

Where is the root of the universal tree of life?

Patrick Forterre^{1*} and Hervé Philippe²

Summary

The currently accepted universal tree of life based on molecular phylogenies is characterised by a prokaryotic root and the sisterhood of archaea and eukaryotes. The recent discovery that each domain (bacteria, archaea, and eucarya) represents a mosaic of the two others in terms of its gene content has suggested various alternatives in which eukaryotes were derived from the merging of bacteria and archaea. In all these scenarios, life evolved from simple prokaryotes to complex eukaryotes. We argue here that these models are biased by overconfidence in molecular phylogenies and prejudices regarding the primitive nature of prokaryotes. We propose instead a universal tree of life with the root in the eukaryotic branch and suggest that many prokaryotic features of the information processing mechanisms originated by simplification through gene loss and non-orthologous displacement. *BioEssays* 21:871–879, 1999. © 1999 John Wiley & Sons, Inc.

Introduction

For a long time, it seemed a hopeless quest to reconstruct the early evolution of life, considering the very scarce fossil record available and the paucity of useful phenotypic characters to define phylogenetic relationships between microorganisms. In the last two decades however, this way of thinking has been turned upside down, and for some time, the prevalent view has been that we have a clear-cut vision of the most ancient history.⁽¹⁾ A universal tree of life has found its way into most textbooks, which depicts a division of the living world into three domains, Archaea, Bacteria, and Eucarya, with a root splitting the Bacteria and a clade grouping Eucarya and Archaea together (the bacterial rooting) (Fig. 1a). This new vision was an apparent triumph of molecular phylogeny: the shape of the universal tree being deduced from ribosomal RNA (rRNA) sequence comparison and the root located using formerly duplicated protein coding genes (Fig. 1b).^(2–4) The new paradigms enshrined in this tree have been widely accepted, since they fitted well with the current prejudice that prokaryotes predated eukaryotes in evolution because they

are simpler and could be represented earlier in the fossil record. Indeed, prokaryotes (both archaea and bacteria) are clustered in the lower branches of this tree, much like monads in the first phylogenetic tree drawn by Haeckel in the nineteenth century.⁽⁵⁾

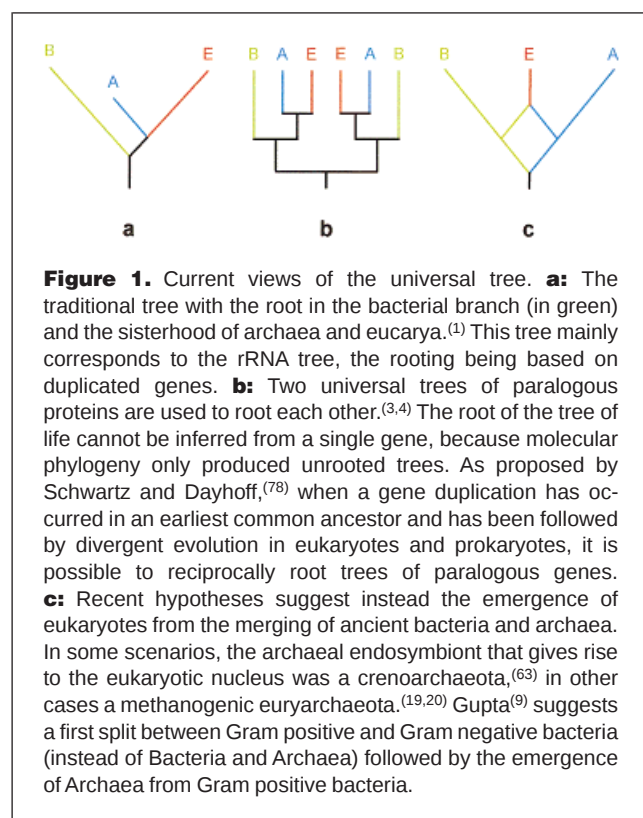
However, the current universal tree has been recently questioned on several grounds. With more and more sequences available, in particular from rapidly expanding genome projects, it turned out that many protein phylogenies contradict the rRNA tree and also each other in terms of relationships between the three domains (reviewed in ^(6–9)). Moreover, many phylogenies became confused, i.e., the three domains topology were no more recovered, with archaeal and bacterial lineages often intermixed^(8,10,11) For example, this turned out to be the case for three out of the six data sets originally used to root the universal tree (ATPases, Ile-tRNA synthetase, and carbamoyl-phosphate synthetases).⁽¹²⁾ Comparative genomics also led to the conclusion that each domain was a mosaic of the two others in terms of gene contents.^(13,14) Surprisingly, it appeared that eukaryotes contained more bacterial genes than archaeal ones, and that archaea contained more bacterial than eukaryotic ones.

The community of molecular evolutionists reacted in different ways to these data. Several authors have proposed alternative hypotheses to the tripartite division of the living world.^(9,13–20) Their scenarios consider only two primary domains and derive the third one from later merging of ancestral members of the two others (Fig. 1c). In most of these

¹Institut de Génétique et Microbiologie, Université Paris-Sud, 91405 Orsay Cedex, France.

²Phylogénie et Evolution Moléculaire, Université Paris-Sud, 91405 Orsay Cedex, France.

