

## LES ARCHAEA : EVOLUTION ET DIVERSITÉ DU TROISIÈME DOMAINE DU VIVANT

Simonetta GRIBALDO<sup>1</sup>, Patrick FORTERRE<sup>1</sup>, Céline BROCHIER-ARMANET<sup>2</sup>

<sup>1</sup>Unité de Biologie Moléculaire du Gène chez les Extrémophiles, Département de Microbiologie, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15

<sup>2</sup>Université de Provence, Equipe Génome, Evolution, Bioinformatique, Laboratoire de Chimie bactérienne (CNRS UPR 9043), Institut de Biologie Structurale et de Microbiologie, 31 Chemin Joseph Aiguier, 13402 Marseille Cedex 20

### I. DÉCOUVERTE DES ARCHAEA

Les travaux réalisés par Georges Fox et Carl Woese à la fin des années 70 (Woese et Fox, 1977) ont remis en cause la classification des organismes cellulaires, basée sur la dichotomie procaryotes/eucaryotes, proposée par Stanier et Van Niels dans les années 60 (Stanier et van Niel, 1962). Fox et Woese ont montré que cette classification basée sur la structure des cellules ne reflétait pas la diversité génétique des organismes (pour un rétrospectif historique de la découverte des *Archaea*, voir Woese, 2007). Leurs analyses de l'ARN de la petite sous-unité du ribosome (ARNr PSU) ont révélé que si ceux des eucaryotes étaient très différents de ceux des procaryotes, ce qui était dans un certain sens attendu, les ARNr PSU des procaryotes formaient deux ensembles distincts, aussi éloignés l'un de l'autre que des eucaryotes. Le premier groupe de procaryotes, qu'ils baptisèrent Eubacteria (eu = vrai) regroupait les ARNr PSU de la plupart des bactéries classiques ainsi que ceux extraits des mitochondries et des chloroplastes eucaryotes (ce résultat a été un argument décisif pour mettre fin au débat sur l'origine symbiotique ou autogène de ces organelles, en tranchant en faveur de la première hypothèse). Le deuxième groupe de procaryotes était composé des ARNr PSU de quelques bactéries anaérobies strictes capables de produire du méthane comme déchet de leur métabolisme (méthanogénèse). A cause de leur métabolisme singulier, supposé être très primitif sur la base de modèles de la composition de l'atmosphère terrestre ancienne (il y a 3-4 milliards d'années) et possiblement présent chez les premières formes de vie, Fox et Woese baptisèrent ce groupe Archaeobacteria et proposèrent que le monde vivant se divise en trois super règnes, les *Eubacteria*, les *Eucarya*, et les *Archaeobacteria*.

A l'époque, Woese avait déjà le sentiment que les méthanogènes n'étaient pas des bactéries classiques à cause de caractères singuliers comme par exemple, une membrane constituée de lipides atypiques (avec un glycérol-1-phosphate, une liaison éther et des chaînes d'isoprénoides au lieu d'un glycérol-3-phosphate, une liaison ester et des chaînes d'acides gras) (Woese, 2007). Il se procura donc d'autres souches de 'bactéries' bizarres en faisant le pari qu'il s'agirait d'*Archaeobacteria* et qu'elles posséderaient ces lipides atypiques. Cette intuition se révéla correcte et les *Archaeobacteria* s'enrichirent ainsi de nouvelles lignées (*Halobacteria*, *Thermoplasma*, *Sulfolobus*). Après la

publication des premiers arbres universels racinés, sur la base de l'utilisation de protéines anciennement dupliquées (voir ci-après), qui indiquaient que les *Archaeobacteria* sont évolutivement plus proches des *Eucarya* que des *Eubacteria* (Gogarten *et al.*, 1989 ; Iwabe *et al.*, 1989), Woese proposa d'enlever le suffixe bacteria au mot archaeobacteria, et les trois domaines devinrent : *Archaea*, *Bacteria*, et *Eucarya* (Woese *et al.*, 1990).

Il est important de noter qu'à l'époque la plupart de ces lignées étaient composées d'organismes vivant dans des niches écologiques considérées comme extrêmes (hyperthermophiles, hyper-acides, hyper salines). Ceci a conduit à populariser l'idée, toujours très ancrée dans une partie de la communauté scientifique, que les *Archaea* sont des organismes restreints aux environnements extrêmes et à leur associer le sobriquet d'extrémophiles. Or cette image est extrêmement biaisée car, en réalité, les *Archaea* sont présentes dans tous les milieux habitables et dominant la biomasse de certains d'entre eux, comme l'océan profond (Schleper *et al.*, 2005). Cette abondance des *Archaea* a été mise en évidence grâce au développement de l'écologie moléculaire au début des années 90. Cette discipline étudie la diversité des micro-organismes d'un écosystème par le clonage et le séquençage de leur ADN (et plus précisément des ARNr PSU). Elle permet donc de s'affranchir de l'étape critique de culture et d'isolement au cours de laquelle une grande partie des micro-organismes est perdue. Or, un grand nombre d'*Archaea* ne sont pas cultivables par des méthodes standard et, malgré dix ans d'efforts pour adapter les techniques de cultures, beaucoup d'*Archaea* restent non cultivées.

### II. PHYLOGÉNIE ET ÉVOLUTION DES ARCHAEA

Depuis leur reconnaissance en tant que troisième domaine du vivant, la classification des *Archaea* a été principalement basée sur les séquences des ARNr PSU. Les phylogénies basées sur ce marqueur montrent une division du domaine *Archaea* en deux lignées principales (phylums), baptisées *Crenarchaeota* et *Euryarchaeota* (Woese *et al.*, 1990). Cette profonde dichotomie, qui n'est observée ni chez les *Bacteria* ni chez les *Eucarya*, a été confirmée par l'analyse d'autres marqueurs phylogénétiques (Gribaldo et Brochier-Armanet, 2006) et par la mise en évidence de différences importantes au niveau de macro-complexes cellulaires, comme l'ARN

