17

Comparative genomics and early cell evolution

Antonio Lazcano

Introduction

The awareness that genes and genomes are extraordinarily rich historical documents from which a wealth of evolutionary information can be retrieved has widened the range of phylogenetic studies to previously unsuspected heights. The development of efficient sequencing techniques, which now allows the rapid sequencing of complete cellular genomes, combined with the simultaneous and independent blossoming of computer science, has led not only to an explosive growth of databases and new sophisticated tools for their exploitation, but also to the recognition that, in spite of many lateral gene-transfer events (LGT), different macromolecules are uniquely suited as molecular chronometers in the construction of nearly universal phylogenies.

Cladistic analysis of rRNA sequences is acknowledged as a prime force in systematics and from its very inception had a major impact to our understanding of early cellular evolution. The comparison of small-subunit ribosomal-RNA (16/18S rRNA) sequences led to the construction of a trifurcated unrooted tree in which all known organisms can be grouped in one of three major monophyletic cell lineages, i.e. the domains bacteria (eubacteria), Archaea (Archaeabacteria) and eucarya (eukaryotes) (Woese *et al*., 1990), which are all derived from an ancestral form, known as the last common ancestor (LCA).

From a cladistic viewpoint, the LCA is merely an inferred inventory of features shared among extant organisms, all of which are located at the tip of the branches of molecular phylogenies. Some time ago it was surmized that the sketchy picture developed with the limited databases would be confirmed by completely sequenced cell genomes from the three primary domains. This has not been the case: the availability of an increasingly large number of completely sequenced cellular genomes has sparked new debates, rekindling the discussion on the nature of the ancestral entity and its predecessors. As reviewed by Becerra *et al*. (2007), this is shown in the diversity of names that have been coined to describe it, which include, among others, progenote (Woese and Fox, 1977), cenancestor (Fitch and Upper, 1987), LUCA, last universal common ancestor (Kyrpides *et al*., 1999) or, later on, last universal cellular ancestor (Philippe and Forterre, 1999), universal ancestor (Doolittle, 2000), LCC, last common community (Line, 2002), and MRCA, most recent common ancestor (Zhaxybayeva and Gogarten, 2004). Close analysis shows that these

terms are not truly synonymous, and in a way they reflect the controversies on the nature of the universal ancestor and the evolutionary processes that shaped it.

Analysis of homologous traits found among the three major lineages suggest that the LCA was not a direct immediate descendant of the RNA World, a protocell or any other pre-life progenitor system. In fact, given the huge gap existing in current descriptions of the evolutionary transition between the prebiotic synthesis of biochemical compounds and the LCA, it is difficult to see how the universal trees can be extended beyond a threshold that corresponds to a period of cellular evolution in which protein biosynthesis was already in operation (Becerra *et al*., 2007).

However, from an evolutionary point of view it is reasonable to assume that at some point in time the ancestors of all forms of life must have been less complex than even the simpler extant cells. Although it is naive to attempt to describe the origin of life and the nature of the first living systems from the available rooted phylogenetic trees, molecular cladistics may provide clues to some very early stages of biological evolution. Indeed, the variations of traits common to extant species can be explained as the outcome of divergent processes from an ancestral life form that existed prior to their separation into the three major biological domains. The purpose of this chapter is to discuss such early stages of cellular evolution, i.e. to discuss some of the possible steps that took place between the origin of life and the last common ancestor of all living beings.

Are there fossil records of the cenancestor?

Although no evolutionary intermediate stages or ancient simplified versions of the basic biological processes have been discovered in contemporary organisms, the differences in the structure and mechanisms of gene expression and replication among the three lineages have provided insights on the stepwise evolution of the replication and translational apparatus, including some late steps in the development of the genetic code. We are now in a position in which it is possible to distinguish the origin of life problem from a whole series of other issues, often confused, that belong to the domain of the evolution of microbial life. Accordingly, the most basic questions pertaining to the origin of life relate to simpler replicating entities predating by a long series of evolutionary events the oldest lineages represented by basal branches in universal molecular phylogenies.

