**:40128-33.
- Hechter, O. 1965. Role of Water Structure in the Molecular Organization of Cell Membranes. *Federation proceedings* **24**:S91-102.
- Hedl, M., L. Taberner, C. V. Stauffacher et V. W. Rodwell. 2004. Class II 3-hydroxy-3-methylglutaryl coenzyme A reductases. *Journal of bacteriology* **186**:1927-32.
- Hemmi, H., K. Shibuya, Y. Takahashi, T. Nakayama et T. Nishino. 2004. (S)-2,3-Di-O-geranylgeranylgeranyl glyceryl phosphate synthase from the thermoacidophilic archaeon *Sulfolobus solfataricus*. Molecular cloning and characterization of a membrane-intrinsic prenyltransferase involved in the biosynthesis of archaeal ether-linked membrane lipids. *The Journal of biological chemistry* **279**:50197-203.
- Henn, F. A., G. L. Decker, J. W. Greenawalt et T. E. Thompson. 1967. Properties of lipid bilayer membranes separating two aqueous phases: Electron microscope studies *Journal of molecular biology* **24**:51-54.
- Henning, T. et F. Salama. 1998. Carbon in the universe. *Science (New York, N.Y)* **282**:2204-10.
- Herdendorf, T. J. et H. M. Miziorko. 2006. Phosphomevalonate kinase: functional investigation of the recombinant human enzyme. *Biochemistry* **45**:3235-42.
- Hilario, E. et J. P. Gogarten. 1993. Horizontal transfer of ATPase genes--the tree of life becomes a net of life. *Biosystems* **31**:111-9.
- Hiltunen, J. K., Z. Chen, A. M. Haapalainen, R. K. Wierenga et A. J. Kastaniotis. 2010. Mitochondrial fatty acid synthesis--an adopted set of enzymes making a pathway of major importance for the cellular metabolism. *Progress in lipid research* **49**:27-45.
- Hiltunen, J. K. et Y. Qin. 2000. beta-oxidation - strategies for the metabolism of a wide variety of acyl-CoA esters. *Biochimica et biophysica acta* **1484**:117-28.
- Hinegardner, R. T. et J. Engelberg. 1963. Rationale for a Universal Genetic Code. *Science* **142**:1083-5.
- Hisano, T., T. Tsuge, T. Fukui, T. Iwata, K. Miki et Y. Doi. 2003. Crystal structure of the (R)-specific enoyl-CoA hydratase from *Aeromonas caviae* involved in polyhydroxyalkanoate biosynthesis. *The Journal of biological chemistry* **278**:617-24.
- Hoang, T. T., S. A. Sullivan, J. K. Cusick et H. P. Schweizer. 2002. Beta-ketoacyl acyl carrier protein reductase (FabG) activity of the fatty acid biosynthetic pathway is a determining factor of 3-oxo-homoserine lactone acyl chain lengths. *Microbiology (Reading, England)* **148**:3849-56.
- Hoffmeister, M., M. Piotrowski, U. Nowitzki et W. Martin. 2005. Mitochondrial trans-2-enoyl-CoA reductase of wax ester fermentation from *Euglena gracilis* defines a new family of enzymes involved in lipid synthesis. *The Journal of biological chemistry* **280**:4329-38.
- Hoja, U., S. Marthol, J. Hofmann, S. Stegner, R. Schulz, S. Meier, E. Greiner et E. Schweizer. 2004. HFA1 encoding an organelle-specific acetyl-CoA carboxylase controls mitochondrial fatty

- acid synthesis in *Saccharomyces cerevisiae*. *The Journal of biological chemistry* **279**:21779-86.
- Holm, L. et C. Sander. 1998. Touring protein fold space with Dali/FSSP. *Nucleic acids research* **26**:316-9.
- Holm, L. et C. Sander. 1996. Mapping the protein universe. *Science (New York, N.Y)* **273**:595-603.
- Hooke, R. 1665. *Micrographia: or some physiological descriptions of minute bodies made by magnifying glasses*. Royal Society, London.
- Hopwood, D. A. et D. H. Sherman. 1990. Molecular genetics of polyketides and its comparison to fatty acid biosynthesis. *Annual review of genetics* **24**:37-66.
- Horiike, T., K. Hamada, S. Kanaya et T. Shinozawa. 2001. Origin of eukaryotic cell nuclei by symbiosis of Archaea in Bacteria is revealed by homology-hit analysis. *Nat Cell Biol* **3**:210-4.
- Houten, S. M. et H. R. Waterham. 2001. Nonorthologous gene displacement of phosphomevalonate kinase. *Molecular genetics and metabolism* **72**:273-6.
- Huber, R., T. Wilharm, D. Huber, A. Trincone, S. Burggraf, H. König, R. Rachel, I. Rockinger, H. Fricke et K. O. Stetter. 1992. *Aquifex pyrophilus* gen. nov. sp. nov., represents a novel group of marine hyperthermophilic hydrogen-oxidizing bacteria. *Syst Appl Microbiol* **15**:340-51.
- Hugler, M., R. S. Krieger, M. Jahn et G. Fuchs. 2003. Characterization of acetyl-CoA/propionyl-CoA carboxylase in *Metallosphaera sedula*. Carboxylating enzyme in the 3-hydroxypropionate cycle for autotrophic carbon fixation. *European journal of biochemistry / FEBS* **270**:736-44.
- Ichihashi, N., T. Matsuura, H. Kita, T. Sunami, H. Suzuki et T. Yomo. 2010. Constructing partial models of cells. *Cold Spring Harbor perspectives in biology* **2**:a004945.
- Icho, T., C. P. Sparrow et C. R. Raetz. 1985. Molecular cloning and sequencing of the gene for CDP-diglyceride synthetase of *Escherichia coli*. *The Journal of biological chemistry* **260**:12078-83.
- Itabashi, Y. et A. Kuksis. 1997. Reassessment of stereochemical configuration of natural phosphatidylglycerols by chiral-phase high-performance liquid chromatography and electrospray mass spectrometry. *Analytical biochemistry* **254**:49-56.
- Iverson, V., R. M. Morris, C. D. Frazar, C. T. Berthiaume, R. L. Morales et E. V. Armbrust. 2012. Untangling genomes from metagenomes: revealing an uncultured class of marine Euryarchaeota. *Science (New York, N.Y)* **335**:587-90.
- Iwabata, H., K. Watanabe, T. Ohkuri, S. Yokobori et A. Yamagishi. 2005. Thermostability of ancestral mutants of *Caldococcus noboribetus* isocitrate dehydrogenase. *FEMS Microbiol Lett* **243**:393-8.
- Iwabe, N., K. Kuma, M. Hasegawa, S. Osawa et T. Miyata. 1989. Evolutionary relationship of archaeobacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes. *Proceedings of the National Academy of Sciences of the United States of America* **86**:9355-9.
- Jackowski, S. et C. O. Rock. 1987. Acetoacetyl-acyl carrier protein synthase, a potential regulator of fatty acid biosynthesis in bacteria. *The Journal of biological chemistry* **262**:7927-31.
- Jacob, F. 1981. *Le jeu des possibles. Essai sur la diversité du vivant*. Fayard, Paris.
- Jacobs, M. H. 1962. Early osmotic history of the plasma membrane. *Circulation* **26**:1013-21.
- Jaffres, J. B. D., G. A. Shields et K. Wallmann. 2007. The oxygen isotope evolution of seawater: a critical review of a long-standing controversy and an improved geological water cycle model for the past 3.4 billion years. *Earth-Science Reviews* **83**:83-122.

- Jahn, U., R. Summons, H. Sturt, E. Grosjean et H. Huber. 2004. Composition of the lipids of *Nanoarchaeum equitans* and their origin from its host *Ignicoccus* sp. strain KIN4/I. *Archives of microbiology* **182**:404-13.
- Jain, R., M. C. Rivera et J. A. Lake. 1999. Horizontal gene transfer among genomes: the complexity hypothesis. *Proceedings of the National Academy of Sciences of the United States of America* **96**:3801-6.
- Jain, R., M. C. Rivera, J. E. Moore et J. A. Lake. 2003. Horizontal gene transfer accelerates genome innovation and evolution. *Molecular biology and evolution* **20**:1598-602.
- Janas, T., T. Janas et M. Yarus. 2004. A membrane transporter for tryptophan composed of RNA. *Rna* **10**:1541-9.
- Jeffares, D. C., A. M. Poole et D. Penny. 1998. Relics from the RNA world. *J Mol Evol* **46**:18-36.
- Jekely, G. 2006. Did the last common ancestor have a biological membrane? *Biol Direct* **1**:35.
- Jenke-Kodama, H., A. Sandmann, R. Muller et E. Dittmann. 2005. Evolutionary implications of bacterial polyketide synthases. *Molecular biology and evolution* **22**:2027-39.
- Jenni, S., M. Leibundgut, D. Boehringer, C. Frick, B. Mikolasek et N. Ban. 2007. Structure of fungal fatty acid synthase and implications for iterative substrate shuttling. *Science (New York, N.Y)* **316**:254-61.
- Jensen, R. A. 1976. Enzyme recruitment in evolution of new function. *Annu Rev Microbiol* **30**:409-25.
- Jiang, C., S. Y. Kim et D. Y. Suh. 2008. Divergent evolution of the thiolase superfamily and chalcone synthase family. *Molecular phylogenetics and evolution* **49**:691-701.
- Jobb, G., A. von Haeseler et K. Strimmer. 2004. TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics. *BMC evolutionary biology* **4**:18.
- Johansson, P., A. Castell, T. A. Jones et K. Backbro. 2006. Structure and function of Rv0130, a conserved hypothetical protein from *Mycobacterium tuberculosis*. *Protein Sci* **15**:2300-9.
- John, U., B. Beszteri, E. Derelle, Y. Van de Peer, B. Read, H. Moreau et A. Cembella. 2008. Novel insights into evolution of protistan polyketide synthases through phylogenomic analysis. *Protist* **159**:21-30.
- Jones, W. J. et G. U. Holzer. 1991. The polar and neutral lipid composition of *Methanosphaera stadtmanae*. *Systematic and Applied Microbiology* **14**:130-34.
- Jornvall, H., J. O. Hoog et B. Persson. 1999. SDR and MDR: completed genome sequences show these protein families to be large, of old origin, and of complex nature. *FEBS letters* **445**:261-4.
- Jung, S., J. G. Zeikus et R. I. Hollingsworth. 1994. A new family of very long chain alpha,omega-dicarboxylic acids is a major structural fatty acyl component of the membrane lipids of *Thermoanaerobacter ethanolicus* 39E. *Journal of lipid research* **35**:1057-65.
- Junglas, B., A. Briegel, T. Burghardt, P. Walther, R. Wirth, H. Huber et R. Rachel. 2008. *Ignicoccus hospitalis* and *Nanoarchaeum equitans*: ultrastructure, cell-cell interaction, and 3D reconstruction from serial sections of freeze-substituted cells and by electron cryotomography. *Archives of microbiology* **190**:395-408.
- Kallberg, Y., U. Oppermann, H. Jornvall et B. Persson. 2002. Short-chain dehydrogenases/reductases (SDRs). *European journal of biochemistry / FEBS* **269**:4409-17.
- Kandler, O. 1995. Cell wall biochemistry in archaea and its phylogenetic implications. *Journal of biological Physics* **20**:165-69.

- Kaneda, K., T. Kuzuyama, M. Takagi, Y. Hayakawa et H. Seto. 2001. An unusual isopentenyl diphosphate isomerase found in the mevalonate pathway gene cluster from *Streptomyces* sp. strain CL190. *Proceedings of the National Academy of Sciences of the United States of America* **98**:932-7.
- Kass, L. R. et K. Bloch. 1967. On the enzymatic synthesis of unsaturated fatty acids in *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America* **58**:1168-73.
- Kasting, J. F. 1993. Earth's early atmosphere. *Science (New York, N.Y)* **259**:920-6.
- Kasting, J. F., M. Tazewell Howard, K. Wallmann, J. Veizer, G. A. Shields et J. Jaffres. 2006. Paleoclimates, ocean depth, and the oxygen isotopic composition of seawater. *Earth and Planetary Science Letters* **252**:82--93.
- Katayama, K., I. Sakurai et H. Wada. 2004. Identification of an *Arabidopsis thaliana* gene for cardiolipin synthase located in mitochondria. *FEBS letters* **577**:193-8.
- Kates, M., B. Palameta, C. N. Joo, D. J. Kushner et N. E. Gibbons. 1966. Aliphatic diether analogs of glyceride-derived lipids. IV. The occurrence of Di-O-dihydroxyphytylglycerol ether containing lipids in extremely halophilic bacteria. *Biochemistry* **5**:4092-99.
- Kates, M., M. K. Wassef et D. J. Kushner. 1968. Radioisotopic studies on the biosynthesis of the glyceryl diether lipids of *Halobacterium cutirubrum*. *Canadian journal of biochemistry* **46**:971-7.
- Kates, M., L. S. Yengoyan et P. S. Sastry. 1965. A Diether Analog of Phosphatidyl Glycerophosphate in *Halobacterium Cutirubrum*. *Biochimica et biophysica acta* **98**:252-68.
- Katoh, K., K. Kuma, H. Toh et T. Miyata. 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic acids research* **33**:511-8.
- Kavanagh, K. L., H. Jornvall, B. Persson et U. Oppermann. 2008. Medium- and short-chain dehydrogenase/reductase gene and protein families : the SDR superfamily: functional and structural diversity within a family of metabolic and regulatory enzymes. *Cell Mol Life Sci* **65**:3895-906.
- Kavanagh, K. L., M. Klimacek, B. Nidetzky et D. K. Wilson. 2003. Crystal structure of *Pseudomonas fluorescens* mannitol 2-dehydrogenase: evidence for a very divergent long-chain dehydrogenase family. *Chemico-biological interactions* **143-144**:551-8.
- Keeling, P. J. 2009. Chromalveolates and the evolution of plastids by secondary endosymbiosis. *The Journal of eukaryotic microbiology* **56**:1-8.
- Keller, Y., F. Bouvier, A. d'Harlingue et B. Camara. 1998. Metabolic compartmentation of plastid prenyl lipid biosynthesis--evidence for the involvement of a multifunctional geranylgeranyl reductase. *European journal of biochemistry / FEBS* **251**:413-7.
- Kellogg, B. A. et C. D. Poulter. 1997. Chain elongation in the isoprenoid biosynthetic pathway. *Current opinion in chemical biology* **1**:570-8.
- Kim, K. H., B. H. Ha, S. J. Kim, S. K. Hong, K. Y. Hwang et E. E. Kim. 2010. Crystal structures of Enoyl-ACP reductases I (FabI) and III (FabL) from *B. subtilis*. *Journal of molecular biology* **406**:403-15.
- Kimber, M. S., F. Martin, Y. Lu, S. Houston, M. Vedadi, A. Dharamsi, K. M. Fiebig, M. Schmid et C. O. Rock. 2004. The structure of (3R)-hydroxyacyl-acyl carrier protein dehydratase (FabZ) from *Pseudomonas aeruginosa*. *The Journal of biological chemistry* **279**:52593-602.
- King, C. C. 1990. Did membrane electrochemistry precede translation? *Orig Life Evol Biosph* **20**:15-25.

- Kito, M. et L. I. Pizer. 1969. Purification and regulatory properties of the biosynthetic L-glycerol 3-phosphate dehydrogenase from *Escherichia coli*. *The Journal of biological chemistry* **244**:3316-23.
- Klein, C., N. J. Beukes, H. D. Holland, J. F. Kasting, L. R. Kump et D. R. Lowe. 1992. Proterozoic atmosphere and ocean. Pp. 135-74 in J. W. Schopf et C. Kelin, eds. *The proterozoic biosphere*. Cambridge University Press, New York.
- Klein, H. P. 1957. Some observations on a cell free lipid synthesizing system from *Saccharomyces cerevisiae*. *Journal of bacteriology* **73**:530-37.
- Klein, M., M. Friedrich, A. J. Roger, P. Hugenholtz, S. Fishbain, H. Abicht, L. L. Blackall, D. A. Stahl et M. Wagner. 2001. Multiple lateral transfers of dissimilatory sulfite reductase genes between major lineages of sulfate-reducing prokaryotes. *Journal of bacteriology* **183**:6028-35.
- Kleinig, H. 1989. The role of plastids in isoprenoid biosynthesis. *Annual Reviews in Plant Physiology and Molecular Biology* **40**:39-59.
- Kluyver, A. J. 1959. *Albert Jan Kluyver. His life and work*. Interscience publishers, New York.
- Knauth, L. P. et D. R. Lowe. 2003. High Archean climatic temperature inferred from oxygen isotope geochemistry of cherts in the 3.5 Ga Swaziland Supergroup, South Africa. *Geological Society of America Bulletin* **115**:566-80.
- Knight, R. D., S. J. Freeland et L. F. Landweber. 2001. Rewiring the keyboard: evolvability of the genetic code. *Nat Rev Genet* **2**:49-58.
- Koga, Y. 2010. Early evolution of membrane lipids: how did the lipid divide occur? *J Mol Evol* **72**:274-82.
- Koga, Y., T. Kyuragi, M. Nishihara et N. Sone. 1998b. Did archaeal and bacterial cells arise independently from noncellular precursors? A hypothesis stating that the advent of membrane phospholipid with enantiomeric glycerophosphate backbones caused the separation of the two lines of descent. *Journal of molecular evolution* **46**:54-63.
- Koga, Y. et H. Morii. 2007. Biosynthesis of ether-type polar lipids in archaea and evolutionary considerations. *Microbiol Mol Biol Rev* **71**:97-120.
- Koga, Y. et H. Morii. 2005. Recent advances in structural research on ether lipids from archaea including comparative and physiological aspects. *Bioscience, biotechnology, and biochemistry* **69**:2019-34.
- Kolbe, M., H. Besir, L. O. Essen et D. Oesterhelt. 2000. Structure of the light-driven chloride pump halorhodopsin at 1.8 Å resolution. *Science (New York, N.Y)* **288**:1390-6.
- Koonin, E. V. 2003. Comparative genomics, minimal gene-sets and the last universal common ancestor. *Nat Rev Microbiol* **1**:127-36.
- Koonin, E. V. 2007. The Biological Big Bang model for the major transitions in evolution. *Biol Direct* **2**:21.
- Koonin, E. V. 2010. The origin and early evolution of eukaryotes in the light of phylogenomics. *Genome biology* **11**:209.
- Koonin, E. V. et W. Martin. 2005. On the origin of genomes and cells within inorganic compartments. *Trends Genet* **21**:647-54.
- Koonin, E. V., T. G. Senkevich et V. V. Dolja. 2006. The ancient Virus World and evolution of cells. *Biol Direct* **1**:29.
- Korn, E. D. 1966. Structure of biological membranes. *Science (New York, N.Y)* **153**:1491-8.

- Koski, K. M., A. M. Haapalainen, J. K. Hiltunen et T. Glumoff. 2005. Crystal structure of 2-enoyl-CoA hydratase 2 from human peroxisomal multifunctional enzyme type 2. *Journal of molecular biology* **345**:1157-69.
- Koski, M. K., A. M. Haapalainen, J. K. Hiltunen et T. Glumoff. 2004. A two-domain structure of one subunit explains unique features of eukaryotic hydratase 2. *The Journal of biological chemistry* **279**:24666-72.
- Kroken, S., N. L. Glass, J. W. Taylor, O. C. Yoder et B. G. Turgeon. 2003. Phylogenomic analysis of type I polyketide synthase genes in pathogenic and saprobic ascomycetes. *Proceedings of the National Academy of Sciences of the United States of America* **100**:15670-5.
- Kuhn, T. S. 1962. La structure des révolutions scientifiques. Flammarion.
- Kunin, V., D. Ahren, L. Goldovsky, P. Janssen et C. A. Ouzounis. 2005a. Measuring genome conservation across taxa: divided strains and united kingdoms. *Nucleic Acids Res* **33**:616-21.
- Kunin, V., L. Goldovsky, N. Darzentas et C. A. Ouzounis. 2005b. The net of life: reconstructing the microbial phylogenetic network. *Genome Res* **15**:954-9.
- Kurland, C. G., B. Canback et O. G. Berg. 2003. Horizontal gene transfer: a critical view. *Proceedings of the National Academy of Sciences of the United States of America* **100**:9658-62.
- Kurland, C. G., L. J. Collins et D. Penny. 2006. Genomics and the irreducible nature of eukaryote cells. *Science* **312**:1011-4.
- Kuzuyama, T. 2002. Mevalonate and nonmevalonate pathways for the biosynthesis of isoprene units. *Bioscience, biotechnology, and biochemistry* **66**:1619-27.
- Kuzuyama, T., J. P. Noel et S. B. Richard. 2005. Structural basis for the promiscuous biosynthetic prenylation of aromatic natural products. *Nature* **435**:983-7.
- Kwan, D. C., J. R. Thomas et A. Bolhuis. 2008. Bioenergetic requirements of a Tat-dependent substrate in the halophilic archaeon *Haloarcula hispanica*. *The FEBS journal* **275**:6159-67.
- Kyrpides, N., R. Overbeek et C. Ouzounis. 1999. Universal protein families and the functional content of the last universal common ancestor. *J Mol Evol* **49**:413-23.
- Lai, C. Y. et J. E. Cronan. 2004. Isolation and characterization of beta-ketoacyl-acyl carrier protein reductase (fabG) mutants of *Escherichia coli* and *Salmonella enterica* serovar Typhimurium. *Journal of bacteriology* **186**:1869-78.
- Lake, J. A. 1988. Origin of the eukaryotic nucleus determined by rate-invariant analysis of rRNA sequences. *Nature* **331**:184-6.
- Lake, J. A., E. Henderson, M. Oakes et M. W. Clark. 1984. Eocytes: a new ribosome structure indicates a kingdom with a close relationship to eukaryotes. *Proceedings of the National Academy of Sciences of the United States of America* **81**:3786-90.
- Lake, J. A., J. A. Servin, C. W. Herbold et R. G. Skophammer. 2008. Evidence for a new root of the tree of life. *Syst Biol* **57**:835-43.
- Lake, J. A., R. G. Skophammer, C. W. Herbold et J. A. Servin. 2009. Genome beginnings: rooting the tree of life. *Philosophical transactions of the Royal Society of London* **364**:2177-85.
- Lam, W. L. et W. F. Doolittle. 1992. Mevinolin-resistant mutations identify a promoter and the gene for a eukaryote-like 3-hydroxy-3-methylglutaryl-coenzyme A reductase in the archaeobacterium *Haloferax volcanii*. *The Journal of biological chemistry* **267**:5829-34.
- Lancet, M. S. et E. Anders. 1970. Carbon isotope fractionation in the Fischer-tropsch synthesis and in meteorites. *Science (New York, N.Y)* **170**:980-2.

- Lange, B. M. et M. Ghassemian. 2003. Genome organization in *Arabidopsis thaliana*: a survey for genes involved in isoprenoid and chlorophyll metabolism. *Plant molecular biology* **51**:925-48.
- Lange, B. M., T. Rujan, W. Martin et R. Croteau. 2000. Isoprenoid biosynthesis: the evolution of two ancient and distinct pathways across genomes. *Proceedings of the National Academy of Sciences of the United States of America* **97**:13172-7.
- Langmuir, I. 1917. The constitution and fundamental properties of solids and liquids. *Journal of the American Chemical Society* **39**:1848-906.
- Langworthy, T. A., G. Holzer, J. G. Zeikus et T. G. Tornabene. 1983. Iso- and anteiso-branched glycerol diethers of the thermophilic anaerobe *Thermodesulfotobacterium commune*. *Syst Appl Microbiol* **4**:1-17.
- Langworthy, T. A., W. R. Mayberry et P. F. Smith. 1974. Long-chain glycerol diether and polyol dialkyl glycerol triether lipids of *Sulfolobus acidocaldarius*. *Journal of bacteriology* **119**:106-16.
- Langworthy, T. A., P. F. Smith et W. R. Mayberry. 1972. Lipids of *Thermoplasma acidophilum*. *Journal of bacteriology* **112**:1193-200.
- Larsen, O., T. Lien et N. K. Birkeland. 1999. Dissimilatory sulfite reductase from *Archaeoglobus profundus* and *Desulfotomaculum thermocisternum*: phylogenetic and structural implications from gene sequences. *Extremophiles* **3**:63-70.
- Lawless, J. G. et G. Yuen. 1979. Quantification of monocarboxylic acids in the Murchison carbonaceous meteorite. *Nature* **282**:396-97.
- Lawrence, J. G. et H. Ochman. 1998. Molecular archaeology of the *Escherichia coli* genome. *Proceedings of the National Academy of Sciences of the United States of America* **95**:9413-7.
- Lawson, F. S., R. L. Charlebois et J. A. Dillon. 1996. Phylogenetic analysis of carbamoylphosphate synthetase genes: complex evolutionary history includes an internal duplication within a gene which can root the tree of life. *Molecular biology and evolution* **13**:970-7.
- Lazcano, A. 2010. Historical development of origins research. *Cold Spring Harb Perspect Biol* **2**:a002089.
- Le, S. Q. et O. Gascuel. 2008. An improved general amino acid replacement matrix. *Molecular biology and evolution* **25**:1307-20.
- Leaver, M., P. Dominguez-Cuevas, J. M. Coxhead, R. A. Daniel et J. Errington. 2009. Life without a wall or division machine in *Bacillus subtilis*. *Nature* **457**:849-53.
- Lebrun, E., J. M. Santini, M. Brugna, A. L. Ducluzeau, S. Ouchane, B. Schoepp-Cothenet, F. Baymann et W. Nitschke. 2006. The Rieske protein: a case study on the pitfalls of multiple sequence alignments and phylogenetic reconstruction. *Molecular biology and evolution* **23**:1180-91.
- Leesong, M., B. S. Henderson, J. R. Gillig, J. M. Schwab et J. L. Smith. 1996. Structure of a dehydratase-isomerase from the bacterial pathway for biosynthesis of unsaturated fatty acids: two catalytic activities in one active site. *Structure* **4**:253-64.
- Leibundgut, M., T. Maier, S. Jenni et N. Ban. 2008. The multienzyme architecture of eukaryotic fatty acid synthases. *Current opinion in structural biology* **18**:714-25.
- Leipe, D. D., L. Aravind et E. V. Koonin. 1999. Did DNA replication evolve twice independently? *Nucleic Acids Res* **27**:3389-401.
- Lenard, J. et S. J. Singer. 1966. Protein conformation in cell membrane preparations as studied by optical rotatory dispersion and circular dichroism. *Proceedings of the National Academy of Sciences of the United States of America* **56**:1828-35.

- Lerat, E., V. Daubin, H. Ochman et N. A. Moran. 2005. Evolutionary origins of genomic repertoires in bacteria. *PLoS Biol* **3**:e130.
- Liang, P. H., T. P. Ko et A. H. Wang. 2002. Structure, mechanism and function of prenyltransferases. *European journal of biochemistry / FEBS* **269**:3339-54.
- Lightner, V. A., T. J. Larson, P. TAILLEUR, G. D. Kantor, C. R. Raetz, R. M. Bell et P. Modrich. 1980. Membrane phospholipid synthesis in *Escherichia coli*. Cloning of a structural gene (plsB) of the sn-glycerol-3-phosphate acyltransferase. *The Journal of biological chemistry* **255**:9413-20.
- Lindas, A. C., E. A. Karlsson, M. T. Lindgren, T. J. Ettema et R. Bernander. 2008. A unique cell division machinery in the Archaea. *Proceedings of the National Academy of Sciences of the United States of America* **105**:18942-6.
- Line, M. A. 2002. The enigma of the origin of life and its timing. *Microbiology* **148**:21-7.
- Liu, T., S. J. Liu, Y. Xue, Y. Ma et P. Zhou. 2002. Purification and characterization of an extremely halophilic acetoacetyl-CoA thiolase from a newly isolated *Halobacterium* strain ZP-6. *Extremophiles* **6**:97-102.
- Lloyd, S. 2001. Measures of complexity. A non-exhaustive list. *Control systems IEEE* **21**:7-8.
- Loeb, J. 1906. The dynamics of living matter. The columbia university press, New York.
- Lombard, J., P. López-García et D. Moreira. 2012a. An ACP-Independent Fatty Acid Synthesis Pathway in Archaea: Implications for the Origin of Phospholipids. *Molecular biology and evolution* **29**:3261-5.
- Lombard, J., P. López-García et D. Moreira. 2012b. Phylogenomic Investigation of Phospholipid Synthesis in Archaea. *Archaea (Vancouver, B.C)*:In press.
- Lombard, J., P. López-García et D. Moreira. 2012c. The early evolution of lipid membranes and the three domains of life. *Nature Reviews Microbiology*. **10**:507-515.
- Lombard, J. et D. Moreira. 2011a. Early evolution of the biotin-dependent carboxylase family. *BMC evolutionary biology* **11**:232.
- Lombard, J. et D. Moreira. 2011b. Origins and early evolution of the mevalonate pathway of isoprenoid biosynthesis in the three domains of life. *Molecular biology and evolution* **28**:87-99.
- Longo, G. S., D. H. Thompson et I. Szleifer. 2007. Stability and phase separation in mixed monopolar lipid/bolalipid layers. *Biophysical journal* **93**:2609-21.
- López-García, P. et D. Moreira. 2006. Selective forces for the origin of the eukaryotic nucleus. *Bioessays* **28**:525-33.
- López-García, P. et D. Moreira. 2008. Tracking microbial biodiversity through molecular and genomic ecology. *Research in microbiology* **159**:67-73.
- López-Lara, I. M. et O. Geiger. 2001. The nodulation protein NodG shows the enzymatic activity of an 3-oxoacyl-acyl carrier protein reductase. *Mol Plant Microbe Interact* **14**:349-57.
- López, P., P. Forterre et H. Philippe. 1999. The root of the tree of life in the light of the covarion model. *J Mol Evol* **49**:496-508.
- Love, S. G. et D. E. Brownlee. 1993. A direct measurement of the terrestrial mass accretion rate of cosmic dust. *Science (New York, N.Y)* **262**:550-3.
- Lovejoy, A. O. 1936. The great chain of being. Harvard university press, Cambridge MA.

- Lu, Y. J., S. W. White et C. O. Rock. 2005. Domain swapping between *Enterococcus faecalis* FabN and FabZ proteins localizes the structural determinants for isomerase activity. *The Journal of biological chemistry* **280**:30342-8.
- Lu, Y. J., Y. M. Zhang, K. D. Grimes, J. Qi, R. E. Lee et C. O. Rock. 2006. Acyl-phosphates initiate membrane phospholipid synthesis in Gram-positive pathogens. *Molecular cell* **23**:765-72.
- Lundin, D., S. Gribaldo, E. Torrents, B. M. Sjöberg et A. M. Poole. 2010. Ribonucleotide reduction - horizontal transfer of a required function spans all three domains. *BMC evolutionary biology* **10**:383.
- Lynen, F. 1964. The biological synthesis of cholesterol. *Nobel Lectures, Physiology or Medicine 1963-1970*.
- Maddy, A. H. et B. R. Malcolm. 1965. Protein conformations in the plasma membrane. *Science* **150**:1616-8.
- Madigan, M. T., J. M. Martinko, P. V. Dunlap et D. P. Clark. 2009. Brock, Biology of microorganisms. Pearson, san Francisco.
- Magrum, L. J., K. R. Luehrs et C. R. Woese. 1978. Are extreme halophiles actually "bacteria"? *J Mol Evol* **11**:1-8.
- Maheen, G., G. Tian, Y. Wang, C. He, Z. Shi, H. Yuan et S. Feng. 2010. Resolving the enigma of prebiotic C-O-P bond formation: prebiotic hydrothermal synthesis of important biological phosphate esters. *Heteroatom Chemistry* **21**:161-67.
- Maier, T., S. Jenni et N. Ban. 2006. Architecture of mammalian fatty acid synthase at 4.5 Å resolution. *Science (New York, N.Y)* **311**:1258-62.
- Maier, T., M. Leibundgut et N. Ban. 2008. The crystal structure of a mammalian fatty acid synthase. *Science (New York, N.Y)* **321**:1315-22.
- Mangold, H. K. et F. Paltauf. 1983. Ether Lipids: Biochemical and Biomedical Aspects. Academic Press.
- Mansy, S. S. 2010. Membrane transport in primitive cells. *Cold Spring Harbor perspectives in biology* **2**:a002188.
- Mansy, S. S., J. P. Schrum, M. Krishnamurthy, S. Tobe, D. A. Treco et J. W. Szostak. 2008. Template-directed synthesis of a genetic polymer in a model protocell. *Nature* **454**:122-5.
- Mansy, S. S. et J. W. Szostak. 2008. Thermostability of model protocell membranes. *Proceedings of the National Academy of Sciences of the United States of America* **105**:13351-5.
- Margoliash, E. 1963. Primary Structure and Evolution of Cytochrome C. *Proceedings of the National Academy of Sciences of the United States of America* **50**:672-9.
- Margulis, L. 1996. Archaeal-eubacterial mergers in the origin of Eukarya: phylogenetic classification of life. *Proceedings of the National Academy of Sciences of the United States of America* **93**:1071-6.
- Margulis, L., M. F. Dolan et R. Guerrero. 2000. The chimeric eukaryote: origin of the nucleus from the karyomastigont in amitochondriate protists. *Proceedings of the National Academy of Sciences of the United States of America* **97**:6954-9.
- Marrakchi, H., K. H. Choi et C. O. Rock. 2002a. A new mechanism for anaerobic unsaturated fatty acid formation in *Streptococcus pneumoniae*. *The Journal of biological chemistry* **277**:44809-16.
- Marrakchi, H., W. E. Dewolf, Jr., C. Quinn et al. 2003. Characterization of *Streptococcus pneumoniae* enoyl-(acyl-carrier protein) reductase (FabK). *The Biochemical journal* **370**:1055-62.

- Marrakchi, H., S. Ducasse, G. Labesse, H. Montrozier, E. Margeat, L. Emorine, X. Charpentier, M. Daffe et A. Quemard. 2002b. MabA (FabG1), a Mycobacterium tuberculosis protein involved in the long-chain fatty acid elongation system FAS-II. *Microbiology (Reading, England)* **148**:951-60.
- Marris, E. 2008. Disputed definitions. Paradigm shift. *Nature* **455**:1023-28.
- Martin, W. et M. Muller. 1998. The hydrogen hypothesis for the first eukaryote. *Nature* **392**:37-41.
- Martin, W. et M. J. Russell. 2003. On the origins of cells: a hypothesis for the evolutionary transitions from abiotic geochemistry to chemoautotrophic prokaryotes, and from prokaryotes to nucleated cells. *Philosophical transactions of the Royal Society of London* **358**:59-83; discussion 83-5.
- Massengo-Tiasse, R. P. et J. E. Cronan. 2008. Vibrio cholerae FabV defines a new class of enoyl-acyl carrier protein reductase. *The Journal of biological chemistry* **283**:1308-16.
- Massengo-Tiasse, R. P. et J. E. Cronan. 2009. Diversity in enoyl-acyl carrier protein reductases. *Cell Mol Life Sci* **66**:1507-17.
- Matsumi, R., H. Atomi, A. J. Driessen et J. van der Oost. 2011. Isoprenoid biosynthesis in Archaea-- biochemical and evolutionary implications. *Research in microbiology* **162**:39-52.
- Matsuzaki, M., H. Kuroiwa, T. Kuroiwa, K. Kita et H. Nozaki. 2008. A cryptic algal group unveiled: a plastid biosynthesis pathway in the oyster parasite Perkinsus marinus. *Molecular biology and evolution* **25**:1167-79.
- Maurer, S. E., D. W. Deamer, J. M. Boncella et P. A. Monnard. 2009. Chemical evolution of amphiphiles: glycerol monoacyl derivatives stabilize plausible prebiotic membranes. *Astrobiology* **9**:979-87.
- Maynard-Smith, J. et E. Szathmary. 1995. The major transitions in evolution. Oxford university press.
- Maynard Smith, J. 1970. Time in the evolutionary process. . *Studium Generale* **23**:266-72.
- Mayr, E. 1998. Two empires or three? *Proceedings of the National Academy of Sciences of the United States of America* **95**:9720-3.
- McCollom, T. M., G. Ritter et B. R. Simoneit. 1999. Lipid synthesis under hydrothermal conditions by Fischer-Tropsch-type reactions. *Orig Life Evol Biosph* **29**:153-66.
- McCollom, T. M. et J. S. Seewald. 2007. Abiotic synthesis of organic compounds in deep-sea hydrothermal environments. *Chem Rev* **107**:382-401.
- McGarvey, D. J. et R. Croteau. 1995. Terpenoid metabolism. *The Plant cell* **7**:1015-26.
- McShea, D. W. 1991. Complexity and evolution: what everybody knows. *Biology and philosophy* **6**:303-24.
- Menendez, C., Z. Bauer, H. Huber, N. Gad'on, K. O. Stetter et G. Fuchs. 1999. Presence of acetyl coenzyme A (CoA) carboxylase and propionyl-CoA carboxylase in autotrophic Crenarchaeota and indication for operation of a 3-hydroxypropionate cycle in autotrophic carbon fixation. *Journal of bacteriology* **181**:1088-98.
- Meyer, T. E., M. A. Cusanovich et M. D. Kamen. 1986. Evidence against use of bacterial amino acid sequence data for construction of all-inclusive phylogenetic trees. *Proceedings of the National Academy of Sciences of the United States of America* **83**:217-20.
- Miinalainen, I. J., Z. J. Chen, J. M. Torkko, P. L. Pirila, R. T. Sormunen, U. Bergmann, Y. M. Qin et J. K. Hiltunen. 2003. Characterization of 2-enoyl thioester reductase from mammals. An ortholog of YBR026p/MRF1'p of the yeast mitochondrial fatty acid synthesis type II. *The Journal of biological chemistry* **278**:20154-61.

- Mileykovskaya, E. et W. Dowhan. 2005. Role of membrane lipids in bacterial division-site selection. *Current opinion in microbiology* **8**:135-42.
- Miller, S. L. 1953. A production of amino acids under possible primitive earth conditions. *Science (New York, N.Y)* **117**:528-9.
- Miller, S. L. 1955. Production of some organic compounds under possible primitive earth conditions. *Journal of the American chemical society* **77**:2351-61.
- Miller, S. L. et A. Lazcano. 1995. The origin of life--did it occur at high temperatures? *J Mol Evol* **41**:689-92.
- Miranda, I., R. Silva et M. A. Santos. 2006. Evolution of the genetic code in yeasts. *Yeast* **23**:203-13.
- Mirkin, B. G., T. I. Fenner, M. Y. Galperin et E. V. Koonin. 2003. Algorithms for computing parsimonious evolutionary scenarios for genome evolution, the last universal common ancestor and dominance of horizontal gene transfer in the evolution of prokaryotes. *BMC evolutionary biology* **3**:2.
- Misra, A., N. Surolia et A. Surolia. 2009. Catalysis and mechanism of malonyl transferase activity in type II fatty acid biosynthesis acyl carrier proteins. *Molecular bioSystems* **5**:651-9.
- Miyazaki, J., S. Nakaya, T. Suzuki, M. Tamakoshi, T. Oshima et A. Yamagishi. 2001. Ancestral residues stabilizing 3-isopropylmalate dehydrogenase of an extreme thermophile: experimental evidence supporting the thermophilic common ancestor hypothesis. *J Biochem* **129**:777-82.
- Mohan, S., T. M. Kelly, S. S. Eveland, C. R. Raetz et M. S. Anderson. 1994. An Escherichia coli gene (FabZ) encoding (3R)-hydroxymyristoyl acyl carrier protein dehydrase. Relation to fabA and suppression of mutations in lipid A biosynthesis. *The Journal of biological chemistry* **269**:32896-903.
- Monnard, P. A., C. L. Apel, A. Kanavarioti et D. W. Deamer. 2002. Influence of ionic inorganic solutes on self-assembly and polymerization processes related to early forms of life: implications for a prebiotic aqueous medium. *Astrobiology* **2**:139-52.
- Monnard, P. A. et D. W. Deamer. 2003. Preparation of vesicles from nonphospholipid amphiphiles. *Methods Enzymol* **372**:133-51.
- Monnard, P. A. et D. W. Deamer. 2001. Nutrient uptake by protocells: a liposome model system. *Orig Life Evol Biosph* **31**:147-55.
- Monnard, P. A. et D. W. Deamer. 2002. Membrane self-assembly processes: steps toward the first cellular life. *Anat Rec* **268**:196-207.
- Monnard, P. A., A. Luptak et D. W. Deamer. 2007. Models of primitive cellular life: polymerases and templates in liposomes. *Philosophical transactions of the Royal Society of London* **362**:1741-50.
- Monod, J. et F. Jacob. 1961. Teleonomic mechanisms in cellular metabolism, growth, and differentiation. *Cold Spring Harb Symp Quant Biol* **26**:389-401.
- Moor, H. et K. Muhlethaler. 1963. Fine Structure in Frozen-Etched Yeast Cells. *J Cell Biol* **17**:609-28.
- Moore, B. S. et J. N. Hopke. 2001. Discovery of a new bacterial polyketide biosynthetic pathway. *Chembiochem* **2**:35-8.
- Moore, J. A. et C. D. Poulter. 1997. Escherichia coli dimethylallyl diphosphate:tRNA dimethylallyltransferase: a binding mechanism for recombinant enzyme. *Biochemistry* **36**:604-14.
- Moore, R. L. et B. J. McCarthy. 1967. Comparative study of ribosomal ribonucleic acid cistrons in enterobacteria and myxobacteria. *J Bacteriol* **94**:1066-74.

- Morange, M. 2011. Some considerations on the nature of LUCA, and the nature of life. *Research in microbiology* **162**:5-9.
- Moreira, D. et P. López-García. 1998. Symbiosis between methanogenic archaea and delta-proteobacteria as the origin of eukaryotes: the syntrophic hypothesis. *J Mol Evol* **47**:517-30.
- Moreno, A. et K. Ruiz-Mirazo. 2009. The problem of the emergence of functional diversity in prebiotic evolution. *Biology & Philosophy* **24**:585-605.
- Morgan-Kiss, R. M. et J. E. Cronan. 2008. The Lactococcus lactis FabF fatty acid synthetic enzyme can functionally replace both the FabB and FabF proteins of Escherichia coli and the FabH protein of Lactococcus lactis. *Archives of microbiology* **190**:427-37.
- Morii, H., S. Kiyonari, Y. Ishino et Y. Koga. 2009. A novel biosynthetic pathway of archaetidyl-myoinositol via archaetidyl-myoinositol phosphate from CDP-archaeol and D-glucose 6-phosphate in methanoarchaeon Methanothermobacter thermoautotrophicus cells. *The Journal of biological chemistry* **284**:30766-74.
- Morii, H. et Y. Koga. 2003. CDP-2,3-Di-O-geranylgeranyl-sn-glycerol:L-serine O-archaetidyltransferase (archaetidylserine synthase) in the methanogenic archaeon Methanothermobacter thermoautotrophicus. *Journal of bacteriology* **185**:1181-9.
- Morii, H., M. Nishihara et Y. Koga. 2000. CTP:2,3-di-O-geranylgeranyl-sn-glycerol-1-phosphate cytidyltransferase in the methanogenic archaeon Methanothermobacter thermoautotrophicus. *The Journal of biological chemistry* **275**:36568-74.
- Morowitz, H. J. 1999. A theory of biochemical organization, metabolic pathways and evolution. *Complexity* **4**:39-53.
- Morowitz, H. J. 1992. Beginnings of cellular life. Metabolism recapitulates biogenesis. Yale university press, New Haven, CT.
- Morowitz, H. J., D. W. Deamer et T. Smith. 1991. Biogenesis as an evolutionary process [letter]. *J Mol Evol* **33**:207-8.
- Morowitz, H. J., B. Heinz et D. W. Deamer. 1988. The chemical logic of a minimum protocell. *Orig Life Evol Biosph* **18**:281-7.
- Moskowitz, G. J. et J. M. Merrick. 1969. Metabolism of poly-beta-hydroxybutyrate. II. Enzymatic synthesis of D-(-)-beta hydroxybutyryl coenzyme A by an enoyl hydratase from Rhodospirillum rubrum. *Biochemistry* **8**:2748-55.
- Moss, S. J., C. J. Martin et B. Wilkinson. 2004. Loss of co-linearity by modular polyketide synthases: a mechanism for the evolution of chemical diversity. *Natural product reports* **21**:575-93.
- Mould, R. M. et C. Robinson. 1991. A proton gradient is required for the transport of two luminal oxygen-evolving proteins across the thylakoid membrane. *The Journal of biological chemistry* **266**:12189-93.
- Moustafa, A., A. Reyes-Prieto et D. Bhattacharya. 2008. Chlamydiae has contributed at least 55 genes to Plantae with predominantly plastid functions. *PLoS ONE* **3**:e2205.
- Mulkidjanian, A. Y., M. Y. Galperin et E. V. Koonin. 2009. Co-evolution of primordial membranes and membrane proteins. *Trends in biochemical sciences* **34**:206-15.
- Mulkidjanian, A. Y., M. Y. Galperin, K. S. Makarova, Y. I. Wolf et E. V. Koonin. 2008. Evolutionary primacy of sodium bioenergetics. *Biology direct* **3**:13.
- Mulkidjanian, A. Y., K. S. Makarova, M. Y. Galperin et E. V. Koonin. 2007. Inventing the dynamo machine: the evolution of the F-type and V-type ATPases. *Nature reviews* **5**:892-9.
- Muller, R. 2004. Don't classify polyketide synthases. *Chemistry & biology* **11**:4-6.

- Murakami, M., K. Shibuya, T. Nakayama, T. Nishino, T. Yoshimura et H. Hemmi. 2007. Geranylgeranyl reductase involved in the biosynthesis of archaeal membrane lipids in the hyperthermophilic archaeon *Archaeoglobus fulgidus*. *The FEBS journal* **274**:805-14.
- Mushegian, A. R. et E. V. Koonin. 1996. A minimal gene set for cellular life derived by comparison of complete bacterial genomes. *Proceedings of the National Academy of Sciences of the United States of America* **93**:10268-73.
- Nagano, N., C. A. Orengo et J. M. Thornton. 2002. One fold with many functions: the evolutionary relationships between TIM barrel families based on their sequences, structures and functions. *Journal of molecular biology* **321**:741-65.
- Nakanishi-Matsui, M. et M. Futai. 2006. Stochastic proton pumping ATPases: from single molecules to diverse physiological roles. *IUBMB life* **58**:318-22.
- Nam, H., N. E. Lewis, J. A. Lerman, D. H. Lee, R. L. Chang, D. Kim et B. O. Palsson. 2012. Network context and selection in the evolution to enzyme specificity. *Science* **337**:1101-4.
- Naraoka, H., A. Shimoyama et K. Harada. 1999. Molecular distribution of monocarboxylic acids in Asuka carbonaceous chondrites from Antarctica. *Orig Life Evol Biosph* **29**:187-201.
- Nassoy, P., M. Goldmann, O. Bouloussa et F. Rondelez. 1995. Spontaneous chiral segregation in bidimensional films. *Physical review letters* **75**:457-60.
- Nemoto, N., T. Oshima et A. Yamagishi. 2003. Purification and characterization of geranylgeranyl glyceryl phosphate synthase from a thermoacidophilic archaeon, *Thermoplasma acidophilum*. *Journal of biochemistry* **133**:651-7.
- Nesbo, C. L., Y. Boucher et W. F. Doolittle. 2001. Defining the core of nontransferable prokaryotic genes: the euryarchaeal core. *J Mol Evol* **53**:340-50.
- Ng, S. Y., B. Chaban, D. J. VanDyke et K. F. Jarrell. 2007. Archaeal signal peptidases. *Microbiology (Reading, England)* **153**:305-14.
- Nicolson, G. L. et S. J. Singer. 1971. Ferritin-conjugated plant agglutinins as specific saccharide stains for electron microscopy: application to saccharides bound to cell membranes. *Proceedings of the National Academy of Sciences of the United States of America* **68**:942-5.
- Nisbet, E. G. et N. H. Sleep. 2001. The habitat and nature of early life. *Nature* **409**:1083-91.
- Nishihara, M. et Y. Koga. 1997. Purification and properties of sn-glycerol-1-phosphate dehydrogenase from *Methanobacterium thermoautotrophicum*: characterization of the biosynthetic enzyme for the enantiomeric glycerophosphate backbone of ether polar lipids of Archaea. *Journal of biochemistry* **122**:572-6.
- Nishihara, M., S. Nagahama, M. Ohga et Y. Koga. 2000. Straight-chain fatty alcohols in the hyperthermophilic archaeon *Pyrococcus furiosus*. *Extremophiles* **4**:275-7.
- Nishihara, M., T. Yamazaki, T. Oshima et Y. Koga. 1999. sn-glycerol-1-phosphate-forming activities in Archaea: separation of archaeal phospholipid biosynthesis and glycerol catabolism by glycerophosphate enantiomers. *Journal of bacteriology* **181**:1330-3.
- Nishimura, Y. et T. Eguchi. 2006. Biosynthesis of archaeal membrane lipids: digeranylgeranyl glycerophospholipid reductase of the thermoacidophilic archaeon *Thermoplasma acidophilum*. *Journal of biochemistry* **139**:1073-81.
- Nomura, C. T., T. Tanaka, T. E. Eguen, A. S. Appah, K. Matsumoto, S. Taguchi, C. L. Ortiz et Y. Doi. 2008. FabG mediates polyhydroxyalkanoate production from both related and nonrelated carbon sources in recombinant *Escherichia coli* LS5218. *Biotechnology progress* **24**:342-51.
- Nooner, D. W., J. M. Gibert, E. Gelpi et J. Oro. 1976. Closed system Fischer-Tropsch synthesis over meteoritic iron, iron ore and nickel-iron alloy. *Geochimica et Cosmochimica Acta* **40**:915-24.

- Nooner, D. W. et J. Oro. 1979. Synthesis of Fatty Acids by a Closed System Fischer-Tropsch Process. *Advances in Chemistry* **178**:159-71.
- Nooner, D. W., E. Sherwood, M. A. More et J. Oro. 1977. Cyanamide mediated syntheses under plausible primitive earth conditions. III. Synthesis of peptides. *J Mol Evol* **10**:211-20.
- Ochman, H., J. G. Lawrence et E. A. Groisman. 2000. Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**:299-304.
- Ohnuma, S., K. Hirooka, H. Hemmi, C. Ishida, C. Ohto et T. Nishino. 1996. Conversion of product specificity of archaeobacterial geranylgeranyl-diphosphate synthase. Identification of essential amino acid residues for chain length determination of prenyltransferase reaction. *The Journal of biological chemistry* **271**:18831-7.
- Ohnuma, S., K. Hirooka, C. Ohto et T. Nishino. 1997. Conversion from archaeal geranylgeranyl diphosphate synthase to farnesyl diphosphate synthase. Two amino acids before the first aspartate-rich motif solely determine eukaryotic farnesyl diphosphate synthase activity. *The Journal of biological chemistry* **272**:5192-8.
- Ohnuma, S., K. Hirooka, N. Tsuruoka, M. Yano, C. Ohto, H. Nakane et T. Nishino. 1998. A pathway where polyprenyl diphosphate elongates in prenyltransferase. Insight into a common mechanism of chain length determination of prenyltransferases. *The Journal of biological chemistry* **273**:26705-13.
- Olendzenski, L. et J. P. Gogarten. 2009. Evolution of genes and organisms: the tree/web of life in light of horizontal gene transfer. *Ann N Y Acad Sci* **1178**:137-45.
- Olsen, G. J. et C. R. Woese. 1996. Lessons from an Archaeal genome: what are we learning from *Methanococcus jannaschii*? *Trends Genet* **12**:377-9.
- Olsen, G. J., C. R. Woese et R. Overbeek. 1994. The winds of (evolutionary) change: breathing new life into microbiology. *J Bacteriol* **176**:1-6.
- Olsen, J. G., A. V. Rasmussen, P. von Wettstein-Knowles et A. Henriksen. 2004. Structure of the mitochondrial beta-ketoacyl-[acyl carrier protein] synthase from *Arabidopsis* and its role in fatty acid synthesis. *FEBS letters* **577**:170-4.
- Oparin, A. I. 1976. Evolution of the concepts of the origin of life, 1924-1974. *Orig Life* **7**:3-8.
- Oparin, A. I. 1938. *Origin of life*. Dover publications, New York.
- Oparin, A. I. 1924. The origin of life (Proiskhozhenie Zhizni) in J. D. Bernal, ed. *The Origin of Life*. The Weidenfeld and Nicolson Natural History. Weidenfeld and Nicolson, London.
- Oro, J. 1961. Comets and the formation of biochemical compounds on the primitive earth. *Nature* **190**:389-90.
- Ourisson, G. et Y. Nakatani. 1994. The terpenoid theory of the origin of cellular life: the evolution of terpenoids to cholesterol. *Chem Biol* **1**:11-23.
- Ouzounis, C. A., V. Kunin, N. Darzentas et L. Goldovsky. 2006. A minimal estimate for the gene content of the last universal common ancestor--exobiology from a terrestrial perspective. *Research in microbiology* **157**:57-68.
- Overton, E. 1899. The probable origin and physiological significance of cellular osmotic properties (Ueber die allgemeinen osmotischen Eigenschaften der Zelle, ihre vermutlichen Ursachen und ihre Bedeutung für die Physiologie) in D. Branton et R. B. Park, eds. *Papers on biological membrane structure*. Little, Brown and Company, Boston.
- Oyaizu, H., B. Debrunner-Vossbrinck, L. Mandelco, J. A. Studier et C. R. Woese. 1987. The green non-sulfur bacteria: a deep branching in the eubacterial line of descent. *Syst Appl Microbiol* **9**:47-53.

- Pace, N. R. 1991. Origin of life--facing up to the physical setting. *Cell* **65**:531-3.
- Pal, C., B. Papp et M. J. Lercher. 2005. Adaptive evolution of bacterial metabolic networks by horizontal gene transfer. *Nat Genet* **37**:1372-5.
- Palmer, T. et B. C. Berks. 2012. The twin-arginine translocation (Tat) protein export pathway. *Nature reviews* **10**:483-96.
- Paoletti, L., Y. J. Lu, G. E. Schujman, D. de Mendoza et C. O. Rock. 2007. Coupling of fatty acid and phospholipid synthesis in *Bacillus subtilis*. *Journal of bacteriology* **189**:5816-24.
- Parish, T., G. Roberts, F. Laval, M. Schaeffer, M. Daffe et K. Duncan. 2007. Functional complementation of the essential gene *fabG1* of *Mycobacterium tuberculosis* by *Mycobacterium smegmatis* *fabG* but not *Escherichia coli* *fabG*. *Journal of bacteriology* **189**:3721-8.
- Park, R. B. 1965. Substructure of chloroplast lamellae. *J Cell Biol* **27**:151-61.
- Park, R. B. et N. G. Pon. 1961. Correlation of structure with function in *Spinacea oleracea* chloroplasts. *Journal of molecular biology* **3**:1-10.
- Park, S. J. et S. Y. Lee. 2003. Identification and characterization of a new enoyl coenzyme A hydratase involved in biosynthesis of medium-chain-length polyhydroxyalkanoates in recombinant *Escherichia coli*. *Journal of bacteriology* **185**:5391-7.
- Pascal, R., H. Martin, M. Gargaud, P. López-García et T. Montmerle. 2009. Le soleil, la Terre... la vie. La quête des origines. Belin.
- Paula, S., A. G. Volkov, A. N. Van Hoek, T. H. Haines et D. W. Deamer. 1996. Permeation of protons, potassium ions, and small polar molecules through phospholipid bilayers as a function of membrane thickness. *Biophys J* **70**:339-48.
- Payandeh, J., M. Fujihashi, W. Gillon et E. F. Pai. 2006. The crystal structure of (S)-3-O-geranylgeranylgeranyl glyceryl phosphate synthase reveals an ancient fold for an ancient enzyme. *The Journal of biological chemistry* **281**:6070-8.
- Penn, O., E. Privman, H. Ashkenazy, G. Landan, D. Graur et T. Pupko. 2010. GUIDANCE: a web server for assessing alignment confidence scores. *Nucleic acids research* **38**:W23-8.
- Penny, D., M. D. Hendy et A. M. Poole. 2003. Testing fundamental evolutionary hypotheses. *Journal of theoretical biology* **223**:377-85.
- Penny, D. et A. Poole. 1999. The nature of the last universal common ancestor. *Curr Opin Genet Dev* **9**:672-7.
- Perepelov, A. V., Q. Wang, S. N. Senchenkova, S. D. Shevelev, G. Zhao, A. S. Shashkov, L. Feng, Y. A. Knirel et L. Wang. 2006. Structure of a teichoic acid-like O-polysaccharide of *Escherichia coli* O29. *Carbohydrate research* **341**:2176-80.
- Peretó, J. 2005. Controversies on the origin of life. *Int Microbiol* **8**:23-31.
- Peretó, J., J. L. Bada et A. Lazcano. 2009. Charles Darwin and the origin of life. *Orig Life Evol Biosph* **39**:395-406.
- Peretó, J., P. López-García et D. Moreira. 2005. Phylogenetic analysis of eukaryotic thiolases suggests multiple proteobacterial origins. *Journal of molecular evolution* **61**:65-74.
- Peretó, J., P. López-García et D. Moreira. 2004. Ancestral lipid biosynthesis and early membrane evolution. *Trends Biochem Sci* **29**:469-77.
- Philippe, H. 1993. MUST, a computer package of Management Utilities for Sequences and Trees. *Nucleic acids research* **21**:5264-72.

- Philippe, H. 2003. L'arbre universel du vivant : du simple au complexe ou du complexe au simple *in* M. Gargaud, D. Despois, J. P. Parisot et J. Reisse, eds. Les traces du vivant. Presses universitaires de Bordeaux, Bordeaux.
- Philippe, H. et P. Forterre. 1999. The rooting of the universal tree of life is not reliable. *J Mol Evol* **49**:509-23.
- Philippe, H. et J. Laurent. 1998. How good are deep phylogenetic trees? *Curr Opin Genet Dev* **8**:616-23.
- Philippe, H., P. López, H. Brinkmann, K. Budin, A. Germot, J. Laurent, D. Moreira, M. Muller et H. Le Guyader. 2000. Early-branching or fast-evolving eukaryotes? An answer based on slowly evolving positions. *Proc Biol Sci* **267**:1213-21.
- Phillips, M. A., P. Leon, A. Boronat et M. Rodriguez-Concepcion. 2008. The plastidial MEP pathway: unified nomenclature and resources. *Trends in plant science* **13**:619-23.
- Pidugu, L. S., K. Maity, K. Ramaswamy, N. Surolia et K. Suguna. 2009. Analysis of proteins with the 'hot dog' fold: prediction of function and identification of catalytic residues of hypothetical proteins. *BMC structural biology* **9**:37.
- Pillai, S., C. Rajagopal, M. Kapoor, G. Kumar, A. Gupta et N. Surolia. 2003. Functional characterization of beta-ketoacyl-ACP reductase (FabG) from *Plasmodium falciparum*. *Biochemical and biophysical research communications* **303**:387-92.
- Pinto da Silva, P. et D. Branton. 1970. MEMBRANE SPLITTING IN FREEZE-ETCHING Covalently Bound Ferritin as a Membrane Marker. *The Journal of Cell Biology* **45**:598-605.
- Pizzarello, S. et E. Shock. 2010. The organic composition of carbonaceous meteorites: the evolutionary story ahead of biochemistry. *Cold Spring Harbor perspectives in biology* **2**:a002105.
- Plobeck, N., S. Eifler, A. Brisson, Y. Nakatani et G. Ourisson. 1992. Sodium di-polyprenyl phosphates form "primitive" membranes. *Tetrahedron* **33**:5249-52.
- Ploskon, E., C. J. Arthur, A. L. Kanari, P. Wattana-amorn, C. Williams, J. Crosby, T. J. Simpson, C. L. Willis et M. P. Crump. 2010. Recognition of intermediate functionality by acyl carrier protein over a complete cycle of fatty acid biosynthesis. *Chemistry & biology* **17**:776-85.
- Plowe, J. Q. 1931. Membranes in the plant cell. Morphological membranes at protoplasmic surfaces. *Protoplasma* **12**:196-221.
- Pohorille, A., M. A. Wilson et C. Chipot. 2003. Membrane peptides and their role in protobiological evolution. *Orig Life Evol Biosph* **33**:173-97.
- Poole, A. M. et D. T. Logan. 2005. Modern mRNA proofreading and repair: clues that the last universal common ancestor possessed an RNA genome? *Molecular biology and evolution* **22**:1444-55.
- Popa, O. et T. Dagan. 2011. Trends and barriers to lateral gene transfer in prokaryotes. *Current opinion in microbiology* **14**:615-23.
- Price, A. C., Y. M. Zhang, C. O. Rock et S. W. White. 2001. Structure of beta-ketoacyl-[acyl carrier protein] reductase from *Escherichia coli*: negative cooperativity and its structural basis. *Biochemistry* **40**:12772-81.
- Pugh, E. L. et M. Kates. 1994. Acylation of proteins of the archaeobacteria *Halobacterium cutirubrum* and *Methanobacterium thermoautotrophicum*. *Biochimica et biophysica acta* **1196**:38-44.
- Pugh, E. L., M. K. Wassef et M. Kates. 1971. Inhibition of fatty acid synthetase in *Halobacterium cutirubrum* and *Escherichia coli* by high salt concentrations. *Canadian journal of biochemistry* **49**:953-8.

- Qin, Y. M., A. M. Haapalainen, S. H. Kilpelainen, M. S. Marttila, M. K. Koski, T. Glumoff, D. K. Novikov et J. K. Hiltunen. 2000. Human peroxisomal multifunctional enzyme type 2. Site-directed mutagenesis studies show the importance of two protic residues for 2-enoyl-CoA hydratase 2 activity. *The Journal of biological chemistry* **275**:4965-72.
- Rahman, O., M. Pfitzenmaier, O. Pester, S. Morath, S. P. Cummings, T. Hartung et I. C. Sutcliffe. 2009. Macroamphiphilic components of thermophilic actinomycetes: identification of lipoteichoic acid in *Thermobifida fusca*. *Journal of bacteriology* **191**:152-60.
- Rajamani, S., A. Vlassov, S. Benner, A. Coombs, F. Olasagasti et D. Deamer. 2008. Lipid-assisted synthesis of RNA-like polymers from mononucleotides. *Orig Life Evol Biosph* **38**:57-74.
- Rambaut, A. et A. Drummond. 2003. Tracer: a program for analysing results from Bayesian MCMC programs such as BEAST&MrBayes. *Distributed by the authors. Edinburgh (United Kingdom) Institute of Evolutionary Biology, University of Edinburgh. Available from: <http://tree.bio.ed.ac.uk/software/tracer/>.*
- Raspail, F. V. 1837. Nouveau système de physiologie végétale et de botanique. J. B. Baillière, Paris.
- Rayleigh, L. 1889. Measurements of the Amount of Oil Necessary in Order to Check the Motions of Camphor upon Water. *Proceedings of the Royal Society of London* **47**:364-67.
- Rayleigh, L. 1891. Surface Tension. *Nature* **43**:437-39.
- Raymond, J., O. Zhaxybayeva, J. P. Gogarten, S. Y. Gerdes et R. E. Blankenship. 2002. Whole-genome analysis of photosynthetic prokaryotes. *Science* **298**:1616-20.
- Reaney, D. C. 1974. On the origin of prokaryotes. *J Theor Biol* **48**:243-51.
- Reardon, D. et G. K. Farber. 1995. The structure and evolution of alpha/beta barrel proteins. *Faseb J* **9**:497-503.
- Rehm, B. H. et A. Steinbuchel. 1999. Biochemical and genetic analysis of PHA synthases and other proteins required for PHA synthesis. *International journal of biological macromolecules* **25**:3-19.
- Reiser, S. E., T. A. Mitsky et K. J. Gruys. 2000. Characterization and cloning of an (R)-specific trans-2,3-enoylacyl-CoA hydratase from *Rhodospirillum rubrum* and use of this enzyme for PHA production in *Escherichia coli*. *Applied microbiology and biotechnology* **53**:209-18.
- Ren, Q., N. Sierro, B. Witholt et B. Kessler. 2000. FabG, an NADPH-dependent 3-ketoacyl reductase of *Pseudomonas aeruginosa*, provides precursors for medium-chain-length poly-3-hydroxyalkanoate biosynthesis in *Escherichia coli*. *Journal of bacteriology* **182**:2978-81.
- Reusch, R. N. 2000. Transmembrane ion transport by polyphosphate/poly-(R)-3-hydroxybutyrate complexes. *Biochemistry (Mosc)* **65**:280-95.
- Richardson, S. H., H. O. Hultin et D. E. Green. 1963. Structural Proteins of Membrane Systems. *Proceedings of the National Academy of Sciences of the United States of America* **50**:821-7.
- Rivera, M. C. et J. A. Lake. 1992. Evidence that eukaryotes and eocyte prokaryotes are immediate relatives. *Science* **257**:74-6.
- Rivera, M. C. et J. A. Lake. 2004. The ring of life provides evidence for a genome fusion origin of eukaryotes. *Nature* **431**:152-5.
- Le Petit Robert. 1984. par Paul Robert, rédaction dirigée par A. Rey et J. Rey-Debove, eds, Paris.
- Robert, F. et M. Chaussidon. 2006. A palaeotemperature curve for the Precambrian oceans based on silicon isotopes in cherts. *Nature* **443**:969-72.
- Robertson, J. D. 1959. The ultrastructure of cell membranes and their derivatives. *Biochemical Society symposium* **16**:3-43.

- Robertson, J. D. 1957. New observations on the ultrastructure of the membranes of frog peripheral nerve fibers. *J Biophys Biochem Cytol* **3**:1043-8.
- Robertson, J. D. 1960. The molecular structure and contact relationships of cell membranes. *Prog Biophys Mol Biol* **10**:343-418.
- Robinow, C. F. 1962. On the plasma membrane of some bacteria and fungi. *Circulation* **26**:1092-104.
- Rodríguez-Concepción, M. et A. Boronat. 2002. Elucidation of the methylerythritol phosphate pathway for isoprenoid biosynthesis in bacteria and plastids. A metabolic milestone achieved through genomics. *Plant Physiol* **130**:1079-89.
- Rohdich, F., S. Hecht, K. Gartner, P. Adam, C. Krieger, S. Amslinger, D. Arigoni, A. Bacher et W. Eisenreich. 2002. Studies on the nonmevalonate terpene biosynthetic pathway: metabolic role of IspH (LytB) protein. *Proceedings of the National Academy of Sciences of the United States of America* **99**:1158-63.
- Rohdich, F., F. Zepeck, P. Adam *et al.* 2003. The deoxyxylulose phosphate pathway of isoprenoid biosynthesis: studies on the mechanisms of the reactions catalyzed by IspG and IspH protein. *Proceedings of the National Academy of Sciences of the United States of America* **100**:1586-91.
- Rohmer, M. 1999. The discovery of a mevalonate-independent pathway for isoprenoid biosynthesis in bacteria, algae and higher plants. *Natural product reports* **16**:565-74.
- Romanowski, M. J., J. B. Bonanno et S. K. Burley. 2002. Crystal structure of the Streptococcus pneumoniae phosphomevalonate kinase, a member of the GHMP kinase superfamily. *Proteins* **47**:568-71.
- Ronquist, F. et J. P. Huelsenbeck. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics (Oxford, England)* **19**:1572-4.
- Rostand, J. 1943. La genèse de la vie. Histoire des idées sur la génération spontanée. Hachette, Paris.
- Rozwarski, D. A., G. A. Grant, D. H. Barton, W. R. Jacobs, Jr. et J. C. Sacchettini. 1998. Modification of the NADH of the isoniazid target (InhA) from Mycobacterium tuberculosis. *Science (New York, N.Y)* **279**:98-102.
- Ruepp, A., W. Graml, M. L. Santos-Martinez, K. K. Koretke, C. Volker, H. W. Mewes, D. Frishman, S. Stocker, A. N. Lupas et W. Baumeister. 2000. The genome sequence of the thermoacidophilic scavenger Thermoplasma acidophilum. *Nature* **407**:508-13.
- Ruse, M. 1996. Monad to man. Harvard university press, Cambridge, MA.
- Rushdi, A. I. et B. R. Simoneit. 2006. Abiotic condensation synthesis of glyceride lipids and wax esters under simulated hydrothermal conditions. *Orig Life Evol Biosph* **36**:93-108.
- Rushdi, A. I. et B. R. Simoneit. 2001. Lipid formation by aqueous Fischer-Tropsch-type synthesis over a temperature range of 100 to 400 degrees C. *Orig Life Evol Biosph* **31**:103-18.
- Russell, M. J. et A. J. Hall. 1997. The emergence of life from iron monosulphide bubbles at a submarine hydrothermal redox and pH front. *J Geol Soc London* **154**:377-402.
- Rutters, H., H. Sass, H. Cypionka et J. Rullkotter. 2001. Monoalkylether phospholipids in the sulfate-reducing bacteria Desulfosarcina variabilis and Desulforhabdus amnigenus. *Archives of microbiology* **176**:435-42.
- Ruzheinikov, S. N., J. Burke, S. Sedelnikova, P. J. Baker, R. Taylor, P. A. Bullough, N. M. Muir, M. G. Gore et D. W. Rice. 2001. Glycerol dehydrogenase. structure, specificity, and mechanism of a family III polyol dehydrogenase. *Structure* **9**:789-802.
- Ruzicka, L. 1953. The isoprene rule and the biogenesis of terpenic compounds. *Experientia* **9**:357-67.

