Evo Edu Outreach (2012) 5:382–388
DOI 10.1007/s12052-012-0444-8

ORIGIN OF LIFE

Cenancestor, the Last Universal Common Ancestor

Luis Delaye · Arturo Becerra

Published online: 2 September 2012

Abstract Darwin suggested that all life on Earth could be phylogenetically related. Modern biology has confirmed Darwin's extraordinary insight; the existence of a universal genetic code is just one of many evidences of our common ancestry. Based on the three domain phylogeny proposed by Woese and Fox in the early 1970s that all living beings can be classified on one of three main cellular lineages (Archaea, Bacteria, and Eukarya), it is possible to reconstruct some of the characteristics of the Last Universal Common Ancestor or cenancestor. Comparative genomics of organisms from the three domains has shown that the cenancestor was not a direct descendant of the prebiotic soup nor a primitive cellular entity where the genotype and the phenotype had an imprecise relationship (i.e., a progenote), rather it was an organism similar in complexity to extant cells. Due to the process of horizontal gene transfer and secondary gene losses, several questions regarding the nature of the cenancestor remain unsolved. However, attempts to infer its nature have led to the identification of a set of universally conserved genes. The research on the nature of the last universal common ancestor promises to shed light on fundamental aspects of living beings.

Keywords Last universal common ancestor · Cenancestor · Progenote · Bacteria · Archaea · Eukarya · Horizontal gene transfer · Early evolution of life

L. Delaye
Departamento de Ingeniería Genética,
Centro de Investigación y Estudios Avanzados del Instituto Politécnico Nacional,
Km. 9.6 Libramiento Norte Carretera Irapuato-León,
Irapuato, Guanajuato 36821, Mexico

A. Becerra (✉)
Facultad de Ciencias, UNAM,
Apdo. Postal 70-407, Ciudad Universitaria,
México, D.F. 04510, Mexico
e-mail: abb@ciencias.unam.mx

One Ancestor "*tous pour un, un pour tous*"

Common ancestry is a central idea in biology; its roots can be traced back to the beginning of evolutionary theory. As proof of this, Charles Darwin wrote in the *Origin of Species*

> All living beings have much in common, in their chemical composition, their cellular structure, their laws of growth, and their liability to injurious influences… Therefore, on the principle of natural selection with divergence of character, it does not seem incredible that, from such low and intermediate form both animals and plants may have been developed; and, if we admit this, we must likewise admit that all the organic beings which have ever lived on this earth may have descended from someone primordial form.

Present-day biology, including biochemistry, molecular phylogeny, and comparative genomics, has confirmed Darwin's extraordinary insight, i.e., that all living beings descent ultimately from a single species.

The modern research on the nature of the last common ancestor (LCA) or cenancestor (Fitch and Upper 1987) is obviously a major trend in present biology (Morange 2009, 2011) and began with the first attempt to reconstruct a universal phylogenetic tree by using a single molecule common to all cells. In the mid-1970s, Woese and Fox (1977) compared the small subunits of ribosomal RNA (16/18S rRNA) sequences from different species, including prokaryotes (cells without a nuclear membrane) and eukaryotes (cell with a nuclear membrane). These comparisons led to the reconstruction of a trifurcated, unrooted tree in which all known organisms can be grouped in one of three major monophyletic cell lineages; these were named as the domains Eubacteria (now Bacteria), Archaeabacteria (now Archaea), and the nucleo-cytoplasmic component of Eukaryotes (now known simply as Eucarya; Fig. 1). As shown, these lineages are derived from a common ancestor (Woese et al. 1990).