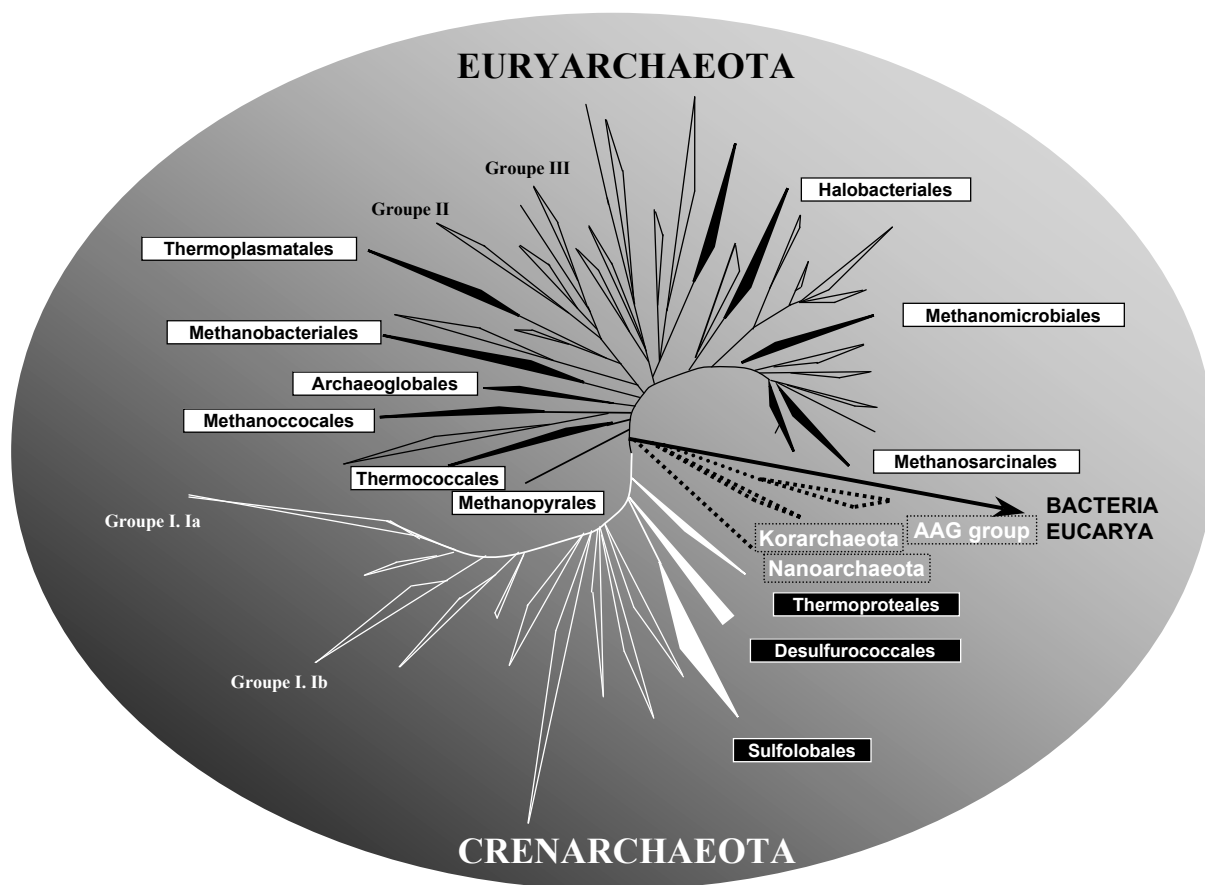


Figure 1: Phylogénie du domaine *Archaea* basée sur l'ARN PSU. Adapté de Schleper *et al.*, 2005.

Les groupes pour lesquels au moins un représentant est cultivé sont indiqués par des triangles pleins; ceux pour lesquels aucun représentant n'est cultivé sont indiqués par des triangles vides. Les *Euryarchaeota* sont représentées en noir (neuf ordres); les *Crenarchaeota* en blanc (trois ordres) et les groupes basaux en pointillés.

polymérase (Huet *et al.*, 1983) ou le ribosome (Lake *et al.*, 1984). L'importance de ces différences a conduit certains scientifiques, comme James Lake, à proposer que les *Crenarchaeota* et les *Euryarchaeota* devraient être considérés comme deux domaines distincts (au même titre que les *Bacteria* ou les *Eucarya*) et non comme deux phylums d'un même domaine. De plus, selon Lake, les *Crenarchaeota* (qu'il propose de renommer Eocytes) seraient plus proches des *Eucarya*, alors que les *Euryarchaeota* seraient spécifiquement apparentées aux *Bacteria* (Lake *et al.*, 1984). Cependant la majorité des analyses phylogénétiques et de génomique comparée ont confirmé que les *Crenarchaeota* et *Euryarchaeota* forment un seul et même domaine, les *Archaea* (Gribaldo et Brochier-Armanet, 2006). En revanche, les analyses génomiques ont confirmé les différences importantes observées au niveau cellulaire entre *Crenarchaeota* et *Euryarchaeota* (Makarova *et al.*, 2007). Par exemple, les *Crenarchaeota* possèdent un certain nombre de protéines impliquées dans les processus informationnels (réplication/réparation, transcription, traduction) qui sont absentes chez les *Euryarchaeota*, et réciproquement (Makarova *et al.*, 2007 ; Brochier-Armanet *et al.*, 2008). Ceci indique qu'après leur

séparation, *Crenarchaeota* et *Euryarchaeota* ont développé des spécificités touchant des mécanismes/structures cellulaires fondamentaux. Une question centrale relative à l'évolution des *Archaea* est donc d'essayer de déterminer dans quelle lignée (*Crenarchaeota* ou *Euryarchaeota*) chacune de ces spécificités est apparue (voir ci-après).

Les phylogénies basées sur l'ARNr PSU ont permis de définir un certain nombre d'ordres, contenant au moins un représentant cultivé, au sein des *Crenarchaeota* et des *Euryarchaeota* (Figure 1). Les *Crenarchaeota* sont actuellement divisées en trois ordres, *Thermoproteales*, *Desulfurococcales* et *Sulfolobales*. Ces trois ordres regroupent des organismes hyperthermophiles ou thermophiles extrêmes (d'où le préfixe crenos=source/origine dans *Crenarchaeota* qui reflète l'idée que les premières formes de vie pourraient avoir été hyperthermophiles), qui peuvent être aérobies (*Sulfolobales*, *Pyrobaculum aerophilum*, *Aeropyrum pernix*...) ou anaérobies (*Hyperthermus*, *Ignicoccus*, *Thermofilum*...) thermoacidophiles (*Sulfolobales*)... Par opposition au nombre restreint d'ordres reconnus chez les *Crenarchaeota*, neuf ordres sont définis au sein des *Euryarchaeota* (*Methanopyrales*, *Thermococcales*, *Methanococcales*, *Methanobacteriales*, *Archaeoglobales*,

*Thermoplasmatales*, *Halobacteriales*, *Methanosarcinales* et *Methanomicrobiales*). Ce nombre plus élevé d'ordres chez les *Euryarchaeota* reflète également une plus grande diversité au niveau écologique (d'où le préfixe euryos = varié dans *Euryarchaeota*). En effet, les *Euryarchaeota* regroupent aussi bien des organismes hyperthermophiles/thermophiles (*Methanopyrales*, *Thermococcales*, *Archaeoglobales*) que mésophiles (*Methanomicrobiales*, *Methanosarcinales*, *Halobacteriales*), des thermoacidophiles ou acidophiles mésophiles (*Thermoplasmatales*), des halophiles extrêmes (*Halobacteriales*), des méthanogènes (*Methanosarcinales*, *Methanococcales*, *Methanomicrobiales*, *Methanopyrales*, *Methanobacteriales*), des aérobies (*Halobacteriales* et *Thermoplasmatales*) ou des anaérobies (méthanogènes, *Thermococcales*, *Archaeoglobales*).

