

## Origine et Evolution de microorganismes

Céline Brochier-Armanet

Coordonnées :

Laboratoire de Chimie Bactérienne (LCB – CNRS UPR9043)

Institut de Biologie Structurale et de Microbiologie, 31 chemin Joseph Aiguier, 13402 Marseille Cedex 20, France.

mail: celine.brochier-armanet@univ-lyon1.fr

### Résumé :

La publication de l'arbre universel du vivant par Woese à la fin des années soixante-dix a bouleversé notre vision du monde vivant par l'abolition de la dichotomie eucaryote/procaryote au profit d'une vision tripartite (Bacteria, Archaea, Eucarya) ou encore par la prise de conscience de la place importante occupée par les microorganismes. Elle a aussi ouvert un champ de recherche centré autour du dernier ancêtre commun à tous les organismes, appelé LUCA (*Last Universal Common Ancestor*). Les études réalisées au cours des vingt dernières années ont révélé que LUCA était un organisme évolué, avec une organisation cellulaire et un génome probablement proche de celui des microorganismes actuels. Cependant un certain nombre de questions fondamentales demeurent, concernant par exemple la nature de ses lipides, du support de son matériel génétique (ADN ou ARN), le milieu dans lequel il vivait (hyperthermophile ou mésophile), la nature de ses descendants directs (posant par la même la question de l'origine des eucaryotes) Ces questions font toujours l'objet de vives discussions, témoignant du dynamisme de ce champ de recherches.

**Texte :**

### **Découverte du microscope**

L'histoire des micro-organismes en biologie commence avec la fabrication du microscope au milieu du 17<sup>ième</sup> siècle (voir pour une histoire détaillée des classifications des microorganismes (Brochier, 2002)). Malgré le perfectionnement des techniques de microscopie et la multiplication des observations, pendant deux siècles, le statut des micro-organismes reste flou, comme le reflète la quasi-absence de débat autour de leur position taxonomique. La plupart des biologistes les interprètent comme des intermédiaires entre le monde minéral et le monde vivant trouvant probablement leur origine par génération spontanée. Les autres se contentent trop souvent de suivre le dogme, stipulant que tout être vivant est soit un animal, soit une plante. Suivant cette logique, les micro-organismes photosynthétiques, sont classés avec les plantes alors que les autres sont intégrés au sein des animaux dans le groupe mal défini des « vers ».

### **Les précurseurs & la naissance de la microbiologie**

Au 19<sup>ième</sup> siècle, la connaissance des microorganismes connaît de grandes avancées et s'accompagne d'une multiplication du nombre d'espèces décrites. Leur rôle en tant qu'agents infectieux, est mis en évidence et leur implication au niveau de processus biochimiques primordiaux est démontrée. Cependant, leur place au sein des êtres vivants reste peu débattue. Tout change en 1866, lorsqu'Haeckel propose de briser la dichotomie Plantae/Animalia en proposant la création d'un troisième règne du vivant, qu'il nomme Protista. Il regroupe sous ce nom tous les organismes unicellulaires, à l'exception des algues bleues-vertes, qu'il maintient au sein des Plantae. A l'intérieur des Protista, Haeckel rassemble plusieurs groupes d'eucaryotes unicellulaires ainsi que le groupe des Monera (organismes unicellulaires dépourvus de noyau). Cette définition des Monera, préfigure de celle que Chatton donnera des Procaryotes cinquante ans plus tard. Peu de temps après Cohn établit un lien entre les Monera et les algues bleues-vertes. En 1904, Haeckel rejoint le point de vue de Cohn et les reclasse au sein des Monera.

### **Les premières classifications bactériennes**

Les premières classifications bactériennes que l'on peut réellement assimiler à des phylogénies sont élaborées au début du 20<sup>ième</sup> siècle. Elles reposent sur un postulat totalement nouveau, selon lequel l'évolution des organismes est principalement conditionnée par celle de l'environnement terrestre. Ceci implique que la connaissance des caractéristiques de l'environnement terrestre au cours du temps peut fournir des indices sur les types d'organismes qui ont pu y vivre. Par exemple, Orla-Jensen suppose que l'émergence de la vie s'est faite dans un milieu dépourvu de matière organique, ce qui implique que les premières formes de vie devaient être autotrophes. Par une succession de

raisonnements basés sur la complexité des nutriments nécessaires au développement des bactéries ainsi que sur les caractéristiques de leur environnement naturel, Orla-Jensen élabore le premier système de classification se voulant naturel. Ces hypothèses seront challengées par celles d'Oparin et d'Horowitz qui proposent une apparition de la vie dans un environnement déjà riche en matière organique. Malheureusement, la plupart des multiples tentatives pour élaborer une classification bactérienne se soldent par des échecs. La raison est que l'extrême simplicité structurale et l'absence de développement qui caractérisent les micro-organismes limitent l'application du principe de subordination des caractères dont l'application chez les animaux et les plantes a permis le développement de classifications évolutives élaborées.

### **Procaryotes & Eucaryotes**

En marge des problèmes de taxonomie, les progrès réalisés en microscopie ont permis des observations de plus en plus fines de la structure des cellules. Chatton publie en 1937 que « *Le caractère distinctif des bactéries et des algues bleues vient de la nature Procaryote de leur cellule.* ». Pour lui, il existe une dichotomie fondamentale du monde vivant entre organismes possédant des cellules nucléées (animaux, plantes et protistes), et organismes à cellules anucléées (algues bleues-vertes et bactéries), qu'il nomme respectivement Eucaryotes et Procaryotes. La connotation des termes Procaryotes (pro = avant ; caryon = noyau) et Eucaryotes (eu = vrai ; caryon = noyau) illustre l'idée d'une évolution dirigée dans le sens d'une complexification croissante, d'une cellule procaryote, plus simple, vers une cellule eucaryote. Cette vision gradiste de la construction cellulaire au cours de l'évolution va marquer très profondément les esprits et surtout définit comme groupe naturel un ensemble d'organismes sur la base d'un caractère négatif qu'est l'absence de noyau. Pourtant, comme le soulignera Copeland, l'absence de noyau chez les procaryotes ne suffit pas à en faire un groupe ayant une signification évolutive. Inversement, la probabilité que le noyau soit apparu indépendamment chez plusieurs organismes est très faible en raison de sa complexité structurale ; ainsi, les eucaryotes pourraient constituer un groupe naturel. Malgré l'importance fondamentale de cette distinction, les termes de Procaryotes et Eucaryotes vont sombrer dans l'oubli jusqu'à leur exhumation par Stanier et Van Niel en 1962 (Stanier and Van Niel, 1962). Leur popularisation amènera la disparition du terme de Monera.