Recognition of the significant differences that exist between the transcriptional and translational machineries of the bacteria, Archaea and eucarya, which were assumed to be the result of independent evolutionary refinements, led to the conclusion that the primary branches were the descendants of a progenote, a hypothetical biological entity in which phenotype and genotype still had an imprecise rudimentary linkage relationship (Woese and Fox, 1977). However, the conclusion that the LCA was a progenote was disputed when the analysis of homologous traits found among some of its descendants suggested that it was not a direct immediate descendant of the RNA World, a protocell or any other pre-life progenitor system. Under the implicit assumption that lateral gene transfer (LGT) had not

been a major driving force in the distribution of homologous traits in the three domains, it was concluded that the LCA was already like extant bacteria (Lazcano *et al*., 1992).

Variations of traits common to the domains bacteria, Archaea and eucarya (Woese *et al*., 1990) can be easily explained as the outcome of divergent processes from ancestral life forms older than the LCA. No paleontological remains will bear testimony of such entity, as the search for a fossil of the cenancestor is bound to prove fruitless. Indeed, there is little or no geological evidence for the environmental conditions on the early Earth at the time of the origin and early evolution of life. It is not possible to assign a precise chronology to the origin and earliest evolution of cells, and identification of the oldest paleontological traces of life remains a contentious issue. The early Archean geological record is scarce and controversial, and most of the sediments preserved from such times have been metamorphosed to a considerable extent. Although the biological origin of the microstructures present in the 3.5×10^9 years Apex cherts of the Australian Warrawoona formation (Schopf, 1993) has been disputed, at this time the weight of evidence favours the idea that life existed 3.5 billion years ago (Altermann and Kazmierczak, 2003).

Isotopic-fractionation data and other biomarkers support the possibility of a metabolically diverse Archean microbial biosphere, which may have included members of the archaeal kingdom. The proposed timing of the onset of microbial methanogenesis based on the low ^{13}C values in methane inclusions found in hydrothermally precipitated quartz in the 3.5-billion-year-old Dresser Formation in Australia (Ueno *et al*., 2006) has been challenged (Lollar and McCollom, 2006). However, sulphur isotope investigations of the same site indicate biological sulphate-reducing activity (Shen *et al*., 2001), and analyses of 3.4×10^9 million-year-old South African cherts suggest that they were inhabited by anaerobic photosynthetic prokaryotes in a marine environment (Tice and Lowe, 2004). These results support the idea that the Early Archean Earth was teeming with prokaryotes, which included anoxygenic phototrophs, sulfate reducers and, perhaps, methanogenic Archaea (Canfield, 2006).

Molecular phylogenies and the last common ancestor

In principle, determination of the evolutionary polarity of character states in universal phylogenies should lead to the recognition of the oldest phenotype. Accordingly, the most parsimonious characterization of the LCA can be achieved by proceeding backwards and summarizing the features of the oldest recognizable node of the universal cladogram, i.e. rooting of the universal tree would provide direct information on the nature of the LCA. However, the plesiomorphic traits found in the space defined by rRNA sequences allow the construction of topologies that specify branching relationships but not the position of the ancestral phenotype. This issue was solved independently by Iwabe *et al*. (1989) and Gogarten *et al*. (1989), who analysed paralogous genes encoding (1) the two elongation factors (EF-G and EF-Tu) that assist in protein biosynthesis; and (2) the α- and β-hydrophilic subunits of F-type ATP synthetases. Using different tree-constructing algorithms, both teams independently placed the root of the universal trees between the eubacteria on the

one side, and the Archaea and the eukaryotic nucleocytoplasm on the other. The rooting of universal phylogenies placed the LCA in the bacterial branch of the universal tree (Gogarten *et al*., 1989; Iwabe *et al*., 1989).