Le nombre relativement limité de phylums (2) et d'ordres (9) défini au sein des *Archaea*, comparé au 20-25 phylums bactériens communément reconnus, peut donner une image très biaisée de leur diversité, alors qu'il est le reflet de notre manque de connaissances concernant ce domaine. En effet, les analyses de la diversité des *Archaea* présentes au sein des écosystèmes par des méthodes d'écologie moléculaire ont révélé un grand nombre de lignées pour lesquels aucun représentant n'a été isolé (Figure 2). Certaines de ces lignées présentent une diversité comparable ou même supérieure à celle de certains ordres. Par exemple, les *Archaea* du groupe I forment un groupe cosmopolite qui semble aussi abondant et diversifié que les *Crenarchaeota* ou les *Euryarchaeota*. Il inclut notamment des lignées présentes dans des écosystèmes très variés tels que les couches superficielles de l'océan, l'océan profond, le sol ou des sources chaudes (Schleper *et al.*, 2005). Beaucoup d'intérêt est porté aux *Archaea* du groupe I. En effet, il a été montré qu'ils sont capables de réaliser l'oxydation aérobie de l'ammoniaque, métabolisme qui était considéré jusqu'à présent comme spécifique des *Bacteria*. Par ailleurs, beaucoup d'entre elles seraient autotrophes. Des études d'écologie microbienne suggèrent que ce groupe pourrait jouer un rôle prépondérant (supérieur à celui des bactéries) dans le cycle de l'azote et du carbone dans certains écosystèmes (Nicol et Schleper, 2006). Bien que quelques écosystèmes soient très étudiés, la biodiversité de beaucoup entre eux reste encore mal connue. Il est donc très probable que notre vision de la biodiversité microbienne, et donc des *Archaea*, soit encore très partielle (Lopez-Garcia et Moreira, 2006). Par ailleurs, il ne faut pas oublier que la connaissance de la biodiversité d'un écosystème ne renseigne pas ou peu sur les caractéristiques des organismes qui le composent, ni sur leurs interactions. Par exemple, une analyse pionnière de fosmidés d'*Archaea* des groupes I, II et III provenant de l'océan profond réalisée par l'équipe de Lopez-Garcia a mis en évidence des différences importantes au niveau des capacités métaboliques, des caractéristiques génomiques et du mode d'évolution de ces groupes (Martin-Cuadrado *et al.*, 2008).

Malgré leur abondance, la presque totalité des lignées d'*Archaea* non cultivées se branchent au sein (ou à proximité) d'ordres de *Crenarchaeota* ou d'*Euryarchaeota* contenant des

organismes cultivés. Par exemple, les *Archaea* du groupe I émergent à proximité des *Crenarchaeota*, alors que celles des groupes II et III sont proches des *Thermoplasmatales* au sein des *Euryarchaeota* (Figure 2). Quelques lignées cependant font exception et, sur la base de phylogénies de l'ARNr PSU, ne se rapprochent ni des *Crenarchaeota*, ni des *Euryarchaeota*. C'est le cas par exemple des séquences isolées dans la source chaude du parc de Yellowstone par le groupe de Norman Pace en 1996 (Barns *et al.*, 1996). La position basale de ces séquences a conduit à la proposition qu'elles proviennent de représentants d'un troisième phylum d'*Archaea* qui fut baptisé *Korarchaeota* (Koros = jeune homme). Depuis leur découverte il y a douze ans, aucune *Korarchaeota* n'a pu être isolée. Cependant, grâce aux efforts intensifs du groupe de Karl Stetter, des cultures enrichies ont pu être obtenues et le premier génome d'un représentant des *Korarchaeota* (*Candidatus Korarchaeum cryptofilum*) séquencé (Elkins *et al.*, 2008). Grâce à ce génome, la position des *Korarchaeota* au sein des *Archaea* a pu être affinée (voir ci-après). Toujours sur la base des séquences d'ARNr PSU, une autre lignée basale d'*Archaea*, les *Nanoarchaeota*, a été élevée au rang de phylum. Les premières observations de *Nanoarchaeota* ont mis en évidence des cellules de très petite taille (400 nm), comptant parmi les plus petites cellules connues, et présentant un mode de vie jamais observé auparavant chez des archées. En effet, elles vivent à la surface d'*Archaea* appartenant au genre *Ignicoccus*, une desulfurococcale hyperthermophile (*Crenarchaeota*), d'où le nom de *Nanoarchaeum equitans* (à cheval) donné au premier représentant décrit. Si la nature de l'association, symbiotique ou parasitaire, entre *N. equitans* et son hôte *Ignicoccus* reste encore mal comprise, il est important de souligner qu'il s'agit de la première association stricte *Archaea/Archaea* décrite. Par ailleurs, le séquençage du génome de *N. equitans* a révélé un génome de très petite taille (490 kilo paires de bases codant 552 gènes). Il s'agit du plus petit génome d'*Archaea* et d'un des plus petits génomes cellulaires connus à ce jour. L'analyse du génome a révélé que presque tous les gènes codés sont impliqués dans des mécanismes informationnels et qu'un grand nombre de gènes impliqués dans le métabolisme et la maintenance cellulaire (lipides, nucléotides, acides aminés...) ont été perdus. Ceci suggère une forte dépendance de *N. equitans* vis-à-vis de son hôte, ce qui est en accord avec les résultats expérimentaux montrant que, si *Ignicoccus* peut être cultivé en absence de *N. equitans*, la réciproque n'est pas vraie.

Il existe d'autres séquences d'hyperthermophiles non cultivés émergeant en position basale dans les arbres de l'ARNr PSU, comme par exemple celles formant le Ancient Archaeal Group (AAG, Figure 2), mais pour lesquelles peu d'information est disponible et qui devront être confirmées par d'autres analyses.