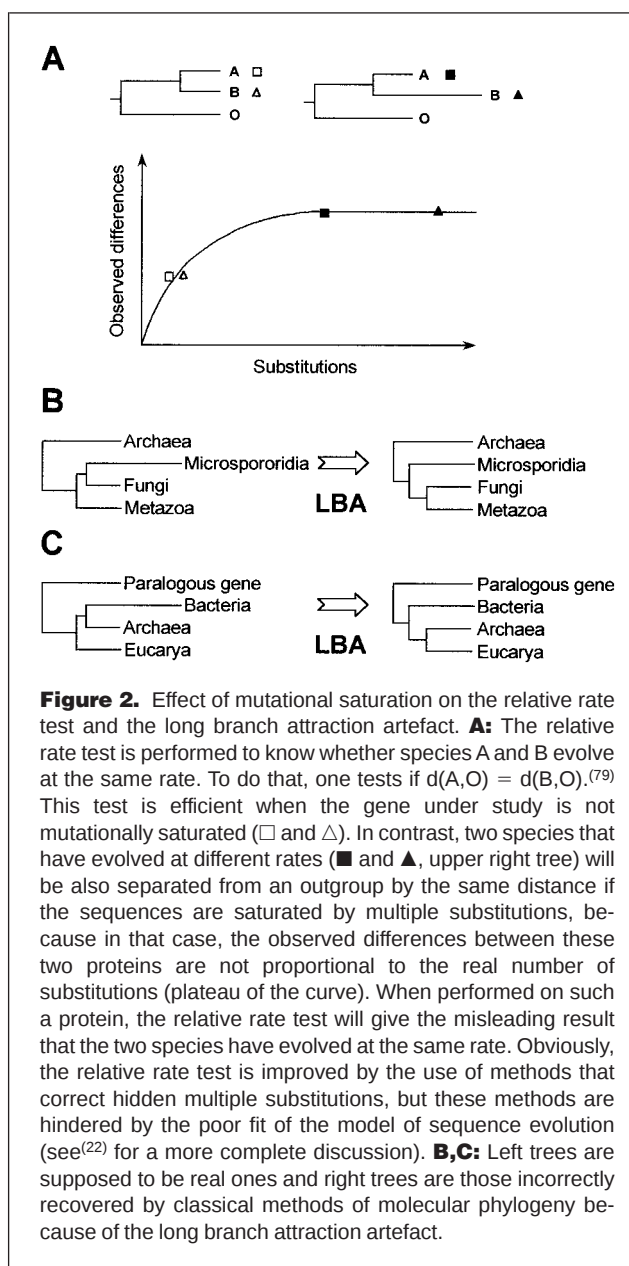
*Correspondence to: Patrick Forterre, Institut de Génétique et Microbiologie, Bat 409, Université Paris-Sud, 91405 Orsay Cedex, France.



hypotheses, the derived domain corresponds to the eukaryotes that originated from a merging of two primitive prokaryotic lineages, but the authors disagree on the nature of these lineages (see legend of Fig. 1).

Alternatively, others have favoured the view that the rRNA tree rooted in the bacterial branch is the good one, and that most, if not all, contradictions observed in other phylogenies should be explained by lateral gene transfers (LGT).^(8,21) For these authors, the choice of the “correct tree” is based on the assumption that proteins involved in translation, transcription, and replication (informational proteins, *sensu*, Rivera et al.,⁽¹⁴⁾ which are often more similar in Archaea and Eucarya, are especially informative in inferring phylogenies because they are less frequently involved in LGT. Moreover, Brown and Doolittle⁽⁶⁾ argued that these proteins have evolved at a similar rate in the three domains, through the application of the relative rate test to a large data set⁽⁹⁾ (see Fig. 2A for explanation).

The common assumption of most people who support one of the two above hypotheses (fusion or LGT) is that trees inferred from molecular phylogenetic analyses are basically correct (and thus require biological explanations). We argue here that this is not a correct assumption and that many molecular phylogenies, in particular those rooting the tree of life, do not reflect the genuine evolutionary history (see for review ⁽²²⁾). Recent studies in eukaryotic evolution support



this view and suggest an explanation for some artefactual results that are strongly supported by classical tree reconstruction methods.^(23–26) In addition, we argue that most authors focus too exclusively on LGT or cell fusion as a possible source of mosaicism between the three domains (often ignoring that mosaicism is a general rule in evolution) and do not pay enough attention to other major evolutionary forces, such as differential gene loss and non-orthologous replacement. Finally we think that most current scenarios about early evolution are highly biased by prejudicial view of prokaryotes as primitive organisms and we consider alternative hypoth-

eses based on evolution from complex to simple at the molecular level.

The crisis of molecular phylogeny

For a long time, some authors have emphasised the potential pitfalls of traditional molecular analyses (distance or parsimony) in inferring ancient phylogenies.⁽²⁷⁾ We have already argued that many contradictions between phylogenies used to infer ancient relationships do not testify for a particular evolutionary scenario, but instead for the absence of valid phylogenetic signal.^(6,7,28) We have challenged the bacterial rooting of the universal tree, noticing that sets of paralogous proteins used to root the tree were plagued with unrecognised paralogies and mostly with unequal rates of evolution. More recently, we have shown that most positions are saturated with respect to amino acid substitutions in the six data sets used to root the universal tree till now.⁽¹²⁾ This makes the relative rate test used by Brown and Doolittle⁽⁸⁾ useless in estimating the rate of protein evolution, since distances estimated from highly saturated sequences tend to be similar, even if the real numbers of substitutions are quite different⁽²²⁾ (Fig. 2A). Besides gene transfer and unrecognised paralogy, the high level of saturation, that is noise, should explain why ancient phylogenies are so often confusing and should make us wary in their interpretation. These difficulties in reconstructing ancient phylogenies are not so surprising, considering that incorrect though statistically supported results have been obtained even in molecular phylogenetic studies of mammalian evolution (see the problem of the monophyly of rodents as a case study^(28–30)).

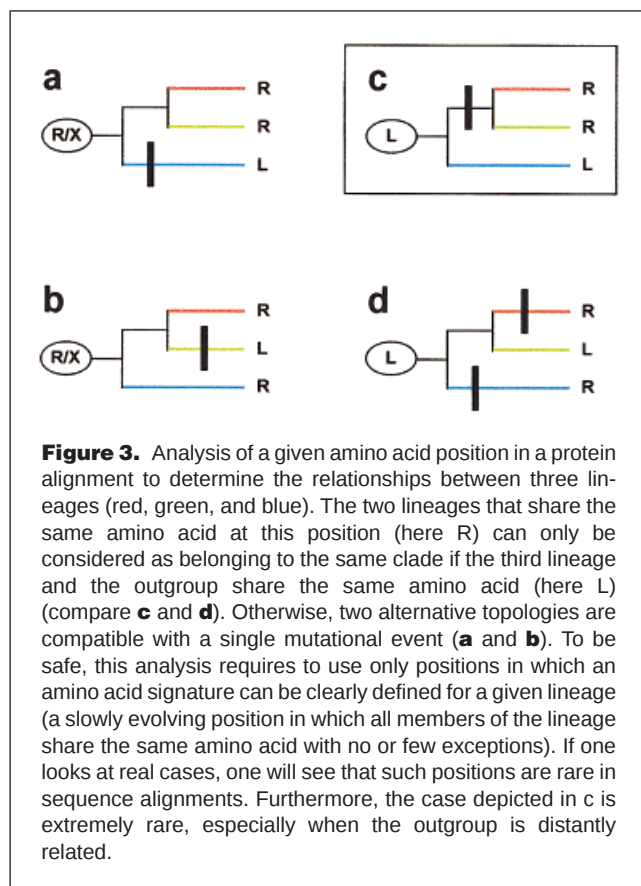
Scientists interested in deciphering ancient phylogenies were reluctant to embrace this conclusion. However, a recent dramatic turn in the study of eukaryotic phylogeny, i.e., the revision of the position of microsporidia, should precipitate a more critical approach of these questions. It was believed for the last decade that microsporidia were “primitive” eukaryotes.⁽³¹⁾ This was mainly based on their early branching in rRNA and elongation factor trees^(32,33) and apparently supported by the lack of mitochondria and the fusion of 5.8 S rRNA with 28 S rRNA (as in prokaryotes) in these organisms.⁽³⁴⁾ However, new data have shown that microsporidia are in fact closely related to fungi (if they are not fungi themselves).⁽³⁵⁾ This is supported by tubulin (α and β), Hsp70, and RNA polymerase phylogenies,^(36–39) but also by phenotypic characters such as the presence of chitin in their endospore wall⁽⁴⁰⁾ and similarities between their cell cycle and those of fungi.⁽⁴¹⁾ Moreover, microsporidia have once contained mitochondria, since they still harbour bona fide mitochondrial genes in their nuclear genome.^(38,42) The incorrect position of microsporidia in the rRNA and elongation factors trees is most likely due to an acceleration of the evolutionary rates of genes involved in the translation machinery, as suggested by their long branches in both trees. This