Il est important de rappeler que ces débats se placent au cœur de la polémique autour de l'origine unique. Malgré la démonstration effectuée par Pasteur en 1861 prouvant que les micro-organismes ne naissent pas par génération spontanée, l'hypothèse d'une origine multiple de la vie n'est toujours pas écartée. Il faudra attendre la seconde moitié du 20<sup>ème</sup> pour que cette idée s'impose, notamment grâce aux travaux de biologie moléculaire qui démontreront l'unité fondamentale du monde vivant par la mise en évidence de caractères fondamentaux partagés par l'ensemble des êtres vivants, telles que (i) l'utilisation de l'ADN comme support de l'information génétique ; (ii) le

mécanisme de réplication de l'ADN semi- conservatif ; (iii) la transcription de l'information génétique via un ARN messager puis sa traduction en protéine par le ribosome ; (iv) l'existence d'un code génétique « universel » ; (v) l'utilisation d'énergie libre stockée sous forme d'ATP et de matériaux bruts nécessaires à l'activité cellulaire et (vi) la délimitation spatiale par une double membrane composée de phospholipides. La présence de ces caractères universels implique que tous les êtres vivants descendent d'un même ancêtre et rend accessible le vieux rêve de Darwin, c'est-à-dire l'établissement d'une phylogénie universelle du monde vivant. Il faudra cependant attendre la fin des années soixante-dix et la révolution apportée par les phylogénies moléculaires pour sa réalisation.

### **L'arbre universel du vivant**

Les prémices de cette révolution se font sentir dans les années cinquante lorsque Sanger développe une méthode de séquençage des protéines et séquence l'insuline. Au cours des années suivantes, d'autres séquences sont obtenues. Progressivement, l'idée que les séquences de protéines puissent contenir une information phylogénétique de qualité se propage. En 1965, Zuckerkandl et Pauling publient leur article « Molecules as Documents of Evolutionary History » qui marque la transition vers l'ère des phylogénies moléculaires. Malgré ces prémices, la révolution apportée par les phylogénies moléculaires ne démarre réellement qu'après la publication en 1977 par Carl Woese et Georges Fox du premier arbre universel du vivant basé sur la comparaison de ARNr de la petite sous-unité du ribosome (ARNrS) de différents organismes (Woese, 2007). Ce choix n'est pas fortuit. En effet l'ARNrS est une molécule universelle de grande taille pouvant fournir un grand nombre caractères comparables entre eucaryotes et procaryotes, sa vitesse d'évolution est suffisamment faible pour avoir conservé la trace d'événements très anciens, comme la séparation des trois domaines. Enfin, son abondance dans les cellules facilite son utilisation au niveau expérimental.

L'arbre universel de Woese bouleverse la vision du monde vivant de l'époque (Adoutte et al., 1996) car il montre que le monde vivant n'est pas divisé en cellules eucaryotes et cellules procaryotes mais en trois domaines distincts dont deux sont de types procaryotes. Il nomme le premier Bacteria qui peut lui-même être subdivisé en dix lignées majeures (phyla). La diversité phénotypique rencontrée à l'intérieur de ces phyla confirme les conclusions de non-validité des caractères physiologiques ou morphologiques, pour l'établissement d'une phylogénie. Seuls quelques groupes, comme les Spirochètes ou les Cyanobactéries, sont confirmés. Le second domaine procaryote, qui regroupe les bactéries méthanogènes, est baptisé Archaea car Woese considère leur phénotype compatible avec les conditions atmosphériques supposées de la terre primitive. Pendant longtemps, et encore actuellement, Archaea a été synonyme d'habitants des milieux extrêmes. Or leur abondance dans tous les environnements s'est révélée être comparable à celle des bactéries et il a été mis en

évidence récemment qu'elles pourraient jouer un rôle prépondérant dans les cycles de l'azote et du carbone les océans et les sols (Nicol and Schleper, 2006).

La publication de l'arbre de Woese va provoquer une modification profonde de la vision de la diversité du monde vivant. En effet, les classifications élaborées au 20<sup>ème</sup> siècle, comme celle publiée par Whittaker, accordent une place centrale aux multicellulaires en raison de leur grande variété de tailles et de formes. Cette classification, figurant encore dans certains manuels de classe, est organisée suivant trois niveaux hiérarchiques illustrant une vision gradiste de l'évolution. Les monères forment un ensemble de petite taille situé en position basale, rappelant qu'il s'agit de cellules simples et donc très primitives. Emergeant à partir des monères, les protistes, forment un ensemble un peu plus diversifié dont sont issus les multicellulaires, représentés par des groupes de très grande taille, situés au sommet de la classification. Les phylogénies universelles basées sur les séquences d'ARNrS renversent cette image en montrant que les distances génétiques séparant les trois domaines sont du même ordre. Ceci implique qu'il y a autant de différences, au niveau génétique, entre une archée et une bactérie qu'entre une bactérie et un eucaryote et qu'il n'existe pas de corrélation entre diversité génétique et divergence phénotypique apparente. De plus, les distances qui séparent les lignées de multicellulaires (Fungi, Metazoa, Plantae, Rhodophytes et Stramenopiles) sont extrêmement faibles, et beaucoup plus faibles que les distances évolutives pouvant séparer deux bactéries. Ceci démontre une apparition récente et probablement indépendante des multicellulaires au cours de l'évolution et que la majorité de l'évolution s'est faite sous forme unicellulaire.

## LUCA

Les travaux de Woese ont confirmé que tous les êtres vivants actuels descendent d'un organisme unique couramment désigné dans la communauté scientifique par l'acronyme LUCA (*Last Universal Common Ancestor*). Ce surnom fut proposé lors d'un congrès, organisé par Patrick Forterre, qui s'est déroulé en 1996 à la fondation des Treilles (France). LUCA, dont l'existence est largement admise dans la communauté scientifique, occupe une place centrale dans l'évolution car c'est à partir de lui que se sont séparées les lignées menant aux trois domaines (Figure 1). Il est important de remarquer que LUCA et la première cellule sont des organismes distincts, i.e. ayant vécu à des époques différentes (Figure 2). LUCA était probablement un organisme évolué et très proche des cellules modernes (Forterre, Gribaldo, and Brochier, 2005). Il est également très probable que la vie terrestre de cette époque, bien que très certainement exclusivement microbienne, était déjà très diversifiée. De nombreuses lignées cellulaires ont du coexister, mais toutes se sont éteintes à l'exception de celle conduisant à LUCA (Figure 2). De même, toutes les lignées issues de descendants de LUCA, à l'exception des trois domaines et des lignées qui les composent, se sont éteintes (Figure 2).