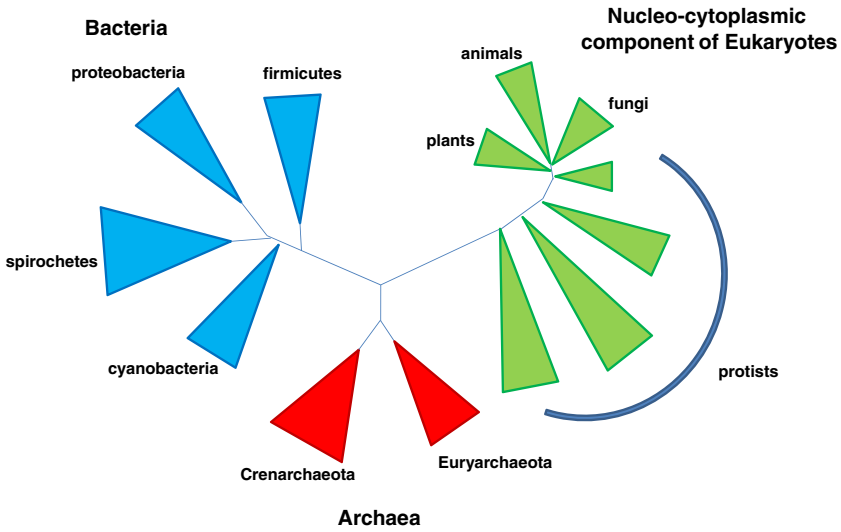

Fig. 1 Three cellular domains. The universal tree of life as suggested by the 16SrRNA molecule

Information from one single molecular marker does not necessarily yield a precise reconstruction of evolutionary processes, but as indicated by many phylogenies constructed from other genes such as those encoding polymerases, ATPase subunits, elongation factors, and ribosomal proteins. The identification of the three major lineages is not an artifact based exclusively on the reductionist extrapolation of information derived from a single gene (i.e., the 16SrRNA) but a true reflection of a common ancestry of all living forms. This is in accordance with the fact that all organisms share the same genetic code and crucial features of genome replication, gene expression, membrane-associated ATPase-mediated energy production, and basic anabolic reactions. Minor variations in the previous process can be easily explained as the outcome of divergent processes from an ancestral life form of the three major biological domains (Delaye et al. 2001; Becerra et al. 2007).

Phylogenetic analysis of rRNA sequences is acknowledged as a prime force in systematics and from its very inception, had a major impact in our understanding of cellular evolution. As exposed by the unrooted rRNA trees, no single domain predates the other two, and all three derive from a common ancestor. Recognition of the 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). That is a biological entity where the phenotype and genotype are the same, i.e. a much simpler biological entity than any extant cell. 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. However, the conclusion that the last common ancestor (LCA) was a progenote was disputed when the analysis of homologous traits found among some of its descendants suggested that it was not a protocell or any other pre-life progenitor system (Lazcano et al. 1992) but an organism similar in complexity to ext\ant prokaryotes.

In those years, the inventory of such shared traits was small, but it was surmised that the sketchy picture developed with the limited data bases would be confirmed when there were 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 (Doolittle 2000). This is shown, for instance, in the diversity of names that have been coined to describe it: progenote (Woese and Fox 1977), cenancestor (Fitch and Upper 1987), last universal cellular ancestor (Philippe and Forterre 1999), and last common community (Line 2002), among others. These terms are not truly synonymous, and they reflect the current controversies on the nature of the universal ancestor and the evolutionary processes that shaped it.

Reconstructing the Cenancestor

As mentioned above, all life on Earth uses exactly the same code to translate the information stored in DNA into proteins (with a few exceptions that are clearly evolutionary novelties). How is it possible that organisms as different as oak trees, *Escherichia coli* bacterium, amoebas, or ourselves share the same set of rules to read (translate) DNA? The answer is common ancestry; much in the same way that sisters and brothers resemble each other, features shared by all living beings were inherited from common ancestral species that lived millions of years ago.

We can use this knowledge to infer some features of the biology of this universal ancestor, or cenancestor. But in order to do such reconstruction, we need an evolutionary tree describing the phylogenetic relationships among all living beings on Earth. As mentioned, such a tree was proposed in the early 70's by Woese and Fox when using the 16SrRNA molecule to infer the phylogenetic relationships among organisms (Woese and Fox 1977). Before the work done by Woese and Fox, there were two main classification systems. In one of them, organisms were classified as Eukaryotes if their genetic material was compartmentalized by a membrane into a nucleus, or Prokaryotes if this structure is absent (Chatton 1938); in the other system, organisms were classified into five Kingdoms (Monera, Protists, Fungi, Plantae, and Animalia) based on their overall biology (Whittaker 1969).