The rooting of universal cladistic trees determines the directionality of evolutionary change and allows for the recognition of ancestral from derived characters. Determination of the root normally imparts polarity to most or all characters. It is, however, important to distinguish between ancient and primitive organisms. Organisms belonging to lineages located near the root of universal rRNA-based trees are cladistically ancient, but they are not endowed with primitive molecular-genetic apparatus, nor do they appear to be more rudimentary in their metabolic abilities than their aerobic counterparts (Islas *et al*., 2003). Primitive living systems would initially refer to pre-RNA Worlds, in which life may have been based on polymers using backbones other than ribose–phosphate and possibly bases different from adenine, uracil, guanine and cytosine (Levy and Miller, 1998), followed by a stage in which life was based on RNA as both the genetic material and as catalysts (Joyce, 2002).

Reticulate phylogenies resulting from LGT complicate the reconstruction of cenancestral traits. Driven in part by the impact of lateral gene acquisition, as revealed by the discrepancies of different gene phylogenies with the canonical rRNA tree, and in part by the surprising complexity of the universal ancestor as suggested by direct-backtrack characterizations, Woese (1998) proposed that the LCA was not a single organismic entity, but rather a highly diverse population of metabolically complementary cellular progenotes which occupied as a whole the root of the tree and that were endowed with multiple, small linear chromosome-like genomes engaged in massive multidirectional horizontal-transfer events.

According to Woese's (1998) model, the essential features of translation and the development of metabolic pathways took place before the earliest branching event. As sequence exchange decreased, the resulting genetic isolation lead, eventually, to bacteria, Archaea and eucarya (Woese, 1998). However, even if the genetic entities that formed such a communal ancestor may have been extremely diverse, their common features, such as the genetic code and the gene-expression machinery, are an indication of their ultimate monophyletic origin (Delaye *et al*., 2004). It is not necessary to assume that the processes that led to the three domains took place inmediatly after the appearance of the code and the establishment of translation. Inventories of LCA genes include sequences that originated in different precenancestral epochs (Delaye and Lazcano, 2000; Anantharaman *et al*., 2002). The origin of the mutant sequences ancestral to those found in all extant species, and the divergence of the bacteria, Archaea and eucarya were not synchronous events, i.e. the separation of the primary domains took place later, perhaps even much later than the appearance of the genetic components of their last common ancestor (Delaye *et al*., 2005).

Was the last common ancestor a hyperthermophile?

The examination of the prokaryotic branches of unrooted rRNA trees had already suggested that the ancestors of both bacteria and Archaea were extreme thermophiles growing optimally at temperatures in the range of 90 °C or above (Achenbach-Richter *et al*.,

1987). Rooted universal phylogenies appeared to confirm this possibility, since heat-loving prokaryotes occupied short branches in the basal portion of molecular cladograms (Stetter, 1994). However, the recognition that the deepest branches in rooted universal phylogenies are occupied by hyperthermophiles does not provide, by itself, conclusive proof of a heat-loving LCA, or much less of a hot origin of life. Analysis of the correlation of the optimal growth temperature of prokaryotes and the G + C nucleotide content of 40 rRNA sequences through a complex Markov model has led Galtier *et al*. (1999) to conclude that the universal ancestor was a mesophile. Further refinements using a model that included both rRNA and highly conserved protein sequences suggest more complex evolutionary history, going from a mesophilic LCA into parallel convergent adaptations to high-temperature regimes in both the ancestors of bacteria and the Archaea. These changes, which may be due to climate changes in the Precambrian Earth, were then lost independently in both lineages (Bousseau *et al*., 2008).

The chemical nature of the cenancestral genome

Since all extant cells are endowed with DNA genomes, the most parsimonious conclusion is that this genetic polymer was already present in the cenancestral population. Although it is possible to recognize the evolutionary relatedness of various orthologous proteins involved with DNA replication and repair (ATP-dependent clamp-loader proteins, topoisomerases, gyrases and 5′-3′ exonucleases) across the entire phylogenetic spectrum (Olsen and Woese, 1997; Edgell and Doolittle, 1997; Leipe *et al*., 1999; Penny and Poole, 1999), comparative proteome analysis has shown that (eu)bacterial replicative polymerases and primases lack homologues in the two other primary kingdoms.