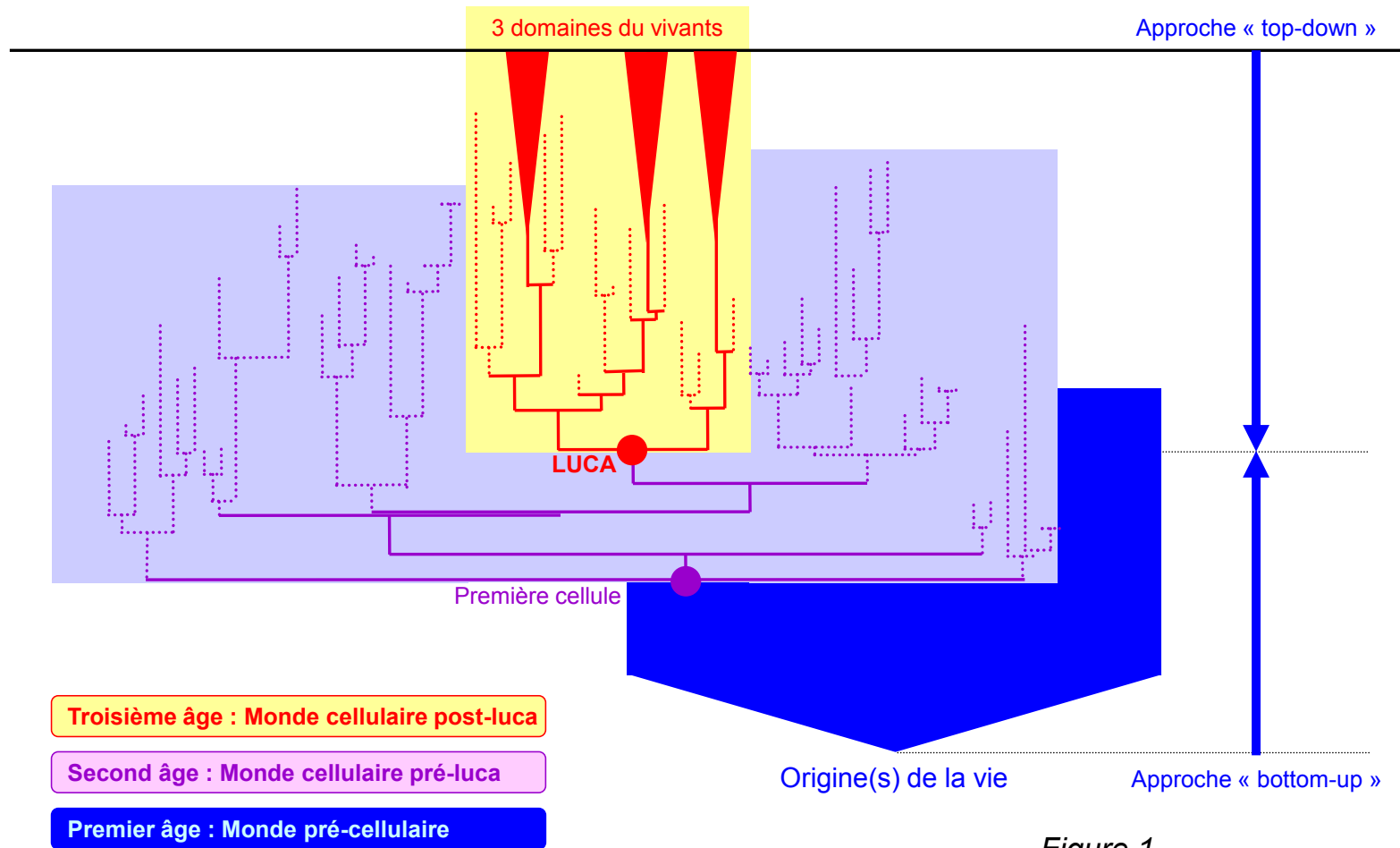


Figure 1

Figure 1.

Schéma représentant l'histoire évolutive du monde vivant (adapté de (Forterre, Gribaldo, and Brochier, 2005)). L'histoire évolutive du monde vivant peut être divisée en trois stades évolutifs, de durée indéterminée, ayant pu coexister dans le temps. Le premier âge commence avec l'origine(s) de la vie et s'achève avec l'émergence de la première cellule. Le second stade commence avec la première cellule et s'achève avec le LUCA. Toutes les lignées évolutives du second âge exceptée celle menant à LUCA se sont éteintes sans laisser de descendants. Le troisième âge commence avec LUCA et l'émergence de deux des trois domaines du vivant. Il est important de noter, que presque toutes les lignées évolutives du second et du troisième âge se sont éteintes sans laisser de descendants (lignées pointillées). Seuls subsistent les lignées qui composent les trois domaines.

Il est donc très vraisemblable que la biodiversité observée actuellement soit bien moindre que celle ayant peuplé la terre au cours de l'évolution.

Il existe deux approches complémentaires pour appréhender l'évolution ancienne. La première, aussi appelée top-down, consiste à comparer les caractères des organismes actuels pour essayer de comprendre leur origine et leur évolution (Figure 2). L'intégration de toutes ces informations dans un contexte évolutif permet de retracer l'évolution des trois domaines à partir de LUCA. La seconde, appelée bottom-up, se focalise sur les grandes étapes qui ont permis le passage de l'inanimé aux premières cellules et leur évolution jusqu'à LUCA.

Une des premières questions posées par ces études est l'époque à laquelle vivait LUCA. Or, il est très difficile de répondre à cette question, car il existe très peu d'archive fossile. De plus même lorsque des fossiles putatifs de microorganismes sont découverts, il est impossible de dire s'il s'agit d'ancêtres ou de descendants de LUCA.

La nature de LUCA est une question tout aussi importante concerne la nature de LUCA. Certains favorisent l'idée que LUCA était un organisme unique. Cette vision de LUCA se justifie par le fait qu'au cours de la division cellulaire une cellule-mère donne naissance à deux cellules-filles. Suivant ce principe, il est théoriquement possible, à partir de toutes les cellules actuelles, de remonter le temps de génération en génération en retraçant leur généalogie (Figure 3A). En remontant suffisamment dans le temps, un point de coalescence va être atteint. Ce point de coalescence représente la cellule dont sont issues toutes les cellules actuelles, c'est-à-dire LUCA. Un autre courant de pensée, initié par Woese, objecte que l'approche précédente est trop réductrice car elle retrace uniquement l'histoire évolutive des enveloppes cellulaires (Woese, 2000). Selon lui, retracer l'histoire évolutive des organismes, implique non seulement de retracer celle des cellules, mais aussi celle de l'ensemble de leurs gènes (Figure 3B). Or les gènes sont susceptibles d'être transférés au cours de l'évolution d'une cellule à l'autre (transferts horizontaux de gènes (THG)). Chaque gène est donc susceptible d'avoir une histoire évolutive propre et en conséquence, il n'existe pas un seul LUCA, mais une multitude de LUCAs (virtuellement un pour chaque gène) qui n'ont pas cohabité dans l'espace et dans le temps.