Although the scheme of three domains (i.e., Bacteria, Archaea, and Eucarya) is incomplete because it does not include the anastomosis of bacteria to conform the mitochondria and chloroplast of Eukaryotes, nor the horizontal gene transfer among Prokaryotes, it does show that during very early stages of cellular evolution, life separated into three main lineages of descent. The classification of three domains (as named by Woese and Fox) is the guide we need to attempt a reconstruction of the biology of the cenancestor. The rationale is simple: if all present-day life derives ultimately from three main lineages of descent, then features (or more precisely, genes) homologous among these three life forms must have been present in the last universal common ancestor (Fig. 2).

This methodology is not infallible, however. Processes such as secondary gene losses or horizontal gene transfers among different cellular lineages have the power to obscure the past (Becerra et al. 1997). This is, if there have been several secondary gene losses after the last common ancestor, then our reconstruction will underestimate the gene content of this hypothetical entity; conversely, if there have been a lot of horizontal gene transfer events during the early evolution of life, we will overestimate the gene content of the cenancestor. The precision of our reconstructions of the genome (and therefore our inferences about their biology) of the last universal common ancestor depends on the relative intensity of previous processes. For instance, the amount of horizontal gene transfer among prokaryotes is still hotly debated among researchers today (Glansdorff 2000; Gogarten and Townsend 2005; Zhaxybayeva and Doolittle 2011).

Despite the methodological difficulties outlined above, different attempts to reconstruct the nature of the last universal common ancestor have led to the identification of a set of highly conserved genes among all cells that very likely have been inherited from the cenancestor (Kyrpides et al. 1999; Doolittle, 2000;Brown et al. 2001; Harris et al. 2003; Mirkin et al. 2003; Yang et al. 2005; Delaye et al. 2005; Moreira and Lopez-Garcia 2006, Ranea et al. 2006, Ouzonis et al. 2006). The set is mainly composed of genes related to transcription and translation (i.e., the beta and beta' prime subunit of RNA polymerase, ribosomal proteins, and elongation factors) (Harris et al. 2003). Notably, the main replicative DNA polymerase is not present in this set. This has led to some authors suggesting that the last universal common ancestor had an RNA genome (Leipe et al. 1999), a dubious conclusion, however, because all present-day cells have DNA genomes.

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 DNA informational proteins across the entire phylogenetic spectrum (Olsen and Woese 1997; Edgell and Doolittle 1997; Leipe et al. 1999; Penny and Poole 1999; Harris et al. 2003), comparative proteome analysis has shown that eubacterial replicative polymerases and primases lack homologues in the two other domains.

The peculiar distribution of the DNA replication machinery has led to suggestions not only of a cenancestor endowed with an RNA genome, but also of the polyphyletic origins of DNA and many of enzymes associated with DNA replication (Leipe et al. 1999; Koonin and Martin 2005) in

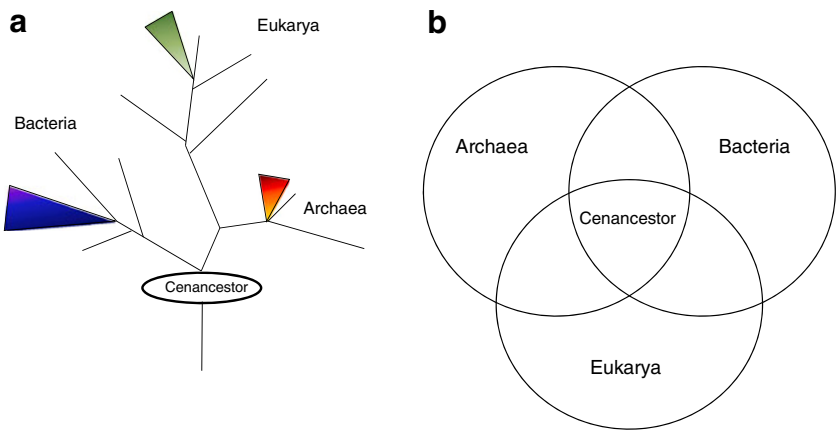

Fig. 2 Reconstructing the cenancestor. **a** Tree of life as suggested by the 16SrRNA molecule; **b** traits present in the *cenancestor* can be inferred by looking at homologous genes among the three cellular domains *Bacteria*, *Archaea*, and *Eukarya*

which viruses may have played a central role (Forterre, 2006). Koonin and Martin (2005) have argued that the cenancestor 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 is difficult to accept these schemes. There are indeed manifold indications that RNA genomes existed during early stages of cellular evolution (Lazcano et al. 1988), but