The peculiar distribution of the DNA replication machinery has led to suggestions not only of an LCA endowed with an RNA genome, but also of the polyphyletic origins of DNA and many of the enzymes associated with its replication (Leipe *et al*., 1999; Koonin and Martin, 2005) in which viruses may have played a central role (Forterre, 2006; Koonin, 2009). Its has been argued by Koonin and Martin (2005) that the LCA was an acellular entity endowed with high numbers of RNA viral-like molecules that had originated abiotically within the cavities of a hydrothermal mound. This idea, which has little, if any, empirical support, does not take into account the problems involved with the abiotic synthesis and accumulation of ribonucleotides and polyribonucleotides, nor does it explain the emergence of functional RNA molecules.

It has also been suggested that the ultimate origins of cellular DNA genomes lie in viral systems, which gave rise to polyphyletic deoxyribonucleotide biosyntheses (Forterre, 2006). According to a rather complex hypothetical scheme, gene transfers mediated by viral takeovers took place three times, giving origin to the DNA genomes of the three primary kingdoms. The invasion of the ancestor of the bacterial domain by a DNA virus eventually led to a replacement of its cellular RNA genes by DNA sequences, while the archaeal and eucaryal DNA-replication enzymes resulted from an invasion by closely related DNA viruses (Forterre, 2006). This proposal is based not only on the assumption that metabolic

pathways can arise in viruses, but also on the possibility of polyphyletic origin of deoxyribonucleotide biosyntheses. This is unlikely: in sharp contrast with other energetically favourable biochemical reactions (such as phosphodiester-backbone hydrolysis or the transfer of amino groups), the direct removal of the oxygen from the 2′-C ribonucleotide pentose ring to form the corresponding deoxy-equivalents is a thermodynamically much less-favoured reaction, considerably reducing the likelihood of multiple independent origins of biological ribonucleotide reduction.

There are indeed manifold indications that RNA genomes existed during early stages of cellular evolution (Lazcano *et al*., 1988) but it is likely that double-stranded DNA genomes had become firmly established prior to the divergence of the three primary domains. It is true that the demonstration of the monophyletic origin of ribonucleotide reductases (RNRs) is greatly complicated by their highly divergent primary sequences and the different mechanisms by which they generate the substrate 3′-radical species required for the removal of the 2′-OH group. However, sequence analysis and biochemical characterization of archaeal RNRs have shown their similarities with their bacterial and eucaryal counterparts, confirming their ultimate monophyletic origin (Stubbe *et al*., 2001).

Strong selection pressures acting over multigene RNA-based mechanisms, such as translation, are responsible for their universal distribution and high conservation, and DNA-based genetic systems were selected for in cells to stabilize earlier RNA replicating systems. This replacement was the outcome of the metabolic evolution of deoxyribonucleotide biosynthesis via the reductive elimination of the 2′-hydroxyl group in ribonucleotides, thymine anabolism and the replacement of uracil, and the selection of editing mechanisms. Indeed, the sequence similarities shared by many ancient large proteins found in all three domains suggest that considerable fidelity existed in the operative genetic system of the LCA. Despite claims to the contrary (Poole and Logan, 2005), such fidelity is unlikely to be found in RNA-based genetic systems (Reanney, 1987; Lazcano *et al*., 1992), which do not replicate using the multiunit cellular DNA-dependent RNA polymerase, but are based on RNA replicases lacking editing mechanisms.

Gene duplications and the evolution of metabolism

Clues to the genetic organization and biochemical complexity of primitive entities from which the LCA evolved may also be derived from the analysis of paralogous gene families. The number of sequences that have undergone such duplications prior to the divergence of the three lineages includes genes encoding for a variety of enzymes that participate in widely different processes such as translation, DNA replication, biosynthetic pathways and energy-producing processes.

Some authors have argued that LCA was an acellular entity, arising directly from abiotic processes (Koga *et al*., 1998; Koonin and Martin, 2005). However, the high conservation and distribution of many membrane proteins, including, for instance, ATPase hydrophilic subunits (Gogarten *et al*., 1989), the signal-recognition particles (SRPs) (Gribaldo and Cammarano, 1998) and ABC transporters (Delaye *et al*., 2005) imply a cellular cenancestor, whose

membrane may have been formed by heterochiral lipids composed of a mixture of glycerol-1-phosphate and glycerol-3-phosphate (Wächstershäuser, 2003; Peretó *et al*., 2004).