En plus de fournir des renseignements sur les relations de parenté, les phylogénies moléculaires permettent de retracer l'histoire évolutive des groupes. Par exemple, les phylogénies de l'ARNr PSU montrent que les lignées les plus basales dans l'arbre des *Archaea* (*Nanoarchaeota*,

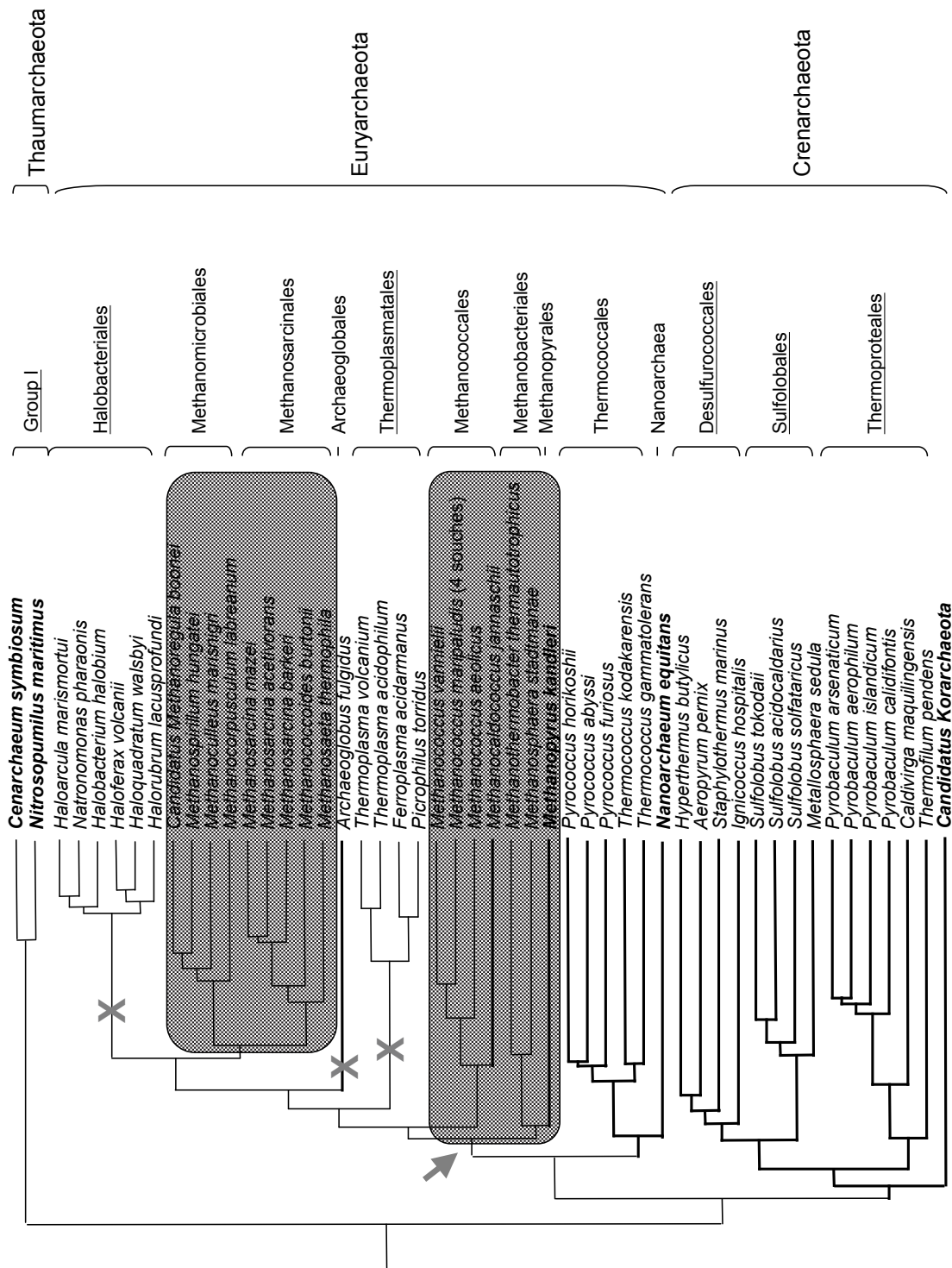


Figure 2: Schéma représentant les relations de parenté entre les *Archaea* dont le génome a été séquencé.

Les relations de parenté indiquées synthétisent l'information phylogénétique apportée par les arbres basés sur les protéines ribosomiques qui ont été publiés ces dernières années (Gribaldo et Brochier-Armanet, 2006 ; Brochier-Armanet *et al.*, 2008). Dans la phylogénie, des branches épaisses désignent des lignées hyperthermophiles, alors que les branches fines indiquent des psychrophiles ou thermophiles modérées. Les groupes d'*Archaea* dont le nom est souligné contiennent des organismes aérobies. Les ovales en fond gris indiquent les *Archaea* méthanogènes. L'apparition de la méthanogénèse est indiquée par une flèche grise et la perte de la méthanogénèse par des X gris. Les noms des organismes marqués en gras indiquent les groupes dont la position a été révisée par rapport à la phylogénie basée sur l'ARNr PSU.

*Korarchaeota*, AAG, les trois ordres de *Crenarchaeota*, ainsi que les *Methanopyrales*, *Thermococcales* et *Archaeoglobales* au sein des *Euryarchaeota*), sont hyperthermophiles (Figure 1). Ce résultat corrobore l'hypothèse selon laquelle LACA (Last Archaeal Common Ancestor) était un organisme hyperthermophile, caractéristique qui aurait été héritée par l'ancêtre des *Crenarchaeota* et l'ancêtre des *Euryarchaeota*. L'émergence plus tardive de lignées mésophiles au sein de ces deux phylums a donc été interprétée par une adaptation indépendante à ces milieux. Par ailleurs, la plupart des représentants de ces groupes étant anaérobies, il a été postulé que LACA était lui-même anaérobie et que l'adaptation à des milieux aérobies serait secondaire. Enfin, l'émergence basale des méthanogènes de l'ordre *Methanopyrales* au sein de *Euryarchaeota* a conduit à l'hypothèse que l'ancêtre des *Euryarchaeota* (et peut-être même celui des *Archaea*) était méthanogène (voir discussion dans Gribaldo et Brochier-Armanet, 2006). Ce métabolisme aurait donc été ancestral au sein de ce groupe, voire même au sein des *Archaea*. L'hypothèse d'un LACA hyperthermophile, anaérobie et méthanogène, s'est imposée rapidement et est encore répandue, car elle s'accorde avec l'idée que les *Archaea* sont un groupe primitif et que la vie est apparue à haute température. Cependant, il faut garder à l'esprit que LACA est l'ancêtre des lignées d'*Archaea* actuelles, c'est-à-dire des *Archaea* modernes. Il est donc le fruit d'une longue évolution qui commence avec l'apparition de la vie, se poursuit avec l'apparition des premières cellules, celle de LUCA (Last Universal Common Ancestor) qui est l'ancêtre commun aux *Archaea*, *Bacteria* et *Eucarya*, continue avec l'apparition de la lignée évolutive conduisant aux *Archaea* et s'achève avec LACA, organisme à partir duquel la diversification des lignées d'*Archaea* vivant actuellement a commencé. Bien qu'il ne soit pas possible de dater ces étapes importantes de l'histoire de la vie sur notre planète, LACA est vraisemblablement très éloigné dans le temps des premières formes de vie. Ses caractéristiques génotypiques et phénotypiques ne peuvent donc pas être corrélées avec les origines de la vie.