acceleration was probably a consequence of the streamlining of the translation apparatus in these parasitic microorganisms, which has relaxed the intermolecular constraints.^(36,38) The long branch of microsporidia produces artefactual rRNA and elongation factor trees because it is attracted by the long branch of the prokaryotic outgroup,⁽³⁸⁾ an artefact enhanced by the high saturation of these data sets (Fig. 2B).^(25,26) The long branch attraction (LBA) phenomenon⁽⁴³⁾ can be explained as follows: species that evolve faster than the others display sequences very divergent from those of their close relatives and, as a result, the fast-evolving sequences appear more distantly related from them than the species are in reality. Indeed, when the outgroup is distantly related, simulations showed that the fast-evolving species emerge too early in the tree because of the LBA artefact (Philippe, unpublished data). Interestingly, the correct position of microsporidia is recovered in the Hsp70 tree, despite of a long microsporidial branch, probably because the branch of the outgroup (the α -proteobacteria) is relatively short in this case, thus reducing the impact of the LBA.⁽³⁸⁾

One could argue that the incorrect position of microsporidia in the elongation factors and rRNA trees is the exception that confirms the rule. However, recent data suggest that it is more likely the tip of an iceberg. For example, different groups of protists with long branches turned out to be located either in the top or at the base of the eukaryotic tree, depending on the gene used as phylogenetic marker (rRNA, actin, or tubulin),⁽²⁵⁾ suggesting that all early branches observed in various eukaryotic trees are artefacts due to their high rate of evolution and LBA. The LBA phenomenon thus seems to play a role that was grossly underestimated in molecular phylogenetic analyses so far.⁽²²⁾ The LBA could in particular explain the bacterial rooting of the universal tree of life using elongation factors, since this data set is saturated, and the two longest branches are the bacterial ones and those of the outgroup (Fig. 2C).⁽¹²⁾

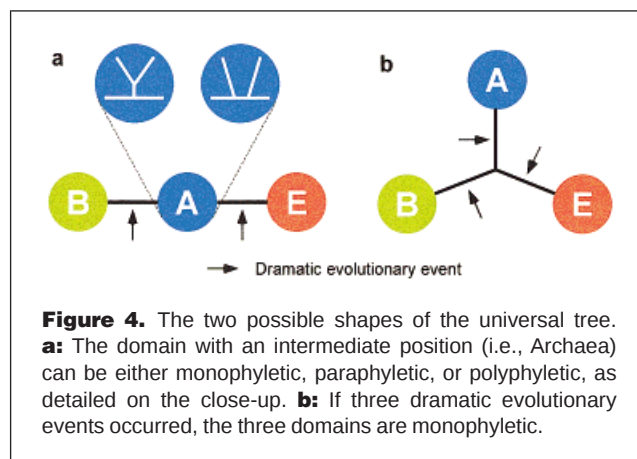
The root of the problem with molecular phylogenies

From our analyses of several examples, the critical point in explaining the frequent failure of traditional methods in inferring molecular phylogeny is the very limited number of positions in an amino acid or nucleotide sequence alignment that contain a real ancient phylogenetic signal.^(7,28,44,45) A careful phylogenetic analysis requires the polarisation of characters common to two lineages, generally using an outgroup, to decide if they testify or not for their sisterhood (for a formal treatment of cladistic analysis, see⁽⁴⁶⁾). The first step, which is not sufficiently done in traditional molecular method, is to identify good characters, i.e., characters stable in all groups studied. These characters need then to be polarised using homologous characters of an outgroup. In the case of the universal tree of life, this turned out to be



extremely difficult. At most positions, it is usually not possible to define a stable character in a given domain (an amino acid conserved in this domain) because these positions have evolved too rapidly, and when it is possible, most of the observed combinations of characters turn out to be uninformative (Fig. 3). Considering the low number of slowly evolving positions in most data sets, it turned out that most classical molecular phylogenies are inferred from fast evolving positions that are exactly those that are not reliable. For example, a single slowly evolving position supports the bacterial rooting in the elongation factor tree^(7,44) and none or a very limited number of positions in other cases. Moreover, the bacterial rooting obtained in the elongation factor tree is clearly due to fast evolving positions, strongly suggesting an LBA artefact⁽⁴⁴⁾ (see also⁽⁴⁷⁾). We have also reanalysed the phylogeny based on SRP (Signal Recognition Particle) protein with the use of a new method that concentrates the ancient phylogenetic signal.⁽⁴⁵⁾ As for elongation factors, the bacterial rooting was only supported by the fast evolving positions, in agreement with our LBA artefact hypothesis (Fig. 2C). In contrast, the slow evolving positions slightly support the eukaryotic rooting.

The paucity of good characters in sequence alignment was not so apparent at the beginning of studies in molecular



phylogeny, since it was not possible to discriminate between good and bad characters when only one or few sequences were available for a given group. However, the problem becomes obvious with the exponential accumulation of sequences.