The conservation of membrane-bound proton-pump ATPase subunits suggests that the cenancestor was able to produce a chemically driven protein gradient across its cell membrane using a variety of oxidized inorganic molecules as molecular acceptors (Castresana and Moreira, 1999), while the high conservation of manifold ABC transporters involved in the import of metabolic substrates is consistent with the possibility of a heterotrophic LCA that depended on external sources of organic compounds (Becerra *et al*., 2007).

As discussed elsewhere (Becerra *et al*., 2007), a survey of the available information shows that sequences that have resulted from early pre-ancestral paralogous expansion may be classified into three major groups:

(1) sequences formed by two tandemly arranged homologous modules which underwent fusion events, such as the (i) protein disulfide oxidoreductase (Ren *et al*., 1998); (ii) large subunit of carbomoyl phosphate synthetase (Alcántara *et al*., 2000); and (iii) HisA, an isomerase involved in histidine biosynthesis (Alifano *et al*., 1996);
(2) gene families which have undergone a major expansion of sequences, such as ABC transporters and other enzymes involved in membrane-transport phenomena (Clayton *et al*., 1997); and
(3) families formed by a relatively small number of paralogous sequences. These include, among others, the pair of homologous genes encoding the EF-Tu and EF-G elongation factors, (Iwabe *et al*., 1989) as well as the duplicated sequences encoding the F-type ATPase hydrophilic α- and β-subunits (Gogarten *et al*., 1989).

The identification of sequences formed by tandemly fused homologous modules provides direct evidence of the existence, during early Precambrian times, of smaller functional genes. Moreover, the families of paralogous duplicates also imply that the LCA was preceded by simpler cells with a smaller genome in which only one copy of each of these genes existed, i.e. by cells in which, for instance, protein synthesis involved only one elongation factor, and with ATPases with limited regulatory abilities. Paralogous families of metabolic genes also support the proposal that anabolic pathways were assembled by the recruitment of primitive enzymes that could react with a wide range of chemically related substrates, i.e. the so-called patchwork assembly of biosynthetic routes (Jensen, 1976; Velasco *et al*., 2002). Such relatively slow unspecific enzymes may have represented a mechanism by which primitive cells with small genomes could have overcome their small coding abilities. How early cells could overcome the bottlenecks imposed by such limitations is an open problem that could be addressed, for instance, by using in vitro systems of anabolic pathways, such as histidine biosynthesis, in which the different homologous enzymes that catalyze several different steps are replaced by one single representative of such a paralogous set.

Genomic evidence for an RNA/protein world

As demonstrated by other analyses, proteins that interact with RNA in one way or another are among the most highly conserved sequences (Delaye and Lazcano, 2000; Anantharaman

et al., 2002). A significant percentage of such highly conserved sequences correspond to proteins that interact directly with RNA (such as ribosomal proteins, DEAD-type helicases, aminoacyl tRNA synthetases and elongation factors, among others), or take part in RNA and nucleotide biosyntheses, including the DNA-dependent RNA polymerase β- and β′-subunits, dimethyladenosine transferase, adenyl-succinate lyases, dihydroorotate oxidase and ribose-phosphate pyrophosphokinase, among many others. Few metabolic genes are part of the conserved ORF product set. These include many sugar metabolism-related sequences, such as the enolase-encoding genes noted above, as well as homologs of sequences involved in nucleotide biosynthesis, such as thioredoxin and phosphoribosyl-pyrophosphate synthase (Delaye *et al*., 2005; Becerra *et al*., 2007).