Les inférences concernant la nature de LACA et l'évolution des *Archaea* mentionnées précédemment sont basées sur les analyses de l'ARNr PSU. Or, bien que ce marqueur possède d'indéniables qualités pour la reconstruction de phylogénies anciennes (ubiquité, faible vitesse d'évolution, faible taux de transfert de gènes, fonction conservée), il peut être sensible à différents artefacts tels que les biais de composition en nucléotides (comme par exemple l'enrichissement en G et C des séquences d'organismes hyperthermophiles) et l'hétérogénéité des vitesses d'évolution entre les séquences (Gribaldo et Philippe, 2002 ; Brochier-Armanet *et al.*, 2008). L'augmentation des données de séquences protéiques, en particulier grâce au séquençage de génomes complets, a permis la reconstruction de phylogénies basées sur des marqueurs protéiques. Parmi ces marqueurs, les protéines ribosomiques se sont imposées comme une alternative de choix à l'ARNr PSU pour l'étude de la phylogénie des *Archaea* (Gribaldo et Brochier-Armanet, 2006). En effet, les phylogénies inférées sont congruentes avec celles basées sur les ARNr PSU, montrent une meilleure résolution

de certains nœuds et se sont révélées moins sensibles aux artefacts de reconstruction. Par exemple, l'analyse de ces protéines a montré que l'émergence précoce des *Methanopyrales* dans les phylogénies de l'ARNr PSU résultait très probablement d'un artefact de reconstruction lié à l'enrichissement des séquences correspondantes en G et C (Gribaldo et Brochier-Armanet, 2006) et que cet ordre était plus probablement apparenté aux *Methanococcales* et *Methanobacteriales* (Figure 2). Cette position phylogénétique est également corroborée par les analyses du contenu en gènes de ces génomes (Slesarev *et al.*, 2002; Gribaldo et Brochier-Armanet, 2006). Cette révision de la position des *Methanopyrales* est importante car elle n'implique plus l'existence d'un ancêtre méthanogène des *Euryarchaeota* (ni des *Archaea*). Ce métabolisme serait vraisemblablement apparu plus tardivement au cours de l'évolution des *Euryarchaeota* (Figure 2).

Une autre révision importante de la phylogénie des *Archaea* concerne la position basale des *Nanoarchaeota* et *Korarchaeota* dans les phylogénies basées sur l'ARNr PSU. L'analyse phylogénétique d'un certain nombre de marqueurs protéiques, ainsi que les analyses de génomique comparée, ont suggéré que les *Nanoarchaeota* représentent très probablement une lignée très divergente d'*Euryarchaeota*, peut-être proche des *Thermococcales* (Gribaldo et Brochier-Armanet, 2006 ; Makarova *et al.*, 2007). La forte divergence des séquences d'ARNr PSU des *Nanoarchaeota* pourrait être la conséquence d'une adaptation à un mode de vie parasite (perte multiple de gènes, enrichissement du génome en bases A et T ...). Leur placement incorrect dans les arbres de l'ARNr PSU serait la conséquence d'un artefact de reconstruction appelé attraction des longues branches, qui est particulièrement accentué lorsque les séquences sont isolées et très divergentes (Philippe et Laurent, 1998). Par ailleurs, des analyses similaires soutiennent une proche parenté entre *Korarchaeota* et *Crenarchaeota* (Elkins *et al.*, 2008). Néanmoins, dans le cas des *Korarchaeota*, d'autres analyses seront nécessaires pour préciser définitivement leur position au sein des *Archaea*.

Enfin, le séquençage des premiers génomes de représentants des *Crenarchaeota* du groupe I (*Cenarchaeum symbiosium* et *Nitrosopumilus maritimus*) a conduit à la remise en cause de la position de ce groupe au sein des *Archaea*. En effet, la position de ce groupe n'est pas totalement résolue sur la base des phylogénies de l'ARNr PSU : ces *Archaea* sont placées comme groupe-frère des *Crenarchaeota* hyperthermophiles ou comme émergeant au sein des *Crenarchaeota*, en tant que groupe frère des *Sulfolobales* (Figure 1) (Brochier-Armanet *et al.*, 2008). Or l'analyse phylogénétique des protéines ribosomiques de ces organismes supporte au contraire une émergence basale du groupe I dans la phylogénie des *Archaea*, c'est-à-dire avant la séparation des *Crenarchaeota* et des *Euryarchaeota* (Figure 2) (Brochier-Armanet *et al.*, 2008). Ceci suggère que ce groupe a vraisemblablement divergé plus tôt que ce qui est suggéré par les phylogénies basées sur l'ARNr PSU. Par ailleurs l'analyse des gènes présents dans les génomes des deux représentants du groupe I a révélé l'absence d'une majorité des gènes spécifiques des

*Crenarchaeota* (e.g. présents chez toutes les *Crenarchaeota* et absents chez les *Euryarchaeota*), et inversement la présence d'un certain nombre de gènes typiquement *Euryarchaeota* (Brochier-Armanet *et al.*, 2008). Ceci implique qu'au niveau génomique une plus grande similarité est observée entre *Euryarchaeota* et membres du groupe I, qu'entre membres du groupe I et *Crenarchaeota* ; similarité qui ne serait pas liée à des échanges de matériel génétique par transfert horizontal (Brochier-Armanet *et al.*, 2008). Enfin, ces mêmes analyses ont montré qu'un certain nombre de gènes conservés chez les *Euryarchaeota* et les *Crenarchaeota* sont absents des génomes des *Crenarchaeota* du groupe I (Brochier-Armanet *et al.*, 2008). Ces résultats issus de la génomique comparée, associés à une émergence basale des *Crenarchaeota* du groupe I sur la base de l'analyse phylogénétique des protéines ribosomales, à leur cosmopolisme, et à leur grande diversité génétique, a conduit à la proposition qu'il pourrait s'agir d'une lignée majeure (phylum) d'*Archaea*, au même titre que les *Crenarchaeota* et les *Euryarchaeota*. Ce nouveau phylum a été baptisé *Thaumarchaeota* (Thaumas = merveilleux). Il est important de noter que même si la position basale des *Thaumarchaeota* devait être remise en cause par l'analyse de nouvelles données (marqueurs protéiques alternatifs, séquences provenant de nouveaux génomes de *Thaumarchaeota*...), leur statut de phylum ne devrait pas être remis en cause. En effet, même s'il s'avérait que les *Thaumarchaeota* sont un groupe-frère des *Crenarchaeota*, les raisons invoquées précédemment devraient assurer leur maintien en tant que phylum au sein des *Archaea*.