- Sacco, E., A. S. Covarrubias, H. M. O'Hare, P. Carroll, N. Eynard, T. A. Jones, T. Parish, M. Daffe, K. Backbro et A. Quemard. 2007a. The missing piece of the type II fatty acid synthase system from *Mycobacterium tuberculosis*. *Proceedings of the National Academy of Sciences of the United States of America* **104**:14628-33.
- Sacco, E., V. Legendre, F. Laval, D. Zerbib, H. Montrozier, N. Eynard, C. Guilhot, M. Daffe et A. Quemard. 2007b. Rv3389C from *Mycobacterium tuberculosis*, a member of the (R)-specific hydratase/dehydratase family. *Biochimica et biophysica acta* **1774**:303-11.
- Sacco, E., N. Slama, K. Backbro, T. Parish, F. Laval, M. Daffe, N. Eynard et A. Quemard. 2010. Revisiting the assignment of Rv0241c to fatty acid synthase type II of *Mycobacterium tuberculosis*. *Journal of bacteriology* **192**:4037-44.
- Sacerdote, M. G. et J. W. Szostak. 2005. Semipermeable lipid bilayers exhibit diastereoselectivity favoring ribose. *Proceedings of the National Academy of Sciences of the United States of America* **102**:6004-8.
- Saitou, N. et M. Nei. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular biology and evolution* **4**:406-25.
- Sapp, J. 2005. The prokaryote-eukaryote dichotomy: meanings and mythology. *Microbiol Mol Biol Rev* **69**:292-305.
- Sato, S., M. Murakami, T. Yoshimura et H. Hemmi. 2008. Specific partial reduction of geranylgeranyl diphosphate by an enzyme from the thermoacidophilic archaeon *Sulfolobus acidocaldarius* yields a reactive prenyl donor, not a dead-end product. *Journal of bacteriology* **190**:3923-9.
- Schleiden. 1837. Contributions to phytogenesis. Sydenham society, London.
- Schliep, K., P. López, F. J. Lapointe et E. Baptiste. 2010. Harvesting evolutionary signals in a forest of prokaryotic gene trees. *Molecular biology and evolution* **28**:1393-405.
- Schmidt, W. J. 1936. Birefringence and fine structure of nerve myelin (Doppelbrechung und Feinbau der markscheide der nervenfaser) in D. Branton et R. B. Park, eds. Papers on biological membrane structure. Little, Brown and Company, Boston.
- Schmitt, F. O., R. S. Bear et G. L. Clark. 1935. X-ray diffraction studies on nerve. *Radiology* **25**:131-51.
- Schneider, R., B. Brors, F. Burger, S. Camrath et H. Weiss. 1997. Two genes of the putative mitochondrial fatty acid synthase in the genome of *Saccharomyces cerevisiae*. *Current genetics* **32**:384-8.
- Schoepp-Cothenet, B., R. van Lis, A. Atteia, F. Baymann, L. Capowiez, A. L. Ducluzeau, S. Duval, F. Ten Brink, M. J. Russell et W. Nitschke. 2012. On the universal core of bioenergetics. *Biochim Biophys Acta*.
- Schrum, J. P., T. F. Zhu et J. W. Szostak. 2010. The origins of cellular life. *Cold Spring Harbor perspectives in biology* **2**:a002212.
- Schwann, T. 1839. Microscopical researches into the accordance in the structure and growth of animals and plants. Sydenham society, London.
- Schwartz, R. M. et M. O. Dayhoff. 1978. Origins of prokaryotes, eukaryotes, mitochondria, and chloroplasts. *Science* **199**:395-403.
- Schweizer, E. et J. Hofmann. 2004. Microbial type I fatty acid synthases (FAS): major players in a network of cellular FAS systems. *Microbiol Mol Biol Rev* **68**:501-17.
- Segre, D., D. Ben-Eli, D. W. Deamer et D. Lancet. 2001. The lipid world. *Orig Life Evol Biosph* **31**:119-45.

- Sherwood, E., A. Joshi et J. Oro. 1977. Cyanamide mediated syntheses under plausible primitive earth conditions. II. The polymerization of deoxythymidine 5'-triphosphate. *J Mol Evol* **10**:193-209.
- Sherwood, E. et J. Oro. 1977. Cyanamide mediated syntheses under plausible primitive earth conditions. I. The synthesis of P1, P2 dideoxythymidine 5'-pyrophosphate. *J Mol Evol* **10**:183-92.
- Shimada, H. et A. Yamagishi. 2011. Stability of heterochiral hybrid membrane made of bacterial sn-G3P lipids and archaeal sn-G1P lipids. *Biochemistry* **50**:4114-20.
- Shimizu, H., S. Yokobori, T. Ohkuri, T. Yokogawa, K. Nishikawa et A. Yamagishi. 2007. Extremely thermophilic translation system in the common ancestor commonote: ancestral mutants of Glycyl-tRNA synthetase from the extreme thermophile *Thermus thermophilus*. *J Mol Biol* **369**:1060-9.
- Simoneit, B. R., A. I. Rushdi et D. Deamer. 2007. Abiotic formation of acylglycerols under simulated hydrothermal conditions and self-assembly properties of such lipid products. *Advance in Space Research* **40**:1649-56.
- Simons, K. et E. Ikonen. 1997. Functional rafts in cell membranes. *Nature* **387**:569-72.
- Singer, S. J. 2004. Some early history of membrane molecular biology. *Annu Rev Physiol* **66**:1-27.
- Singer, S. J. et G. L. Nicolson. 1972. The fluid mosaic model of the structure of cell membranes. *Science (New York, N.Y)* **175**:720-31.
- Sinninghe Damsté, J. S., W. I. Rijpstra, E. C. Hopmans, S. Schouten, M. Balk et A. J. Stams. 2007. Structural characterization of diabolic acid-based tetraester, tetraether and mixed ether/ester, membrane-spanning lipids of bacteria from the order Thermotogales. *Archives of microbiology* **188**:629-41.
- Sinninghe Damsté, J. S., M. Strous, W. I. Rijpstra, E. C. Hopmans, J. A. Geenevasen, A. C. van Duin, L. A. van Niftrik et M. S. Jetten. 2002. Linearly concatenated cyclobutane lipids form a dense bacterial membrane. *Nature* **419**:708-12.
- Skophammer, R. G., C. W. Herbold, M. C. Rivera, J. A. Servin et J. A. Lake. 2006. Evidence that the root of the tree of life is not within the Archaea. *Molecular biology and evolution* **23**:1648-51.
- Smit, A. et A. Mushegian. 2000. Biosynthesis of isoprenoids via mevalonate in Archaea: the lost pathway. *Genome research* **10**:1468-84.
- Smith, H. W. 1962. The plasma membrane, with notes on the history of botany. *Circulation* **26**:987-1012.
- Smith, J. L. et D. H. Sherman. 2008. Biochemistry. An enzyme assembly line. *Science (New York, N.Y)* **321**:1304-5.
- Smith, S. et S. C. Tsai. 2007. The type I fatty acid and polyketide synthases: a tale of two megasynthases. *Natural product reports* **24**:1041-72.
- Snel, B., P. Bork et M. A. Huynen. 1999. Genome phylogeny based on gene content. *Nat Genet* **21**:108-10.
- Snel, B., P. Bork et M. A. Huynen. 2002. Genomes in flux: the evolution of archaeal and proteobacterial gene content. *Genome Res* **12**:17-25.
- Sober, E. et M. Steel. 2002. Testing the hypothesis of common ancestry. *Journal of theoretical biology* **218**:395-408.
- Soderberg, T., A. Chen et C. D. Poulter. 2001. Geranylgeranylglyceryl phosphate synthase. Characterization of the recombinant enzyme from *Methanobacterium thermoautotrophicum*. *Biochemistry* **40**:14847-54.

- Sogin, S. J., M. L. Sogin et C. R. Woese. 1971. Phylogenetic measurement in procaryotes by primary structural characterization. *J Mol Evol* **1**:173-84.
- Sohlenkamp, C., I. M. López-Lara et O. Geiger. 2003. Biosynthesis of phosphatidylcholine in bacteria. *Progress in lipid research* **42**:115-62.
- Sorek, R., Y. Zhu, C. J. Creevey, M. P. Francino, P. Bork et E. M. Rubin. 2007. Genome-wide experimental determination of barriers to horizontal gene transfer. *Science* **318**:1449-52.
- Sousa, F. L., R. J. Alves, M. A. Ribeiro, J. B. Pereira-Leal, M. Teixeira et M. M. Pereira. 2012. The superfamily of heme-copper oxygen reductases: types and evolutionary considerations. *Biochimica et biophysica acta* **1817**:629-37.
- Sparrow, C. P. et C. R. Raetz. 1985. Purification and properties of the membrane-bound CDP-diglyceride synthetase from *Escherichia coli*. *The Journal of biological chemistry* **260**:12084-91.
- Stanier, R. Y. et C. B. Van Niel. 1962. The concept of a bacterium. *Arch Mikrobiol* **42**:17-35.
- Staunton, J. et K. J. Weissman. 2001. Polyketide biosynthesis: a millennium review. *Natural product reports* **18**:380-416.
- Stephens, R. S., S. Kalman, C. Lammel *et al.* 1998. Genome sequence of an obligate intracellular pathogen of humans: *Chlamydia trachomatis*. *Science (New York, N.Y)* **282**:754-9.
- Stetter, K. O. 1996. Hyperthermophiles in the history of life in G. R. Bock et J. A. Goode, eds. *Evolution of hydrothermal ecosystems on earth (and mars?)*. Wiley
- Steussy, C. N., A. D. Robison, A. M. Tetrack, J. T. Knight, V. W. Rodwell, C. V. Stauffacher et A. L. Sutherland. 2006. A structural limitation on enzyme activity: the case of HMG-CoA synthase. *Biochemistry* **45**:14407-14.
- Stoeckenius, W. 1962. Structure of the plasma membrane. An electron-microscope study. *Circulation* **26**:1066-9.
- Stoltzfus, A., D. F. Spencer, M. Zuker, J. M. Logsdon, Jr. et W. F. Doolittle. 1994. Testing the exon theory of genes: the evidence from protein structure. *Science* **265**:202-7.
- Streiff, S., N. Ribeiro, Z. Wu, E. Gumienna-Kontecka, M. Elhabiri, A. M. Albrecht-Gary, G. Ourisson et Y. Nakatani. 2007. "Primitive" membrane from polyprenyl phosphates and polyprenyl alcohols. *Chem Biol* **14**:313-9.
- Stubbe, J. 2000. Ribonucleotide reductases: the link between an RNA and a DNA world? *Curr Opin Struct Biol* **10**:731-6.
- Studier, M. H. 1968. Origin of organic matter in early solar system - I. Hydrocarbons. *Geochimica et Cosmochimica Acta* **32**:151-73.
- Stuible, H. P., S. Meier, C. Wagner, E. Hannappel et E. Schweizer. 1998. A novel phosphopantetheine:protein transferase activating yeast mitochondrial acyl carrier protein. *The Journal of biological chemistry* **273**:22334-9.
- Sugino, H., M. Sasaki, H. Azakami, M. Yamashita et Y. Murooka. 1992. A monoamine-regulated *Klebsiella aerogenes* operon containing the monoamine oxidase structural gene (*maoA*) and the *maoC* gene. *Journal of bacteriology* **174**:2485-92.
- Summers, R. G., A. Ali, B. Shen, W. A. Wessel et C. R. Hutchinson. 1995. Malonyl-coenzyme A:acyl carrier protein acyltransferase of *Streptomyces glaucescens*: a possible link between fatty acid and polyketide biosynthesis. *Biochemistry* **34**:9389-402.
- Swarnamukhi, P. L., S. K. Sharma, P. Bajaj, N. Surolia, A. Surolia et K. Suguna. 2006. Crystal structure of dimeric FabZ of *Plasmodium falciparum* reveals conformational switching to active hexamers by peptide flips. *FEBS letters* **580**:2653-60.

- Syvänen, M. 1985. Cross-species gene transfer; implications for a new theory of evolution. *J Theor Biol* **112**:333-43.
- Szabó, P., I. Scheuring, T. Czárán et E. Szathmáry. 2002. In silico simulations reveal that replicators with limited dispersal evolve towards higher efficiency and fidelity. *Nature* **420**:340-43.
- Szathmáry, E. 2007. Coevolution of metabolic networks and membranes: the scenario of progressive sequestration. *Philos Trans R Soc Lond B Biol Sci* **362**:1781-7.
- Szostak, J. W. 2011. An optimal degree of physical and chemical heterogeneity for the origin of life? *Philosophical transactions of the Royal Society of London* **366**:2894-901.
- Szostak, J. W., D. P. Bartel et P. L. Luisi. 2001. Synthesizing life. *Nature* **409**:387-90.
- Tachibana, A., Y. Yano, S. Otani, N. Nomura, Y. Sako et M. Taniguchi. 2000. Novel prenyltransferase gene encoding farnesylgeranyl diphosphate synthase from a hyperthermophilic archaeon, *Aeropyrum pernix*. Molecular evolution with alteration in product specificity. *European journal of biochemistry / FEBS* **267**:321-8.
- Taguchi, K., Y. Aoyagi, H. Matsusaki, T. Fukui et Y. Doi. 1999. Co-expression of 3-ketoacyl-ACP reductase and polyhydroxyalkanoate synthase genes induces PHA production in *Escherichia coli* HB101 strain. *FEMS microbiology letters* **176**:183-90.
- Takakura, K. et T. Sugawara. 2004. Membrane dynamics of a myelin-like giant multilamellar vesicle applicable to a self-reproducing system. *Langmuir* **20**:3832-4.
- Tan, H. H., A. Makino, K. Sudesh, P. Greimel et T. Kobayashi. 2012. Spectroscopic evidence for the unusual stereochemical configuration of an endosome-specific lipid. *Angewandte Chemie (International ed)* **51**:533-5.
- Tanaka, R., U. Oster, E. Kruse, W. Rudiger et B. Grimm. 1999. Reduced activity of geranylgeranyl reductase leads to loss of chlorophyll and tocopherol and to partially geranylgeranylated chlorophyll in transgenic tobacco plants expressing antisense RNA for geranylgeranyl reductase. *Plant physiology* **120**:695-704.
- Tanford, C. 1973. The hydrophobic effect: formation of micelles and biological membranes. Wiley, New York.
- Tanford, C. 2004. Ben Franklin stilled the waves. Oxford university press, New York.
- Taron, D. J., W. C. Childs, 3rd et F. C. Neuhaus. 1983. Biosynthesis of D-alanyl-lipoteichoic acid: role of diglyceride kinase in the synthesis of phosphatidylglycerol for chain elongation. *Journal of bacteriology* **154**:1110-6.
- Taroncher-Oldenburg, G., K. Nishina et G. Stephanopoulos. 2000. Identification and analysis of the polyhydroxyalkanoate-specific beta-ketothiolase and acetoacetyl coenzyme A reductase genes in the cyanobacterium *Synechocystis* sp. strain PCC6803. *Applied and environmental microbiology* **66**:4440-8.
- Tarshis, L. C., P. J. Proteau, B. A. Kellogg, J. C. Sacchettini et C. D. Poulter. 1996. Regulation of product chain length by isoprenyl diphosphate synthases. *Proceedings of the National Academy of Sciences of the United States of America* **93**:15018-23.
- Tatusov, R. L., E. V. Koonin et D. J. Lipman. 1997. A genomic perspective on protein families. *Science* **278**:631-7.
- Tawfik, D. S. et A. D. Griffiths. 1998. Man-made cell-like compartments for molecular evolution. *Nat Biotechnol* **16**:652-6.
- Tekaia, F., A. Lazcano et B. Dujon. 1999. The genomic tree as revealed from whole proteome comparisons. *Genome Res* **9**:550-7.

- Thaddeus, P. 2006. The prebiotic molecules observed in the interstellar gas. *Philosophical transactions of the Royal Society of London* **361**:1681-7.
- Theobald, D. L. 2010. A formal test of the theory of universal common ancestry. *Nature* **465**:219-22.
- Thomas, J. A. et F. R. Rana. 2007. The influence of environmental conditions, lipid composition, and phase behavior on the origin of cell membranes. *Orig Life Evol Biosph* **37**:267-85.
- Tornabene, T. G. et T. A. Langworthy. 1979. Diphytanyl and dibiphytanyl glycerol ether lipids of methanogenic archaeobacteria. *Science (New York, N.Y)* **203**:51-3.
- Tornabene, T. G., R. S. Wolfe, W. E. Balch, G. Holzer, G. E. Fox et J. Oro. 1978. Phytanyl-glycerol ethers and squalenes in the archaeobacterium *Methanobacterium thermoautotrophicum*. *Journal of molecular evolution* **11**:259-66.
- Treharne, K. J., E. I. Mercer et T. W. Goodwin. 1966. Incorporation of [14C] carbon dioxide and [2-14C] mevalonic acid into terpenoids of higher plants during chloroplast development. *The Biochemical journal* **99**:239-45.
- Tsuge, T., T. Hisano, S. Taguchi et Y. Doi. 2003a. Alteration of chain length substrate specificity of *Aeromonas caviae* R-enantiomer-specific enoyl-coenzyme A hydratase through site-directed mutagenesis. *Applied and environmental microbiology* **69**:4830-6.
- Tsuge, T., K. Taguchi, T. Seiichi et Y. Doi. 2003b. Molecular characterization and properties of (R)-specific enoyl-CoA hydratases from *Pseudomonas aeruginosa*: metabolic tools for synthesis of polyhydroxyalkanoates via fatty acid beta-oxidation. *International journal of biological macromolecules* **31**:195-205.
- Tsukahara, H., E. Imai, H. Honda, K. Hatori et K. Matsuno. 2002. Prebiotic oligomerization on or inside lipid vesicles in hydrothermal environments. *Orig Life Evol Biosph* **32**:13-21.
- Tucci, S. et W. Martin. 2007. A novel prokaryotic trans-2-enoyl-CoA reductase from the spirochete *Treponema denticola*. *FEBS letters* **581**:1561-6.
- Turner, W. 1890. The Cell Theory, Past and Present. *J Anat Physiol* **24**:253-87.
- Uragami, M., Y. Miyake et S. L. Regen. 2000. Influence of headgroup chirality on the mixing behaviour of phosphatidylglycerol mimics in fluid bilayers. *Langmuir* **16**:3491-96.
- Van Wyhe, J. 2002. The Complete Work of Charles Darwin Online (<http://darwin-online.org.uk>).
- Vignais, P. M., B. Billoud et J. Meyer. 2001. Classification and phylogeny of hydrogenases. *FEMS microbiology reviews* **25**:455-501.
- Vlassov, A. 2005. How was membrane permeability produced in an RNA world? *Orig Life Evol Biosph* **35**:135-49.
- Vlassov, A., A. Khvorova et M. Yarus. 2001. Binding and disruption of phospholipid bilayers by supramolecular RNA complexes. *Proceedings of the National Academy of Sciences of the United States of America* **98**:7706-11.
- Vollhardt, D., G. Emrich, T. Gutberlet et J. H. Fuhrhop. 1996. Chiral discrimination and pattern formation in N-Dodecylmannonamide monolayers at the air-water interface. *Langmuir* **12**:5659-63.
- Voynova, N. E., S. E. Rios et H. M. Miziorko. 2004. *Staphylococcus aureus* mevalonate kinase: isolation and characterization of an enzyme of the isoprenoid biosynthetic pathway. *Journal of bacteriology* **186**:61-7.
- Wächtershäuser, G. 2003. From pre-cells to Eukarya--a tale of two lipids. *Molecular microbiology* **47**:13-22.

- Wächtershäuser, G. 2006. From volcanic origins of chemoautotrophic life to Bacteria, Archaea and Eukarya. *Philosophical transactions of the Royal Society of London* **361**:1787-806; discussion 806-8.
- Wächtershäuser, G. 1988. Before enzymes and templates: theory of surface metabolism. *Microbiol Rev* **52**:452-84.
- Wagner, M. et M. Horn. 2006. The Planctomycetes, Verrucomicrobia, Chlamydiae and sister phyla comprise a superphylum with biotechnological and medical relevance. *Current opinion in biotechnology* **17**:241-9.
- Wagner, M., A. J. Roger, J. L. Flax, G. A. Brusseau et D. A. Stahl. 1998. Phylogeny of dissimilatory sulfite reductases supports an early origin of sulfate respiration. *Journal of bacteriology* **180**:2975-82.
- Wakil, S. J. 1961. Mechanism of Fatty Acid Synthesis. *Journal of lipid research* **2**:1-22.
- Wakil, S. J. 1958. A malonic acid derivative as an intermediate in fatty acid synthesis. *Journal of the American Chemical Society* **80**:6465.
- Walde, P., A. Goto, P. A. Monnard, M. Wessicken et P. L. Luisi. 1994. Oparin's reactions revisited: enzymatic synthesis of poly(adenylic acid) in micelles and self-reproducing vesicles. *J. Am. Chem. Soc.* **116**:7541-47.
- Waldrop, M. M. 2008. Disputed definitions. Complexity. *Nature* **455**:1023-28.
- Wallach, D. F. et P. H. Zahler. 1966. Protein conformations in cellular membranes. *Proceedings of the National Academy of Sciences of the United States of America* **56**:1552-9.
- Walter, A. et J. Gutknecht. 1986. Permeability of small nonelectrolytes through lipid bilayer membranes. *J Membr Biol* **90**:207-17.
- Wang, H. et J. E. Cronan. 2004a. Functional replacement of the FabA and FabB proteins of Escherichia coli fatty acid synthesis by Enterococcus faecalis FabZ and FabF homologues. *The Journal of biological chemistry* **279**:34489-95.
- Wang, H. et J. E. Cronan. 2004b. Only one of the two annotated Lactococcus lactis fabG genes encodes a functional beta-ketoacyl-acyl carrier protein reductase. *Biochemistry* **43**:11782-9.
- Wang, H., J. Guo, H. Pang et X. Zhang. 2010. Crystallization and preliminary X-ray analysis of the MaoC-like dehydratase from Phytophthora capsici. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **66**:272-4.
- Weijers, J. W., S. Schouten, E. C. Hopmans, J. A. Geenevasen, O. R. David, J. M. Coleman, R. D. Pancost et J. S. Sinninghe Damste. 2006. Membrane lipids of mesophilic anaerobic bacteria thriving in peats have typical archaeal traits. *Environmental microbiology* **8**:648-57.
- Wesener, S. R., V. Y. Potharla et Y. Q. Cheng. 2011. Reconstitution of the FK228 biosynthetic pathway reveals cross talk between modular polyketide synthases and fatty acid synthase. *Applied and environmental microbiology* **77**:1501-7.
- Whereat, A. F., F. E. Hull et M. W. Orishimo. 1967. The role of succinate in the regulation of fatty acid synthesis by heart mitochondria. *The Journal of biological chemistry* **242**:4013-22.
- White, S. W., J. Zheng, Y. M. Zhang et Rock. 2005. The structural biology of type II fatty acid biosynthesis. *Annual review of biochemistry* **74**:791-831.
- Whittaker, R. H. 1969. New concepts of kingdoms or organisms. Evolutionary relations are better represented by new classifications than by the traditional two kingdoms. *Science (New York, N.Y)* **163**:150-60.
- Wick, R., P. Walde et P. L. Luisi. 1995. Light microscopic investigations of the autocatalytic self-reproduction of giant vesicles. *J Am Chem Soc* **117**:1435-36.

- Wilding, E. I., J. R. Brown, A. P. Bryant, A. F. Chalker, D. J. Holmes, K. A. Ingraham, S. Iordanescu, C. Y. So, M. Rosenberg et M. N. Gwynn. 2000. Identification, evolution, and essentiality of the mevalonate pathway for isopentenyl diphosphate biosynthesis in gram-positive cocci. *Journal of bacteriology* **182**:4319-27.
- Wilkins, A. S. 1996. Are there 'Kuhnian' revolutions in Biology? *Bioessays* **18**:695-96.
- Williams, P. D., D. D. Pollock, B. P. Blackburne et R. A. Goldstein. 2006. Assessing the accuracy of ancestral protein reconstruction methods. *PLoS Comput Biol* **2**:e69.
- Woese, C. 1998. The universal ancestor. *Proceedings of the National Academy of Sciences of the United States of America* **95**:6854-9.
- Woese, C. R. 1987. Bacterial evolution. *Microbiological reviews* **51**:221-71.
- Woese, C. R. 1967. The present status of the genetic code. *Progress in Nucleic Acid Research and Molecular Biology* **7**:107-72.
- Woese, C. R. 2000. Interpreting the universal phylogenetic tree. *Proceedings of the National Academy of Sciences of the United States of America* **97**:8392-6.
- Woese, C. R. 2002. On the evolution of cells. *Proceedings of the National Academy of Sciences of the United States of America* **99**:8742-7.
- Woese, C. R. et G. E. Fox. 1977a. The concept of cellular evolution. *J Mol Evol* **10**:1-6.
- Woese, C. R. et G. E. Fox. 1977b. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proceedings of the National Academy of Sciences of the United States of America* **74**:5088-90.
- Woese, C. R., O. Kandler et M. L. Wheelis. 1990. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proceedings of the National Academy of Sciences of the United States of America* **87**:4576-9.
- Woese, C. R., L. J. Magrum et G. E. Fox. 1978. Archaeobacteria. *Journal of molecular evolution* **11**:245-51.
- Wolf, Y. I., L. Aravind, N. V. Grishin et E. V. Koonin. 1999. Evolution of aminoacyl-tRNA synthetases--analysis of unique domain architectures and phylogenetic trees reveals a complex history of horizontal gene transfer events. *Genome research* **9**:689-710.
- Wong, J. T., J. Chen, W. K. Mat, S. K. Ng et H. Xue. 2007. Polyphasic evidence delineating the root of life and roots of biological domains. *Gene* **403**:39-52.
- Wright, L. D. 1961. Biosynthesis of isoprenoid compounds. *Annual Review of Biochemistry* **30**:525-48.
- Xue, H., K. L. Tong, C. Marck, H. Grosjean et J. T. Wong. 2003. Transfer RNA paralogs: evidence for genetic code-amino acid biosynthesis coevolution and an archaeal root of life. *Gene* **310**:59-66.
- Yahr, T. L. et W. T. Wickner. 2001. Functional reconstitution of bacterial Tat translocation in vitro. *The EMBO journal* **20**:2472-9.
- Yamamura, A., T. Ichimura, F. Mimoto, J. Ohtsuka, K. Miyazono, M. Okai, M. Kamo, W. C. Lee, K. Nagata et M. Tanokura. 2008. A unique catalytic triad revealed by the crystal structure of APE0912, a short-chain dehydrogenase/reductase family protein from *Aeropyrum pernix* K1. *Proteins* **70**:1640-5.
- Yang, D., L. W. Shipman, C. A. Roessner, A. I. Scott et J. C. Sacchettini. 2002. Structure of the *Methanococcus jannaschii* mevalonate kinase, a member of the GHMP kinase superfamily. *The Journal of biological chemistry* **277**:9462-7.

- Yen, M. R., K. T. Harley, Y. H. Tseng et M. H. Saier, Jr. 2001. Phylogenetic and structural analyses of the oxal family of protein translocases. *FEMS microbiology letters* **204**:223-31.
- Yen, M. R., Y. H. Tseng, E. H. Nguyen, L. F. Wu et M. H. Saier, Jr. 2002. Sequence and phylogenetic analyses of the twin-arginine targeting (Tat) protein export system. *Archives of microbiology* **177**:441-50.
- Youngleson, J. S., D. T. Jones et D. R. Woods. 1989. Homology between hydroxybutyryl and hydroxyacyl coenzyme A dehydrogenase enzymes from *Clostridium acetobutylicum* fermentation and vertebrate fatty acid beta-oxidation pathways. *Journal of bacteriology* **171**:6800-7.
- Yuan, J., J. C. Zweers, J. M. van Dijl et R. E. Dalbey. 2010. Protein transport across and into cell membranes in bacteria and archaea. *Cell Mol Life Sci* **67**:179-99.
- Yuen, G., N. Blair, D. J. Des Marais et S. Chang. 1984. Carbon isotope composition of low molecular weight hydrocarbons and monocarboxylic acids from Murchison meteorite. *Nature* **307**:252-4.
- Yuen, G. et K. A. Kvenvolden. 1973. Monocarboxylic acids in Murray and Murchison carbonaceous meteorites. *Nature* **246**:301-02.
- Yuen, G., J. G. Lawless et E. H. Edelson. 1981. Quantification of monocarboxylic acids from a spark discharge synthesis. *Journal of molecular evolution* **17**:43-47.
- Zepik, H. H., S. Rajamani, M. C. Maurel et D. Deamer. 2007. Oligomerization of thioglutamic acid: encapsulated reactions and lipid catalysis. *Orig Life Evol Biosph* **37**:495-505.
- Zhang, L., A. K. Joshi et S. Smith. 2003a. Cloning, expression, characterization, and interaction of two components of a human mitochondrial fatty acid synthase. Malonyltransferase and acyl carrier protein. *The Journal of biological chemistry* **278**:40067-74.
- Zhang, Y. M. et C. O. Rock. 2008. Membrane lipid homeostasis in bacteria. *Nature reviews* **6**:222-33.
- Zhang, Y. M., B. Wu, J. Zheng et C. O. Rock. 2003b. Key residues responsible for acyl carrier protein and beta-ketoacyl-acyl carrier protein reductase (FabG) interaction. *The Journal of biological chemistry* **278**:52935-43.
- Zhaxybayeva, O. et J. P. Gogarten. 2004. Cladogenesis, coalescence and the evolution of the three domains of life. *Trends Genet* **20**:182-7.
- Zhaxybayeva, O., P. Lapierre et J. P. Gogarten. 2005. Ancient gene duplications and the root(s) of the tree of life. *Protoplasma* **227**:53-64.
- Zhou, D. et R. H. White. 1991. Early steps of isoprenoid biosynthesis in *Escherichia coli*. *The Biochemical journal* **273** (Pt 3):627-34.
- Zhu, K., K. H. Choi, H. P. Schweizer, C. O. Rock et Y. M. Zhang. 2006. Two aerobic pathways for the formation of unsaturated fatty acids in *Pseudomonas aeruginosa*. *Molecular microbiology* **60**:260-73.
- Zhu, L., J. Cheng, B. Luo, S. Feng, J. Lin, S. Wang, J. E. Cronan et H. Wang. 2009. Functions of the *Clostridium acetobutylicum* FabF and FabZ proteins in unsaturated fatty acid biosynthesis. *BMC microbiology* **9**:119.
- Zhu, T. F. et J. W. Szostak. 2009. Coupled growth and division of model protocell membranes. *J Am Chem Soc* **131**:5705-13.
- Zuckerandl, E. et L. Pauling. 1965. Molecules as documents of evolutionary history. *J Theor Biol* **8**:357-66.

TABLE DES ANNEXES

Annexe 1. Liste des génomes utilisés pour chaque analyse.

Annexe 2*. Article : *Origins and early evolution of the mevalonate pathway of isoprenoid biosynthesis in the three domains of life.*

Annexe 3. Article : *Early evolution of the biotin-dependent carboxylase family.*

Inclut la correction de cet article.

Annexe 4. Chapitre *in press* du livre “*LUCA: The Last Universal Cellular Ancestor of life*”: *The membranes of the cenancestor.*

Annexe 5. Article : *An ACP independent fatty acid synthesis pathway in archaea: implications for the origin of phospholipids.*

Annexe 6. Article : *The early evolution of lipid membranes and the three domains of life.*

Annexe 7. Article *in press* : *Phylogenomic investigation of phospholipid synthesis in archaea.*

* Les articles sont dans l'ordre dans lequel ils ont été écrits.

THESE DE DOCTORAT DE L'UNIVERSITE PARIS SUD

Ecole doctorale Gènes, Génomes, Cellules (ED 426)

Présentée par :

Jonathan LOMBARD

Pour l'obtention du grade de :

Docteur en Sciences de l'Université Paris Sud

**ORIGINES ET EVOLUTION DES VOIES DE SYNTHÈSE DES PHOSPHOLIPIDES
DANS LES TROIS DOMAINES DU VIVANT.**

IMPLICATIONS POUR LA NATURE DES MEMBRANES DU CENANCETRE

ANNEXES

Thèse soutenue le 17 décembre 2012 devant le jury composé de :

Vincent DAUBIN	Rapporteur
Simonetta GRIBALDO	Rapporteur
Pierre CAPY	Examineur
Olivier LESPINET	Examineur
Michel MORANGE	Examineur
David MOREIRA	Directeur de thèse

TABLE DES ANNEXES

Annexe 1. Liste des génomes utilisés pour chaque analyse.

Annexe 2*. Article : *Origins and early evolution of the mevalonate pathway of isoprenoid biosynthesis in the three domains of life.*

Annexe 3. Article : *Early evolution of the biotin-dependent carboxylase family.*

Inclut la correction de cet article.

Annexe 4. Chapitre *in press* du livre “*LUCA: The Last Universal Cellular Ancestor of life*”:
The membranes of the cenancestor.

Annexe 5. Article : *An ACP independent fatty acid synthesis pathway in archaea: implications for the origin of phospholipids.*

Annexe 6. Article : *The early evolution of lipid membranes and the three domains of life.*

Annexe 7. Article *in press* : *Phylogenomic investigation of phospholipid synthesis in archaea.*

* Les articles apparaissent dans l'ordre chronologique d'écriture.

Annexe 1. Liste des génomes utilisés par analyse

	Acides Gras - Assemblage phospholipides	Voie du méthylethritol phosphate	Voie du mévalonate
Bactéries			
Assembled genomes – Voir M&M			*
<i>Acaryochloris marina</i>	*	*	*
<i>Acholeplasma laidlawii</i>	*	*	*
<i>Acidimicrobium ferrooxidans</i> DSM 10331	*	*	*
<i>Acidiphilium cryptum</i> JF-5	*	*	*
<i>Acidithiobacillus ferrooxidans</i>	*	*	*
<i>Acidobacteria bacterium</i> Ellin345		*	*
<i>Acidobacterium capsulatum</i> ATCC 51196	*		
<i>Acidothermus cellulolyticus</i>		*	*
<i>Acidovorax avenae</i>		*	*
<i>Acidovorax citrulli</i> AAC00-1	*		
<i>Acinetobacter baumannii</i>		*	*
<i>Acinetobacter johnsonii</i> SH046	*		
<i>Actinobacillus pleuropneumoniae</i>	*	*	*
<i>Actinomyces coleocanis</i> DSM 15436	*		
<i>Actinomyces odontolyticus</i>		*	*
<i>Aeromonas hydrophila</i>	*	*	*
<i>Agrobacterium tumefaciens</i> str. C58		*	*
<i>Akkermansia muciniphila</i>	*	*	*
<i>Alcanivorax borkumensis</i> SK2		*	*
<i>Algoriphagus</i> sp. PR1	*	*	*
<i>Alicyclobacillus acidocaldarius</i>	*	*	*
<i>Aliivibrio salmonicida</i> LFI1238		*	*
<i>Alistipes putredinis</i>		*	*
<i>Alkalilimnicola ehrlichei</i> MLHE-1		*	*
<i>Alkaliphilus metalliredigens</i>	*	*	*
<i>Alteromonas macleodii</i>	*	*	*
<i>Aminobacterium colombiense</i> DSM 12261	*		
<i>Aminomonas paucivorans</i>	*		
<i>Anabaena variabilis</i>	*	*	*
<i>Anaerobaculum hydrogeniformans</i>	*		
<i>Anaerocellum thermophilum</i>		*	*
<i>Anaerococcus hydrogenalis</i>		*	*
<i>Anaerofustis stercorihominis</i>		*	*
<i>Anaerolinea thermophila</i>	*		
<i>Anaeromyxobacter dehalogenans</i>		*	*
<i>Anaerostipes caccae</i>		*	*
<i>Anaerotruncus colihominis</i>		*	*
<i>Anaplasma marginale</i> str. Florida	*		
<i>Anaplasma marginale</i> str. St. Marie		*	*
<i>Anoxybacillus flavithermus</i> WK1		*	*
<i>Aquifex aeolicus</i>	*	*	*
<i>Arcobacter butzleri</i> RM4018	*	*	*

<i>Aromatoleum aromaticum EbN1</i>	*	*
<i>Arthrobacteraurescens</i>	*	*
<i>Arthrospira maxima</i>	*	*
<i>Atopobium parvulum DSM 20469</i>	*	*
<i>Aurantimonas sp. SI85-9A1</i>	*	*
<i>Azoarcus sp. BH72</i>	*	*
<i>Azorhizobium caulinodans ORS 571</i>	*	*
<i>Azotobacter vinelandii</i>	*	*
<i>Bacillus cereus</i>	*	*
<i>Bacillus subtilis</i>	*	*
<i>bacterium Ellin514</i>	*	*
<i>Bacteroides fragilis</i>	*	*
<i>Bacteroides uniformis ATCC 8492</i>	*	*
<i>Bacteroides vulgatus</i>	*	*
<i>Bartonella henselae str. Houston-1</i>	*	*
<i>Baumannia cicadellinicola</i>	*	*
<i>Bdellovibrio bacteriovorus HD100</i>	*	*
<i>Beggiatoa sp. PS</i>	*	*
<i>Beijerinckia indica</i>	*	*
<i>Bifidobacterium adolescentis</i>	*	*
<i>Bifidobacterium bifidum</i>	*	*
<i>Blastopirellula marina</i>	*	*
<i>Blautia hansenii</i>	*	*
<i>Bordetella bronchiseptica RB50</i>	*	*
<i>Borrelia burgdorferi</i>	*	*
<i>Brachyspira murdochii DSM 12563</i>	*	*
<i>Bradyrhizobium japonicum USDA 110</i>	*	*
<i>Brevibacterium linens</i>	*	*
<i>Brevundimonas sp. BAL3</i>	*	*
<i>Brevundimonas subvibrioides ATCC 15264</i>	*	*
<i>Brucella abortus</i>	*	*
<i>Bryantella formatexigens</i>	*	*
<i>Buchnera aphidicola</i>	*	*
<i>Burkholderia cenocepacia</i>	*	*
<i>Caldicellulosiruptor saccharolyticus</i>	*	*
<i>Calditerrivibrio nitroreducens</i>	*	*
<i>Caminibacter mediatlanticus TB-2</i>	*	*
<i>Campylobacter gracilis RM3268</i>	*	*
<i>Campylobacter jejuni</i>	*	*
<i>Candidatus Carsonella ruddii PV</i>	*	*
<i>Candidatus Desulfurudis audaxviator</i>	*	*
<i>Candidatus Koribacter versatilis Ellin345</i>	*	*
<i>Candidatus Nitrospira defluvii</i>	*	*
<i>Candidatus Pelagibacter ubique</i>	*	*
<i>Candidatus Phytoplasma australiense</i>	*	*
<i>Candidatus Ruthia magnifica</i>	*	*
<i>Candidatus Solibacter usitatus Ellin6076</i>	*	*
<i>Candidatus Vesicomysocius okutanii HA</i>	*	*
<i>Capnocytophaga sputigena Capno</i>	*	*
<i>Carboxydibrachium pacificum</i>	*	*
<i>Carboxydotherrmus hydrogenoformans</i>	*	*
<i>Carnobacterium sp. AT7</i>	*	*
<i>Caulobacter crescentus CB15</i>	*	*
<i>Cellvibrio japonicus Ueda107</i>	*	*
<i>Chlamydia trachomatis</i>	*	*
<i>Chlamydophila pneumoniae</i>	*	*
<i>Chlorobaculum parvum</i>	*	*
<i>Chlorobium limicola</i>	*	*
<i>Chlorobium luteolum DSM 273</i>	*	*

<i>Chlorobium phaeobacteroides</i>		*	*
<i>Chlorobium tepidum</i> TLS	*		
<i>Chloroflexus aurantiacus</i>	*	*	*
<i>Chloroherpeton thalassium</i>	*	*	*
<i>Chromobacterium violaceum</i> ATCC12472		*	*
<i>Chromohalobacter salexigens</i> DSM 3043		*	*
<i>Chthoniobacter flavus</i>	*	*	*
<i>Citrobacter koseri</i> ATCC BAA-895		*	*
<i>Clavibacter michiganensis</i>		*	*
<i>Clostridium acetobutylicum</i>		*	*
<i>Clostridium botulinum</i>		*	*
<i>Collinsella aerofaciens</i>	*	*	*
<i>Colwellia psychrerythraea</i> 34H		*	*
<i>Comamonas testosterone</i> KF-1		*	*
<i>Congregibacter litoralis</i> KT71		*	*
<i>Coprococcus eutactus</i>	*	*	*
<i>Coprothermobacter proteolyticus</i>		*	*
<i>Coralimargarita akajimensis</i> DSM 45221	*		
<i>Corynebacterium diphtheriae</i>		*	*
<i>Coxiella burnetii</i>		*	*
<i>Croceibacter atlanticus</i>		*	*
<i>Crocospaera watsonii</i>	*	*	*
<i>Cupriavidus taiwanensis</i>		*	*
<i>Cyanobium</i> sp.	*	*	*
<i>Cyanothece</i> sp. ATCC 51142	*	*	*
<i>Cytophaga hutchinsonii</i>	*	*	*
<i>Dechloromonas aromatica</i>	*	*	*
<i>Deferribacter desulfuricans</i>	*		
<i>Dehalococcoides ethenogenes</i>	*	*	*
<i>Dehalogenimonas lykanthroporepellens</i> BL-DC-9	*		
<i>Deinococcus geothermalis</i>	*	*	*
<i>Deinococcus radiodurans</i>	*	*	*
<i>Delftia acidovorans</i> SPH-1		*	*
<i>Denitrovibrio acetiphilus</i>	*		
<i>Desulfarculus baarsii</i> DSM 2075	*		
<i>Desulfatibacillum alkenivorans</i> AK-01	*	*	*
<i>Desulfitobacterium hafniense</i>		*	*
<i>Desulfococcus oleovorans</i> Hxd3	*	*	*
<i>Desulfohalobium retbaense</i> DSM 5692	*		
<i>Desulfotalea psychrophila</i> LSv54		*	*
<i>Desulfotomaculum reducens</i>		*	*
<i>Desulfovibrio vulgaris</i>		*	*
<i>Desulfurispirillum indicum</i>	*		
<i>Desulfuromonas acetoxidans</i> DSM 684		*	*
<i>Dethiosulfovibrio peptidovorans</i> DSM 11002	*		
<i>Diaphorobacter</i> sp. TPSY		*	*
<i>Dichelobacter nodosus</i> VCS1703A		*	*
<i>Dictyoglomus thermophilum</i>	*	*	*
<i>Dictyoglomus turgidum</i> DSM 6724	*		
<i>Dinoroseobacter shibae</i> DFL12		*	*
<i>Dokdonia donghaensis</i>	*	*	*
<i>Dorea formicigenerans</i>		*	*
<i>Ehrlichia canis</i> str. Jake		*	*
<i>Elusimicrobium minutum</i>	*	*	*
<i>Endoriftia persephone</i>		*	*
<i>Enterobacter sakazakii</i> ATCC BAA-894	*	*	*
<i>Enterococcus faecalis</i>	*	*	*
<i>Epulopiscium</i> sp.		*	*
<i>Erwinia tasmaniensis</i> Et1/99		*	*

<i>Erythrobacter litoralis</i> HTCC2594		*	*
<i>Escherichia coli</i> O157:H7	*	*	*
<i>Eubacterium dolichum</i>	*		
<i>Eubacterium siraeum</i>		*	*
<i>Exiguobacterium sibiricum</i>		*	*
<i>Faecalibacterium prausnitzii</i>		*	*
<i>Fervidobacterium nodosum</i>	*	*	*
<i>Fibrobacter succinogenes</i>	*		
<i>Finegoldia magna</i>		*	*
<i>Flavobacterium johnsoniae</i>	*	*	*
<i>Francisella tularensis</i>	*	*	*
<i>Frankia alni</i> ACN14a	*	*	*
<i>Fulvimarina pelagi</i> HTCC2506		*	*
<i>Fusobacterium nucleatum</i>	*	*	*
<i>Gallionella capsiferriformans</i> ES-2	*		
<i>Gemmata obscuriglobus</i>	*	*	*
<i>Gemmatimonas aurantiaca</i> T-27	*		
<i>Geobacillus kaustophilus</i>		*	*
<i>Geobacter metallireducens</i> GS-15	*	*	*
<i>Glaciecola</i> sp. HTCC2999		*	*
<i>Gloeobacter violaceus</i>	*	*	*
<i>Gluconacetobacter diazotrophicus</i> PAI 5		*	*
<i>Gluconobacter oxydans</i> 621H		*	*
<i>Gramella forsetii</i>		*	*
<i>Granulibacter bethesdensis</i> CGDNIH1		*	*
<i>Haemophilus influenzae</i>		*	*
<i>Hahella chejuensis</i> KCTC 2396		*	*
<i>Halomonas elongata</i> DSM 2581	*		
<i>Halorhodospira halophila</i> SL1		*	*
<i>Halothermothrix orenii</i>		*	*
<i>Helicobacter pylori</i>	*	*	*
<i>Heliobacterium modesticaldum</i>	*	*	*
<i>Herminiimonas arsenicoxydans</i>		*	*
<i>Herpetosiphon aurantiacus</i>	*	*	*
<i>Histophilus somni</i>		*	*
<i>Hoeflea phototrophica</i> DFL-43		*	*
<i>Hydrogenivirga</i> sp. 128-5-R1-1	*	*	*
<i>Hydrogenobaculum</i> sp. Y04AAS1	*	*	*
<i>Hyphomonas neptunium</i> ATCC 15444		*	*
<i>Idiomarina baltica</i> OS145		*	*
<i>Ilyobacter polytropus</i> DSM 2926	*		
<i>Isosphaera pallida</i> ATCC 43644	*		
<i>Janibacter</i> sp. HTCC2649		*	*
<i>Jannaschia</i> sp. CCS1		*	*
<i>Janthinobacterium</i> sp. Marseille		*	*
<i>Kineococcus radiotolerans</i>		*	*
<i>Klebsiella pneumoniae</i>		*	*
<i>Kocuria rhizophila</i>		*	*
<i>Kordia algicida</i>		*	*
<i>Kosmotoga olearia</i> TBF 19.5.1	*		
<i>Ktedonobacter racemifer</i> DSM 44963	*		
<i>Labrenzia aggregata</i> IAM 12614		*	*
<i>Lactobacillus acidophilus</i> NCFM		*	*
<i>Lactococcus lactis</i>	*	*	*
<i>Lawsonia intracellularis</i> PHE/MN1-00	*	*	*
<i>Leeuwenhoekiiella blandensis</i>		*	*
<i>Legionella pneumophila</i>		*	*
<i>Leifsonia xyli</i>		*	*
<i>Lentisphaera araneosa</i>	*	*	*

<i>Leptolyngbya valderiana</i>	*	*	*
<i>Leptospira interrogans</i>	*	*	*
<i>Leptospirillum ferrooxidans</i>	*		
<i>Leptothrix cholodnii SP-6</i>		*	*
<i>Leptotrichia hofstadii</i>	*		
<i>Leuconostoc citreum KM20</i>		*	*
<i>Leuconostoc mesenteroides</i>		*	*
<i>Limnobacter sp. MED105</i>		*	*
<i>Listeria innocua</i>		*	*
<i>Loktanella vestfoldensis SKA53</i>		*	*
<i>Lyngbya sp. PCC 8106</i>		*	*
<i>Lysinibacillus sphaericus</i>		*	*
<i>Magnetococcus sp. MC-1</i>	*	*	*
<i>Magnetospirillum magneticum AMB-1</i>		*	*
<i>Mannheimia haemolytica PHL213</i>		*	*
<i>Maricaulis maris MCS10</i>	*	*	*
<i>Marinitoga piezophila KA3</i>	*	*	*
<i>Marinobacter algicola DG893</i>		*	*
<i>Mariprofundus ferrooxydans PV-1</i>	*	*	*
<i>Meiothermus ruber DSM 1279</i>	*		
<i>Meiothermus silvanus DSM 9946</i>	*		
<i>Mesoplasma florum LI</i>	*	*	*
<i>Mesorhizobium loti MAFF303099</i>		*	*
<i>Methylacidiphilum infernorum V4</i>		*	*
<i>Methylibium petroleiphilum PM1</i>	*	*	*
<i>Methylobacillus flagellatus KT</i>	*	*	*
<i>Methylobacterium extorquens PA1</i>		*	*
<i>Methylocella silvestris BL2</i>		*	*
<i>Methylococcus capsulatus str. Bath</i>		*	*
<i>Methylophaga sp. DMS010</i>		*	*
<i>Micrococcus luteus</i>		*	*
<i>Microcoleus chthonoplastes</i>		*	*
<i>Microcystis aeruginosa</i>		*	*
<i>Microscilla marina</i>	*	*	*
<i>Mitsuokella multacida</i>	*	*	*
<i>Moorella thermoacetica</i>	*	*	*
<i>Moritella sp. PE36</i>		*	*
<i>Mycobacterium avium</i>		*	*
<i>Mycobacterium tuberculosis</i>		*	*
<i>Mycoplasma agalactiae PG2</i>	*		
<i>Mycoplasma genitalium G37</i>		*	*
<i>Mycoplasma pneumoniae M129</i>	*	*	*
<i>Mycoplasma pulmonis UAB CTIP</i>		*	*
<i>Myxococcus xanthus DK 1622</i>	*	*	*
<i>Natranaerobius thermophilus</i>		*	*
<i>Nautilia profundicola AmH</i>	*		
<i>Neisseria elongata ATCC 29315</i>	*		
<i>Neisseria gonorrhoeae</i>		*	*
<i>Neorickettsia sennetsu str. Miyayama</i>		*	*
<i>Neptuniibacter caesariensis</i>		*	*
<i>Nitratifactor salsuginis</i>	*		
<i>Nitratiruptor sp. SB155-2</i>	*	*	*
<i>Nitrobacter hamburgensis X14</i>		*	*
<i>Nitrosococcus oceani</i>		*	*
<i>Nitrosomonas europaea ATCC 19718</i>		*	*
<i>Nitrosomonas eutropha C91</i>	*		
<i>Nitrospira multififormis ATCC 25196</i>		*	*
<i>Nocardia farcinica</i>		*	*
<i>Nocardioides sp. JS614</i>		*	*

<i>Nodularia spumigena</i>		*	*
<i>Nostoc punctiforme</i> PCC 73102	*	*	*
<i>Novosphingobium aromaticivorans</i> DSM12444		*	*
<i>Oceanibulbus indolifex</i> HEL-45		*	*
<i>Oceanicaulis alexandrii</i> HTCC2633		*	*
<i>Oceanicola batsensis</i> HTCC2597		*	*
<i>Oceanithermus profundus</i>	*		
<i>Oceanobacillus iheyensis</i>		*	*
<i>Oceanobacter</i> sp. RED65		*	*
<i>Ochrobactrum anthropi</i> ATCC 49188		*	*
<i>Octadecabacter antarcticus</i>		*	*
<i>Oenococcus oeni</i>		*	*
<i>Oligotropha carboxidovorans</i> OM5		*	*
<i>Opitutus terrae</i> PB90-1	*	*	*
<i>Orientia tsutsugamushi</i>		*	*
<i>Paenibacillus larvae</i>		*	*
<i>Parabacteroides johnsonii</i>		*	*
<i>Parachlamydia acanthamoebae</i> str. Hall's coccus	*		
<i>Paracoccus denitrificans</i> PD1222		*	*
<i>Parvibaculum lavamentivorans</i> DS-1		*	*
<i>Parvimonas micra</i>		*	*
<i>Parvularcula bermudensis</i> HTCC2503		*	*
<i>Pasteurella multocida</i> subsp. <i>multocida</i> str. Pm70	*	*	*
<i>Pasteuria nishizawae</i>		*	*
<i>Pectobacterium atrosepticum</i> SCRI1043		*	*
<i>Pediococcus pentosaceus</i>		*	*
<i>Pedobacter</i> sp. BAL39	*	*	*
<i>Pelobacter propionicus</i> DSM 2379		*	*
<i>Pelodictyon luteolum</i>		*	*
<i>Pelotomaculum thermopropionicum</i>		*	*
<i>Persephonella marina</i>	*		
<i>Petrotoga mobilis</i> SJ95	*	*	*
<i>Phaeobacter gallaeciensis</i>		*	*
<i>Phenyllobacterium zucineum</i> HLK1		*	*
<i>Photobacterium angustum</i> S14	*		
<i>Photobacterium profundum</i>		*	*
<i>Photorhabdus luminescens</i> Laumondii TTO1		*	*
<i>Pirellula staleyi</i> DSM 6068	*		
<i>Planctomyces limnophilus</i> DSM 3776	*		
<i>Planctomyces maris</i> DSM 8797		*	*
<i>Plesiocystis pacifica</i> SIR-1		*	*
<i>Polaribacter irgensii</i>		*	*
<i>Polaromonas naphthalenivorans</i> CJ2		*	*
<i>Polynucleobacter necessarius</i>		*	*
<i>Porphyromonas gingivalis</i>	*	*	*
<i>Prevotella amnii</i> CRIS 21A-A	*		
<i>Prochlorococcus marinus</i>	*	*	*
<i>Propionibacterium acnes</i>		*	*
<i>Prosthecochloris aestuarii</i>	*	*	*
<i>Proteus mirabilis</i> HI4320		*	*
<i>Protochlamydia amoebophila</i>		*	*
<i>Providencia alcalifaciens</i> DSM 30120		*	*
<i>Pseudoalteromonas atlantica</i> T6c		*	*
<i>Pseudomonas aeruginosa</i>		*	*
<i>Pseudovibrio</i> sp. JE062		*	*
<i>Psychrobacter arcticus</i> 273-4		*	*
<i>Psychroflexus torquis</i>		*	*
<i>Psychromonas ingrahamii</i> 37		*	*
<i>Pyramidobacter piscolens</i>	*		

<i>Ralstonia eutropha</i> H16	*	*	*
<i>Reinekea</i> sp. MED297		*	*
<i>Renibacterium salmoninarum</i>		*	*
<i>Rhizobium leguminosarum</i>	*	*	*
<i>Rhodobacter sphaeroides</i>	*	*	*
<i>Rhodococcus jostii</i>		*	*
<i>Rhodoferax ferrireducens</i> T118		*	*
<i>Rhodopirellula baltica</i> SH 1	*	*	*
<i>Rhodopseudomonas palustris</i>		*	*
<i>Rhodospirillum centenum</i> SW		*	*
<i>Rhodospirillum rubrum</i> ATCC 11170	*		
<i>Rickettsia prowazekii</i> str. Madrid E		*	*
<i>Rickettsiella grylli</i>		*	*
<i>Robiginitalea biformata</i>		*	*
<i>Roseburia intestinalis</i>		*	*
<i>Roseiflexus castenholzii</i> DSM 13941	*	*	*
<i>Roseobacter denitrificans</i> OCh 114		*	*
<i>Roseovarius nubinhibens</i> ISM		*	*
<i>Rubroacter xylanophilus</i>	*	*	*
<i>Ruegeria</i> sp. R11		*	*
<i>Ruminococcus gnavus</i>		*	*
<i>Saccharophagus degradans</i> 2-40		*	*
<i>Saccharopolyspora erythraea</i>		*	*
<i>Sagittula stellata</i> E-37		*	*
<i>Salinibacter ruber</i>	*	*	*
<i>Salinispora arenicola</i>		*	*
<i>Salmonella typhimurium</i> LT2		*	*
<i>Sebaldella termitidis</i> ATCC 33386	*		
<i>Shewanella baltica</i>		*	*
<i>Shigella flexneri</i>		*	*
<i>Silicibacter pomeroyi</i> DSS-3		*	*
<i>Sinorhizobium meliloti</i> 1021		*	*
<i>Sodalis glossinidius</i> str. 'morsitans'		*	*
<i>Solibacter usitatus</i> Ellin6076	*	*	*
<i>Sorangium cellulosum</i> 'So ce 56'		*	*
<i>Sphingomonas wittichii</i> RW1	*	*	*
<i>Sphingopyxis alaskensis</i> RB2256	*	*	*
<i>Spirochaeta smaragdinae</i> DSM 11293	*		
<i>Spiroplasma citri</i>	*		
<i>Staphylococcus aureus</i>		*	*
<i>Stenotrophomonas maltophilia</i>	*	*	*
<i>Stigmatella aurantiaca</i> DW4/3-1		*	*
<i>Streptobacillus moniliformis</i> DSM 12112	*		
<i>Streptococcus agalactiae</i>	*	*	*
<i>Streptococcus thermophilus</i>		*	*
<i>Streptomyces albus</i>		*	*
<i>Streptomyces coelicolor</i>		*	*
<i>Sulfitobacter</i> sp. EE-36		*	*
<i>Sulfurihydrogenibium</i> sp. YO3AOP1		*	*
<i>Sulfurihydrogenibium yellowstonense</i>	*		
<i>Sulfurimonas denitrificans</i> DSM 1251	*	*	*
<i>Sulfurovum</i> sp. NBC37-1	*	*	*
<i>Symbiobacterium thermophilum</i>		*	*
<i>Synechococcus elongatus</i>		*	*
<i>Synechocystis</i> sp. PCC 6803		*	*
<i>Syntrophobacter fumaroxidans</i> MPOB	*	*	*
<i>Syntrophomonas wolfei</i>		*	*
<i>Syntrophus aciditrophicus</i> SB	*	*	*
<i>Terriglobus saanensis</i> SP1PR4	*		

<i>Thauera</i> sp. MZ1T	*	*	*
<i>Thermanaerovibrio acidaminovorans</i>	*		
<i>Thermoanaerobacter tengcongensis</i>		*	*
<i>Thermobifida fusca</i>		*	*
<i>Thermocrinis albus</i> DSM 14484	*		
<i>Thermodesulfovibrio yellowstonii</i> DSM 11347	*	*	*
<i>Thermomicrobium roseum</i>	*		
<i>Thermosinus carboxydivorans</i>		*	*
<i>Thermosipho africanus</i> TCF52B	*	*	*
<i>Thermosynechococcus elongatus</i>		*	*
<i>Thermotoga maritima</i> MSB8	*	*	*
<i>Thermovibrio ammonificans</i>	*		
<i>Thermus thermophilus</i>	*	*	*
<i>Thioalkalivibrio</i> sp. HL-EbGR7		*	*
<i>Thiobacillus denitrificans</i> ATCC 25259	*	*	*
<i>Thiomicrospira crunogena</i> XCL-2		*	*
<i>Treponema brennaborensis</i> DSM 12168	*		
<i>Treponema denticola</i> ATCC 35405	*		
<i>Treponema pallidum</i>		*	*
<i>Trichodesmium erythraeum</i>		*	*
<i>Tropheryma whipplei</i>		*	*
<i>Truepera radiovictrix</i>	*		
<i>Ureaplasma parvum</i> serovar 1 str. ATCC 27813	*		
<i>Verminephrobacter eiseniae</i> EF01-2		*	*
<i>Verrucomicrobium spinosum</i> DSM 4136	*	*	*
<i>Vibrio cholerae</i>		*	*
<i>Victivallis vadensis</i>	*	*	*
<i>Waddlia chondrophila</i> WSU 86-1044	*		
<i>Wigglesworthia glossinidia</i>		*	*
<i>Wolinella succinogenes</i> DSM 1740		*	*
<i>Xanthobacter autotrophicus</i> Py2	*	*	*
<i>Xanthomonas oryzae</i>	*	*	*
<i>Xylella fastidiosa</i>		*	*
<i>Zymomonas mobilis</i> subsp. <i>mobilis</i> ZM4		*	*

Archées

Banque de données nr de genbank – Voir M&M	*		
<i>Acidilobus saccharovorans</i> 345-15	*		
<i>Aciduliprofundum boonei</i> T469	*		
<i>Aeropyrum pernix</i> K1	*	*	*
<i>Archaeoglobus fulgidus</i> DSM 4304	*	*	*
<i>Archaeoglobus profundus</i> DSM 5631	*		
<i>Caldivirga maquilingsensis</i> IC-167	*	*	*
<i>Candidatus Caldiarchaeum subterraneum</i>	*		
<i>Candidatus Korarchaeum cryptofilum</i> OPF8	*		
<i>Candidatus Methanoregula boonei</i> 6A8	*	*	*
<i>Cenarchaeum symbiosum</i> A	*	*	*
<i>Desulfurococcus kamchatkensis</i> 1221n	*	*	*
<i>Ferroglobus placidus</i> DSM 10642	*		
<i>Ferroplasma acidarmanus</i> fer1	*	*	*
<i>Halalkalicoccus jeotgali</i> B3	*		
<i>Haloarcula marismortui</i> ATCC 43049	*	*	*
<i>Halobacterium salinarum</i>	*	*	*
<i>Haloferax volcanii</i> DS2	*		
<i>Halogeometricum borinquense</i> DSM 11551	*		
<i>Halomicrobium mukohataei</i> DSM 12286	*		
<i>Haloquadratum walsbyi</i> DSM 16790	*	*	*
<i>Halorhabdus utahensis</i> DSM 12940	*		

<i>Halorubrum lacusprofundi</i> ATCC 49239	*	*	*
<i>Haloterrigena turkmenica</i> DSM 5511	*		
<i>Hyperthermus butylicus</i> DSM 5456	*	*	*
<i>Ignicoccus hospitalis</i> KIN4/1	*	*	*
<i>Ignisphaera aggregans</i> DSM 17230	*		
<i>Metallosphaera sedula</i> DSM 5348	*	*	*
<i>Methanobrevibacter ruminantium</i> M1	*		
<i>Methanobrevibacter smithii</i>	*	*	*
<i>Methanocaldococcus fervens</i> AG86	*		
<i>Methanocaldococcus infernus</i> ME	*		
<i>Methanocaldococcus jannaschii</i> DSM 2661	*	*	*
<i>Methanocaldococcus</i> sp. FS406-22	*		
<i>Methanocaldococcus vulcanius</i> M7	*		
<i>Methanocella paludicola</i> SANAE	*		
<i>Methanococcoides burtonii</i> DSM 6242	*	*	*
<i>Methanococcus aeolicus</i> Nankai-3	*	*	*
<i>Methanococcus maripaludis</i>	*	*	*
<i>Methanococcus vannielii</i> SB	*	*	*
<i>Methanococcus voltae</i> A3	*	*	*
<i>Methanocorpusculum labreanum</i> Z	*	*	*
<i>Methanoculleus marisnigri</i> JR1	*	*	*
<i>Methanohalobium evestigatum</i> Z-7303	*		
<i>Methanohalophilus mahii</i> DSM 5219	*		
<i>Methanoplanus petrolearius</i> DSM 11571	*		
<i>Methanopyrus kandleri</i> AV19	*	*	*
<i>Methanosaeta thermophila</i> PT	*	*	*
<i>Methanosarcina acetivorans</i> C2A	*	*	*
<i>Methanosarcina barkeri</i> str. Fusaro	*	*	*
<i>Methanosarcina mazei</i> Go1	*	*	*
<i>Methanosphaera stadtmanae</i> DSM 3091	*	*	*
<i>Methanosphaerula palustris</i> E1-9c	*		
<i>Methanospirillum hungatei</i> JF-1	*	*	*
<i>Methanothermobacter marburgensis</i> str. Marburg	*		
<i>Methanothermobacter thermautotrophicus</i> Delta H	*	*	*
<i>Methanothermococcus okinawensis</i> IH1	*		
<i>Methanothermus fervidus</i>	*		
<i>Nanoarchaeum equitans</i> Kin4-M	*	*	*
<i>Natrialba magadii</i> ATCC 43099	*	*	*
<i>Natronomonas pharaonis</i> DSM 2160	*	*	*
<i>Nitrosopumilus maritimus</i> SCM1	*	*	*
<i>Picrophilus torridus</i> DSM 9790	*	*	*
<i>Pyrobaculum aerophilum</i> str. IM2	*	*	*
<i>Pyrobaculum arsenaticum</i> DSM 13514	*	*	*
<i>Pyrobaculum calidifontis</i> JCM 11548	*	*	*
<i>Pyrobaculum islandicum</i> DSM 4184	*	*	*
<i>Pyrococcus abyssi</i> GE5	*	*	*
<i>Pyrococcus furiosus</i> DSM 3638	*	*	*
<i>Pyrococcus horikoshii</i> OT3	*	*	*
<i>Staphylothermus hellenicus</i> DSM 12710	*		
<i>Staphylothermus marinus</i> F1	*	*	*
<i>Sulfolobus acidocaldarius</i> DSM 639	*	*	*
<i>Sulfolobus solfataricus</i> P2	*	*	*
<i>Sulfolobus tokodaii</i> str. 7	*	*	*
<i>Thermococcus barophilus</i> MP	*	*	*
<i>Thermococcus gammatolerans</i> EJ3	*		
<i>Thermococcus kodakarensis</i> KOD1	*	*	*
<i>Thermococcus onnurineus</i> NA1	*	*	*
<i>Thermococcus sibiricus</i> MM 739	*	*	*
<i>Thermococcus</i> sp. AM4	*	*	*

<i>Thermofilum pendens</i> Hrk 5	*	*	*
<i>Thermoplasma acidophilum</i> DSM 1728	*	*	*
<i>Thermoplasma volcanium</i> GSS1	*	*	*
<i>Thermoproteus neutrophilus</i> V24Sta	*	*	*
<i>Thermoproteus tenax</i> Kra1	*		
<i>Thermosphaera aggregans</i> DSM 11486	*		
<i>Vulcanisaeta distributa</i> DSM 14429	*		

Eucaryotes

Banque de données nr de genbank		*	*
Banques de données EST – Voir Matériel et méthodes		*	*
<i>Acanthamoeba castellanii</i>	*		
<i>Arabidopsis thaliana</i>	*		
<i>Aureococcus anophagefferens</i>	*		
<i>Babesia bovis</i> T2Bo	*		
<i>Bigelowiella natans</i>	*		
<i>Blastocladia emersonii</i>	*		
<i>Bryopsis hypnoides</i>	*		
<i>Chlamydomonas reinhardtii</i>	*		
<i>Chlorella vulgaris</i>		*	
<i>Chlorokybus atmophyticus</i>	*		
<i>Chromera velia</i>	*		
<i>Chrysodidymus synuroideus</i>	*		
<i>Cryptococcus neoformans</i> var. <i>grubii</i>	*		
<i>Cryptomonas paramecium</i>	*		
<i>Cryptosporidium hominis</i> TU502	*		
<i>Cyanidioschyzon merolae</i>		*	
<i>Cyanophora paradoxa</i>	*		
<i>Desmarestia viridis</i>	*		
<i>Dictyostelium discoideum</i> AX4	*		
<i>Durinskia baltica</i>	*		
<i>Eimeria tenella</i> str. Houghton	*		
<i>Emiliana huxleyi</i>	*		
<i>Encephalitozoon cuniculi</i> GB-M1	*		
<i>Entamoeba histolytica</i> HM-1:IMSS	*		
<i>Euglena gracilis</i>	*		
<i>Galdieria sulphuraria</i>		*	
<i>Giardia intestinalis</i>	*		
<i>Glaucocystis nostochinearum</i>	*		
<i>Glomus irregulare</i>	*		
<i>Gracilaria chilensis</i>	*		
<i>Guillardia theta</i>	*		
<i>Haemoproteus columbae</i>	*		
<i>Hartmannella vermiformis</i>	*		
<i>Hemiselmis andersenii</i>	*		
<i>Igernella notabilis</i>	*		
<i>Kryptoperidinium foliaceum</i>	*		
<i>Leishmania major</i>	*		
<i>Leucocytozoon majoris</i>	*		
<i>Malawimonas jakobiformis</i>	*		
<i>Mesostigma viride</i>	*		
<i>Micromonas pusilla</i> CCMP1545	*	*	
<i>Monosiga brevicollis</i> ATCC 50154	*		
<i>Naegleria gruberi</i>	*		
<i>Nephroselmis olivacea</i>	*		
<i>Ochromonas danica</i>	*		
<i>Odontella sinensis</i>	*		
<i>Pan troglodytes verus</i>	*		

<i>Parahaemoproteus vireonis</i>	*	
<i>Paramecium tetraurelia</i> strain d4-2	*	
<i>Paulinella chromatophora</i>	*	
<i>Perkinsus marinus</i>	*	
<i>Phakopsora pachyrhizi</i>	*	
<i>Physarum polycephalum</i>	*	
<i>Phytophthora infestans</i> T30-4	*	
<i>Plasmodium falciparum</i> 3D7	*	
<i>Porphyra purpurea</i>	*	
<i>Reclinomonas americana</i>	*	
<i>Rhodomonas salina</i>	*	
<i>Saccharomyces cerevisiae</i> A364A	*	
<i>Schizosaccharomyces pombe</i> 972h-	*	
<i>Staurastrum punctulatum</i>	*	
<i>Synedra acus</i> Cafeteria roenbergensis	*	
<i>Talaromyces stipitatus</i> ATCC 10500	*	
<i>Tetrahymena thermophila</i>	*	
<i>Thalassiosira oceanica</i> CCMP1005	*	
<i>Theileria annulata</i> strain Ankara	*	
<i>Toxoplasma gondii</i> ME49	*	
<i>Trichomonas vaginalis</i> G3	*	
<i>Trichoplax adhaerens</i>	*	
<i>Trypanosoma cruzi</i>	*	
<i>Ustilago maydis</i> 521	*	
<i>Volvox carteri</i>	*	*

ANNEXE 2.

Origins and early evolution of the mevalonate pathway of isoprenoid biosynthesis in the three domains of life.

Lombard, J. et D. Moreira. 2011.

Molecular biology and evolution **28**:87-99.

Origins and Early Evolution of the Mevalonate Pathway of Isoprenoid Biosynthesis in the Three Domains of Life

Jonathan Lombard and David Moreira*

Unité d'Ecologie, Systématique et Evolution, Centre National de la Recherche Scientifique, Université Paris-Sud, Orsay, France

*Corresponding author: E-mail: david.moreira@u-psud.fr.

Associate editor: Andrew Roger

Abstract

Isoprenoids are a very diverse family of organic compounds widespread in the three domains of life. Although they are produced from the condensation of the same precursors in all organisms (isopentenyl pyrophosphate and dimethylallyl diphosphate), the evolutionary origin of their biosynthesis remains controversial. Two independent nonhomologous metabolic pathways are known: the mevalonate (MVA) pathway in eukaryotes and archaea and the methylerythritol phosphate (MEP) pathway in bacteria and several photosynthetic eukaryotes. The MVA pathway is also found in a few bacteria, what has been explained in previous works by recent acquisition by horizontal gene transfer (HGT) from eukaryotic or archaeal donors. To reconsider the question of the evolutionary origin of the MVA pathway, we have studied the origin and the evolution of the enzymes of this pathway using phylogenomic analyses upon a taxon-rich sequence database. On the one hand, our results confirm the conservation in archaea of an MVA pathway partially different from eukaryotes. This implies that each domain of life possesses a characteristic major isoprenoid biosynthesis pathway: the classical MVA pathway in eukaryotes, a modified MVA pathway in archaea, and the MEP pathway in bacteria. On the other hand, despite the identification of several HGT events, our analyses support that the MVA pathway was ancestral not only in archaea and eukaryotes but also in bacteria, in contradiction with previous claims that the presence of this pathway in bacteria was due to HGT from other domains. Therefore, the MVA pathway is likely an ancestral metabolic route in all the three domains of life, and hence, it was probably present in the last common ancestor of all organisms (the cenancestor). These findings open the possibility that the cenancestor had membranes containing isoprenoids.

Key words: isoprenoids, mevalonate pathway, early evolution, cenancestor.