Although RNA hydrolysis is an exergonic process, degradosome-mediated mRNA turnover plays a key role as a regulatory mechanism for gene expression in both prokaryotes and eukaryotes (Blum *et al*., 1997). A possible explanation for the vey high conservation of DEAD-type RNA helicases may lie in their role in protein biosynthesis and in mRNA degradation. This possibility is supported by the phylogenetic relatedness of the RhlB and DeaD sequences (Schmid and Linder, 1992) and by the surprising conservation of the *eno*-like sequences. If this interpretation is correct, then it could be argued that degradosome-mediated mRNA turnover is an ancient control mechanism at RNA level that was established prior to the divergence of the three primary kingdoms. Together with other lines of evidence, including the observation that the most highly conserved gene clusters in several (eu)bacterial genomes are regulated at RNA-level (Siefert *et al*., 1997), these results are consistent with the hypothesis that during early stages of cell evolution RNA molecules played a more conspicuous role in cellular processes (Becerra *et al*., 2007).

Conclusions

Analysis of the increasingly large database of completely sequenced cellular genomes from the three major domains in order to define the set of the most conserved protein-encoding sequences to characterize the gene complement of the last common ancestor of extant life results in a set dominated by different putative ATPases and by molecules involved in gene expression and RNA metabolism (Delaye *et al*., 2005). DEAD-type RNA helicase and enolase genes, which are known to be part of the RNA degradosome, are as conserved as many transcription and translation genes. This suggests the early evolution of a control mechanism for gene expression at the RNA level, providing additional support to the hypothesis that during early cellular evolution RNA molecules played a more prominent role. Conserved sequences related to biosynthetic pathways include those encoding putative phosphoribosyl-pyrophosphate synthase and thioredoxin, which participate in nucleotide metabolism. Although the information contained in the available databases corresponds only to a minor portion of biological diversity, these sequences are likely to be part of an essential and highly conserved pool of protein domains common to all organisms (Becerra *et al*., 2007).

The high levels of genetic redundancy detected in all sequenced genomes imply that duplication has played a major role in the accretion of the complex genomes found in extant cells. They also show that prior to the early duplication events revealed by the large protein families, simpler living systems existed which lacked the large sets of enzymes and the sophisticated regulatory abilities of contemporary organisms. Once it appeared, the LCA would have been in the company of its siblings, a population of entities similar to it that existed throughout the same period. They may have not survived, but some of their genes did if they became integrated via lateral transfer into the LCA genome. The cenancestor is thus one of the last evolutionary outcomes of a series of ancestral events including lateral gene transfer, gene losses and paralogous duplications that took place before the separation of bacteria, Archaea and eucarya (cf. Becerra *et al*., 2007).

References

Achenbach-Richter, L., Gupta, R., Stetter., K. O., and Woese, C. R. (1987). Were the original eubacteria thermophiles? Systematic and Applied Microbiology, 9, 34–39.

Alcántara, C., Cervera, J., and Rubio, V. (2000). Carbamate kinase can replace in vivo carbamoyl phosphate synthetase. Implications for the evolution of carbamoyl phosphate biosynthesis. FEBS Letters, 484, 261–264.

Alifano, P., Fani, R., Liò, P., Lazcano, A. and Bazzicalupo, M. (1996). Histidine biosynthetic pathway and genes, structure, regulation, and evolution. *Microbiology Reviews*, **60**, 44–69.

Altermann, W. and Kazmierczak, J. (2003). Archean microfossils, a reappraisal of early life on Earth. *Research in Microbiology*, **154**, 611–7.

Anantharaman V., Koonin, E. V. and Aravind, L. (2002). Comparative genomics and evolution of proteins involved in RNA metabolism. *Nucleic Acid Research*, **30**, 1427–64.

Becerra, A., Delaye, L., Islas, A. and Lazcano A. (2007). Very early stages of biological evolution related to the nature of the last common ancestor of the three major cell domains. *Annual Review of Ecology and Evolutionary Systematics*, **38**, 361–79.

Bousseau, B., Blanquart, S., Necsulea, A., Lartillot, N. And Gouy, M. (2008). Parallel adaptations to high temperatures in the Archaean Eon. *Nature*, **456**, 942–5.

Canfield, D. E. (2006). Biochemistry, gas with an ancient history. *Nature*, **440**, 426–7.

Castresana, J. and Moreira, D. (1999). Respiratory chains in the last common ancestor of living organisms. *Journal of Molecular Evolution*, **49**, 453–60.