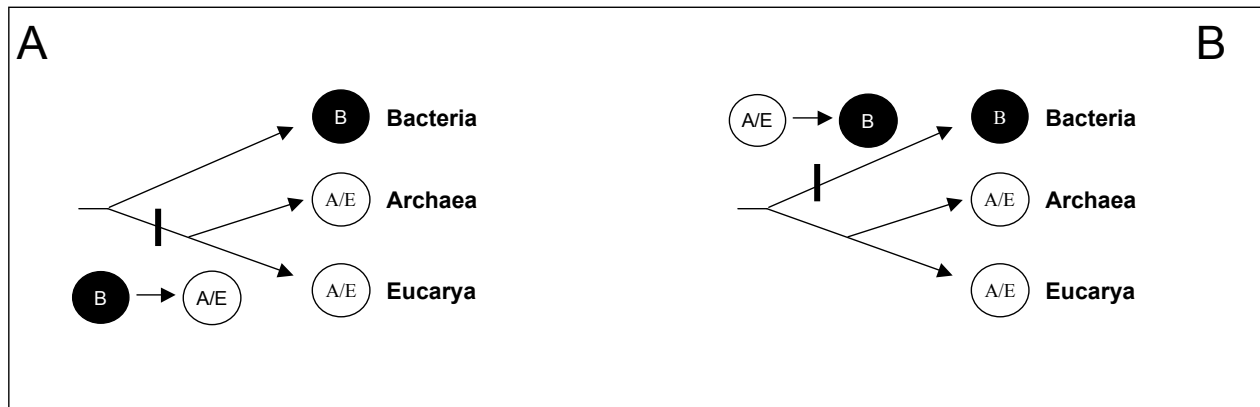
### III. ORIGINE DES ARCHAEA

A de très rares exceptions, l'idée que les micro-organismes ont dominé l'histoire de la vie sur Terre depuis son apparition est largement acceptée. En effet, sur la base de traces indirectes d'activité biologique et de microfossiles présumés, il est admis que les premières communautés microbiennes ont pu apparaître à l'Archéen, plus précisément au cours d'une période située entre 3,4 et 2,5 milliards d'années, les macro-organismes apparaissant dans le registre fossile beaucoup plus tard (à partir de 1,2 milliards d'années). Cependant, même si de telles communautés ont effectivement existé, il est impossible de savoir précisément quels organismes les composaient : s'agissait-il d'archées, de bactéries ou d'eucaryotes ? de leurs ancêtres ? ou bien de formes de vie cellulaires encore plus anciennes, ayant précédé LUCA ? A cause de leur nom, les *Archaea* sont souvent considérées comme les micro-organismes les plus anciens, c'est-à-dire plus anciens que les *Bacteria* ou les *Eucarya*. Cependant cet *a priori* n'est pas prouvé et les *Bacteria* pourraient être aussi anciennes ou même plus anciennes. Par ailleurs, il est important de rappeler que l'interprétation des traces de vie les plus anciennes est compliquée par le fait qu'au cours des milliards d'années d'évolution des lignées évolutives (qui n'étaient ni des *Archaea*, ni des *Bacteria* ni des *Eucarya*) aujourd'hui disparues ont pu exister (Forterre *et al.*, 2005). Si ces lignées se sont éteintes sans laisser de descendants vivant actuellement, il est également probable qu'elles ont possédé des caractéristiques (métabolismes, structures cellulaires...) qui ont disparu avec elles.

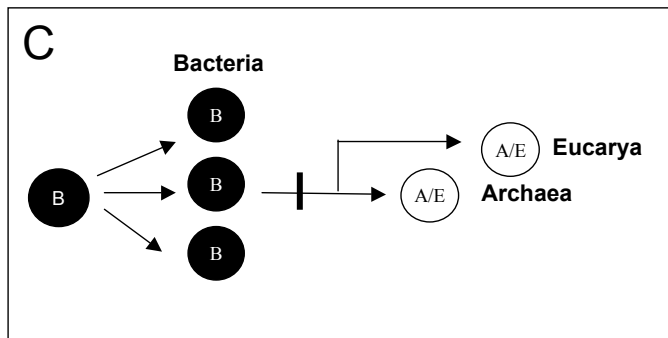
Une question importante relative à l'évolution des *Archaea* concerne leur position par rapport aux *Eucarya* et *Bacteria*. En effet, bien que, depuis leur découverte, les *Archaea* soient considérées presque unanimement comme un domaine à part entière, la question de leurs relations de parenté avec les deux autres domaines (*Bacteria* et *Eucarya*) se pose. Résoudre cette question revient à déterminer si l'événement de spéciation qui a donné naissance à la lignée des *Archaea* se situe au niveau de LUCA, ou si au contraire cette spéciation est plus récente, impliquant que la première lignée à avoir divergé à partir de LUCA était soit celle des *Eucarya*, soit celle des *Bacteria*. Le séquençage des premiers génomes d'*Archaea* a mis en évidence que ces dernières partagent de manière spécifique un nombre important de caractères avec les *Eucarya* (e.g. forte ressemblance des machineries informationnelles). Ceci pourrait être interprété en faveur de l'hypothèse d'une parenté étroite entre ces deux domaines. Cependant, les *Archaea* partagent également des caractères avec les *Bacteria* (organisation cellulaire de type procaryote, présence d'opérons...) et les *Eucarya* partagent des caractères avec les *Bacteria* (lipides...). Par ailleurs, il n'est pas possible de déterminer quels sont, parmi les caractères évoqués, ceux qui sont primitifs et ceux qui sont dérivés. Il n'est donc pas possible de préciser les relations de parenté entre les trois domaines sur la base de l'analyse des caractères partagés par les domaines deux à deux. L'utilisation de la phylogénie moléculaire est donc nécessaire. Cependant pour raciner un arbre phylogénétique, il faut inclure dans l'analyse des séquences provenant d'un groupe extérieur (séquences homologues aux séquences dont on veut retracer l'histoire). Par exemple, pour retracer l'histoire évolutive des mammifères par phylogénie moléculaire, un groupe extérieur possible pourrait être constitué de séquences provenant de vertébrés non mammifères (serpents, crocodiles, oiseaux...). Or comment trouver un groupe extérieur quand l'objet d'étude est l'ensemble des organismes vivants ? Une solution a été proposée dès 1978 (Schwartz et Dayhoff, 1978) pour tenter de répondre à cette question. Concrètement, elle consiste à reconstruire la phylogénie de protéines paralogues très anciennes (protéines apparentées issues d'un événement de duplication de gène antérieur à LUCA et qui ont été conservées dans les trois domaines du vivant). Chacun des deux paralogues présents chez LUCA a enregistré lors de son évolution l'ordre d'émergence des trois domaines. Pour pouvoir révéler cette histoire évolutive, il est nécessaire d'analyser les deux paralogues présents dans chacun des trois domaines de manière simultanée. Il en résulte une phylogénie en miroir, dont chaque partie correspond à l'histoire évolutive d'un des paralogues montrant les relations de parenté entre les trois domaines.