The present crisis of molecular phylogeny will be overcome only if one recognises that the most important step in the analytical process is the search for good characters.⁽²²⁾ Since useful amino acid or nucleotidic positions are rare in sequence alignments, it will be important both to design new methods to extract these few good characters from a sequence data set⁽⁴⁴⁾ and to screen other types of characters at the molecular level such as insertions or deletions (indels) in macromolecular sequences,⁽⁹⁾ specific tri-dimensional structures, or complex and integrated molecular mechanisms. A good example of the validity of this “hennigian” approach based on character analysis is provided again by microsporidia. Indeed, the grouping of microsporidia with fungi and animals is supported by the presence of a common insertion in the sequences of the elongation factor EF-1 α from microsporidia, fungi, and metazoa.⁽³³⁾ This is striking, since traditional methods of phylogenetic analysis based on the same data set (distance, parsimony, and even maximum likelihood with improved models) all place microsporidia at the base of the eukaryotic tree.⁽³³⁾

It is often stated that molecular phylogenetic methods are intrinsically better than those based on phenotypic characters since they deal with many more characters, i.e., with all homologous positions in a sequence alignment,⁽⁴⁸⁾ and thus can be quantified and submitted to statistical tests. The case of microsporidia clearly shows that the opposite can be true, since a single good phenotypic character at the molecular level (here an insertion) can give the good answer, while the statistical treatment of hundreds of amino acids gives the wrong one. The failure of the statistical approach lies in the fact that the model of sequence evolution used is far from reflecting the real evolutionary mechanisms.⁽²²⁾

The uniqueness of the three domains

The incorrect position of microsporidia in both rRNA and elongation factor trees seriously challenges all hypotheses based on these markers, such as the three domains concept and the bacterial rooting of the universal tree of life. The three-domain concept has been indeed recently put into question by several authors who proposed alternative classifications and phylogenetic relationships between prokaryotes and eukaryotes.^(9,13) However, one cannot deny that many unique molecular features testify for the uniqueness of each domain. In particular, recent analyses of several completely sequenced archaeal and bacterial genomes, and their comparison with the yeast genome, indicate that each domain is characterised by a particular set of essential proteins, which are universally distributed in one domain and absent in the two others, as well as by a unique pattern of distribution of proteins common to two domains.^(49–51) For example, all bacteria are characterised by a unique DNA replication apparatus, involving a replicase and an initiator protein that have no homologues in the two other domains.⁽⁴⁹⁾ All archaea are characterised by a unique subset of eukaryotic-like replication and repair proteins, as well as unique proteins such as an atypical type II DNA topoisomerase (Topo VI).⁽⁵²⁾ Finally, eucarya also contain unique informational proteins such as a ubiquitous DNA topoisomerase I unrelated to their prokaryotic counterparts (for review see⁽⁵³⁾). It is especially important to emphasise the uniqueness of eukaryotes at the molecular level, since this strongly argues against current fashionable hypotheses viewing them as simply originating from the merging of archaea and bacteria. Eukaryotes contain many essential proteins that have no homologues in the two other domains, such as a plethora of transcription factors, proteins of the cytoskeleton, all components of nuclear pores, spliceosomes, polyadenylation systems, telomerases, and so on. The phenotypic coherence of each domain indicates that the three-domain concept is operational, at least in terms of classification, and reflects some ancient fundamental events in the history of life.

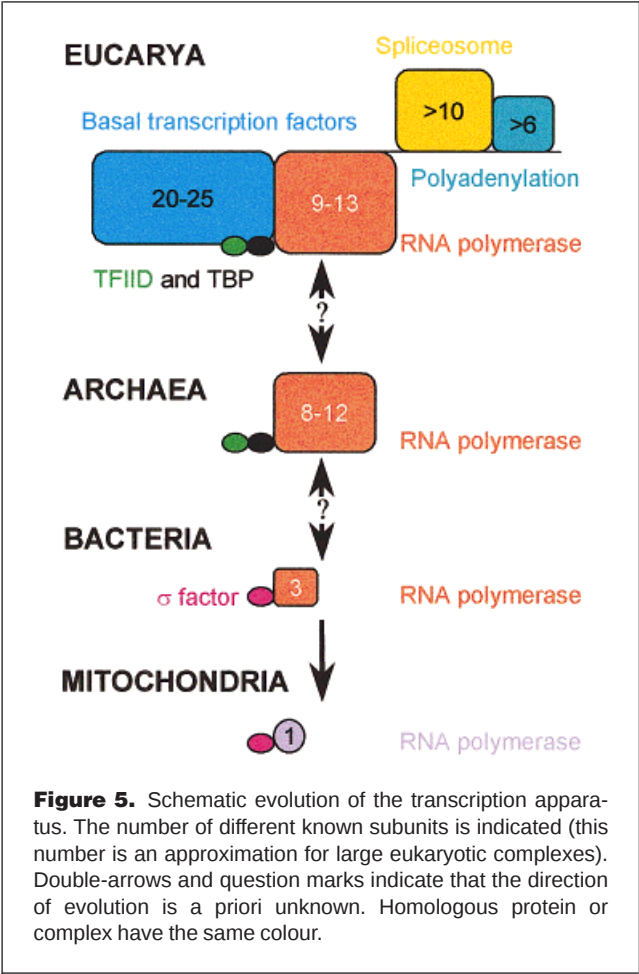
Although unique, each domain exhibits a mosaic of features present in the two others. For example, whereas many informational proteins are eukaryotic-like in Archaea (however with many exceptions), their proteins involved in metabolic pathways (operational proteins *sensu* Rivera et al.⁽¹⁴⁾) are often bacterial-like.^(54,55) The same observation can be made with eukaryotes that contain both informational proteins of archaeal type and operational proteins of bacterial type (again with many exceptions). These observations have inspired the recent popular fusion hypotheses and suggested that LGT has played an essential role in cellular evolution. One case of massive LGT is indeed well recognised. The presence of many bacterial-like genes in eukaryotes can be easily explained by a massive transfer to the nucleus of bacterial genes from the α -proteobacterial endosymbiont that

gave rise to mitochondria.⁽⁵⁶⁾ Indeed, many of these bacterial-like genes group with proteobacteria in phylogenetic tree (even when these bacterial genes are present in eucarya without organelles). Additional prokaryotic genes could easily have been acquired by eukaryotes, given the propensity of eukaryotes to eat Bacteria.⁽⁵⁷⁾ In fact, eukaryotes might even have incorporated some archaeal genes in this way, since some eukaryotes also contain archaeal endosymbionts and since for an eukaryotic predator archaea should have as good a taste as bacteria.