Introduction

Isoprenoids constitute one of the largest families of biological compounds, encompassing around 30,000 known products in the three domains of life (bacteria, archaea, and eukaryotes). They fulfill diverse biochemical functions as major structural membrane components in archaea, photosynthetic pigments, hormones, quinones acting in electron transport chains, and plant defense compounds (McGarvey and Croteau 1995; Lange et al. 2000). All these molecules are oligomers synthesized by organisms of the three domains of life through successive condensations of two activated forms of isoprene: the isopentenyl pyrophosphate (IPP) and the dimethylallyl diphosphate (DMAPP) (Ruzicka 1953). Although all known organisms seem to employ IPP and DMAPP, these two isoprenoid precursors can be synthesized by two independent and nonhomologous metabolic pathways, namely, the mevalonate (MVA) and the methylerythritol phosphate (MEP) pathways (fig. 1 and supplementary fig. 1, Supplementary Material online).

The MVA pathway (fig. 1) was first discovered in yeasts and animals in the 1950s, and during some time, it was assumed to be responsible for the production of IPP and DMAPP from acetyl-coenzyme A (CoA) in all organisms (McGarvey and Croteau 1995), even if it seemed to be absent in most bacteria (Zhou and White 1991). The

MEP pathway, capable of generating IPP and DMAPP from D-glyceraldehyde-3-phosphate (G3P) and pyruvate, was discovered in the 1990s in bacteria and plants (Rodríguez-Concepción and Boronat 2002; Phillips et al. 2008). Today, the MVA pathway is considered to be the major route of IPP and DMAPP synthesis in eukaryotes and archaea (Kuzuyama 2002; Boucher et al. 2004), whereas the MEP pathway would be characteristic of bacteria (Lange et al. 2000). However, several exceptions to this general rule are known: First, some bacteria do possess the MVA pathway (Bochar et al. 1999; Romanowski et al. 2002; Voynova et al. 2004; Steussy et al. 2006), although this has been commonly explained as a horizontal gene transfer (HGT) acquisition from archaeal or eukaryotic donors (Boucher and Doolittle 2000; Wilding et al. 2000). Second, plants, other photosynthetic eukaryotes, and apicomplexa (a clade of parasitic protists derived from a photosynthetic ancestor) have been found to contain the MEP pathway in addition to the MVA one. All these eukaryotic groups carry plastids (either photosynthetic or, in the case of apicomplexa, highly derived), suggesting that they obtained the MEP pathway through the transfer of genes from the original cyanobacterial endosymbiont that originated the plastids (Lange et al. 2000). However, although the phylogenies of some of the enzymes of this pathway support the expected cyanobacterial origin, some others appear to support

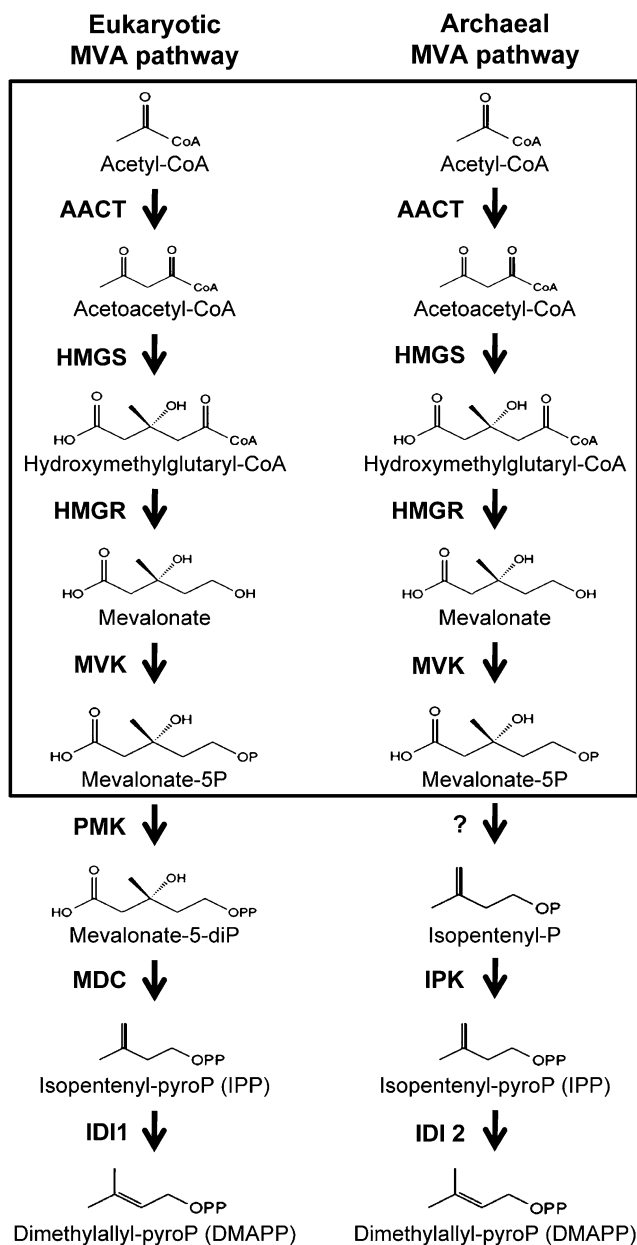


Fig. 1. The partially different mevalonate pathways of eukaryotes and archaea. The classical mevalonate pathway discovered in eukaryotes (left) and the new mevalonate pathway proposed in archaea (right) only share the first four steps (boxed area).

different noncyanobacterial origins (Brinkman et al. 2002; Matsuzaki et al. 2008; Moustafa et al. 2008). Finally, most archaeal species lack the three last enzymes of the classic (eukaryotic-like) MVA pathway, as the analysis of some archaeal genomes already revealed several years ago (Smit and Mushegian 2000; Boucher et al. 2004). These enzymes are the phosphomevalonate kinase (PMK), the mevalonate-5-decarboxylase (MDC), and the isopentenyl diphosphate isomerase (IDI), which are only present in some restricted archaeal clades maybe as a result of proposed HGTs from eukaryotes or bacteria (Boucher et al. 2004).

The modifications of the MVA pathway in archaea have been studied in particular detail. To explain the absence of

some MVA pathway enzymes in these organisms, some authors have underlined the existence of examples of nonhomologous enzyme recruitment to replace some steps of the MVA pathway in other groups. This is the case of the nonorthologous PMK similar to viral nucleoside monophosphate kinases used by metazoa instead of the regular gene from the galactokinase, homoserine kinase, mevalonate kinase (MVK), and PMK family (Houten and Waterham 2001; Herdendorf and Miziorko 2006). These findings have led to propose the same kind of nonhomologous replacements in archaea (Smit and Mushegian 2000), and several studies have tried to identify the enzymes carrying out the analogous reactions required to complete the pathway in this domain of life. A nonhomologous enzyme able to catalyze the same reaction as IDI has been discovered in *Streptomyces* sp. and called IDI2 (Kaneda et al. 2001); this IDI2 has not to be mistaken for another enzyme of mammals, also named IDI2, which is a paralogue of the first IDI to have been isolated, IDI1. Subsequently, IDI2 was found to be encoded in the genomes of different archaea and characterized in several of them (Barkley et al. 2004; Dutoit et al. 2008). In addition, when trying to isolate enzymes in *Methanocaldococcus jannaschii* able to phosphorylate the mevalonate-5-phosphate instead of the typical PMK, Grochowski et al. (2006) found a putative alternative mechanism to produce IPP. They identified an enzyme able to phosphorylate the isopentenyl phosphate into IPP (an isopentenyl phosphate kinase, IPK), and they proposed that it could be part of an alternative pathway to produce IPP in *M. jannaschii* in which the order of the phosphorylation and decarboxylation steps would be interchanged (fig. 1).

Despite the ubiquity of isoprenoids, the uneven taxonomic distribution of the two major biosynthetic pathways (MVA in archaea and eukaryotes and MEP in bacteria) opens several questions about their evolutionary origin. Some authors have proposed that the two nonhomologous pathways would have appeared twice independently, one in the bacterial lineage and the other one in the archaeal descent. According to this scenario, the last common ancestor of all organisms (the cenancestor) would have had no isoprenoids (Lange et al. 2000; Martin and Russell 2003). As isoprenoids are essential components of the archaeal membranes, their late emergence might imply, in the opinion of some of these authors (Lange et al. 2000; Martin and Russell 2003), the absence of lipid membranes in the cenancestor. Thereby, the cenancestor has been proposed to be either noncellular (Koga et al. 1998) or limited by a mineral structure instead of a lipid membrane (Martin and Russell 2003).

Some recent phylogenomic analyses have already confirmed the MEP pathway to be restricted to bacteria and plastid-bearing eukaryotes (Matsuzaki et al. 2008). In this work, we have carried out detailed phylogenetic analyses of each enzyme involved in the other major route of isoprenoid biosynthesis, the MVA pathway. Our results complete the data about the existence of an alternative MVA pathway in archaea as proposed by some authors

(Boucher et al. 2004; Grochowski et al. 2006) and suggest that each major pathway for the synthesis of the isoprenoid precursors evolved in each domain of life (the MEP pathway in bacteria and the two MVA pathways in archaea and eukaryotes, respectively). Interestingly, our phylogenies disagree with the hypothesis that all the bacteria bearing the MVA pathway acquired it from archaea or eukaryotes by HGT. On the contrary, this pathway appears to have been present in the last common ancestor of bacteria and, most likely, in the cenancestor. This finding has important implications concerning the nature of lipids in the ancestral membranes and their subsequent evolution during the diversification of the three domains of life.

Materials and Methods

Sequence Retrieval and Alignment

For each domain of life, one protein sequence of each of the MVA pathway enzymes was retrieved from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (<http://www.genome.jp/kegg>), with a few exceptions: The sequences of *Emiliana huxleyi*, *Aureococcus anophagefferens*, and *Naegleria gruberi* were obtained from the Joint Genome Institute (<http://genome.jgi-psf.org>), the 3-hydroxy-3-methylglutaryl-CoA reductase (HMGR) of *Pseudomonas mevalonii* was obtained from the Protein Data Bank (<http://www.rcsb.org/pdb/home/home.do>), and the 3-hydroxy-3-methylglutaryl-CoA synthase (HMGS) of *Halobacterium salinarum* and the isopentenyl kinase (IPK) of *M. jannaschii* were obtained from GenBank (<http://www.ncbi.nlm.nih.gov/Genbank>). Similarity searches with BLASTp (Altschul et al. 1990) were done with these representative sequences as queries against their respective domain of life. In cases where a particular enzyme was missing in KEGG for one domain, we used sequences from the other domains as queries. Similarity searches in archaea and bacteria were done against the completely sequenced genomes available in GenBank (supplementary table 1, Supplementary Material online) and completed with the ongoing genome projects (http://www.ncbi.nlm.nih.gov/sutils/genom_table.cgi). In eukaryotes, all searches were done against the non-redundant eukaryote-annotated GenBank database.

Sequences found by these searches in the three domains of life were aligned with Muscle 3.6 (Edgar 2004), and alignments were edited and manually refined with the program ED of the MUST package (Philippe 1993). When information was available about functional sites important for the biochemical function of a particular enzyme, it was used in order to identify and remove likely nonorthologous enzymes or enzymes unable to catalyze a given reaction. Redundant and partial sequences were also removed at this step.

Phylogenetic Analyses

The complete sequence data set for each enzyme of the MVA pathway was first analyzed by Neighbor-Joining (Saitou and Nei 1987) using the MUST package (Philippe 1993) to obtain preliminary phylogenetic trees in order to select some representative sequences with which to carry out more detailed maximum likelihood (ML) and Bayesian

inference (BI) phylogenetic analyses. ML tree reconstructions were done with the program TREEFINDER (Jobb et al. 2004) with the LG + Γ model (Le and Gascuel 2008) and four rate categories, which was selected as the best-fit model for all our data sets by the model selection tool implemented in TREEFINDER (Jobb et al. 2004). Node support was assessed by 1,000 bootstrap replicates with the same model. BI trees were reconstructed using the program MrBayes v. 3.0b4 (Ronquist and Huelsenbeck 2003) with a mixed substitution model and a Γ distribution of substitution rates with four categories. Searches were run with 4 chains of 1,000,000 generations for which the first 250,000 generations were discarded as “burn in,” trees being sampled every 100 generations. Stabilization of the chain parameters was verified using the program TRACER (Rambaut and Drummond 2003).

Results and Discussion

As mentioned above, the MVA pathway, described for the first time in eukaryotes, has traditionally been considered as the main way to produce IPP and DMAPP in eukaryotes and archaea, whereas the MEP pathway has been affiliated to bacteria and plastid-bearing eukaryotes (Lange et al. 2000; Kuzuyama 2002). Nevertheless, two observations raise doubts about this general assumption: archaea lack the three last enzymes of the eukaryotic MVA pathway (Smit and Mushegian 2000; Boucher et al. 2004) and some bacteria, notably some Firmicutes, do possess the genes of the MVA pathway (Boucher and Doolittle 2000; Wilding et al. 2000). In the next sections, we summarize the results of our phylogenetic analyses and propose an alternative view of the evolution of those pathways in the three domains of life.

An Alternative MVA Pathway Is Well Conserved in Archaea

We carried out molecular phylogenetic studies using a selection of representative species for all the enzymes involved in the MVA pathway except for its first enzyme, the acetoacetyl-CoA thiolase (AACT). This enzyme is not specific of this pathway, and it belongs to a large thiolase superfamily with a very complex evolutionary history that has already been studied in detail (Peretó et al. 2005). Our analyses show that all archaeal complete genome sequences encode many of the enzymes needed to accomplish the MVA pathway, with the exception of *Nanoarchaeum equitans* (table 1). This small archaeon lives in an obligatory association with the hyperthermophilic crenarchaeote *Ignicoccus hospitalis* and is known to obtain its lipids from this host (Jahn et al. 2004), which explains the absence of these biosynthetic genes.

Apart from AACT, the first three enzymes of the MVA pathway are shared by eukaryotes and archaea. Their phylogenies showed a general topology congruent with the accepted phylogeny of organisms. As an example, we show the phylogenetic tree of the HMGS responsible for the second step of the pathway (fig. 2). It supported the expected monophyly of groups down to the level of phyla or classes

Table 1. Archaeal Genera Bearing Enzymes of the MVA Pathway.

		HMGS	HMGR	MVK	PMK	MDC	IPK	IDI1	IDI2	
Crenarchaeota	<i>Aeropyrum</i>	+	I	+	–	–	+	–	+	
	<i>Hyperthermus</i>	+	I	+	–	–	+	–	+	
	<i>Staphylothermus</i>	+	II	+	–	–	+	–	+	
	<i>Ignicoccus</i>	+	I	+	–	–	+	–	+	
	<i>Metallosphaera</i>	+	I	+	+	+	+	–	+	
	<i>Sulfolobus</i>	+	I	+	+	+	§	–	+	
	<i>Caldivirga</i>	+	I	+	–	–	+	–	+	
	<i>Pyrobaculum</i>	+	I	+	–	–	+	–	+	
	<i>Thermofilum</i>	+	II	+	–	–	+	–	+	
	<i>Desulfurococcus</i>	+	II	+	–	–	+	–	+	
	<i>Thermoproteus</i>	+	I	+	–	–	+	–	+	
	Thaumarchaeota	<i>Cenarchaeum</i>	+	II	+	–	–	+	+	–
	(Crenarchaeota group I)	<i>Nitrosopumilus</i>	+	II	+	–	–	+	+	–
	Euryarchaeota	<i>Haloarcula</i>	+	I	+	–	+	+	+	–
	<i>Halobacterium</i>	+	I	+	–	+	+	+	+	
	<i>Haloquadratum</i>	+	I	+	–	+	+	+	–	
	<i>Halorubrum</i>	+	I	+	–	+	+	+	+	
	<i>Natronomonas</i>	+	I	+	–	+	+	+	+	
	<i>Natrialba</i>	+	I	+	–	+	+	+	+	
	<i>Candidatus Methanoregula</i>	+	I	+	–	–	+	–	+	
	<i>Methanospirillum</i>	+	I	+	–	–	+	–	+	
	<i>Methanocorpusculum</i>	+	I	+	–	–	+	–	+	
	<i>Methanoculleus</i>	+	I	+	–	–	+	–	+	
	<i>Methanococcoides</i>	+	I	+	–	–	+	–	+	
	<i>Methanosaeta</i>	+	II	+	–	–	+	–	+	
	<i>Methanosarcina</i>	+	I	+	–	–	+	–	+	
	<i>Archaeoglobus</i>	+	II	+	–	–	+	–	+	
	<i>Methanobrevibacter</i>	+	I	+	–	–	+	–	+	
	<i>Methanothermobacter</i>	+	I	+	–	–	+	–	+	
	<i>Methanosphaera</i>	+	I	+	–	–	+	–	+	
	<i>Methanocaldococcus</i>	+	I	+	–	–	+	–	+	
	<i>Methanococcus</i>	+	I	+	–	–	+	–	+	
	<i>Pyrococcus</i>	+	I	+	–	–	+	–	+	
	<i>Thermococcus</i>	+	I	+	–	–	+	–	+	
	<i>Methanopyrus</i>	+	I	+	–	–	+	–	+	
	<i>Nanoarchaeum</i>	–	–	–	–	–	–	–	–	
	<i>Ferroplasma</i>	+	II	?	–	–	?	–	+	
	<i>Picrophilus</i>	+	II	?	–	+	+	–	+	
	<i>Thermoplasma</i>	+	II	?	–	+	+	–	+	

(+) A homologue of the enzyme can be detected; (–) No detection of homologous sequences; (I) class I HMGR; (II) class II HMGR; (§) our analysis detected IPK homologues in all Sulfolobales but *Sulfolobus acidocaldarius* and *S. tokodaii*; (?) a homologue can be detected but it remains unclear if it is an actual orthologue of the enzyme.

with only two clear exceptions concerning the bacterial phylum Chloroflexi and the bacteroidete *Flavobacterium johnsoniae*, which branched among the archaea, supporting ancient HGT events from this domain. The HMGS phylogenetic tree also allowed the distinction of the mitochondrial and cytosolic paralogues described in metazoa (Ayte et al. 1990) and showed the intriguing very long branch basal to most archaeal HMGS sequences discussed by Boucher et al. (2004) and Jiang et al. (2008). The two following enzymes of the pathway, HMGR (fig. 3) and MVK (fig. 4), also yielded phylogenetic trees with a good global agreement with the accepted phylogeny of organisms. Notably, except for a few clear cases of HGT like the ones commented above, all these trees retrieved a well-supported separation of the eukaryotic, archaeal, and bacterial domains.

The last three enzymes of the classical MVA pathway, PMK, MDC, and IDI 1 (IDI1), did not follow this “three domain” distribution because they are absent from most

archaeal species, as noticed in previous works (Smit and Mushegian 2000; Boucher et al. 2004). Our sequence similarity searches (table 1) confirmed their absence in a larger archaeal taxonomic spectrum. The exceptions are Haloarchaea and Thermoplasmatales, which have MDC sequences, and Haloarchaea and Thaumarchaeota, which contain homologues of IDI1. However, these archaeal homologues of MDC and IDI1 branched among the bacteria, suggesting HGT events (Boucher et al. 2004) (supplementary figs. 3 and 4, Supplementary Material online). In contrast, the class Sulfolobales (Crenarchaeota) contains both PMK and MDC homologues branching with a strong support in an intermediate position between the eukaryotic and the bacterial sequences (table 1, supplementary figs. 2 and 3, Supplementary Material online). This position makes it difficult to consider them as the result of an HGT, neither from bacteria nor from eukaryotes, as it was proposed by Boucher et al. (2004). On the contrary, the intermediate position showed by these sequences

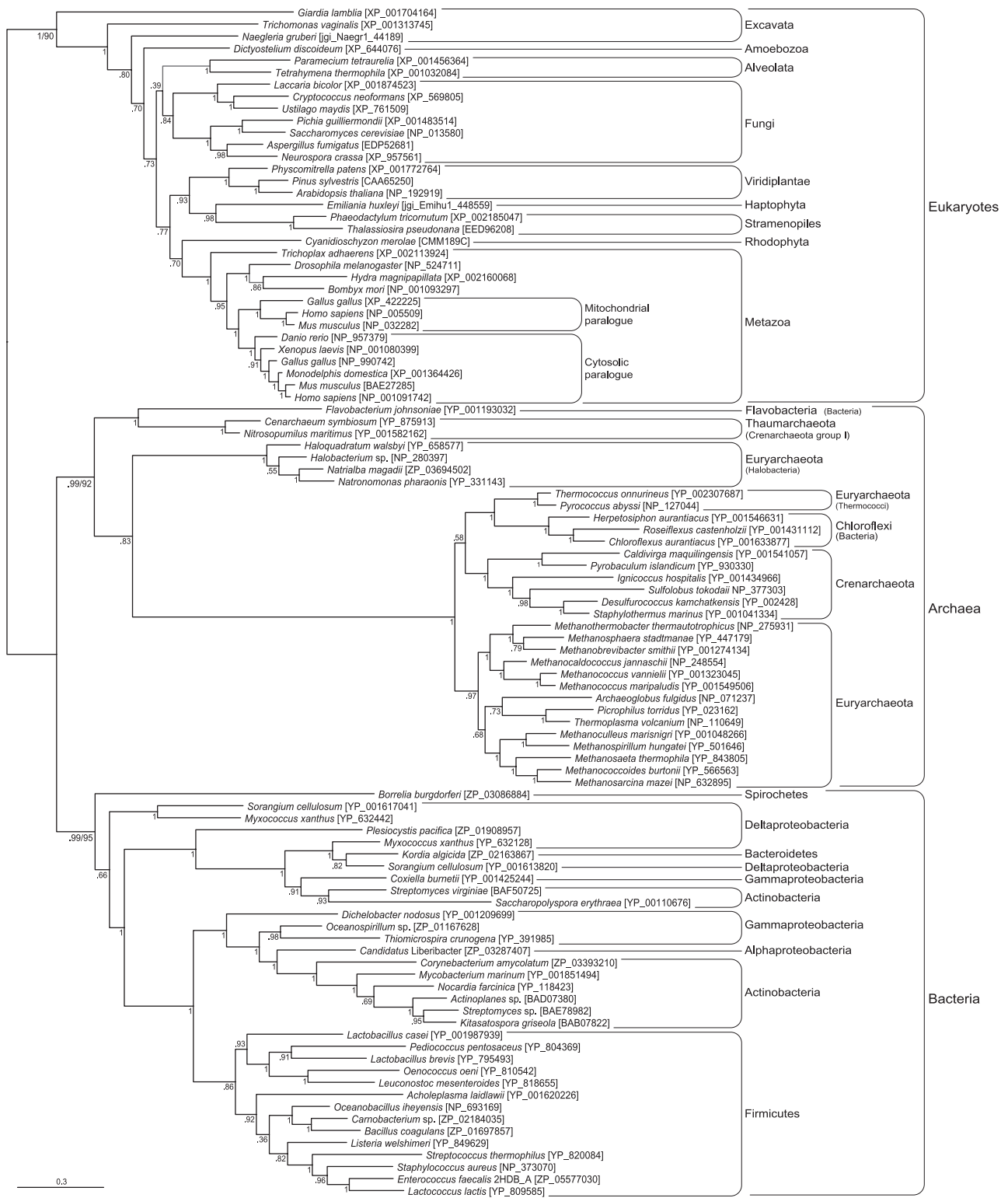


Fig. 2. Bayesian phylogenetic tree of the HMGS. This tree was reconstructed using 99 representative sequences and 311 conserved sites. Numbers at nodes are posterior probabilities. Maximum likelihood bootstrap values are also provided for nodes concerning the monophyly of major clades.

opens the possibility for an ancestral presence of PMK and MDC in archaea, subsequently conserved only in the Sulfolobales.

As mentioned above, two enzymes have been proposed to participate in an alternative pathway of conversion of

phosphomevalonate to IPP and DMAPP in archaea (Barkley et al. 2004; Grochowski et al. 2006; Dutoit et al. 2008): the IPK and the IDI 2 (IDI2). We observe that archaea containing other genes of the MVA pathway also have orthologues of IPK and IDI2, except for *Sulfolobus tokodaii*, *Cenarchaeum*



Fig. 3. Bayesian phylogenetic tree of the HMGR. This tree was reconstructed using 110 representative sequences and 274 conserved sites. Numbers at nodes are posterior probabilities. ML bootstrap values are also provided for nodes concerning the monophyly of major clades.

symbiosum, *Nitrosopumilus maritimus* and some halophilic archaea. Interestingly, these archaeal genomes lacking IDI2 have IDI1 instead, suggesting nonhomologous replacement of IDI2. The phylogeny of IDI2 supported the separation between archaeal and bacterial sequences, except for the Firmicutes, which branched among archaea, indicating a likely HGT event from this domain ([supplementary fig. 5, Supplementary Material](#) online). With the exception of 14 eukaryotes and 9 bacteria, homologues of the archaeal IPK could not be detected in the genomes of most species from the other domains of life (not shown). IPK sequences were less

conserved than those of the other MVA pathway enzymes, in particular in the case of the bacterial and eukaryotic homologues. As a result, phylogenetic trees for this enzyme were poorly supported (low bootstrap values and posterior probabilities) and less congruent with the presumed organismal phylogeny. Removal of the very divergent bacterial and eukaryotic homologues led to better alignments and more supported trees, where most of the different archaeal classes were monophyletic ([supplementary fig. 6, Supplementary Material](#) online). These results are consistent with the existence of a modified MVA pathway in archaea ([Grochowski](#)

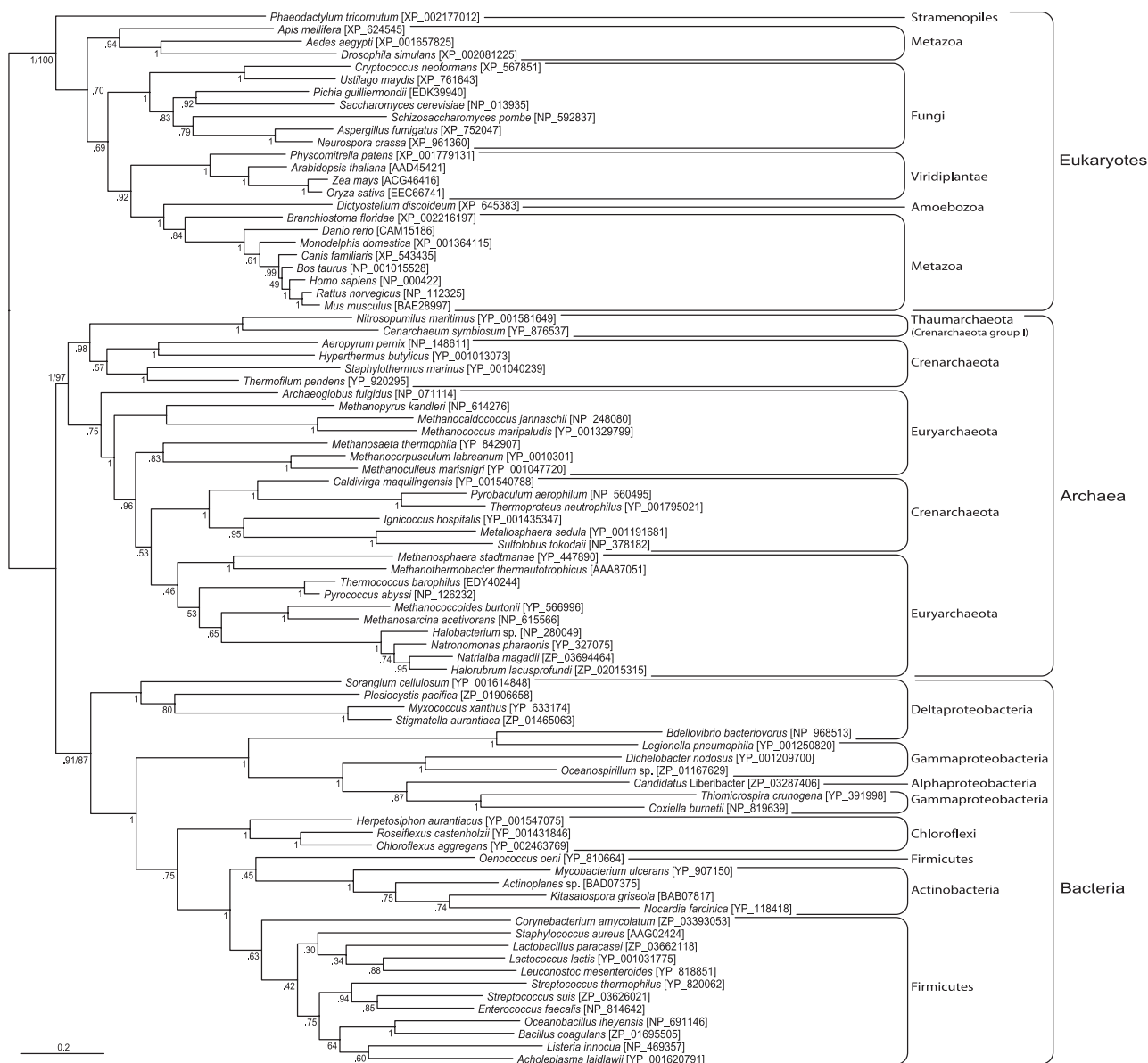


Fig. 4. Bayesian phylogenetic tree of the mevalonate kinase (MVK). This tree was reconstructed using 84 representative sequences and 232 conserved sites. Numbers at nodes are posterior probabilities. Maximum likelihood bootstrap values are also provided for nodes concerning the monophyly of major clades.

et al. 2006; Chen and Poulter 2010), and, even if the phosphomevalonate decarboxylase required to complete this metabolic route remains to be identified, our results support the alternative pathway to be extensive and characteristic of the whole archaeal domain. This suggests that the alternative pathway, including the recently characterized enzymes IPK and IDI2, was ancestral to archaea. This has implications on the classical bipartite vision of the distribution of isoprenoid biosynthesis pathways in the three domains of life. Actually, despite several sporadic cases of HGT, three ways to produce IPP and DMAPP would exist, each of which is widespread within a domain of life: the MEP pathway in bacteria, the classical MVA pathway in eukaryotes, and the alternative MVA pathway in archaea (fig. 1). The characterization of the missing archaeal phosphomevalonate decarboxylase, including its distribution and phylogeny, will be a major help

to understand the differentiation of those three pathways in the tree of life. This evolutionary history may have been rather complex because our results concerning the Sulfolobales raise the possibility that the eukaryotic-type MVA pathway was present in ancestral archaea before its replacement in most archaeal groups by the archaeal-type pathway.

An Ancestral MVA Pathway in Bacteria

As previously mentioned, even if it is generally assumed that bacteria do not possess the MVA pathway, this metabolic route has already been described in some Firmicutes (Bochar et al. 1999; Wilding et al. 2000; Romanowski et al. 2002; Voynova et al. 2004; Steussy et al. 2006). It has been recently shown that the MVA pathway may complement defects in the synthesis of IPP involved in iron-sulfur

Table 2. Bacteria Bearing Enzymes of the MVA Pathway.

		HMGS	HMGR	MVK	PMK	MDC	IPK	
Actinobacteria	<i>Corynebacterium amycolatum</i>	+	I	+	+	+	–	
	<i>Gardnerella vaginalis</i>	+	II	+	+	+	–	
	<i>Mycobacterium marinum</i>	+	I	+	+	+	–	
	<i>Nocardia farcinica</i>	+	I	+	+	+	–	
	<i>Streptomyces griseoflavus</i>	+	I	+	+	+	–	
Bacteroidetes	<i>Croceibacter atlanticus</i>	–	II	–	–	+	–	
	<i>Dokdonia donghaensis</i>	–	II	–	–	+	–	
	<i>Flavobacteria bacterium</i>	+	II	–	–	+	–	
	<i>Flavobacterium johnsoniae</i>	+	II	–	–	+	–	
	<i>Gramella forsetii</i>	–	II	–	–	+	–	
	<i>Kordia algicida</i>	+	II	–	–	+	–	
	<i>Leeuwenhoekiella blandensis</i>	–	II	–	–	+	–	
	<i>Polaribacter irgensii</i>	–	II	–	–	+	–	
	<i>Psychroflexus torques</i>	–	II	–	–	+	–	
	<i>Robiginitalea biformata</i>	–	II	–	–	+	–	
	Unidentified eubacterium	–	II	–	–	+	–	
	Firmicutes	<i>Bacillus coagulans</i>	+	II	+	+	+	–
		<i>Gemella haemolysans</i>	+	II	+	+	+	–
<i>Listeria grayi</i>		+	II	+	+	+	–	
<i>Oceanobacillus iheyensis</i>		+	II	+	+	+	–	
<i>Staphylococcus capitis</i>		+	II	+	+	+	–	
<i>Catonella morbi</i>		+	II	+	+	+	–	
<i>Carnobacterium</i> sp.		+	II	+	+	+	–	
<i>Enterococcus faecalis</i>		+	II	+	+	+	–	
<i>Lactobacillus johnsonii</i>		+	II	+	+	+	–	
<i>Lactococcus lactis</i>		+	II	+	+	+	–	
<i>Leuconostoc citreum</i>		+	II	+	+	+	–	
<i>Oenococcus oeni</i>		+	II	+	+	+	–	
<i>Pediococcus pentosaceus</i>		+	II	+	+	+	–	
<i>Streptococcus agalactiae</i>		+	II	+	+	+	–	
<i>Acholeplasma laidlawii</i>		+	II	+	+	+	–	
Chloroflexi	<i>Chloroflexus aggregans</i>	+	II	+	–	+	+	
	<i>Herpetosiphon aurantiacus</i>	+	II	+	–	+	+	
	<i>Roseiflexus castenholzii</i>	+	II	+	–	+	+	
Proteobacteria	Rhodobacteraceae bacterium	+	I	+	+	+	–	
	<i>Candidatus Liberibacter</i>	+	I	+	+	+	–	
	<i>Bdellovibrio bacteriovorus</i>	–	II	+	+	+	–	
	<i>Myxococcus xanthus</i>	+	II	+	+	+	–	
	<i>Plesiocystis pacifica</i>	+	II	+	–	+	–	
	<i>Sorangium cellulosum</i>	+	II	+	+	+	–	
	<i>Stigmatella aurantiaca</i>	+	II	+	+	+	–	
	<i>Coxiella burnetii</i>	+	I	+	+	+	–	
	<i>Dichelobacter nodosus</i>	+	I	+	+	+	–	
	<i>Thiomicrospira crunogena</i>	+	I	+	+	+	–	
	<i>Borrelia burgdorferi</i>	+	II	+	?	+	–	

This table summarizes the 93 species where homologous sequences of MVA pathway enzymes were systematically detected in 43 bacterial representatives of their respective genera. (+) A homologue of the enzyme can be detected; (–) no detection of homologous sequences; (I) class I HMGR; (II) class II HMGR; (?) a homologue can be detected but it remains unclear if it is an actual orthologue of the enzyme.

protein biogenesis in certain bacteria (Vinella et al. 2009), but the major role of this pathway in Firmicutes is related to isoprenoid biosynthesis (Wilding et al. 2000; Voynova et al. 2004). Contrary to previous studies where the presence of this pathway seemed to be limited to a very reduced range of bacterial taxa, especially Firmicutes (Lange et al. 2000; Boucher et al. 2004), the current availability of a larger sampling of complete genome sequences has allowed us to find that its occurrence appears far from being restricted to this bacterial phylum. In fact, we found homologues of the MVA pathway enzymes in 93 bacterial species belonging to 43 different genera of 6 phyla (table 2). In particular, Firmicutes, Actinobacteria, Proteobacteria, and the Spirochete genus *Borrelia* encode in their genomes

most of the enzymes of this pathway. In addition, Bacteroidetes possess only some of the enzymes, and the Chloroflexi contain IPK instead of PMK. The case of the HMGR will be discussed below, but we can highlight the existence of two different classes of enzymes, one present in some Proteobacteria and Actinobacteria and the other widespread in the six MVA pathway-bearing bacterial phyla.

The distribution of the isopentenyl isomerases IDI1 and IDI2 is much more complex. The reaction carried out by the IDI enzymes is necessary in the MVA pathway to produce DMAPP from IPP, but it is not compulsory in the MEP pathway because the last enzyme of this pathway (HMGR—4-hydroxy-3-methylbut-2-enyl diphosphate reductase) catalyzes the production of both IPP and DMAPP

at the same time (Rohdich et al. 2001, 2002). However, IDI genes are present in the genomes of a variety of bacteria, probably because they may allow these organisms to isomerize IPP and DMAPP according to their metabolic requirements at a given moment. Thus, although IDI1 and IDI2 are characteristic of eukaryotes and archaea, respectively, the distribution of these enzymes among bacteria is much less clear-cut. Bacteroidetes, most actinobacteria, and some proteobacterial species have IDI1, whereas Firmicutes, Cyanobacteria, Chlorobi, Chloroflexi, *Deinococcus*–*Thermus*, Spirochetes, Deltaproteobacteria, and some other proteobacterial and actinobacterial species contain IDI2 (supplementary figs. 4 and 5, Supplementary Material online).

One of the most intriguing phylogenies is the one of HMGR. There are two different homologous classes of this enzyme (Scher and Rodwell 1989; Bochar et al. 1999; Friesen and Rodwell 2004; Hedl et al. 2004). The class I HMGR is present in eukaryotes, archaea, and several proteobacteria and actinobacteria, whereas the class II is characteristic of most bacteria and a small number of archaea. Biochemical studies of HMGR showed that both classes share similar active sites and form homodimers. However, the respective function of each monomer is different from one class to the other, the two classes do not use exactly the same cofactors, and they are differently inhibited by statins. Because class II HMGR is widespread and characteristic of many bacterial phyla, it may be considered as ancestral to the bacterial domain. The clade of class II HMGR also contains a few archaeal representatives, which were proposed to have received their genes by HGT from bacteria (Boucher et al. 2001). With respect to class I bacterial sequences, they clearly branch within the class I clade despite their divergence (fig. 3), which was interpreted as a complex history of HGTs and specialization leading these sequences to be present in several MVA pathway-lacking organisms (Gophna et al. 2006). Nevertheless, these bacterial sequences did not branch within the eukaryotic or archaeal clades, as could be expected if they had been acquired by recent HGT from one of these domains. Intriguingly, a second HMGR copy found in the haloarchaeon *Natronomonas pharaonis* branched sister to this bacterial clade far from the other haloarchaeal sequences (which include a second copy of this same species).

In summary, the phylogenetic analyses of the eukaryotic-like MVA pathway enzymes in a large taxonomic sampling produced topologies supporting the monophyly of major groups (figs. 2–4 and supplementary figs. 2–5, Supplementary Material online). In particular, this includes the emergence of the bacterial sequences as a monophyletic group distinct from archaea and eukaryotes (i.e., the three domains topology). In fact, for each enzyme, the vast majority of bacterial sequences form an independent monophyletic group, as, for example, in the phylogenies of HMGS (fig. 2) and MVK (fig. 4). If, as mentioned above, bacteria had obtained their MVA pathway enzymes by HGT as traditionally assumed (Boucher and Doolittle 2000; Wilding et al. 2000), all the bacterial sequences should branch among those of another domain of life and share similar biochemical prop-

erties with their donors. On the contrary, most bacterial sequences for each enzyme form monophyletic groups separated from the archaeal and eukaryotic clades, and, when well characterized biochemically, they have their own sequence signatures and biochemical characteristics. Our phylogenetic analyses only revealed the expected pattern of HGTs in the case of the HMGS of Chloroflexi and *F. johnsoniae* and the IDI2 of Firmicutes branching within the archaeal clade (fig. 2 and supplementary fig. 5, Supplementary Material online), which supports that these bacteria obtained at least part of their MVA pathway from archaea. Conversely, we also observed cases of HGT from bacteria to archaea, for example, the class II HMGR of certain euryarchaeota and thaumarchaeota, which emerged well nested within the bacterial clade (fig. 3). Phylogenetic trees constructed using only the bacterial sequences yielded topologies supporting the monophyly of several bacterial phyla, which argued against the hypothesis of multiple HGT events among distantly related bacteria (data not shown).

Three different scenarios can be suggested to explain the segregation of bacterial sequences from those of archaea and eukaryotes observed in the phylogenetic trees of the MVA pathway enzymes. The first one would entail that an ancient bacterial lineage obtained all the enzymes by HGT from another domain of life, archaea or eukaryotes, followed by a period of fast evolution that would be responsible of the high divergence of the bacterial enzymes from the original ones. Then, new HGT events would have occurred independently from this bacterial lineage toward all the current MVA pathway-bearing bacterial classes (fig. 5A). The other two possibilities would imply that the MVA pathway was ancestral to bacteria, which would be the consequence either of the acquisition by an ancestral bacterium of the pathway from another domain before the radiation of different bacterial phyla (fig. 5B) or of the direct vertical inheritance of an MVA pathway already present in the last common ancestor of all organisms, the cenancestor (fig. 5C). These two latter hypotheses imply a massive loss of the MVA pathway in many bacterial lineages, but this can easily be explained by the functional redundancy with the MEP pathway, which is present in most bacteria.

Conclusions: The Early Origin of the MVA Pathway

The distribution of the different isoprenoid biosynthesis pathways in the three domains of life is very complex but displays a clear preferential relationship: the MEP pathway in bacteria, the classical MVA pathway in eukaryotes, and the alternative MVA pathway in archaea. Our work supports that each route is ancestral to one of the domains, but it is not enough to decide whether these pathways were present in the cenancestor or not. If we exclude the extremely improbable possibility that the pathway was assembled in the three domains by independent recruitment of homologous enzymes carrying out other functions, three other possibilities concerning this question can

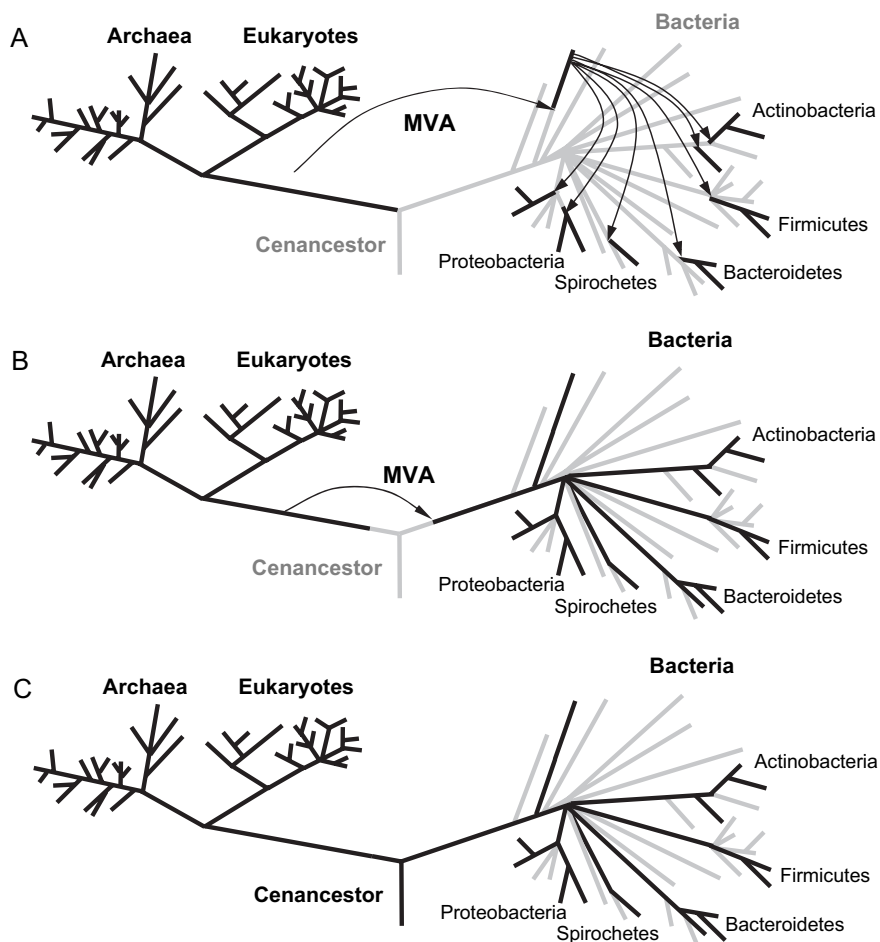


Fig. 5. Hypotheses for the origin of the MVA pathway of isoprenoid synthesis in bacteria. (A) Transfer of the complete MVA pathway from eukaryotic or archaeal donors to a bacterial lineage followed by multiple transfers from this lineage to all the bacterial groups currently possessing the MVA pathway (black branches). (B) Ancient transfer of the complete MVA pathway to a bacterial ancestor before the divergence of the contemporary bacterial phyla. (C) Inheritance of the MVA pathway from the cenancestor. Scenarios (B and C) entail multiple losses of the MVA pathway in several bacterial phyla (gray branches). For convenience, the classical bacterial rooting of the tree of life has been adopted.

be proposed. The first one would be the lack of any isoprenoid biosynthesis pathway in the cenancestor (fig. 6A), an idea that has already been advanced and used by some authors as an argument to discard the presence of lipid membranes in the cenancestor (Martin and Russell 2003). This is, however, at odds with the present-day ubiquitous presence of isoprenoids in the three domains of life. The mirror hypothesis would be the presence of ancestral forms of the three pathways in the cenancestor followed by selective losses in each domain of life (fig. 6B). Nevertheless, the MEP pathway is completely absent in archaea and in all non-plastid-bearing eukaryotes, so we can assume that it was not present in the ancestors of these two domains.

In the light of our results, a more parsimonious third hypothesis emerges: the classical MVA pathway would have been present in the cenancestor and inherited not only by archaea and eukaryotes but also by bacteria, which would have replaced it by the MEP pathway in a variety of bacterial phyla (fig. 6C). As we said previously, it is even possible that the last common ancestor of archaea also had the entire classical MVA pathway, including the last enzymes (PMK and MDC) that have been lost in the vast

majority of archaeal species, but not in Sulfolobales. Thus, we can speculate that the classical MVA pathway was present in the cenancestor and replaced or modified in most bacteria and archaea by their respective current characteristic pathways. If this hypothesis turned out to be correct, the MVA pathway would represent a rare case where a metabolic pathway would have retained more ancestral characteristics in eukaryotes than in bacteria and most archaea (the Sulfolobales being an exception among the available archaeal genomes, as mentioned above).

The possibility of an isoprenoid biosynthesis pathway operating in the cenancestor has interesting implications on the open question about the presence and nature of the membrane of this ancestral organism. Indeed, the differences between the membranes of archaea, on the one hand, and those shared by eukaryotes and bacteria, on the other hand, have led to a hot controversy on the nature of the cenancestral membrane. The major differences are the presence of fatty acid lateral chains bound by ester links to *sn*-glycerol-3-phosphate (G3P) in eukaryotes and bacteria, in contrast with archaeal membranes containing isoprenoid lateral chains and *sn*-glycerol-1-phosphate (G1P)

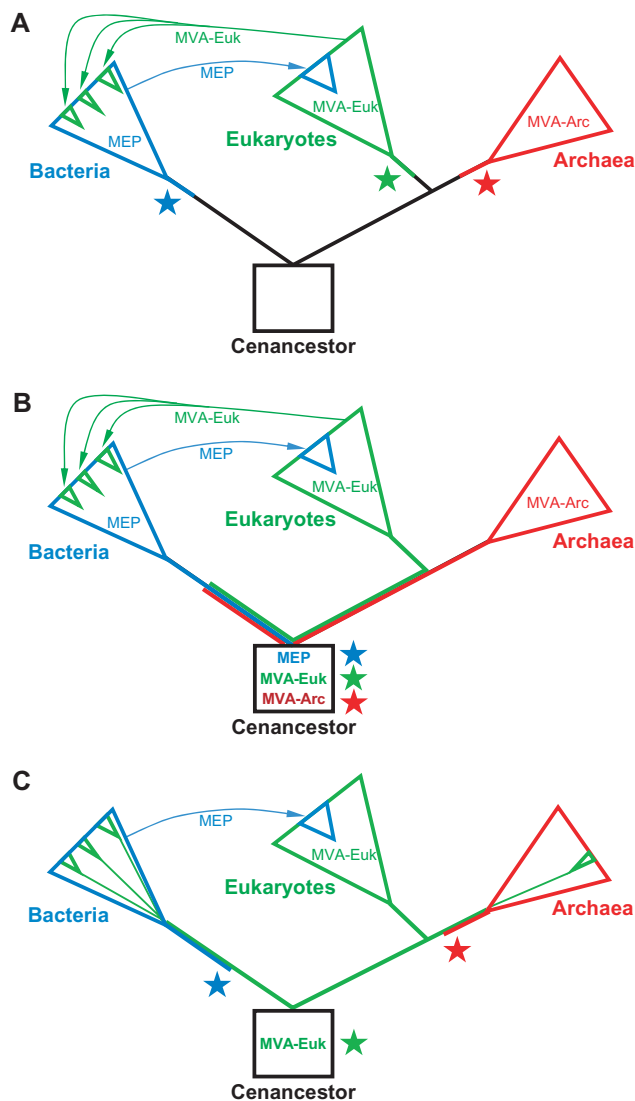


FIG. 6. Hypotheses for the evolution of the different isoprenoid biosynthesis pathways on a schematic three domains of life tree. The MEP pathway is indicated in blue, and the eukaryotic and archaeal MVA (MVA-Euk and MVA-Arc) pathways are indicated in green and red, respectively. Stars filled with these colors indicate the origin of the respective pathways. (A) Each pathway would have appeared in the ancestors of each domain of life. Late HGT events would explain the presence of MEP and MVA-Euk pathways in several eukaryotes and bacteria, respectively. (B) The three pathways would have been present in the cenancestor, but only one pathway would have been conserved in each domain of life. Presence of supplementary pathways would be explained by HGT acquisitions. (C) The classical MVA pathway would have been present in the cenancestor and replaced in most bacteria and archaea by the MEP and MVA-Arc pathways, respectively. This figure shows the classical rooting for convenience, but the hypothesis remains valid if eukaryotes are considered as derived from a symbiotic event between archaea and bacteria.

bound by ether links. Although there are exceptions concerning the nature of the lateral chains (fatty acids or isoprenoids) and the type of link (ester or ether), the glycerol-phosphate stereoisomer is always different between archaea (G1P) and bacteria and eukaryotes (G3P). This, together with the apparent nonhomologous biosynthesis of

isoprenoids in archaea and bacteria, has led some authors to propose either that the cenancestor was acellular, namely, that it was not bounded by any type of membrane (Koga et al. 1998), or that it was surrounded by mineral membranes instead of lipid ones (Martin and Russell 2003; Mulikjanian et al. 2009). Nevertheless, fatty acids have been shown to be present in archaea in variable proportions (Gattinger et al. 2002) and homologues of the bacterial and eukaryotic genes encoding the enzymes involved in fatty acid biosynthesis and degradation are present in most archaeal genome sequences, suggesting that they are ancestral (Peretó et al. 2004). In addition, Peretó et al. (2004) also showed that the dehydrogenases involved in the synthesis of G1P and G3P in archaea and bacteria evolved from universally spread enzymatic families that were most likely present in the cenancestor. The different glycerol stereoisomers would have been selected afterward in the different lineages. Our results, supporting the possibility of an ancestral MVA pathway, provide an additional argument for the hypothesis that all the major components of contemporary membranes (phospholipids composed of glycerol-phosphate bound to lateral chains of fatty acids and/or isoprenoids) existed in the cenancestor and that the specialization observed in current cellular membranes is most likely secondary and linked to the divergence of the three domains of life.

Supplementary Material

Supplementary table 1 and figures 1–6 are available at *Molecular Biology and Evolution* online <http://mbe.oxfordjournals.org/>.

Acknowledgments

This work was supported by the interdisciplinary programs Origine des Planètes et de la Vie and InTerrVie (Interactions Terre—Vie) of the French Centre National de la Recherche Scientifique and Institut National des Sciences de l'Univers. J.L. is recipient of a PhD fellowship of the French Research Ministry. We thank Puri López-García for critical reading of the manuscript and three anonymous reviewers for helpful comments.

References

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol.* 215:403–410.
- Ayte J, Gil-Gomez G, Haro D, Marrero PF, Hegardt FG. 1990. Rat mitochondrial and cytosolic 3-hydroxy-3-methylglutaryl-CoA synthases are encoded by two different genes. *Proc Natl Acad Sci U S A.* 87:3874–3878.
- Barkley SJ, Cornish RM, Poulter CD. 2004. Identification of an archaeal type II isopentenyl diphosphate isomerase in *Methanothermobacter thermautotrophicus*. *J Bacteriol.* 186:1811–1817.
- Bochar DA, Stauffacher CV, Rodwell VW. 1999. Sequence comparisons reveal two classes of 3-hydroxy-3-methylglutaryl coenzyme A reductase. *Mol Genet Metab.* 66:122–127.
- Boucher Y, Doolittle WF. 2000. The role of lateral gene transfer in the evolution of isoprenoid biosynthesis pathways. *Mol Microbiol.* 37:703–716.

- Boucher Y, Huber H, L'Haridon S, Stetter KO, Doolittle WF. 2001. Bacterial origin for the isoprenoid biosynthesis enzyme HMG-CoA reductase of the archaeal orders Thermoplasmatales and Archaeoglobales. *Mol Biol Evol.* 18:1378–1388.
- Boucher Y, Kamekura M, Doolittle WF. 2004. Origins and evolution of isoprenoid lipid biosynthesis in archaea. *Mol Microbiol.* 52:515–527.
- Brinkman FS, Blanchard JL, Cherkasov A, et al. (14 co-authors). 2002. Evidence that plant-like genes in *Chlamydia* species reflect an ancestral relationship between Chlamydiaceae, cyanobacteria, and the chloroplast. *Genome Res.* 12:1159–1167.
- Chen M, Poulter CD. 2010. Characterization of thermophilic archaeal isopentenyl phosphate kinases. *Biochemistry* 49:207–217.
- Dutoit R, de Ruyck J, Durisotti V, Legrain C, Jacobs E, Wouters J. 2008. Overexpression, physicochemical characterization, and modeling of a hyperthermophilic *Pyrococcus furiosus* type 2 IPP isomerase. *Proteins* 71:1699–1707.
- Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32:1792–1797.
- Friesen JA, Rodwell VW. 2004. The 3-hydroxy-3-methylglutaryl coenzyme-A (HMG-CoA) reductases. *Genome Biol.* 5:248.
- Gattinger A, Schlöter M, Munch JC. 2002. Phospholipid etherlipid and phospholipid fatty acid fingerprints in selected euryarchaeotal monocultures for taxonomic profiling. *FEMS Microbiol Lett.* 213:133–139.
- Gophna U, Thompson JR, Boucher Y, Doolittle WF. 2006. Complex histories of genes encoding 3-hydroxy-3-methylglutaryl-coenzyme reductase. *Mol Biol Evol.* 23:168–178.
- Grochowski LL, Xu H, White RH. 2006. *Methanocaldococcus jannaschii* uses a modified mevalonate pathway for biosynthesis of isopentenyl diphosphate. *J Bacteriol.* 188:3192–3198.
- Hedl M, Taberner L, Stauffacher CV, Rodwell VW. 2004. Class II 3-hydroxy-3-methylglutaryl coenzyme A reductases. *J Bacteriol.* 186:1927–1932.
- Herdendorf TJ, Miziorko HM. 2006. Phosphomevalonate kinase: functional investigation of the recombinant human enzyme. *Biochemistry* 45:3235–3242.
- Houten SM, Waterham HR. 2001. Nonorthologous gene displacement of phosphomevalonate kinase. *Mol Genet Metab.* 72:273–276.
- Jahn U, Summons R, Sturt H, Grosjean H, Huber H. 2004. Composition of the lipids of *Nanoarchaeum equitans* and their origin from its host *Ignicoccus* sp. strain KIN4/l. *Arch Microbiol.* 182:404–413.
- Jiang C, Kim SY, Suh DY. 2008. Divergent evolution of the thiolase superfamily and chalcone synthase family. *Mol Phylogenet Evol.* 49:691–701.
- Jobb G, von Haeseler A, Strimmer K. 2004. TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics. *BMC Evol Biol.* 4:18.
- Kaneda K, Kuzuyama T, Takagi M, Hayakawa Y, Seto H. 2001. An unusual isopentenyl diphosphate isomerase found in the mevalonate pathway gene cluster from *Streptomyces* sp. strain CL190. *Proc Natl Acad Sci U S A.* 98:932–937.
- Koga Y, Kyuragi T, Nishihara M, Sone N. 1998. Did archaeal and bacterial cells arise independently from noncellular precursors? A hypothesis stating that the advent of membrane phospholipid with enantiomeric glycerophosphate backbones caused the separation of the two lines of descent. *J Mol Evol.* 46:54–63.
- Kuzuyama T. 2002. Mevalonate and nonmevalonate pathways for the biosynthesis of isoprene units. *Biosci Biotechnol Biochem.* 66:1619–1627.
- Lange BM, Rujan T, Martin W, Croteau R. 2000. Isoprenoid biosynthesis: the evolution of two ancient and distinct pathways across genomes. *Proc Natl Acad Sci U S A.* 97:13172–13177.
- Le SQ, Gascuel O. 2008. An improved general amino acid replacement matrix. *Mol Biol Evol.* 25:1307–1320.
- Martin W, Russell MJ. 2003. On the origins of cells: a hypothesis for the evolutionary transitions from abiotic geochemistry to chemoautotrophic prokaryotes, and from prokaryotes to nucleated cells. *Philos Trans R Soc Lond B Biol Sci.* 358:59–83.
- Matsuzaki M, Kuroiwa H, Kuroiwa T, Kita K, Nozaki H. 2008. A cryptic algal group unveiled: a plastid biosynthesis pathway in the oyster parasite *Perkinsus marinus*. *Mol Biol Evol.* 25:1167–1179.
- McGarvey DJ, Croteau R. 1995. Terpenoid metabolism. *Plant Cell.* 7:1015–1026.
- Moustafa A, Reyes-Prieto A, Bhattacharya D. 2008. Chlamydiae has contributed at least 55 genes to Plantae with predominantly plastid functions. *PLoS ONE.* 3:e2205.
- Mulkidjanian AY, Galperin MY, Koonin EV. 2009. Co-evolution of primordial membranes and membrane proteins. *Trends Biochem Sci.* 34:206–215.
- Peretó J, López-García P, Moreira D. 2004. Ancestral lipid biosynthesis and early membrane evolution. *Trends Biochem Sci.* 29:469–477.
- Peretó J, López-García P, Moreira D. 2005. Phylogenetic analysis of eukaryotic thiolases suggests multiple proteobacterial origins. *J Mol Evol.* 61:65–74.
- Philippe H. 1993. MUST, a computer package of Management Utilities for Sequences and Trees. *Nucleic Acids Res.* 21:5264–5272.
- Phillips MA, Leon P, Boronat A, Rodríguez-Concepción M. 2008. The plastidial MEP pathway: unified nomenclature and resources. *Trends Plant Sci.* 13:619–623.
- Rambaut A, Drummond AJ. 2003. Tracer: a program for analysing results from Bayesian MCMC programs such as BEAST & MrBayes. Distributed by the authors. Edinburgh (United Kingdom) Institute of Evolutionary Biology, University of Edinburgh. Available from: <http://evolve.zoo.ox.ac.uk/software.html?id=tracer>.
- Rodríguez-Concepción M, Boronat A. 2002. Elucidation of the methylerythritol phosphate pathway for isoprenoid biosynthesis in bacteria and plastids. A metabolic milestone achieved through genomics. *Plant Physiol.* 130:1079–1089.
- Rohdich F, Hecht S, Gartner K, Adam P, Krieger C, Amslinger S, Arigoni D, Bacher A, Eisenreich W. 2002. Studies on the nonmevalonate terpene biosynthetic pathway: metabolic role of IspH (LytB) protein. *Proc Natl Acad Sci U S A.* 99:1158–1163.
- Rohdich F, Kis K, Bacher A, Eisenreich W. 2001. The non-mevalonate pathway of isoprenoids: genes, enzymes and intermediates. *Curr Opin Chem Biol.* 5:535–540.
- Romanowski MJ, Bonanno JB, Burley SK. 2002. Crystal structure of the *Streptococcus pneumoniae* phosphomevalonate kinase, a member of the GHMP kinase superfamily. *Proteins* 47:568–571.
- Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572–1574.
- Ruzicka L. 1953. The isoprene rule and the biogenesis of terpenic compounds. *Experientia* 9:357–367.
- Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol.* 4:406–425.
- Scher DS, Rodwell VW. 1989. 3-Hydroxy-3-methylglutaryl coenzyme A lyase from *Pseudomonas mevalonii*. *Biochim Biophys Acta.* 1003:321–326.
- Smit A, Mushegian A. 2000. Biosynthesis of isoprenoids via mevalonate in Archaea: the lost pathway. *Genome Res.* 10:1468–1484.
- Steussy CN, Robison AD, Tetrack AM, Knight JT, Rodwell VW, Stauffacher CV, Sutherland AL. 2006. A structural limitation on enzyme activity: the case of HMG-CoA synthase. *Biochemistry* 45:14407–14414.

- Vinella D, Brochier-Armanet C, Loiseau L, Talla E, Barras F. 2009. Iron-sulfur (Fe/S) protein biogenesis: phylogenomic and genetic studies of A-type carriers. *PLoS Genet.* 5:e1000497.
- Voynova NE, Rios SE, Miziorko HM. 2004. *Staphylococcus aureus* mevalonate kinase: isolation and characterization of an enzyme of the isoprenoid biosynthetic pathway. *J Bacteriol.* 186:61–67.
- Wilding EI, Brown JR, Bryant AP, Chalker AF, Holmes DJ, Ingraham KA, Iordanescu S, So CY, Rosenberg M, Gwynn MN. 2000. Identification, evolution, and essentiality of the mevalonate pathway for isopentenyl diphosphate biosynthesis in gram-positive cocci. *J Bacteriol.* 182:4319–4327.
- Zhou D, White RH. 1991. Early steps of isoprenoid biosynthesis in *Escherichia coli*. *Biochem J.* 273:627–634.

ANNEXE 3.

Early evolution of the biotin-dependent carboxylase family.

Lombard, J. et D. Moreira. 2011.

BMC evolutionary biology **11**:232.

Correction : *BMC Evolutionary Biology* 2012, **12**:117

RESEARCH ARTICLE

Open Access

Early evolution of the biotin-dependent carboxylase family

Jonathan Lombard and David Moreira*

Abstract

Background: Biotin-dependent carboxylases are a diverse family of carboxylating enzymes widespread in the three domains of life, and thus thought to be very ancient. This family includes enzymes that carboxylate acetyl-CoA, propionyl-CoA, methylcrotonyl-CoA, geranyl-CoA, acyl-CoA, pyruvate and urea. They share a common catalytic mechanism involving a biotin carboxylase domain, which fixes a CO₂ molecule on a biotin carboxyl carrier peptide, and a carboxyl transferase domain, which transfers the CO₂ moiety to the specific substrate of each enzyme. Despite this overall similarity, biotin-dependent carboxylases from the three domains of life carrying their reaction on different substrates adopt very diverse protein domain arrangements. This has made difficult the resolution of their evolutionary history up to now.

Results: Taking advantage of the availability of a large amount of genomic data, we have carried out phylogenomic analyses to get new insights on the ancient evolution of the biotin-dependent carboxylases. This allowed us to infer the set of enzymes present in the last common ancestor of each domain of life and in the last common ancestor of all living organisms (the cenancestor). Our results suggest that the last common archaeal ancestor had two biotin-dependent carboxylases, whereas the last common bacterial ancestor had three. One of these biotin-dependent carboxylases ancestral to Bacteria most likely belonged to a large family, the CoA-bearing-substrate carboxylases, that we define here according to protein domain composition and phylogenetic analysis. Eukaryotes most likely acquired their biotin-dependent carboxylases through the mitochondrial and plastid endosymbioses as well as from other unknown bacterial donors. Finally, phylogenetic analyses support previous suggestions about the existence of an ancient bifunctional biotin-protein ligase bound to a regulatory transcription factor.

Conclusions: The most parsimonious scenario for the early evolution of the biotin-dependent carboxylases, supported by the study of protein domain composition and phylogenomic analyses, entails that the cenancestor possessed two different carboxylases able to carry out the specific carboxylation of pyruvate and the non-specific carboxylation of several CoA-bearing substrates, respectively. These enzymes may have been able to participate in very diverse metabolic pathways in the cenancestor, such as in ancestral versions of fatty acid biosynthesis, anaplerosis, gluconeogenesis and the autotrophic fixation of CO₂.

Background

Biotin-dependent carboxylases are a group of enzymes present in the three domains of life (Archaea, Bacteria and Eucarya) able to catalyze the fixation of CO₂ on different specific substrates. They participate in many essential metabolic functions as diverse as the autotrophic fixation of CO₂, the biosynthesis and degradation of fatty acids, the gluconeogenesis, the anaplerotic

production of oxaloacetate or the degradation of some amino acids [1-10]. These enzymes belong to the larger biotin-enzyme family that also contains some biotin-dependent decarboxylases and transcarboxylases [11]. Members of the biotin-enzyme family share functional domains and reaction mechanisms and are characterized by their dependence on covalently bound biotin as a cofactor. Biotin, also called vitamin H, is a prosthetic group made up of a valerate side chain attached to a bicyclic ring consisting of one ureido and one thiophan

* Correspondence: david.moreira@u-psud.fr

Unité d'Ecologie, Systématique et Evolution, UMR CNRS 8079, Univ. Paris-Sud, 91405 Orsay Cedex, France

rings (Figure 1). In this work, we focus on the evolution of the biotin-dependent carboxylases.

The biotin-dependent carboxylase family includes the acetyl-CoA carboxylases, the propionyl-CoA carboxylases, the methylcrotonyl-CoA carboxylases, the geranyl-CoA carboxylases, the acyl-CoA carboxylases, the pyruvate carboxylases and the urea carboxylases. They share a common catalytic mechanism and three functional components: a biotin carboxyl carrier protein (BCCP), a biotin

carboxylase (BC) domain and a carboxyl transferase (CT) domain [11]. BCCP is the biotinylated element, which is also shared with the decarboxylases and transcarboxylases of the large biotin-enzyme family. BC catalyzes the ATP-dependent fixation of CO₂ to the BCCP-bound biotin, and thus the intermediate formation of carboxybiotin. Finally, CT binds the carboxyl group from the carboxybiotin to the specific substrate of each carboxylase (Figure 1).

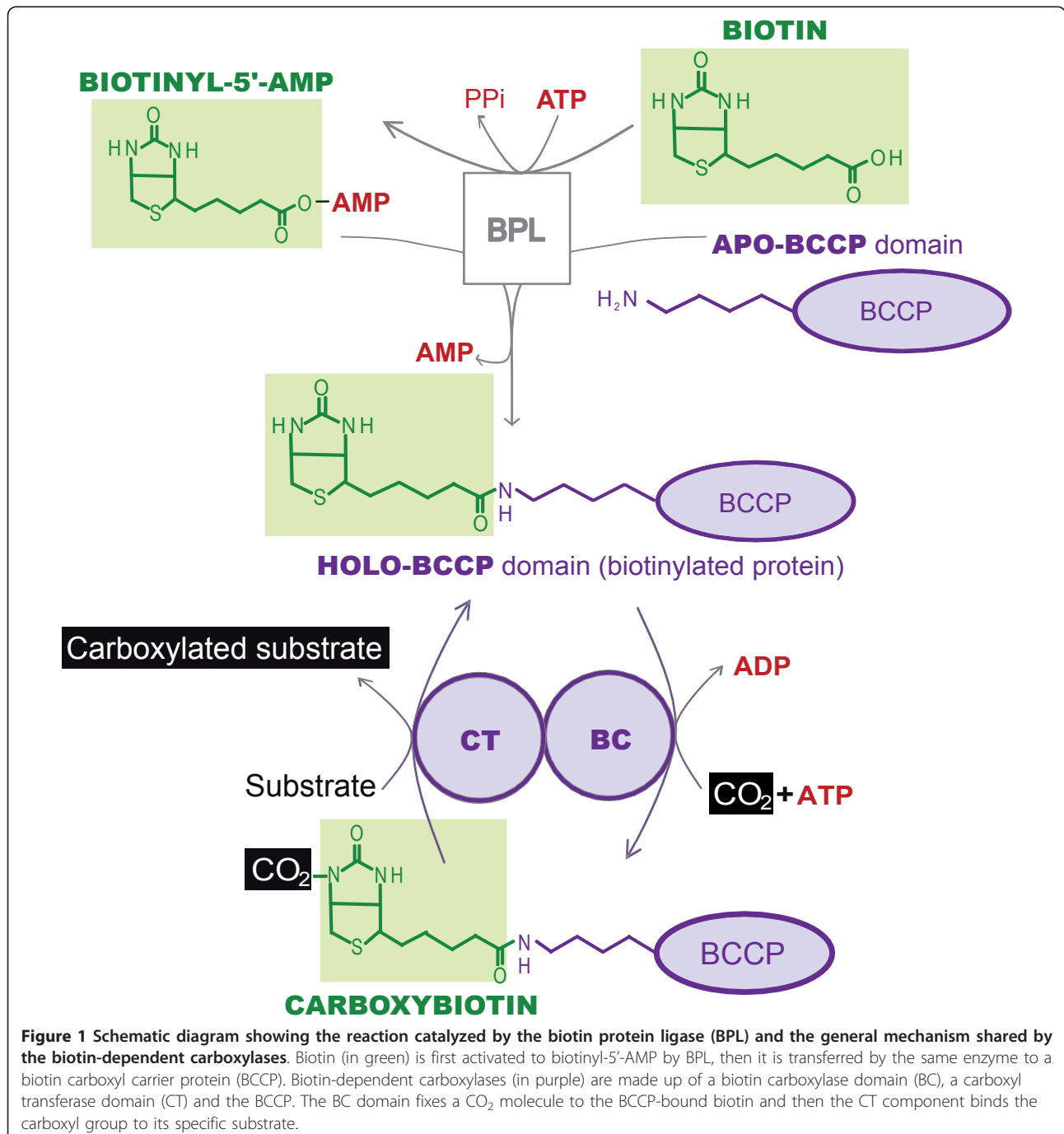


Figure 1 Schematic diagram showing the reaction catalyzed by the biotin protein ligase (BPL) and the general mechanism shared by the biotin-dependent carboxylases. Biotin (in green) is first activated to biotinyl-5'-AMP by BPL, then it is transferred by the same enzyme to a biotin carboxyl carrier protein (BCCP). Biotin-dependent carboxylases (in purple) are made up of a biotin carboxylase domain (BC), a carboxyl transferase domain (CT) and the BCCP. The BC domain fixes a CO₂ molecule to the BCCP-bound biotin and then the CT component binds the carboxyl group to its specific substrate.