Figure 2

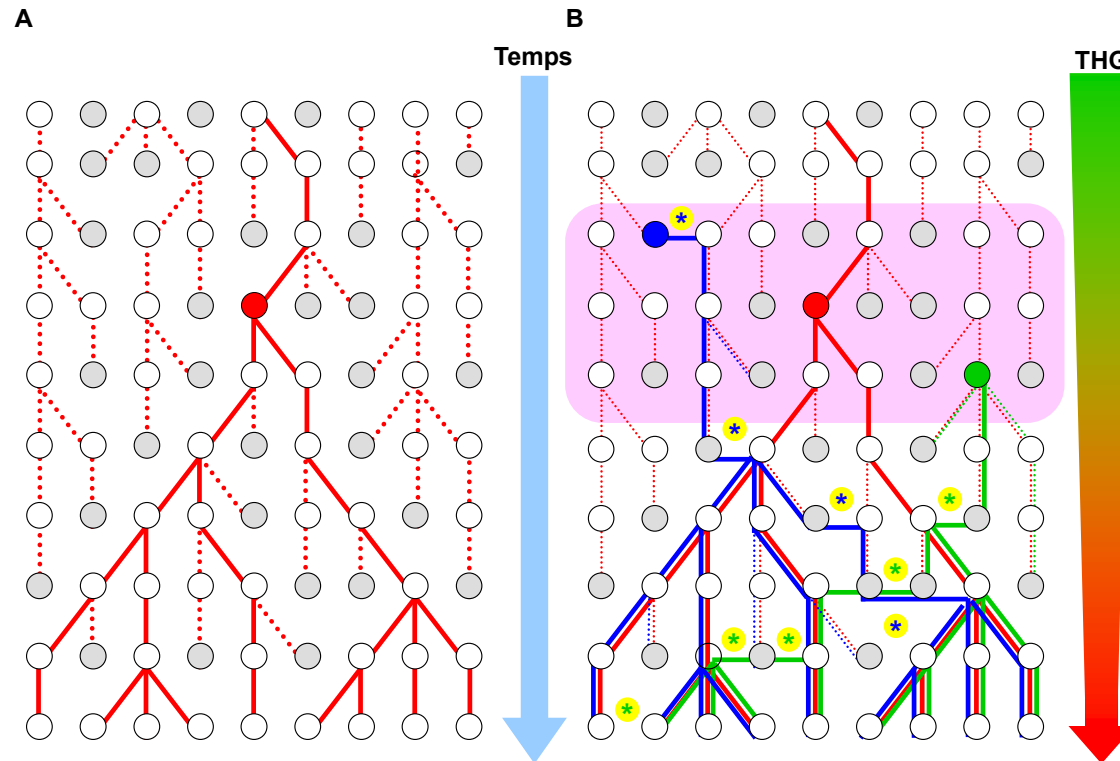


Figure 2.

**A.** Schéma illustrant le principe de coalescence. Le modèle utilisé suppose une population cellulaire de taille, dans laquelle chaque cellule peut se reproduire et laisse en moyenne un descendant à chaque génération (certaines cellules laisseront un descendant, d'autres pourront en laisser plusieurs et certaines aucun). En remontant le temps de génération en génération en suivant la généalogie des cellules, un point de coalescence est atteint. Ce point représente l'ancêtre commun le plus récent dans le temps de toutes les cellules de la population étudiée. Il est important de noter que cet ancêtre, lui-même fruit d'une évolution (i.e. ayant possédé de nombreux ancêtres), ne vivait pas seul mais au sein d'une population diversifiée. Si la population étudiée est l'ensemble des cellules vivantes sur terre actuellement, le point de coalescence représente LUCA.

**B.** Si on intègre à l'histoire de LUCA l'histoire de tous ses composants (en particulier celle de ses gènes qui sont sujets à THG (astérisques jaunes)) en plus de la généalogie cellulaire, alors plusieurs points de coalescence distincts peuvent être atteints (virtuellement un pour chaque composant étudié). LUCA dans ce cas ne peut pas être considéré comme un organisme unique. Il s'agirait d'un ensemble de cellules ayant vécu dans des lieux et à des époques différentes. Cette approche est souvent couplée à l'idée que les THG étaient beaucoup fréquents qu'actuellement. La fréquence de ces derniers se serait réduite lorsque les trois domaines du vivant se seraient individualisés (flèche située à droite).



## **LUCA organisme cellulaire ou non ?**

Si l'étude de LUCA soulève des débats parfois très vifs dans la communauté scientifique, un consensus semble cependant avoir été atteint quant à son statut d'organisme cellulaire. En effet, pour beaucoup, l'apparition des membranes a eu lieu tôt au cours de l'évolution, permettant la concentration des éléments essentiels à l'émergence de la vie. Une hypothèse provocatrice a cependant été émise récemment (Martin and Russell, 2003). Ces auteurs proposent que la vie et une part importante de son évolution s'est déroulée dans une cheminée hydrothermale. L'océan et l'activité géochimique au niveau de la base de cette cheminée auraient fourni l'énergie et les constituants nécessaires à l'émergence de la chimie prébiotique. Un confinement au niveau de micro-cavités minérales aurait compensé l'absence de compartiment membranaires. Les premières membranes seraient postérieures à LUCA et seraient apparues indépendamment dans les lignées bactérienne et archée. Ce modèle, jugé peut convainquant, avait été proposé pour expliquer la présence de lipides membranaires différents chez les bactéries et eucaryotes d'une part, et chez les archées d'autre part. Si tous sont constitués d'une molécule de glycérol-phosphate associée à deux chaînes hydrocarbonées, chez les bactéries et les eucaryotes, c'est un glycérol-3-phosphate (G3P) qui est lié à deux molécules d'acides gras par des liaisons ester, alors que chez les archées, il s'agit de glycérol-1-phosphate (G1P) relié par des liaisons éther à des molécules d'isoprénoïdes. Il existe néanmoins des exceptions car des lipides avec liaison éther ont été décrits chez des eucaryotes et des bactéries thermophiles alors que des acides gras ont été découverts chez des archées. Seule la différence au niveau de la stéréochimie du glycérol-phosphate semble n'avoir aucune exception jusqu'à présent. La présence de phospholipides de nature différente chez les bactéries/eucaryotes et archées a été interprétée comme un argument en faveur d'une apparition indépendante des lipides et donc des membranes. Cependant, la génomique comparée a mis en évidence que des protéines membranaires impliquées dans la sécrétion (SRP) ou dont l'activité est liée à l'exploitation ou la création d'un gradient de protons (ATP synthases, cytochrome oxidases ...) étaient présentes chez LUCA (Forterre, Gribaldo, and Brochier, 2005). Tous ces arguments penchent en faveur de la présence de membranes chez LUCA dont les lipides étaient peut-être synthétisés de manière non stéréo-spécifique (Pereto, Lopez-Garcia, and Moreira, 2004). La stéréo-spécificité observée actuellement pourrait être apparu secondairement par le recrutement d'enzymes différentes.

## **Le génome de LUCA**

De nombreuses études évolutives centrées sur LUCA ont pour objectif la caractérisation de son génome. Cette question est très importante, car la connaissance du nombre et du type de gènes présents chez LUCA permettrait d'établir un portrait robot de cet organisme et de répondre à des

questions telles que : Dans quel type d'environnement vivait-il ? Quelles étaient ses capacités physiologiques ?... Les hypothèses relatives au génome LUCA reflètent l'idée que les évolutionnistes se font de lui. Pour certains, LUCA était un organisme simple (proche des toutes premières cellules), il devait donc posséder un génome de petite taille (<200 gènes). L'évolution post-LUCA se serait faite principalement dans le sens d'une complexification croissante, par gain de matériel génétique dans les trois domaines. Pour d'autres au contraire LUCA possédait un génome de très grande taille et l'évolution se serait faite principalement par perte de gènes dans les trois domaines. Enfin, certains, plus modérés, imaginent LUCA avec un génome d'une taille similaire à ceux des cellules modernes. Dans ce cadre, l'évolution aurait été équilibrée entre gains et pertes de gènes dans les trois lignées.