However, it needs to be recalled that mosaicism, which has been called heterobathmy by Hennig,⁽⁴⁶⁾ is the rule for any group of organisms when it is compared to its relatives and in general does not require LGT explanation. For example, the platypus shares characters with reptiles (laying of eggs, lack of dug) and with mammals (presence of hair, producing of milk, and jaw constituted of a single bone, the dentary) as well as it displays unique characters (beak and venomous spur). Yet nobody has suggested that this mosaic is due to LGT. In the case of the three domains, mosaicism might have been produced by a variety of processes including different rates of evolution of a character in different domains, gene duplication and gene loss producing unrecognised paralogy, or non-orthologous replacement (see below from complex to simple). In any case, it should be pointed out that these have not been of sufficient intensity and unidirectionality to obscure the uniqueness of the three domains at the molecular level.

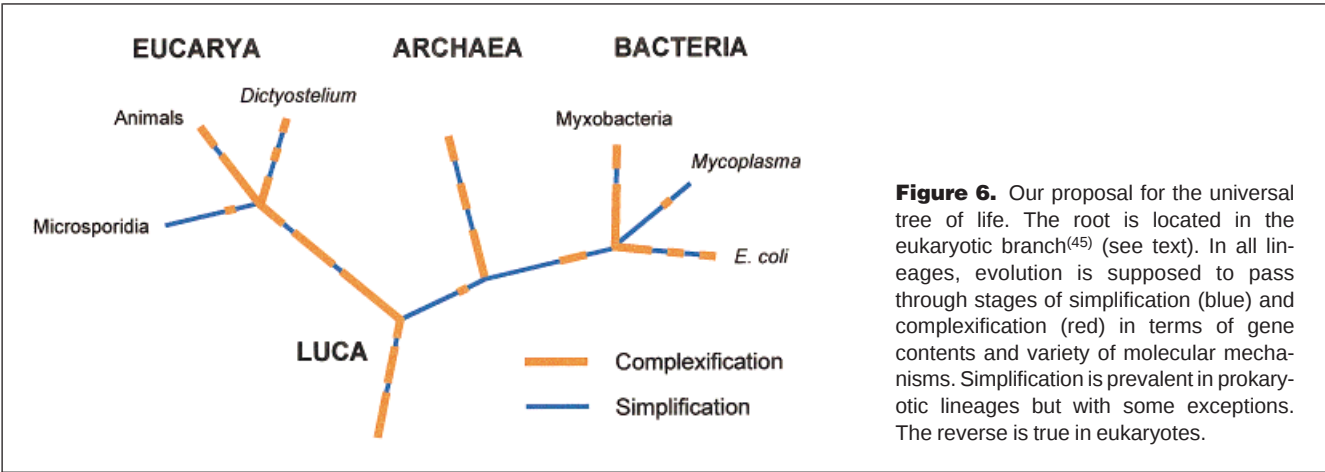
The shape of the tree

To determine the shape of the universal tree, one has first to determine if the three domains are monophyletic or paraphyletic. The monophyly of eukaryotes is not controversial since it is supported by the recognition that all of them originated from a common ancestor that already contained a bacterial mitochondrial endosymbiont.⁽⁵⁸⁾ In addition, many molecular phylogenies and numerous indel signatures support this monophyly. The monophyly of bacteria has rarely been questioned, except by Gupta,⁽⁹⁾ who derives archaea from several groups of Gram positive bacteria. However, his hypothesis is based on few phylogenies, which are biased by probable LGT.^(59,60) Furthermore, it cannot explain why archaea and Gram positive bacteria are so different at the molecular level, whilst Gram positives and Gram negatives are so similar. The monophyly of archaea is much more controversial.⁽⁶¹⁾ The rRNA tree supports this monophyly, but this could be due to an LBA effect since the bacterial and the eukaryotic branches are the longest ones in this tree. In particular, Lake⁽⁶²⁾ has argued for a long time that archaea are paraphyletic with eukaryotes deriving from an archaeal subgroup (the eocytes *sensu* Lake, crenoarchaeota *sensu* Woese). This paraphyly is strongly supported by a specific



indel in the elongation factor EF1 α present in eukaryotes and crenotes.⁽⁶³⁾

The phenotypic homogeneity of a group of organisms is sometimes used as an argument to support their monophyly. However, this can be misleading. For example, the emergence of birds and mammals led to the appearance of three phenotypically coherent groups: reptiles, birds, and mammals; but only the latter two are monophyletic, reptiles being paraphyletic! The phenotypic coherence of the three domains at the molecular level does not in fact impose a strong constraint on evolutionary scenarios except to assume two (Fig. 4a) or three (Fig. 4b) major morphological and physiological changes (dramatic evolutionary events), during which fundamental molecular mechanisms evolved independently from each others. In a scenario with three dramatic evolutionary events, the three groups are related by a trifurcation in an unrooted tree (Fig. 4b), whilst in the case of two dramatic evolutionary events, one of the three groups has an “intermediate” position between the two others, but this group can be either mono, poly, or paraphyletic (Fig. 4). The possibility that archaea have an intermediate position between the two other domains is suggested by recent data from comparative genomics. Indeed, one notices a reduction in the number of proteins involved in transcription, translation, and replication machineries, with a gradient from Eucarya to Bacteria, and Archaea in-between. For example, the replication factor that loads the DNA replicase onto the RNA primer is encoded by five homologous genes in Eucarya, two in Archaea and one in Bacteria.⁽⁴⁹⁾ The same trends can be recognised in the number of RNA polymerase subunits and basal transcription factors, of ribosomal proteins, and of translation initiation factors. This suggests an evolutionary trend from eukaryotes to bacteria or from bacteria to eukaryotes, with archaea in between. A striking feature of this process is the huge gap in complexity between archaea and eukaryotes, as exemplified by the evolution of the transcription apparatus (Fig. 5),



suggesting that extreme simplification or complexification occurred as a dramatic evolutionary event between eukaryotes and prokaryotes.

The problem of the root and the prokaryotic dogma

The second parameter to determine the shape of the tree would be to locate its root. We have already mentioned that the bacterial rooting claimed from traditional phylogenetic analyses was not reliable. It is often implicitly argued that the presence of informational proteins of eukaryotic-type in archaea supports the bacterial rooting of the universal tree.⁽⁶⁾ However, one could have a root in the eukaryotic or the archaeal branches and still have features only common to archaea and eukaryotes if these traits are primitive and have been lost in bacteria. In fact, the presence of homologous characters specific to two domains is compatible with any rooting if these characters are primitive.⁽⁶⁾ Knowing the location of the root indeed would be helpful in identifying some characters as primitive (those shared by the two groups that are not in the same clade). For example, if the root turns out to be in one of the two prokaryotic branches, this would suggest that “prokaryotic characters” are primitive, while if the root is in the eukaryotic branch, this will not tell us if they are primitive or derived.