Clayton, R. A., White, O., Ketchum, K. A. and Venter, C. J. (1997). The genome from the third domain of life. *Nature*, **387**, 459–62.

Delaye, L. and Lazcano, A. (2000). RNA-binding peptides as molecular fossils. In *Origins from the Big-Bang to Biology*, eds. J. Chela-Flores, G. Lemerchand and J. Oró. Proceedings of the First Ibero-American School of Astrobiology. Dordrecht: Klüwer Academic Publishers, pp. 285–8.

Delaye, L., Becerra, A. and Lazcano, A. (2004). The nature of the last common ancestor. In *The Genetic Code and the Origin of Life*, ed. Lluis Ribas de Pouplana. Georgetown: Landes Bioscience, pp. 34–47.

Delaye, L., Becerra, A. and Lazcano, A. (2005). The last common ancestor, what's in a name? *Origins of Life and Evolution of the Biosphere*, **35**, 537–54.

Edgell, R. D. and Doolittle, W. F. (1997). Archaea and the origins of DNA replication proteins. *Cell*, **89**, 995–8.
Fitch, W.M. and Upper, K. (1987). The phylogeny of tRNA sequences provides evidence of ambiguity reduction in the origin of the genetic code. Cold Spring Harbor Symposium. *Quantitative Biology*, **52**, 759–67.
Forterre, P. (2006). Three RNA cells for ribosomal lineages and three DNA viruses to replicate their genomes, a hypothesis for the origin of cellular domain. *Proceedings of the National Academy of Sciences USA*, **103**, 3669–74.
Galtier, N., Tourasse, N. and Gouy, M. (1999). A nonhyperthermophilic common ancestor to extant life forms. *Science*, **283**, 220–1.
Gogarten, J. P., Kibak, H., Dittrich, P., Taiz, L. and Bowman, E. J. (1989). Evolution of the vacuolar H^+-ATPase; implications for the origin of eukayotes. *Proceedings of the National Academy of Sciences USA*, **86**, 6661–5.
Islas, S., Velasco, A. M., Becerra, A., Delaye, L. and Lazcano, A. (2003). Hyperthermophily and the origin and earliest evolution of life. *International Microbiology*, **6**, 87–94.
Iwabe, N., Kuma, K., Hasegawa, M., Osawa, S. and Miyata, T. (1989). Evolutionary relationship of archaebacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes. *Proceedings of the National Academy of Sciences USA*, **86**, 9355–9.
Jensen, R. A. (1976). Enzyme recruitment in evolution of new function. *Annual Review of Microbiology*, **30**, 409–25.
Joyce, G. F. (2002). The antiquity of RNA-based evolution. *Nature*, **418**, 214–21.
Koga, Y., Kyuragi, T., Nishihara, M. and Sone, N. (1998). Did archaeal and bacterial cells arise independently from noncellular precursors? A hypothesis stating that the advent of membrane phospholipids with enantiomeric glycerophosphate backbones caused the separation of the two lines of descent. *Journal of Molecular Evolution*, **46**, 54–63.
Koonin, E. V. (2009). On the origin of cells and virases, primordial virus world scenario. *Annals of the New York Academy of Sciences*, **1178**, 47–64.
Koonin E.V. and Martin, W. (2005). On the origin of genomes and cells within in organic compartments. *Trends in Genetics*, **21**, 647–54.
Kyrpides, N., Overbeek, R. and Ouzonis, C. (1999). Universal protein families and the functional content of the last universal common ancestor. Journal *of Molecular Evolution*, **49**, 413–23.
Lazcano, A., Fox, G.E. and Oró, J. (1992). Life before DNA, the origin and early evolution of early Archean cells. In *The Evolution of Metabolic Function*, ed. R. P. Mortlock. Boca Raton, FL: CRC Press, pp. 237–95.
Leipe, D. D., Aravind, L. and Koonin, E. V. (1999). Did DNA replication evolve twice independently? *Nucleic Acid Research*, **27**, 3389–401.
Levy, M. and Miller, S. L. (1998). The stability of the RNA bases: implications for the origin of life. *Proceedings of the National Academy of Sciences USA*, **95**, 7933–38.
Line, M. A. (2002). The enigma of the origin of life and its timing. *Microbiology*, **148**, 21–7.
Lollar, B. S. and McCollom, T. M. (2006). Biosignatures and abiotic constrainst on early life. *Nature*, **444**, E18.
Olsen, G. and Woese, C. R. (1997). archaeal genomics: an overview. *Cell*, **89**, 991–4.
Penny, D. and Poole, A. (1999). The nature of the last common ancestor. *Current Opinion in Genetics and Development*, **9**, 672–7.