Before the carboxylation reaction can occur, an independent protein called biotin protein ligase (BPL, also named holocarboxylase synthetase or BirA; E.C. 6.3.4.9; 6.3.4.10; 6.3.4.11; 6.3.4.15) is in charge of the post-transcriptionally attachment of the biotin to a specific conserved lysine residue of the BCCP [12,13]. BPL first adenylates the carboxyl group of the valerate chain of the biotin molecule, then the resulting biotinyl-5'-AMP molecule is used to transfer the biotin moiety to the ϵ -amino group of the lysine residue of BCCP [14,15]. As each carboxylase has its own BCCP counterpart, it was first thought that each BCCP had one specific BPL devoted to its biotinylation, but subsequent studies have shown that each organism has usually only one BPL protein able to biotinylate the BCCP counterparts from different carboxylases. BPL has even been reported to be able to biotinylate heterologous BCCP elements from other species [12,16,17]. Four types of BPL have been described (Figure 2A, [18]). In prokaryotes, there is one monofunctional sequence bearing only the BPL function [14] and one bifunctional BPL-regulatory gene (BirA). The latter is known to carry, in addition to the BPL catalytic domain, an N-terminus regulatory domain that participates in the transcriptional control of genes involved in biotin biosynthesis [19,20]. Structural comparison and meticulous similarity searches have shown that the BPL biotinylating domain is related to diverse enzymes, including lipoyl protein ligases (LPLs), asparagine synthetases and class II aminoacyl tRNA synthetases [21]. The BPL catalytic domain and the regulatory N-terminal motif have been proposed to have emerged and fused before the radiation of Archaea and Bacteria [22-24]. Concerning the eukaryotes, the N-terminal halves of their BPLs are very diverse: plant BPL contains an N-terminal domain of unknown function and, in contrast with the bacterial N-terminal motif, unable to control the expression of biotin biosynthesis genes [25]. In *Saccharomyces cerevisiae*, a longer N-terminal domain interacting with biotin appears to be involved in the regulation of certain genes (biotin sensing, see Pirner *et al.* [26]).

As previously mentioned, the BCCP domain is characteristic of all the biotin enzyme family, including the decarboxylases and transcarboxylases, whereas the BC domain is limited to the carboxylases. The CT domain gives its substrate specificity to each biotin-dependent carboxylase. Although all biotin-dependent carboxylases bear these three types of protein domains, their arrangement is unequal among different carboxylases and from one domain of life to another. This arrangement will be briefly summarized below (Figure 2B).

Acetyl-CoA carboxylases (ACC; E.C. 6.4.1.2) catalyze the carboxylation of acetyl-CoA to malonyl-CoA in different metabolic processes, such as the first step of fatty acid synthesis or the autotrophic 3-hydroxypropionate

pathway. In *Escherichia coli*, which bears a model bacterial ACC, the three protein domains are encoded by four independent peptides: one BC subunit, one BCCP peptide and two distantly related CT subunits interacting together to ensure the carboxyl transfer function [8]. It is noteworthy that these two distantly related CT subunits are part of a single peptide in all the other biotin-dependent carboxylases bearing them, so here we will refer to the fused peptide as a CT domain and explicitly point to the separation of the CT subunits in bacterial ACC sequences when required. In eukaryotes, the three domains are encoded in a single polypeptide arranged as BC-BCCP-CT from the N- to the C-terminus [1]. Propionyl-CoA carboxylases (PCC; E.C. 6.4.1.3) attach one carboxyl group to the propionyl-CoA to synthesize the methylmalonyl-CoA taking part in the 3-hydroxypropionate pathway of CO₂ fixation, in the synthesis of polyketides and in the degradation of fatty acids and branched-chain amino acids. In bacteria and eukaryotes, PCCs are made up of two polypeptides: a BC-BCCP α -subunit, and a CT-bearing β -subunit. Recent work has shown an additional domain (BC-CT interaction domain, called BT) involved in subunit interactions and located between the BC and the BCCP motifs in the α -subunit [27]. Archaea were thought for a long time to lack ACC and PCC since fatty acids were unknown in this domain of life. Nevertheless, fatty acid synthesis and degradation has been recently detected in archaea and one biotin-dependent carboxylase, using both acetyl-CoA and propionyl-CoA as substrates, has been described in the 3-hydroxypropionate/4-hydroxybutyrate pathway of CO₂ fixation of the archaeal order Sulfolobales [6,28,29]. This archaeal ACC/PCC is composed of three different types of subunits bearing respectively BC, BCCP and CT domains.

The 3-methylcrotonyl-CoA carboxylase (MCC; E.C. 6.4.1.4) and the geranyl-CoA carboxylase (GCC; E.C. 6.4.1.5) synthesize 3-methylglutaconyl-CoA and γ -carboxygeranyl-CoA by adding one CO₂ molecule to their respective substrates. MCC is involved in the breakdown of leucine and acyclic monoterpenes of the citronellol family, whereas GCC appears to participate only in the latter catabolism [9]. While MCC activity has been demonstrated in bacteria and eukaryotes, GCC has only been observed in bacteria. GCC has also been shown to use substrates different from geranyl-CoA [30]. Both MCC and GCC are composed of BC/BCCP-containing α -subunits and CT-bearing β -subunits.

The acyl-CoA carboxylases are a heterogeneous group of enzymes involved in the same pathways as ACC and PCC and also carrying out important functions in the synthesis of secondary metabolites in bacteria. Acyl-CoA carboxylases are basically made up of two types of subunits, the first bearing the BC and the BCCP domains and the second carrying the CT function [7]. As previously shown, this structure is also shared with PCCs,

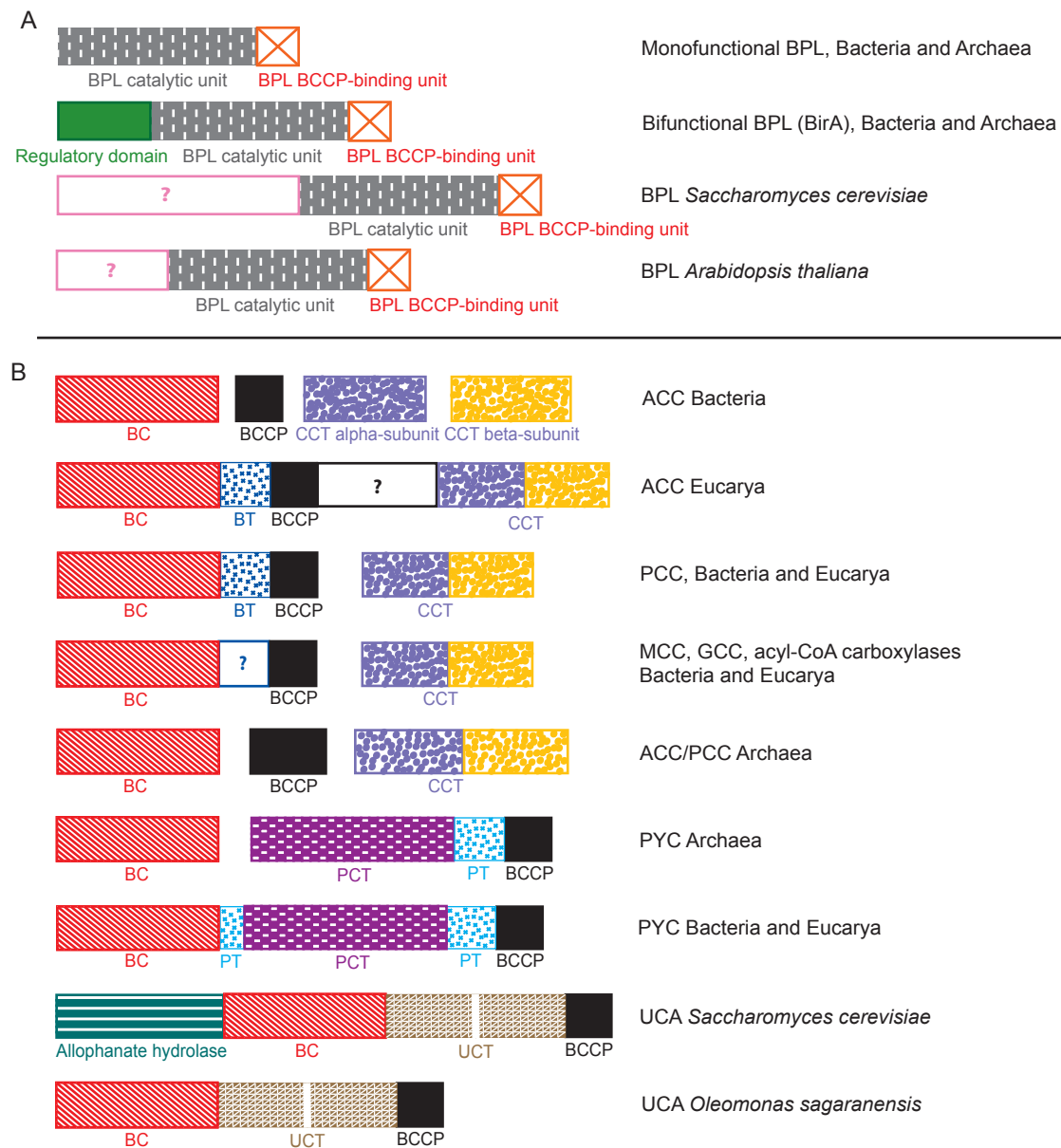


Figure 2 Protein domain arrangements of different biotin protein ligases and biotin-dependent carboxylases with respect to their functions and phylogenetic origins. (A) Biotin protein ligases (BPL) and (B) biotin-dependent carboxylases. Homologous protein domains in different enzymes are shown in the same colors and filling. ACC: acetyl-CoA carboxylase; PCC: propionyl-CoA carboxylase; MCC: methylcrotonyl-CoA carboxylase; GCC: geranyl-CoA carboxylase; PYC: pyruvate carboxylase; UCA: urea carboxylase; BC: biotin carboxylase; BCCP: biotin carboxyl carrier protein; CCT: CoA-substrate related carboxyl transferase; PCT: PYC-related carboxyl transferase; UCT: UCA-related carboxyl transferase; BT: BC-CT interaction domain; PT: PYC tetramerization.

MCCs and GCCs. Indeed, acyl-CoA carboxylases are known to carboxylate acetyl-CoA, propionyl-CoA or butyryl-CoA and, very often, several of these substrates. These promiscuous acyl-CoA carboxylases have been reported in Actinomycetes and Delta-proteobacteria [7,31,32]. Such metabolic promiscuity has led to some confusion in the literature. In the one hand, some authors sometimes consider acyl-CoA carboxylases as

PCCs or ACCs according to their preferential substrate, regardless of their possible promiscuous nature or origin [7,31,32]. Moreover, these enzymes can easily change their favorite substrate during evolution, as illustrated by studies showing that the replacement of one single precise residue in the CT sequence of one acyl-CoA carboxylase is enough to shift its substrate specificity [33,34]. In the other hand, most comparative studies on

ACC, PCC and MCC activities very often consider each function as being characteristic of one particular carboxylase, ignoring the potential promiscuity typical of these enzymes [11]. Although such functional classifications can be useful to study enzymes within particular metabolic pathways, they may lead to confusion between the well-characterized one-substrate specific carboxylases and promiscuous enzymes carrying out one particular function in a given metabolic context. As an alternative, a phylogenetic approach can complement the functional studies to get new insights on the emergence of functions and classification of this family of enzymes.

Pyruvate carboxylases (PYC; E.C. 6.4.1.1) catalyze the carboxylation of pyruvate to oxaloacetate, an important function for anaplerosis, gluconeogenesis and fatty acid synthesis. A significant feature concerning the protein domain composition of PYC is the homology of the BC and BCCP domains with those of the rest of biotin-dependent carboxylases, in contrast with the independent origin of the CT element. Most bacteria and eukaryotes bear a polypeptidic PYC carrying the BC domain in its N-terminal end, the CT in the central part and the BCCP domain in the C-terminus [35]. Recent studies have discovered a domain situated between the BC and CT domains and between the CT and BCCP domains in *Rhizobium etli* [36-38]. The new domain, referred to as the PYC tetramerization (PT) or allosteric domain, would play a role in tetramerization of PYC subunits and allosteric regulation of PYC by acetyl-CoA. Finally, archaea and some bacteria contain an acetyl-CoA-independent PYC with two different subunits instead of one polypeptide: the BC α -subunit and the CT/BCCP β -subunit [39-42].

Urea carboxylases (UCA; E.C. 6.3.4.6) fix a carboxyl group in urea to form allophanate, an intermediate product of a two-steps process of urea degradation. Allophanate is subsequently hydrolyzed by the allophanate hydrolase to ammonia and CO₂. While in *Saccharomyces cerevisiae* the allophanate hydrolase and the urea carboxylase are fused within the same polypeptide [43], in the alpha-proteobacterium *Oleomonas sagaranensis* and green algae these two functions are carried out by two independent enzymes homologous to the unique *S. cerevisiae* peptide [44]. The *O. sagaranensis* urea carboxylase is a polypeptide by itself containing a BC domain in its N-terminal end, a BCCP domain in its C-terminus and a predicted central CT domain non-homologous to the previously cited CTs.

These examples illustrate the very complex modular architecture of the biotin-dependent carboxylases, which has been studied for long [45]. This family has been proposed to have evolved by duplication, fusion and recombination events from small monofunctional precursors to produce the modern multifunctional polypeptides

[11,45]. The ubiquity of the biotin-dependent carboxylases in the three domains of life and the diversification of their elements have been pointed as arguments for the ancient origin of this family [46,47]. However, hitherto the phylogenetic analyses of biotin-dependent carboxylases have been focused in relatively restricted groups and functions [9,32] or have been carried out with a very limited taxonomic sampling [4,47]. In this work, we attempt to reconstruct the early evolution of the biotin-dependent carboxylases using a global phylogenomic approach in a wide range of taxonomic groups. Our results support previous hypothesis concerning the modular emergence of enzymes of the biotin carboxylase family but also challenge current views on the relationships among different groups of enzymes, especially those related to the acyl-CoA carboxylases. On the basis of phylogenetic evidence, we have tried to define the ancestral content of enzymes from the biotin-dependent carboxylase family in the respective last common ancestor of each domain of life and in the last common ancestor of all living organisms (the cenancestor).

Results

As mentioned above, the polypeptide composition can largely vary from one biotin-dependent carboxylase to another. Therefore, we have studied each functional protein domain separately. Unlike the biotin carboxylase (BC) domain, which is shared among all the biotin-dependent carboxylases, three non-homologous carboxyl transferase (CT) domains exist in these enzymes: one is common to the ACC, all the PCC, the MCC, the GCC and the acyl-CoA carboxylases and will be called here the CoA-substrate related carboxyl transferase (CCT); another is characteristic of the PYC and will be called the pyruvate carboxylase carboxyl transferase (PCT); finally, UCAs seem to use their own CT (UCT).

Phylogenetic analysis of the BC domain

The biotin carboxylase domain (BC) is the main distinctive feature shared among all the biotin-dependent carboxylases with respect to the biotin-dependent decarboxylases and transcarboxylases. A critical step in our study was the assignment of each one of the BC domain sequences to their specific function. As the BC domain is a very well conserved motif, it is difficult to distinguish the actual substrate specificity of one precise BC-encoding sequence only on the basis of primary sequence characteristics. Consequently, we constructed preliminary BC domain phylogenetic trees (see Methods and additional file 1) in order to classify each sequence according to a phylogenetic framework. Functional assignment of the bacterial sequences was summarized in a presence-and-absence pattern (Table 1, for complete data see additional file 2).

Table 1 Bacterial phyla bearing biotin-dependent carboxylase sequences

	Protein function (protein subunit)										
	ACC (BC)	ACC (CT-alpha)	ACC (CT-beta)	PCC (alpha)	PCC (beta)	MCC (alpha)	MCC (beta)	XCC (alpha)	XCC (beta)	PYC	UCA (BC domain)
Alphaproteobacteria (62)	+++	+++	+++	+++	+++	+++	+++	+	+	++	+
Betaproteobacteria (27)	+++	+++	+++	+++	+++	+++	+++	+++	+++	-	+
Gammaaproteobacteria (69)	+++	+++	+++	-	-	+	+	+	+	-	+
Deltaproteobacteria (16)	+++	+++	+++	-	-	-	-	+++	+++	+	-
Epsilonaproteobacteria (8)	+++	+++	+++	-	-	-	-	+++	-	-	-
Acidobacteria (3)	+++	+++	+++	-	-	-	-	+++	+++	+	+
Actinobacteria (29)	+	+	+	-	-	-	+++	+++	+++	++	+
Firmicutes (59)	+++	+++	+++	-	-	-	-	+	+	++	-
Aquificae (4)	+++	+++	+++	-	-	-	-	+++	-	-	-
Bacteroidetes (20)	+++	+++	+++	-	-	-	-	+++	+++	+	-
Chlamydia/ Verrucomicrobia (11)	+++	+++	+++	-	-	-	-	-	-	+	+
Cyanobacteria (18)	+++	+++	+++	-	-	-	-	-	-	-	-
Chloroflexi (5)	+++	+++	+++	-	-	-	+++	+++	+++	-	-
Chlorobi (6)	+++	+++	+++	-	-	-	-	+++	+++	-	-
Planctomycetes (4)	+++	+++	+++	-	-	-	-	-	-	++	-
Deinococcus-Thermus (3)	+++	+++	+++	-	-	-	-	+	+++	-	-
Spirochaetes (3)	-	-	-	-	-	-	-	+	+	-	-
Thermotogales (5)	-	-	-	-	-	-	-	-	+++	-	-
Fusobacteria (1)*	+++	+++	+++	-	-	-	-	-	+++	-	-
Elusimicrobia (1)*	+++	-	-	-	-	-	-	-	-	-	+++
Dictyoglomi (1) *	+++	+++	+++	-	-	-	-	-	-	-	-
Nitrospira (1)*	+++	+++	+++	-	-	-	-	+++	-	-	-

Numbers next to phyla names correspond to the number of complete genome sequences upon which the search was carried out (see additional file 2 for the complete list). +++, homologues could be found in more than 50% of the genome sequences; +, homologues could be found in less than 50% of the complete genome sequences; -, no homologues found; *, only one genome sequence from these phyla was analyzed.

It is important to note that the BC domain related to the ACC function could be found in almost all bacterial groups, with the exception of Actinobacteria, Spirochaetes and Thermotogales. The PYC polypeptide bearing the BC domain was less widespread but could be found in Alpha- and Delta-proteobacteria, Acidobacteria, Actinobacteria, Firmicutes, Bacteroidetes, Planctomycetes and some organisms from the Chlamydiae-Verrucomicrobia group. Instead of this fused polypeptide, Aquificales and Epsilon-proteobacteria bear an independent PYC-BC domain forming a separate subunit from the CT-BCCP domains. In contrast to the largely distributed ACC and PYC, MCC was limited to Alpha-, Beta- and Gamma-proteobacteria, whereas monofunctional PCC was restricted to Alpha- and Beta-proteobacteria. UCA was found scattered in some representatives of Alpha-, Beta- and Gamma-proteobacteria, of the Chlamydiae-Verrucomicrobia group and of Acidobacteria. Finally,

acyl-CoA carboxylases were found to be moderately distributed in all groups of Proteobacteria (except for the Epsilon division), Acidobacteria, Actinobacteria, Bacteroidetes, Chloroflexi, Chlorobi, Deinococcus-Thermus group and one spirochete. In addition to these bacterial sequences, the eukaryotic BC sequences formed well defined functional groups closely related to their bacterial homologues. The only exception to this observation in our preliminary results were the very divergent eukaryotic polypeptidic ACCs, which formed an independent clade probably induced by a long branch attraction artifact due to the acceleration of evolutionary rates subsequent to the polypeptidic fusion of ACC. ACC, PYC, MCC and PCC are widely distributed throughout eukaryotes, while UCA and acyl-CoA carboxylases are limited to fungi and some green algae.

In order to improve the resolution of the BC domain phylogeny, we selected representative sequences from

the complete dataset to carry out maximum likelihood (ML) and Bayesian inference (BI) phylogenetic analyses. The resulting trees were similar to each other, so only the ML tree was shown here (Figure 3, the same holds for the trees of the other protein domains and their respective figures). The BC domain phylogeny confirmed the previous functional classification and supported the monophyly of most bacterial sequences within each functional group. Multifunctional acyl-CoA carboxylase sequences formed different clusters branching at deep nodes of the tree. Eukaryotic sequences clustered again within the groups of their functional bacterial counterparts.

In all trees shown above, the archaeal sequences did not form a monophyletic group but were split in several clades within the paraphyletic group of acyl-CoA carboxylases. Nevertheless, these archaeal clades were close to each other, suggesting that they could actually be monophyletic but artificially divided into different clades because of a lack of phylogenetic signal or a phylogenetic reconstruction artifact. As there was little doubt about the position of the eukaryotic sequences with respect to the bacterial ones, we proceeded to the removal of all eukaryotic sequences in order to try to improve the resolution for the prokaryotic sequences, in particular the archaeal ones (Figure 4). As in previous phylogenies, a bacterial ACC clade could be observed in the resulting tree. MCC, UCA and monofunctional-PCC clades branched within a paraphyletic acyl-CoA carboxylase assemblage, and there was a cluster containing all archaeal sequences and bacterial PYCs. Archaeal sequences were not monophyletic but archaeal phyla were retrieved. They occupied closely related basal branches of the PYC cluster with relatively weak support. Consequently, the possibility of an artificial division of archaeal sequences related to a phylogenetic reconstruction artifact could not yet be excluded. We tested this alternative using an unbiased AU test to statically compare the topology shown in Figure 4 with a tree in which the monophyly of archaea was constrained (see Methods). The monophyly of archaea could not be rejected by the test ($P = 0.44$), so archaeal sequences may be monophyletic, as it will be discussed later (see the PCT domain phylogeny section). Interestingly, the BC domain sequences of archaeal ACC-PCC sequences did not cluster among the acyl-CoA carboxylases as could have been expected, but branched among the archaeal PYC sequences. Several bacterial sequences from Aquificales and Epsilon-proteobacteria branched very close to the archaeal sequences, suggesting that they may have acquired these genes by horizontal gene transfer (HGT) from archaeal donors. One representative of these bacterial sequences (*Aquifex aeolicus*) has indeed been deeply studied and shown to adopt an archaeal PYC structure

(i.e., two types of subunits instead of one sole polypeptide) and to actually carry out the PYC function [38,42].

Phylogenetic analysis of the CT domains: The CoA-related carboxyl transferase (CCT)

Since CCT is shared by enzymatic complexes catalyzing carboxylation of different substrates, the first stage of our analysis was to predict the function of each sequence. As for the BC domain analysis described above, a phylogenetic analysis (additional file 1) allowed the functional assignment of CCT sequences based on their position in the resulting phylogenetic tree (results summarized in Table 1, for complete data see additional file 2).

All bacterial species, except the Thermotogales, Spirochaetes and most Actinobacteria, harbored the two CCT-bearing subunits of the bacterial ACC. These independent subunits were very divergent from the rest of the fused CCT domains, probably as a consequence of the separation of the two units in independent peptides. The preliminary phylogenies were congruent with the main bacterial groups within the bacterial ACC clusters. CCT homologues of MCC could be detected in Alpha-, Beta-, and Gamma-proteobacteria, Actinobacteria and Chloroflexi, whereas the CCT domain of monofunctional PCC could only be found in Alpha- and Beta-proteobacteria. A CCT protein domain of acyl-CoA carboxylases (including GCC and likely multifunctional PCCs) could be found in all the proteobacterial groups except the Epsilon-division, in Acidobacteria, Actinobacteria (which possess a high number of paralogues), Firmicutes, Bacteroidetes, Chloroflexi, some Chlorobi, the Deinococcus-Thermus group, Thermotogales and Fusobacteria. Despite a few differences between the BC-domain and the CCT-domain detection (see additional file 2), these results largely matched our previous observations on the presence-absence patterns of the BC domain, supporting the validity of our phylogenetic approach to discriminate sequences according to their functions. Concerning the CCT sequences from eukaryotes, the results from the preliminary phylogenetic reconstructions coincided with those of the BC domain in that all the eukaryotic sequences but the ACC branched in close relationship with their bacterial homologues. As in the BC domain phylogeny, eukaryotic ACC-CCTs formed a highly divergent cluster, what may probably be explained as the result of sequence divergence subsequent to the fusion of different ACC domains into one unique polypeptide.

To improve the resolution of the CCT phylogeny, we selected representative sequences to carry out ML and BI phylogenetic analyses (the CCT subunits of the bacterial ACC and the eukaryotic ACC were removed because of their extreme sequence divergence), with



Figure 3 Maximum likelihood tree of the biotin carboxylase (BC) domain. This tree is based on 284 representative sequences and 384 conserved sites and was arbitrarily rooted on the bacterial ACC-related sequences. Numbers at nodes indicate bootstrap values higher than 50. Triangles correspond to collapsed groups of eukaryotes and Proteobacteria. Colors on leaves represent the affiliation of the sequences to their respective domain of life: archaea (red), bacteria (blue) and eukaryotes (green). Bars on the right report the functional assignment of the sequences; sequences that are not in front of any bar are assumed to bear an acyl-CoA carboxylase activity.

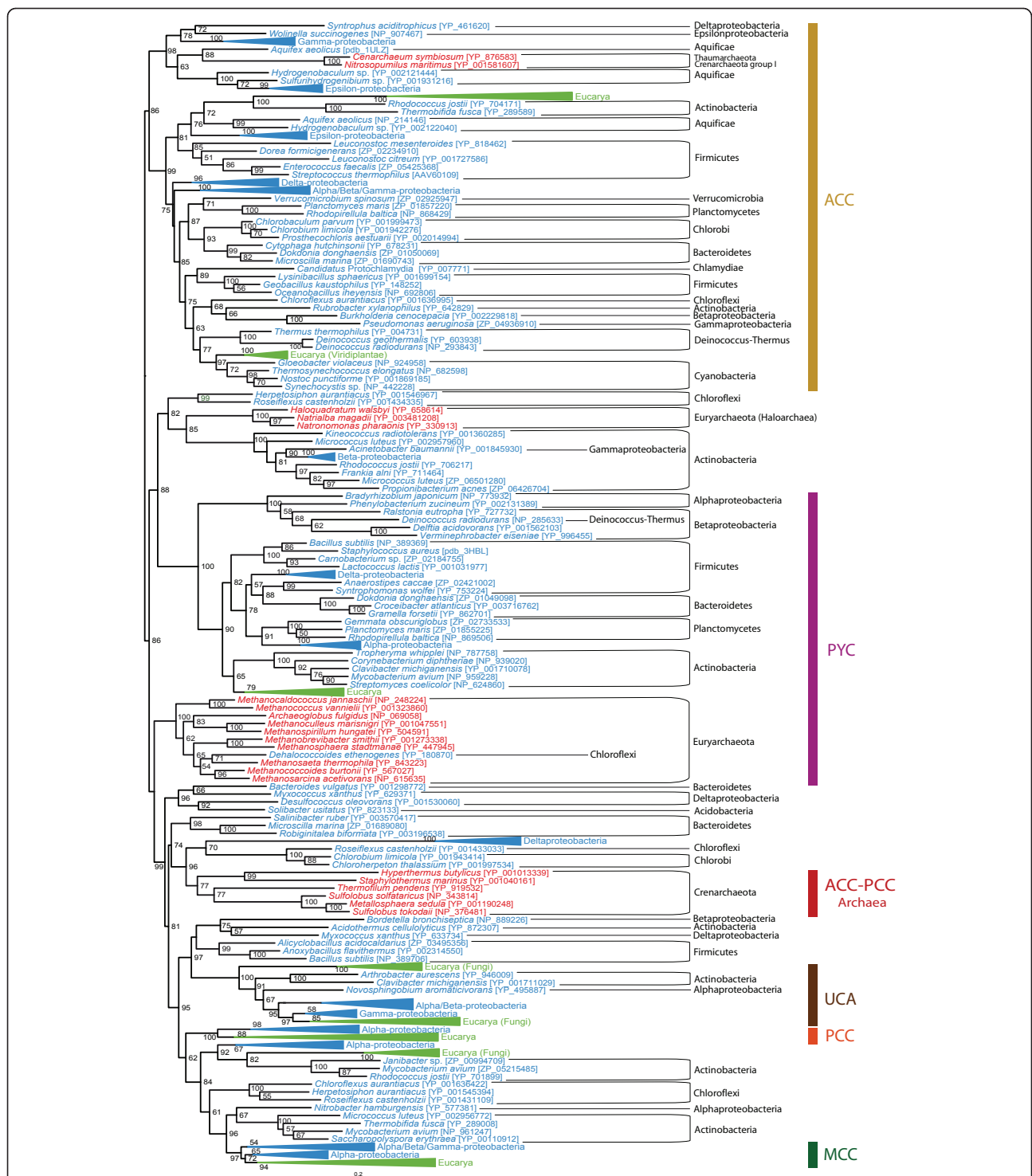


Figure 4 Maximum likelihood tree of archaeal and bacterial biotin carboxylase (BC) domain sequences. This tree is based on 196 representative sequences and 322 conserved sites and was arbitrarily rooted on the PYC-related sequences. Numbers at nodes indicate bootstrap robustness values higher than 50. Colors on leaves represent the affiliation of the sequences to their respective domain of life: archaea (red), bacteria (blue) and eukaryotes (green). Bars on the right report the functional assignment of the sequences; sequences that are not in front of any bar are assumed to bear an acyl-CoA carboxylase activity.

similar results between both techniques (Figure 5). CCT sequences have diverged from each other more than the BC sequences did and, thus, CCT trees are characterized by long branches. Although some HGTs could be identified, major accepted bacterial groups were found within the functional groups described before, suggesting vertical inheritance as a major process to explain their presence in the bacterial genomes. Eukaryotic CCT sequences were also monophyletic within their functional groups and strongly related to their bacterial counterparts.

Finally, although CCT homologues could be detected in a wide diversity of archaeal genomes, we did not retrieve the monophyly of these archaeal sequences in the CCT phylogenies. The statistical support for the separation into several archaeal subgroups was weak, so we tested the possibility that they actually form a monophyletic assemblage by using an unbiased AU test to compare the tree shown in Figure 5 with a tree in which we forced the monophyly of the archaeal sequences (see Methods). The latter could not be rejected by the AU test ($P = 0.12$), opening the possibility that the archaeal sequences may be monophyletic and that the topology observed could be affected by a phylogenetic reconstruction artifact.

Phylogenetic analysis of the CT domains: The pyruvate carboxylase carboxyl transferase (PCT)

Although the PCT domain is characteristic of PYC and is not shared by any other biotin-dependent carboxylase, the oxaloacetate decarboxylase (ODC; E.C. 4.1.1.3, which takes part in Na^+ transport in some bacteria) is also known to use an α -subunit homologous to the PCT domain [48]. Consequently, ODC α -subunit sequences were incorporated into our phylogenetic analyses. In addition, some extremely divergent bacterial homologues showed similarity to transcarboxylases, but we will not treat them here because of their very poor sequence conservation.

A preliminary tree was used to select a set of representative sequences to reconstruct accurate phylogenies with ML and BI methods (Figure 6). PCT phylogenetic trees showed high divergence between two types of sequences. On the one hand, the PCT domain of bacteria and eukaryotes, which is integrated in their PYC polypeptides. On the other hand, the archaeal-type PCT, found in the β -subunit of the archaeal PYC, in the PCT α -subunit of ODC and in divergent sequences of unknown function in some bacteria, especially in Chlorobi and Epsilon-proteobacteria species. As expected, the PCT domain of the polypeptidic PYC sequences showed the same distribution and phylogeny as the BC domain of the same PYC sequences. In particular, the PCT phylogeny supported the monophyly of the bacterial phyla and the position of

the eukaryotic lineage within the bacterial radiation. The phylogeny of the archaeal PCT sequences was also congruent with the accepted archaeal groups. The archaeal cluster contained the bacteria known to carry the archaeal BC-PYC subunit (see above) and some Firmicutes. This result supported an HGT from these archaea to the concerned bacterial groups. Well-characterized ODC homologues are distributed in a limited group of Gamma-proteobacteria and some sporadic bacteria, what supports the restriction of this function to a very small diversity of organisms.

Phylogenetic analysis of the CT domains: The urea carboxylase carboxyl transferase (UCT)

Compared to the rest of enzymes of the biotin-dependent carboxylase family, little is known about the reaction mechanism and biochemical properties of UCA. Although significant similarities to the BC and BCCP domains have been identified in the UCA gene from the alpha-proteobacterium *Oleomonas sagaranensis*, almost no information exists for the predicted central UCT domain [44]. Similarity searches carried out with this central domain retrieved a protein domain annotated as allophanate hydrolase, probably because the first identified UCA gene was the one of *Saccharomyces cerevisiae*, in which the allophanate hydrolase and the UCA genes are part of a single polypeptide. Thus, the proposed *O. sagaranensis* UCT domain has been annotated as an allophanate hydrolase although it most likely takes part in the UCT function instead of in the allophanate hydrolase.

Fused *S. cerevisiae*-like UCA-allophanate hydrolase polypeptides were found only in some eukaryotes, whereas *O. sagaranensis*-like UCAs containing a BC, a BCCP and the hypothetical UCT domain were detected in some Alpha-, Beta- and Gamma-proteobacteria, some Actinobacteria and some eukaryotes. Finally, our UCT searches led to the finding of two hypothetical homologous genes in some bacterial and archaeal (Thermococcales) genomes. One of these genes was similar to the N-terminal region of the predicted UCT whereas the other was similar to the C-terminal region. These two genes were frequently found adjacent to each other in the concerned genomes, but none of them has been studied so far, and thus, functional data are not available.

Analysis of the BT/PT domain

As mentioned above, extra protein domains involved in subunit interactions have been recently reported in PYC sequences of *Rhizobium etli*, *Staphylococcus aureus* and *Homo sapiens* [36-38] and in PCC of *Ruegeria pomeroyi* [27]. In the former, the extra domain was called PT (PYC tetramerization) and was shown to contribute to allosteric regulation in addition to subunit interaction.

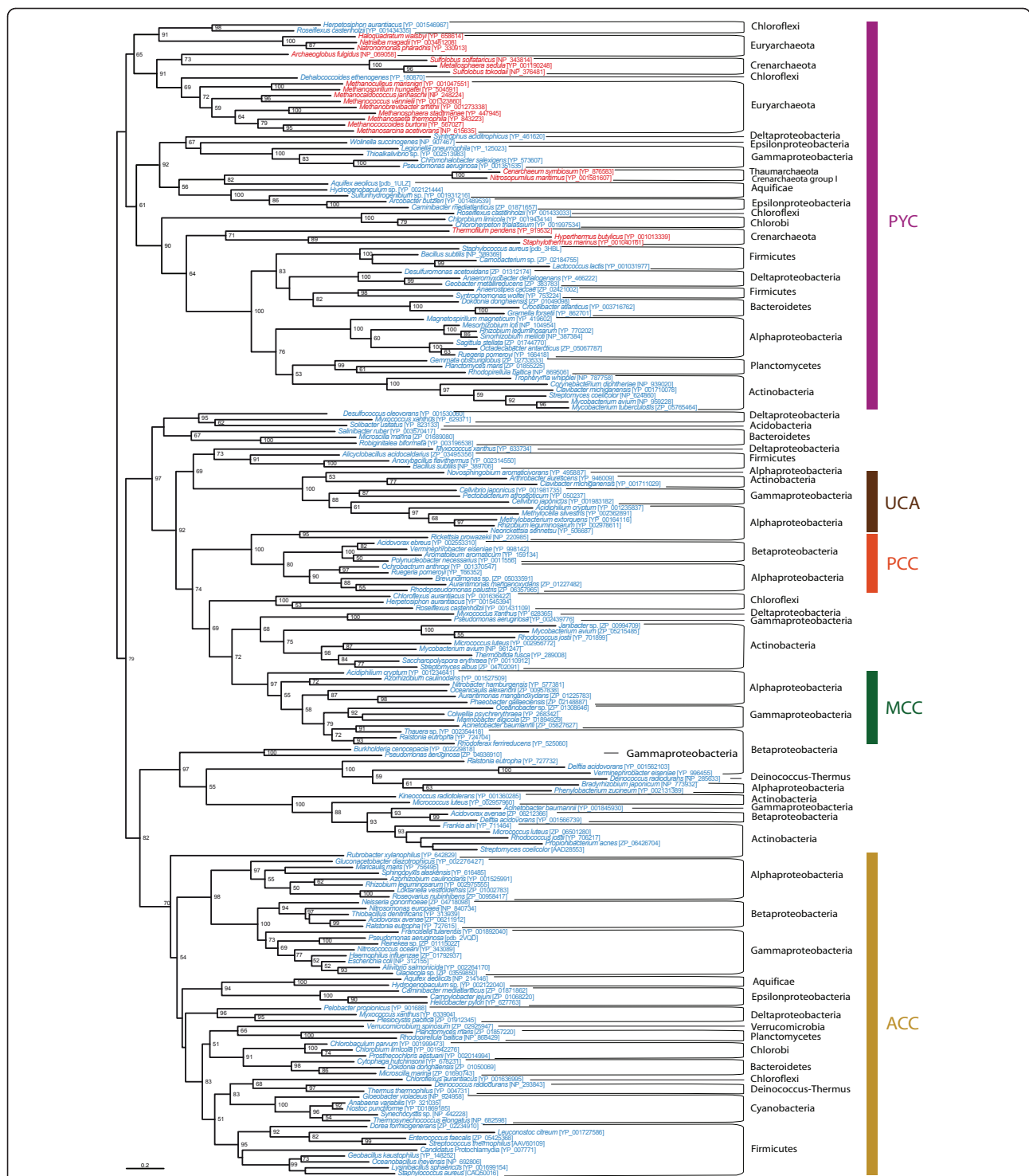


Figure 5 Maximum likelihood tree of CoA-substrate related carboxyl transferase (CCT) domain sequences. This tree is based on 179 representative sequences and 438 conserved sites and was midpoint rooted. Numbers at nodes indicate bootstrap robustness values higher than 50. Triangles correspond to collapsed groups of eukaryotes. Colors on leaves represent the affiliation of the sequences to their respective domain of life: archaea (red), bacteria (blue) and eukaryotes (green). Bars on the right report the functional assignment of the sequences; sequences that are not in front of any bar are assumed to bear an acyl-CoA carboxylase activity.



Figure 6 Maximum likelihood tree of pyruvate carboxylase carboxyl transferase (PCT) domain sequences. This tree is based on 126 representative sequences and 432 conserved sites and was midpoint rooted. Numbers at nodes indicate bootstrap robustness values higher than 50. Colors on leaves represent the affiliation of the sequences by their respective domain of life: archaea (red), bacteria (blue) and eukaryotes (green). Bars on the right report the functional assignment of the sequences; sequences that are not in front of any bar have unknown function.

PT is split in two parts, the first between the BC and the PCT domains, and the second between the PCT and the BCCP domains (Figure 2). In PCC, the protein domain relevant for subunit interaction was called BT (BC-CT interaction domain) and found to be located between the BC and the BCCP domains of the PCC α -subunit. Although the BT and the PT domains shared little primary sequence similarity, structural comparisons revealed striking resemblances concerning above all one helix structure followed by several β -strands. Further structural predictions suggested that the intermediate region between the BC and the BCCP domain of eukaryotic ACC is similar to BT/PT domains [27], supporting that this feature may be widely shared among biotin-dependent carboxylases.

Consequently, we attempted to identify BT/PT-like domains within other biotin-dependent carboxylases. High sequence similarity to BT and PT could only be identified among PCC and PYC sequences, respectively, confirming previous results [27,38]. In addition, we observed a protein region of 100-140 residues flanked by BC and BCCP domains in α -subunits from MCC, GCC and acyl-CoA carboxylases (reported in Figure 2 as domains of unknown function). We used two different algorithms (see Methods) to carry out secondary structure searches on the linkers of the MCC, GCC, archaeal and bacterial PYC and acyl-CoA carboxylase sequences. The helix structure characteristic of the BT function in PCC [27] was predicted in all linkers tested, except the acyl-CoA carboxylase characterized by Rodriguez *et al.* [7], for which the two methods applied were not consistent with each other. This enzyme is known to use a supplementary subunit called accE that is involved in subunit interaction [49]. Interestingly, a helix structure was detected in this accE that might carry out the same function as in the BT/PT domain.

Phylogenetic analysis of the BPL-BirA genes

The biotin carboxyl carrier protein (BCCP) domain is a short biotinylated peptide (~ 90 amino acids) often fused to the other constitutive domains of the biotin-dependent carboxylases. Its short size and participation in very different polypeptide architectures make difficult to reconstruct accurate phylogenies using the BCCP sequences. Consequently, instead of analyzing the short BCCP domain itself, we decided to study the phylogeny of the biotin protein ligase (BPL) responsible of its biotinylation.

In eukaryotes, BPL occurs in a polypeptide in which the C-terminal end bears the BPL function, whereas little is known about the function of the very diverse N-terminal region. As previously mentioned, two main types of BPL can be found in prokaryotes: the so-called BPL that solely bears a BPL function, and BirA, that has a DNA-binding domain in its N-terminus responsible

for a transcriptional regulatory role. Although quaternary structure differences among different taxa have been major concerns in previous literature [23,24,50,51], we will only consider here the major distinction concerning the bifunctionality of BirA with respect to the monofunctionality of BPL [22]. Accordingly, we classified prokaryotic sequences into the BPL or BirA groups on the basis of the absence or presence of the DNA-binding N-terminal domain, respectively. Noteworthy, we observed that several BirA sequences lacked the conserved positions important for the regulatory function as established by Mukhopadhyay *et al.* [18]. Except for some rare cases that had already been reported (*Clostridium acetobutylicum*, *Lactococcus lactis*, *Halobacterium* sp., *Pyrococcus abyssi*, *Pyrococcus furiosus*), our BPL-BirA searches confirmed the presence of only one BPL homologue in each prokaryotic genome. The BPL protein was widespread in Archaea, Bacteria and Eucarya and its phylogeny was congruent with the accepted main taxonomic groups within each domain (Figure 7), supporting that an ancient BPL function existed in the three domains of life. Nevertheless, instead of finding monophyletic groups containing uniform functional clades, our results showed the monofunctional and the bifunctional sequences as being mixed up all over the prokaryotic part of the tree whatever the reconstruction method or representative sequence selection dataset was used. Some species bearing both BPL and BirA genes, as *Halobacterium* sp. and *C. acetobutylicum*, had their respective BPL close to their own BirA in phylogenetic trees, whereas *P. abyssi*, *P. furiosus* and *L. lactis* BPL and BirA genes branched very far from each other.

Discussion

The large distribution of biotin-dependent carboxylases and their remarkable mechanistic conservation has been interpreted as an evidence for the ancient origin of this enzyme family [46,47]. However, phylogenetic analyses of biotin-dependent carboxylases carried so far were either restricted to some functions and taxonomic groups or used very limited taxonomic sampling [4,9,32,47]. For the first time, here we studied the early evolution of the biotin-dependent carboxylases using a global phylogenomic approach in a wide range of taxonomic groups. Our results support the ancient origin of this family and give new insights on the evolutionary history of these protein domains and enzymes. In particular, we have observed that the domains of life are different with regard to the importance of the biotin-dependent carboxylases in their metabolism and evolution. Consequently, we will first discuss the evolution of these enzymes in each domain of life, prior to conclude on how our results provide information about the biotin-dependent carboxylase content in the cenancestor.

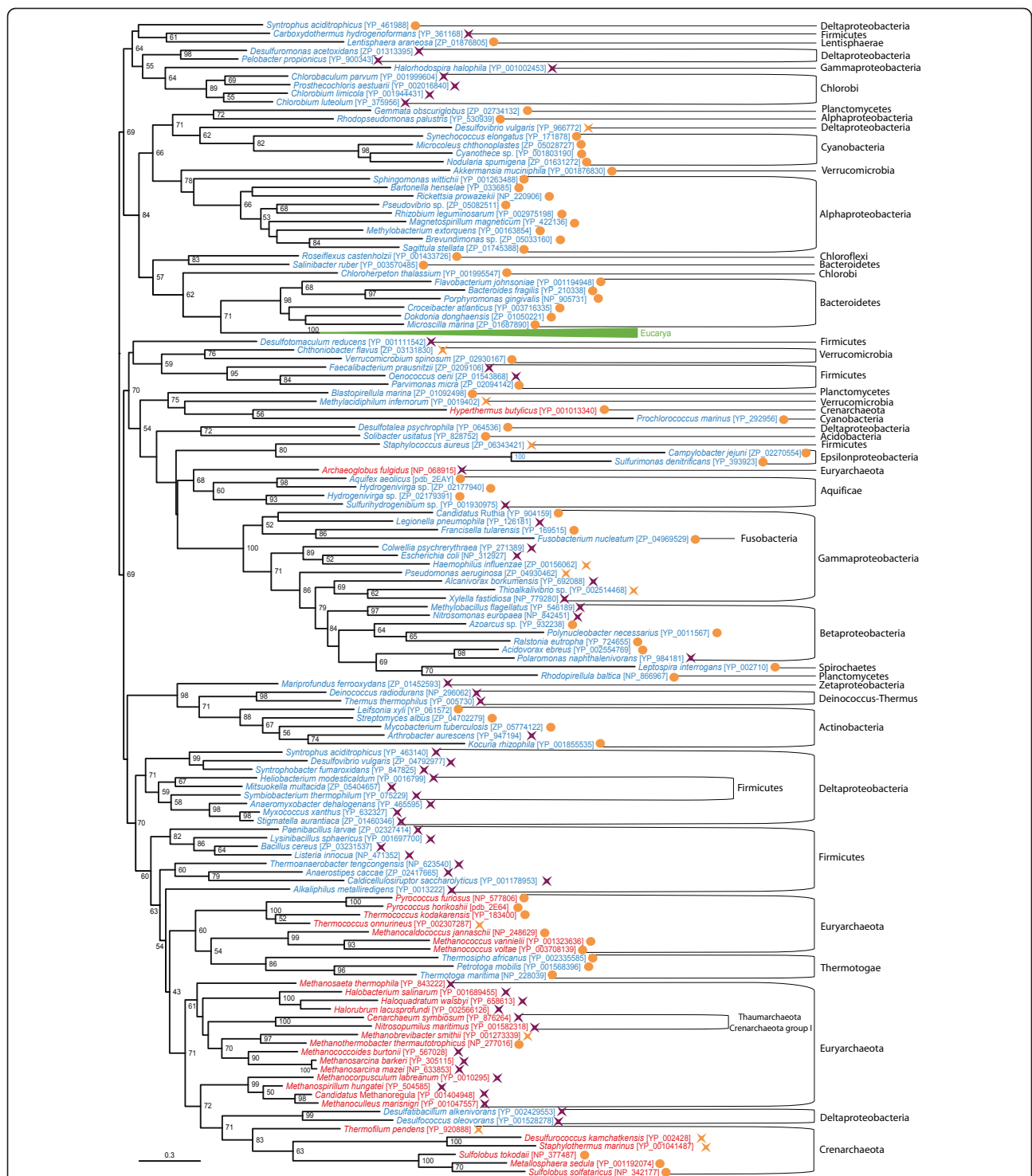


Figure 7 Maximum likelihood tree of the biotin carboxyl ligase (BPL). This tree is based on 156 representative sequences and 164 conserved sites and was arbitrarily rooted. Numbers at nodes indicate bootstrap robustness values higher than 50. Triangles correspond to collapsed groups of eukaryotes. Colors on leaves represent the affiliation of the sequences to their respective domain of life: archaea (red), bacteria (blue) and eukaryotes (green). Purple crosses correspond to sequences bearing a complete N-terminal domain and thus assumed to carry out a BirA regulatory function in addition to the BPL activity; orange circles correspond to sequences lacking the N-terminal domain and thus assumed to carry only the BPL function; orange crosses correspond to sequences that bear an N-terminal domain but lacking conserved positions important for the regulatory function.

Biotin-dependent carboxylases in Bacteria

The high diversity of biotin-dependent carboxylase paralogues found in the bacterial domain reflects the major role of these enzymes in bacterial metabolism. The presence-absence study combined to the BC and CCT phylogenies supported the wide distribution and monophyly of bacterial ACC, suggesting that it is ancestral to this domain of life. Similarly, the PCT phylogeny suggested that the polypeptidic PYC is monophyletic and widespread enough to support the ancestral presence of this carboxylase in Bacteria. In contrast, the BC domain devoted to UCA activity is limited to a restricted number of bacteria branching within the acyl-CoA carboxylases from which it probably emerged. Proteins homologous to UCT are distributed among a wider range of bacteria, but these are hypothetical genes and their function is unknown. In order to explain why the distribution of UCT is larger than that of the UCA-related BC, two possibilities exist: either UCT could be specific of biotin-dependent carboxylation and collaborate with other BCs to specifically carry out this function, or the UCT domain could have been recruited from another biochemical reaction, so that certain UCT homologues that we can detect would be representatives of the original function. Thus, there is not enough information to advance a confident hypothesis on UCAs origins, but available evidence could hardly support an ancient origin of UCA in Bacteria and rather suggests its recent emergence and spread by frequent HGT among bacteria.

Sequences annotated as acyl-CoA carboxylases were found to be paraphyletic in BC and CCT phylogenies (Figures 3, 4 and 5). Duplication and HGT events have undoubtedly been of major importance in their evolution, as paralogues and xenologues could be detected in our analyses. Nevertheless, the acyl-CoA carboxylases were distributed in a wide range of bacteria and most often clustered into clades congruent with the accepted main bacterial groups. Moreover, several arguments supported that GCC, MCC and PCC are parts of the acyl-CoA carboxylases: i) GCC, MCC and PCC sequences branched among sequences previously annotated as acyl-CoA carboxylases in our BC and CCT domain phylogenetic trees (Figures 3, 4 and 5); ii) MCCs and monofunctional PCCs were restricted to some proteobacterial groups, supporting the emergence of MCC and PCC functions from an original acyl-CoA carboxylase in a proteobacterial ancestor; iii) the GCC, MCC and PCC sequences shared a common domain architecture with the acyl-CoA carboxylases (Figure 2); iv) some GCC and PCC have been reported to carry out the carboxylation of different substrates, a promiscuity which is characteristic of the acyl-CoA carboxylases [7,30,32]. As a result, our phylogeny-based classification of GCC, MCC and PCC sequences into the acyl-CoA carboxylase group is

reliable and strongly supports that these functions are specific adaptations from a promiscuous acyl-CoA carboxylase-like protein according to the metabolic needs of each species, a common process in enzyme evolution [52,53]. Indeed, functional flexibility of acyl-CoA carboxylases may be of major importance in relatively rapid adaptation strategies concerning bacterial metabolism [5,54,55] since minor precise sequence modifications have been shown to easily change their function [33]. This enlarges our perspective on the emergence of GCCs, MCCs, PCCs and acyl-CoA carboxylases with respect to previous considerations [9,11]

In the available literature, promiscuous acyl-CoA carboxylases are called ACCases [7], but this name can be easily confused with ACCs that specifically carboxylate acetyl-CoA. Thus, a more explicit name should be given to the whole group including PCC, MCC, GCC and acyl-CoA carboxylase sequences in order to point out their common origin whatever their substrate specificity would have become throughout evolution. Therefore, we propose to call this group XCC, for "any CoA-bearing-substrate carboxylase". XCC enzymes are characterized by two subunit types (one BC-BCCP and one CCT), the specificity and quaternary structure of which would have evolved from one generic biotin-dependent enzyme ancestral to bacteria.

Recent works have reported a BT domain located in between the BC and the BCCP domains of the PCC α -subunit [27]. The BT domain has been proposed to be of major relevance for subunit interaction and even though our sequence similarity searches did not identify unequivocally a BT domain in other XCC enzymes, a region of unknown function could be detected between the BC and BCCP domains of all XCC α -subunits. That region shared low sequence similarity with the PCC BT domain but was predicted to contain a helix structure as in the BT/PT domain. Acyl-CoA carboxylases that lack that characteristic helix have been shown to need an additional accE subunit for normal domain interaction [7,32,49], and we have found that the accE also contains a helix structure. Although functional confirmation is required, these results suggest that this unknown region in XCC might be an essential homologue of the BT domain. Therefore, the accE subunit of some acyl-CoA carboxylases would be a substitute developed in BT-lacking α -subunits.

In summary, the presence-absence analyses and the phylogenies of BC, CCT, PCT and UCT support the ancestral presence of ACC, PYC and XCC in bacteria. In addition, specific functions would have arisen from XCCs by duplication, subfunctionalization and HGT events across bacterial evolution. Nevertheless, the absence of some of these enzymes in several bacterial lineages that have been proposed to have diverged early in the bacterial domain (such as the Thermotogae and Aquificae)

opens the possibility that they may have evolved after the last common ancestor of all Bacteria. Therefore, the analysis of the content of biotin-dependent carboxylases in the other domains of life is necessary to infer whether these enzymes are ancestral or not.

Biotin-dependent carboxylases in eukaryotes

In most of our biotin-dependent carboxylase phylogenies, eukaryotic sequences branch within their bacterial homologues, what strongly supports HGTs events from bacteria to an ancestor of the eukaryotes as the most likely way for them to have acquired these enzymes. Some of these HGTs probably took place through endosymbiosis, as suggested by the cellular localization of eukaryotic PCCs and MCCs in mitochondria [56,57] and the mitochondria-targeting signal peptides observed in their N-terminal regions. In addition to the typical eukaryotic ACC polypeptide, plastid-bearing eukaryotes also carry a bacterial-like multi-subunit ACC [8] with an N-terminal region containing a plastid-targeting sequence. Contrary to typical eukaryotic ACC polypeptides that form an independent group, these eukaryotic bacterial-like ACCs branch among cyanobacterial ACCs, strongly supporting an endosymbiotic origin (Figure 3).

Targeting sequences seem to be absent from eukaryotic PYC and ACC. Both BC- and PCT-domain phylogenies strongly support the branching of eukaryotic PYC sequences among their bacterial homologues, what indicates a bacterial origin even though the vector of this HGT cannot be specified. Surprisingly, concerning the eukaryotic ACCs, the phylogenetic reconstructions are at odds with protein domain comparison. On the one hand, eukaryotic ACCs branch among bacterial ACC sequences in preliminary CCT and accurate BC phylogenies (additional file 1 and Figure 3, respectively), suggesting that bacterial donors are at the origin to these sequences. However, eukaryotic ACC sequences form extremely divergent groups in both BC and CCT phylogenies, probably as a consequence of rapid evolution subsequent to the fusion of these domains to generate the eukaryotic ACC polypeptide. Thus, reconstruction artifacts are likely and little confidence can be given to the position of eukaryotic sequences in these phylogenies. On the other hand, protein domain composition of eukaryotic ACC is very different from those of PYCs and bacterial ACCs, but shares strong similarity with that of XCC (Figure 2). Huang *et al.* proposed that the most likely way to explain the unique domain composition of eukaryotic ACC is a fusion event between the α - and β -subunits of XCC on both sides of a central linker [27]. The elucidation of the function of the BT/PT domain of eukaryotic ACCs will be necessary to decide between the XCC and the bacterial ACC origin for the eukaryotic ACC polypeptides.

Finally, the BC phylogeny indicates that UCA genes were acquired recently and independently by fungi and green algae from bacteria. Whereas eukaryotes have maintained the same domain structure than their bacterial counterparts for PYC, MCC, PCC and bacterial-type ACC, the components of eukaryotic ACC and UCA have fused to generate large polypeptides.

Biotin-dependent carboxylases in Archaea

Although recent studies have revealed the importance of biotin-dependent carboxylases in Archaea, little is known about their origin and evolution [28,29,40]. These enzymes are much less abundant and diversified in Archaea than in the other two domains of life, but their study could be of major interest in understanding the emergence and ancient evolution of several central metabolic pathways in these organisms [58]. In Archaea, the BC domain is encoded by an independent subunit in both the PYC and the ACC-PCC complexes, instead of being a part of wider polypeptides. BC sequences are widespread among archaea and their phylogeny shows that major archaeal groups form monophyletic clades, which are more closely related among them than to any other taxon. This evidence supports the ancestral presence of the BC domain in Archaea and points to its vertical inheritance in this domain of life. Surprisingly, archaeal BC sequences do not cluster together with regard to the PYC and ACC-PCC functional groups, but rather following the accepted phylogeny of Archaea. This strongly supports the hypothesis for the presence of one unique promiscuous BC subunit in the last common archaeal ancestor that would have been recruited in both, PYC and ACC-PCC functions. In the other hand, the PCT phylogeny and its distribution in archaea also indicate that the PYC function may have existed in the archaeal common ancestor. Concerning the ACC-PCC function, the wide distribution of the CCT domain in archaea suggests that it could also be ancestral. Moreover, even though archaeal CCT sequences were not found to be monophyletic in our CCT phylogenies, the AU test could not reject the hypothesis of the archaeal monophyly, and thus the presence of a CCT subunit in the archaeal common ancestor can be hypothesized. As a result, one unique promiscuous BC subunit that collaborated with PCT and CCT subunits to catalyze their specific reactions may have existed in the last common archaeal ancestor. Interestingly, such a collaboration of a BC-encoding subunit with both the CCT and PCT systems has been suggested in *Archaeoglobus fulgidus* [59] although, to our knowledge, this has not yet been experimentally tested. Our analyses support to extend this possibility to the ancestor of Archaea.

An alternative origin of archaeal biotin carboxylases through ancient HGTs from bacteria cannot be completely

excluded. However, except for some particular cases limited to some specific organisms, the distribution and the phylogenetic trees of the BC and the PCT domains are congruent with the expected phylogeny of Archaea. As a result, the hypothetical HGT events responsible for the presence of the BC and PCT domains in Archaea should have been ancient enough to predate the last common archaeal ancestor, and thus HGT cannot be favored over the simple vertical inheritance from the cenancestor. Concerning the CCT domain, some groups, as for example the Halobacteriales, may have inherited their sequences by HGT from bacterial donors, but our results do not provide enough support neither to the monophyly of all archaeal CCTs nor to the putative HGT-mediated origin of these sequences, so this issue remains open.

Relatively recent HGT events cannot be ignored in the study of the evolutionary history of biotin-dependent carboxylases. They may have played an important role, both in the case of closely related organisms (for example, HGTs of XCC among Actinobacteria), as well as for distantly related taxa (e.g., the archaeal-type PYC transferred to some aquificales and epsilon-proteobacterial species). HGTs may have also been at the origin of completely new functions. For instance, bacterial ODC has been recently proposed to use one decarboxylating subunit common to all decarboxylases and a CT domain from any other origin [60]. ODC-CT is restricted to a limited number of anaerobic bacteria and is closely related to archaeal PCT, what supports the archaeal origin of ODC-CT. Possible criticism against this HGT hypothesis could be that the bacterial ODC sequences do not branch within the archaeal PCT sequences but as a sister group (Figure 6). Nevertheless, the divergence of the bacterial sequences in order to acquire their new function is likely to have induced a long branch attraction artifact in this phylogeny.

Occurrence of BCCP: BPL-BirA evolution in the three domains of life

Contrarily to the rest of the protein domains involved in biotin-dependent carboxylation, we did not analyze the evolution of the BCCP domain because of the limited phylogenetic signal due to its small size (~ 90 amino acids) and low sequence conservation, which make the BCCP phylogenies very sensitive to reconstruction artifacts. In addition, this domain can be found in many different polypeptide architectures, including transcarboxylases and decarboxylases that are not part of our work. A phylogenetic analysis of BCCPs published some years ago [59] supported the occurrence of many different fusion and fission events with the PCT domains and only one fusion event with the BC domain in the XCC sequences. Such complex fusion and fission scenario should be balanced by the recent discovery of the PT domains in PYC sequences

[36], which makes independent fusions much less parsimonious than previously suggested.

We analyzed the biotin protein ligase (BPL) protein domain instead of the BCCP domain itself in order to reconstruct a more reliable phylogeny that could provide indirect evidence of the BCCP content in the cenancestor. BPL is related to the lipoyl protein ligases [21] and is responsible for the specific biotinylation of BCCP. In prokaryotes, some species bear a BPL gene, whereas others contain BirA genes. BirA proteins consist of a transcriptional regulation domain in the N-terminus and a BPL domain in the C-terminus. Because BirA could be found in both bacteria and archaea, it has been proposed to be the most ancient known transcriptional factor, already present in the cenancestor [22]. Nevertheless, universal phylogenetic analyses of BPL and BirA had never been carried out to test this hypothesis.

Our results confirmed previous observations of single BPL or BirA genes in each species. Except for some likely HGTs events, our phylogenetic reconstructions of the BPL protein domain showed a phylogeny congruent with the main accepted taxonomic groups in the three domains of life, suggesting a dominant vertical inheritance from the common ancestor to all living organisms (Figure 7). The most puzzling feature of this phylogeny was the distribution of the sequences bearing the N-terminal transcriptional regulation domain characteristic of BirA all over the prokaryotic part of the BPL tree and mixed up with the sequences lacking that domain. This observation could be explained either by an ancestral BPL gene that would have fused many independent times with the N-terminal transcriptional regulator to form different BirA genes, or by an ancestral BirA gene that would have lost its N-terminal regulatory domain in many lineages independently. As the latter hypothesis appears more parsimonious than the former, our results tend to support the BirA gene to be ancestral to all organisms, as previously proposed [22]. Finally, this evidence supporting the presence of the BPL function in the cenancestor is an additional indirect argument in favor of the presence of the BCCP domain proteins in this ancestral organism.

Conclusions

To summarize, our data support that ACC, XCC and PYC are ancestral to Bacteria. MCC, PCC, GCC and UCA emerged subsequently from XCC in the bacterial lineage. Eukaryotes most likely inherited their biotin-dependent carboxylases through different HGTs from bacteria: MCC and PCC are related to alphaproteobacterial homologues and therefore seem to have been acquired through the mitochondrial endosymbiosis, whereas plant bacterial-like ACC is related to cyanobacterial homologues and thus of probable plastid origin. The remaining eukaryotic enzymes (ACC, PYC and UCA) have likely

been also transferred from bacteria but their donors are unknown. The last common archaeal ancestor likely used a promiscuous BC subunit with PCT and CCT subunits to ensure PYC and ACC/PCC functions, respectively. Finally, the cenancestor likely bore a BirA gene able to biotinylate BCCPs. We used these results to infer the set of biotin-dependent carboxylases present in the cenancestor. It is important to note here that the precise nature of the cenancestor remains debated, in particular whether it was a single organism or a community of more or less related different organisms experiencing a high frequency of HGT [61,62]. When we infer that a particular biotin-dependent carboxylase may have been ancestral, this is applicable to both views.

Since eukaryotes obtained their biotin-dependent carboxylases from bacteria, we ignore them for the discussion concerning the cenancestor complement and we focus specifically on the respective ancestors of Archaea and Bacteria as intermediate steps between present-day species and the cenancestor. The components of biotin-dependent carboxylases have been duplicated, recombined and fused many times across evolution and, thus, many different evolutionary scenarios can be proposed. As it would be too long to discuss all of them, we will focus only on the one that we consider to be the most parsimonious (for examples of other scenarios see additional file 3).

In this hypothesis (Figure 8), the cenancestor would have had a relatively simple biotin-dependent carboxylase content similar to that of modern archaea: one promiscuous BC-bearing subunit would have interacted with a PCT-PT-BCCP subunit to catalyze the PYC function and also with independent BCCP and CCT subunits to carry out the carboxylation of CoA-substrates. Ancient archaea would have inherited, conserved and adapted this content to their low biotinylating needs, keeping one PYC and one ACC-PCC, either conserved or lost across subsequent evolution of the different archaeal phyla. In the bacterial lineage, the duplication of the BC domain allowed the emergence of two different biotin-dependent carboxylases, a polypeptidic PYC and one ancient CoA-substrate carboxylase made up of three subunits. The polypeptidic PYC was vertically inherited in bacteria whereas a duplication of the three-subunit CoA-substrate carboxylase took place before the last common bacterial ancestor. One of the resulting CoA-substrate carboxylases had its CCT subunit split in two parts that became the two CCT subunits in the bacterial ACCs. In the other three-subunit enzyme there was a fusion between the BC and BCCP domains to give rise to an ancestral promiscuous XCC. Later duplications and subfunctionalizations lead to the emergence of the very diverse XCC family that we know in contemporary bacteria.

A strong point of this scenario is its relative simplicity, relying on the general assumption of aggregative peptide domain architecture as a major force in protein evolution [63]. Noteworthy, this hypothesis assumes the independent emergence of the PT (ancient PYC) domain and the BT (bacterial XCC) domain. Thus, a convergent evolution has to be invoked to explain BT and PT structural resemblances. Despite that little is known concerning the characteristics and conservation of the BT/PT domains, their shared structure consisting of a conserved helix and several β -strands seems simple enough to hypothesize that it could have emerged twice independently and their conserved position between the BC and the BCCP domains could be the result of structural constraints related to their common subunit-interaction roles.

Although some other scenarios could be proposed based on the results of our phylogenetic and protein domain composition analyses, all of them support the presence of a relatively complex biotin-dependent carboxylase complement in the cenancestor. It would have been able to carboxylate the pyruvate and several CoA-linked substrates. Contemporary biotin-dependent carboxylases contribute to very diverse major biological functions, such as fatty acid biosynthesis, anaplerosis, gluconeogenesis, and autotrophic CO₂ fixation. Therefore, identifying the precise activity of the ancient biotin-dependent carboxylases proposed to have been present in the cenancestor is a very difficult question. Phylogenetic analyses alone do not provide enough information to clarify this issue. A promising approach would be the bioinformatic inference of the sequences of the ancestral enzymes to synthesize them in order to characterize their activity in detail (what is often called "enzyme resurrection" [64-66]). That would be particularly interesting to check for potential activities involved in fatty acid biosynthesis or autotrophic CO₂ fixation since the presence of these metabolic pathways in the cenancestor remains a hotly debated question [67-70].

Methods

Sequence retrieval and alignment

For each domain of life for which sequence data was available, we retrieved one representative of each of the different biotin-dependent carboxylases and BPL/BirA enzymes from the KEGG database (<http://www.genome.jp/kegg>) to be used as seeds for further similarity searches. Since some biotin-dependent carboxylases were absent from this database, we completed with archaeal ACC/PCC [29], bacterial GCC [9], one bacterial UCA [44] and one proteobacterial ODC [60] sequence obtained from GenBank (<http://www.ncbi.nlm.nih.gov/Genbank>). Similarity searches with BLASTp [71] were done with the well-characterized protein domains

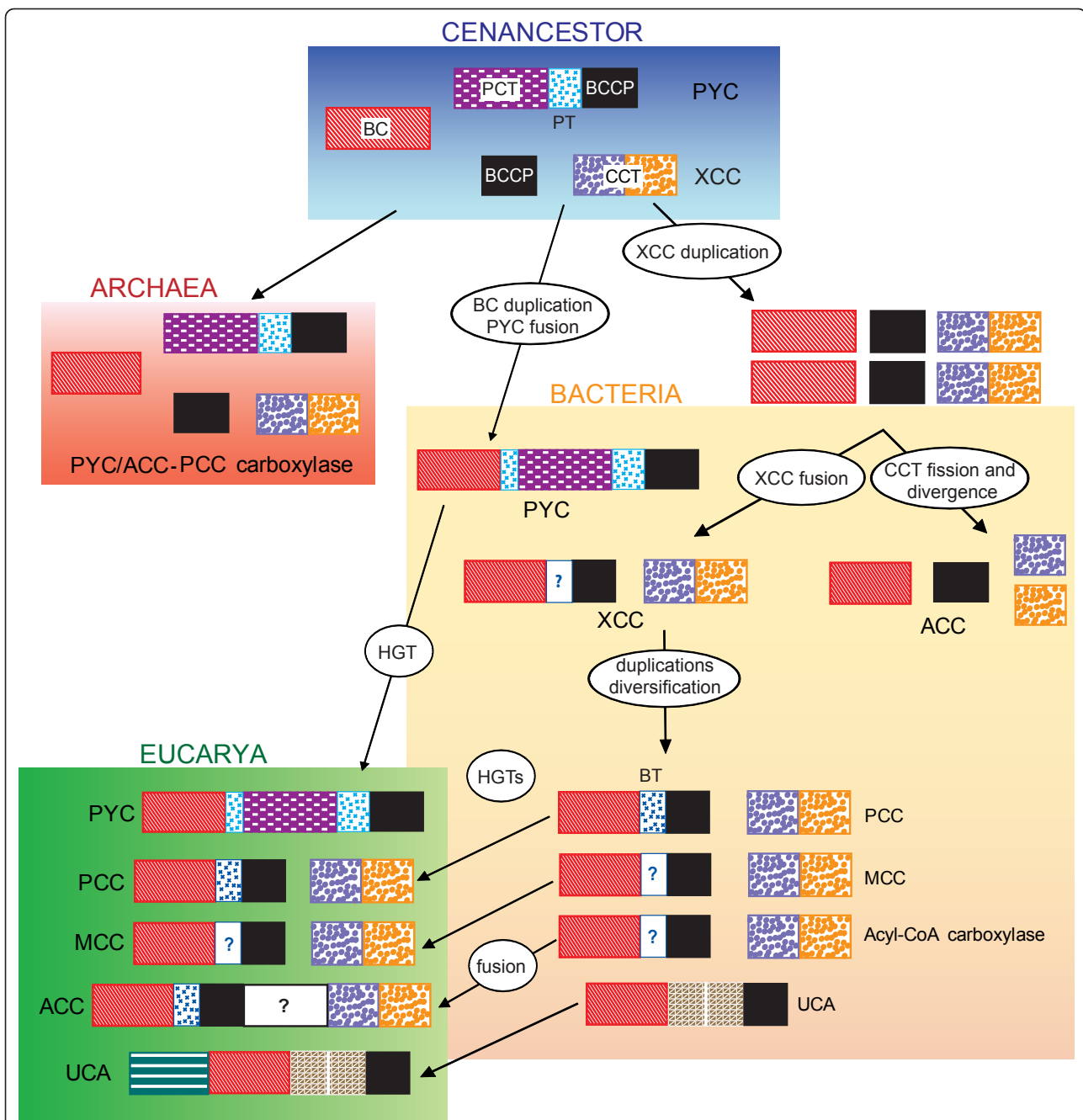


Figure 8 Evolutionary hypothesis for the early diversification of the biotin-dependent carboxylases. Colored blocks represent functional protein domains. A biotin-dependent carboxylase content similar to that of modern archaea was present in the cenacestor. It was conserved in Archaea but in Bacteria these enzymes were duplicated and fused several times, bearing to the very diverse modern biotin-dependent carboxylase family found in these organisms. Eukaryotes received their enzymes by horizontal gene transfer (HGT).

contained in these representative sequences as queries against their respective domain of life. In cases where a particular enzyme was missing in KEGG for one domain of life, we used sequences from the other domains as queries. Similarity searches in archaea and bacteria were done against a list of completely sequenced genomes

available in GenBank (298 bacteria and 55 archaea, additional file 4). In eukaryotes, all searches were done against the complete non-redundant (nr) eukaryote-annotated GenBank database.