Many authors bypass the difficulty to polarise molecular characters between the three domains by considering features present in eukaryotes to be *de facto* more recently (evolved) than the corresponding ones in prokaryotes. For these authors, the Last Universal Cellular Ancestor (LUCA) was a typical prokaryote endowed with all putative plesiomorphies common to bacteria and archaea, such as the mode of cell division, organisation of the genome and so on.⁽⁶⁴⁾ This line of reasoning comes from the prejudice of viewing prokaryotes (meaning *before nucleus*) as more primitive than eukaryotes (the prokaryotic dogma).^(65–67) However, it is not known to what extent the “simplicity” of prokaryotes is due to their “primitive characters” or to the elimination of ancient more complex characters by reductive evolution.^(66–69) There is no good reason to consider that bacterial systems are primitive as compared to archaeal and eukaryotic ones simply because they lack one or more components. They might be simpler because they have been streamlined towards more efficiency. It is striking for example that the rate of DNA chain elongation at the replication forks is about ten times higher in bacteria than in eucarya.⁽⁷⁰⁾ Accordingly, it is always possible to imagine at least two opposite scenarios when one considers the evolution between a bacterial trait and an eucaryal one (with archaea in the middle). Figure 5 shows as an example the transcription apparatus in which the simpler bacterial apparatus can be viewed either as a primitive one or as the result of a reductive evolution (as indicated by the double arrow).

From complex to simple

Both complexification and simplification have occurred during evolution (Fig. 6). For example, the evolution from the origin of life to LUCA obviously must have been globally from simple to complex, whereas the evolution from Gram positive bacteria to *Mycoplasma*, Gram negative bacteria to mitochondria and chloroplasts, fungi to yeasts or microsporidia, are examples of evolution from complex to simple. What happened in the case of the eukaryote/prokaryote transition?

Two arguments can be put forward to suggest a “complex to simple” scenario in which a eukaryotic-like LUCA (in terms of its basic molecular biology) evolved by simplification to give birth to present-day prokaryotes (Fig. 6). First, reductive evolution of central molecular mechanisms still occurs, as indicated by the fact that some proteins involved in the information processing systems and present in the three domains are missing in several lineages of a particular domain. For example, homologues of the eukaryotic translation initiation factor eIF-1 are present in all archaea and a few bacteria but are missing in the majority of known bacteria.⁽⁷¹⁾ This suggests a loss of this factor in most bacterial lineages, demonstrating a recent streamlining of the translation machinery in bacterial evolution. Second, many “non-orthologous displacements” associated to a reduction event have probably occurred during the evolution of the three domains. Koonin and co-workers⁽⁷²⁾ have coined the term “non-orthologous displacement” to design the functional replacement in a lineage of a given protein by a paralogous or unrelated protein with the same activity. Such displacements are usually assumed to be rare for proteins involved in multiple molecular interactions. However, the occurrence of such events is shown by the well documented displacement of the original proteobacterial RNA polymerase by a bacteriophage-like RNA polymerase in the evolution of mitochondria and chloroplasts (Fig. 5).^(73,74) Non-orthologous displacement could actually have been much more frequent than suspected in the evolution of the information processing apparatus, explaining for example why bacterial and eucaryal DNA replicases belong to different DNA polymerase families (C and B, respectively) whilst they interact with homologous accessory proteins. The same phenomenon could explain why homologous bacterial and archaeal RNA polymerases use non-homologous transcription initiation factors (Fig. 5), or else why bacterial IF-2 and eukaryotic eIF-2 are also not homologous, despite their central role in the initiation of translation and the fact that they interact with “refinement factors” that are homologous in the three domains.⁽⁷⁵⁾

It is unlikely that non orthologous displacement can lead to an increase in the number of components in the system, since it is very difficult to imagine the displacement of a single component by several at once. In contrast, a replacement of several components by a single one is much easier, if the latter can perform the same task with similar or better

efficiency. This actually happened in the case of mitochondrial and chloroplastic RNA polymerases, since a four subunits RNA polymerase has been displaced by a monomeric one.⁽⁷⁴⁾ Accordingly, the two archaeal transcription factors (TBP and TFIID) have more likely been replaced by a single bacterial one (the σ factor) than the reverse. The same reasoning can be used for the displacement of the three different eucaryal DNA replicases (DNA polymerases α , γ and η) by a single one at the bacterial replication forks (Pol III) and so on. These considerations make a scenario from a complex eukaryotic-like LUCA to simpler (but more efficient) prokaryotic systems more appealing than the classical hypothesis viewing prokaryotes as primitive organisms.

The hypothesis of a eukaryotic-like LUCA was proposed long time ago by Reaney and a few others^(66,67,76,77) who considers many RNA molecules typical of eukaryotes to be relics of the RNA world and thus should have been present in LUCA. This idea has been recently put forward again and elaborated by Poole and co-workers.⁽⁶⁹⁾ As a matter of fact, many eukaryotic features of the information systems involve RNA components, such as telomerase guide RNA, which are absent from both archaea and bacteria. The hypothesis of a eukaryotic-like molecular biology for LUCA is indeed appealing since it can explain both the existence of specific features in the central information processing systems of eukaryotes (those which have been lost in the prokaryotic lineage) and the existence of many proteins common to archaea and bacteria that are not found in eucarya. The presence of many eucaryal traits in archaeal proteins involved in the core of the information processing system could be easily explained by both extensive streamlining of these systems and further non-orthologous displacement in bacteria. Interestingly, such streamlining would have precisely accelerated the rate of evolution of bacterial proteins involved in these systems (as in the case of microsporidia), producing the LBA artefact responsible for the bacterial rooting obtained using traditional phylogenetic analyses (Fig. 2B and C).

Acknowledgments

We thank Philippe Lopez, David Moreira and Miklós Müller for the critical reading of the manuscript.