Peretó, J., López-García, P. and Moreira, D. (2004). Ancestral lipid biosynthesis and early membrane evolution. *Trends in Biochemical Sciences*, **29**, 469–77.
Philippe, H. and Forterre, P. (1999). The rooting of the universal tree of life is not reliable. *Journal of Molecular Evolution*, **49**, 509–23.
Poole, A. and Logan, D. T. (2005). Modern mRNA proofreading and repair, clues that the last universal common ancestor possessed an RNA genome? Molecular *Biology and Evolution*, **22**, 1444–55.
Reanney, D. C. (1987). Genetic error and genome design. Cold Spring Harbor Symposium. *Quantitative Biology*, **52**, 751–7.
Ren, B., Tibbelin, G., de Pascale, D., Rossi, M., Bartolucci, S. and Ladenstein, R. (1998). A protein disulfide oxidoreductase from the archaeon *Pyrococcus furiosus* contains two thioredoxin fold units. *Nature Structural Biololgy*, **7**,602–11.
Schmid, S. R. and Linder, P. (1992). D-E-A-D protein family of putative RNA helicases. *Molecular Microbiology*, **6**, 283–92.
Schopf, J. W. (1993). Microfossils of the early Archaean Apex chert: new evidence for the antiquity of life. *Science*, **260**, 640–6.
Shen, Y., Buick, R. and Canfield, D. E. (2001). Isotopic evidence for microbial sulphate reduction in the early Archaean Era. *Nature*, **410**,77–81.
Siefert, J. L., Martin, K. A., Abdi, F., Wagner, W. R. and Fox, G. E. (1997). Conserved gene clusters in bacterial genomes provide further support for the primacy of RNA. *Journal of Molecular Evolution*, **45**, 467–72.
Stetter, K. O. (1994). The lesson of archaebacteria. In *Early Life on Earth*, Nobel Symposium No. 84, ed. S. Bengtson. New York: Columbia University Press, pp. 114–22.
Stubbe, J., Ge, J. and Yee, C. S. (2001). The evolution of ribonucleotide reduction revisited. *Trends in Biochemical Sciences*, **26**, 93–9.
Ueno, Y., Yamada, K., Yoshida, N., Maruyama, S. and Isozaki, Y. (2006). Evidence from fluid inclusions for microbial methanogenesis in the early Archaean Era. *Nature*, **440**, 516–19.
Velasco, A. M., Leguina, J. I. and Lazcano, A. (2002). Molecular evolution of the lysine biosynthetic pathways. *Journal of Molecular Evolution*, **55**, 445–9.
Wächstershäuser, G. (2003). From pre-cells to Eukarya – a tale of two lipids. *Molecular Microbiology*, **47**, 13–22.
Woese, C. R. (1998). The universal ancestor. *Proceedings of the National Academy of Sciences USA*, **95**, 6854–9.
Woese, C. R. and Fox, G. E. (1977) The concept of cellular evolution. *Journal of Molecular Evolution*, **10**, 1–6.
Woese, C. R., Kandler, O. and Wheelis, M. L. (1990). Towards a natural system of organisms: proposal for the domains Archaea, bacteria, and eucarya. *Proceedings of the National Academy of Sciences USA*, **87**, 4576–9.
Zhaxybayeva, O. and Gogarten, P. J. (2004). Cladogenesis, coalescence and the evolution of the three domains of life. *Trends in Genetics*, **20**, 182–7.