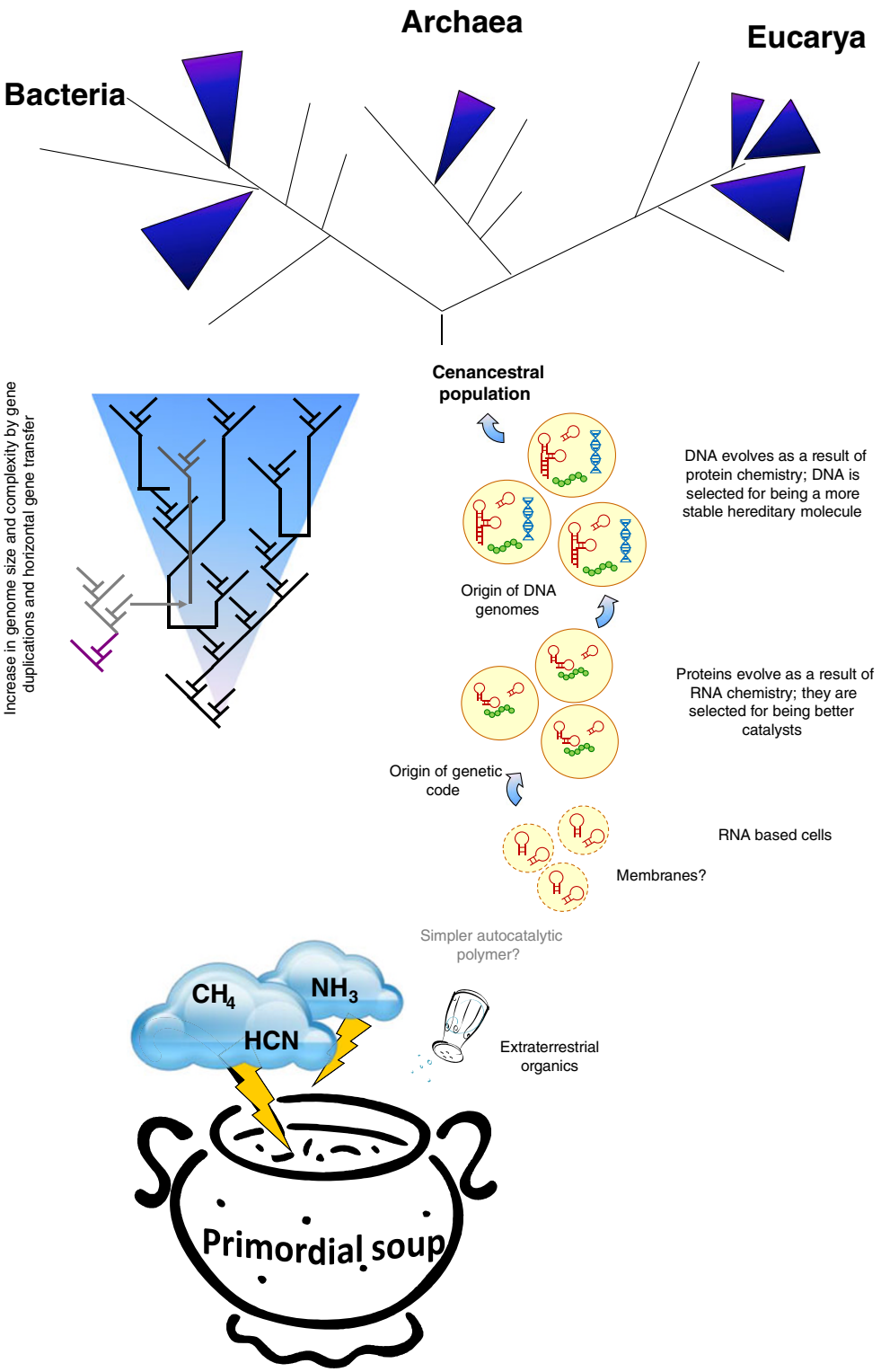

Fig. 3 Early evolution of life on Earth. Life originated from prebiotic chemistry. First stages of cellular evolution may have included replicative polymers other than DNA and RNA; the RNA world refers to a time when the RNA molecule acted as the hereditary as well as catalytic molecule of cells; eventually, RNA chemistry originated proteins (a relic from these days is the RNA-mediated synthesis of proteins in extant ribosomes); it is thought that cells capable of synthesizing proteins were selected for having superior catalytic molecules; finally, protein chemistry-originated DNA and cells with DNA genomes were selected for having a more stable hereditary molecule; the last universal common ancestor or *cenancestor* was very likely similar to extant cells in their metabolic and hereditary capacities

it is likely that double-stranded DNA genomes had become firmly established prior to the divergence of the three primary domains. It's especially likely, considering the sequence similarities shared by many ancient, large proteins found in all three domains that suggests considerable fidelity existed in the operative genetic system of their common ancestor, but such fidelity is unlikely to be found in RNA-based genetic systems (Reanney 1987; Lazcano et al. 1992)

Echoes from Ancient Worlds

Current descriptions of the cenacestor are limited by the scant information available: it is hard to understand the evolutionary forces that acted on our distant ancestors, whose environments and detailed biological characteristics are forever beyond our knowledge. By definition, the node located at the bottom of the cladogram is the root of a phylogenetic tree and corresponds to the common ancestor of the group under study. But names may be misleading. What we have been calling the root of the universal tree is in fact the tip of its trunk: inventories of cenancestor genes include sequences that originated in different pre-cenancestral epochs. Biological evolution prior to the divergence of the three domains was not a continuous, unbroken chain of progressive transformation steadily proceeding towards the LCA (Fig. 3).