La génomique comparative s'intéresse à ces questions en recherchant les gènes homologues communs aux trois domaines. Leur analyse phylogénétique permet de déterminer ceux qui étaient présents chez LUCA, qui ont été transmis aux trois domaines et qui ont été conservés au cours de l'évolution (Figure 3A). La comparaison de 21 génomes complets bactériens et d'archées a mis en évidence 51 à 81 gènes de ce type (Koonin, 2003). Ce sont principalement de gènes impliqués dans la formation du ribosome et la synthèse des protéines. Cet ensemble de gènes est insuffisant pour permettre une vie cellulaire. Une explication proposée est que certains homologues ont trop divergé pour être identifiés (Figure 3B), ou ont été perdus ou remplacés au cours de l'évolution (Figures 3C-D). L'inclusion dans ces analyses des génomes de symbiotes ou de parasites, caractérisés par des processus de réduction génomique, peut également contribuer à réduire ce set de gènes universel. Enfin, il est aussi possible que des homologues détectés dans les trois domaines n'étaient pas présents chez LUCA (Figures 3E-F). Pour affiner ces analyses, des approches moins restrictives, se basant sur la recherche de fonction (et non pas de gènes) conservées dans les trois domaines, ont abouti à la proposition qu'au moins 300/400 fonctions pourraient avoir été présentes chez LUCA (Kyrpides, Overbeek, and Ouzounis, 1999). Ce nombre proche de ce qui est observé au niveau des plus petits génomes est donc compatible avec une vie cellulaire. Cependant, certaines catégories fonctionnelles sont mystérieusement sous représentées au niveau du set de gènes universels. C'est le cas par exemple des gènes impliqués dans des fonctions liées à l'ADN, tant au niveau de la synthèse des nucléotides que de la réplication (Leipe, Aravind, and Koonin, 1999). Ceci a été interprété par certains par le fait que LUCA était dépourvu d'ADN et possédait un génome ARN (Mushegian and Koonin, 1996). Ceci implique que le recrutement de l'ADN comme support du matériel génétique se serait produit plusieurs fois indépendamment. D'autres au contraire continuent de soutenir l'hypothèse d'un LUCA avec un génome ADN avançant que certaines des protéines dont les fonctions sont liées à l'ADN ont pu être remplacées au cours de l'évolution. Enfin, un rôle possible des virus dans la transition du monde ARN au monde ADN a été postulé récemment (Forterre, 2005).

Figure 3

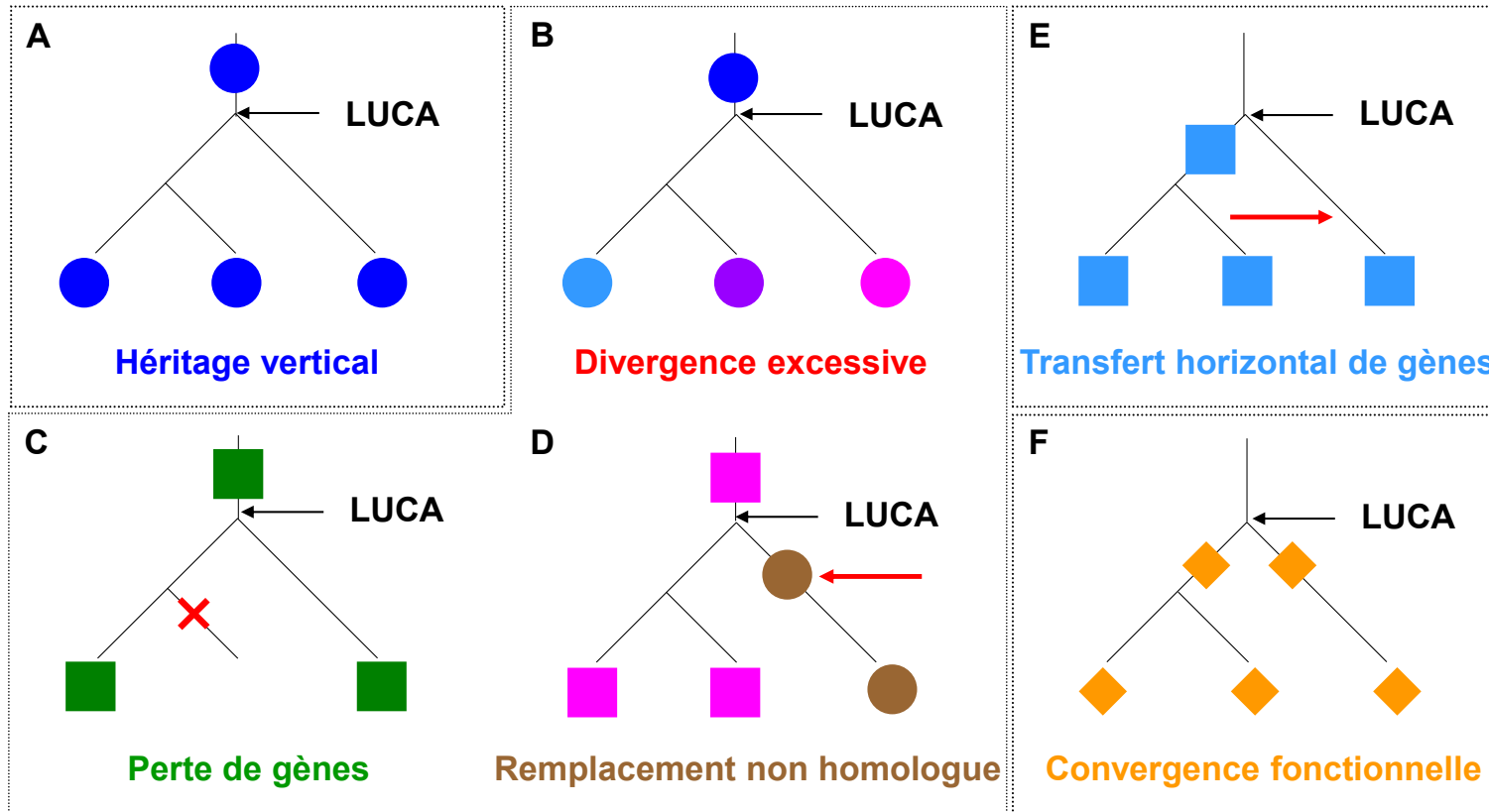


Figure 3.

A. Gènes orthologues hérités de LUCA et conservés dans les trois domaines

B-D. Gènes orthologues hérités de LUCA, conservés dans les trois domaines mais présentant des séquences être détectés (B), perdu dans l'un des trois domaines (C), remplacé dans l'un des domaines suite à un recrutement secondaire ou un THG, par un gène non homologue (D).

E. Gènes homologues, non hérités de LUCA, présents dans les trois domaines suite à des THG.

F. Fonction, non héritée de LUCA, présente dans les trois domaines suite à une convergence évolutive (i.e. apparition indépendante d'un même caractère au cours de l'évolution).

## **LUCA hyperthermophile ou mésophile ?**

L'émergence précoce des lignées hyperthermophiles au sein des archées et des bactéries a été interprétée par le fait que LUCA était lui-même hyperthermophile (pour plus de détail autour de cette question voir (Brochier, 2002)). Cette hypothèse, très rapidement acceptée s'accorde avec les hypothèses d'émergence de la vie à haute température suivie d'une adaptation progressive et indépendante à des environnements mésophiles dans chacun des trois domaines (Stetter, 1996). Il est important de souligner que la question de l'hyperthermophilie de LUCA est en fait déconnectée de celle de l'origine de la vie. En effet, les deux époques concernées étant probablement très éloignées dans le temps (Figure 1), le milieu dans lequel vivait LUCA était probablement très différent de celui dans lequel la vie est apparue. Cependant, l'hypothèse d'un LUCA hyperthermophile est toujours âprement débattue. Tout d'abord, l'analyse des séquences d'ARNrS par des méthodes plus adaptées a montré que la position basale des phylums hyperthermophiles bactériens résultait d'une ALB (Brochier and Philippe, 2002). De plus, des estimations de la composition en bases des séquences d'ARNr de LUCA par des modèles de Markov aboutissent à des taux de G+C incompatibles avec une vie à haute température (Galtier, Tourasse, and Gouy, 1999). Enfin, la génomique comparée a mis en évidence qu'un grand nombre de THG se sont vraisemblablement produit au cours de l'évolution d'archées hyperthermophiles vers les bactéries hyperthermophiles et auraient pu favoriser une adaptation secondaire de ces dernières à un mode de vie hyperthermophile.