Sequences for each protein domain found by these searches in the three domains of life were aligned with

Muscle 3.6 [72] or MAFFT v6.814c-b [73]. Alignments were edited with the program ED of the MUST package [74] and redundant and partial sequences were removed at this step. Ambiguously aligned regions were removed prior to phylogenetic analyses using the NET program from the MUST package. Alignments are available in Nexus format as additional files 5, 6, 7, 8 and 9. Preliminary secondary structure searches on MCC, GCC, PYC, XCC and accE (see results) were carried out using APSSP (Advanced Protein Secondary Structure Prediction Server, <http://imtech.res.in/raghava/apssp/>) and GOR4 [75].

Phylogenetic analyses

Preliminary trees based on the complete sequence dataset for each enzyme were constructed by the approximately maximum likelihood approach with FastTree 2.1.3 [76] in order to classify sequences in functional classes with respect to well-characterized proteins (see additional file 1). Neighbor joining trees (NJ) [77] using the MUST package [74] were also reconstructed to select representative sequences with which carrying out more detailed maximum likelihood (ML) and Bayesian inference (BI) phylogenetic analyses. ML tree reconstructions were done with the program TREEFINDER [78] with the LG + Γ model [79] and 4 rate categories, which was selected as the best-fit model for all our datasets by the model selection tool implemented in TREEFINDER [78]. Node support was assessed by 1,000 bootstrap replicates with the same model. BI trees were reconstructed using the program MrBayes v. 3.0b4 [80] with a mixed substitution model and a Γ distribution of substitution rates with 4 categories. Searches were run with 4 chains of 1,000,000 generations for which the first 2,500 generations were discarded as "burn in", trees being sampled every 100 generations. Stabilization of the chain parameters was verified using the program TRACER [81]. Approximately unbiased tests [82] were carried out using the test tool implemented in TREEFINDER [78].

Additional material

Additional file 1: Schematic phylogenetic trees of biotin carboxylase (BC) sequences and CoA-substrate related carboxyl transferase (CCT) sequences. These preliminary FastTree reconstructions allowed the functional assignment of each sequence according to a phylogenetic framework and functional data from the available literature.

Additional file 2: Biotin-dependent carboxylase content in bacterial complete genome sequences. This is an Excel file showing the presence or absence of homologues of the different biotin-dependent carboxylases in a collection of complete genome sequences.

Additional file 3: Alternative evolutionary scenarios for the early diversification of the biotin-dependent carboxylases. Colored blocks represent functional protein domains. The first scenario postulates the duplication of an ancient biotin-dependent carboxylase and the insertion of a pyruvate carboxylase carboxyl transferase (PCT) within the BT/PT

domain predated the cenacestator. This organism had a polypeptidic PYC and a two-subunit XCC that evolved through several duplications and splits in Archaea and Bacteria. The second scenario explores the possibility that the cenacestator bore a CCT divided in two subunits as in contemporary bacterial ACC.

Additional file 4: List of all complete genome sequences used in this work. This is a text file containing the complete list of species upon which we carried out sequence similarity searches to detect biotin-dependent carboxylase homologues.

Additional file 5: Masked multiple sequence alignment of biotin carboxylase (BC) domain sequences (NEXUS format). This alignment was used to reconstruct the tree shown in Figure 3.

Additional file 6: Masked multiple sequence alignment of archaeal and bacterial biotin carboxylase (BC) domain sequences (NEXUS format). This alignment was used to reconstruct the tree shown in Figure 4.

Additional file 7: Masked multiple sequence alignment of CoA-substrate related carboxyl transferase (CCT) domain sequences (NEXUS format). This alignment was used to reconstruct the tree shown in Figure 5.

Additional file 8: Masked multiple sequence alignment of pyruvate carboxylase carboxyl transferase (PCT) domain sequences (NEXUS format). This alignment was used to reconstruct the tree shown in Figure 6.

Additional file 9: Masked multiple sequence alignment of biotin carboxyl ligase (BPL) sequences (NEXUS format). This alignment was used to reconstruct the tree shown in Figure 7.

Acknowledgements

This work was supported by the interdisciplinary programs OPV (Origine des Planètes et de la Vie) and InTerVie (Interactions Terre-Vie) of the French Centre National de la Recherche Scientifique and Institut National des Sciences de l'Univers. JL is recipient of a PhD fellowship of the French Research Ministry. We thank Philippe Deschamps for critical reading of the manuscript and four anonymous reviewers for useful suggestions.

Authors' contributions

JL and DM designed research; JL carried out phylogenetic analyses, and JL and DM wrote the manuscript. All authors read and approved the final version.

Received: 4 March 2011 Accepted: 9 August 2011
Published: 9 August 2011

References

1. Barber MC, Price NT, Travers MT: **Structure and regulation of acetyl-CoA carboxylase genes of metazoa.** *Biochim Biophys Acta* 2005, **1733**:1-28.
2. Bramwell H, Hunter IS, Coggins JR, Nimmo HG: **Propionyl-CoA carboxylase from *Streptomyces coelicolor* A3(2): cloning of the gene encoding the biotin-containing subunit.** *Microbiology* 1996, **142**:649-655.
3. Dunn MF, Encarnacion S, Araiza G, Vargas MC, Davalos A, Peralta H, Mora Y, Mora J: **Pyruvate carboxylase from *Rhizobium etli*: mutant characterization, nucleotide sequence, and physiological role.** *J Bacteriol* 1996, **178**:5960-5970.
4. Kondo H, Kazuta Y, Saito A, Fuji K: **Cloning and nucleotide sequence of *Bacillus stearothermophilus* pyruvate carboxylase.** *Gene* 1997, **191**:47-50.
5. Kimura Y, Sato R, Mimura K, Sato M: **Propionyl coenzyme A carboxylase is required for development of *Myxococcus xanthus*.** *J Bacteriol* 1997, **179**:7098-7102.
6. Menendez C, Bauer Z, Huber H, Gad'on N, Stetter KO, Fuchs G: **Presence of acetyl coenzyme A (CoA) carboxylase and propionyl-CoA carboxylase in autotrophic Crenarchaeota and indication for operation of a 3-hydroxypropionate cycle in autotrophic carbon fixation.** *J Bacteriol* 1999, **181**:1088-1098.
7. Rodriguez E, Banchio C, Diacovich L, Bibb MJ, Gramajo H: **Role of an essential acyl coenzyme A carboxylase in the primary and secondary**

- metabolism of *Streptomyces coelicolor* A3(2). *Applied Environ Microbiol* 2001, **67**:4166-4176.
8. Cronan JE Jr, Waldrop GL: Multi-subunit acetyl-CoA carboxylases. *Progress Lipid Research* 2002, **41**:407-435.
 9. Aguilar JA, Zavala AN, Diaz-Perez C, Cervantes C, Diaz-Perez AL, Campos-Garcia J: The *atu* and *liu* clusters are involved in the catabolic pathways for acyclic monoterpenes and leucine in *Pseudomonas aeruginosa*. *Applied Environ Microbiol* 2006, **72**:2070-2079.
 10. Tong L: Acetyl-coenzyme A carboxylase: crucial metabolic enzyme and attractive target for drug discovery. *Cell Mol Life Sci* 2005, **62**:1784-1803.
 11. Jitrapakdee S, Wallace JC: The biotin enzyme family: conserved structural motifs and domain rearrangements. *Current Protein Peptide Sci* 2003, **4**:217-229.
 12. Chapman-Smith A, Cronan JE Jr: In vivo enzymatic protein biotinylation. *Biomol Engineering* 1999, **16**:119-125.
 13. Pardini NR, Bailey LM, Booker GW, Wilce MC, Wallace JC, Polyak SW: Microbial biotin protein ligases aid in understanding holocarboxylase synthetase deficiency. *Biochim Biophys Acta* 2008, **1784**:973-982.
 14. Bagautdinov B, Kuroishi C, Sugahara M, Kunishima N: Crystal structures of biotin protein ligase from *Pyrococcus horikoshii* OT3 and its complexes: structural basis of biotin activation. *J Mol Biol* 2005, **353**:322-333.
 15. Bagautdinov B, Matsuura Y, Bagautdinova S, Kunishima N: Protein biotinylation visualized by a complex structure of biotin protein ligase with a substrate. *J Biol Chem* 2008, **283**:14739-14750.
 16. McAllister HC, Coon MJ: Further studies on the properties of liver propionyl coenzyme A holocarboxylase synthetase and the specificity of holocarboxylase formation. *J Biol Chem* 1966, **241**:2855-2861.
 17. Leon-Del-Rio A, Gravel RA: Sequence requirements for the biotinylation of carboxyl-terminal fragments of human propionyl-CoA carboxylase alpha subunit expressed in *Escherichia coli*. *J Biol Chem* 1994, **269**:22964-22968.
 18. Mukhopadhyay B, Purwantini E, Kreder CL, Wolfe RS: Oxaloacetate synthesis in the methanarchaeon *Methanosarcina barkeri*: pyruvate carboxylase genes and a putative *Escherichia coli*-type bifunctional biotin protein ligase gene (*bpl/birA*) exhibit a unique organization. *J Bacteriol* 2001, **183**:3804-3810.
 19. Barker DF, Campbell AM: The *birA* gene of *Escherichia coli* encodes a biotin holoenzyme synthetase. *J Mol Biol* 1981, **146**:451-467.
 20. Bower S, Perkins J, Yocum RR, Serron P, Sorokin A, Rahaim P, Howitt CL, Prasad N, Ehrlich SD, Pero J: Cloning and characterization of the *Bacillus subtilis* *birA* gene encoding a repressor of the biotin operon. *J Bacteriol* 1995, **177**:2572-2575.
 21. Reche PA: Lipoylating and biotinylation enzymes contain a homologous catalytic module. *Protein Sci* 2000, **9**:1922-1929.
 22. Rodionov DA, Mironov AA, Gelfand MS: Conservation of the biotin regulon and the BirA regulatory signal in Eubacteria and Archaea. *Genome Res* 2002, **12**:1507-1516.
 23. Wood ZA, Weaver LH, Brown PH, Beckett D, Matthews BW: Co-repressor induced order and biotin repressor dimerization: a case for divergent followed by convergent evolution. *J Mol Biol* 2006, **357**(2):509-523.
 24. Daniels KG, Beckett D: Biochemical properties and biological function of a monofunctional microbial biotin protein ligase. *Biochemistry* 2010, **49**:5358-5365.
 25. Alban C: Is plant biotin holocarboxylase synthetase a bifunctional enzyme? *C R Acad Sci III* 2000, **323**:681-688.
 26. Pirner HM, Stolz J: Biotin sensing in *Saccharomyces cerevisiae* is mediated by a conserved DNA element and requires the activity of biotin-protein ligase. *J Biol Chem* 2006, **281**:12381-12389.
 27. Huang CS, Sadre-Bazzaz K, Shen Y, Deng B, Zhou ZH, Tong L: Crystal structure of the alpha(6)beta(6) holoenzyme of propionyl-coenzyme A carboxylase. *Nature* 2010, **466**:1001-1005.
 28. Chuakrut S, Arai H, Ishii M, Igarashi Y: Characterization of a bifunctional archaeal acyl coenzyme A carboxylase. *J Bacteriol* 2003, **185**:938-947.
 29. Hugler M, Krieger RS, Jahn M, Fuchs G: Characterization of acetyl-CoA/propionyl-CoA carboxylase in *Metallosphaera sedula*. Carboxylating enzyme in the 3-hydroxypropionate cycle for autotrophic carbon fixation. *Europ J Biochem* 2003, **270**:736-744.
 30. Aguilar JA, Diaz-Perez C, Diaz-Perez AL, Rodriguez-Zavala JS, Nikolau BJ, Campos-Garcia J: Substrate specificity of the 3-methylcrotonyl coenzyme A (CoA) and geranyl-CoA carboxylases from *Pseudomonas aeruginosa*. *J Bacteriol* 2008, **190**:4888-4893.
 31. Kimura Y, Kojito T, Kimura I, Sato M: Propionyl-CoA carboxylase of *Myxococcus xanthus*: catalytic properties and function in developing cells. *Arch Microbiol* 1998, **170**:179-184.
 32. Gago G, Kurth D, Diacovich L, Tsai SC, Gramajo H: Biochemical and structural characterization of an essential acyl coenzyme A carboxylase from *Mycobacterium tuberculosis*. *J Bacteriol* 2006, **188**:477-486.
 33. Diacovich L, Mitchell DL, Pham H, Gago G, Melgar MM, Khosla C, Gramajo H, Tsai SC: Crystal structure of the beta-subunit of acyl-CoA carboxylase: structure-based engineering of substrate specificity. *Biochemistry* 2004, **43**:14027-14036.
 34. Arabolaza A, Shillito ME, Lin TW, Diacovich L, Melgar M, Pham H, Amick D, Gramajo H, Tsai SC: Crystal structures and mutational analyses of acyl-CoA carboxylase beta subunit of *Streptomyces coelicolor*. *Biochemistry* 2010, **49**:7367-7376.
 35. Jitrapakdee S, Wallace JC: Structure, function and regulation of pyruvate carboxylase. *Biochem J* 1999, **340**:1-16.
 36. St Maurice M, Reinhardt L, Surinya KH, Attwood PV, Wallace JC, Cleland WW, Rayment I: Domain architecture of pyruvate carboxylase, a biotin-dependent multifunctional enzyme. *Science* 2007, **317**:1076-1079.
 37. Xiang S, Tong L: Crystal structures of human and *Staphylococcus aureus* pyruvate carboxylase and molecular insights into the carboxyltransfer reaction. *Nature Struct Mol Biol* 2008, **15**:295-302.
 38. Jitrapakdee S, St Maurice M, Rayment I, Cleland WW, Wallace JC, Attwood PV: Structure, mechanism and regulation of pyruvate carboxylase. *Biochem J* 2008, **413**:369-387.
 39. Goss JA, Cohen ND, Utter MF: Characterization of the subunit structure of pyruvate carboxylase from *Pseudomonas citronellolis*. *J Biol Chem* 1981, **256**:11819-11825.
 40. Mukhopadhyay B, Stoddard SF, Wolfe RS: Purification, regulation, and molecular and biochemical characterization of pyruvate carboxylase from *Methanobacterium thermoautotrophicum* strain deltaH. *J Biol Chem* 1998, **273**:5155-5166.
 41. Mukhopadhyay B, Patel VJ, Wolfe RS: A stable archaeal pyruvate carboxylase from the hyperthermophile *Methanococcus jannaschii*. *Archiv Microbiol* 2000, **174**:406-414.
 42. Kondo S, Nakajima Y, Sugio S, Yong-Biao J, Sueda S, Kondo H: Structure of the biotin carboxylase subunit of pyruvate carboxylase from *Aquifex aeolicus* at 2.2 Å resolution. *Acta Crystallograph* 2004, **60**:486-492.
 43. Sumrada RA, Cooper TG: Urea carboxylase and allophanate hydrolase are components of a multifunctional protein in yeast. *J Biol Chem* 1982, **257**:9119-9127.
 44. Kanamori T, Kanou N, Atomi H, Imanaka T: Enzymatic characterization of a prokaryotic urea carboxylase. *J Bacteriol* 2004, **186**:2532-2539.
 45. Obermayer M, Lynen F: Structure of biotin enzymes. *Trends Biochem Sci* 1976, **1**:169-171.
 46. Wood HG, Barden RE: Biotin enzymes. *Annual Rev Biochem* 1977, **46**:385-413.
 47. Toh H, Kondo H, Tanabe T: Molecular evolution of biotin-dependent carboxylases. *Europ J Biochem* 1993, **215**:687-696.
 48. Balsera M, Buey RM, Li XD: Quaternary structure of the oxaloacetate decarboxylase membrane complex and mechanistic relationships to pyruvate carboxylases. *J Biol Chem* 2011, **286**:9457-9467.
 49. Diacovich L, Peiru S, Kurth D, Rodriguez E, Podesta F, Khosla C, Gramajo H: Kinetic and structural analysis of a new group of Acyl-CoA carboxylases found in *Streptomyces coelicolor* A3(2). *J Biol Chem* 2002, **277**:31228-31236.
 50. Tron CM, McNaie IW, Nutley M, Clarke DJ, Cooper A, Walkinshaw MD, Baxter RL, Campopiano DJ: Structural and functional studies of the biotin protein ligase from *Aquifex aeolicus* reveal a critical role for a conserved residue in target specificity. *J Mol Biol* 2009, **387**:129-146.
 51. Gupta V, Gupta RK, Khare G, Salunke DM, Suroliya A, Tyagi AK: Structural ordering of disordered ligand-binding loops of biotin protein ligase into active conformations as a consequence of dehydration. *PLoS One* 2010, **5**:e9222.
 52. Jensen RA: Enzyme recruitment in evolution of new function. *Annual Rev Microbiol* 1976, **30**:409-425.
 53. Khersovany O, Tawfik DS: Enzyme promiscuity: a mechanistic and evolutionary perspective. *Annual Rev Biochem* 2010, **79**:471-505.
 54. Gande R, Dover LG, Krumbach K, Besra GS, Sahn H, Oikawa T, Eggeling L: The two carboxylases of *Corynebacterium glutamicum* essential for fatty acid and mycolic acid synthesis. *J Bacteriol* 2007, **189**:5257-5264.

55. Demirev AV, Khanal A, Sedai BR, Lim SK, Na MK, Nam DH: **The role of acyl-coenzyme A carboxylase complex in lipstatin biosynthesis of *Streptomyces toxytricini***. *Appl Microbiol Biotech* 2010, **87**:1129-1139.
56. Anderson MD, Che P, Song J, Nikolau BJ, Wurtele ES: **3-Methylcrotonyl-coenzyme A carboxylase is a component of the mitochondrial leucine catabolic pathway in plants**. *Plant Physiol* 1998, **118**:1127-1138.
57. Jiang H, Rao KS, Yee VC, Kraus JP: **Characterization of four variant forms of human propionyl-CoA carboxylase expressed in *Escherichia coli***. *J Biol Chem* 2005, **280**:27719-27727.
58. Berg IA, Kockelkorn D, Ramos-Vera WH, Say RF, Zarzycki J, Hugler M, Alber BE, Fuchs G: **Autotrophic carbon fixation in archaea**. *Nat Rev Microbiol* 2010, **8**:447-460.
59. Jordan IK, Henze K, Fedorova ND, Koonin EV, Galperin MY: **Phylogenomic analysis of the *Giardia intestinalis* transcarboxylase reveals multiple instances of domain fusion and fission in the evolution of biotin-dependent enzymes**. *J Mol Microbiol Biotech* 2003, **5**:172-189.
60. Studer R, Dahinden P, Wang WW, Auchli Y, Li XD, Dimroth P: **Crystal structure of the carboxyltransferase domain of the oxaloacetate decarboxylase Na⁺ pump from *Vibrio cholerae***. *J Mol Biol* 2007, **367**:547-557.
61. Zhaxybayeva O, Gogarten JP: **Cladogenesis, coalescence and the evolution of the three domains of life**. *Trends Genet* 2004, **20**:182-187.
62. Moreira D, López-García P: **Prebiotic chemistry-Biochemistry-Emergence of life, 4.2-2 Ga: The last common ancestor**. *Earth Moon Planets* 2006, **98**:187-197.
63. Doolittle RF: **The multiplicity of domains in proteins**. *Ann Rev Biochem* 1995, **64**:287-314.
64. Perez-Jimenez R, Ingles-Prieto A, Zhao ZM, Sanchez-Romero I, Alegre-Cebollada J, Kosuri P, Garcia-Manyes S, Kappock TJ, Tanokura M, Holmgren A, et al: **Single-molecule paleoenzymology probes the chemistry of resurrected enzymes**. *Nat Struct Mol Biol* 2011, **18**:592-596.
65. Gaucher EA, Thomson JM, Burgan MF, Benner SA: **Inferring the palaeoenvironment of ancient bacteria on the basis of resurrected proteins**. *Nature* 2003, **425**:285-288.
66. Gaucher EA, Govindarajan S, Ganesh OK: **Palaeotemperature trend for Precambrian life inferred from resurrected proteins**. *Nature* 2008, **451**:704-707.
67. Martin W, Russell MJ: **On the origins of cells: a hypothesis for the evolutionary transitions from abiotic geochemistry to chemoautotrophic prokaryotes, and from prokaryotes to nucleated cells**. *Philos T R Soc* 2003, **358**:59-83.
68. Lombard J, Moreira D: **Origins and early evolution of the mevalonate pathway of isoprenoid biosynthesis in the three domains of life**. *Mol Biol Evol* 2011, **28**:87-99.
69. Pereto JG, Velasco AM, Becerra A, Lazcano A: **Comparative biochemistry of CO₂ fixation and the evolution of autotrophy**. *Int Microbiol* 1999, **2**:3-10.
70. Pereto J, Lopez-García P, Moreira D: **Ancestral lipid biosynthesis and early membrane evolution**. *Trends Biochem Sci* 2004, **29**:469-477.
71. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool**. *J Mol Biol* 1990, **215**:403-410.
72. Edgar RC: **MUSCLE: a multiple sequence alignment method with reduced time and space complexity**. *BMC Bioinformatics* 2004, **5**:113.
73. Katoh K, Misawa K, Kuma K, Miyata T: **MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform**. *Nucleic Acids Res* 2002, **30**:3059-3066.
74. Philippe H: **MUST, a computer package of Management Utilities for Sequences and Trees**. *Nucleic Acids Res* 1993, **21**:5264-5272.
75. Garnier J, Gibrat JF, Robson B: **GOR method for predicting protein secondary structure from amino acid sequence**. *Methods Enzymol* 1996, **266**:540-553.
76. Price MN, Dehal PS, Arkin AP: **FastTree 2—approximately maximum-likelihood trees for large alignments**. *PLoS One* 2010, **5**:e9490.
77. Saitou N, Nei M: **The neighbor-joining method: a new method for reconstructing phylogenetic trees**. *Mol Biol Evol* 1987, **4**:406-425.
78. Jobb G, von Haeseler A, Strimmer K: **TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics**. *BMC Evol Biol* 2004, **4**:18.
79. Le SQ, Gascuel O: **An improved general amino acid replacement matrix**. *Mol Biol Evol* 2008, **25**:1307-1320.
80. Ronquist F, Huelsenbeck JP: **MrBayes 3: Bayesian phylogenetic inference under mixed models**. *Bioinformatics* 2003, **19**:1572-1574.
81. Rambaut A, AJ D: **Tracer: a program for analysing results from Bayesian MCMC programs such as BEAST&MrBayes**. [http://tree.bio.ed.ac.uk/software/tracer/].
82. Shimodaira H: **An approximately unbiased test of phylogenetic tree selection**. *Syst Biol* 2002, **51**:492-508.

doi:10.1186/1471-2148-11-232

Cite this article as: Lombard and Moreira: Early evolution of the biotin-dependent carboxylase family. *BMC Evolutionary Biology* 2011 **11**:232.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit



This Provisional PDF corresponds to the article as it appeared upon acceptance. Fully formatted PDF and full text (HTML) versions will be made available soon.

Correction: Early evolution of the biotin-dependent carboxylase family

BMC Evolutionary Biology 2012, **12**:117 doi:10.1186/1471-2148-12-117

Jonathan Lombard (jonathan.lombard@u-psud.fr)
David Moreira (david.moreira@u-psud.fr)

ISSN 1471-2148

Article type Correction

Submission date 2 May 2012

Acceptance date 11 May 2012

Publication date 20 July 2012

Article URL <http://www.biomedcentral.com/1471-2148/12/117>

Like all articles in BMC journals, this peer-reviewed article was published immediately upon acceptance. It can be downloaded, printed and distributed freely for any purposes (see copyright notice below).

Articles in BMC journals are listed in PubMed and archived at PubMed Central.

For information about publishing your research in BMC journals or any BioMed Central journal, go to

<http://www.biomedcentral.com/info/authors/>

Correction: Early evolution of the biotin-dependent carboxylase family

Jonathan Lombard¹
Email: jonathan.lombard@u-psud.fr

David Moreira¹
* Corresponding author
Email: david.moreira@u-psud.fr

¹ Unité d'Ecologie, Systématique et Evolution, UMR CNRS 8079, Univ. Paris-Sud, Paris-Sud, Orsay Cedex 91405, France

Correction

After publication of our work [1], we noticed several major mistakes in the figure images provided for final publication: although the main text and the legends are correct, Figure 3 has been replaced by an image present in the Addition file 1 and Figures 4, 5 and 6 are displaced with regard to their correct numbers and legends. Please, accept our apologies and refer to the correct corresponding Figures 3, 4, 5 and 6 that we provide in this erratum. Legends are the same as in the original article.

Figure 1 Schematic diagram showing the reaction catalyzed by the biotin protein ligase (BPL) and the general mechanism shared by the biotin-dependent carboxylases. Biotin (in green) is first activated to biotinyl-5'-AMP by BPL, then it is transferred by the same enzyme to a biotin carboxyl carrier protein (BCCP). Biotin-dependent carboxylases (in purple) are made up of a biotin carboxylase domain (BC), a carboxyl transferase domain (CT) and the BCCP. The BC domain fixes a CO₂ molecule to the BCCP-bound biotin and then the CT component binds the carboxyl group to its specific substrate

Figure 2 Protein domain arrangements of different biotin protein ligases and biotin-dependent carboxylases with respect to their functions and phylogenetic origins. (A) Biotin protein ligases (BPL) and (B) biotin-dependent carboxylases. Homologous protein domains in different enzymes are shown in the same colors and filling. ACC: acetyl-CoA carboxylase; PCC: propionyl-CoA carboxylase; MCC: methylcrotonyl-CoA carboxylase; GCC: geranyl-CoA carboxylase; PYC: pyruvate carboxylase; UCA: urea carboxylase; BC: biotin carboxylase; BCCP: biotin carboxyl carrier protein; CCT: CoA-substrate related carboxyl transferase; PCT: PYC-related carboxyl transferase; UCT: UCA-related carboxyl transferase; BT: BC-CT interaction domain; PT: PYC tetramerization

Figure 3 Maximum likelihood tree of the biotin carboxylase (BC) domain. This tree is based on 284 representative sequences and 384 conserved sites and was arbitrarily rooted on the bacterial ACC-related sequences. Numbers at nodes indicate bootstrap values higher than 50. Triangles correspond to collapsed groups of eukaryotes and Proteobacteria. Colors on leaves represent the affiliation of the sequences to their respective domain of life: archaea (red), bacteria (blue) and eukaryotes (green). Bars on the right report the functional assignment of the sequences; sequences that are not in front of any bar are assumed to bear an acyl-CoA carboxylase activity

Figure 4 Maximum likelihood tree of archaeal and bacterial biotin carboxylase (BC) domain sequences. This tree is based on 196 representative sequences and 322 conserved sites and was arbitrarily rooted on the PYC-related sequences. Numbers at nodes indicate bootstrap robustness values higher than 50. Colors on leaves represent the affiliation of the sequences to their respective domain of life: archaea (red), bacteria (blue) and eukaryotes (green). Bars on the right report the functional assignment of the sequences; sequences that are not in front of any bar are assumed to bear an acyl-CoA carboxylase activity

Figure 5 Maximum likelihood tree of CoA-substrate related carboxyl transferase (CCT) domain sequences. This tree is based on 179 representative sequences and 438 conserved sites and was midpoint rooted. Numbers at nodes indicate bootstrap robustness values higher than 50. Triangles correspond to collapsed groups of eukaryotes. Colors on leaves represent the affiliation of the sequences to their respective domain of life: archaea (red), bacteria (blue) and eukaryotes (green). Bars on the right report the functional assignment of the sequences; sequences that are not in front of any bar are assumed to bear an acyl-CoA carboxylase activity

Figure 6 Maximum likelihood tree of pyruvate carboxylase carboxyl transferase (PCT) domain sequences. This tree is based on 126 representative sequences and 432 conserved sites and was midpoint rooted. Numbers at nodes indicate bootstrap robustness values higher than 50. Colors on leaves represent the affiliation of the sequences to their respective domain of life: archaea (red), bacteria (blue) and eukaryotes (green). Bars on the right report the functional assignment of the sequences; sequences that are not in front of any bar have unknown function

Competing interests

The authors declare no competing interests.

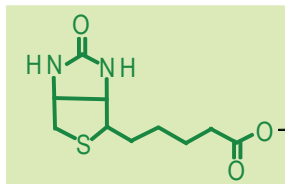
Authors' contributions

JL and DM designed research; JL carried out phylogenetic analyses, and JL and DM wrote the manuscript. All authors read and approved the final version.

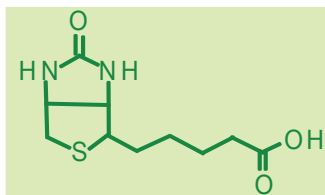
Reference

1. Lombard J, Moreira D: **Early evolution of the biotin-dependent carboxylase family.** *BMC Evol Biol* 2011, **11**:232.

BIOTINYL-5'-AMP



BIOTIN



PPI ATP

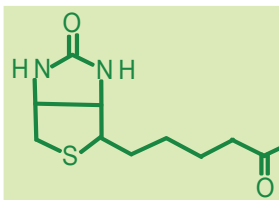
BPL

AP0-BCCP domain

H₂N

BCCP

AMP



BCCP

HOLO-BCCP domain (biotinylated protein)

Carboxylated substrate

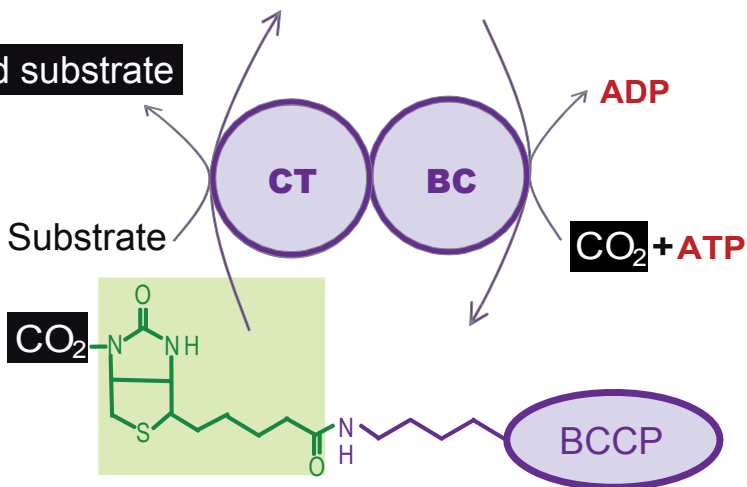


Figure 1

CARBOXYBIOTIN

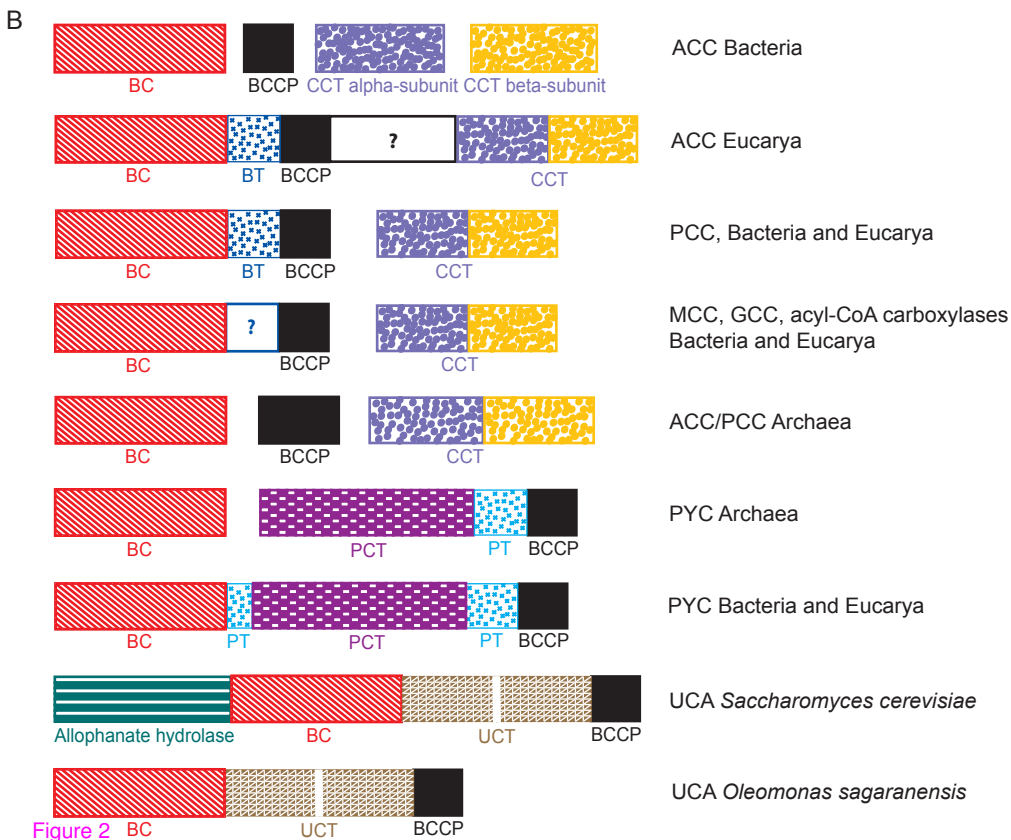
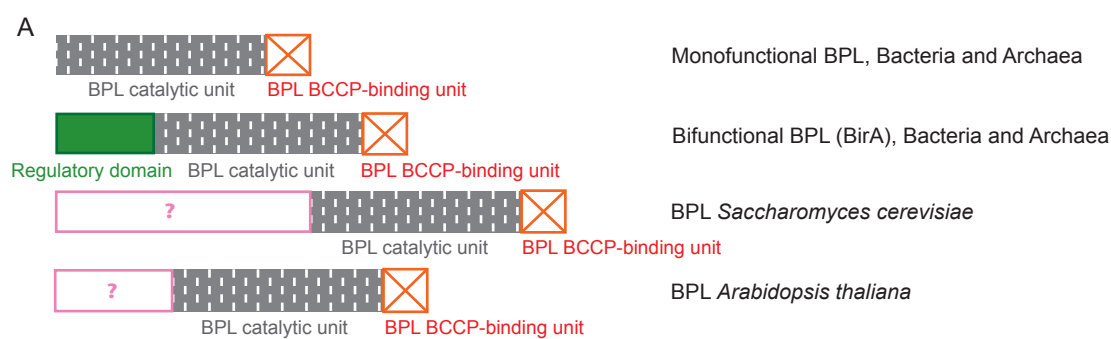


Figure 2 BC UCT BCCP

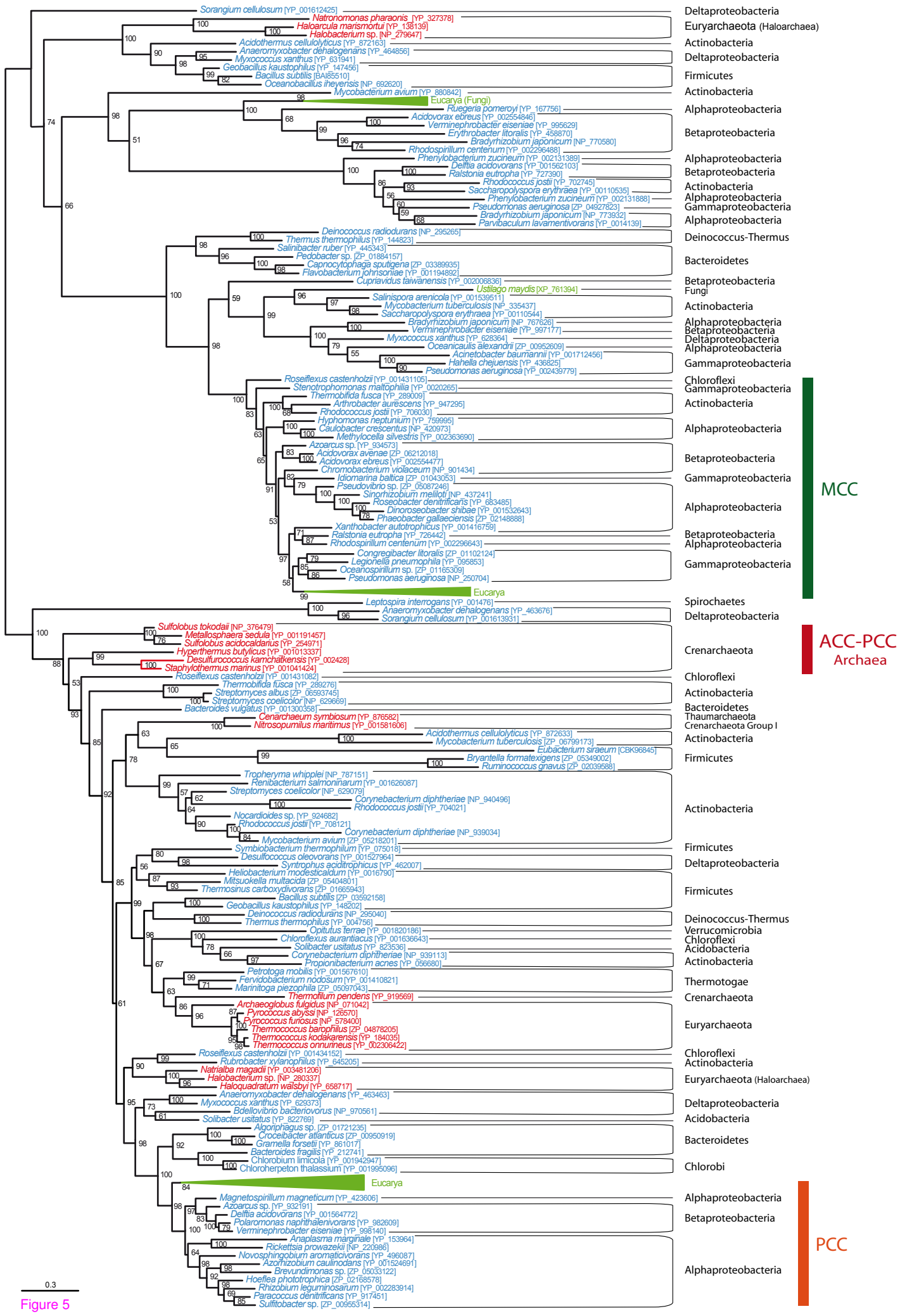


Figure 5

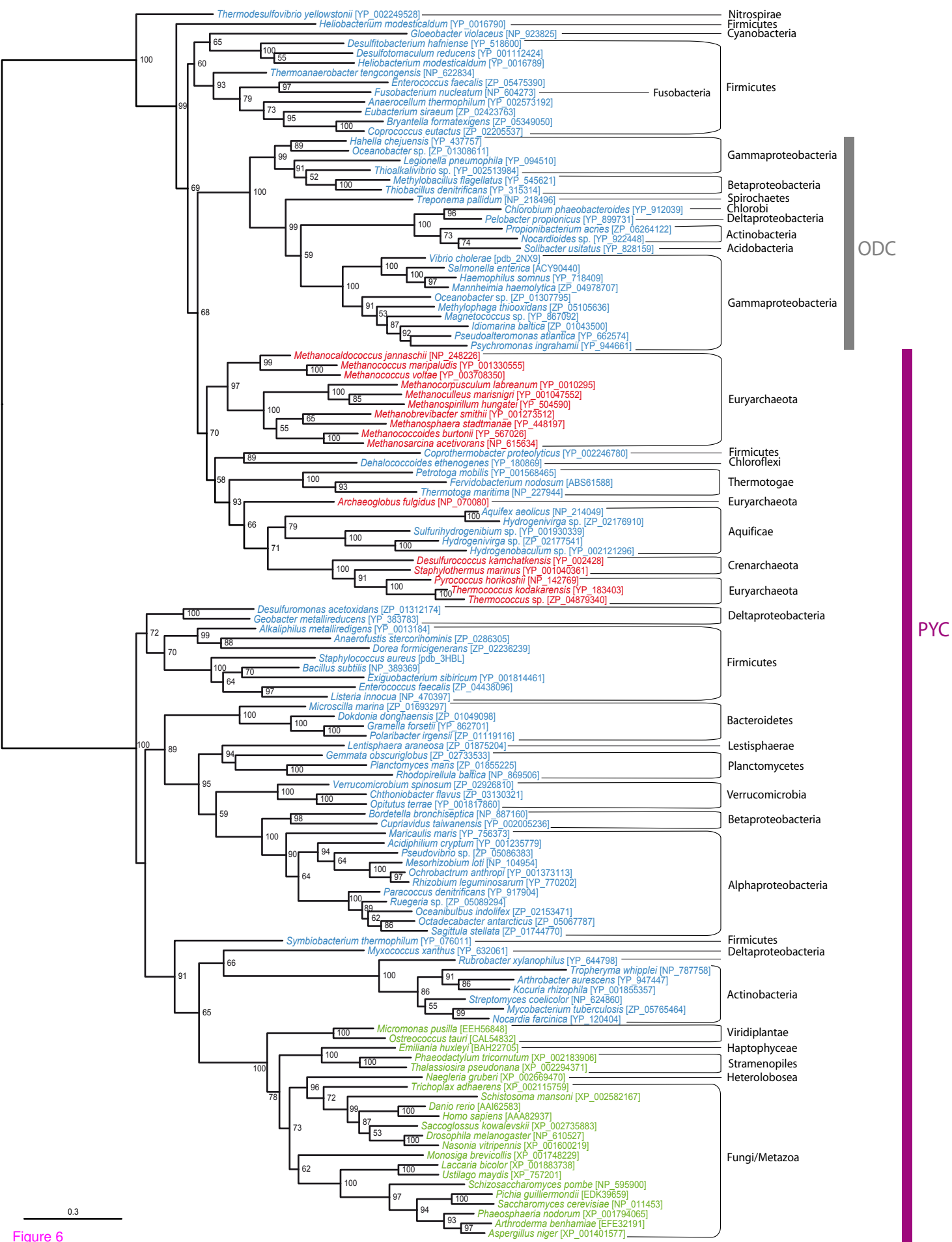


Figure 6

ANNEXE 4.

The membranes of the cenancestor.

Lombard, J., López-García P et D. Moreira. 2011.

In “LUCA: The Last Universal Cellular Ancestor of life”, Editeur : E. V. Koonin.

In press.

The membranes of the cenancestor

Jonathan Lombard, Purificación López-García and David Moreira

Abstract The encapsulation of a genetic system was a crucial early step in the origin and early evolution of life. Genome compartmentalization is indeed required for natural selection to act upon individual entities and to maintain a certain stability of individualized genomes over generations. Compartmentalization in all contemporary cells is made possible by lipid membranes carrying out essential functions. Consequently, the most parsimonious hypothesis to explain their origin would be that the last common ancestor of living organisms (the cenancestor) already had a lipid membrane. Although all cell membranes are made up of phospholipids, which are amphiphilic molecules sharing a common general structure, there are two types of phospholipids unevenly distributed in the three domains of life. Archaeal phospholipids consist on *sn*-glycerol-1-phosphate bound to isoprenoid chains through ether links whereas Bacteria and Eucarya use *sn*-glycerol-3-phosphate bound to fatty acids via ester links. Moreover, these dissimilarities are paralleled with differences in the corresponding biosynthetic pathways. As a result, a debate has emerged between authors postulating that membranes evolved independently in the modern domains of life from a cenancestor devoid of lipid membranes and authors supporting membrane continuity between a fully cellular cenancestor and present-day lineages. This question is of major importance since it strongly influences our vision of the cenancestor nature and properties. Here, we review this controversy as well as the main arguments that have been used for and against each opposing hypothesis, and show that, in our view, recent phylogenomic results tend to confirm the presence of lipid membranes in a complex modern-like cenancestor.

D. Moreira (✉)

CNRS UMR 8079, Université Paris-Sud, 91405 Orsay Cedex, France

e-mail: david.moreira@u-psud.fr

Introduction

Compartmentalization of a genetic system was a crucial, early step in the origin of life and the evolution of genomes. Not only it was probably necessary to secure a certain stability of the genetic information in primordial living entities, but it was also compulsory to individualize genomes into discrete entities that can evolve through generations (Szostak et al. 2001). With no exception, all contemporary cells are compartmentalized by membranes made up of a lipid matrix with embedded proteins in it (Singer and Nicolson 1972). These boundaries ensure the cellular integrity, separate the inside from the outside of the cell, and maintain the internal pH, osmotic conditions and the gradients of different molecules. They are responsible for selective exchanges with the environment, as they allow the diffusion of certain molecules and the active transport of others. They also take part in energy transduction because they carry both, the primary gradient-producing electron transport chains and the protein structures able to exploit the electrochemical gradient across the membranes. Finally, membranes participate in signal transduction from the environment and harbor some key enzymes that catalyze metabolic reactions such as the synthesis of the membrane lipids themselves (Berg et al. 2007).

Although it can be speculated that membranes may have had less functions in ancestral organisms with regard to modern cells, membrane boundaries are essential for life as we know it: by separating the internal medium from the environment and from other organisms, they permit self-maintenance (metabolism) and information storage that allow organisms to reproduce and undergo natural selection and, thus, Darwinian evolution. For these reasons, many authors consider that a given entity can only be considered as a living organism not only when it is self-reproducing but also when it is compartmented (Deamer 1997; Fleischaker 1990; Luisi 1997; Mavelli and Ruiz-Mirazo 2007; Moreno and Ruiz-Mirazo 2009; Morowitz et al. 1988; Oparin 1976; Pereto 2005; Szostak et al. 2001). Moreover, the chemical properties of membrane components have lead to scenarios in which the first steps of the origin of life could have been triggered by lipids and lipid-mediated compartmentalization (Chen and Szostak 2004; Deamer 1986; Deamer 1997; Hanczyc et al. 2003; Segre et al. 2001; Szathmary 2007), hence supporting a continuity between the prebiotic chemistry and the modern cells (Morowitz et al. 1991).

The outstanding chemical properties of the lipid membranes that turn them into entities of major interest in prebiotic chemistry, together with their universal distribution,

function and structure in cells from the three domains of life, strongly suggest that some kind of boundary similar to contemporary membranes may have existed in the last common ancestor of living organisms (the cenancestor). Nevertheless, the unexpected diversity of the chemical components that make up the contemporary cell membranes has raised doubts about the presence of lipid compartments in the cenancestor. Indeed, the main components of lipid membranes, the phospholipids, are made up of *sn*-glycerol-3-phosphate usually linked to fatty acid chains via ester links in bacteria and eukaryotes, whereas their archaeal counterparts use *sn*-glycerol-1-phosphate most often linked to isoprenoid chains via ether links (these differences will be detailed below). Such strong dissimilarities between contemporary membranes in archaea and bacteria/eukaryotes led some authors to propose that these lineages independently developed each type of membrane and that the cenancestor was acellular (Koga et al. 1998) or bound by mineral (iron monosulfide) bubbles (Martin and Russell 2003), whereas others invoke a greater continuity between the cenancestor and the subsequent three domains of life by proposing that the cenancestor bore mixed membranes with properties that can be inferred from modern organisms (Peretó et al. 2004; Wächtershäuser 2003).

The presence or absence of a membrane in the cenancestor, its chemical nature and the early evolutionary process which originated the subsequent domains of life are major issues when trying to infer the characteristics of the cenancestor and to determine the functions that it could have been able to carry out. Addressing these questions requires a thorough revision of current knowledge on lipid diversity and its impact on the different suggested scenarios. In this chapter, we will first describe some features common to all contemporary membranes and, afterwards, the differences that have been put forward as contradictory with the presence of lipid membranes in the cenancestor and the scenarios that have been proposed to take into consideration these observations. We will revisit the apparent dissimilarities that have been observed in biological membranes and highlight a cellular hypothesis that we think to be the most plausible based on current data.

Universal characteristics of biological membranes

Together with membrane proteins, the main components of modern membranes are phospholipids (although other types of lipids are also of major importance). Phospholipids are amphiphilic molecules, i.e. they bear in the same molecule hydrophilic regions interacting with water and hydrophobic regions segregating from the aqueous medium. The hydrophilic

part is called the head, and the hydrophobic parts are the tails (Fig. 1). Phospholipids are built around a glycerol backbone linked to a phosphate group either in its first or its third carbon, what raises two different enantiomers: *sn*-glycerol-1-phosphate (G1P) and *sn*-glycerol-3-phosphate (G3P), respectively. Several different polar groups can be linked to the phosphate residue: serine, *myo*-inositol, ethanolamine, glycerol, etc. The phosphate and the polar group form the hydrophilic head of the phospholipid. The remaining two free carbons of glycerol are usually linked to long carbon chains either by ether or ester bonds, forming the characteristic hydrophobic tails. These carbon chains are very diverse, but they mostly belong to two main families: fatty acids synthesized from the condensation of acetyl-CoA units, and isoprenoid chains synthesized from universal activated forms of isoprene that serve as precursors (isopentenyl pyrophosphate, IPP, and dimethylallyl diphosphate, DMAPP).

The currently accepted fluid mosaic model of membranes was born from the attempt to minimize the thermodynamic requirements of phospholipids and membrane proteins or, more precisely, maximize the hydrophobic interactions in the one hand and the hydrophilic interactions in the other hand (Singer 2004; Singer and Nicolson 1972). Phospholipids form bilayers in which the hydrophilic heads interact among them and with water in both sides of the membrane, while the hydrophobic tails face each other in the central part of it and, thus, maximize the hydrophobic links away from water. Essential integral proteins are inserted into this bilayer: the hydrophobic parts of proteins interact with the hydrophobic central part of the phospholipid matrix whereas the polar regions of the proteins are excluded from the membrane to one or both sides of it.

The term “mosaic” in this model refers to the mixture of phospholipids and proteins within the plane of the membrane, contrary to competitor models (Benson 1966; Davson and Danielli 1952; Robertson 1964). The model is also “fluid” since the membrane proteins and the phospholipids can diffuse in it, allowing random movements but also forming specific aggregates when required. Importantly, phospholipids can freely move within their half-bilayer, although the transfer from one side to the other is thermodynamically unlikely and must be mediated by specific integral proteins called flippases (Daleke and Lyles 2000). In the same way, integral proteins can move within the plane of the membrane but most often cannot change from one polarization to the other because of thermodynamic constraints. This is relevant to another major feature of contemporary biological membranes: they are asymmetric, that is the composition of each half-bilayer (inner/outer) is different from the other. The most common example is the asymmetric distribution of oligosaccharides, which

are specifically found in the outside of the cell linked to lipids and proteins in so-called glycolipids and glycoproteins, and absent from the inner half-bilayer. Other examples are the specific polarities of integral proteins and the presence of peripheral proteins bound only in one side of the membrane.

Although here we have focused our attention on phospholipids and proteins, other components take part in the membrane composition, as for example many isoprenoid derivatives, which bear essential functions both for the cell (pigments) and membrane (fluidity control by sterols) functioning. However, as the truly minimal components of contemporary membranes are phospholipids, most of our discussion will concern their characteristics and evolution.

Two different phospholipid types in the three domains of life

Despite all phospholipids share the aforementioned common structure that confers them their chemical and thermodynamic properties, major differences have been observed between the phospholipids occurring in archaea and those from bacteria and eukaryotes (Fig. 1). The most important one, which is also the only one without any known exception, is the utilization of opposite glycerol phosphate enantiomers: archaea use G1P whereas bacteria and eukaryotes use G3P in their phospholipids. This is correlated with their enzymatic stereospecificities, namely the presence in archaea of a *sn*-glycerol-1-phosphate dehydrogenase (G1PDH) only able to produce G1P, while bacteria and eukaryotes have a *sn*-glycerol-3-phosphate dehydrogenase (G3PDH) synthesizing the opposite enantiomer (Kito and Pizer 1969; Koga et al. 1998; Nishihara and Koga 1997). G1PDH and G3PDH are non-homologous enzymes that belong to totally unrelated protein families (Daiyasu et al. 2002; Koga et al. 1998). This raises the question of which enzyme, if any, was present in the cenancestor.

Moreover, archaea are known to bear isoprenoid chains in their phospholipids and have been thought to completely lack fatty acids until very recently (De Rosa et al. 1986; Kates 1992), whereas bacterial and eukaryotic phospholipids usually incorporate fatty acids instead of isoprenoid chains. These differences are paralleled with dissimilarities in the related biosynthetic pathways. On the one hand, the fatty acid biosynthesis pathway was unknown in archaea; on the other hand, isoprenoid precursors, IPP and DMAPP, although universally distributed, are synthesized by two independent non-homologous pathways that were thought

to be exclusively related to some groups of organisms: the mevalonate pathway (MVA) in archaea and eukaryotes and the methylerythritol phosphate pathway (MEP) in bacteria and plastid-bearing eukaryotes (Lange et al. 2000). All these contradictions among the three domains of life appeared to belie the presence of hydrophobic tails in the hypothetical phospholipids of the cenancestor.

Finally, in the archaeal phospholipid archetype, isoprenoid chains are bound to G1P via ether links, whereas in bacteria and eukaryotes fatty acids are usually bound to G3P through ester links (Fig. 1). However, many exceptions to this splitting are known in modern organisms: ether-linked lipids have been detected in proportions up to 25% in animal membranes (Mangold and Paltauf 1983) as well as in at least six different bacterial phyla, including Firmicutes, Proteobacteria, Planctomycetes, Thermotogae, Aquificae and Thermodesulfobacteria (Damste et al. 2007; Huber et al. 1992; Jung et al. 1994; Langworthy et al. 1983; Rutters et al. 2001; Sinninghe Damste et al. 2002; Weijers et al. 2006).

First uncertainties: was the cenancestor compartmented? The acellular hypothesis

First of all, it must be noted that most of the hypotheses proposed only discuss the descent of the archaeal and the bacterial lineages from the cenancestor, the eukaryotes being assumed to have evolved subsequently from these primordial lineages. In fact, many hypotheses consider that eukaryotes originated from a symbiosis event between an archaeal and a bacterial organism (e.g., Martin and Muller 1998; Moreira and Lopez-Garcia 1998). This is based on the apparent chimerical nature of eukaryotic genomes, which show a mix of genes of putative archaeal origin (usually the informational genes) and genes of putative bacterial origin (most operational or metabolic genes). Independently of the debate about the nature and origin of the first eukaryotes (Gribaldo et al. 2010), discussions about ancestral membranes traditionally either accommodate only chimeric hypotheses or simply ignore the problem of the evolution of the eukaryotic membranes.

Until the late 1990's, the cellular nature of the cenancestor was not a matter of hot debate, contrarily to what it has become since then. The fact that all organisms were limited by lipid membranes made up of phospholipid and proteins and respecting the fluid mosaic model was simply regarded as a consequence of the vertical inheritance of these characteristics from early ancestors. Nevertheless, the different composition in archaeal

phospholipids with respect to their bacterial counterparts was already known (see above). Of these dissimilarities, the only one for which any exception had been detected was the stereochemistry of the glycerol phosphate backbone (G1P in archaeal *versus* G3P in bacteria and eukaryotes).

In this context, the isolation and sequencing of the archaeal G1PDH gene (Koga et al. 1998; Nishihara and Koga 1997), followed by the discovery that G1PDH and G3PDH were non-homologous, led to the idea that the two enantiomer biosynthesis pathways had independent origins. On this basis, Koga *et al.* (1998) proposed that each glycerol phosphate dehydrogenase (and its related phospholipid type) had emerged independently in the ancestors of Archaea and Bacteria, triggering the differentiation of these two domains of life from a phospholipid-lacking cenancestor. This hypothesis entails that the first compartmented organisms were the respective ancestors of Bacteria and Archaea and, thus, that the cenancestor was acellular (Koga et al. 1998). In order to describe the nature of the cenancestor, these authors invoked previous suggestions that primitive metabolism could have emerged and concentrated in a chemical manner on pyrite (FeS₂) surfaces (Wächtershäuser 1988). Accordingly, they proposed that soluble metabolites, some enzymes and even the genetic machineries originated in this type of non-compartmented system. At this chemical stage, a racemic mixture of G1P and G3P could have accumulated, providing phospholipid precursors prior to the independent emergence of stereospecific catalysts and the subsequent cellularization of the bacterial and archaeal lineages. Eukaryotes would lately have evolved from an endosymbiosis event between the two primordial domains of life (Fig. 2A).

The most interesting point of this hypothesis is that it straight addresses the question of the presence and the nature of a membrane in the cenancestor, which became a debated open matter since then. Nevertheless, this scenario raises several major issues: (i) although pyrite surfaces could be interesting concentrating hypotheses for the prebiotic chemistry that could have provided the building blocks involved in the first steps of life, the cenancestor was an organism that certainly had transcription, translation (included the genetic code) and some protein enzymes (Delaye et al. 2005; Gogarten et al. 1989; Koonin 2003; Kyrpides et al. 1999; Ouzounis et al. 2006). Therefore, it is very difficult to conceive that such complexity may have occurred in free solution without the diffusion of the intermediates (de Duve and Miller 1991; Oparin 1976); (ii) reproduction of genetic elements must be linked to the entities that confer them a reproductive advantage to allow Darwinian evolution; therefore, spatial confinement is indispensable (de Duve 2005; Oparin 1976); (iii) amphiphilic molecules have

been obtained under plausible prebiotic conditions (Apel and Deamer 2005; McCollom et al. 1999), and amphiphilic compounds able to spontaneously form membranous vesicles under appropriate conditions have been detected in meteorites (Deamer 1986; Deamer and Pashley 1989), so the prebiotic synthesis of lipid molecules and membranes does not seem to be a restricting step that should absolutely be dependent on biosynthesis by biological catalysts (Deamer et al. 2002).

To summarize, the acellular scenario could be seen at first sight as the simplest response to the puzzling observation that two very different phospholipid types exist in the domains of life, which are correlated to non-homologous biosynthesis pathways. Nevertheless, this no-compartment hypothesis raises more problems than it gives answers, since it requires the emergence of an extraordinary complexity from very simple chemical mechanisms without providing neither a satisfying concentration mechanism nor a plausible driving force. Moreover, this idea is contradictory in that it proposes an ancestral system that could have been able to reach complex autoreplication and complex metabolism but not to produce amphiphilic lipid molecules that, paradoxically, can be easily synthesized in prebiotic conditions.

Did the cenancestor have boundaries different from lipid membranes? The inorganic compartments hypothesis

As we have just seen, the apparent irreconcilable differences between archaeal and bacterial/eukaryotic phospholipids first led to the proposal of an acellular cenancestor. However, non-compartmented systems can hardly account for the probable complexity attained by the cenancestor and are not able to put together the genetic elements (genotype) and metabolic functions (phenotype) essential to apply Darwinian selection and evolution (de Duve 2005; Tawfik and Griffiths 1998). Since compartments were required in the cenancestor but lipid membranes seemed impossible to some authors, the obvious third possibility was that the cenancestor was compartmented in some other kind of structure. As pioneers of this idea, Oparin and coworkers (1976) tried to prepare spherical aggregates (coacervates) from macromolecular components such as Arabic gum, gelatin and histones, but these structures were unsuitable to model primordial cells because their synthesis in prebiotic conditions is very unlikely and their thermodynamic stability is low (Monnard and Deamer 2002; Walde et al. 1994). Alternatively, hydrothermal vents had been proposed as likely sites of prebiotic

synthesis of organic molecules since they provide a concentrating mechanism (physical limits to slow down free diffusion in the ocean), Fe-S and Fe-Ni-S centers that may participate in primordial catalysts, and a continuous redox energy source to allow the synthesis of possible biochemical building blocks (Russell 1997). Moreover, in these sites, inorganic ‘bubbles’ are formed by the spontaneous precipitation of iron monosulfide. This led some authors to speculate that not only some prebiotic chemistry but all the evolutionary process from the first prebiotic chemistry up to the cenancestor had taken place in a submarine hydrothermal vent, and that the respective ancestors of archaea and bacteria cellularized and escaped the original vent only after two independent lipid inventions in a very late period of this process (Fig. 2B, Koonin and Martin 2005; Martin and Russell 2003).

In this scenario, the transitions between the RNA, RNA-protein and DNA worlds, as well as the development of transcription and translation, all took place in these hydrothermal compartments. The cenancestor itself would have been part of this hydrothermal scenario: it would correspond to a network of iron monosulfide compartments containing retrovirus-like RNA chromosomes (the cenancestor is here thought to be incapable of DNA replication, since the replication machineries are not orthologous in archaea and bacteria (Koonin et al 2006)). These genetic elements were the agents of variation (mutation) and natural selection for self-replication. Within the mineral boundaries, they would entail the emergence of intricate autocatalytic networks able to make DNA, RNA, and proteins, and to sustain a core and an intermediate enzymatic metabolism (nitrogen metabolism, amino acid, nucleic acid and cofactor biosyntheses). As the hydrothermal vent grew through the formation of new iron monosulfide cavities, the most successful genetic elements (the fastest replicators) would have diffused into them, allowing the propagation of such ensembles within the spreading mineral compartment network. In this context, two distinct systems for phospholipid biosynthesis might have evolved and coupled with the emerging DNA genomes. Once membranes, cell walls, DNA genomes and bioenergetic pathways came into place, they independently allowed the emergence of the first free-living organisms, namely the respective ancestors of bacteria and archaea. As in the previous scenario, eukaryotes would have emerged from the symbiosis between an archaeon and a bacterium (Martin and Russell 2003).

This hypothesis is of major interest, especially in the early prebiotic era, as a mechanism of production and concentration of biochemical building blocks. Nevertheless, it encounters major problems when it is applied to subsequent evolutionary stages, in particular the one corresponding to the cenancestor. First of all, in contrast with claims formulated by

the authors of this hypothesis (Koonin and Martin 2005), this hypothesis does not allow Darwinian evolution since the replication of the autocatalytic units is not necessarily coupled with the formation of new compartments. In other words, in this hypothesis the most efficient autoreplicative units would be able to control their own replication and would freely diffuse into the new formed ‘bubbles’, but, as they would have no control at all on the formation of those bubbles, there would be no coupling between genetic material replication and organismal multiplication. Therefore, those compartments could neither be considered as autopoietic entities (Fleischaker 1990) nor as individuals undergoing natural selection (de Duve 2005). Another problem related to this scenario concerns the emergence of lipid membranes: the biosynthesis pathways producing the lipid components that would integrate the future archaeal and bacterial membranes are proposed to have evolved and colonized the inorganic components in the same way as the DNA genomes and other elements. However, this hypothesis fails to propose a realistic function for those molecules, and, thus, a selective advantage to explain the success of their biosynthetic pathways. It could be naturally assumed that the presence of membranes allowed the encapsulation of genetic or metabolic units. But, if these elements were already encapsulated in lipid membranes, the iron monosulfide compartments at the center of this hypothesis would become unnecessary.

There are other arguments strongly related to the problem of lipid membranes that weaken Martin and Russell’s hypothesis this model: as we will discuss later, several membrane proteins (e.g., ATPases and the membrane-targeting SRP system) most likely existed in the cenancestor (Gogarten et al. 1989; Gribaldo and Cammarano 1998; Mulkidjanian et al. 2007), what raises the question of their location and function in the cenancestor if it was bounded by inorganic compartments (Jekely 2006; Pereto et al. 2004). To answer to this concern, the authors of this scenario refined it by incorporating the idea that some kind of aliphatic chains could have been produced in the hydrothermal conditions that allowed the formation of hydrophobic patches in which the ATPases and the SRP systems could be embedded (Koonin and Martin 2005). But, again, if lipid vesicles able to bear ATPases and the SRP system existed, why these lipid components did not encapsulate also the genetic entities? And even if the genetic entities were external to the vesicles, what was the driving force to evolve membrane-targeting systems (the SRP complexes) if the genetic elements encoding these proteins were not linked to them? Only two ways to link together a lipid vesicle and some genetic elements can be proposed: either to covalently fix them to each other (and there is no evidence for this) or that the genetic elements be encapsulated by the

lipid layers, as in modern cells. Finally, it has been shown that other metal sulfides different from iron monosulfides would do better in providing the energy for the synthesis of biomolecules, questioning the likely importance of FeS cavities in prebiotic steps (Mulkiđjanian 2009; Mulkiđjanian and Galperin 2009). However, regardless to the nature of the metal involved in the construction of such a mineral boundary, the criticisms about the presence of mineral boundaries in the cenancestor remain.

To conclude, the hypothesis based on hydrothermal mineral compartments may provide an interesting scenario for prebiotic chemistry, but it seems implausible for subsequent evolutionary stages, including the cenancestor. Indeed, since the proposed inorganic boundaries were not linked to the autoreplicative entities, this hypothesis falls into the same problems as the acellular hypothesis: a boundary able to link together all genomic elements is necessary to explain what we know about the cenancestor. Moreover, it is noteworthy that the inorganic boundaries proposal needs to assume some kind of lipid structures prior to the divergence of the prokaryotic lineages. A more parsimonious hypothesis would be that the cenancestor already bore a lipid membrane.

Was the cenancestor bounded by heterochiral membranes? The pre-cell theory

As we have seen, trying to explain the dissimilarities between the archaeal and the bacterial/eukaryotic phospholipids first led to hypotheses assuming the lack of lipid membranes in the cenancestor. However, other authors addressed the possibility that both types of phospholipids actually existed in the cenancestor and that the contemporary differences in archaeal and bacterial cells are secondary and related to the divergence of these two lineages from the last common ancestor.

This hypothesis was first postulated as an extension to Kandler's pre-cell theory. Kandler's hypothesis (1994) depicts the cenancestor not as an organism but as a population of self-reproducing entities endowed with metabolism and genetic information, and unable to limit frequent mutual exchange of genes. In spite of the frequent gene exchange, not all cells would have been identical within the population. These entities are called pre-cells because of their incapacity to keep their genetic information apart from the rest of the community, but from a membrane point of view, these entities limited by lipid membranes can be considered as perfectly cellular. Kandler assumes that the bacterial lineage diverged first from this community, which continued to evolve in its own way. A second lineage of organisms would

have diverged later to become the Archaea, then a third lineage started a symbiotic relationship with a bacterium to originate the Eucarya (Kandler 1994).

Based on this scenario of a ‘cenancestral population’, Wächtershäuser addressed the question of the chemical nature of the boundaries of the pre-cells (Wächtershäuser 2003). First, he assumed that hypothetical hybrid heterochiral membranes (i.e., with a mix of G1P and G3P) may have existed although they would have been less stable than homochiral ones. The pre-cells are proposed to have borne such heterochiral membranes, the components of which could have been synthesized in a racemic way either by inorganic catalysts or by primitive non-stereospecific enzymes (Fig. 2C). In these pre-cells, spontaneous lipid segregation due to physico-chemical stability differences between the two phospholipid types would have occurred. Fusions and fissions among cells are proposed to explain the mutual genetic exchange and also to explain how the heterochiral membranes progressively segregated into two opposite homochiral subpopulations. The emergence of one G3PDH stereospecific enzyme would have definitely fixed one kind of homochiral membrane in the bacterial lineage, whereas the subsequent advent of the G1PDH stereospecific enzyme would have set the opposite homochiral membranes in archaeal ancestors. Shifts from one chirality to the other would have been prevented by the lower stability of the heterochiral intermediates (Wächtershäuser 2003). This hypothesis of lipid segregation based on chirality has been recently extended to the hydrophobic tails: whereas G1P-fatty acids and G3P-fatty acids are enantiomers, phospholipids made up of G1P-isoprenoids and G3P-isoprenoids are diastereomers (non-enantiomeric stereoisomers with at least two chiral centers). Since diastereoisomers have different physicochemical characteristics that enantiomers have not, different combinations of these components are not supposed to segregate in the same way. Consequently, the hydrocarbon chains are proposed to play a supplementary role in the segregation of heterochiral membranes in addition to the chirality of the glycerol phosphate backbones (Koga 2011).

This scenario takes into account both the existence of ancestral compartmented organisms and the characteristics of modern membranes. Nevertheless, it does not provide an explanation for the synthesis of phospholipid components, keeping the door open to either an inorganic biosynthesis or an enzymatic non-stereospecific pathway. The possibility of inorganic biosynthesis, such as the anabolism on colloidal surfaces or microcrystalline particles with metal-sulfide structures proposed by Wächtershäuser, closely brings together the origin of life and the cenancestor time, and addresses the question of the capabilities of the

cenancestor and its independence with respect to its environment. The possibility of the enzymatic non-stereospecific synthesis of phospholipids in the cenancestor has been criticized by arguing that it would make the cenancestor the most versatile organism that never existed (Koonin and Martin 2005). Nevertheless, this argument is weakened if we envisage that the putative enantiomeric lipid mix in the cenancestor was the result of a limited number of non-stereospecific enzymes. When Wächtershäuser hypothesis was first postulated, there was no available information concerning the early evolutionary history of the phospholipid biosynthesis pathways. We will treat this aspect later (see below) to show that actually the majority of the enzymatic machinery responsible for phospholipid biosynthesis in modern cells can most likely be traced back to the cenancestor.

In addition to the biotic or abiotic nature of the source of phospholipids, there is another important concern about this hypothesis. The whole hypothesis depends on the assumption that heterochiral membranes are less stable than their homochiral counterparts. At first sight, this looks plausible since, for example, racemic mixtures of D- and L-myristoylalanine undergo rapid segregation (Nassoy et al. 1995) and mixed bacterial-archaeal liposomes appear to be less stable than uniform ones (Longo et al. 2007). However, the opposite observation (a higher stability of hybrid eukaryotic/archaeal liposomes) has also been reported in other independent studies (Fan et al. 1995). To solve this ambiguity, new experiments have recently been carried out, showing that hybrid liposomes are as stable as homochiral liposomes and the stability is much more correlated to the hydrocarbon chain length than to the backbone chirality or chain nature (Shimada and Yamagishi 2011). These results, which explain differences among previous works, are at odds with the theoretical premise of Wächtershäuser's pre-cell scenario.

A phylogenomic exploration of the enzymatic capacities of the cenancestor to synthesize phospholipid backbones

In the previous sections, we have discussed the differences between the two archetypical phospholipids and the hypotheses that have been proposed to take them into account when inferring the nature of the cenancestor membranes. The last hypothesis discussed above bumped into a lack of information concerning the enzymatic capacities of the cenancestor, preventing the presentation of a plausible way to synthesize its membranes. In the following,

we will explore the evolutionary history of the enzymes involved in this process in contemporary cells.