La mise en œuvre de ces analyses a été réalisée indépendamment par deux groupes de recherche en 1989 qui ont travaillé sur deux sous-unités d'une famille universelle d'ATPases (les sous-unités 70 and 60 kDa des H<sup>+</sup>-ATPases vacuolaires chez les *Eucarya*, et les sous-unités a et b des F<sub>0</sub>F<sub>1</sub>-ATPases chez les bactéries), et sur les facteurs d'élongation de la traduction EF-1a et Ef-2 (Gogarten *et al.*, 1989; Iwabe *et al.*, 1989). Ces deux équipes ont obtenu des

Arbre universel du vivant enraciné dans la branche bactérienne



Un seul domaine primaire - Bacteria



Modèles chimériques

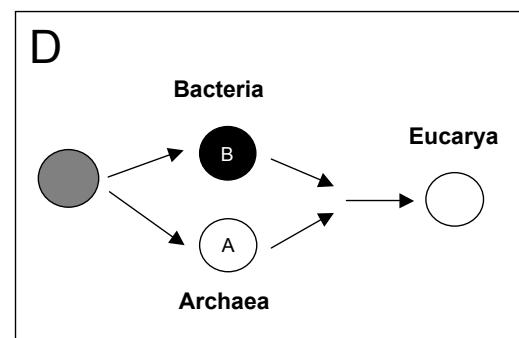


Figure 3: Relations de parenté entre les *Archaea* et les autres deux domaines du vivant.

A et B. Schémas montrant l'enracinement de l'arbre du vivant dans la branche bactérienne. Cet enracinement ne permet pas de déterminer si (schéma A) l'ancêtre des trois domaines possédait des caractères typiquement bactériens (représentés par un cercle noir annoté un B) qui ont été remplacés (barre verticale noire) par des caractères *Archaea/Eucarya* (représentés avec un cercle blanc annoté A/E), avant la séparation de ces deux lignées, ou au contraire si (schéma B) l'ancêtre des trois domaines possédait des caractères typiques *Archaea/Eucarya* (représentés avec un cercle blanc annoté A/E) qui ont été remplacés (barre verticale noire) par des caractères bactériens (représentés par un cercle noir annoté B) dans la lignée bactérienne.

C. Schéma illustrant l'hypothèse de Cavalier-Smith. Selon ce modèle, il n'existe qu'un domaine primaire (les *Bacteria*). Les *Archaea* et les *Eucarya* seraient des domaines secondaires issus des *Bacteria*, et leurs caractères partagés (représentés avec un cercle blanc avec un A/E) seraient donc dérivés (barre verticale noire).

D. Schéma illustrant les hypothèses chimériques de l'origine des *Eucarya*. Selon ces modèles, les *Eucarya* seraient issus d'une association entre *Bacteria* et *Archaea*. C'est pourquoi ils présentent un mélange de caractères bactériens (représentés par un cercle noir avec un B) et archéens (représentés par un cercle noir avec un A) qui seraient dérivés de ces deux ancêtres (mélange représenté par un cercle hachuré noir et blanc). Ces hypothèses ne font aucune hypothèse quant à la présence de traits bactériens ou archéens chez LUCA qui est représenté par un rond gris.

résultats similaires, c'est-à-dire le placement de la racine de l'arbre universel du vivant dans la branche menant aux *Bacteria* (Figure 3A et 3B); résultats qui ont été confirmés par les analyses ultérieures de quelques autres protéines paralogues (Gribaldo et Philippe, 2002). Ceci implique que le premier événement de spéciation qui a eu lieu au niveau de LUCA correspond à la séparation de la lignée qui a donné naissance aux *Bacteria* d'une part, et à une lignée ancestrale aux *Archaea* et aux *Eucarya* d'autre part (l'apparition des lignées conduisant à ces deux domaines résulte donc d'un événement de spéciation plus tardif). Selon ce scénario,

*Archaea* et *Eucarya* sont donc plus proches parents qu'ils ne le sont des *Bacteria* (Figure 3A et 3B). Par contre, il est important de remarquer que cela ne signifie pas que les *Archaea* sont les ancêtres des *Eucarya*, erreur fréquemment retrouvée dans la littérature scientifique (Yutin *et al.*, 2008). En effet, les *Eucarya* n'émergent pas au sein des *Archaea*, mais ils en sont un groupe-frère, c'est-à-dire qu'ils partagent un ancêtre commun qui était probablement très différent des *Archaea* et *Eucarya* tels qu'on les définit actuellement. Une autre erreur fréquente est de considérer que la proche parenté *Archaea/Eucarya* implique que les caractères communs aux *Archaea*

et aux *Eucarya* sont des caractères dérivés, c'est-à-dire des caractères qui sont apparus dans la lignée qui conduit de LUCA à l'ancêtre des *Archaea* et des *Eucarya* (Figure 3A). Suivant cette logique, LUCA aurait possédé des traits bactériens (Figure 3A). Or cette lecture de l'arbre universel n'est pas correcte, car il serait tout aussi juste de proposer que LUCA possédait les caractéristiques communes aux *Archaea* et aux *Eucarya* et que les traits typiques des *Bacteria* sont apparus dans la branche bactérienne (Figure 3B). Bien que la re-analyse phylogénétique de ces protéines par des méthodes phylogénétiques plus perfectionnées suggère que l'enracinement dans la branche bactérienne puisse résulter d'artefacts de reconstruction (Gribaldo et Philippe, 2002), il est admis par une majorité de scientifiques car il postule un LUCA procaryote (cellule à l'organisation plus simple que les eucaryotes), ce qui s'accorde avec une conception de l'évolution qui va du simple au complexe, héritée de la *scala naturae* aristotélicienne.

Quelques alternatives à la proche parenté *Eucarya/Archaea* ont été proposées, comme celle des Eocytes de Lake mentionnée auparavant, ou celle de Cavalier-Smith qui considère qu'il n'existe qu'un seul domaine primaire, les *Bacteria*, les *Archaea* et les *Eucarya* représentant des domaines secondaires (Figure 3C). Selon ce scénario, les *Archaea* et les *Eucarya* sont deux groupes frères qui ont émergé à partir de bactéries à Gram positif, plus précisément à partir des *Actinobacteria*, en réponse à une pression de sélection pour la résistance aux antibiotiques, ou encore l'apparition d'histones de type *Archaea/Eucarya* pour protéger l'ADN dans des milieux hyperthermophiles (Cavalier-Smith, 2002). Une faiblesse majeure de cette hypothèse est que des exemples d'adaptation à l'hyperthermophilie à partir d'ancêtres mésophiles sont connus chez les bactéries (*Thermotogales*, *Aquificales*...) sans que ces dernières n'aient subi une évolution aussi drastique que celle qui aurait donné naissance aux *Archaea*, évoquée dans le modèle de Cavalier-Smith. Le même argument peut être employé pour la résistance aux antibiotiques.