## **Origine et relations de parentés entre les trois domaines**

L'existence de trois domaines pose la question de l'ordre relatif d'émergence des trois domaines à partir de LUCA, c'est-à-dire de la position de la racine de l'arbre universel du vivant. Les progrès de la biologie moléculaire ont mis en évidence l'existence de caractères partagés par les domaines deux à deux. Malheureusement, ils ne peuvent pas être utilisés pour déterminer les relations de parentés des trois domaines car ils ne sont pas polarisés dans le temps. Ainsi, comment interpréter la présence de processus informationnels de type archées/eucaryote ? Est un caractère ancestral, présent chez LUCA et perdu dans la lignée bactérienne ? Ou bien est-ce un caractère dérivé qui a remplacé ceux de type bactérien hérités de LUCA chez un ancêtre des archées et des eucaryotes ? En absence d'une polarisation des ces caractères, seules les approches de phylogénie moléculaire permettent de résoudre l'ordre d'émergence des trois domaines.

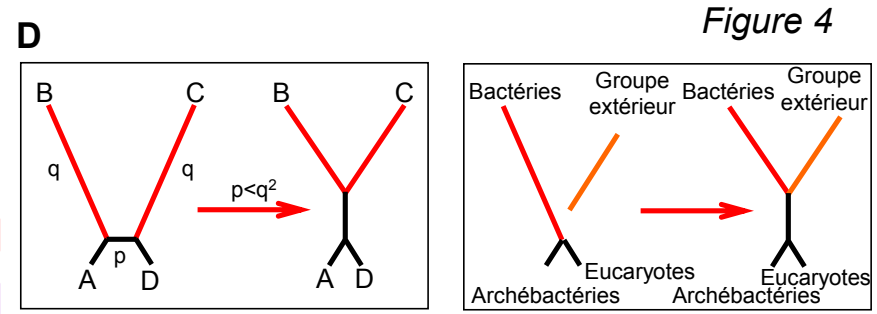
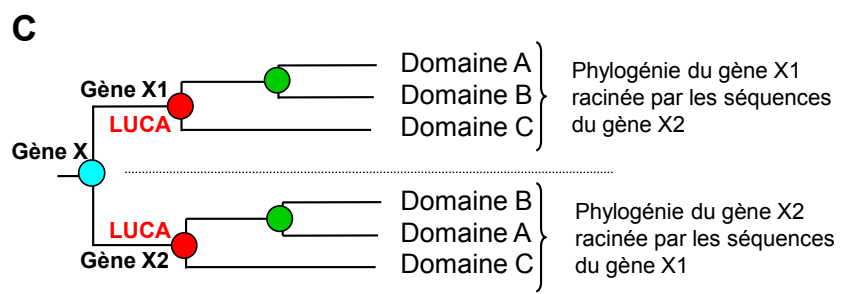
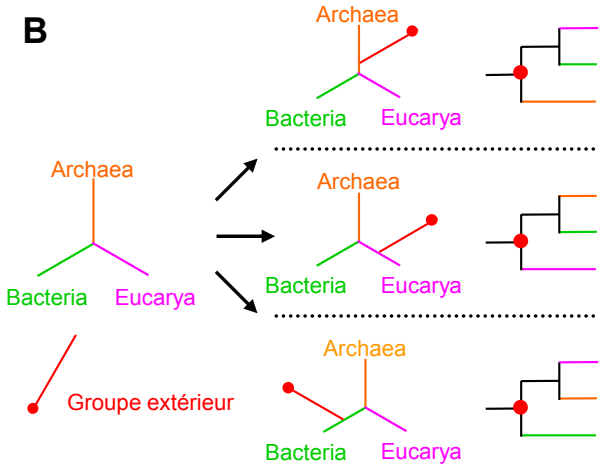
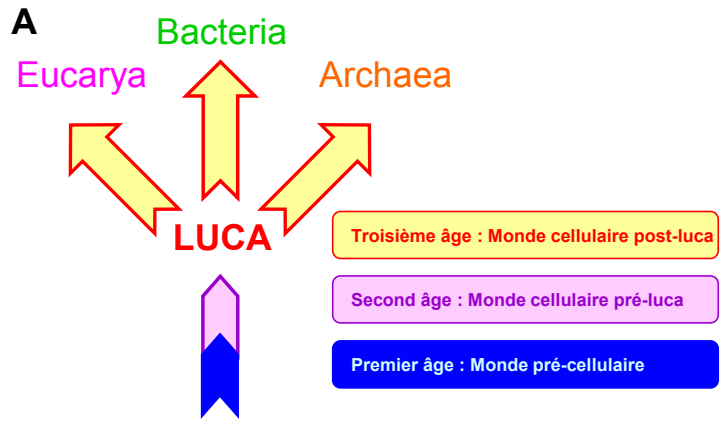
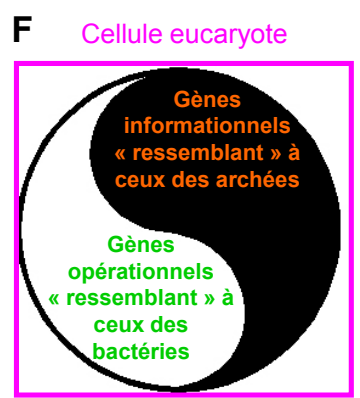
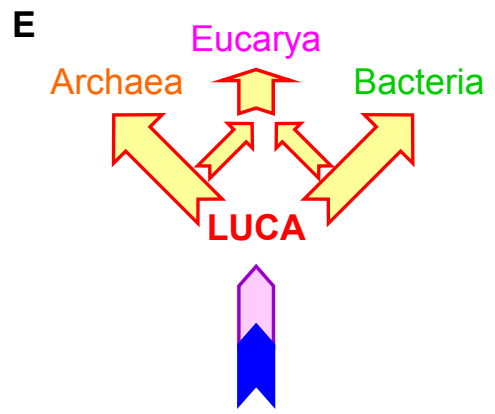


Figure 4



**Figure 4.**

**A.** Schéma représentant la vision traditionnelle de l'émergence des trois domaines du vivant suite à deux événements de spéciation. Les relations de parentés entre les trois domaines ne pas précisées ici. Le code couleur des flèches est similaire celui utilisé pour les trois âges de la figure 2.

**B.** Schéma montrant les trois enracinements possibles de l'arbre universel du vivant grâce à l'inclusion d'un groupe extérieur.