As already mentioned, the most distinctive feature between the archaeal and the bacterial/eukaryotic phospholipids is the glycerol phosphate stereochemistry (G1P and G3P, respectively). This is the only characteristic without any known exception so far. G3P is a very common metabolite that contemporary cells synthesize from multiple substrates, such as dihydroxyacetone phosphate by a flavin-dependent G3P-dehydrogenase (*glp*), or glycerol by a glycerol kinase (*glpK*). If we solely consider the presence of glycerol phosphate in cells, G3P can actually be detected *in vivo* in heterotrophic archaeal cells incubated with glycerol (Nishihara et al. 1999). Moreover, genes similar to G3P-producing enzymes have been detected not only in bacteria and eukaryotes, but also in archaea (Peretó et al. 2004). These arguments suggest that G3P is a universal metabolite but, as archaea do not use it for their phospholipids, this cannot be accepted as enough evidence to infer that G3P participated in cenacestral phospholipids. As a result, much interest has focused on glycerol phosphate dehydrogenases responsible for G1P/G3P supply specifically related to phospholipid biosynthesis in modern cells. Primary sequence comparisons and tertiary structure analyses show that archaeal G1PDH and bacterial/eukaryotic G3PDH derive from different ancestral enzymes, despite they share the same substrate and coenzyme requirements (Daiyasu et al. 2002; Koga et al. 1998). Nevertheless, the evolutionary history of both enzymes is very complex since each one belongs to different large enzyme superfamilies.

G1PDH is part of a superfamily that also contains glycerol dehydrogenase (GDH), alcohol dehydrogenase (ADH) and 3-dehydroquinate synthase (DHQS). All of them share a great functional and structural resemblance, including the oxidoreduction of NAD^+ (Carpenter 1998, Ruzheinikov 2001). A phylogeny of this superfamily shows that horizontal gene transfers (HGT) do exist but are limited in number and do not mask the phylogenetic origin of each sub-family: GDH is widespread among bacteria, ADH and DHGS are characteristic of bacteria and eukaryotes and G1PDH of archaea. Thus, although neither function seems to be universal, the ancestral existence of members from this superfamily in the three domains of life is supported, so that the presence of at least one primitive NAD^+ -dependent dehydrogenase of this superfamily can be inferred for the cenacestor. Since contemporary enzymes from the three domains of life are not congruent with regard to their substrate specificities, it can be reasonably proposed that this ancestral enzyme had a nonspecific

dehydrogenase activity from which modern characteristics rose, and G1P synthesis cannot be excluded (nor definitely confirmed) in the cenancestor (Peretó et al. 2004).

Very similar arguments can be provided for G3PDH, which is distantly related to two other enzymes, UDP-glucose 6-dehydrogenase (UDPGDH) and 3-hydroxyacyl-CoA dehydrogenase (HACDH) (Kavanagh et al. 2003). All these enzymes also share a NAD⁺-dependent oxidoreduction mechanism. Phylogenetic analyses of this superfamily show three groups of sequences, each of them corresponding to one of the three functions previously cited (Peretó et al. 2004). Some HGTs can be detected, but they do not hide the origin of each sub-family gene. UDPGDH is a universally distributed subfamily involved in synthesis of different polymers in eukaryotes and bacteria (Arrecubieta et al. 1994; Binari et al. 1997; Dalessandro and Northcote 1977). In addition, the presence of D-glucuronic acid in some archaeal cell walls (Kandler 1994) suggests that these enzymes could be involved in its biosynthesis. This implies major metabolic roles of this subfamily in the three domains of life. Similarly, HACDH is widely distributed in the three domains of life and is involved in different polypeptidic enzymes carrying out many diverse functions (Hiltunen and Qin 2000; Peretó et al. 2004; Youngleson et al. 1989). Thus, among members of the G3PDH/UDPGDH/HACDH superfamily, only G3PDH is limited to bacteria and eukaryotes (the rare homologous genes found in archaeal genomes have probably been acquired by HGT). As the two other subfamilies are universally distributed, the whole superfamily is likely to have had at least two representative genes in the cenancestor (maybe more), despite their function cannot precisely be established. Consequently, it cannot be excluded that at least one of these ancestral promiscuous enzymes was capable of synthesis of G3P among other functions, before functional divergence happened. In that case, G3PDH function could have been lost in the archaeal lineage just because catabolic G3P formation could be catalyzed by other enzymes (glycerol kinase and flavin-dependent G3PDH, as previously seen) and specific-G3P synthesis for phospholipid biosynthesis became unnecessary in this lineage (Peretó et al. 2004).

In summary, although G1PDH and G3PDH belong to two separate superfamilies, ancestral representatives from both were most likely present in the cenancestor, where they could have provided G1P and G3P non-specifically. Thus, the cenancestor would have been endowed with hybrid, heterochiral membranes (Fig. 2D).

Could the cenancestor synthesize hydrocarbon lateral chains?

Apart from the backbone chirality, another major difference between archaeal and bacterial/eukaryotic phospholipids is the utilization of isoprenoids as hydrophobic chains in the former whereas fatty acids are used in the latter. This is stressed by the observation that, although isoprenoids exist in all organisms and their precursors (IPP and DMAPP) are the same in the three domains of life, two different non-homologous isoprenoid biosynthesis pathways are unevenly distributed in the three domains of life: the mevalonate pathway (MVA) is typical of archaea and eukaryotes, whereas the methylerythritol phosphate pathway (MEP) is characteristic of bacteria and plastid-bearing eukaryotes. This disparity is surprising if we take into account that isoprenoids are present in all organisms as essential components of phospholipids, electron transport chains, pigments or hormones (Lange et al. 2000; McGarvey and Croteau 1995).

Recent work has tried to reconcile the ubiquitous utilization of isoprenoids with the apparent lack of an ancestral way to synthesize their precursors. First, the MEP pathway has been confirmed to be restricted to bacteria and plastid-bearing eukaryotes (Matsuzaki et al. 2008). This pathway is widely distributed in bacterial genomes, supporting that it was present in the last common ancestor of bacteria. Plastids are known to have evolved from endosymbiotic cyanobacteria and transmitted major metabolic routes to their eukaryotic hosts (Martin et al. 1998). Although some phylogenies of the MEP pathway genes do not support the expected cyanobacterial origin, all of them support a relatively recent bacterial origin of these eukaryotic enzymes (Brinkman et al. 2002; Matsuzaki et al. 2008; Moustafa et al. 2008). This limited taxonomic distribution makes the presence of the MEP pathway in the cenancestor unlikely.

Therefore, the search of plausible routes to produce isoprenoids in the cenancestor has focused in the alternative MVA pathway. Although it was supposed to be limited to eukaryotes and archaea, some bacterial representatives have been characterized and shown to participate in isoprenoid precursor synthesis among other functions (Voynova et al. 2004; Wilding et al. 2000). These bacterial genes were first thought to be the result of HGTs from the two other domains (Boucher and Doolittle 2000; Wilding et al. 2000). Surprisingly, when similarity searches were carried out in bacterial genomes, a much more wide diversity of bacteria bearing enzymes of the MVA pathway was identified (representatives are found in 6 different phyla). Phylogenetic analyses of each enzyme of the pathway showed that these

genes did not branch among the sequences from the two other domains of life as it would be expected for HGTs, but systematically clustered together in clades separated from eukaryotic and archaeal sequences. Moreover, although some cases of HGT among bacterial species were identified, bacterial sequences in these phylogenies most often respected the monophyly of main bacterial taxa, supporting vertical inheritance as the major mechanism to explain the presence of the pathway in these bacteria. Finally, biochemical properties only shared among bacteria and absent from the two other domains of life have been pointed out (Bochar et al. 1999; Friesen and Rodwell 2004; Hedl et al. 2004; Scher and Rodwell 1989), supporting that these characteristic are enzymatic synapomorphies that evolved in the bacterial lineage after the divergence of the three domains of life (Lombard and Moreira 2011b).

This result was of major relevance in resolving the apparent opposition between the universal distribution of isoprenoid precursors and the so far uneven distribution of their biosynthetic pathways: although the MEP pathway was really limited to bacteria and plastid-bearing eukaryotes, supporting its bacterial origin, the mevalonate pathway was shown to be ancestral to each domain of life, suggesting a common inheritance from the cenancestor and subsequent modifications to explain specificities in each domain of life. Many bacteria would have lost the mevalonate pathway in favor of the functionally redundant MEP pathway, as well as archaea are known to have changed the last steps of the MVA pathway with regard to the typical eukaryotic/bacterial counterpart (Lombard and Moreira 2011b).

The assumption that fatty acids are missing in archaea (De Rosa et al. 1986; Kates 1992), where isoprenoids are the main hydrophobic chains in phospholipids, represented a similar difficulty than the uneven distribution of the isoprenoid precursor biosynthesis pathways. Though, significant amounts of the total phospholipid side chains (from 11.3 to 89%) were recently determined to contain fatty acids in five different euryarchaeotal classes (Gattinger et al. 2002). Furthermore, genes involved in fatty acid biosynthesis were described as the largest functional gene family in the crenarchaeote *Sulfolobus solfataricus* (She et al. 2001). These arguments suggest that not only fatty acids can be found in archaea, but also that they are widespread in this domain of life and can occasionally be major components of archaeal membranes. In agreement with this, sequence similarity searches carried out to identify archaeal homologues of fatty acid biosynthesis and degradation genes found that all genes, except those of the acyl-carrier protein machinery (ACP), have more or less distantly related homologues in a wide diversity of archaeal genomes. Genes involved in the fatty acid biosynthesis pathway have very complex evolutionary histories including many duplication

and fusion events with other enzymatic units, what makes the early evolution of these families very difficult to study. Nevertheless, recent phylogenomic analyses showed that the presence of the first step of the fatty acid biosynthesis (acetyl-CoA carboxylation) in the last common ancestor of archaea can not be excluded (Lombard and Moreira 2011a).

In summary, the observation that significant amounts of fatty acids exist in membranes from a wide diversity of archaea, the finding of fatty acid biosynthesis genes in a wide diversity of archaeal genomes and the possibility that some of these genes were present in the last common archaeal ancestor imply that ancient assumptions of the lack of fatty acids in archaea must be revisited. Indeed, these arguments may imply that fatty acid synthesis was inherited in archaeal organisms from their last common ancestor, and because of the similarity between the bacterial and the archaeal counterparts, it would seem reasonable to propose that this machinery was present and able to produce hydrocarbon chains in the cenancestor.

Ether bonds could have existed in the cenancestor

Once the glycerol phosphate backbones and the hydrophobic tails are synthesized, these elements are specifically linked together, typically through ether links in archaea and ester links in bacteria. As for the specific utilization of G1P and G3P discussed above, this difference has also been considered as an important distinctive feature and as argument to postulate the absence of lipid membranes in the cenancestor. However, as previously reported, many exceptions are known in a variety of modern bacteria and eukaryotes (Damste et al. 2007; Huber et al. 1992; Jung et al. 1994; Langworthy et al. 1983; Mangold and Paltauf 1983; Rutters et al. 2001; Sinninghe Damste et al. 2002; Weijers et al. 2006). From an enzymatic point of view, (*S*)-3-*O*-geranylgeranyl glyceryl phosphate synthase (GGGPS), the enzyme that transfers the first isoprenoid chain to G1P in archaea (Fig. 3), has been argued to be restricted to the archaeal domain, despite the existence of some bacterial homologues interpreted as HGT acquisitions (Boucher et al. 2004). The enzyme responsible of the second ether-related isoprenoid transfer to the G1P backbone, called (*S*)-2,3-Di-*O*-geranylgeranyl glyceryl phosphate synthase (DGGGPS), has also been thought to be limited to archaea (Hemmi et al. 2004). Since these enzymes catalyze the crossroad steps at which the three major characteristics of archaeal polar lipids are assembled in one molecule (G1P and isoprenoid chains bound by ether links), they have independently been proposed to have triggered the emergence of archaea (Hemmi et al. 2004; Payandeh et al. 2006). Nevertheless, both GGGPS

and DGGGPS belong to respective large families of prenyltransferases widespread in the three domains of life: GGGPS is part of a group of enzymes known to transfer isoprenoid groups to a wide diversity of non-isoprenoid acceptors (Soderberg et al. 2001), whereas DGGGPS is related to the UbiA prenyltransferase family involved in biosynthesis of diverse molecules like respiratory quinones, hemes and pigments in the three domains of life (Hemmi et al. 2004).

Although accurate phylogenies of these enzyme families would provide stronger arguments to confirm or deny the utilization of ether bonds in the cenancestor, (i) the wide dispersal of ether-linked phospholipids in organisms from the three domains of life; (ii) the apparent ubiquity of phospholipid ether-forming archaeal enzyme families; (iii) and the very large substrate range of these enzymes, provide preliminary support to the idea that ether-linked phospholipids may have a very ancient origin in the three domains of life. If information about the early evolution of ether link synthesis remains scarce, to our knowledge the analogous bacterial/eukaryotic ester-forming phosphatidic synthesis has not been studied with respect to its evolutionary origin yet.

Phospholipid head groups are conserved across the three domains of life

As we have shown, most controversies about the presence of phospholipids in the cenancestor have focused on dissimilarities between archaeal and bacterial/eukaryotic glycerol phosphate enantiomers, hydrophobic chains and bonds among them. Polar head group synthesis has been less considered because of their apparent scattered taxonomic distribution. However, first surveys suggest that at least part of polar head group biosynthesis is shared among the three domains of life and could provide interesting insights on phospholipid early evolution (Koga and Morii 2007).

Once glycerol phosphate moieties and hydrophobic chains have been linked together into archaetidic or phosphatidic acids, these intermediates are respectively activated into subsequent CDP-archaeol or CDP-diacylglycerol (Fig. 3). Analogous reactions have been described both in bacteria (Dowhan 1997) and archaea (Morii et al. 2000). The primary sequence of the archaeal enzyme has not been identified so far, but the inability to detect a homologue of the bacterial one in the genome of the archaeon *Methanothermobacter thermoautotrophicus* (Morii et al. 2000) could be interpreted as evidence of the non-

homology of the respective enzymatic mechanisms. Nonetheless, information about this step is yet too scarce to confidently infer its presence or absence in the cenancestor.

In contrast, some promising work has been carried out on the replacement of CDP by the polar head residues. Although polar groups may be very diverse in the three domains of life (Koga and Morii 2005; Morii et al. 2010), L-serine, *myo*-inositol, glycerol and ethanolamine have been demonstrated to occur in bacteria, eukaryotes and archaea (Koga and Morii 2007). The CDP-alcohol phosphatidyltransferase family is a group of homologous enzymes responsible of the polar head group link during phospholipid synthesis (Fig. 4). Enzymes from this family are well characterized in bacteria (Cronan 2003), and two archaeal representatives (archaeatidylserine synthase and archaetidyl-*myo*-inositol synthase) have already been isolated and sequenced, showing homology between the archaeal and the bacterial counterparts (Morii et al. 2009; Morii and Koga 2003). Furthermore, similarity searches carried out on archaeal genomes allowed the detection of hypothetical proteins similar to the bacterial CDP-alcohol phosphatidyltransferase family. These homologues were widespread in archaea and they could be classified into two groups related to their bacterial homologues (Daiyasu et al. 2005). Phylogenetic reconstructions of the CDP-alcohol phosphatidyltransferase family supported their ancestral presence in both bacteria and archaea, which favors the hypothesis of vertical inheritance of the polar lipid biosynthesis machinery from the cenancestor (Daiyasu et al. 2005; Koga and Morii 2007).

Beyond phospholipids: ancestral protein membrane components

The uncertainty of the presence of a lipid membrane in the cenancestor transcends the simple question of the existence of a boundary in this organism. Indeed, as contemporary membranes are major sites of cellular and metabolic essential processes, this issue also addresses our view of other functions in the cenancestor: Was it endowed with an energy-transducing system? Did it bear a metabolite transport complex able to provide necessary molecules and release waste products? Was it equipped with a protein membrane-targeting mechanism in charge of protein insertion within the phospholipid bilayer and peptide secretion?

The first membrane protein complexes that were studied with regard to their presence in the cenancestor were ATPases. Representatives from this family can either use a proton or a sodium gradient to synthesize ATP, or hydrolyze ATP to carry out solute transport

(Nakanishi-Matsui and Futai 2006). They are ubiquitous complex machineries made up, among others, of two subunit types that are mutual paralogues. As these paralogues were universally distributed, they were used to reconstruct a rooted tree of life (Gogarten et al. 1989). Consequently, these complexes are among the minimal components that are thought to have been present in the cenancestor. Now, these enzymes need hydrophobic bilayers to function and they are always related to ion gradients, so their presence in the cenancestor suggests that this organism had a lipid bilayer more or less impermeable to ions, although the nature of the ions (Na^+ or H^+) has been recently called into question (Mulkidjanian et al. 2009; Mulkidjanian et al. 2008).

Other noteworthy membrane systems related to energy metabolism are the respiratory chains. In modern cells, respiratory chains transduce energy by the transfer of electrons from a donor to a final acceptor through membrane-associated redox chains, and couple it to proton translocation from one side of the membrane to the other. Many different respiratory chains are currently known to take place in very diverse organisms. Four of them have been analyzed and proposed to have been present in the cenancestor based on their widespread distribution in archaea and bacteria (Castresana and Moreira 1999). Aerobic respiration has been the most debated one, since the presence of an O_2 -consuming pathway in the cenancestor seemed contradictory with the assumption that O_2 was absent in the early atmosphere and became highly concentrated much later thanks to cyanobacterial oxygenic photosynthesis (Klein et al. 1992). Nevertheless, localized traces of oxygen may have existed in the early atmosphere due to water photolysis (Kasting 1993) and recent detailed phylogenetic analyses support that one of the known dioxygen reductases (A- O_2 Red) predated the divergence of the three domains of life (Brochier-Armanet et al. 2009). In the same way, nitrate, sulfate and sulfur respiration were also proposed to have been present in the cenancestor based on the occurrence of genes coding the proteins involved in electron transfer to inorganic compounds (the terminal acceptors) in a wide bacterial and archaeal diversity (Castresana and Moreira 1999), although in some cases controversies have raised about the possibility of frequent HGTs instead of the inheritance from the cenancestor (e.g. see the debate about sulfate respiration in Fukuba et al. 2003; Klein et al. 2001; Larsen et al. 1999; Wagner et al. 1998). Apart from these controversies, it is noteworthy that the intermediate ubiquitous electron carriers shared by respiratory chains (as for example cytochrome b, the Rieske protein, the blue copper protein, 2F-2S ferredoxin, 4Fe-4S ferredoxin, rubredoxin and flavodoxin), have also been proposed to be ancestral to all living organisms based on the detection of their homologues in a wide

diversity of genomes (Castresana and Moreira 1999), and supplementary data and phylogenies have provided additional support to several of these assumptions (Baymann et al. 2003; Lebrun et al. 2006; Vignais et al. 2001). Therefore, according with the accumulation of new data, the proteins belonging to the respiratory chains may have been present in the cenancestor. Since all known respiratory chains are embedded within lipid membranes, their presence in the cenancestor also supports the existence of a lipid boundary in this organism.

Besides energy metabolism, the protein targeting machinery and the protein membrane translocation systems have also attracted much interest because of their widespread distribution in modern cells and the importance of their functions. The signal recognition particle (SRP) and its SRP receptor (SR) are respectively involved in the recognition of newly synthesized peptides under elongation and the binding to their elongating ribosomes. Thus, this mechanism participates in both unfolded protein insertion into membranes and protein secretion to the outside of the cell. The major components of this system, SRP and SR, are mutual paralogues which are among the few known genes to be shared by all modern cells. As a result, they have been used, similarly to the ATPases described above, to reconstruct universal phylogenies in an attempt to root the tree of life. Their phylogenies confirmed their vertical inheritance from a common ancestor to all domains of life, what definitely supports their presence in the cenancestor (Gribaldo and Cammarano 1998).

In addition to the SRP/SR system, three protein translocation mechanisms can be predicted to have existed in the cenancestor: (i) the Sec pathway is the main machinery of protein export and membrane insertion in many bacteria, and although the archaeal counterparts share little similarity, some of them can complement their bacterial homologues (Auer et al. 1991). Moreover, phylogenies and functional comparisons support the ancestral presence of several Sec components in the cenancestor (Cao and Saier 2003); (ii) the YidC pathway is principally related to protein insertion in bacterial membranes, but it has also been detected in a wide diversity of euryarchaeotal genomes. Although functional studies on archaeal genes have not been reported yet, the phylogeny of this family of translocases was congruent with their presence in the cenancestor (Yen et al. 2001); (iii) the Tat pathway is hitherto the only translocation system known to be able to transport folded proteins across the membranes. The complex is made up of at least two subunits, the pore-forming unit TatA and the peptide-signal recognition TatC, which are present in bacteria, chloroplasts and archaea (Yuan et al. 2010). Although archaeal components of this system have only been described in some halophilic archaea (Rose et al. 2002), genomic surveys show that these elements are

present in a wide diversity of bacteria and archaea and phylogenies support an ancient origin in both lineages, what suggests that the cenancestor likely bore the components of this translocating system (Yen et al. 2002). In line with these observations, two families of signal peptidases responsible for the removal of signal sequences from the precursor proteins and likely also signal peptide peptidases that cleave signal peptides after their partition from the pre-proteins have been observed to be shared by organisms of the three domains of life (Eichler 2002; Ng et al. 2007). Thus, despite the lack of phylogenies for these enzymes, it is tempting to postulate that the structures responsible for the last steps of protein translocation were also vertically inherited in the three domains of life. Together, these results strongly support the presence of several protein membrane-targeting and translocation systems in the cenancestor, implying the existence of a membrane in this organism.

Some other protein components related to membranes have been proposed to be present in the cenancestor, as the ABC transporter family and the ion-coupled permeases (Delaye et al. 2005). However, these propositions are based on genome comparisons of only one organism from each domain of life, and consequently, further phylogenomic analyses are essential to test them. Meanwhile, a conservative proposition of the membrane proteins present in the cenancestor would include ATPases, respiratory chains and several protein membrane targeting and translocation mechanisms.

Conclusion

We have reported here a historical analysis of the evolution of the ideas concerning the presence of lipid membranes (or their absence) in the cenancestor, as well as depicted the *état de l'art* with respect to modern phylogenomic knowledge on this subject. In spite of the major efforts that have been done to answer to this capital question about the cenancestor, we have seen that many uncertainties remain.

On the one hand, the main objection to the arguments suggesting the presence of a lipid membrane in the cenancestor so far is the fact that some of them are indicative instead of demonstrative: (i) the cenancestor bore genes that may have catalyzed the synthesis of glycerol phosphate enantiomers and ether links, but this does not exclude the possibility that they carried out other functions in early evolutionary times; (ii) similarly, it is not possible to exclude that fatty acids were present in the cenancestor but current evidence does not totally

prove that they were there; (iii) some metabolite transport system could have existed in the cenancestor (ABC transporters and permeases), as well as several respiratory chains, but further phylogenomic analyses must be carried out in order to ensure this possibility. On the other hand, some indirect elements look very likely: (i) ATPases, membrane-targeting and membrane translocation mechanisms existed in the cenancestor, as well as some membrane proteins linked to respiratory chains; (ii) the isoprenoid precursors could enzymatically be produced in this organism; (iii) at least some polar head groups binding to hypothetical phospholipids can be strongly assumed to have been present at that time.

In this context, it is difficult to definitely assure the exact type of lipid membrane that could have existed in the cenancestor, what explains why this question remains a hot topic in the debates about early evolution and the nature of the cenancestor. Probably, many insights could be provided through new approaches, as for example the so-called ‘enzyme resurrection’, which consists in the bioinformatic inference of ancestral enzyme sequences in view of their synthesis to allow biochemical characterization (Gaucher et al. 2008; Gaucher et al. 2003; Perez-Jimenez et al. 2011).

Anyway, pending new evidence about this subject, we have shown here that scenarios invoking an acellular ancestor or inorganic boundaries in the cenancestor fail to provide a reliable explanation for the whole set of characters inferred to have existed in this organism. Lipid-free hypotheses were proposed because modern membrane components were thought to be absent in the cenancestor, but as long as more recent results cannot discard their occurrence in it, the most parsimonious hypothesis remains that the last common ancestor of all living cells was already a cell, endowed with some kind of lipid boundary, and bearing some energy-producing and translocation functions. We cannot yet provide a detailed representation of the precise chemical nature of this membrane, but the likely presence of a membrane supports that the cenancestor was already a complex organism very distant from the early steps of the origin of life. It probably had an energy transduction system, some essential metabolic pathways, the machinery for the syntheses of proteins, lipids and nucleic acids and, thus, very likely a genome of considerable size. Nevertheless, these metabolic abilities were probably less specific in the cenancestor than they are now in contemporary cells, so the noteworthy divergences among modern lineages may have been the result of optimization through time from a quite large common unspecific background, as evolution has been doing since that time.

References

- Apel CL, Deamer DW (2005) The formation of glycerol monodecanoate by a dehydration condensation reaction: increasing the chemical complexity of amphiphiles on the early Earth. *Orig Life Evol Biosph* 35:323-32
- Arrecubieta C, Lopez R, Garcia E (1994) Molecular characterization of cap3A, a gene from the operon required for the synthesis of the capsule of *Streptococcus pneumoniae* type 3: sequencing of mutations responsible for the unencapsulated phenotype and localization of the capsular cluster on the pneumococcal chromosome. *J Bacteriol* 176:6375-83
- Auer J, Spicker G, Bock A (1991) Presence of a gene in the archaeobacterium *Methanococcus vannielii* homologous to secY of eubacteria. *Biochimie* 73:683-8
- Baymann F, Lebrun E, Brugna M, Schoepp-Cothenet B, Giudici-Ortoni MT, Nitschke W (2003) The redox protein construction kit: pre-last universal common ancestor evolution of energy-conserving enzymes. *Philos Trans R Soc Lond B Biol Sci* 358:267-74
- Benson AA (1966) On the orientation of lipids in chloroplast and cell membranes. *J Am Oil Chem Soc* 43:265-70
- Berg JM, Tymoczko JL, Stryer L (2007) *Biochemistry*. W. H. Freeman and Co., New York
- Binari RC, Staveley BE, Johnson WA, Godavarti R, Sasisekharan R, Manoukian AS (1997) Genetic evidence that heparin-like glycosaminoglycans are involved in wingless signaling. *Development* 124:2623-32
- Bochar DA, Stauffacher CV, Rodwell VW (1999) Sequence comparisons reveal two classes of 3-hydroxy-3-methylglutaryl coenzyme A reductase. *Mol Genet Metab* 66:122-7
- Boucher Y, Doolittle WF (2000) The role of lateral gene transfer in the evolution of isoprenoid biosynthesis pathways. *Mol Microbiol* 37:703-16
- Boucher Y, Kamekura M, Doolittle WF (2004) Origins and evolution of isoprenoid lipid biosynthesis in archaea *Mol Microbiol*, p 515-27
- Brinkman FS, Blanchard JL, Cherkasov A, Av-Gay Y, Brunham RC, Fernandez RC, Finlay BB, Otto SP, Ouellette BF, Keeling PJ, Rose AM, Hancock RE, Jones SJ, Greberg H (2002) Evidence that plant-like genes in *Chlamydia* species reflect an ancestral relationship between Chlamydiaceae, cyanobacteria, and the chloroplast. *Genome Res* 12:1159-67
- Brochier-Armanet C, Talla E, Gribaldo S (2009) The multiple evolutionary histories of dioxygen reductases: Implications for the origin and evolution of aerobic respiration. *Mol Biol Evol* 26:285-97
- Cao TB, Saier MH, Jr. (2003) The general protein secretory pathway: phylogenetic analyses leading to evolutionary conclusions. *Biochim Biophys Acta* 1609:115-25
- Castresana J, Moreira D (1999) Respiratory chains in the last common ancestor of living organisms. *J Mol Evol* 49:453-60
- Chen IA, Szostak JW (2004) A kinetic study of the growth of fatty acid vesicles. *Biophys J* 87:988-98
- Cronan JE (2003) Bacterial membrane lipids: where do we stand? *Annu Rev Microbiol* 57:203-24
- Daiyasu H, Hiroike T, Koga Y, Toh H (2002) Analysis of membrane stereochemistry with homology modeling of sn-glycerol-1-phosphate dehydrogenase. *Protein Eng* 15:987-95
- Daiyasu H, Kuma K, Yokoi T, Morii H, Koga Y, Toh H (2005) A study of archaeal enzymes involved in polar lipid synthesis linking amino acid sequence information, genomic contexts and lipid composition. *Archaea* 1:399-410
- Daleke DL, Lyles JV (2000) Identification and purification of aminophospholipid flippases. *Biochim Biophys Acta* 1486:108-27
- Dalessandro G, Northcote DH (1977) Changes in enzymic activities of nucleoside diphosphate sugar interconversions during differentiation of cambium to xylem in pine and fir. *Biochem J* 162:281-8
- Damste JS, Rijpstra WI, Hopmans EC, Schouten S, Balk M, Stams AJ (2007) Structural characterization of diabolic acid-based tetraester, tetraether and mixed ether/ester, membrane-spanning lipids of bacteria from the order Thermotogales. *Arch Microbiol* 188:629-41

- Davson H, Danielli JF (1952) The permeability of Natural Membranes. Cambridge Univ. Press, London/New York
- de Duve C (2005) The onset of selection. *Nature* 433:581-2
- de Duve C, Miller SL (1991) Two-dimensional life? *Proc Natl Acad Sci U S A* 88:10014-7
- De Rosa M, Gambacorta A, Gliozzi A (1986) Structure, biosynthesis, and physicochemical properties of archaeobacterial lipids. *Microbiol Rev* 50:70-80
- Deamer D, Dworkin JP, Sandford SA, Bernstein MP, Allamandola LJ (2002) The first cell membranes. *Astrobiology* 2:371-81
- Deamer DW (1986) Role of amphiphilic compounds in the evolution of membrane structure on the early earth. *Orig Life Evol Biosph* 17:3-25
- Deamer DW (1997) The first living systems: a bioenergetic perspective. *Microbiol Mol Biol Rev* 61:239-61
- Deamer DW, Pashley RM (1989) Amphiphilic components of the Murchison carbonaceous chondrite: surface properties and membrane formation. *Orig Life Evol Biosph* 19:21-38
- Delaye L, Becerra A, Lazcano A (2005) The last common ancestor: what's in a name? *Orig Life Evol Biosph* 35:537-54
- Dowhan W (1997) CDP-diacylglycerol synthase of microorganisms. *Biochim Biophys Acta* 1348:157-65
- Eichler J (2002) Archaeal signal peptidases from the genus *Thermoplasma*: structural and mechanistic hybrids of the bacterial and eukaryal enzymes. *J Mol Evol* 54:411-5
- Fan Q, Relini A, Cassinadri D, Gambacorta A, Gliozzi A (1995) Stability against temperature and external agents of vesicles composed of archaeal bolaform lipids and egg PC. *Biochim Biophys Acta* 1240:83-8
- Fleischaker GR (1990) Origins of life: an operational definition. *Orig Life Evol Biosph* 20:127-137
- Friesen JA, Rodwell VW (2004) The 3-hydroxy-3-methylglutaryl coenzyme-A (HMG-CoA) reductases. *Genome Biol* 5:248
- Fukuba T, Ogawa M, Fujii T, Naganuma T (2003) Phylogenetic diversity of dissimilatory sulfite reductase genes from deep-sea cold seep sediment. *Mar Biotechnol (NY)* 5:458-68
- Gattinger A, Schloter M, Munch JC (2002) Phospholipid etherlipid and phospholipid fatty acid fingerprints in selected euryarchaeotal monocultures for taxonomic profiling. *FEMS Microbiol Lett* 213:133-9
- Gaucher EA, Govindarajan S, Ganesh OK (2008) Palaeotemperature trend for Precambrian life inferred from resurrected proteins. *Nature* 451:704-7
- Gaucher EA, Thomson JM, Burgan MF, Benner SA (2003) Inferring the palaeoenvironment of ancient bacteria on the basis of resurrected proteins. *Nature* 425:285-8
- Gogarten JP, Kibak H, Dittrich P, Taiz L, Bowman EJ, Bowman BJ, Manolson MF, Poole RJ, Date T, Oshima T, et al. (1989) Evolution of the vacuolar H⁺-ATPase: implications for the origin of eukaryotes. *Proc Natl Acad Sci U S A* 86:6661-5
- Gribaldo S, Cammarano P (1998) The root of the universal tree of life inferred from anciently duplicated genes encoding components of the protein-targeting machinery. *J Mol Evol* 47:508-16
- Gribaldo S, Poole AM, Daubin V, Forterre P, Brochier-Armanet C (2010) The origin of eukaryotes and their relationship with the Archaea: are we at a phylogenomic impasse? *Nat Rev Microbiol* 8:743-52
- Hanczyc MM, Fujikawa SM, Szostak JW (2003) Experimental models of primitive cellular compartments: encapsulation, growth, and division. *Science* 302:618-22
- Hedl M, Taberero L, Stauffacher CV, Rodwell VW (2004) Class II 3-hydroxy-3-methylglutaryl coenzyme A reductases. *J Bacteriol* 186:1927-32
- Hemmi H, Shibuya K, Takahashi Y, Nakayama T, Nishino T (2004) (S)-2,3-Di-O-geranylgeranylgeranyl glyceryl phosphate synthase from the thermoacidophilic archaeon *Sulfolobus solfataricus*. Molecular cloning and characterization of a membrane-intrinsic prenyltransferase involved in the biosynthesis of archaeal ether-linked membrane lipids. *J Biol Chem* 279:50197-203
- Hiltunen JK, Qin Y (2000) Beta-oxidation - strategies for the metabolism of a wide variety of acyl-CoA esters. *Biochim Biophys Acta* 1484:117-28

- Huber R, Wilharm T, Huber D, Trincone A, Burggraf S, Konig H, Rachel R, Rockinger I, Fricke H, Stetter KO (1992) *Aquifex pyrophilus* gen. nov. sp. nov., represents a novel group of marine hyperthermophilic hydrogen-oxidizing bacteria. *Syst Appl Microbiol* 15:340-351
- Jekely G (2006) Did the last common ancestor have a biological membrane? *Biol Direct* 1:35
- Jung S, Zeikus JG, Hollingsworth RI (1994) A new family of very long chain alpha,omega-dicarboxylic acids is a major structural fatty acyl component of the membrane lipids of *Thermoanaerobacter ethanolicus* 39E. *J Lipid Res* 35:1057-65
- Kandler O (1994) Cell wall biochemistry in archaea and its phylogenetic implications. *Journal of biological physics* 20:165-169
- Kasting JF (1993) Earth's early atmosphere. *Science* 259:920-6
- Kates M (1992) Archaeobacterial lipids: structure, biosynthesis and function. *Biochem Soc Symp* 58:51-72
- Kavanagh KL, Klimacek M, Nidetzky B, Wilson DK (2003) Crystal structure of *Pseudomonas fluorescens* mannitol 2-dehydrogenase: evidence for a very divergent long-chain dehydrogenase family. *Chem Biol Interact* 143-144:551-8
- Kito M, Pizer LI (1969) Purification and regulatory properties of the biosynthetic L-glycerol 3-phosphate dehydrogenase from *Escherichia coli*. *J Biol Chem* 244:3316-23
- Klein C, Beukes NJ, Holland HD, Kasting JF, Kump LR, Lowe DR (1992) Proterozoic atmosphere and ocean. In: Schopf JW, Kelin C (eds) *The proterozoic biosphere*. Cambridge University Press, New York, p 135-174
- Klein M, Friedrich M, Roger AJ, Hugenholtz P, Fishbain S, Abicht H, Blackall LL, Stahl DA, Wagner M (2001) Multiple lateral transfers of dissimilatory sulfite reductase genes between major lineages of sulfate-reducing prokaryotes. *J Bacteriol* 183:6028-35
- Koga Y (2011) Early evolution of membrane lipids: how did the lipid divide occur? *J Mol Evol* 72:274-82
- Koga Y, Kyuragi T, Nishihara M, Sone N (1998) Did archaeal and bacterial cells arise independently from noncellular precursors? A hypothesis stating that the advent of membrane phospholipid with enantiomeric glycerophosphate backbones caused the separation of the two lines of descent. *J Mol Evol* 47:631
- Koga Y, Morii H (2005) Recent advances in structural research on ether lipids from archaea including comparative and physiological aspects. *Biosci Biotechnol Biochem* 69:2019-34
- Koga Y, Morii H (2007) Biosynthesis of ether-type polar lipids in archaea and evolutionary considerations. *Microbiol Mol Biol Rev* 71:97-120
- Koonin EV (2003) Comparative genomics, minimal gene-sets and the last universal common ancestor. *Nat Rev Microbiol* 1:127-36
- Koonin EV, Martin W (2005) On the origin of genomes and cells within inorganic compartments. *Trends Genet* 21:647-54
- Koonin EV, Senkevich TG, Dolja VV (2006) The ancient Virus World and evolution of cells. *Biol Direct* 1:29
- Kyrpides N, Overbeek R, Ouzounis C (1999) Universal protein families and the functional content of the last universal common ancestor. *J Mol Evol* 49:413-23
- Lange BM, Rujan T, Martin W, Croteau R (2000) Isoprenoid biosynthesis: the evolution of two ancient and distinct pathways across genomes. *Proc Natl Acad Sci U S A* 97:13172-7
- Langworthy TA, Holzer G, Zeikus JG, Tornabene TG (1983) Iso- and anteiso-branched glycerol diethers of the thermophilic anaerobe *Thermodesulfotobacterium commune*. *Syst Appl Microbiol* 4:1-17
- Larsen O, Lien T, Birkeland NK (1999) Dissimilatory sulfite reductase from *Archaeoglobus profundus* and *Desulfotomaculum thermocisternum*: phylogenetic and structural implications from gene sequences. *Extremophiles* 3:63-70
- Lebrun E, Santini JM, Brugna M, Ducluzeau AL, Ouchane S, Schoepp-Cothenet B, Baymann F, Nitschke W (2006) The Rieske protein: a case study on the pitfalls of multiple sequence alignments and phylogenetic reconstruction. *Mol Biol Evol* 23:1180-91
- Lombard J, Moreira D (2011a) Early evolution of the biotin-dependent carboxylase family. *BMC Evol Biol* 11:232

- Lombard J, Moreira D (2011b) Origins and early evolution of the mevalonate pathway of isoprenoid biosynthesis in the three domains of life. *Mol Biol Evol* 28:87-99
- Longo GS, Thompson DH, Szleifer I (2007) Stability and phase separation in mixed monopolar lipid/bolalipid layers. *Biophys J* 93:2609-21
- Luisi PL (1997) About various definitions of life. *Origins of Life and Evolution of the Biosphere* 28:613-622
- Mangold HK, Paltauf F (1983) *Ether Lipids: Biochemical and Biomedical Aspects*. Academic Press
- Martin W, Muller M (1998) The hydrogen hypothesis for the first eukaryote. *Nature* 392:37-41
- Martin W, Russell MJ (2003) On the origins of cells: a hypothesis for the evolutionary transitions from abiotic geochemistry to chemoautotrophic prokaryotes, and from prokaryotes to nucleated cells. *Philos Trans R Soc Lond B Biol Sci* 358:59-83; discussion 83-5
- Martin W, Stoebe B, Goremykin V, Hapsmann S, Hasegawa M, Kowallik KV (1998) Gene transfer to the nucleus and the evolution of chloroplasts. *Nature* 393:162-5
- Matsuzaki M, Kuroiwa H, Kuroiwa T, Kita K, Nozaki H (2008) A cryptic algal group unveiled: a plastid biosynthesis pathway in the oyster parasite *Perkinsus marinus*. *Mol Biol Evol* 25:1167-79
- Mavelli F, Ruiz-Mirazo K (2007) Stochastic simulations of minimal self-reproducing cellular systems. *Philos Trans R Soc Lond B Biol Sci* 362:1789-802
- McCollom TM, Ritter G, Simoneit BR (1999) Lipid synthesis under hydrothermal conditions by Fischer-Tropsch-type reactions. *Orig Life Evol Biosph* 29:153-66
- McGarvey DJ, Croteau R (1995) Terpenoid metabolism. *Plant Cell* 7:1015-26
- Monnard PA, Deamer DW (2002) Membrane self-assembly processes: steps toward the first cellular life. *Anat Rec* 268:196-207
- Moreira D, Lopez-Garcia P (1998) Symbiosis between methanogenic archaea and delta-proteobacteria as the origin of eukaryotes: the syntrophic hypothesis. *J Mol Evol* 47:517-30
- Moreno A, Ruiz-Mirazo K (2009) The problem of the emergence of functional diversity in prebiotic evolution. *Biol Philos* 24:585-605
- Morii H, Kiyonari S, Ishino Y, Koga Y (2009) A novel biosynthetic pathway of archaetidyl-myoinositol via archaetidyl-myoinositol phosphate from CDP-archaeol and D-glucose 6-phosphate in methanoarchaeon *Methanothermobacter thermoautotrophicus* cells. *J Biol Chem* 284:30766-74
- Morii H, Koga Y (2003) CDP-2,3-Di-O-geranylgeranyl-sn-glycerol:L-serine O-archaetidyltransferase (archaetidylserine synthase) in the methanogenic archaeon *Methanothermobacter thermoautotrophicus*. *J Bacteriol* 185:1181-9
- Morii H, Nishihara M, Koga Y (2000) CTP:2,3-di-O-geranylgeranyl-sn-glycero-1-phosphate cytidyltransferase in the methanogenic archaeon *Methanothermobacter thermoautotrophicus*. *J Biol Chem* 275:36568-74
- Morii H, Ogawa M, Fukuda K, Taniguchi H, Koga Y (2010) A revised biosynthetic pathway for phosphatidylinositol in Mycobacteria. *J Biochem* 148:593-602
- Morowitz HJ, Deamer DW, Smith T (1991) Biogenesis as an evolutionary process. *J Mol Evol* 33:207-8
- Morowitz HJ, Heinz B, Deamer DW (1988) The chemical logic of a minimum protocell. *Orig Life Evol Biosph* 18:281-7
- Moustafa A, Reyes-Prieto A, Bhattacharya D (2008) Chlamydiae has contributed at least 55 genes to Plantae with predominantly plastid functions. *PLoS ONE* 3:e2205
- Mulkidjanian AY (2009) On the origin of life in the zinc world: 1. Photosynthesizing, porous edifices built of hydrothermally precipitated zinc sulfide as cradles of life on Earth. *Biol Direct* 4:26
- Mulkidjanian AY, Galperin MY (2009) On the origin of life in the zinc world. 2. Validation of the hypothesis on the photosynthesizing zinc sulfide edifices as cradles of life on Earth. *Biol Direct* 4:27
- Mulkidjanian AY, Galperin MY, Koonin EV (2009) Co-evolution of primordial membranes and membrane proteins. *Trends Biochem Sci* 34:206-15
- Mulkidjanian AY, Galperin MY, Makarova KS, Wolf YI, Koonin EV (2008) Evolutionary primacy of sodium bioenergetics. *Biol Direct* 3:13

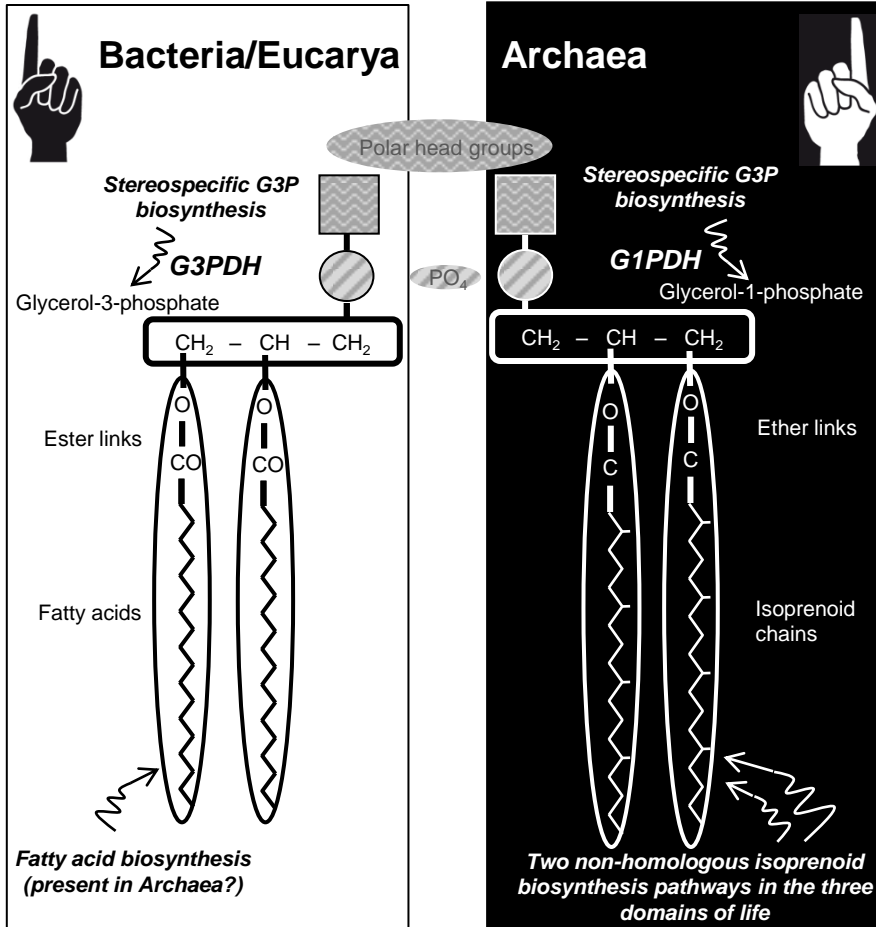
- Mulkidjanian AY, Makarova KS, Galperin MY, Koonin EV (2007) Inventing the dynamo machine: the evolution of the F-type and V-type ATPases. *Nat Rev Microbiol* 5:892-9
- Nakanishi-Matsui M, Futai M (2006) Stochastic proton pumping ATPases: from single molecules to diverse physiological roles. *IUBMB Life* 58:318-22
- Nassoy P, Goldmann M, Bouloussa O, Rondelez F (1995) Spontaneous chiral segregation in bidimensional films. *Phys Rev Lett* 75:457-460
- Ng SY, Chaban B, VanDyke DJ, Jarrell KF (2007) Archaeal signal peptidases. *Microbiology* 153:305-14
- Nishihara M, Koga Y (1997) Purification and properties of sn-glycerol-1-phosphate dehydrogenase from *Methanobacterium thermoautotrophicum*: characterization of the biosynthetic enzyme for the enantiomeric glycerophosphate backbone of ether polar lipids of Archaea. *J Biochem* 122:572-6
- Nishihara M, Yamazaki T, Oshima T, Koga Y (1999) sn-glycerol-1-phosphate-forming activities in Archaea: separation of archaeal phospholipid biosynthesis and glycerol catabolism by glycerophosphate enantiomers. *J Bacteriol* 181:1330-3
- Oparin AI (1976) Evolution of the concepts of the origin of life, 1924-1974. *Orig Life* 7:3-8
- Ouzounis CA, Kunin V, Darzentas N, Goldovsky L (2006) A minimal estimate for the gene content of the last universal common ancestor--exobiology from a terrestrial perspective. *Res Microbiol* 157:57-68
- Payandeh J, Fujihashi M, Gillon W, Pai EF (2006) The crystal structure of (S)-3-O-geranylgeranylgeranyl glyceryl phosphate synthase reveals an ancient fold for an ancient enzyme. *J Biol Chem* 281:6070-8
- Pereto J (2005) Controversies on the origin of life. *Int Microbiol* 8:23-31
- Pereto J, Lopez-Garcia P, Moreira D (2004) Ancestral lipid biosynthesis and early membrane evolution. *Trends Biochem Sci* 29:469-77
- Perez-Jimenez R, Ingles-Prieto A, Zhao ZM, Sanchez-Romero I, Alegre-Cebollada J, Kosuri P, Garcia-Manyes S, Kappock TJ, Tanokura M, Holmgren A, Sanchez-Ruiz JM, Gaucher EA, Fernandez JM (2011) Single-molecule paleoenzymology probes the chemistry of resurrected enzymes. *Nat Struct Mol Biol* 18:592-6
- Robertson JD (1964) Unit membranes: a review with recent new studies of experimental alterations and a new subunit structure in synaptic membranes. In: Locke M (ed) *Cellular Membranes in Development*. Academic, New York/London, p 1-81
- Rose RW, Bruser T, Kissinger JC, Pohlschroder M (2002) Adaptation of protein secretion to extremely high-salt conditions by extensive use of the twin-arginine translocation pathway. *Mol Microbiol* 45:943-50
- Rutters H, Sass H, Cypionka H, Rullkötter J (2001) Monoalkylether phospholipids in the sulfate-reducing bacteria *Desulfosarcina variabilis* and *Desulforhabdus amnigenus*. *Arch Microbiol* 176:435-42
- Scher DS, Rodwell VW (1989) 3-Hydroxy-3-methylglutaryl coenzyme A lyase from *Pseudomonas mevalonii*. *Biochim Biophys Acta* 1003:321-6
- Segre D, Ben-Eli D, Deamer DW, Lancet D (2001) The lipid world. *Orig Life Evol Biosph* 31:119-45
- She Q, Singh RK, Confalonieri F, Zivanovic Y, Allard G, Awayez MJ, Chan-Weiher CC, Clausen IG, Curtis BA, De Moors A, Erauso G, Fletcher C, Gordon PM, Heikamp-de Jong I, Jeffries AC, Kozera CJ, Medina N, Peng X, Thi-Ngoc HP, Redder P, Schenk ME, Theriault C, Tolstrup N, Charlebois RL, Doolittle WF, Duguet M, Gaasterland T, Garrett RA, Ragan MA, Sensen CW, Van der Oost J (2001) The complete genome of the crenarchaeon *Sulfolobus solfataricus* P2. *Proc Natl Acad Sci U S A* 98:7835-40
- Shimada H, Yamagishi A (2011) Stability of heterochiral hybrid membrane made of bacterial sn-G3P lipids and archaeal sn-G1P lipids. *Biochemistry* 50:4114-20
- Singer SJ (2004) Some early history of membrane molecular biology. *Annu Rev Physiol* 66:1-27
- Singer SJ, Nicolson GL (1972) The fluid mosaic model of the structure of cell membranes. *Science* 175:720-31
- Sinninghe Damste JS, Strous M, Rijpstra WI, Hopmans EC, Geenevasen JA, van Duin AC, van Niftrik LA, Jetten MS (2002) Linearly concatenated cyclobutane lipids form a dense bacterial membrane. *Nature* 419:708-12

- Soderberg T, Chen A, Poulter CD (2001) Geranylgeranylgeranyl phosphate synthase. Characterization of the recombinant enzyme from *Methanobacterium thermoautotrophicum*. *Biochemistry* 40:14847-54
- Szathmary E (2007) Coevolution of metabolic networks and membranes: the scenario of progressive sequestration. *Philos Trans R Soc Lond B Biol Sci* 362:1781-7
- Szostak JW, Bartel DP, Luisi PL (2001) Synthesizing life. *Nature* 409:387-90
- Tawfik DS, Griffiths AD (1998) Man-made cell-like compartments for molecular evolution. *Nat Biotechnol* 16:652-6
- Vignais PM, Billoud B, Meyer J (2001) Classification and phylogeny of hydrogenases. *FEMS Microbiol Rev* 25:455-501
- Voynova NE, Rios SE, Miziorko HM (2004) *Staphylococcus aureus* mevalonate kinase: isolation and characterization of an enzyme of the isoprenoid biosynthetic pathway. *J Bacteriol* 186:61-7
- Wächtershäuser G (1988) Before enzymes and templates: Theory of surface metabolism. *Microbiological Reviews* 52:452-484
- Wächtershäuser G (2003) From pre-cells to Eukarya--a tale of two lipids. *Mol Microbiol* 47:13-22
- Wagner M, Roger AJ, Flax JL, Brusseau GA, Stahl DA (1998) Phylogeny of dissimilatory sulfite reductases supports an early origin of sulfate respiration. *J Bacteriol* 180:2975-82
- Walde P, Goto A, Monnard PA, Wessicken M, Luisi PL (1994) Oparin's reactions revisited: enzymatic synthesis of poly(adenylic acid) in micelles and self-reproducing vesicles. *Journal of the American Chemical Society* 116:7541-7547
- Weijers JW, Schouten S, Hopmans EC, Geenevasen JA, David OR, Coleman JM, Pancost RD, Sinninghe Damste JS (2006) Membrane lipids of mesophilic anaerobic bacteria thriving in peats have typical archaeal traits. *Environ Microbiol* 8:648-57
- Wilding EI, Brown JR, Bryant AP, Chalker AF, Holmes DJ, Ingraham KA, Iordanescu S, So CY, Rosenberg M, Gwynn MN (2000) Identification, evolution, and essentiality of the mevalonate pathway for isopentenyl diphosphate biosynthesis in gram-positive cocci. *J Bacteriol* 182:4319-27
- Yen MR, Harley KT, Tseng YH, Saier MH, Jr. (2001) Phylogenetic and structural analyses of the oxal family of protein translocases. *FEMS Microbiol Lett* 204:223-31
- Yen MR, Tseng YH, Nguyen EH, Wu LF, Saier MH, Jr. (2002) Sequence and phylogenetic analyses of the twin-arginine targeting (Tat) protein export system. *Arch Microbiol* 177:441-50
- Youngleson JS, Jones DT, Woods DR (1989) Homology between hydroxybutyryl and hydroxyacyl coenzyme A dehydrogenase enzymes from *Clostridium acetobutylicum* fermentation and vertebrate fatty acid beta-oxidation pathways. *J Bacteriol* 171:6800-7
- Yuan J, Zweers JC, van Dijl JM, Dalbey RE (2010) Protein transport across and into cell membranes in bacteria and archaea. *Cell Mol Life Sci* 67:179-99

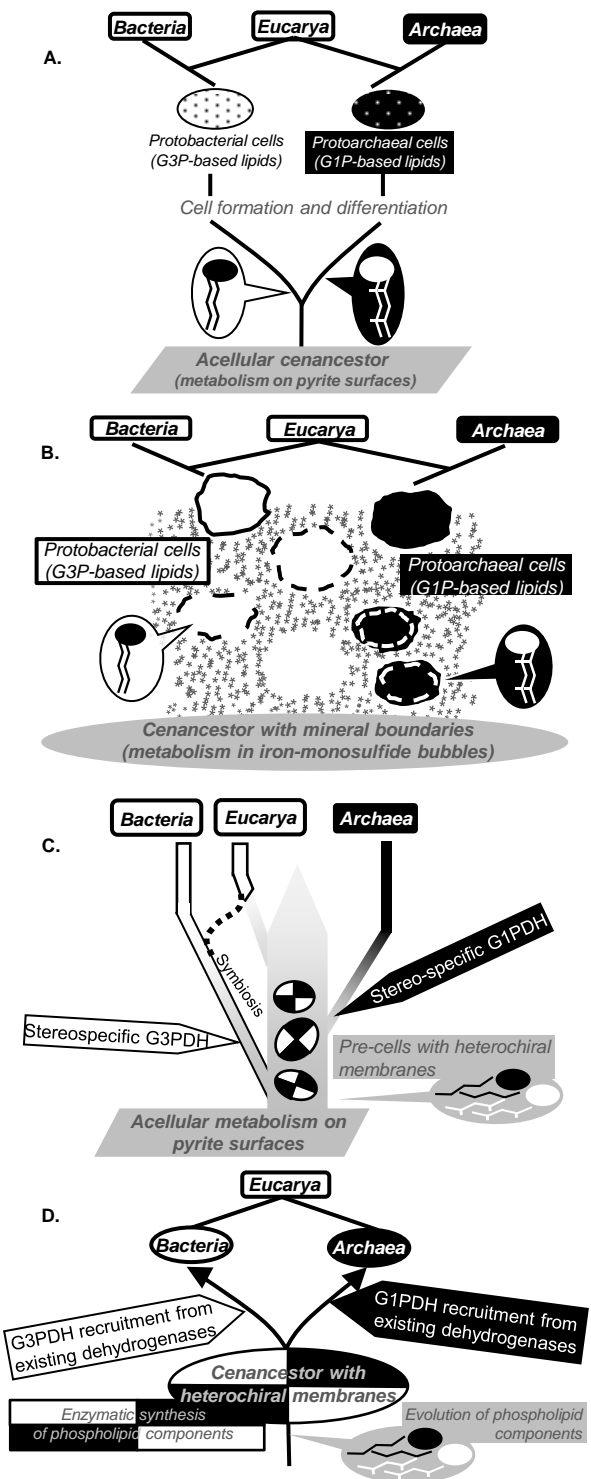
Fig. 1. Two types of phospholipids in the three domains of life.

Fig. 2. The boundary status in the cenancestor.

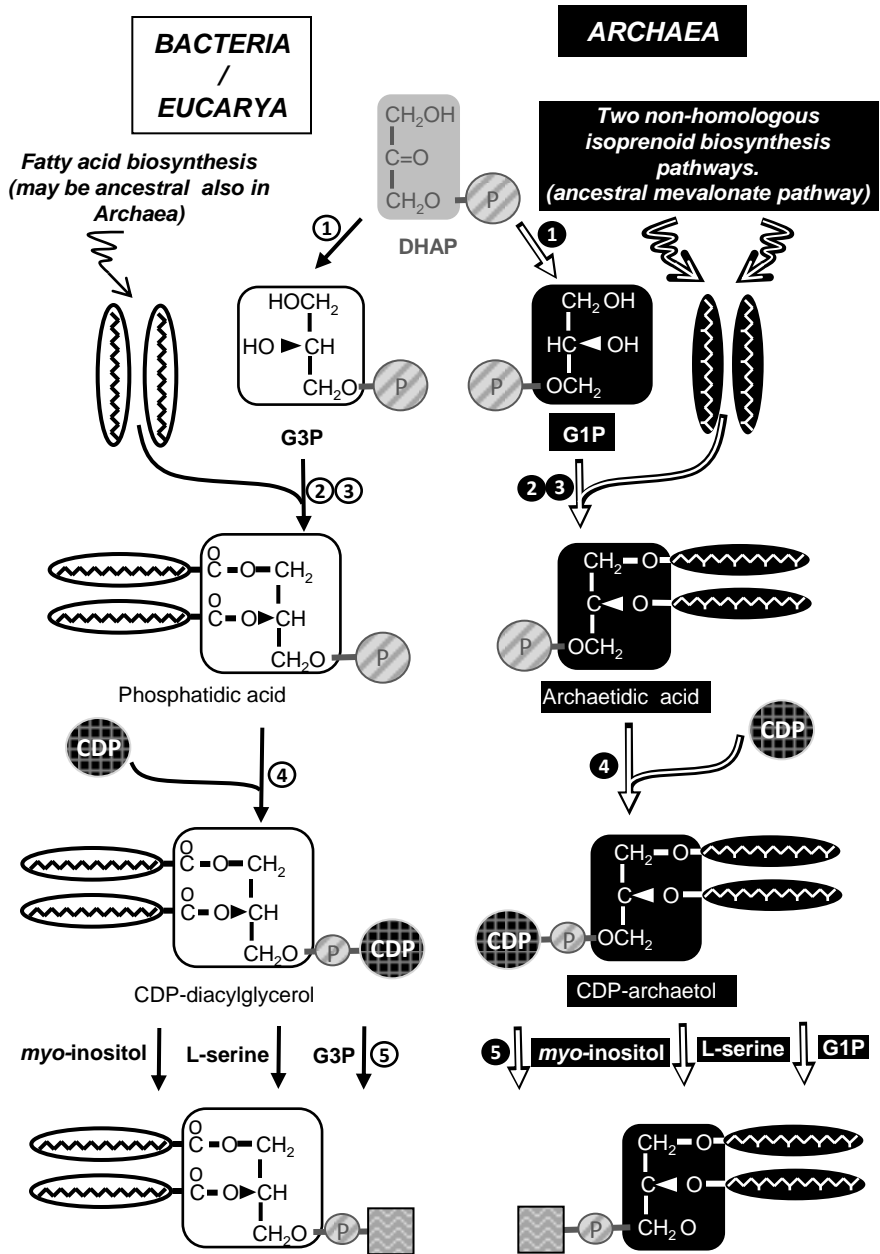
Fig. 3. Some representative steps of the phospholipid biosynthetic pathways in bacteria/eukaryotes and archaea. In bacteria and eukaryotes: (1) G-3-P dehydrogenase; (2) G3P acyltransferase; (3) 1-acylglycero-3-phosphate acyltransferase; (4) CDP-diacylglycerol synthase; (5) CDP alcohol phosphatidyltransferase. In archaea: (1) G1P dehydrogenase; (2) (S)-3-O-geranylgeranylgeranyl phosphate synthase; (3) (S)-2,3-Di-O-geranylgeranylgeranyl phosphate synthase; (4) CDP-archaeol synthase; (5) archeatidylserine synthase (and other homologues with related functions).



Lombard et al. Figure 1.



Lombard et al. Figure 2.



Lombard et al. Figure 3.

ANNEXE 5.

An ACP independent fatty acid synthesis pathway in archaea: implications for the origin of phospholipids.

Lombard, J., López-García P et D. Moreira. 2011.

Molecular biology and evolution **29**:3261-5

An ACP-Independent Fatty Acid Synthesis Pathway in Archaea: Implications for the Origin of Phospholipids

Jonathan Lombard, Purificación López-García, and David Moreira*

Unité d'Ecologie, Systématique et Evolution, Université Paris-Sud, Orsay, France

*Corresponding author: E-mail: david.moreira@u-psud.fr.

Associate editor: Andrew Roger

Abstract

Fatty acids (FAs) are major building blocks of membrane phospholipids in bacteria and eukaryotes. Their presumed absence in archaea led to propose a late origin in bacteria and eukaryotes and that the last common ancestor of living organisms (the cenancestor) was devoid of both FA and phospholipid membranes. However, small FA amounts and homologs of bacterial FA biosynthesis enzymes are found in archaea. We have investigated the origin of these archaeal enzymes using phylogenomic analyses of all enzymes of the main bacterial FA biosynthesis pathway. Our results suggest that modern archaea and their last common ancestor possessed a complete pathway except for the acyl carrier protein (ACP) processing machinery, which evolved in the bacterial lineage. This has not only implications for archaeal physiology but also opens the possibility for the presence of ACP-independent FA synthesis in the cenancestor, which may have been endowed with FA-phospholipid membranes.

Key words: fatty acid synthesis, acyl carrier protein, archaea, early evolution, cenancestor.

Although fatty acids (FAs) play major roles as storage molecules, secondary metabolites, and components of bacterial and eukaryotic membrane phospholipids, little is known about them in archaea. This reflects that, most often, archaeal membrane phospholipids are considered to lack FA and to carry isoprenoids instead (Boucher 2007). In fact, although bacterial and eukaryotic phospholipids contain FA bound to *sn*-glycerol-3-phosphate with ester links, in archaea, isoprenoids are linked to *sn*-glycerol-1-phosphate through ether bounds (De Rosa et al. 1986; Boucher 2007). These striking differences figured among the main cellular characteristics that supported archaea as an independent domain of life (Woese et al. 1978).

However, experimental data show that FAs also exist in archaea. Biochemical evidence for FA synthase machineries (Pugh et al. 1971; Pugh and Kates 1994) and small amounts of free FA or their derivatives (Kates et al. 1968; Langworthy et al. 1974; Tornabene et al. 1978; Carballeira et al. 1997; Nishihara et al. 2000) have been reported in diverse archaea. Phospholipids bearing FA have been detected in very diverse euryarchaeotes (Gattinger et al. 2002), and, as in bacteria and eukaryotes, archaeal FAs are sometimes involved in protein structure (Colella et al. 1998) and acylation (Pugh and Kates 1994). Finally, archaeal homologs of bacterial FA biosynthesis enzymes have been reported in archaeal genomes (Peretó et al. 2004; Iverson et al. 2012). Despite these arguments, the significance of FA in archaea (including the last archaeal common ancestor [LACA]) remains poorly acknowledged in recent scientific literature (Boucher 2007; Dibrova et al. 2011). Similarly, the assumption that the last common ancestor of living organisms (the cenancestor) lacked FA has even served as a supplementary argument to propose that this organism could not synthesize phospholipids and, therefore, was devoid of lipid membranes (Martin and Russell 2003).

Determining the evolutionary origin of archaeal FA has implications both to envisage early evolution scenarios and to better understand archaeal physiology. FA biosynthesis occurs through repetitive condensation of acyl (malonyl-CoA) groups in a series of cyclic steps (fig. 1) catalyzed by enzymes that may be individual (FA synthase type II [FAS-II]) or fused in one unique polypeptide (FAS-I). FAS-II is the main FA biosynthesis pathway in most bacteria, mitochondria, and chloroplasts (White et al. 2005), whereas FAS-I is more rare in bacteria but predominant in the eukaryotic cytoplasm (Schweizer and Hofmann 2004). As we were particularly interested by the presence of FAS homologs in archaea and only individual FAS-II but not polypeptidic FAS-I homologs were detected in the available archaeal genome sequences, we removed the FAS-I sequences from our analyses and restricted our systematic phylogenomic analyses to the FAS-II individual peptides. Our results support that the actual distinctive feature between archaea and bacteria is not the presence of the FAS system itself, but the lack of the FAS-related acyl carrier protein (ACP) processing machinery in archaea.

FAS-II Homologs in Archaea

In bacteria, several reactions supply the malonyl-ACP building blocks necessary to FAS-II: 1) acetyl-coenzyme A (CoA) is converted into malonyl-CoA by the acetyl-CoA carboxylase (ACC), 2) a phosphopantetheine group from CoA is attached by an ACP synthase (ACPS) to the apo-ACP, and 3) the malonyl-CoA:ACP transacylase (MCAT) charges malonyl-CoA to holo-ACP (fig. 1). A wide diversity of archaea bear ACC homologs of possible monophyletic origin, suggesting their vertical inheritance from the cenancestor (Lombard and Moreira 2011a). ACP is a peptide cofactor that covalently binds all FA intermediates, channeling them among the FAS-II enzymes. Except for a few unrelated species that

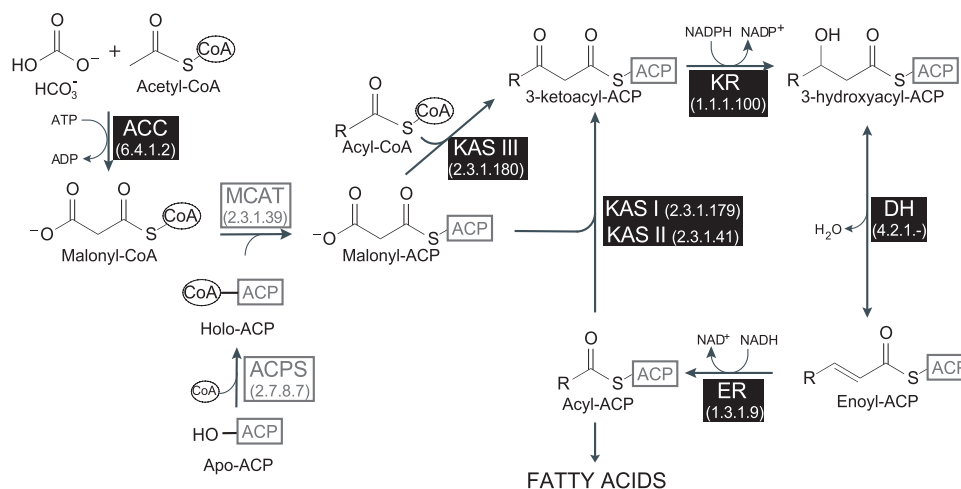


Fig. 1. Archaeal homologs of the archetypal bacterial FAS-II pathway. The bacterial FAS-II enzymes are represented. Protein names and EC numbers are indicated within boxes. Black boxes indicate the existence of archaeal homologs, potentially involved in an alternative archaeal CoA-dependent pathway, whereas gray boxes denote bacterial enzymes absent in most archaea. ACC, acetyl-CoA carboxylase; ACP, acyl-carrier protein; ACPS, acyl carrier protein synthase; CoA, coenzyme A; DH, beta-hydroxyacyl dehydratase; ER, enoyl reductase; KAS, beta-ketoacyl synthase; KR, beta-ketoacyl reductase; MCAT, malonyl-coenzymeA:acyl-carrier-protein transacylase.

probably acquired ACP by independent horizontal gene transfers (HGTs), we did not retrieve homologs of the ACP-processing machinery (ACP, ACPS, and MCAT) in archaea. This contrasted with the presence of homologs of all the other FAS genes in a large diversity of archaea (supplementary table S1, Supplementary Material online).

Once malonyl-ACP is produced, it is condensed with acetyl-CoA (beta-ketoacyl synthase III [KAS III], FabH) to yield ketoacyl-ACP, an intermediate that is sequentially reduced (beta-ketoacyl reductase [KR], FabG), dehydrated (beta-hydroxyacyl dehydratase [DH], FabZ or FabA), and reduced again (enoyl reductase [ER], FabI). A subsequent extension with malonyl-ACP (beta-ketoacyl synthases KAS I or II, respectively, FabB and FabF) conducts a new elongation cycle (fig. 1). Although nonhomologous replacements exist in some bacteria (Marrakchi et al. 2003), we have focused our search on the most common archetypal FAS-II enzymes.

A phylogenetic analysis of the thiolase superfamily, to which KAS I, KAS II, and KAS III belong, has recently been published (Jiang et al. 2008). It separated this superfamily into five clusters based on phylogenetic trees and active site architecture: archaeal thiolases, bacterial + eukaryotic thiolases, bacterial + eukaryotic KAS I–KAS II, archaeal and eukaryotic hydroxymethylglutaryl-CoA synthase (HMGS), and bacterial+eukaryotic KAS III-chalcone synthases (CHS). However, this can be questioned as only one very divergent archaeal sequence was used as an outgroup, and HMGS was recently shown to be universal and probably present in the cenancestor (Lombard and Moreira 2011b). Thus, we carried out a more complete analysis of the thiolase superfamily (supplementary fig. S1, Supplementary Material online). Although these enzymes preserve strong structural similarity, primary sequence conservation is low. However, similarities within each group of sequences point to four different clusters with distinctive enzymatic functions: 1) universally distributed thiolases, 2) bacterial KAS I–KAS II, 3) bacterial KAS III–CHS,

and 4) universally distributed HMGS. Thiolase and HMGS sequences supported the separation of most bacterial and archaeal sequences into two monophyletic groups, strongly suggesting that the cenancestor was endowed with these two enzymes. In contrast, KAS is widespread in bacteria and absent in archaea. Thiolases and HMGS catalyze the removal of an acetyl-CoA moiety from an acyl-CoA chain or the opposite condensation reaction, so the major difference between KAS and these other enzymes is the use of substrates linked to ACP instead of CoA. Thus, our results support that the cenancestor was able to (de)acetylate acyl-CoA chains, whereas the use of acyl-ACP seems to be a more recent bacterial derived character.