D'autres modèles dits chimériques proposent que les *Eucarya* sont issus d'une association entre *Archaea* et *Bacteria* (Figure 3D). Ces hypothèses essaient d'expliquer l'existence de caractères partagés par les *Eucarya* et les *Archaea* d'une part, et ceux partagés par les *Eucarya* et les *Bacteria*, d'autre part. Par exemple, Lynn Margulis a proposé que l'association aurait impliqué une *Archaea* membre de la famille des *Thermoplasmatales* (qui aurait fourni le matériel génétique informationnel) et une bactérie Spirochète (qui aurait été à l'origine du flagelle eucaryote) (Martin, 2005). Cependant, les analyses génomiques viennent contredire ce scénario évolutif car très peu de gènes, évolutivement plus proches des *Thermoplasmatales* ou des *Spirochaetes*, ont été mis en évidence dans les génomes eucaryotes. Enfin, les scénarios les plus intéressants sont ceux qui proposent que les *Eucarya* sont issus d'une association syntrophique entre *Archaea* méthanogènes (donc consommatrices d'hydrogène) et une proteobacteria productrice d'hydrogène (Martin, 2005). Bien que plusieurs variantes de cette hypothèse aient été proposées, celle de López-García et Moreira est

probablement la plus complète (Lopez-Garcia et Moreira, 2006). Les hypothèses syntrophiques pour expliquer l'origine des *Eucarya* restent difficiles à tester, mais il est probable que l'amélioration des connaissances relatives aux associations microbiennes devrait apporter des informations capitales.

## CONCLUSIONS

Bien que la connaissance des *Archaea*, tant du point de vue biologique que du point de vue évolutif, soit encore parcellaire, une phylogénie stable, pouvant servir de base à l'élaboration d'une classification reflétant les relations de parenté, émerge depuis quelques années. Cependant, beaucoup de questions restent ouvertes, comme celles des relations de parenté entre les trois domaines du vivant et la nature primitive ou dérivée des caractères communs aux *Archaea* et aux *Eucarya*. L'obtention de nouvelles données génomiques, la découverte et la caractérisation de nouvelles lignées, ainsi que l'amélioration des techniques d'analyse vont certainement contribuer à y répondre.

## REFERENCES

- Barns S.M., Delwiche C.F., Palmer J.D., Pace N.R. (1996). Perspectives on archaeal diversity, thermophily and monophyly from environmental rRNA sequences. *Proc. Natl. Acad. Sci. U S A*, **93**, 9188-9193.
- Brochier-Armanet C., Boussau B., Gribaldo S., Forterre P. (2008). Mesophilic *Crenarchaeota*: proposal for a third archaeal phylum, the *Thaumarchaeota*. *Nat. Rev. Microbiol.*, **6**, 245-252.
- Cavalier-Smith T. (2002). The neomuran origin of archaeobacteria, the negibacterial root of the universal tree and bacterial megaclassification. *Int. J. Syst. Evol. Microbiol.*, **52**, 7-76.
- Elkins J.G., Podar M., Graham D.E., Makarova K.S., Wolf Y., Randau L., et al. (2008) A korarchaeal genome reveals insights into the evolution of the *Archaea*. *Proc. Natl. Acad. Sci. U S A*, **105**, 8102-8107.
- Forterre P., Gribaldo S., Brochier C. (2005). Luca: the last universal common ancestor. *Med. Sci. (Paris)*, **21**, 860-865.
- Gogarten J.P., Kibak H., Dittrich P., Taiz L., Bowman E.J., Bowman B.J., et al. (1989) Evolution of the vacuolar H<sup>+</sup>-ATPase: implications for the origin of eukaryotes. *Proc. Natl. Acad. Sci. U S A*, **86**, 6661-6665.
- Gribaldo S., Philippe H. (2002) Ancient phylogenetic relationships. *Theor. Popul. Biol.*, **61**, 391-408.
- Gribaldo S., Brochier-Armanet, C. (2006). The origin and evolution of *Archaea*: a state of the art. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, **361**, 1007-1022.
- Huet J., Schnabel R., Sentenac A., Zillig W. (1983). Archaeobacteria and eukaryotes possess DNA-dependent RNA polymerases of a common type. *The Embo Journal*, **2**, 1291-1294.
- 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. U S A*, **86**, 9355-9359.
- Lake J.A., Henderson E., Oakes M., Clark M.W. (1984). Eocytes: a new ribosome structure indicates a kingdom with a close relationship to eukaryotes. *Proc. Natl. Acad. Sci. U S A*, **81**, 3786-3790.



- Lopez-Garcia P., Moreira D. (2006) Selective forces for the origin of the eukaryotic nucleus. *Bioessays*, 28, 525-533.
- Makarova K.S., Sorokin A.V., Novichkov P.S., Wolf Y.I., Koonin E.V. (2007). Clusters of orthologous genes for 41 archaeal genomes and implications for evolutionary genomics of archaea. *Biol. Direct.*, 2, 33.
- Martin W. (2005). Archaeobacteria (*Archaea*) and the origin of the eukaryotic nucleus. *Curr. Opin. Microbiol.*, 8, 630-637.
- Martin-Cuadrado A.B., Rodriguez-Valera F., Moreira D., Alba J.C., Ivars-Martinez E., Henn M.R., *et al.* (2008) Hindsight in the relative abundance, metabolic potential and genome dynamics of uncultivated marine archaea from comparative metagenomic analyses of bathypelagic plankton of different oceanic regions. *Isme J.*, 2, 865-886.
- Nicol G.W., Schleper C. (2006). Ammonia-oxidising *Crenarchaeota*: important players in the nitrogen cycle? *Trends Microbiol.*, 14, 207-212.
- Philippe H., Laurent J. (1998). How good are deep phylogenetic trees? *Curr. Opin. Genet. Dev.*, 8, 616-623.
- Schleper C., Jurgens G., Jonscheit M. (2005). Genomic studies of uncultivated archaea. *Nat. Rev. Microbiol.*, 3, 479-488.
- Schwartz R.M., Dayhoff M.O. (1978). Origins of prokaryotes, eukaryotes, mitochondria, and chloroplasts. *Science*, 199, 395-403.
- Slesarev A.I., Mezhevaya K.V., Makarova K.S., Polushin N.N., Shcherbinina O.V., Shakhova V.V., *et al.* (2002). The complete genome of hyperthermophile *Methanopyrus kandleri* AV19 and monophyly of archaeal methanogens. *Proc. Natl. Acad. Sci. U S A*, 99, 4644-4649.
- Stanier R., van Niel C. (1962). The concept of a bacterium. *Archives of Microbiology*, 42, 17-35.
- Woese C. (2007). The birth of *Archaea*: a personal retrospective, Blackwell.
- Woese C.R., Fox G.E. (1977). Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc. Natl. Acad. Sci. U S A*, 74, 5088-5090.
- Woese C.R., Kandler O., Wheelis M.L. (1990). Towards a natural system of organisms: proposal for the domains *Archaea*, *Bacteria*, and *Eucarya*. *Proc. Natl. Acad. Sci. U S A*, 87, 4576-4579.
- Yutin N., Makarova K.S., Mekhedov S.L., Wolf Y.I., Koonin E.V. (2008). The deep archaeal roots of eukaryotes. *Mol. Biol. Evol.*, 25, 1619-1630.