**C.** Utilisation des protéines anciennement dupliquée pour raciner l'arbre universel du vivant. Soit X, un gène ayant été dupliqué chez un ancêtre de LUCA (rond bleu). Les deux copies résultantes vont évoluer indépendamment l'une de l'autre, donnant naissance aux deux gènes X1 et X2. A partir de LUCA deux lignées s'individualisent suite à un événement de spéciation (rond rouge), puis l'une des deux lignées va subir à son tour un événement de spéciation (rond vert). Ces deux événements de spéciation sont à l'origine des trois domaines du vivant et tous les organismes qui les composent possèdent une copie de X1 et une de X2. La reconstruction d'une phylogénie moléculaire de X1 et X2 permet d'obtenir une phylogénie en miroir (i.e. symétrique) dans laquelle la phylogénie de X1 est racinée par les séquences de X2 et réciproquement.

**D.** Artefact d'attraction des longues branches. Soit 4 taxons A, B, C et D dont les relations de parentés réelles sont connues. Si les séquences de B et D évoluent beaucoup plus vite que celle de A et C ( $q^2 > p$ ), et que la branche interne séparant les deux groupes A+B et C+D est courte, alors les phylogénies obtenues avec la plupart des méthodes classiques rapprocheront A de D et B de C.

**E.** Scénario alternatif de l'origine des eucaryotes : LUCA a donné naissance à deux domaines : les archées et les bactéries. Les eucaryotes sont issus d'une association entre archées et bactéries. Le code couleur des flèches est similaire celui utilisé pour les trois âges de la figure 2.

**F.** Le séquençage des génomes complets a révélé la nature chimérique des eucaryotes : les gènes eucaryotes impliqués dans les processus de transmission et d'expression de l'information génétique (i.e. gènes informationnels) sont proches de ceux des archées alors que ceux impliqués dans la maintenance et le métabolisme de la cellule (i.e. gènes opérationnels) sont proches des ceux des bactéries.

**4**

Pour une majorité de scientifiques, la divergence des trois domaines à partir de LUCA s'est faite par événements de spéciation successifs : LUCA donne naissance à deux lignées distinctes, un des trois domaines et une lignée qui subira ultérieurement un événement de spéciation qui donnera naissance aux deux autres domaines (Figure 4A). Or il existe trois manières de raciner l'arbre universel du vivant (Figure 4B), chaque placement impliquant une histoire différente. Par exemple, si la racine est placée dans la branche eucaryote, cela signifie qu'à partir de LUCA, ont émergé deux lignées qui vont donner naissance aux procaryotes (qui forment un groupe monophylétique) d'une part et aux eucaryotes d'autre part. Dans ce cas de figure, la nature procaryote ou eucaryote de LUCA reste indéterminée. En revanche, il possédait très vraisemblablement des lipides de type bactérien et des processus informationnels de type archées/eucaryotes. Les deux autres enracinements possibles sont dans la branche des archées ou dans celle des bactéries. Dans les deux cas, les procaryotes ne forment plus un groupe monophylétique, mais sont paraphylétiques et l'hypothèse la plus parcimonieuse est de supposer que LUCA possédait une organisation cellulaire de type procaryote ; les eucaryotes se différenciant à partir d'un ancêtre procaryote. Si la racine est placée dans la branche bactérienne, alors LUCA aurait pu posséder des processus informationnels de type bactérien ou archées/eucaryote, alors que ces lipides étaient probablement de type bactérien. En revanche si la racine est placée dans la branche des archées, la présence de processus informationnels de type archées/eucaryotes chez LUCA est probable alors que la nature de ses lipides reste incertaine.

Pour raciner une phylogénie, il faut disposer d'un groupe de taxons extérieur au groupe étudié, or par définition l'arbre universel du vivant englobe tous les organismes vivants. En apparence insoluble ce problème a été contourné de manière élégante par Schwartz et Dayhoff, à la fin des années soixante-dix, qui proposent d'utiliser des gènes qui ont subi une duplication chez un ancêtre de LUCA, c'est-à-dire avant la séparation des trois domaines (Figure 4C). Les deux copies résultantes sont appelées paralogues. Les premières phylogénies basées sur de telles protéines aboutissent au même résultat et positionnent la racine de l'arbre universel du vivant dans la branche bactérienne (Gogarten et al., 1989; Iwabe et al., 1989). Ceci implique que les archées sont plus spécifiquement apparentées aux eucaryotes qu'aux bactéries. Ce résultat a été confirmé l'analyse d'autres protéines anciennement dupliquées. La position cette racine, si elle a été majoritairement acceptée, demeure fortement controversée. La première critique est que les protéines utilisées ne contiennent plus suffisamment de signal phylogénétique pour résoudre les relations de parenté entre les domaines. De plus, les arbres obtenus sont caractérisés par des branches bactériennes significativement plus longues que celles des archées ou des eucaryotes, suggérant que les séquences bactériennes évoluent plus vite (Philippe and Forterre, 1999). Or les différences de vitesse d'évolution entre les séquences sont à

l'origine d'artefacts de reconstruction comme l'Attraction des Longues Branches (ALB) qui a pour conséquence le regroupement des séquences évoluant vite d'une part et séquences évoluant lentement d'autre part (Figure 4D). Dans le cas des phylogénies de protéines anciennement dupliquées, une ALB entre la longue branche bactérienne dans la phylogénie du premier paralogue et la longue branche menant au deuxième paralogue (branches rouge et orange, respectivement, Figure 4D) a été démontrée par la ré-analyse de certains jeux de données. Ces analyses favorisent une racine située dans la branche eucaryote, mais de manière non statistiquement significative (Lopez, Forterre, and Philippe, 1999). La conclusion de toutes ces analyses, est que sur la base des protéines anciennement dupliquées, il n'est pas possible résoudre la position de la racine de manière certaine. Il faut donc chercher de nouvelles pistes pour tenter de résoudre cette question fondamentale sans laquelle la nature de LUCA reste incertaine (Baptiste and Brochier, 2004).

Pour une autre partie de la communauté scientifique, les débats autour de la position de la racine n'ont aucun sens, car LUCA aurait donné naissance à deux domaines : les bactéries et les archées (Figure 4E). Les eucaryotes seraient quant à eux issus d'une association entre archées et bactéries (pour une revue détaillée voir (Lopez-Garcia and Moreira, 2006)). Cette hypothèse repose sur les premières données génomiques eucaryotes, qui ont montré que les gènes des eucaryotes pouvaient être divisés en trois grandes catégories. La première catégorie regroupe les gènes spécifiques des eucaryotes qui sont probablement apparus après l'émergence de ce domaine. La seconde catégorie regroupe des gènes très similaires à ceux des archées et qui sont impliqués principalement dans les processus informationnels. La dernière catégorie est constituée de gènes très proches des gènes bactériens impliqués majoritairement dans le métabolisme.