We obtained similar results on KR sequences that, together with ER, belong to the short-chain dehydrogenases/reductases superfamily (Kallberg et al. 2002). As this family is very diverse, we reconstructed independent phylogenies for KR and ER. The KR phylogeny incorporated the closely related PhaB acetoacetyl-CoA reductases, which are involved in polyhydroxyalkanoate (PHA) synthesis in bacteria and haloarchaea. The main enzymatic difference between PhaB and KR is the use of CoA- instead of ACP-linked substrates, although some enzymes are known to use both (Taguchi et al. 1999; Ren et al. 2000). Although PhaB sequences were scattered among KR homologs, the bacterial KR-PhaB sequences clustered according to main bacterial phyla (fig. 2). Some archaeal sequences were dispersed among the bacterial KR-PhaB, indicating independent HGTs from bacteria. Nevertheless, a wide diversity of uncharacterized archaeal sequences branched at the base of the bacterial KR-PhaB cluster. Maximum likelihood (ML) phylogenetic reconstruction of these archaeal sequences showed at least two paralogs in crenarchaeota (with some HGTs from them to other groups), sisters of a basal group of euryarchaeota (supplementary fig. S2, Supplementary Material online). This result indicates the ancient presence of these genes not only in the last

et al. 2011) or that it only bore abiotically synthesized FA (Koonin and Martin 2005). In this context, the preeminence of FA in present-day bacterial membranes could be explained by the evolution of the very efficient ACP-processing machinery in this lineage, whereas archaea would have recruited other ancestral membrane-synthesizing mechanisms (e.g., isoprenoids [Lombard and Moreira 2011b]) and relegated FA to different cellular functions.

Materials and Methods

Similarity searches were done with basic local alignment search tool-p (Altschul et al. 1990) or psi-BLAST for DH analyses (Altschul et al. 1997) against complete genomes available in GenBank (supplementary table S2, Supplementary Material online) using seed sequences retrieved from KEGG (<http://www.genome.jp/kegg>). For each seed, homologs were aligned with Muscle 3.6 (Edgar 2004). Redundant sequences and sequences out of the size range known from the literature were removed. Preliminary trees were constructed by approximate ML with FastTree 2.1.3 (Price et al. 2010) and refined by iterative steps to remove the sequences the most divergent from the archaeal ones. These divergent sequences and ambiguously aligned regions were detected with Guidance (Penn et al. 2010) and discarded using NET from the MUST package (Philippe 1993). More accurate phylogenies were inferred by ML on the refined sequence data sets using TREEFINDER (Jobb et al. 2004) with the LG+ Γ model (Le and Gascuel 2008) and four rate categories. Statistical support was assessed by approximate bootstrap (100 replicates).

Supplementary Material

Supplementary tables S1 and S2 and figures S1–S4 are available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>).

Acknowledgments

This work was supported by the interdisciplinary programs OPV (Origine des Planètes et de la Vie) and InTerrVie (Interactions Terre-Vie) of the French CNRS and INSU. J.L. is recipient of a PhD fellowship of the French Research Ministry.

References

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol.* 215:403–410.
- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25: 3389–3402.
- Autio KJ, Guler JL, Kastaniotis AJ, Englund PT, Hiltunen JK. 2008. The 3-hydroxyacyl-ACP dehydratase of mitochondrial fatty acid synthesis in *Trypanosoma brucei*. *FEBS Lett.* 582:729–733.
- Boucher Y. 2007. Lipids: biosynthesis, function, and evolution. In: Cavicchioli R, editor. *Archaea, molecular and cellular biology*. Washington, DC: ASM Press. p. 341–353.
- Carballeira NM, Reyes M, Sostre A, Huang H, Verhagen MF, Adams MW. 1997. Unusual fatty acid compositions of the hyperthermophilic archaeon *Pyrococcus furiosus* and the bacterium *Thermotoga maritima*. *J Bacteriol.* 179:2766–2768.
- Colella M, Lobasso S, Babudri F, Corcelli A. 1998. Palmitic acid is associated with halorhodopsin as a free fatty acid. Radiolabeling of halorhodopsin with 3H-palmitic acid and chemical analysis of the reaction products of purified halorhodopsin with thiols and NaBH₄. *Biochim Biophys Acta.* 1370:273–279.
- De Rosa M, Gambacorta A, Gliozzi A. 1986. Structure, biosynthesis and physicochemical properties of archaeobacterial lipids. *Microbiol Rev.* 50:70–80.
- Dibrova D, Makarova K, Galperin M, Koonin EV, Mulikidjanian A. 2011. Comparative analysis of lipid biosynthesis in Archaea and bacteria: what was the structure of first membrane lipids? Proceedings of the International Moscow Conference on Computational Biology; 2011 Jul 21–24; Moscow (Russia): Moscow State University.
- Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32:1792–1797.
- Gattinger A, Schlöter M, Munch JC. 2002. Phospholipid etherlipid and phospholipid fatty acid fingerprints in selected euryarchaeotal monocultures for taxonomic profiling. *FEMS Microbiol Lett.* 213: 133–139.
- Iverson V, Morris RM, Frazar CD, Berthiaume CT, Morales RL, Armbrust EV. 2012. Untangling genomes from metagenomes: revealing an uncultured class of marine Euryarchaeota. *Science* 335:587–590.
- Jiang C, Kim SY, Suh DY. 2008. Divergent evolution of the thiolase superfamily and chalcone synthase family. *Mol Phylogenet Evol.* 49: 691–701.
- Jobb G, von Haeseler A, Strimmer K. 2004. TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics. *BMC Evol Biol.* 4:18.
- Kallberg Y, Oppermann U, Jornvall H, Persson B. 2002. Short-chain dehydrogenases/reductases (SDRs). *Eur J Biochem.* 269:4409–4417.
- Kates M, Wassef MK, Kushner DJ. 1968. Radioisotopic studies on the biosynthesis of the glyceryl diether lipids of *Halobacterium cutirubrum*. *Can J Biochem.* 46:971–977.
- Koonin EV, Martin W. 2005. On the origin of genomes and cells within inorganic compartments. *Trends Genet.* 21:647–654.
- Langworthy TA, Mayberry WR, Smith PF. 1974. Long-chain glycerol diether and polyol dialkyl glycerol triether lipids of *Sulfolobus acidocaldarius*. *J Bacteriol.* 119:106–116.
- Le SQ, Gascuel O. 2008. An improved general amino acid replacement matrix. *Mol Biol Evol.* 25:1307–1320.
- Lombard J, Moreira D. 2011a. Early evolution of the biotin-dependent carboxylase family. *BMC Evol Biol.* 11:232.
- Lombard J, Moreira D. 2011b. Origins and early evolution of the mevalonate pathway of isoprenoid biosynthesis in the three domains of life. *Mol Biol Evol.* 28:87–99.
- Marrakchi H, Dewolf WE Jr, Quinn C, et al. (12 co-authors). 2003. Characterization of *Streptococcus pneumoniae* enoyl-(acyl-carrier protein) reductase (FabK). *Biochem J.* 370:1055–1062.
- Martin W, Russell MJ. 2003. On the origins of cells: a hypothesis for the evolutionary transitions from abiotic geochemistry to chemoautotrophic prokaryotes, and from prokaryotes to nucleated cells. *Philos Trans R Soc Lond B Biol Sci.* 358:59–83; discussion 83–55.
- Nishihara M, Nagahama S, Ohga M, Koga Y. 2000. Straight-chain fatty alcohols in the hyperthermophilic archaeon *Pyrococcus furiosus*. *Extremophiles* 4:275–277.
- Penn O, Privman E, Ashkenazy H, Landan G, Graur D, Pupko T. 2010. GUIDANCE: a web server for assessing alignment confidence scores. *Nucleic Acids Res.* 38:W23–W28.

- Peretó J, López-García P, Moreira D. 2004. Ancestral lipid biosynthesis and early membrane evolution. *Trends Biochem Sci.* 29:469–477.
- Philippe H. 1993. MUST, a computer package of management utilities for sequences and trees. *Nucleic Acids Res.* 21:5264–5272.
- Ploskon E, Arthur CJ, Kanari AL, Wattana-amorn P, Williams C, Crosby J, Simpson TJ, Willis CL, Crump MP. 2010. Recognition of intermediate functionality by acyl carrier protein over a complete cycle of fatty acid biosynthesis. *Chem Biol.* 17:776–785.
- Price MN, Dehal PS, Arkin AP. 2010. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS One.* 5:e9490.
- Pugh EL, Kates M. 1994. Acylation of proteins of the archaeobacteria *Halobacterium cutirubrum* and *Methanobacterium thermoautotrophicum*. *Biochim Biophys Acta.* 1196:38–44.
- Pugh EL, Wassef MK, Kates M. 1971. Inhibition of fatty acid synthetase in *Halobacterium cutirubrum* and *Escherichia coli* by high salt concentrations. *Can J Biochem.* 49:953–958.
- Ren Q, Sierro N, Witholt B, Kessler B. 2000. FabG, an NADPH-dependent 3-ketoacyl reductase of *Pseudomonas aeruginosa*, provides precursors for medium-chain-length poly-3-hydroxyalkanoate biosynthesis in *Escherichia coli*. *J Bacteriol.* 182:2978–2981.
- Schweizer E, Hofmann J. 2004. Microbial type I fatty acid synthases (FAS): major players in a network of cellular FAS systems. *Microbiol Mol Biol Rev.* 68:501–517.
- Taguchi K, Aoyagi Y, Matsusaki H, Fukui T, Doi Y. 1999. Co-expression of 3-ketoacyl-ACP reductase and polyhydroxyalkanoate synthase genes induces PHA production in *Escherichia coli* HB101 strain. *FEMS Microbiol Lett.* 176:183–190.
- Tornabene TG, Wolfe RS, Balch WE, Holzer G, Fox GE, Oro J. 1978. Phytanyl-glycerol ethers and squalenes in the archaeobacterium *Methanobacterium thermoautotrophicum*. *J Mol Evol.* 11:259–266.
- Tsuge T, Taguchi K, Seiichi T, Doi Y. 2003. Molecular characterization and properties of (R)-specific enoyl-CoA hydratases from *Pseudomonas aeruginosa*: metabolic tools for synthesis of polyhydroxyalkanoates via fatty acid beta-oxidation. *Int J Biol Macromol.* 31:195–205.
- White SW, Zheng J, Zhang YM, Rock CO. 2005. The structural biology of type II fatty acid biosynthesis. *Annu Rev Biochem.* 74:791–831.
- Woese CR, Magrum LJ, Fox GE. 1978. Archaeobacteria. *J Mol Evol.* 11:245–251.

ANNEXE 6.

The early evolution of lipid membranes and the three domains of life.

Lombard, J., López-García P et D. Moreira. 2011.

Nature Reviews Microbiology. **10**:507-515.

OPINION

The early evolution of lipid membranes and the three domains of life

Jonathan Lombard, Purificación López-García and David Moreira

Abstract | All cell membranes are composed of glycerol phosphate phospholipids, and this commonality argues for the presence of such phospholipids in the last common ancestor, or cenancestor. However, phospholipid biosynthesis is very different between bacteria and archaea, leading to the suggestion that the cenancestor was devoid of phospholipid membranes. Recent phylogenomic studies challenge this view, suggesting that the cenancestor did possess complex phospholipid membranes. Here, we discuss the implications of these recent findings for membrane evolution in archaea and bacteria, and for the origin of the eukaryotic cell.

Omnis cellula e cellula ('Every cell originates from a cell')¹.

With this eloquent epigram more than 150 years ago, Virchow established the basis of contemporary cell theory, which states that cells are the basic units of life and invariably originate from other cells by division¹. This was in perfect agreement with Darwin's 'common descent' hypothesis². That cells come from cells, and organisms from organisms, implies an unavoidable physical continuity, justifying the Darwinian idea that all species are inter-related within a tree of life (that is, a tree of cells or organisms), with its deepest node occupied by the last universal common ancestor, or cenancestor (according to Fitch)³.

All cells are bound by lipid membranes that ensure the individuality and integrity of cells and mediate their interactions with the surrounding environment⁴. Despite the crucial role of membranes in allowing the genetic and metabolic systems to interact and evolve together, most studies on the origin and early evolution of life have focused on the emergence of either the genetic system or energy and carbon metabolism^{5,6}. This long-standing dichotomous debate

— replication first versus metabolism first — left little room for membranes and, consequently, the origin and evolution of membranes has received much less attention than the origin of the genetic material or of energy and carbon metabolism. This is particularly surprising in the context of metabolism-first views, as the establishment of an electrochemical gradient across membranes to yield free energy that can be chemically stored⁷ is a universal feature that links membranes to energy metabolism.

Historically, the origin of membranes has been mostly approached from a bottom-up perspective, focusing on how amphiphilic molecules form vesicles under prebiotic conditions and serve as primordial boundaries for protocells (BOX 1). By contrast, a top-down approach, allowing the characteristics of the cenancestor's boundaries to be inferred by comparing present-day organisms, came much later, after the discovery of archaea and their distinct membranes. This led to a paradox. According to the cell theory, as cells come from cells and modern cells are bounded by lipid membranes composed of similar molecules (phospholipids), a cenancestor with phospholipid-based membranes is the most parsimonious

inference. However, two different, albeit structurally similar, kinds of phospholipids exist in nature (FIG. 1). Bacteria and eukaryotes have the same membrane biochemistry, with ester-linked fatty acid phospholipids that are based on glycerol-3-phosphate (G3P). These G3P phospholipids were thought to be universal, but the surprise came when pioneering studies of archaeal biochemistry showed that archaeal phospholipids are made of glycerol-1-phosphate (G1P) that is ether linked to isoprenoid chains^{8–10}. This chemical disparity mirrors the use of different phospholipid biosynthesis pathways in archaea and bacteria, and in particular the use of a distinctive glycerol phosphate dehydrogenase to synthesize G1P¹¹. When they were discovered, these archaeal pathways were considered to be unique and non-homologous to those of bacteria and eukaryotes¹¹.

How can these findings be reconciled with the logical inference that phospholipids are ancestral membrane components? Did their biosynthesis evolve independently in archaea and bacteria–eukaryotes? Does this imply that the cenancestor lacked lipid membranes? If so, what was the nature of the cenancestral membranes? These questions have led to controversy and raised additional, rarely explicit, issues on the evolution of eukaryotes. In this Opinion article, we explore the different hypotheses that have been proposed to answer these questions and discuss them in the light of recent phylogenomic data.

Phospholipid biosynthesis: evolved twice?

Various hypotheses have addressed the fundamental differences between archaeal and bacterial–eukaryotic phospholipids and, more specifically, the apparently unrelated nature of the pathways that synthesize the two opposed glycerol phosphate stereoisomers (FIG. 1). Koga *et al.*¹¹ openly deserted the cell theory by proposing that the cenancestor was acellular (that is, it had no membrane) (FIG. 2a). According to this radical view, phospholipid biosynthesis emerged late (relative to other views) and independently in the ancestral lineages that led to contemporary archaea and bacteria. Although this hypothesis accounts directly

for the differences between archaeal and bacterial membrane phospholipids, it is at odds with the increasing evidence pointing to a complex cenancestor that contained several

hundred genes¹². Such complexity would have required Darwinian evolution (based on natural selection and other evolutionary forces) to operate on individualized entities,

a situation that would have been achieved by membrane compartmentation¹³. To reconcile this requirement with the apparent lack of homology between the archaeal and bacterial lipid biosynthesis pathways and, hence, the absence of lipids in the cenancestor, Martin and Russell¹⁴ envisaged that the cenancestor had mineral, instead of lipid, membranes. In this model, the first cells would correspond to three-dimensional iron monosulphide compartments in a submarine chimney in which the redox, pH and temperature gradients were established by hydrothermal venting. Geochemistry would have been replaced progressively by biogeochemistry, leading to a complex cenancestor possessing ribosomes and other universally conserved features enclosed by mineral membranes. Phospholipid biosynthesis would have evolved independently during the evolution of the archaeal and bacterial lineages, allowing their respective release from the maternal chimney (FIG. 2b). However, hydrothermal systems are largely transient in nature, with timescales ranging between 1 and 10,000 years for complete hydrothermal fields and typically of less than 100 years for the individual chimneys¹⁵. As mineral-bounded cells would not have had the capacity to move between different chimneys, the whole evolutionary pathway between the origin of life and the emergence of complex archaeal and bacterial cells would have to have occurred in the same unique hydrothermal chimney at a surprising speed. More importantly, this proposal fails to postulate a mechanism to couple the formation of these mineral compartments with the replication of the inner biological components, and therefore compromises the link between the different constituents that is necessary for Darwinian evolution to act on individuals¹⁶.

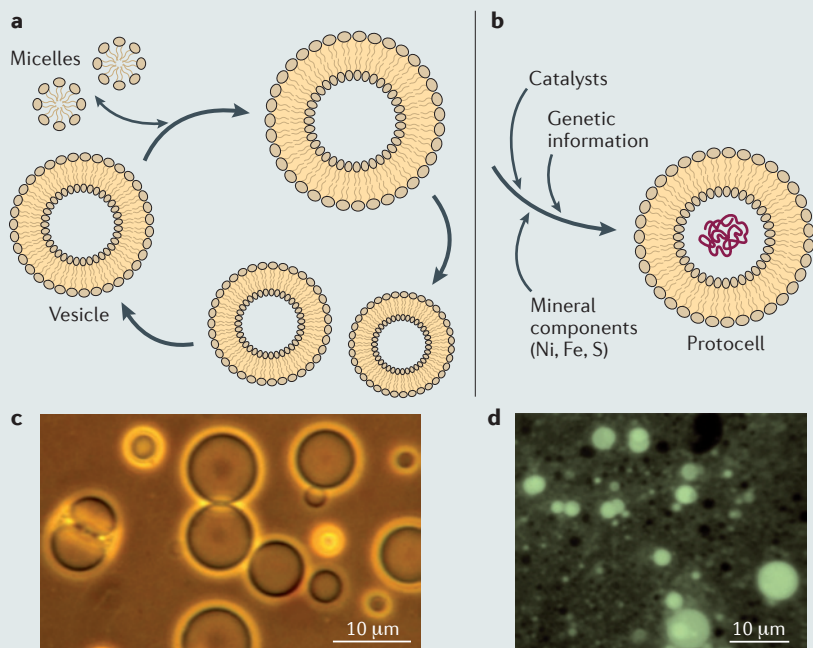
In contrast to these 'late membrane origin' hypotheses, other models invoke a much earlier origin of phospholipids. In an extension to his 'iron-sulphur world' hypothesis, Wächtershäuser¹⁷ speculated that early cellularization occurred via membranes composed of simple lipids that were synthesized non-enzymatically by either inorganic transition metals or primitive non-stereospecific enzymes, leading to a community of cenancestral pre-cells (FIG. 2c). Pre-cell heterochiral membranes would have been replaced by more stable homochiral membranes when stereospecific enzymes appeared, triggering the divergence of archaea and bacteria. Therefore, despite acknowledging the presence of lipid membranes in pre-cells, Wächtershäuser

Box 1 | Prebiotic origin of membranes

Although modern cell membranes are bilayers of glycerol phospholipids, the first cell membranes probably self-assembled from simple, single-chain amphiphilic molecules, such as monocarboxylic acids or alcohols⁷³. In contrast with monolayered micelles, vesicles expose hydrophilic groups to both the exterior and interior of their bilayer boundary, being able to encapsulate a certain volume of solution⁷⁴ and, depending on the bilayer permeability, creating gradients of particular molecules and ions⁷⁵. Furthermore, vesicles can grow and divide spontaneously⁷⁶ (see the figure). From an origin-of-life perspective, these are interesting properties that made vesicles of amphiphilic compounds replace Oparin's 'coacervates' (proteinaceous aggregates)⁷⁷ in models of primeval-cell formation^{78,79}. The possibility of incorporating the building blocks of replicating genetic polymers inside such vesicles has converted these vesicles into an attractive and tractable model for synthetic-biology experiments and protocell formation *in vitro*^{13,67}. This avenue of research is progressing quickly, to the point that the traditional focus on self-maintenance (metabolism) as a major property of life, together with self-replication (a genetic system), is shifting towards a focus on self-assembly (membranes) in contemporary origin-of-life thinking^{67,80,81}.

What kinds of amphiphilic compounds were available on the early Earth and could serve for the self-assembly of protomembranes? Two sources of such compounds are known. The first source is extraterrestrial and consists of the organic matter delivered by carbonaceous chondrites. These primitive meteorites are enriched in organic compounds, including amino acids and a variety of amphiphilic molecules; such amphiphilic molecules can assemble into vesicles spontaneously, as Deamer⁷⁸ showed in 1985. The second source is terrestrial and corresponds to the abiotic formation of hydrocarbons by Fisher-Tropsch synthesis, involving the reaction of carbon monoxide and hydrogen to form hydrocarbons in the presence of iron catalysts under hydrothermal conditions. These serpentinization reactions may have been very active in the early Archean ocean⁴⁵. Hydrocarbons can easily oxidize into mixtures of long-chain carboxylic acids and alcohols that, in the presence of glycerol, can form phospholipids⁷⁹. Membranes with an increasing presence of phospholipids may have triggered new selective pressures for the evolution of metabolism and transport⁴⁵.

The figure shows the vesicle growth and division cycle (part a) and the formation of protocells (part b). The micrographs show self-assembled vesicles formed from the amphiphilic C₃-C₁₁ carboxylic acids and polycyclic hydrocarbon derivatives found in the Murchison meteorite (part c) and from decanoic acid (part d). The vesicles formed from decanoic acid have incorporated short, fluorescently labelled DNA fragments during wet-dry-wet cycles. Micrographs courtesy of D. Deamer (University of California, Santa Cruz, USA).



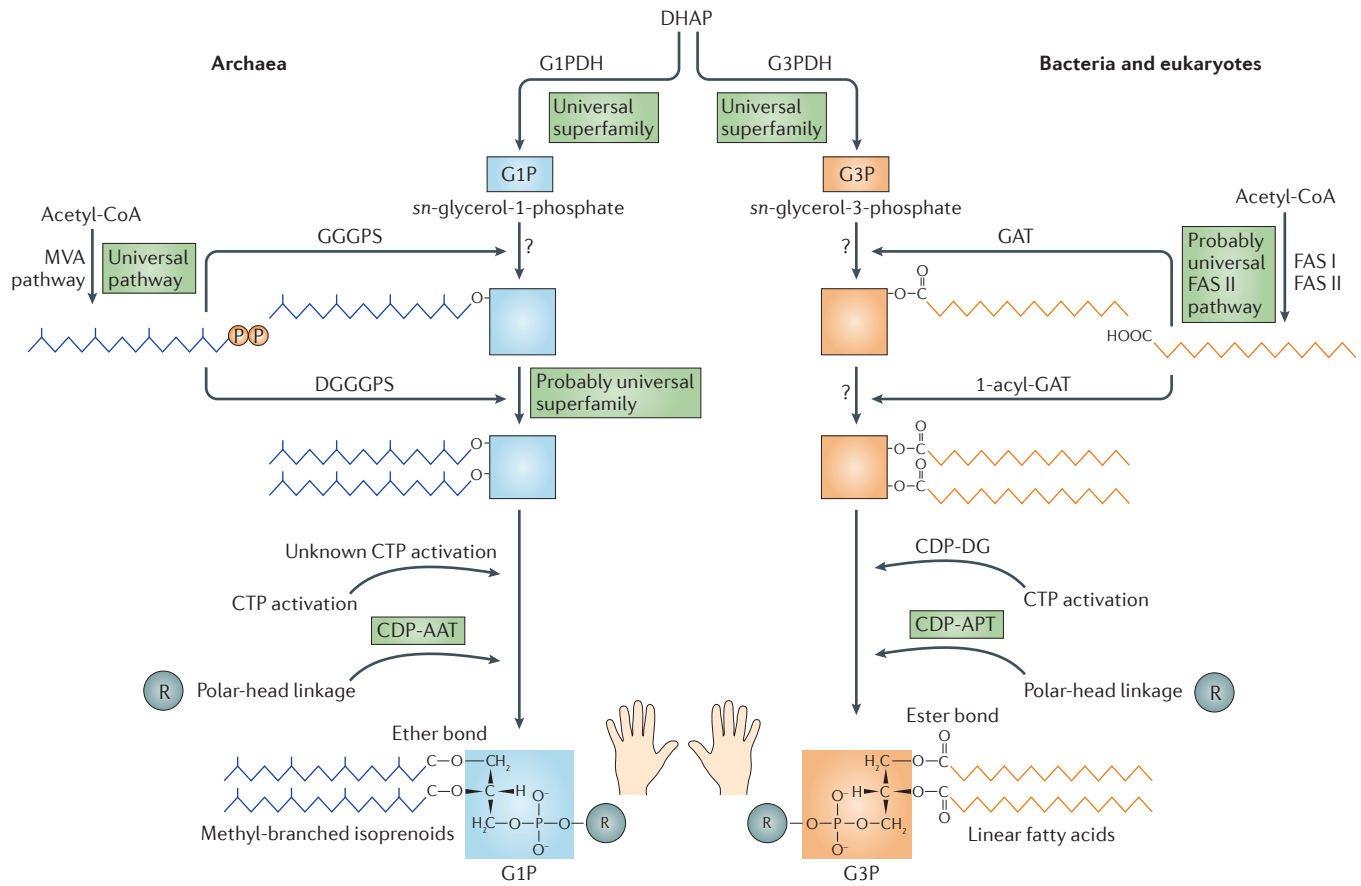


Figure 1 | Phospholipid biosynthesis pathways in archaea, bacteria and eukaryotes. Phospholipid components and the enzymes that synthesize them are different in modern archaea versus modern bacteria and eukaryotes. For some steps in the pathways, there is phylogenomic evidence either supporting the hypothesis that homologous enzymes carried out a particular step in the cenacestor (universal proteins or pathways) or indicating that the presence of the relevant enzymes in the cenacestor cannot be excluded (probably universal proteins or pathways). Cytidine diphosphate-alcohol archaetidyltransferase (CDP-AAT) and CDP-alcohol

phosphatidyltransferase (CDP-APT) are homologous in the two pathways. Polar head radicals can be serine, ethanolamine or glycerol, among others. A question mark indicates that information is unknown. CDP-DG, CDP-diacylglycerol synthase; CTP, cytidine triphosphate; DGGGPS, digerylgeranylgeranyl glyceryl phosphate synthase; DHAP, dihydroxyacetone phosphate; G1P, *sn*-glycerol-1-phosphate; G1PDH, G1P dehydrogenase; G3P, *sn*-glycerol-3-phosphate; G3PDH, G3P dehydrogenase; GAT, G1P acyltransferase; GGGPS, geranylgeranylgeranyl glyceryl phosphate synthase; MVA, mevalonate.

thought that specific lipid biosynthesis pathways evolved independently in archaea and bacteria.

Phylogenomics of lipid biosynthesis

Although textbooks often emphasize the differences between archaeal and bacterial phospholipids, the distinction is actually not so sharp. Ether links are found in bacterial (and eukaryotic) phospholipids^{18,19}, fatty acids have been detected in archaea^{20–22}, and isoprenoids are universally distributed membrane components^{23,24} (TABLE 1). The asymmetry of the glycerol phosphate stereoisomers — G1P in archaea and G3P in bacteria and eukaryotes — that are synthesized by non-homologous glycerol phosphate dehydrogenases¹¹ is the only inviolate difference. However, phylogenomic approaches, based on molecular phylogenetic analyses of genes

from complete genome sequences, have questioned the strength of this distinction, as they have uncovered the fact that both G1P dehydrogenases and G3P dehydrogenases belong to large multi-enzymatic superfamilies that are widespread in the three domains of life, and that at least one member of each superfamily probably evolved before the separation of archaea and bacteria²⁵. Therefore, the cenacestor might have used those ancestral enzymes to synthesize a mix of both G1P and G3P (FIG. 2d). Subsequent duplication of the ancestral enzymes and the recruitment of different copies in archaea and bacteria would have led to the evolution of the specific G1P dehydrogenases and G3P dehydrogenases as the two domains diverged.

Recent phylogenomic analyses also revealed that the mevalonate pathway of

isoprenoid biosynthesis, which is highly conserved in all archaea and eukaryotes and in several bacterial phyla, was probably present in the cenacestor and was lost secondarily in most bacteria, in which it was replaced by the non-homologous methylerythritol phosphate pathway²⁶. Similarly, archaeal genomes have homologues of bacterial fatty acid biosynthesis genes²⁵, and although these genes generally belong to large multigene families with complex evolutionary histories, an ancestral origin cannot be excluded²⁷. For example, a biotin-dependent carboxylase, which catalyses the incorporation of a CO₂ moiety into biotin-bearing substrates and is required for fatty acid biosynthesis, was probably present in the cenacestor²⁸. Finally, in addition to the enzymes that synthesize these phospholipid building blocks, those enzymes that link glycerol

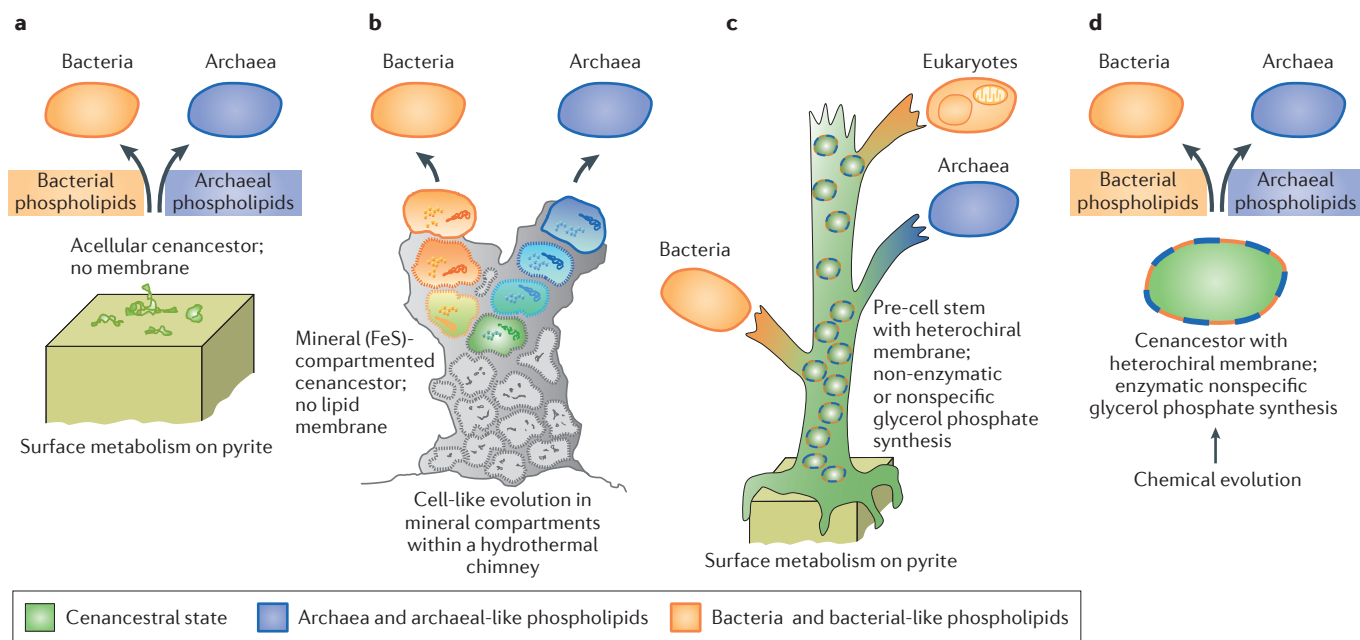


Figure 2 | Models explaining the early evolution of archaeal and bacterial phospholipid biosynthesis. **a** | Independent evolution of phospholipid biosynthesis from an acellular cenancestor¹¹. **b** | Independent evolution of phospholipid biosynthesis from a mineral-bounded cenacestral compartment¹⁴. **c** | Evolution of archaeon-specific and bacterium-specific

phospholipid biosynthesis pathways from a stem of pre-cells with heterochiral membranes¹⁷. **d** | Evolution of archaeon-specific and bacterium-specific phospholipid biosynthesis pathways from a cellular cenancestor with heterochiral membranes that are synthesized via universal but substrate-nonspecific enzymes²⁵.

phosphate to hydrocarbon chains and polar head groups also belong to universal gene families and are probably also ancestral^{29–31} (TABLE 1).

Phospholipid membranes in the cenancestor

The phylogenomic evidence discussed above suggests that the cenancestor possessed a complete toolkit for making both isoprenoid- and fatty acid-based phospholipids, possibly using a mixture of G1P and G3P stereoisomers^{25–27} (TABLE 1). The early origin of phospholipids is further supported by the universal conservation of several contemporary membrane-embedded proteins^{12,32}. These include proteins involved in membrane bioenergetics — notably an ATPase that specifically exploits transmembrane ion gradients^{29,33}, and integral membrane proteins (for example, hydrogenases and dioxygen reductases) that are involved in respiratory chains^{34–37} — and also proteins of the secretion and membrane-targeting machineries, such as the signal recognition particle and the Sec, YidC and Tat (twin-arginine translocation) protein export and membrane insertion pathways^{38–41}. In an attempt to reconcile the compelling evidence that these membrane proteins were cenacestral with the proposal for an iron monosulphide-bounded cenancestor, Koonin and Martin⁴² proposed that lipid

patches, but not a continuous lipid membrane, accumulated on mineral walls to host these proteins. However, the co-occurrence of a respiratory chain and ATPases in the cenancestor strongly suggests that ATPases had already evolved to exploit a transmembrane proton (or sodium, according to recent suggestions⁴³) gradient, which requires a continuous membrane²⁹.

In summary, in our opinion, the cenancestor probably had lipid membranes and the enzymatic machinery to synthesize modern phospholipid components, including G1P, G3P, isoprenoids and fatty acids. Contrary to previous assumptions¹⁷, heterochiral membranes formed by G1P- and G3P-based phospholipids do not appear to be intrinsically less stable than homochiral ones⁴⁴. Therefore, on the basis of phylogenomic and physicochemical considerations, we propose that the cenancestor possessed a heterochiral, complex, modern-like phospholipid membrane (FIG. 3). The large number of enzymes required to synthesize it is compatible with a complex cenancestor having a large genome¹². Differences between archaeal and bacterial membranes would have evolved as these two domains diverged from the cenancestor. From an ecological standpoint, a single ancient origin for lipid membranes seems realistic. In fact, it is difficult to imagine that

phospholipid membranes, and thus true cellularization, originated twice because when a given ancestral organism acquired a lipid membrane it would gain a strong selective advantage and supersede less efficient competitors⁴⁵. This probably did occur but would have been earlier in evolution, when cellularization first appeared, at a moment that might be considered the true origin of life.

The archaea–bacteria ‘lipid divide’

Most early evolution scenarios^{11,13,15,23} consider Archaea and Bacteria as the primary domains (that is, the two domains that diverged directly from the cenancestor), and the domain Eukarya as having originated secondarily, containing chimeric organisms that were derived from the symbiosis of an archaeon (or a member of a protoeukaryotic sister lineage to archaea) and at least one bacterium, the ancestor of mitochondria.

If the cenancestor had complex heterochiral membranes, what was the driving force triggering the ‘lipid divide’ — the differentiation of archaeal and bacterial membranes? The first possible explanation to be invoked was the instability of mixed membranes^{17,46}, but experiments with liposomes containing archaeal and bacterial phospholipids showed that the stability of homochiral

Table 1 | **Phylogenomic arguments in favour of a cenancestral phospholipid membrane**

Observations	Implications for the cenancestor
Phospholipid components	
<ul style="list-style-type: none"> • Ancestral presence of at least one member of the two dehydrogenase superfamilies to which the contemporary dehydrogenases synthesizing G1P (archaea) and G3P (bacteria and eukaryotes)²⁵ belong. 	<ul style="list-style-type: none"> • Possible enzymatic synthesis of G1P and G3P
<ul style="list-style-type: none"> • Presence of isoprenoids in the three domains of life • Conservation of the mevalonate pathway of isoprenoid synthesis in archaea, bacteria and eukaryotes²⁶ 	<ul style="list-style-type: none"> • Biosynthesis of isoprenoids
<ul style="list-style-type: none"> • Presence of fatty acids in the three domains of life • Conservation of key enzymes for fatty acid synthesis and degradation in archaea, bacteria and eukaryotes^{25,27} 	<ul style="list-style-type: none"> • Probable biosynthesis and degradation of fatty acids
<ul style="list-style-type: none"> • Broad distribution of ether-linked phospholipids and of homologues of the archaeal enzyme superfamily responsible for ether link formation in the three domains of life³⁰ 	<ul style="list-style-type: none"> • Possible hydrocarbon chain attachment to glycerol phosphate (at least) via ether links
<ul style="list-style-type: none"> • Presence of one representative of the CDP-APT family involved in polar head group attachment in archaea and bacteria³¹ 	<ul style="list-style-type: none"> • Phospholipid head group attachment in the cenancestor
Lipid membrane-associated proteins	
<ul style="list-style-type: none"> • Universally conserved H⁺ (or Na⁺) ATPase^{32,33,83} 	<ul style="list-style-type: none"> • Synthesis of ATP exploiting a transmembrane ion gradient • Need for a continuous lipid membrane
<ul style="list-style-type: none"> • Universally conserved components of respiratory chains (cytochrome <i>b</i>, Rieske protein, hydrogenases and dioxygen reductases)^{34–37} 	<ul style="list-style-type: none"> • Likely presence of a respiratory chain in the membrane that could generate a proton gradient across the membrane
<ul style="list-style-type: none"> • Universally conserved SRP system (SRP domain and its receptor)³⁸ 	<ul style="list-style-type: none"> • Targeting of proteins to the membrane
<ul style="list-style-type: none"> • Universally conserved Sec, YidC and Tat pathways^{39–41} 	<ul style="list-style-type: none"> • Protein export and insertion into the membrane

CDP-APT, cytidine diphosphate-alcohol phosphatidyltransferase; G1P, sn-glycerol-1-phosphate; G3P, sn-glycerol-3-phosphate; SRP, signal recognition particle; Tat, twin-arginine translocation.

and heterochiral mixed membranes is similar, challenging this idea⁴⁴. The past evolutionary constraints faced by the two domains might constitute another explanation. It is widely accepted that the last common ancestor of archaea was hyperthermophilic⁴⁷, so the composition of archaeal membranes may result from ancestral adaptation to extremely hot environments. Extreme physicochemical conditions induce chronic energy stress, which has to be managed by the tight regulation of membrane permeability. Thus, archaeal membranes have evolved to prevent proton leakage at high temperature and to control transmembrane electrochemical gradients at extreme pH and high salinity⁴⁸. In the bacterial lineage, the evolution of the acyl-carrier protein allowed efficient fatty acid synthesis, and this pathway was recruited for phospholipid synthesis, relegating isoprenoid biosynthesis to other cellular functions²⁷. Finally, the idea of a 'frozen accident' cannot be discarded. Components of bacterial phospholipids are also present in archaea and vice versa, but these components function in alternative cellular processes in the other domain: G1P is used in bacterial envelopes^{49,50}, fatty acids are used in archaeal metabolism⁵¹ and the universally distributed isoprenoids are involved in a range of functions^{23,24}. This suggests again

that these components are ancestral but were recruited for different uses in archaea and bacteria. Their recruitment may have occurred accidentally or by drift. It could be that enzymes from different ancestral dehydrogenase families specialized in the synthesis of opposed glycerol phosphate stereoisomers.

The origin of eukaryotic membranes

The origin of eukaryotic membranes is a problem that is rarely addressed by the different hypotheses that have been proposed to explain the emergence of eukaryotes (FIG. 4). Eukaryotic membranes have typical bacterial-like phospholipids. By contrast, the apparent conservation of the isoprenoid biosynthesis mevalonate pathway in archaea and eukaryotes, and its loss in most bacteria, could support a relationship between archaea and eukaryotes. However, recent phylogenomic analyses show that there are major differences between the archaeal and eukaryotic mevalonate pathways; archaea have the most divergent pathway, whereas eukaryotes and several bacteria appear to have retained the ancestral version²⁶. This suggests that eukaryotes inherited their membranes directly from bacteria or from a common ancestor of bacteria and eukaryotes to the exclusion of archaea. This is at odds

with the classical Woesean three-domain phylogeny rooted on the bacterial branch⁵². With regard to the eukaryotes, this phylogeny implies that the last common ancestor of archaea and eukaryotes would have had either an archaeal-like membrane that was subsequently replaced by bacterial-like phospholipids in eukaryotes (FIG. 4a), or an ancestral mixed membrane with both G1P and G3P phospholipids that evolved towards a modern archaeal-like membrane in archaea and towards a bacterial-like membrane in eukaryotes after the divergence of both lineages (the pre-cell-like model) (FIG. 4b).

Both options are problematic. Unless considering massive horizontal transfer of all the necessary genes, the mixed-membrane model implies the less parsimonious assumption that bacterial-like membranes evolved twice from the ancestral mixed membrane, in bacteria and eukaryotes independently. The fact that no archaeal-to-bacterial membrane transition has been identified so far also undermines the hypothesis that an archaeal-like membrane was secondarily replaced in eukaryotes. This also affects the view of Cavalier-Smith⁵³ that archaea and eukaryotes evolved from a Gram-positive bacterial ancestor and that archaea replaced the

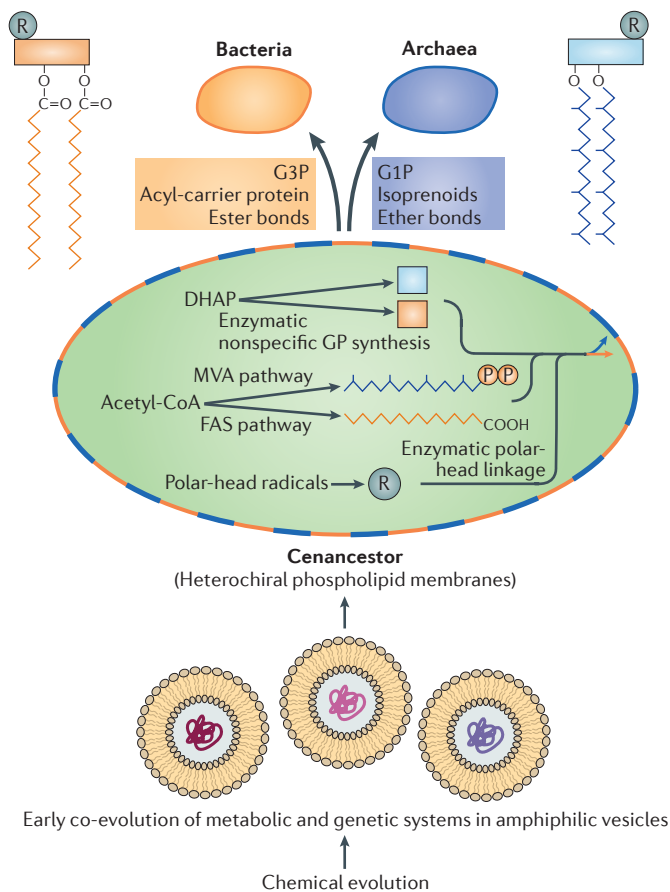


Figure 3 | Complement of enzymes involved in the biosynthesis of phospholipid components in the cenancestor, and their evolution during the archaea–bacteria split. This complement of enzymes is inferred by phylogenomic analysis of complete genome sequences of contemporary species. The cenancestor would have been able to synthesize heterochiral phospholipid membranes with a mix of *sn*-glycerol-1-phosphate (G1P) (blue) and *sn*-glycerol-3-phosphate (G3P) (orange) produced from dihydroxyacetone phosphate (DHAP), bound to isoprenoid and fatty acid lateral chains and to polar head radicals. We propose that the first cells were surrounded by amphiphilic vesicles that were synthesized abiotically and that the cenancestor already possessed a sophisticated enzymatic machinery for lipid biosynthesis. The divergence of bacteria and archaea from the cenancestor was paralleled by the specialization of their membranes. Bacteria use G3P that is bound via an ester link to fatty acids which are synthesized in an efficient way owing to the acyl-carrier protein. By contrast, archaea use G1P that is bound via an ether link to isoprenoids. GP, glycerol phosphate; MVA, mevalonate; FAS, fatty acid synthesis.

bacterial membrane secondarily (FIG. 4c). Cavalier-Smith postulates that the cenancestor was the last common bacterial ancestor and was bounded by two membranes (that is, was Gram negative), and that there was a single transition of double-membrane cells to single-membrane cells during evolution. Consequently, in his view, archaea, eukaryotes and Gram-positive bacteria, being bounded by single membranes, are monophyletic. However, in addition to the problem of a bacterial-to-archaeal membrane transition, the recent discovery of an archaeon with a double membrane, *Ignicoccus hospitalis*⁵⁴, invalidates the idea that the number of cell membranes is an

extremely conserved characteristic that was altered only once during the history of the three domains. Thus, this characteristic cannot be used to support the monophyly of archaea, eukaryotes and Gram-positive bacteria.

Today, there is little doubt that the eukaryotic cell is a chimera that is derived from the endosymbiosis of the alphaproteobacterial ancestor of mitochondria within a host cell⁵⁵. However, the nature of that host cell is still highly debated. The classical Woesean models propose that the host was a member of an independent proto-eukaryotic sister lineage to the archaea, with a nucleus and phagocytic capacity^{56–58}. However,

direct evidence for the existence of this amitochondriate proto-eukaryotic lineage is lacking, so several models (often called chimeric or symbiogenetic models) propose that eukaryotes derive directly from a symbiosis between archaea and bacteria^{56,59}. This would account for the mosaic distribution of characteristics that is observed in eukaryotes, with the genes involved in replication, transcription and translation being of archaeal origin and most genes involved in metabolism and other cellular functions having bacterial homologues^{60,61}.

How do these models explain the bacterial-like nature of eukaryotic membranes? The models can be divided into two types, which are best illustrated by, respectively, the hydrogen hypothesis and the syntrophy hypothesis for the origin of eukaryotes. The hydrogen hypothesis proposes that eukaryotes originated from the endosymbiosis of an alphaproteobacterium within an archaeon⁶², implying, similarly to some classical models, that there was a transformation of the original archaeal cell membrane into a bacterial-like membrane⁶³. In this model, the nuclear membrane would form *de novo* from nucleus-encoded bacterial phospholipids that would form vesicles encapsulating the archaeon-derived nuclear genome (FIG. 4d). The second type of chimeric models, such as the syntrophy hypothesis^{64,65}, postulate that eukaryotes derive from the endosymbiosis of an archaeon within a bacterium, whereas mitochondria derive from a second endosymbiotic event^{64–66}. In this case, the outer membranes of the consortium would be bacterial, eliminating the need for the aforementioned transition (FIG. 4e). This model additionally proposes that the external bacterial-like membrane developed an extensive endomembrane system with secretory functions, leading to the formation of the future nuclear membrane in an analogous manner to that postulated by classical autogenous models for the origin of the nucleus⁶⁵. The membrane of the ancestral archaeon would then simply be lost.

In summary, most classical and chimeric models for the origin of eukaryotes require an archaeal-to-bacterial membrane transition. However, traces of this type of transition have never been found in contemporary bacteria or archaea, indicating that either this transition was extremely rare or it never occurred. In that case, the distribution of phospholipids in contemporary membranes would actually argue in favour of a bacterium at the origin of the cytoplasmic membrane of the proto-eukaryotic host that acquired mitochondria.

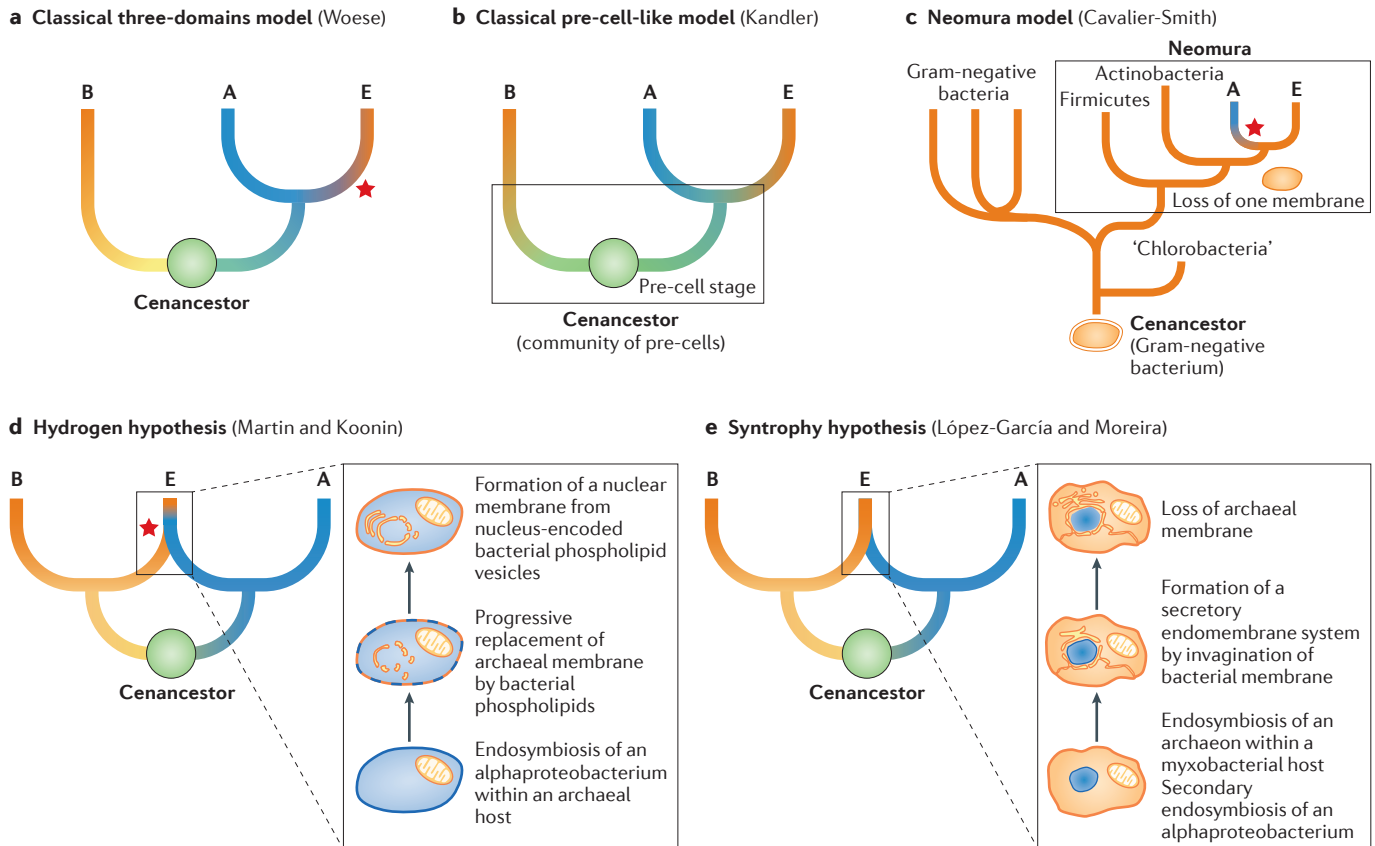


Figure 4 | Models explaining the bacterial-like nature of phospholipid membranes in eukaryotes. Different views of the evolutionary relationships among the three domains of life are depicted as simplified phylogenetic trees. The cenancestor (green) is placed at the root of the trees. Orange branches correspond to organisms with bacterial-like phospholipids, and blue branches correspond to organisms with archaeal-like phospholipids. Red stars indicate transitions from one type of phospholipid (archaeal or

bacterial) to the other. Insets in the hydrogen and syntrophy hypotheses provide details about lipid evolution after the chimeric origin of eukaryotes by a symbiosis between archaea and bacteria. **a** | The classical three-domain model from Woese⁵². **b** | The classical pre-cell-like model from Kandler⁸². **c** | The Neomura model from Cavalier-Smith⁵³. **d** | The hydrogen hypothesis as detailed by Martin and Koonin⁶³. **e** | The syntrophy hypothesis as detailed by López-García and Moreira⁶⁵. A, Archaea; B, Bacteria; E, Eukarya.

Conclusion and perspectives

Recent phylogenomic results suggest that pathways for the biosynthesis of the different phospholipid moieties and, hence, of phospholipid membranes, are ancient (TABLE 1). They date back at least to the cenancestor, which in our opinion probably possessed phospholipid membranes that were predated by simpler membranes in earlier evolutionary periods. As a corollary, and in contrast with hypotheses proposing a late origin of lipid membranes, this opens the possibility that lipid membranes have existed all along the evolution of life, from prebiotic times, when the first protocells were formed by vesicles of amphiphilic molecules of abiotic origin (BOX 1), to fully modern-like cells endowed with an enzyme set for phospholipid synthesis. This would extend the cell theory back to a protocellular era. Therefore, bottom-up approaches based on prebiotic chemistry studies can be enriched by top-down phylogenomic

analyses that shed light on the characteristics of early cell membranes. Such synergy is necessary to clarify two of the most challenging and interlinked issues in origin-of-life research, namely the origin of the first cells, and the coupling of metabolism and genetic information⁵. The field of synthetic biology may contribute to filling this gap in our knowledge by fostering the study of vesicles that are built with prebiotically plausible lipids containing simple replicative polymers, and using these vesicles as primordial life models^{13,67}.

In addition to the information that can be inferred about the cenancestor and even earlier life forms, membranes are crucial to understanding the evolution of the three domains of life. Bacteria and archaea have different phospholipids, but also different DNA replication mechanisms⁶⁸. Could the driving forces behind the lipid and DNA replication machineries be the same? Considering that DNA replication,

chromosome segregation and cell division are intimately coupled in bacteria and archaea^{69,70}, this possibility is worth exploring. Also, an accurate biochemical description of modern membranes, considering not only the major but also the rare lipids, is still missing. This information could overcome the supposedly insurmountable difficulties in explaining the differences between archaeal and bacterial lipids. For example, the presence of fatty acids in archaea, as described several decades ago^{20,21,71,72}, is often neglected, as is the presence of archaeal-like ether lipids in some bacteria^{18,19}. Phylogenomic analyses, which so far have mostly focused on the evolution of glycerol phosphate dehydrogenases, isoprenoids and fatty acids, must also address the origin and evolution of the enzymes that catalyse the formation of links among these elements and the lipid modifications such as glycosylation. Advances in these areas will improve our understanding not only of cell

membrane evolution but also of the origin of life and the diversification of the three domains.

Jonathan Lombard, Purificación López-García and David Moreira are at the Unité d'Ecologie, Systématique et Evolution, Centre National de la Recherche Scientifique Unité Mixte de Recherche 8079, Université Paris-Sud, 91405 Orsay, France.

Correspondence to D.M.

e-mail: david.moreira@u-psud.fr

doi:10.1038/nrmicro2815

Published online 11 June 2012

- Virchow, R. *Die Cellularpathologie in ihrer Begründung auf physiologische und pathologische Gewebelehre.* (Hirschwald, 1858).
- Darwin, C. *The Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle For Life* (John Murray, 1859).
- Fitch, W. M. & Upper, K. The phylogeny of tRNA sequences provides evidence for ambiguity reduction in the origin of the genetic code. *Cold Spring Harb. Symp. Quant. Biol.* **52**, 759–767 (1987).
- Eididin, M. Lipids on the frontier: a century of cell-membrane bilayers. *Nature Rev. Mol. Cell Biol.* **4**, 414–418 (2003).
- Pereto, J. Controversies on the origin of life. *Int. Microbiol.* **8**, 23–31 (2005).
- Fry, I. The origins of research into the origins of life. *Endeavour* **30**, 24–28 (2006).
- Mitchell, P. Coupling of phosphorylation to electron and hydrogen transfer by a chemi-osmotic type of mechanism. *Nature* **191**, 144–148 (1961).
- Kates, M., Yengoyan, L. S. & Sastry, P. S. A. diether analog of phosphatidyl glycerophosphate in *Halobacterium cutirubrum*. *Biochim. Biophys. Acta* **98**, 252–268 (1965).
- Langworthy, T. A., Smith, P. F. & Mayberry, W. R. Lipids of *Thermoplasma acidophilum*. *J. Bacteriol.* **112**, 1193–1200 (1972).
- Woese, C. R., Magrum, L. J. & Fox, G. E. Archaeobacteria. *J. Mol. Evol.* **11**, 245–251 (1978).
- Koga, Y., Kyuragi, T., Nishihara, M. & Sone, N. Did archaeal and bacterial cells arise independently from noncellular precursors? A hypothesis stating that the advent of membrane phospholipid with enantiomeric glycerophosphate backbones caused the separation of the two lines of descent. *J. Mol. Evol.* **46**, 54–63 (1998).
- Koonin, E. V. Comparative genomics, minimal gene-sets and the last universal common ancestor. *Nature Rev. Microbiol.* **1**, 127–136 (2003).
- Szostak, J. W., Bartel, D. P. & Luisi, P. L. Synthesizing life. *Nature* **409**, 387–390 (2001).
- Martin, W. & Russell, M. J. On the origins of cells: a hypothesis for the evolutionary transitions from abiotic geochemistry to chemoautotrophic prokaryotes, and from prokaryotes to nucleated cells. *Phil. Trans. R. Soc. Lond. B* **358**, 59–83 (2003).
- Lowell, R. P., Rona, P. A. & Herzen, R. P. V. Seafloor hydrothermal systems. *J. Geophys. Res.* **100**, 327–352 (1995).
- de Duve, C. The onset of selection. *Nature* **433**, 581–582 (2005).
- Wächtershäuser, G. From pre-cells to Eukarya – a tale of two lipids. *Mol. Microbiol.* **47**, 13–22 (2003).
- Damste, J. S. *et al.* Structural characterization of diabolic acid-based tetraester, tetraether and mixed ether/ester, membrane-spanning lipids of bacteria from the order *Thermotogales*. *Arch. Microbiol.* **188**, 629–641 (2007).
- Weijers, J. W. *et al.* Membrane lipids of mesophilic anaerobic bacteria thriving in peats have typical archaeal traits. *Environ. Microbiol.* **8**, 648–657 (2006).
- Kates, M., Wassef, M. K. & Kushner, D. J. Radioisotopic studies on the biosynthesis of the glyceryl diether lipids of *Halobacterium cutirubrum*. *Can. J. Biochem.* **46**, 971–977 (1968).
- Langworthy, T. A., Mayberry, W. R. & Smith, P. F. Long-chain glycerol diether and polyol dialkyl glycerol triether lipids of *Sulfolobus acidocaldarius*. *J. Bacteriol.* **119**, 106–116 (1974).
- Gattinger, A., Schloter, M. & Munch, J. C. Phospholipid etherlipid and phospholipid fatty acid fingerprints in selected euryarchaeal monocultures for taxonomic profiling. *FEMS Microbiol. Lett.* **213**, 133–139 (2002).
- Lange, B. M., Rujan, T., Martin, W. & Croteau, R. Isoprenoid biosynthesis: the evolution of two ancient and distinct pathways across genomes. *Proc. Natl Acad. Sci. USA* **97**, 13172–13177 (2000).
- McGarvey, D. J. & Croteau, R. Terpenoid metabolism. *Plant Cell* **7**, 1015–1026 (1995).
- Pereto, J., Lopez-Garcia, P. & Moreira, D. Ancestral lipid biosynthesis and early membrane evolution. *Trends Biochem. Sci.* **29**, 469–477 (2004).
- Lombard, J. & Moreira, D. Origins and early evolution of the mevalonate pathway of isoprenoid biosynthesis in the three domains of life. *Mol. Biol. Evol.* **28**, 87–99 (2011).
- Lombard, J., López-García, P. & Moreira, D. An ACP-independent fatty acid synthesis pathway in archaea? Implications for the origin of phospholipids. *Mol. Biol. Evol.* (in the press).
- Lombard, J. & Moreira, D. Early evolution of the biotin-dependent carboxylase family. *BMC Evol. Biol.* **11**, 232 (2011).
- Soderberg, T., Chen, A. & Poulter, C. D. Geranylgeranyl glyceryl phosphate synthase. Characterization of the recombinant enzyme from *Methanobacterium thermoautotrophicum*. *Biochemistry* **40**, 14847–14854 (2001).
- Hemmi, H., Shibuya, K., Takahashi, Y., Nakayama, T. & Nishino, T. [S]-2,3-Di-O-geranylgeranyl glyceryl phosphate synthase from the thermoacidophilic archaeon *Sulfolobus solfataricus*. Molecular cloning and characterization of a membrane-intrinsic prenyltransferase involved in the biosynthesis of archaeal ether-linked membrane lipids. *J. Biol. Chem.* **279**, 50197–50203 (2004).
- Daiyasu, H. *et al.* A study of archaeal enzymes involved in polar lipid synthesis linking amino acid sequence information, genomic contexts and lipid composition. *Archaea* **1**, 399–410 (2005).
- Jekely, G. Did the last common ancestor have a biological membrane? *Biol. Direct* **1**, 35 (2006).
- Gogarten, J. P. *et al.* Evolution of the vacuolar H⁺-ATPase: implications for the origin of eukaryotes. *Proc. Natl Acad. Sci. USA* **86**, 6661–6665 (1989).
- Brochier-Armanet, C., Talla, E. & Gribaldo, S. The multiple evolutionary histories of dioxygen reductases: implications for the origin and evolution of aerobic respiration. *Mol. Biol. Evol.* **26**, 285–297 (2009).
- Baymann, F. *et al.* The redox protein construction kit: pre-last universal common ancestor evolution of energy-conserving enzymes. *Phil. Trans. R. Soc. Lond. B* **358**, 267–274 (2003).
- Lebrun, E. *et al.* The Rieske protein: a case study on the pitfalls of multiple sequence alignments and phylogenetic reconstruction. *Mol. Biol. Evol.* **23**, 1180–1191 (2006).
- Vignais, P. M., Billoud, B. & Meyer, J. Classification and phylogeny of hydrogenases. *FEMS Microbiol. Rev.* **25**, 455–501 (2001).
- Gribaldo, S. & Cammarano, P. The root of the universal tree of life inferred from anciently duplicated genes encoding components of the protein-targeting machinery. *J. Mol. Evol.* **47**, 508–516 (1998).
- Cao, T. B. & Saier, M. H. Jr. The general protein secretory pathway: phylogenetic analyses leading to evolutionary conclusions. *Biochim. Biophys. Acta* **1609**, 115–125 (2003).
- Yen, M. R., Harley, K. T., Tseng, Y. H. & Saier, M. H. Jr. Phylogenetic and structural analyses of the oxa1 family of protein translocases. *FEMS Microbiol. Lett.* **204**, 225–231 (2001).
- Yen, M. R., Tseng, Y. H., Nguyen, E. H., Wu, L. F. & Saier, M. H. Jr. Sequence and phylogenetic analyses of the twin-arginine targeting (Tat) protein export system. *Arch. Microbiol.* **177**, 441–450 (2002).
- Koonin, E. V. & Martin, W. On the origin of genomes and cells within inorganic compartments. *Trends Genet.* **21**, 647–654 (2005).
- Mulkidjanian, A. Y., Makarova, K. S., Galperin, M. Y. & Koonin, E. V. Inventing the dynamo machine: the evolution of the F-type and V-type ATPases. *Nature Rev. Microbiol.* **5**, 892–899 (2007).
- Shimada, H. & Yamagishi, A. Stability of heterochiral hybrid membrane made of bacterial sn-G3P lipids and archaeal sn-G1P lipids. *Biochemistry* **50**, 4114–4120 (2011).
- Budin, I. & Szostak, J. W. Physical effects underlying the transition from primitive to modern cell membranes. *Proc. Natl Acad. Sci. USA* **108**, 5249–5254 (2011).
- Koga, Y. Early evolution of membrane lipids: how did the lipid divide occur? *J. Mol. Evol.* **72**, 274–282 (2011).
- Boussau, B., Blanquart, S., Neccsulea, A., Lartillot, N. & Gouy, M. Parallel adaptations to high temperatures in the Archaean eon. *Nature* **456**, 942–945 (2008).
- Valentine, D. L. Adaptations to energy stress dictate the ecology and evolution of the Archaea. *Nature Rev. Microbiol.* **5**, 316–323 (2007).
- Taron, D. J., Childs, W. C., 3rd & Neuhaus, F. C. Biosynthesis of D-alanyl-lipoteichoic acid: role of diglyceride kinase in the synthesis of phosphatidylglycerol for chain elongation. *J. Bacteriol.* **154**, 1110–1116 (1983).
- Itabashi, Y. & Kuksis, A. Reassessment of stereochemical configuration of natural phosphatidylglycerols by chiral-phase high-performance liquid chromatography and electrospray mass spectrometry. *Anal. Biochem.* **254**, 49–56 (1997).
- Kolbe, M., Besir, H., Essen, L. O. & Oesterheld, D. Structure of the light-driven chloride pump halorhodopsin at 1.8 Å resolution. *Science* **288**, 1390–1396 (2000).
- Woese, C. R., Kandler, O. & Wheelis, M. L. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Natl Acad. Sci. USA* **87**, 4576–4579 (1990).
- Cavalier-Smith, T. Deep phylogeny, ancestral groups and the four ages of life. *Phil. Trans. R. Soc. B* **365**, 111–132 (2010).
- Junglas, B. *et al.* *Ignicoccus hospitalis* and *Nanoarchaeum equitans*: ultrastructure, cell-cell interaction, and 3D reconstruction from serial sections of freeze-substituted cells and by electron cryotomography. *Arch. Microbiol.* **190**, 395–408 (2008).
- Gray, M. W. & Doolittle, W. F. Has the endosymbiotic hypothesis been proven? *Nature Rev. Microbiol.* **4**, 1–42 (2006).
- Embley, T. M. & Martin, W. Eukaryotic evolution, changes and challenges. *Nature* **440**, 623–630 (2006).
- Poole, A. M. & Penny, D. Evaluating hypotheses for the origin of eukaryotes. *Bioessays* **29**, 74–84 (2007).
- Gribaldo, S., Poole, A. M., Daubin, V., Forterre, P. & Brochier-Armanet, C. The origin of eukaryotes and their relationship with the Archaea: are we at a phylogenetic impasse? *Nature Rev. Microbiol.* **8**, 743–752 (2010).
- Lopez-Garcia, P. & Moreira, D. Metabolic symbiosis at the origin of eukaryotes. *Trends Biochem. Sci.* **24**, 88–95 (1999).
- Rivera, M. C., Jain, R., Moore, J. E. & Lake, J. A. Genomic evidence for two functionally distinct gene classes. *Proc. Natl Acad. Sci. USA* **95**, 6239–6244 (1998).
- Ribeiro, S. & Golding, G. B. The mosaic nature of the eukaryotic nucleus. *Mol. Biol. Evol.* **15**, 779–788 (1998).
- Martin, W. & Muller, M. The hydrogen hypothesis for the first eukaryote. *Nature* **392**, 37–41 (1998).
- Martin, W. & Koonin, E. V. Introns and the origin of nucleus-cytosol compartmentalization. *Nature* **440**, 41–45 (2006).
- Moreira, D. & Lopez-Garcia, P. Symbiosis between methanogenic archaea and δ-proteobacteria as the origin of eukaryotes: the syntrophic hypothesis. *J. Mol. Evol.* **47**, 517–530 (1998).
- Lopez-Garcia, P. & Moreira, D. Selective forces for the origin of the eukaryotic nucleus. *Bioessays* **28**, 525–533 (2006).
- Horiike, T., Hamada, K., Kanaya, S. & Shinozawa, T. Origin of eukaryotic cell nuclei by symbiosis of Archaea in Bacteria is revealed by homology-hit analysis. *Nature Cell Biol.* **3**, 210–214 (2001).
- Mansy, S. S. *et al.* Template-directed synthesis of a genetic polymer in a model protocell. *Nature* **454**, 122–125 (2008).
- Edgell, D. R. & Doolittle, W. F. Archaea and the origin(s) of DNA replication proteins. *Cell* **89**, 995–998 (1997).
- Rothfield, L., Taghbalout, A. & Shih, Y. L. Spatial control of bacterial division-site placement. *Nature Rev. Microbiol.* **3**, 959–968 (2005).
- Mileykovskaya, E. & Dowhan, W. Role of membrane lipids in bacterial division-site selection. *Curr. Opin. Microbiol.* **8**, 135–142 (2005).
- Tornabene, T. G. *et al.* Phytanyl-glycerol ethers and squalenes in the archaeobacterium *Methanobacterium thermoautotrophicum*. *J. Mol. Evol.* **11**, 259–266 (1978).

72. Tornabene, T. G. & Langworthy, T. A. Diphytanyl and dibiphytanyl glycerol ether lipids of methanogenic archaeobacteria. *Science* **203**, 51–53 (1979).
73. Monnard, P. A. & Deamer, D. W. Membrane self-assembly processes: steps toward the first cellular life. *Anat. Rec.* **268**, 196–207 (2002).
74. Deamer, D. W. & Barchfeld, G. L. Encapsulation of macromolecules by lipid vesicles under simulated prebiotic conditions. *J. Mol. Evol.* **18**, 203–206 (1982).
75. Chen, I. A. & Szostak, J. W. Membrane growth can generate a transmembrane pH gradient in fatty acid vesicles. *Proc. Natl Acad. Sci. USA* **101**, 7965–7970 (2004).
76. Hanczyc, M. M. & Szostak, J. W. Replicating vesicles as models of primitive cell growth and division. *Curr. Opin. Chem. Biol.* **8**, 660–664 (2004).
77. Oparin, A. I. *The Origin of Life* (Macmillan, 1938).
78. Deamer, D. W. Boundary structures are formed by organic components of the Murchison carbonaceous chondrite. *Nature* **317**, 792–794 (1985).
79. Pohorille, A. & Deamer, D. Self-assembly and function of primitive cell membranes. *Res. Microbiol.* **160**, 449–456 (2009).
80. Mansy, S. S. & Szostak, J. W. Reconstructing the emergence of cellular life through the synthesis of model protocells. *Cold Spring Harb. Symp. Quant. Biol.* **74**, 47–54 (2009).
81. Segre, D., Ben-Eli, D., Deamer, D. W. & Lancet, D. The lipid world. *Orig. Life Evol. Biosph.* **31**, 119–145 (2001).
82. Kandler, O. Cell wall biochemistry in archaea and its phylogenetic implications. *J. Biol. Phys.* **20**, 165–169 (1994).
83. Mulkidjanian, A. Y., Galperin, M. Y., Makarova, K. S., Wolf, Y. I. & Koonin, E. V. Evolutionary primacy of sodium bioenergetics. *Biol. Direct* **3**, 13 (2008).