References

1. Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci USA* 1990;87:4576–4579.
2. Woese CR, Olsen GJ. Archaeobacterial phylogeny: perspective on the urkingdoms. *System Appl Microbiol* 1985;7:161–177.
3. Gogarten JP, Kibak H, Dittrich P, Taiz L, Bowman EJ, Bowman BJ, Manolson MF, Poole RJ, Date T, Oshima T, et al. Evolution of the vacuolar H⁺-ATPase: implications for the origin of eukaryotes. *Proc Natl Acad Sci USA* 1989;86:6661–6665.
4. Iwabe N, Kuma K, Hasegawa M, Osawa S, Miyata T. Evolutionary relationship of archaeobacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes. *Proc Natl Acad Sci USA* 1989;86:9355–9359.
5. Haeckel E. *Generelle Morphologie der Organismen: Allgemeine Grundzüge der organischen Formen-Wissenschaft, mechanisch begründet durch die von Charles Darwin reformirte Descendenz-Theorie*. Berlin: Georg Reimer. 1866.
6. Forterre P. Protein versus rRNA: rooting the universal tree of life? *ASM news* 1997;63:89–92.
7. Forterre P, Benachenhou-Lahfa N, Confalonieri F, Duguet M, Elie C, Labedan B. The nature of the last universal ancestor and the root of the tree of life, still open questions. *Biosystems* 1992;28:15–32.
8. Brown JR, Doolittle WF. Archaea and the prokaryote-to-eukaryote transition. *Microbiol Mol Biol Rev* 1997;61:456–502.
9. Gupta RS. 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* 1998;229:695–708.
10. Pennisi E. Genome data shake tree of life. *Science* 1998;280:672–674.
11. Doolittle RF. Microbial genomes opened up. *Nature* 1998;392:339–342.
12. Philippe H, Forterre P. The rooting of the universal tree of life is not reliable. *J Mol Evol* 1999; in press.
13. Koonin EV, Mushegian AR, Galperin MY, Walker DR. Comparison of archaeal and bacterial genomes: computer analysis of protein sequences predicts novel functions and suggests a chimeric origin for the archaea. *Mol Microbiol* 1997;25:619–637.
14. Rivera MC, Jain R, Moore JE, Lake JA. Genomic evidence for two functionally distinct gene classes. *Proc Natl Acad Sci USA* 1998;95:6239–6244.
15. Zillig W. Eukaryotic traits in Archaeobacteria. Could the eukaryotic cytoplasm have arisen from archaeobacterial origin? *Ann NY Acad Sci* 1987;503:78–82.
16. Sogin ML. Early evolution and the origin of eukaryotes. *Curr Opin Genet Dev* 1991;1:457–463.
17. Lake JA, Rivera MC. Was the nucleus the first endosymbiont? *Proc Natl Acad Sci USA* 1994;91:2880–2881.
18. Gupta RS, Golding GB. The origin of the eukaryotic cell. *Trends Biochem Sci* 1996;21:166–171.
19. Martin W, Muller M. The hydrogen hypothesis for the first eukaryote. *Nature* 1998;392:37–41.
20. Moreira D, Lopez-Garcia P. Symbiosis between Methanogenic Archaea and delta-proteobacteria as the origin of Eukaryotes: the syntrophic hypothesis. *J Mol Evol* 1998;47:517–530.
21. Woese C. The universal ancestor. *Proc Natl Acad Sci USA* 1998;95:6854–6859.
22. Philippe H, Laurent J. How good are deep phylogenetic trees? *Curr Opin Genet Dev* 1998;8:616–623.
23. Germot A, Philippe H. Critical analysis of eukaryotic phylogeny: a case study based on HSP70 family. *J Euk Microbiol* 1999; in press.
24. Budin K, Philippe H. New insights into the phylogeny of eukaryotes based on ciliate HSP70 sequences. *Mol Biol Evol* 1998;15:943–956.
25. Philippe H, Adoutte A. The molecular phylogeny of Eukaryota: solid facts and uncertainties. In: Coombs G, Vickerman K, Sleigh M, Warren A, editors. *Evolutionary relationships among Protozoa*. London: Chapman & Hall; 1998. p 25–56.
26. Roger AJ, Sandblom O, Doolittle WF, Philippe H. An evaluation of elongation factor 1a as a phylogenetic marker for eukaryotes. *Mol Biol Evol* 1999;16:218–233.
27. Meyer TE, Cusanovich MA, Kamen MD. Evidence against use of bacterial amino acid sequence data for construction of all-inclusive phylogenetic trees. *Proc Natl Acad Sci USA* 1986;83:217–220.
28. Philippe H. Rodent monophyly: pitfalls of molecular phylogenies. *J Mol Evol* 1997;45:712–715.
29. Cao Y, Okada N, Hasegawa M. Phylogenetic position of guinea pigs revisited. *Mol Biol Evol* 1997;14:461–464.
30. Sullivan J, Swofford DL. Are guinea pigs rodents? The importance of adequate models in molecular phylogenetics. *J Mam Evol* 1997;4:77–86.
31. Cavalier-Smith T. Eukaryotes with no mitochondria. *Nature* 1987;326:332–333.
32. Vossbrinck CR, Maddox JV, Friedman S, Debrunner-Vossbrinck BA, Woese CR. Ribosomal RNA sequence suggests microsporidia are extremely ancient eukaryotes. *Nature* 1987;326:411–414.