Les mécanismes d'association entre bactéries et archées varient généralement d'un auteur à l'autre (fusion, englobement ou symbiose). Les théories les plus abouties sont certainement les hypothèses qualifiées d'hypothèses hydrogènes qui reposent sur le transfert interspécifique d' $H_2$  (Lopez-Garcia and Moreira, 2006). L'hypothèse proposée repose sur une association syntrophique en anaérobiose stricte entre des archées méthanogènes, produisant du méthane à partir de  $CO_2$  et d' $H_2$  produits par des delta-protéobactéries fermentatives hétérotrophes. Cette association fait intervenir un troisième partenaire plus tardif, des alpha-protéobactéries capables d'oxyder une grande variété de composés (méthane, fumarate, oxygène...). Une étape de symbiose obligatoire aurait suivi avec (i) l'apparition de compartiments membranaires isolant les différentes activités métaboliques et cataboliques, qui vont donner naissance à des compartiments spécialisés (noyau, réticulum endoplasmique, golgi...), (ii) des THG dirigés principalement vers les archées qui prend le progressivement contrôle de la quasi-totalité de l'expression génétique, et (iii) la perte de la méthanogénèse consécutive à une intégration stable des alpha-protéobactéries qui prennent



progressivement le contrôle du métabolisme. Les hypothèses chimériques ont souvent du mal à expliquer le découplage physique et temporel de la transcription et de la traduction chez les eucaryotes. La pression évolutive avancée ici, est liée à l'invasion progressive du génome de l'archée par des introns provenant du génome de l'alpha-protéobactérie. La présence d'introns implique une étape d'épissage du messager avant sa traduction qui pourrait avoir favorisé la séparation physique de la transcription et de la traduction.

Les arguments en faveur de l'hypothèse syntrophique sont que de telles associations existent actuellement. Cependant l'analyse des génomes eucaryotes n'a pas révélé une grande proportion de gènes proches de gènes de delta-protéobactéries ou de méthanogènes. L'analyse de nouvelles données devraient permettre de tester ces hypothèses très prochainement.

### **Conclusion**

L'étude de LUCA est un exercice difficile en raison du peu de données disponibles et des problèmes méthodologiques rencontrés pour leur analyse. Par conséquent, de nombreuses questions relatives à LUCA restent actuellement sans réponse. Beaucoup d'espoirs sont nés avec l'avènement deux nouvelles disciplines : l'écologie moléculaire et la métagénomique, qui sont en train de bouleverser notre vision de la biodiversité terrestre et de son évolution d'une manière comparable à celle provoquée par la publication de la première phylogénie universelle du vivant il y a 30 ans. Outre une meilleure connaissance de la biodiversité à l'intérieur de lignées eucaryotes, bactériennes et archées déjà connues, elles ont permis la découverte de nouveaux groupes majeurs. Les analyses de métagénomiques ont quant à elles permis la découverte de nouveaux métabolismes. Il est très probable que l'amélioration de notre connaissance de la biodiversité, tant du point de vue biologique que génétique, va apporter de nouvelles pièces qui permettront de poursuivre la reconstruction de ce puzzle fascinant que sont l'origine et l'évolution de la vie sur notre planète.

### **Remerciements**

Je tiens à remercier chaleureusement Patrick Forterre, Simonetta Gribaldo, David Moreira et Purificacion Lopez-Garcia pour les nombreuses discussions que nous avons eues au sujet de LUCA ces dernières années ; Simonetta Gribaldo, Eric Armanet et Violaine Brochier pour leur relecture critique de ce manuscrit ; Jean-Claude Bertrand et Manolo Gouy pour le travail collectif que nous avons réalisé dans le cadre de la rédaction d'un chapitre dédié à LUCA qui paraîtra dans le livre *Ecologie Microbienne* ; La Société Française de Microbiologie qui m'a invitée à son VIIième congrès.

### **Bibliographie**

- Adoutte A., Germot A., Le Guyader H., Philippe H. (1996). Que savons-nous de l'histoire évolutive des eucaryotes ? 2. De la diversification des protistes à la radiation des multicellulaires. *médecine/sciences* **12**:I-XIX.
- Baptiste E., Brochier C. (2004). On the conceptual difficulties in rooting the tree of life. *Trends Microbiol* **12**:9-13.
- Brochier C. (2002). Phylogénie des Eubactéries et mise en évidence des transferts horizontaux de gènes. Thèse de doctorat, Université Paris-Sud, Orsay.
- Brochier C., Philippe H. (2002). A non-hyperthermophilic ancestor for bacteria. *Nature* **417**:244.
- Forterre P. (2005). The two ages of the RNA world, and the transition to the DNA world: a story of viruses and cells. *Biochimie* **87**:793-803.
- Forterre P., Gribaldo S., Brochier C. (2005). [Luca: the last universal common ancestor]. *Med Sci (Paris)* **21**:860-865.
- Galtier N., Tourasse N., Gouy M. (1999). A nonhyperthermophilic common ancestor to extant life forms. *Sciences* **283**:220-221.
- Gogarten J.P., Kibak H., Dittrich P., Taiz L., Bowman E.J., Bowman B.J., Manolson M.F., Poole R.J., Date T., Oshima T., Konishi J., Denda K., Yoshida M. (1989). Evolution of the vacuolar H<sup>+</sup>-ATPase: implications for the origin of eukaryotes. *Proc. Natl. Acad. Sci. USA* **86**:6661-6665.
- Iwabe N., Kuma K., Hasegawa M., Osawa S., Miyata T. (1989). Evolutionary relationship of archaeobacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes. *Proc. Natl. Acad. Sci. USA* **86**:9355-9359.
- Koonin E.V. (2003). Comparative genomics, minimal gene-sets and the last universal common ancestor. *Nat Rev Microbiol* **1**:127-136.
- Kyrpides N., Overbeek R., Ouzounis C. (1999). Universal Protein Families and the Functional Content of the Last Universal Common Ancestor. *J Mol Evol* **49**:413-423.
- Leipe D.D., Aravind L., Koonin E.V. (1999). Did DNA replication evolve twice independently? *Nucleic Acids Res* **27**:3389-3401.
- Lopez P., Forterre P., Philippe H. (1999). The root of the tree of life in the light of the covarion model. *J. Mol. Evol.* **49**:496-508.
- Lopez-Garcia P., Moreira D. (2006). Selective forces for the origin of the eukaryotic nucleus. *Bioessays* **28**:525-533.
- Martin W., Russell M.J. (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.
- Mushegian A.R., Koonin E.V. (1996). A minimal gene set for cellular life derived by comparison of complete bacterial genomes. *Proc. Natl. Acad. Sci. USA* **93**:10268-10273.
- Nicol G.W., Schleper C. (2006). Ammonia-oxidising Crenarchaeota: important players in the nitrogen cycle? *Trends Microbiol* **14**:207-212.
- Pereto J., Lopez-Garcia P., Moreira D. (2004). Ancestral lipid biosynthesis and early membrane evolution. *Trends Biochem Sci* **29**:469-477.
- Philippe H., Forterre P. (1999). The rooting of the universal tree of life is not reliable. *J. Mol. Evol.* **49**:509-523.
- Stanier R.Y., Van Niel C.B. (1962). The Concept of a Bacterium. *Arch. Mikrobiol.* **42**:17-35.
- Stetter K.O. (1996). Hyperthermophiles in the history of life. *Ciba Found Symp* **202**:1-10.
- Woese C.R. (2000). Interpreting the universal phylogenetic tree. *Proc Natl Acad Sci U S A* **97**:8392-8396.
- Woese C.R. (2007). The birth of the Archaea: a personal retrospective. In "Archaea" (Garrett R.A., Klenk H.P., eds), Blakwell publishing, Oxford, pp. 1-15.**