Acknowledgements

This work was supported by the interdisciplinary programmes Origine des Planètes et de la Vie and InTerrVie (Interactions Terre/Vie, both of the French Centre National de la Recherche Scientifique and the French Institut National des Sciences de l'Univers. J.L. is a recipient of a Ph.D. fellowship from the French Research Ministry. The authors thank the anonymous referees for useful comments.

Competing interests statement

The authors declare no competing financial interests.

FURTHER INFORMATION

David Moreira's homepage:

<http://www.ese.u-psud.fr/rubrique7.html?lang=en>

ALL LINKS ARE ACTIVE IN THE ONLINE PDF

ANNEXE 7.

Phylogenomic investigation of phospholipid synthesis in archaea.

Lombard, J., López-García P et D. Moreira. 2011.

Archaea: In press.

Phylogenomic Investigation of Phospholipid Synthesis in Archaea

Jonathan Lombard, Purificación López-García, and David Moreira

*Unité d'Ecologie Systématique et Evolution, CNRS UMR 8079, Université Paris-Sud,
91405 Orsay Cedex, France*

Correspondence should be addressed to David Moreira, david.moreira@u-psud.fr

Archaea have idiosyncratic cell membranes usually based on phospholipids containing glycerol-1-phosphate linked by ether bonds to isoprenoid lateral chains. Since these phospholipids strongly differ from those of bacteria and eukaryotes, the origin of the archaeal membranes (and by extension, of all cellular membranes) was enigmatic and called for accurate evolutionary studies. In this article we review some recent phylogenomic studies that have revealed a modified mevalonate pathway for the synthesis of isoprenoid precursors in archaea and suggested that this domain uses an atypical pathway of synthesis of fatty acids devoid of any acyl carrier protein, which is essential for this activity in bacteria and eukaryotes. In addition, we show new or updated phylogenetic analyses of enzymes likely responsible for the isoprenoid chain synthesis from their precursors and the phospholipid synthesis from glycerol phosphate, isoprenoids and polar head groups. These results support that most of these enzymes can be traced back to the last archaeal common ancestor and, in many cases, even to the last common ancestor of all living organisms.

1. Introduction

Archaea were identified as an independent domain of life in the late 1970s thanks to the characterization of differences between their ribosomal RNA molecules and those of bacteria and eukaryotes [1]. At that time, it was already known that the so-called "archaeobacteria" held several atypical biochemical characteristics, of which the most remarkable was their unusual membrane phospholipids. Whereas all bacteria and eukaryotes were known to have membranes based on fatty acids linked by ester bonds to glycerol-phosphate, archaea appeared to have phospholipids composed of isoprenoid chains condensed with glycerol-phosphate by ether linkages [2-6]. Moreover, the archaeal glycerol ethers contained *sn*-glycerol-1-phosphate (G1P), whereas bacterial and eukaryotic ones contained *sn*-glycerol-3-phosphate (G3P) (for review, see [7, 8]). These differences, which have been at the center of an intense debate about the nature of the first cell membranes [9], have progressively been shown to be not so sharp: ether-linked lipids are common both in eukaryotes (up to 25% of the total lipids in certain animal cells, [10]) and in several thermophilic bacteria [11-13]; fatty acid phospholipids have been found in diverse archaea [14]; conversely, isoprenoids are known to be universal although they are synthesized by non-homologous pathways in the three domains of life, [15]; even the stereochemistry of the glycerol phosphate has some exceptions, as shown by the recent discovery of archaeal-like *sn*-glycerol-1-phosphate specific lipids in some bacteria [16] and in eukaryotic endosomes [17].

Today, thanks to the significant accumulation of complete genome sequence data for a wide variety of species, it is possible to study the major pathways of phospholipid synthesis and their exceptions by looking at the presence or absence of the corresponding genes in the different genomes. In addition, these data allows reconstructing the evolutionary history of each gene by combining comparative genomics and phylogenetics, namely, by using a phylogenomic approach. In this article, we summarize the results concerning the phylogenomic studies of the pathways of synthesis of the different phospholipid components in archaea and provide some new data about the enzymes described in Koga and Morii [18] and Matsumi et al. [19] as likely responsible for the assembly of archaeal lipids from these building blocks.

2. Archaea Possess an Atypical Pathway of Biosynthesis of Isoprenoid Precursors

Isoprenoids are chains of isoprene units and their derivatives and are found ubiquitously in all living beings. They are involved in very diverse functions, such as photosynthetic pigments, hormones, quinones acting in electron transport chains, plant defense compounds, etc. [15, 20]. In archaea, isoprenoids also make up the hydrophobic lateral chains of phospholipids [7]. Isoprenoid biosynthesis requires two isoprene activated precursors, called isopentenyl pyrophosphate (IPP) and dimethylallyl diphosphate (DMAPP), which act as building parts. The first metabolic pathway responsible for the biosynthesis of isoprenoid precursors was discovered in yeasts and animals in the 1950s and named mevalonate (MVA) pathway in reference to its first committed precursor. Since its first description, the MVA pathway was thought to synthesize the isoprenoid precursors in all organisms [20], but closer scrutiny showed its absence in most bacteria [21]. An elegant series of multidisciplinary approaches allowed in the late 1990s and early 2000s the description of an independent and non-homologous pathway in bacteria, the methylerythritol phosphate (MEP) pathway (see [22] for review). The MEP pathway was found to be widespread in bacteria and plastid-bearing eukaryotes, whereas the MVA pathway was assumed to operate in archaea and eukaryotes [15, 23]. In the early 2000's, archaeal isoprenoids were thought to be synthesized through the MVA pathway because the ¹⁴C-labelled intermediates of this pathway were shown to be incorporated into the archaeal phospholipids [24]. However, only two of the enzymes of the eukaryotic pathway have been described in archaea [25, 26]. Moreover, the search of MVA pathway enzyme sequences in archaeal genomes revealed that from the seven enzymes acting on this pathway, most archaeal species lack the last three of them: phosphomevalonate kinase (PMK), mevalonate diphosphate decarboxylase (MDC), and isopentenyl diphosphate isomerase (IDI1), involved in the conversion of phosphomevalonate into IPP and DMAPP [27]. Attempts to characterize these missing archaeal enzymes revealed two new enzymes: the isopentenyl phosphate kinase (IPK) and an alternative isopentenyl diphosphate isomerase (IDI2) [28-30]. Although the decarboxylation reaction required to synthesize isopentenyl phosphate from phosphomevalonate has not been described yet, an alternative final MVA pathway involving IPK and IDI2 was then proposed to exist in archaea (Figure 1, [29]).

Taking advantage of the recent accumulation of genomic data, we recently carried out a thorough phylogenomic analysis of the pathways of isoprenoid precursor synthesis [31]. Concerning archaea, our presence-and-absence analysis extended the presence of the proposed archaeal alternative pathway to a wide archaeal diversity. The main exception to these observations was *Nanoarchaeum equitans*, what explains the dependence of this organism to obtain its lipids from its crenarchaeal host, *Ignicoccus hospitalis* [32]. Phylogenetic reconstructions of the MVA pathway enzymes present in archaea systematically revealed a clade independent from the other domains of life, representative of the archaeal diversity and generally respecting the main archaeal taxonomic groups [31]. These results are consistent with the presence of a distinctive archaeal MVA pathway in the last archaeal common ancestor (LACA) and support that different isoprenoid precursor biosynthesis pathways are characteristic of each domain of life: the classical MVA pathway in eukaryotes, the alternative MVA pathway in archaea and the MEP pathway in bacteria.

The origin of the archaeal MVA pathway appears to be composite. The enzymes shared with the classical eukaryotic MVA pathway were also ancestral in eukaryotes and bacteria and, therefore, they can be inferred to have been present in the last common ancestor of all living organisms (the cenancestor), from which the archaeal lineage would have inherited them. The last enzymes of the eukaryotic MVA pathway were most likely present in the respective ancestors of eukaryotes and bacteria, but their distribution in archaea is scarce and complex: homologues of MDC can be detected in Haloarchaea and Thermoplasmatales and some IDI1 genes have been identified in Haloarchaea and Thaumarchaeota, although they most likely reflect recent horizontal gene transfer (HGT) events from bacterial donors, as deduced from their position well nested among bacterial sequences in phylogenetic analyses [23, 31]. In contrast, the class Sulfolobales contains PMK and MDC homologues which robustly branch in an intermediate position between the eukaryotic and the bacterial sequences, suggesting that these sequences could be ancestral versions that would have been lost in the rest of archaea [31]. If this is the case, a eukaryotic-like pathway (except for the IDI function, which remains ambiguous) can be proposed to have existed in the cenancestor. This pathway would have been replaced by the MEP pathway in the bacterial lineage, whereas in archaea only the last steps were

replaced by non-homologous enzymes. We do not know the evolutionary forces that drove these changes; however, the mevalonate kinase (MVK), PMK and MDC belong to a large family of kinases, namely the GHMP kinases [33], which is characterized by a high structural and mechanistic conservation that contrasts with the large range of substrates that they can use [34]. If we assume that ancestors of these enzymes were not very specific, this could have allowed some tolerance to recruitment of evolutionary unrelated enzymes not only in archaea but also in some eukaryotes that have replaced their ancestral PMK by a non-homologous one [35, 36]. In agreement with this idea, the archaeal IPK itself appears to be able to use different substrates [37], what could have made the replacement easier.

3. Early Evolution of Isoprenoid Chain Synthesis

Once the isoprenoid precursors, IPP and DMAPP, have been synthesized, they still have to be assembled to make isoprenoid chains. The prenyltransferases responsible of this function are the isoprenyl diphosphate synthases (IPPS). Since many more complete genome sequences are now available than at the time of previous IPPS phylogenetic analyses, we present here an updated survey of the evolution of these enzymes. Although many different enzymes are required to synthesize the wide diversity of isoprenoids, the first steps are widely shared among the three domains of life [38]. They consist of the progressive addition of IPP units (5 carbons each) to an elongating allyl polyisoprenoid diphosphate molecule. Starting with DMAPP (5 carbons), consecutive condensation reactions produce geranylgeranyl diphosphate (GPP, 10 carbons), farnesyl diphosphate (FPP, 15 carbons), geranylgeranyl diphosphate (GGPP, 20 carbons), etc. Different IPPSs are characterized by the allylic substrate that they accept (DMAPP, GPP, FPP, ...) and by the stereochemistry of the double bonds, but the main characteristic used to classify them is the size of the products that they synthesize: they can be short-chain IPPS (~up to 25 carbons) or long-chain IPPS (>25 carbons). All IPPS are homologous and share a common reaction mechanism. Short-chain IPPS mainly differ in the size of their substrate-binding hydrophobic pocket (the smaller the pocket, the shorter the final product, [38]), but point mutations have been shown to importantly impact the size of the pocket and, thus, of their final products [39, 40]. Description of equivalent natural

mutations have been reported in archaea [41], what argues against the possibility of inferring the precise product of a given IPPS only based on its phylogenetic position [41, 42].

The first IPPS phylogenetic trees published [42] only used 13 sequences (of which 12 were short-chain enzymes) and supported a split between prokaryotic and eukaryotic enzymes. At that time, archaeal IPPS were assumed to be more ancient because they were known to provide isoprenoids for several pathways while bacteria and eukaryotes were thought to have more specialized enzymes. Later phylogenetic analyses with more sequences showed two groups corresponding to the functional split between short and long-chain enzymes [41]. This was expected since the long-chain IPPS use supplementary proteins to stabilize the hydrophobic substrate [38], what would logically impact the protein structure and therefore trigger the large divergence between the short- and the long-chain enzymes. In that work, the short-chain enzymes formed three groups according to the three domains of life, what could be considered as evidence for the presence of one short-chain enzyme in the cenancestor and its subsequent inheritance in modern lineages, including archaea. Furthermore, in that phylogeny archaea did not branch as a basal lineage, disproving their previously assumed ancient character. In addition, enzymes with different product specificities branched mixed all over the tree, supporting the product plasticity of the ancestral short-chain enzyme and the subsequent evolution of particular specificities according to biological requirements in modern organisms [41]. Archaeal long-chain IPPS had not been detected in that study, but some of them were incorporated in phylogenetic studies some years later [23].

We have searched for homologues of IPPS in a representative set of 348 complete genomes from the three domains of life (including 88 archaea). A preliminary phylogenetic analysis showed that, in agreement with previous reports [23], the resulting sequences split into two clades mainly related to the short- or long-chain product specificity (data not shown). To avoid the phylogenetic artifacts that can be introduced by the high sequence divergence between short- and long-chain enzymes, we carried out independent phylogenetic analyses for each one of the two paralogues, to which we will refer as short- or long-chain IPPS with regard to dominant functions of the characterized enzymes. However, as previously mentioned [39-41], substrate specificity exchange

between short and long substrates appears to be relatively common, so these partial trees must be acknowledged mainly as phylogenetic groups related to the most widespread biochemical function, but do not definitely determine the substrate specificity of all their sequences, which can only be established by biochemical studies. In preliminary short-chain IPPS phylogenies, long branches at the base of several eukaryotic paralogues evidenced for the high divergence of these sequences. Since we were here mainly interested in the archaeal sequences, we removed the divergent eukaryotic sequences from our analyses. Figure 2 and Supplementary Figure 1 present the phylogeny of some representative prokaryotic short-chain enzyme sequences. Most archaeal sequences branch together in a monophyletic group largely congruent with the main archaeal phyla and orders, what suggests that this enzyme was present in LACA and was vertically inherited in most archaeal lineages. Most bacterial sequences also cluster together according to the main bacterial taxonomic groups, suggesting the ancestral presence of this enzyme in the last bacterial common ancestor and, given its presence also in LACA, most likely also in the cenancestor. However, some recent HGTs can also be pointed out from this phylogeny. Concerning archaea, Thermoproteales, some Desulfurococcales, some Thermoplasmatales, *Methanocella paludicola* and *Aciduliprofundum boonei* branch within the bacterial group, so several HGTs can be invoked to explain this pattern, although in most cases the support is weak and does not allow confidently determining the identity of the bacterial donors (Figure 2).

As in the case of the short-chain enzymes, most archaeal sequences also group together in our long-chain IPPS phylogenies and, despite the weakly supported paraphyly of the euryarchaeotal sequences, the main archaeal taxonomic groups are observed (Figure 3 and Supplementary Figure 2). Bacterial sequences also group together according to main taxa, supporting that the respective common ancestors of bacteria and archaea, and thus likely also the cenancestor, had a long-chain IPPS enzyme that was inherited in modern organisms. Contrary to short-chain IPPS, eukaryotic long-chain IPPS are less divergent, so they were conserved in our analyses. Surprisingly, all the eukaryotic sequences, including those from plastid-lacking eukaryotes, branch together as the sister group of cyanobacteria, except for some algae that branch within the cyanobacterial group. In addition, some other more recent HGTs are observed, especially in the archaeal

Thermococcus genus, which branches close to a very divergent sequence from the delta-proteobacterium *Bdellovibrio bacteriovorus*.

Altogether, these results support that in spite of some recent HGTs from bacteria to archaea, most archaea have homologues of both the short- and the long-chain IPPS that were inherited from LACA and that, most likely, were already present in the cenancestor.

4. Archaea May Have an ACP-independent Fatty Acid Biosynthesis Pathway

In bacteria and eukaryotes, fatty acids are components of membrane phospholipids, energy-storage molecules, substrates for post-translational modifications of proteins, secondary metabolites and components of coenzymes and messenger compounds. Despite the general assumption that archaeal lipid metabolism is based on isoprenoids, a variety of experimental approaches have shown that fatty acids (FA) also exist in these organisms. Archaea show variable concentrations of free FA or their derivatives [4, 24, 43-45], what has stimulated some attempts to biochemically characterize an archaeal FA synthase [46, 47]. Archaeal FA can participate in protein structure [48-50] and acylation [47] but they have also been found as components of membrane phospholipids in very diverse euryarchaeotes [14]. Therefore, it is not surprising that homologues of several of the enzymes involved FA biosynthesis in bacteria had been detected in archaeal genomes [8, 51]. In bacteria, FA biosynthesis occurs through multiple condensations of acyl groups (malonyl-CoA) in a series of cyclical steps. The building blocks necessary to the activity of the FA synthases are provided by several reactions. First, the acetyl-CoA carboxylase (ACC) converts acetyl-coenzyme A (CoA) into malonyl-CoA. Second, the peptide cofactor acyl carrier protein (ACP), which is required to channel the elongating intermediates among the FA synthase enzymes, has to be activated through the addition of a phosphopantetheine group from CoA to the apo-ACP by an ACP synthase. Finally, the malonyl-CoA:ACP transacylase (MCAT) charges the malonyl-CoA to holo-ACP, resulting in malonyl-ACP.

Since no archaeal FA synthase system has been described in detail yet, we recently studied the evolution of the archaeal homologues of the bacterial genes involved in FA synthesis. First, many archaea possess ACC homologs that are closely related in phylogenetic analyses, suggesting their possible monophyletic origin and, consequently,

the presence of this enzyme in LACA [52, 53]. Second, only a few unrelated archaeal species have ACP and the ACP-processing machinery (ACP synthase and MCAT) and phylogenetic analysis suggests that they acquired them by HGTs from bacteria. As a result, ACP and its related enzymes are missing in most archaea and appear not to have been present in LACA [54]. For the rest of enzymes involved in the cyclic steps of FA synthesis (see Figure 1), phylogenetic analyses support that the archaeal sequences are more closely related to bacterial enzymes that are active on substrates linked to CoA than to enzymes that use substrates linked to ACP. Taking this into account, we have proposed that the ACP processing system has specifically evolved in the bacterial lineage and that archaea carry out FA synthesis using a likely ancient ACP-independent pathway [54].

Interestingly, it has been unclear for a long time if either the bacterial FA synthesis pathway specifically recognizes each acyl-ACP intermediate or if its enzymes randomly fix these intermediates, modifying the correct ones and releasing the inappropriate ones. Recent work has shown that ACP carries out its channeling function by adopting unique conformations for each enzyme of the FA elongation cycle [55]. Yet, it can be reasonably assumed that a putative ACP-independent mechanism would rather use random interactions between intermediate metabolites and enzymes, what might be assumed to be less efficient than the bacterial ACP-mediated system. Although this remains speculative and needs biochemical confirmation, the high efficiency of the ACP-mediated machinery may explain the preeminence of FA in bacterial membranes, whereas archaea would have opted by the alternative ancestral mechanism of synthesis of lateral chains for membrane phospholipids, namely, the isoprenoids [31], relegating FA to different cellular functions and, only to a small extent, to membrane synthesis [54].

5. Linking Glycerol Phosphate with the Lateral Chains

Although archaeal phospholipids were already known to use G1P instead of G3P as bacteria and eukaryotes do, the recent characterization and sequencing of the enzyme responsible for its synthesis in archaea (G1P dehydrogenase [56, 57]) was astonishing because this enzyme appeared to be totally unrelated to the canonical G3P dehydrogenase [58]. Since then, very little exceptions to this clear-cut distinction between archaea and bacteria have been described (the most remarkable is probably the description of an

archaeal-like G1P dehydrogenase in the bacterium *Bacillus subtilis*, [59]), but the fact that the two dehydrogenases belong to large enzymatic superfamilies from which they were recruited has provided a model for the independent origin of these dehydrogenases from likely ancient promiscuous enzymes [8].

Geranylgeranylglyceryl diphosphate and di-O-geranylgeranylglyceryl phosphate synthases (GGGPS and DGGGPS, respectively) are the enzymes that subsequently link isoprenoids to glycerol phosphate in archaea [18]. Boucher et al. [23] found that GGGPS are widespread in archaea whereas, among bacteria, it was found only in Bacillales and in one Bacteroidetes species. They also described a very divergent homolog in halophilic archaea and in *Archaeoglobus*. Despite their divergence, these enzymes appear to carry out the same function [60]. Using an updated genome sequence database, we also retrieve the very divergent GGGPS in Methanomicrobiales, confirm that GGGPS is present in Bacillales and extend this observation to a surprising diversity of Bacteroidetes (Supplementary Figure 3). This suggests that GGGPS plays an important role in these two bacterial groups and, indeed, it has recently been shown to participate in the synthesis of archaeal-type lipids of unknown function in these bacteria [16]. In conclusion, at least one GGGPS gene appears to have been present in LACA, whereas the divergent copy of this gene probably emerged later in a more recent group of euryarchaeota. This gene was independently transferred to the respective ancestors of Bacillales and Bacteroidetes.

Hemmi et al. [61] carried out the first phylogenetic analysis of the DGGGPS sequences and their superfamily, the UbiA prenyltransferase family (17 sequences branching in 6 different groups). Using all sequences available at present, we retrieve similar but much more diversified groups (Supplementary Figure 4). Archaeal sequences cluster together in a monophyletic assemblage, suggesting common ancestry, although crenarchaeota are paraphyletic probably as a result of reconstruction artifacts. A number of bacteroidetes sequences branch among the crenarchaeota, reflecting an HGT event to an ancestor of this bacterial group. In addition several bacterial phyla (Chlorobi, Cyanobacteria, Chloroflexi, Planctomycetales and some proteobacterial species) possess UbiA-related homologs. However, in several of these bacterial species, these enzymes are involved in photosynthesis (they take part in the synthesis of respiratory quinones, hemes,

chlorophylls, and vitamin E [61]) and it is difficult to determine if the widespread distribution in bacteria is due to the ancestral presence of a homolog of this enzyme in the last common bacterial ancestor or to several recent HGTs to these bacterial groups from archaeal donors.

The search of homologues of the bacterial glycerol phosphate acyl transferases PlsX, PlsY, PlsB and PlsC, which carry out the addition of fatty acids to glycerol phosphate [62, 63], in the available archaeal genome sequences allowed retrieving very few homologs, all of them most likely acquired by HGT from bacterial donors (not shown).

6. Linking the Polar Head Groups

Once the two hydrocarbon chains are linked to the phospholipid backbone, two additional steps are required to add the polar head group (Figure 1). First, an enzyme replaces the phosphate linked to the glycerol moiety by a CDP group; then, this CDP is replaced by the final polar head group, which can be very diverse (glycerol, ethanolamine, serine, etc) [18]. The first step has been biochemically described in archaea but the enzyme that carries out this function remains unknown [64]. The bacterial counterpart is the CDP diglyceride synthetase (CdsA) [65, 66]. We looked for archaeal homologs of the bacterial enzyme and retrieved several sequences that we used as queries for exhaustive searches in archaeal genomes. This allowed us to observe that this enzyme is widespread both in bacteria and archaea. Our phylogenetic trees show two clades that correspond to archaea and bacteria and the phylogenetic relationships within each of them are congruent with the main accepted taxonomic groups (Figure 4). This supports the vertical inheritance of this gene from the cenancestor and, consequently, its presence in LACA. However, to our knowledge no archaeal CdsA has been biochemically characterized so far, which would be required to confirm that these archaeal homologs are responsible for the biochemical activity described by Morii et al. [64].

Two enzymes involved in the second step of the attachment of polar head groups have been characterized in archaea [67, 68]. These enzymes belong to the large family of the CDP alcohol phosphatidyltransferases and have homologues in bacteria [69]. The bacterial members of this enzyme family are known to have different specificities in order to add different polar head groups on phospholipids. The phylogenomic survey

carried out by Daiyasu et al. [69] showed that the different sequences group in phylogenetic trees according to their predicted substrates. In addition, the groups of bacterial sequences were in agreement with the main bacterial taxonomic groups. Our analysis with a larger taxonomic sample confirms that homologs of these genes are widespread in archaea. All sequences group into three main categories (data not shown), one related to the addition of serine as a polar head group, another related to inositol-phosphate transfer (and maybe also glycerol in archaea, according to classification in [69]) and the last one using glycerol. Archaea are found in the first two groups but not in the last one. In order to avoid artifacts due to extreme sequence divergence, we carried out phylogenies for each of the two groups containing archaeal representatives. In the first phylogeny, archaeal sequences predicted to use serine [67] as a substrate are limited to Euryarchaeota (Supplementary Figure 5). This tree shows a poorly supported group of slow evolving bacteria together with several divergent bacterial, eukaryotic and archaeal (*Methanococcoides burtonii* and haloarchaea) sequences. Such mixed distribution dominated by strong sequence divergence and HGTs contrasts with the rest of the phylogeny, which is congruent with the main accepted taxonomic groups. Especially, one monophyletic group of euryarchaeota can be pointed out, supporting that this enzyme probably existed at least in the last euryarchaeotal common ancestor. In the second tree, a larger group of archaeal enzymes, predicted to use glycerol phosphate or myo-inositol-phosphate as substrates [68, 69], cluster in our analysis with bacterial enzymes that use myo-inositol, but these bacterial sequences are scarce and appear to have been acquired from archaea by HGT (Supplementary Figure 6). The phylogeny of archaeal sequences supports the monophyly of Crenarchaeota and a patchy distribution of several paralogues in Euryarchaeota, what makes difficult the analysis of the evolution of these enzymes in archaea without supplementary biochemical information. At any rate, the wide distribution of this enzyme family in archaea strongly suggests that at least one representative of these enzymes was already present in LACA. Whereas it was conserved in Crenarchaeota, it was subjected to several duplication and neofunctionalization events in Euryarchaeota, which would explain the complex distribution pattern observed in euryarchaeotal species.

7. Saturation of Isoprenoid Chains

So far, we have described the main synthesis and link mechanisms of archaeal phospholipid components, but a wide diversity of phospholipids actually exists in archaea [70]. A substantial characteristic of archaeal membranes is their saturation rate, since double bonds have a prominent influence on membrane stability [71]. Most archaea contain saturated phospholipids and the geranylgeranyl reductase (GGR), the enzyme responsible for the reduction of isoprenoid chains, has been recently described in the euryarchaeota *Thermoplasma acidophilum* and *Archaeoglobus fulgidus* and the crenarchaeote *Sulfolobus acidocaldarius* [72-74]. These studies show that archaeal GGR are able to use geranylgeranyl chains either isolated or attached to phospholipids as substrates. When GGPP is used as a substrate, all bonds but the one in position C2 are reduced, allowing the incorporation in phospholipids, but isoprenoids chains that are already attached to glycerol phosphate can have all their double bonds reduced by this enzyme. As a result, isoprenoid reduction can happen at different steps of the phospholipid biosynthesis pathway (Figure 1).

Archaeal GGRs are homologous to previously reported cyanobacterial and plant GGRs involved in chlorophyll synthesis [75, 76] and many different GGR paralogues have been identified in archaeal genomes that could carry out independent isoprenoid chain reductions [19]. Our phylogeny of GGRs confirms that GGRs are widespread among crenarchaeota and that at least two paralogues exist in a wide diversity of euryarchaeota, although some archaeal groups also bear supplementary GGR genes (Supplementary Figure 7). This supports the ancestral presence of at least one GGR gene in archaea, followed by a complex history of duplications and HGTs. GGRs are also present in many bacterial genomes, but several HGTs can be observed, probably related to the function of this gene in photosynthesis, making uncertain the primary origin of this gene in bacteria. Among eukaryotes, only sequences from plastid-bearing organisms were detected and these sequences clearly branched within one group of cyanobacterial sequences, thus supporting the plastidial origin of these genes in eukaryotes.

8. Discussion

In contrast with the rather impressive knowledge about the biochemistry and biosynthesis of bacterial cell membranes, many aspects of their archaeal counterparts remain to be elucidated. Even the synthesis of the most canonical archaeal membrane lipids, the phospholipids based on isoprenoid lateral chains, still has some unresolved points, such as the identity of the enzyme carrying out the phosphomevalonate decarboxylase activity necessary for the final steps of the synthesis of isoprenoid precursors. Moreover, it is becoming clear that archaeal membranes incorporate components that were supposed to be restricted to the other two domains of life, bacteria and eukaryotes. This is notably the case of fatty acids that, despite being relatively widespread in membrane phospholipids of euryarchaeotal species [14], are synthesized by a pathway that, most likely, has important differences with the bacterial counterpart [54].

Comparative genomics and phylogenomics provide a powerful way to address these questions, in particular by the detection of potential candidates to carry out missing enzymatic functions thanks to similarities with bacterial and eukaryotic enzymes. This has allowed us to propose, for example, the existence of an ACP-independent pathway of fatty acid synthesis in archaea [54] and allows us to propose here that archaeal CdsA homologues could carry the same function as in bacteria. This approach has also provided evidence supporting that lipid membranes were already evolved long ago, at the time of the cenancestor [9]. Thus, rather than radical inventions of new phospholipid biochemistries, bacteria and archaea appear to have specialized their cell membranes by tuning the relative importance of the different components, with isoprenoids becoming dominant in archaea and fatty acids in bacteria. Nevertheless, these bioinformatic approaches have limitations and biochemical investigation remains crucial to characterize the different missing activities (the uncharacterized MVA pathway enzymes in archaea, the hypothetical ACP-independent FA synthesis pathway, the characterization of the archaeal CdsA and a larger diversity of CDP alcohol phosphatidyltransferases) in order to complete the puzzle of archaeal membrane synthesis.

9. Material and Methods

Sequence seeds for similarity searches were retrieved from the KEGG database (<http://www.genome.jp/kegg>). Searches were carried out with BLASTp [77] against a list of completely sequenced genomes available in GenBank (Supplementary Table 1). The resulting sequences were aligned with by default parameters with Muscle 3.6 [78]. Redundant and partial sequences were removed and ambiguously aligned regions were discarded prior to phylogenetic analyses using the NET program from the MUST package [79]. Phylogenetic trees were reconstructed with the approximately maximum likelihood approach with FastTree 2.1.3 [80].

Acknowledgements

We thank Angela Corcelli for the invitation to contribute this article. This work was supported by the interdisciplinary programs Origine des Planètes et de la Vie and InTerrVie (Interactions Terre-Vie) of the French Centre National de la Recherche Scientifique and Institut National des Sciences de l'Univers and by French National Agency for Research (EVOLDEEP project, contract number ANR-08-GENM-024-002). J.L. is recipient of a PhD fellowship of the French Research Ministry.

References

- [1] C.R. Woese and G.E. Fox, "Phylogenetic structure of the prokaryotic domain: the primary kingdoms," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 74, no. 11, pp. 5088-5090, 1977.
- [2] M. Kates, L.S. Yengoyan and P.S. Sastry, "A Diether Analog of Phosphatidyl Glycerophosphate in *Halobacterium Cutirubrum*," *Biochimica et biophysica acta*, vol. 98, pp. 252-268, 1965.
- [3] T.A. Langworthy, P.F. Smith and W.R. Mayberry, "Lipids of *Thermoplasma acidophilum*," *Journal of bacteriology*, vol. 112, no. 3, pp. 1193-1200, 1972.
- [4] T.A. Langworthy, W.R. Mayberry and P.F. Smith, "Long-chain glycerol diether and polyol dialkyl glycerol triether lipids of *Sulfolobus acidocaldarius*," *Journal of bacteriology*, vol. 119, no. 1, pp. 106-116, 1974.
- [5] C.R. Woese, L.J. Magrum and G.E. Fox, "Archaeobacteria," *Journal of molecular evolution*, vol. 11, no. 3, pp. 245-251, 1978.

- [6] T.G. Tornabene and T.A. Langworthy, "Diphytanyl and dibiphytanyl glycerol ether lipids of methanogenic archaeobacteria," *Science (New York, N.Y)*, vol. 203, no. 4375, pp. 51-53, 1979.
- [7] M. De Rosa, A. Gambacorta and A. Gliozzi, "Structure, biosynthesis, and physicochemical properties of archaeobacterial lipids," *Microbiological reviews*, vol. 50, no. 1, pp. 70-80, 1986.
- [8] J. Peretó, P. López-García and D. Moreira, "Ancestral lipid biosynthesis and early membrane evolution," *Trends in biochemical sciences*, vol. 29, no. 9, pp. 469-477, 2004.
- [9] J. Lombard, P. López-García and D. Moreira, "The early evolution of lipid membranes and the three domains of life," *Nature reviews microbiology*, vol. 10, no. 7, pp. 507-515, 2012.
- [10] H.K. Mangold and F. Paltauf, *Ether Lipids: Biochemical and Biomedical Aspects*, Academic Press, 1983.
- [11] R. Huber, T. Wilharm, D. Huber, A. Trincone, S. Burggraf, H. König, R. Rachel, I. Rockinger, H. Fricke and K.O. Stetter, "*Aquifex pyrophilus* gen. nov. sp. nov., represents a novel group of marine hyperthermophilic hydrogen-oxidizing bacteria," *Systematic and applied microbiology*, vol. 15, pp. 340-351, 1992.
- [12] J.S. Sinninghe Damsté, W.I. Rijpstra, E.C. Hopmans, S. Schouten, M. Balk and A.J. Stams, "Structural characterization of diabolic acid-based tetraester, tetraether and mixed ether/ester, membrane-spanning lipids of bacteria from the order Thermotogales," *Archives of microbiology*, vol. 188, no. 6, pp. 629-641, 2007.
- [13] T.A. Langworthy, G. Holzer, J.G. Zeikus and T.G. Tornabene, "Iso- and anteiso-branched glycerol diethers of the thermophilic anaerobe *Thermodesulfotobacterium commune*," *Systematic and applied microbiology*, vol. 4, pp. 1-17, 1983.
- [14] A. Gattinger, M. Schloter and J.C. Munch, "Phospholipid etherlipid and phospholipid fatty acid fingerprints in selected euryarchaeotal monocultures for taxonomic profiling," *FEMS microbiology letters*, vol. 213, no. 1, pp. 133-139, 2002.
- [15] B.M. Lange, T. Rujan, W. Martin and R. Croteau, "Isoprenoid biosynthesis: the evolution of two ancient and distinct pathways across genomes," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 97, no. 24, pp. 13172-13177, 2000.
- [16] H. Guldan, F.M. Matysik, M. Bocola, R. Sterner and P. Babinger, "Functional Assignment of an Enzyme that Catalyzes the Synthesis of an Archaea-Type Ether Lipid in Bacteria," *Angewandte Chemie (International ed)*, vol. 50, no. 35, pp. 8188-8191, 2011.
- [17] H.H. Tan, A. Makino, K. Sudesh, P. Greimel and T. Kobayashi, "Spectroscopic evidence for the unusual stereochemical configuration of an endosome-specific lipid," *Angewandte Chemie (International ed)*, vol. 51, no. 2, pp. 533-535, 2012.
- [18] Y. Koga and H. Morii, "Biosynthesis of ether-type polar lipids in archaea and evolutionary considerations," *Microbiology and molecular biology reviews*, vol. 71, no. 1, pp. 97-120, 2007.

- [19] R. Matsumi, H. Atomi, A.J. Driessen and J. van der Oost, "Isoprenoid biosynthesis in Archaea--biochemical and evolutionary implications," *Research in microbiology*, vol. 162, no. 1, pp. 39-52, 2011.
- [20] D.J. McGarvey and R. Croteau, "Terpenoid metabolism," *The Plant cell*, vol. 7, no. 7, pp. 1015-1026, 1995.
- [21] D. Zhou and R.H. White, "Early steps of isoprenoid biosynthesis in *Escherichia coli*," *The Biochemical journal*, vol. 273 (Pt 3), pp. 627-634, 1991.
- [22] M. Rodríguez-Concepción and A. Boronat, "Elucidation of the methylerythritol phosphate pathway for isoprenoid biosynthesis in bacteria and plastids. A metabolic milestone achieved through genomics," *Plant physiology*, vol. 130, no. 3, pp. 1079-1089, 2002.
- [23] Y. Boucher, M. Kamekura and W.F. Doolittle, "Origins and evolution of isoprenoid lipid biosynthesis in archaea," *Molecular Microbiology*, vol. 52, pp. 515-527, 2004.
- [24] M. Kates, M.K. Wassef and D.J. Kushner, "Radioisotopic studies on the biosynthesis of the glyceryl diether lipids of *Halobacterium cutirubrum*," *Canadian journal of biochemistry*, vol. 46, no. 8, pp. 971-977, 1968.
- [25] W.L. Lam and W.F. Doolittle, "Mevinolin-resistant mutations identify a promoter and the gene for a eukaryote-like 3-hydroxy-3-methylglutaryl-coenzyme A reductase in the archaebacterium *Haloferax volcanii*," *The Journal of biological chemistry*, vol. 267, no. 9, pp. 5829-5834, 1992.
- [26] D. Yang, L.W. Shipman, C.A. Roessner, A.I. Scott and J.C. Sacchettini, "Structure of the *Methanococcus jannaschii* mevalonate kinase, a member of the GHMP kinase superfamily," *The Journal of biological chemistry*, vol. 277, no. 11, pp. 9462-9467, 2002.
- [27] A. Smit and A. Mushegian, "Biosynthesis of isoprenoids via mevalonate in Archaea: the lost pathway," *Genome research*, vol. 10, no. 10, pp. 1468-1484, 2000.
- [28] S.J. Barkley, R.M. Cornish and C.D. Poulter, "Identification of an Archaeal type II isopentenyl diphosphate isomerase in *Methanothermobacter thermoautotrophicus*," *Journal of bacteriology*, vol. 186, no. 6, pp. 1811-1817, 2004.
- [29] L.L. Grochowski, H. Xu and R.H. White, "*Methanocaldococcus jannaschii* uses a modified mevalonate pathway for biosynthesis of isopentenyl diphosphate," *Journal of bacteriology*, vol. 188, no. 9, pp. 3192-3198, 2006.
- [30] R. Dutoit, J. de Ruyck, V. Durisotti, C. Legrain, E. Jacobs and J. Wouters, "Overexpression, physicochemical characterization, and modeling of a hyperthermophilic *Pyrococcus furiosus* type 2 IPP isomerase," *Proteins*, vol. 71, no. 4, pp. 1699-1707, 2008.
- [31] J. Lombard and D. Moreira, "Origins and early evolution of the mevalonate pathway of isoprenoid biosynthesis in the three domains of life," *Molecular biology and evolution*, vol. 28, no. 1, pp. 87-99, 2011.
- [32] U. Jahn, R. Summons, H. Sturt, E. Grosjean and H. Huber, "Composition of the lipids of *Nanoarchaeum equitans* and their origin from its host *Ignicoccus* sp. strain KIN4/I," *Archives of microbiology*, vol. 182, no. 5, pp. 404-413, 2004.

- [33] P. Bork, C. Sander and A. Valencia, "Convergent evolution of similar enzymatic function on different protein folds: the hexokinase, ribokinase, and galactokinase families of sugar kinases," *Protein Science*, vol. 2, no. 1, pp. 31-40, 1993.
- [34] J.L. Andreassi, 2nd and T.S. Leyh, "Molecular functions of conserved aspects of the GHMP kinase family," *Biochemistry*, vol. 43, no. 46, pp. 14594-14601, 2004.
- [35] S.M. Houten and H.R. Waterham, "Nonorthologous gene displacement of phosphomevalonate kinase," *Molecular genetics and metabolism*, vol. 72, no. 3, pp. 273-276, 2001.
- [36] T.J. Herdendorf and H.M. Mizioro, "Phosphomevalonate kinase: functional investigation of the recombinant human enzyme," *Biochemistry*, vol. 45, no. 10, pp. 3235-3242, 2006.
- [37] M. Chen and C.D. Poulter, "Characterization of thermophilic archaeal isopentenyl phosphate kinases," *Biochemistry*, vol. 49, no. 1, pp. 207-217, 2010.
- [38] B.A. Kellogg and C.D. Poulter, "Chain elongation in the isoprenoid biosynthetic pathway," *Current opinion in chemical biology*, vol. 1, no. 4, pp. 570-578, 1997.
- [39] S. Ohnuma, K. Hirooka, H. Hemmi, C. Ishida, C. Ohto and T. Nishino, "Conversion of product specificity of archaeobacterial geranylgeranyl-diphosphate synthase. Identification of essential amino acid residues for chain length determination of prenyltransferase reaction," *The Journal of biological chemistry*, vol. 271, no. 31, pp. 18831-18837, 1996.
- [40] S. Ohnuma, K. Hirooka, N. Tsuruoka, M. Yano, C. Ohto, H. Nakane and T. Nishino, "A pathway where polyprenyl diphosphate elongates in prenyltransferase. Insight into a common mechanism of chain length determination of prenyltransferases," *The Journal of biological chemistry*, vol. 273, no. 41, pp. 26705-26713, 1998.
- [41] A. Tachibana, Y. Yano, S. Otani, N. Nomura, Y. Sako and M. Taniguchi, "Novel prenyltransferase gene encoding farnesylgeranyl diphosphate synthase from a hyperthermophilic archaeon, *Aeropyrum pernix*. Molecular evolution with alteration in product specificity," *European journal of biochemistry / FEBS*, vol. 267, no. 2, pp. 321-328, 2000.
- [42] A. Chen, P.A. Kroon and C.D. Poulter, "Isoprenyl diphosphate synthases: protein sequence comparisons, a phylogenetic tree, and predictions of secondary structure," *Protein Science*, vol. 3, no. 4, pp. 600-607, 1994.
- [43] T.G. Tornabene, R.S. Wolfe, W.E. Balch, G. Holzer, G.E. Fox and J. Oro, "Phytanyl-glycerol ethers and squalenes in the archaeobacterium *Methanobacterium thermoautotrophicum*," *Journal of molecular evolution*, vol. 11, no. 3, pp. 259-266, 1978.
- [44] M. Nishihara, S. Nagahama, M. Ohga and Y. Koga, "Straight-chain fatty alcohols in the hyperthermophilic archaeon *Pyrococcus furiosus*," *Extremophiles*, vol. 4, no. 5, pp. 275-277, 2000.
- [45] N.M. Carballeira, M. Reyes, A. Sostre, H. Huang, M.F. Verhagen and M.W. Adams, "Unusual fatty acid compositions of the hyperthermophilic archaeon *Pyrococcus furiosus* and the bacterium *Thermotoga maritima*," *Journal of bacteriology*, vol. 179, no. 8, pp. 2766-2768, 1997.

- [46] E.L. Pugh, M.K. Wassef and M. Kates, "Inhibition of fatty acid synthetase in *Halobacterium cutirubrum* and *Escherichia coli* by high salt concentrations," *Canadian journal of biochemistry*, vol. 49, no. 8, pp. 953-958, 1971.
- [47] E.L. Pugh and M. Kates, "Acylation of proteins of the archaeobacteria *Halobacterium cutirubrum* and *Methanobacterium thermoautotrophicum*," *Biochimica et biophysica acta*, vol. 1196, no. 1, pp. 38-44, 1994.
- [48] M. Colella, S. Lobasso, F. Babudri and A. Corcelli, "Palmitic acid is associated with halorhodopsin as a free fatty acid. Radiolabeling of halorhodopsin with 3H-palmitic acid and chemical analysis of the reaction products of purified halorhodopsin with thiols and NaBH₄," *Biochimica et biophysica acta*, vol. 1370, no. 2, pp. 273-279, 1998.
- [49] A. Corcelli, S. Lobasso, M. Colella, M. Trotta, A. Guerrieri and F. Palmisano, "Role of palmitic acid on the isolation and properties of halorhodopsin," *Biochimica et biophysica acta*, vol. 1281, no. 2, pp. 173-181, 1996.
- [50] M. Kolbe, H. Besir, L.O. Essen and D. Oesterhelt, "Structure of the light-driven chloride pump halorhodopsin at 1.8 Å resolution," *Science (New York, N.Y.)*, vol. 288, no. 5470, pp. 1390-1396, 2000.
- [51] V. Iverson, R.M. Morris, C.D. Frazar, C.T. Berthiaume, R.L. Morales and E.V. Armbrust, "Untangling genomes from metagenomes: revealing an uncultured class of marine Euryarchaeota," *Science (New York, N.Y.)*, vol. 335, no. 6068, pp. 587-590, 2012.
- [52] J. Lombard and D. Moreira, "Early evolution of the biotin-dependent carboxylase family," *BMC evolutionary biology*, vol. 11, pp. 232, 2011.
- [53] J. Lombard and D. Moreira, "Correction: Early evolution of the biotin-dependent carboxylase family," *BMC evolutionary biology*, vol. 12, no. 1, pp. 117, 2012.
- [54] J. Lombard, P. López-García and D. Moreira, "An ACP-independent fatty acid synthesis pathway in archaea: Implications for the Origin of Phospholipids," *Molecular biology and evolution*, doi: 10.1093/molbev/mss160, 2012.
- [55] E. Ploskon, C.J. Arthur, A.L. Kanari, P. Wattana-amorn, C. Williams, J. Crosby, T.J. Simpson, C.L. Willis and M.P. Crump, "Recognition of intermediate functionality by acyl carrier protein over a complete cycle of fatty acid biosynthesis," *Chemistry & biology*, vol. 17, no. 7, pp. 776-785, 2010.
- [56] M. Nishihara and Y. Koga, "Purification and properties of sn-glycerol-1-phosphate dehydrogenase from *Methanobacterium thermoautotrophicum*: characterization of the biosynthetic enzyme for the enantiomeric glycerophosphate backbone of ether polar lipids of Archaea," *Journal of biochemistry*, vol. 122, no. 3, pp. 572-576, 1997.
- [57] Y. Koga, T. Kyuragi, M. Nishihara and N. Sone, "Did archaeal and bacterial cells arise independently from noncellular precursors? A hypothesis stating that the advent of membrane phospholipid with enantiomeric glycerophosphate backbones caused the separation of the two lines of descent," *Journal of molecular evolution*, vol. 47, no. 5, pp. 631, 1998.
- [58] H. Daiyasu, T. Hiroike, Y. Koga and H. Toh, "Analysis of membrane stereochemistry with homology modeling of sn-glycerol-1-phosphate dehydrogenase," *Protein engineering*, vol. 15, no. 12, pp. 987-995, 2002.

- [59] H. Guldan, R. Sterner and P. Babinger, "Identification and characterization of a bacterial glycerol-1-phosphate dehydrogenase: Ni(2+)-dependent AraM from *Bacillus subtilis*," *Biochemistry*, vol. 47, no. 28, pp. 7376-7384, 2008.
- [60] J. Payandeh, M. Fujihashi, W. Gillon and E.F. Pai, "The crystal structure of (S)-3-O-geranylgeranyl glyceryl phosphate synthase reveals an ancient fold for an ancient enzyme," *The Journal of biological chemistry*, vol. 281, no. 9, pp. 6070-6078, 2006.
- [61] H. Hemmi, K. Shibuya, Y. Takahashi, T. Nakayama and T. Nishino, "(S)-2,3-Di-O-geranylgeranyl glyceryl phosphate synthase from the thermoacidophilic archaeon *Sulfolobus solfataricus*. Molecular cloning and characterization of a membrane-intrinsic prenyltransferase involved in the biosynthesis of archaeal ether-linked membrane lipids," *The Journal of biological chemistry*, vol. 279, no. 48, pp. 50197-50203, 2004.
- [62] M. Yoshimura, T. Oshima and N. Ogasawara, "Involvement of the YneS/YgiH and PlsX proteins in phospholipid biosynthesis in both *Bacillus subtilis* and *Escherichia coli*," *BMC microbiology*, vol. 7, pp. 69, 2007.
- [63] Y.J. Lu, Y.M. Zhang, K.D. Grimes, J. Qi, R.E. Lee and C.O. Rock, "Acyl-phosphates initiate membrane phospholipid synthesis in Gram-positive pathogens," *Molecular cell*, vol. 23, no. 5, pp. 765-772, 2006.
- [64] H. Morii, M. Nishihara and Y. Koga, "CTP:2,3-di-O-geranylgeranyl-sn-glycerol-1-phosphate cytidyltransferase in the methanogenic archaeon *Methanothermobacter thermoautotrophicus*," *The Journal of biological chemistry*, vol. 275, no. 47, pp. 36568-36574, 2000.
- [65] T. Icho, C.P. Sparrow and C.R. Raetz, "Molecular cloning and sequencing of the gene for CDP-diglyceride synthetase of *Escherichia coli*," *The Journal of biological chemistry*, vol. 260, no. 22, pp. 12078-12083, 1985.
- [66] C.P. Sparrow and C.R. Raetz, "Purification and properties of the membrane-bound CDP-diglyceride synthetase from *Escherichia coli*," *The Journal of biological chemistry*, vol. 260, no. 22, pp. 12084-12091, 1985.
- [67] H. Morii and Y. Koga, "CDP-2,3-Di-O-geranylgeranyl-sn-glycerol:L-serine O-archaetidyltransferase (archaetidylserine synthase) in the methanogenic archaeon *Methanothermobacter thermoautotrophicus*," *Journal of bacteriology*, vol. 185, no. 4, pp. 1181-1189, 2003.
- [68] H. Morii, S. Kiyonari, Y. Ishino and Y. Koga, "A novel biosynthetic pathway of archaetidyl-myo-inositol via archaetidyl-myo-inositol phosphate from CDP-archaeol and D-glucose 6-phosphate in methanoarchaeon *Methanothermobacter thermoautotrophicus* cells," *The Journal of biological chemistry*, vol. 284, no. 45, pp. 30766-30774, 2009.
- [69] H. Daiyasu, K. Kuma, T. Yokoi, H. Morii, Y. Koga and H. Toh, "A study of archaeal enzymes involved in polar lipid synthesis linking amino acid sequence information, genomic contexts and lipid composition," *Archaea (Vancouver, B.C)*, vol. 1, no. 6, pp. 399-410, 2005.
- [70] Y. Koga and H. Morii, "Recent advances in structural research on ether lipids from archaea including comparative and physiological aspects," *Bioscience, biotechnology, and biochemistry*, vol. 69, no. 11, pp. 2019-2034, 2005.

- [71] O. Dannenmuller, K. Arakawa, T. Eguchi, K. Kakinuma, S. Blanc, A.M. Albrecht, M. Schmutz, Y. Nakatani and G. Ourisson, "Membrane properties of archaeal macrocyclic diether phospholipids," *Chemistry (Weinheim an der Bergstrasse, Germany)*, vol. 6, no. 4, pp. 645-654, 2000.
- [72] Y. Nishimura and T. Eguchi, "Biosynthesis of archaeal membrane lipids: digeranylgeranyl glycerophospholipid reductase of the thermoacidophilic archaeon *Thermoplasma acidophilum*," *Journal of biochemistry*, vol. 139, no. 6, pp. 1073-1081, 2006.
- [73] M. Murakami, K. Shibuya, T. Nakayama, T. Nishino, T. Yoshimura and H. Hemmi, "Geranylgeranyl reductase involved in the biosynthesis of archaeal membrane lipids in the hyperthermophilic archaeon *Archaeoglobus fulgidus*," *The FEBS journal*, vol. 274, no. 3, pp. 805-814, 2007.
- [74] S. Sato, M. Murakami, T. Yoshimura and H. Hemmi, "Specific partial reduction of geranylgeranyl diphosphate by an enzyme from the thermoacidophilic archaeon *Sulfolobus acidocaldarius* yields a reactive prenyl donor, not a dead-end product," *Journal of bacteriology*, vol. 190, no. 11, pp. 3923-3929, 2008.
- [75] H.A. Addlesee, L.C. Gibson, P.E. Jensen and C.N. Hunter, "Cloning, sequencing and functional assignment of the chlorophyll biosynthesis gene, *chlP*, of *Synechocystis* sp. PCC 6803," *FEBS letters*, vol. 389, no. 2, pp. 126-130, 1996.
- [76] R. Tanaka, U. Oster, E. Kruse, W. Rudiger and B. Grimm, "Reduced activity of geranylgeranyl reductase leads to loss of chlorophyll and tocopherol and to partially geranylgeranylated chlorophyll in transgenic tobacco plants expressing antisense RNA for geranylgeranyl reductase," *Plant physiology*, vol. 120, no. 3, pp. 695-704, 1999.
- [77] S.F. Altschul, W. Gish, W. Miller, E.W. Myers and D.J. Lipman, "Basic local alignment search tool," *Journal of molecular biology*, vol. 215, no. 3, pp. 403-410, 1990.
- [78] R.C. Edgar, "MUSCLE: multiple sequence alignment with high accuracy and high throughput," *Nucleic acids research*, vol. 32, no. 5, pp. 1792-1797, 2004.
- [79] H. Philippe, "MUST, a computer package of Management Utilities for Sequences and Trees," *Nucleic acids research*, vol. 21, no. 22, pp. 5264-5272, 1993.
- [80] M.N. Price, P.S. Dehal and A.P. Arkin, "FastTree 2--approximately maximum-likelihood trees for large alignments," *PLoS one*, vol. 5, no. 3, pp. e9490, 2010.

Figure Legends

FIGURE 1: Biosynthesis pathways of phospholipid components in archaea. Abbreviations for the archaeal mevalonate (MVA) pathway: AACT, acetoacetyl-CoA thiolase; HMGS, 3-hydroxy-3-methylglutaryl-CoA synthase; HMGR, 3-hydroxy-3-methylglutaryl-CoA reductase; MVK, mevalonate kinase; IPK, isopentenyl phosphatase kinase; IDI2, isopentenyl diphosphate isomerase type II. GHMP, galactokinase-homoserine kinase-mevalonate kinase-phosphomevalonate kinase; IPPS, isoprenyl diphosphate synthases (asterisks indicate enzymes shared with the eukaryotic MVA pathway). Abbreviations for the

hypothetical archaeal fatty acid (FA) synthesis pathway: ACC, acetyl-CoA carboxylase; PCC, propionyl-CoA carboxylase; KR-PhaB, beta-ketoacyl reductase; DH-MaoC/PhaJ, beta-hydroxyacyl dehydratase; ER, enoyl reductase; SDR, short-chain dehydrogenases/reductases. Abbreviations for the *sn*-glycerol-1-phosphate synthesis pathway: G1PDH, glycerol-1-phosphate dehydrogenase; DHQS, 3-dehydroquinone synthase; GDH, glycerol dehydrogenase; ADH, alcohol dehydrogenase. Abbreviations for the phospholipid assembly pathway: GGGPS, (S)-3-O-geranylgeranyl glyceryl phosphate synthase; DGGGPS, (S)-2,3-di-O-geranylgeranyl glyceryl phosphate synthase; GGR, geranylgeranyl reductase; CDSA, CDP diglyceride synthetase. Names between parentheses indicate the family or superfamily to which belong the archaeal enzymes postulated to carry out particular functions on the basis of phylogenomic analyses.

FIGURE 2: Short-chain IPPS phylogenetic tree reconstructed using 136 representative sequences and 244 conserved sites. Multifurcations correspond to branches with support values <0.50. Triangles correspond to well supported-bacterial clades (numbers in parentheses correspond to the number of sequences included in these clades). For the complete phylogeny, see Supplementary Figure 1.

FIGURE 3: Long-chain IPPS phylogenetic tree reconstructed using 218 representative sequences and 241 conserved sites. Multifurcations correspond to branches with support values <0.50. Triangles correspond to well-supported clades outside Archaea (numbers in parentheses correspond to the number of sequences included in these clades). For the complete phylogeny, see Supplementary Figure 2.

FIGURE 4: CdsA phylogenetic tree reconstructed using 133 representative sequences and 87 conserved sites. Branches with support values <0.50 have been collapsed.

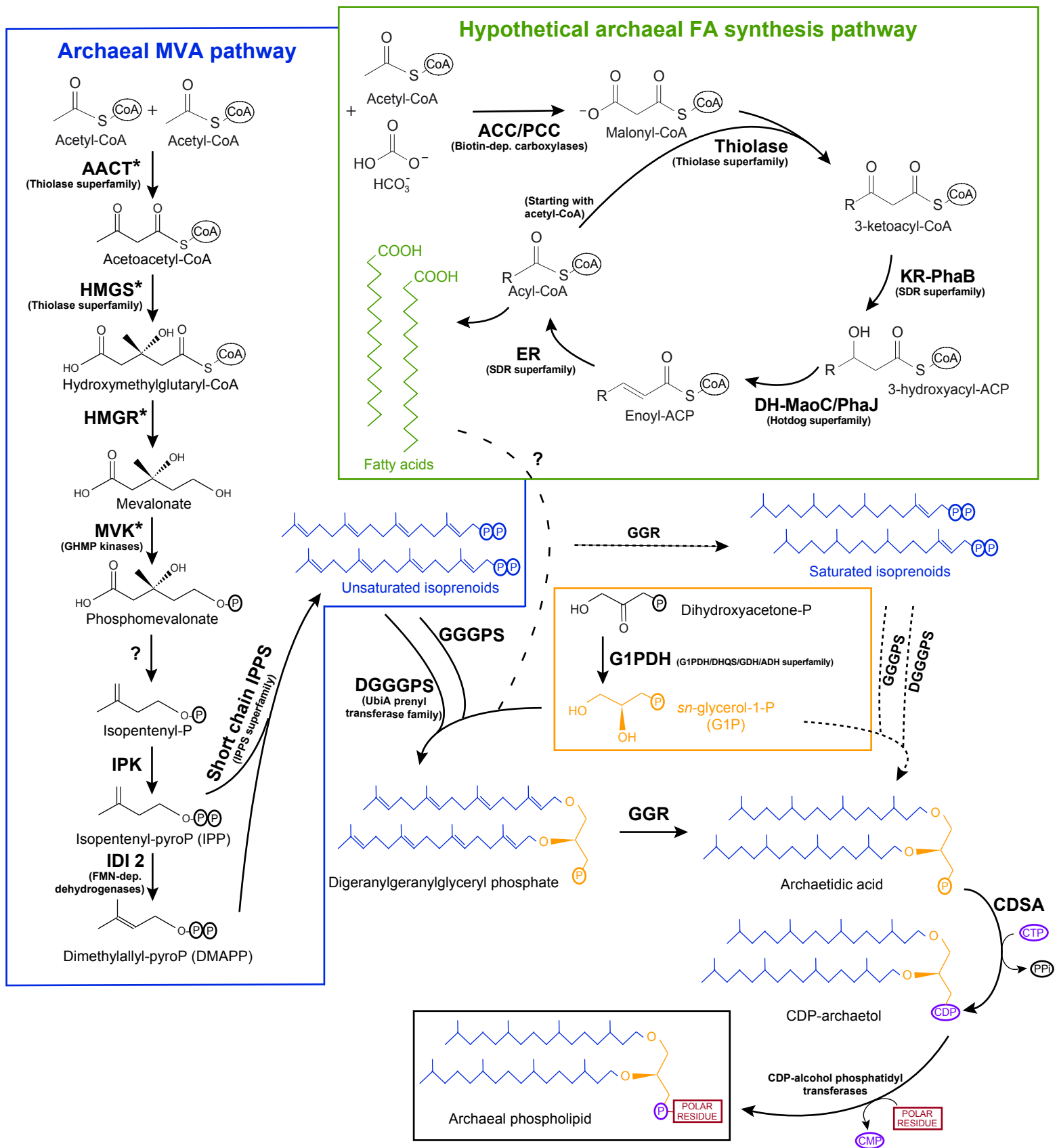


Figure 1. Lombard et al.

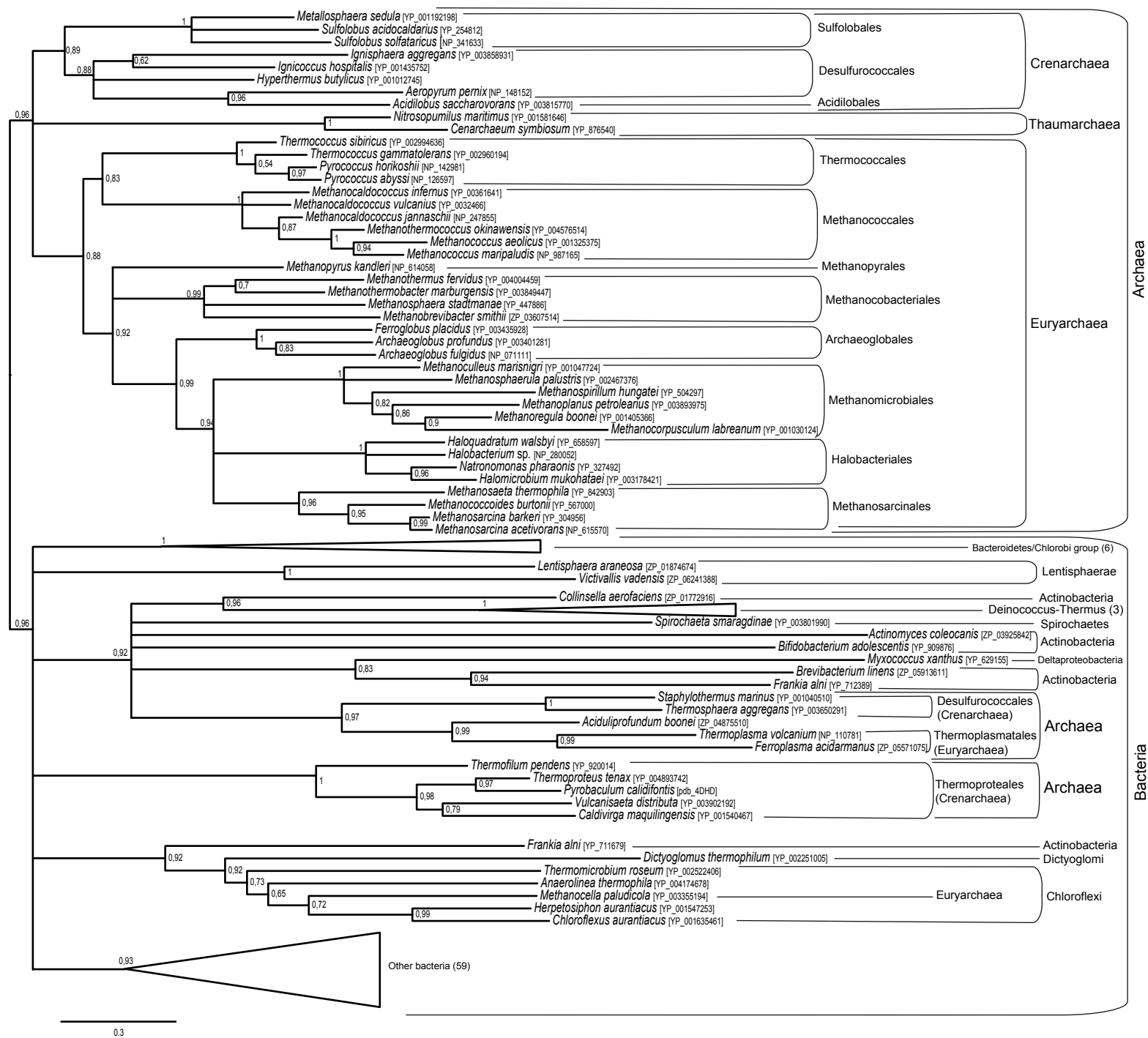


Figure 2. Lombard et al.

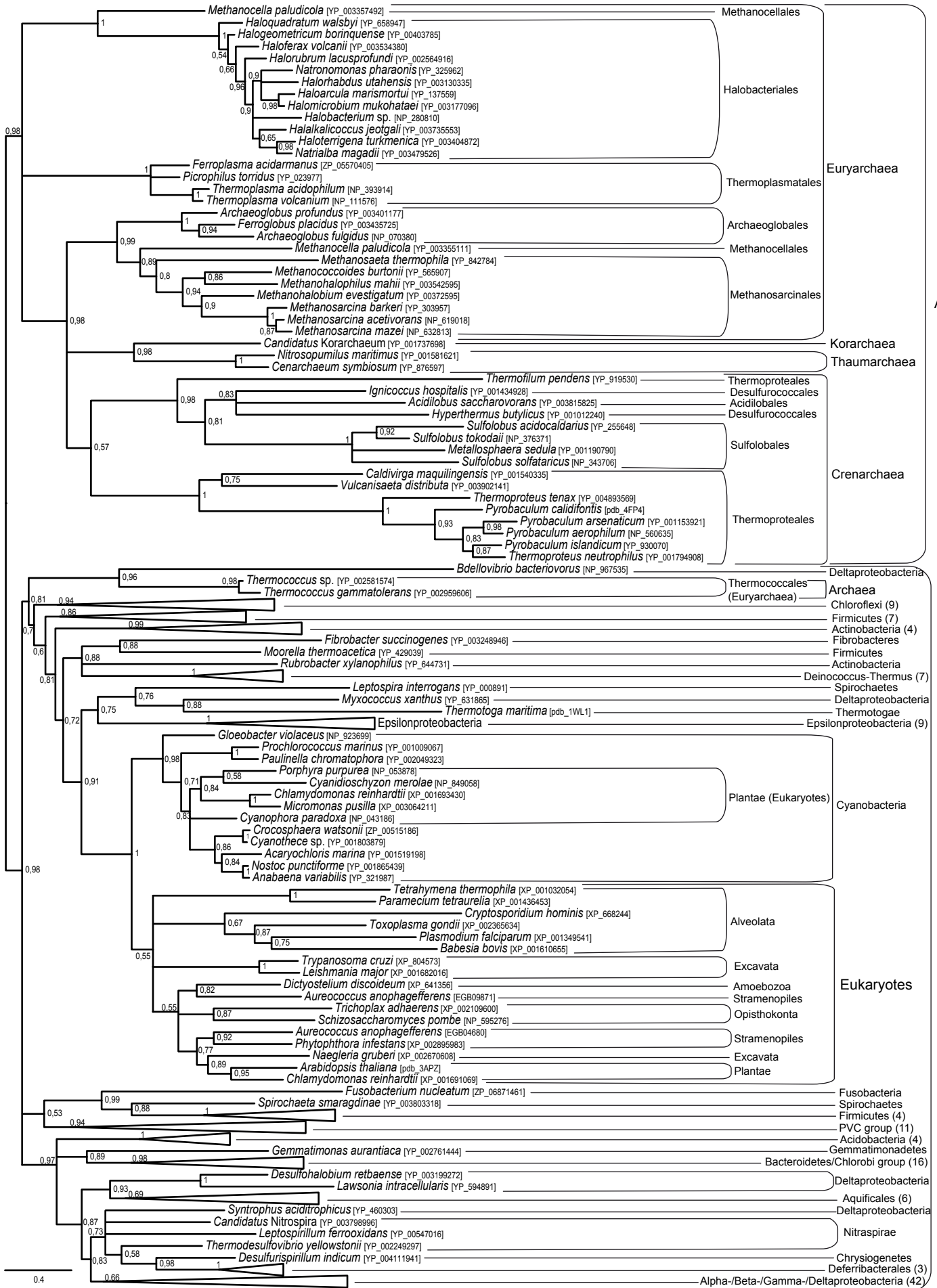


Figure 3. Lombard et al.

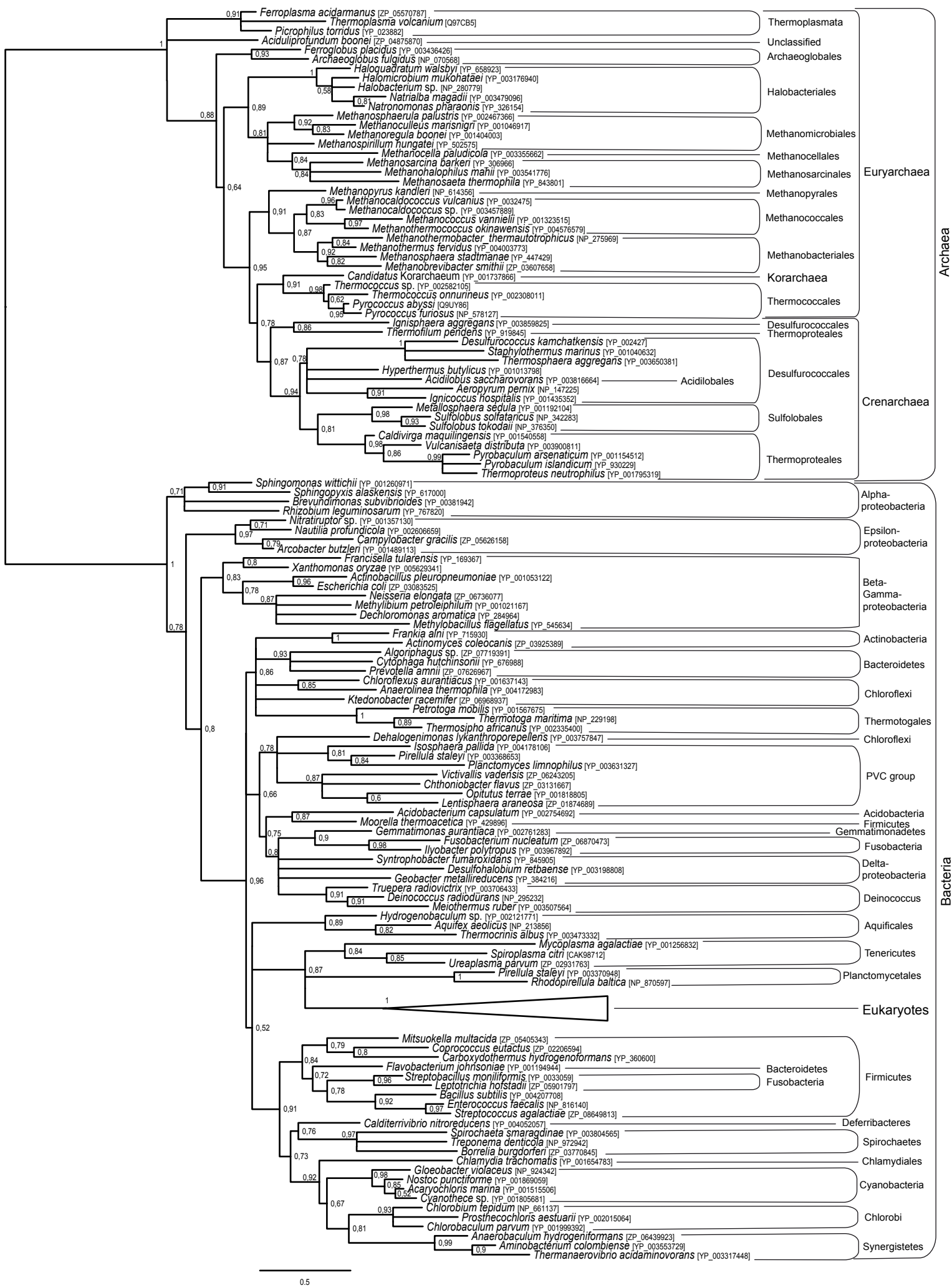


Figure 4. Lombard et al.