33. Kamaishi T, Hashimoto T, Nakamura Y, Masuda Y, Nakamura F, Okamoto K, Shimizu M, Hasegawa M. Complete nucleotide sequences of the genes encoding translation elongation factors 1 alpha and 2 from a microsporidian parasite, *Glugea plecoglossi*: implications for the deepest branching of eukaryotes. *J Biochem (Tokyo)* 1996;120:1095–1103.
34. Vossbrinck CR, Woese CR. Eukaryotic ribosomes that lack a 5.8S RNA. *Nature* 1986;320:287–288.
35. Müller M. What are the Microsporidia? *Parasitology Today* 1997;13:455–456.
36. Edlind TD, Li J, Visvesvara GS, Vodkin MH, McLaughlin GL, Katiyar SK. Phylogenetic analysis of beta-tubulin sequences from amitochondrial protozoa. *Mol Phylogenet Evol* 1996;5:359–367.
37. Keeling PJ, Doolittle WF. Alpha-tubulin from early-diverging eukaryotic lineages and the evolution of the tubulin family. *Mol Biol Evol* 1996;13:1297–1305.
38. Germot A, Philippe H, Le Guyader H. Evidence for loss of mitochondria in Microsporidia from a mitochondrial-type HSP70 in *Nosema locustae*. *Mol Biochem Parasitol* 1997;87:159–168.
39. Hirt RP, Logsdon JM, Jr, Healy B, Dorey MW, Doolittle WF, Embley TM. Microsporidia are related to fungi: evidence from the largest subunit of RNA polymerase II and other proteins. *Proc Natl Acad Sci USA* 1999;96:580–585.
40. Canning EU. Phylum Microsporidiae. In: Margulis L, Corliss JO, Melkonian M, Chapman DJ, editors. *Handbook of protista*. Boston: Jones and Bartlett Publishers; 1990. p 53–72.
41. Flegel TW, Pasharawipat T. A proposal for typical eukaryotic meiosis in microsporidians. *Can J Microbiol* 1995;41:1–11.
42. Hirt RP, Healy B, Vossbrinck CR, Canning EU, Embley TM. A mitochondrial Hsp70 orthologue in *Vairimorpha necatrix*: molecular evidence that microsporidia once contained mitochondria. *Curr Biol* 1997;7:995–998.
43. Felsenstein J. Cases in which parsimony or compatibility methods will be positively misleading. *Syst Zool* 1978;27:401–410.
44. Lopez P, Forterre P, Philippe H. The root of the tree of life in the light of the covarian model. *J Mol Evol* 1999; in press.
45. Brinkmann H, Philippe H. Archaea sister-group of Bacteria? Indications from tree reconstruction artefacts in ancient phylogenies. *Mol Biol Evol* 1999; in press.
46. Hennig W. *Phylogenetic systematics*. Urbana: University of Illinois Press. 1966.
47. Lake JA. Optimally recovering rate variation information from genomes and sequences: pattern filtering. *Mol Biol Evol* 1998;15:1224–1231.
48. Graur D. Towards a molecular resolution of the ordinal phylogeny of the eutherian mammals. *FEBS letter* 1993;325:152–159.
49. Edgell DR, Doolittle WF. Archaea and the origin(s) of DNA replication proteins. *Cell* 1997;89:995–998.
50. Forterre P. Archaea: what can we learn from their sequences? *Curr Opin Genet Dev* 1997;7:764–770.
51. Olsen GJ, Woese CR. Archaeal genomics: an overview. *Cell* 1997;89:991–994.
52. Bergerat A, de Massy B, Gadelle D, Varoutas PC, Nicolas A, Forterre P. An atypical topoisomerase II from Archaea with implications for meiotic recombination. *Nature* 1997;386:414–417.
53. Forterre P, Bergerat A, Lopez-Garcia P. The unique DNA topology and DNA topoisomerases of hyperthermophilic archaea. *FEMS Microbiol Rev* 1996;18:237–248.
54. Bult CJ, White O, Olsen GJ, Zhou L, Fleischmann RD, Sutton GG, Blake JA, FitzGerald LM, Clayton RA, Gocayne JD, et al. Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. *Science* 1996;273:1058–1073.
55. Klenk HP, Clayton RA, Tomb JF, White O, Nelson KE, Ketchum KA, Dodson RJ, Gwinn M, Hickey EK, Peterson JD, et al. The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*. *Nature* 1997;390:364–370.
56. Martin W, Herrmann RG. Gene transfer from organelles to the nucleus: how much, what happens, and why? *Plant Physiol* 1998;118:9–17.
57. Doolittle WF. You are what you eat: a gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes. *Trends Genet* 1998;14:307–311.
58. Hashimoto T, Sanchez LB, Shirakura T, Muller M, Hasegawa M. Secondary absence of mitochondria in *Giardia lamblia* and *Trichomonas vaginalis* revealed by valyl-tRNA synthetase phylogeny. *Proc Natl Acad Sci USA* 1998;95:6860–6865.
59. Roger AJ, Brown JR. A chimeric origin for eukaryotes re-examined. *Trends Biochem Sci* 1996;21:370–372.
60. Philippe H, Budin K, Moreira D. Horizontal transfers confuse the prokaryotic phylogeny based on the HSP70 protein family. *Molecular Microbiology* 1999;31:1007–1009.
61. Baldauf SL, Palmer JD, Doolittle WF. The root of the universal tree and the origin of eukaryotes based on elongation factor phylogeny. *Proc Natl Acad Sci USA* 1996;93:7749–7754.
62. Lake JA. Prokaryotes and archaeobacteria are not monophyletic: rate invariant analysis of rRNA genes indicates that eukaryotes and eocytes form a monophyletic taxon. *Cold Spring Harb Symp Quant Biol* 1987;52:839–846.
63. Rivera MC, Lake JA. Evidence that eukaryotes and eocyte prokaryotes are immediate relatives. *Science* 1992;257:74–76.
64. Doolittle WF. Fun with genealogy. *Proc Natl Acad Sci USA* 1997;94:12751–12753.
65. Forterre P. Neutral terms. *Nature* 1992;335:305.
66. Reaney DC. On the origin of prokaryotes. *Theor Biol* 1974;48:243–251.
67. Carlisle M. Prokaryotes and eukaryotes: strategies and successes. *TIBS* 1982;7:128–130.
68. Forterre P. Thermoreduction, a hypothesis for the origin of prokaryotes. *C R Acad Sci III* 1995;318:415–422.
69. Poole AM, Jeffares DC, Penny D. The path from the RNA world. *J Mol Evol* 1998;46:1–17.
70. Adams RLP. DNA replication. In: Rickwood D, editor. *Focus*, Oxford: IRL press; 1991. p 1–85.
71. Kyrpides NC, Woese CR. Universally conserved translation initiation factors. *Proc Natl Acad Sci USA* 1998;95:224–228.
72. Koonin EV, Mushegian AR, Bork P. Non-orthologous gene displacement. *Trends Genet* 1996;12:334–336.
73. Cermakian N, Ikeda TM, Miramontes P, Lang BF, Gray MW, Cedergren R. On the evolution of the single-subunit RNA polymerases. *J Mol Evol* 1997;45:671–681.
74. Gray MW, Lang BF. Transcription in chloroplasts and mitochondria: a tale of two polymerases. *Trends Microbiol* 1998;6:1–3.
75. Kyrpides NC, Woese CR. Archaeal translation initiation revisited: the initiation factor 2 and eukaryotic initiation factor 2B alpha-beta-delta subunit families. *Proc Natl Acad Sci USA* 1998;95:3726–3730.
76. Doolittle W. Genes in pieces: were they ever together? *Nature* 1978;272:581–582.
77. Darnell JE. Implications of RNA-RNA splicing in evolution of eukaryotic cells. *Science* 1978;202:1257–1260.
78. Schwartz RM, Dayhoff MO. Origins of prokaryotes, eukaryotes, mitochondria, and chloroplasts. *Science* 1978;199:395–403.
79. Sarich VM, Wilson AC. Generation time and genomic evolution in primates. *Science* 1973;179:1144–1147.