Is important to note that the features that were present in the cenancestor would not be present in the first living systems (origin-of-life period). The notable coincidence between the monomeric constituents of living organisms and those synthesized in laboratory simulations of the prebiotic environment appears to be too striking to be fortuitous, and the discovery of catalytically active RNA molecules has given considerable credibility to prior suggestions of an evolutionary stage prior to the development of proteins and DNA genomes during which early life forms largely based on ribozymes may have existed. The difficulties involved with the synthesis and accumulation of ribonucleotides and RNA molecules in the prebiotic environment have led to the suggestion that the RNA world itself was the evolutionary outcome of some predecessor primordial living systems of what are now referred to as pre-RNA worlds (Fig. 4; Joyce 2002). However, the chemical nature of the first genetic polymers and the catalytic agents that may have formed the hypothetical pre-RNA worlds can only be surmised and cannot be deduced from comparative genomics or deep phylogenies (Becerra et al. 2007).

Slight or no geological evidence of the environmental conditions on the early Earth at the time of the origin and early evolution of life, nor any molecular or physical vestiges that preceded the appearance of the first cellular organisms are found in the Archean fossil record. Also, the 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 year-old Apex Cherts of the Australian Warrawoona formation (Schopf 1993) has been disputed, at the time being, the weight of evidence favors the idea that life existed 3.5 billion years ago (Altermann and Kazmierczak 2003, Brasier et al. 2004, 2006).

Comparative genomics may provide signs to the genetic organization and biochemical complexity of the earlier entities from which the cenancestor evolved. Genes involved in RNA metabolism, i.e., genes whose products synthesize, degrade, or interact with RNA, are among the most highly conserved sequences common to all known genomes, and provide insights into an early stage in cell evolution during which RNA played a much more conspicuous biological role (Tekaia et al. 1999, Delaye and Lazcano 2000, Anantharaman et al. 2002). However, it is difficult to see how the applicability of comparative genomics can be extended beyond a threshold that corresponds to a period of cellular evolution in which protein biosynthesis was already in operation. Older stages are not yet amenable to molecular phylogenetic analysis.

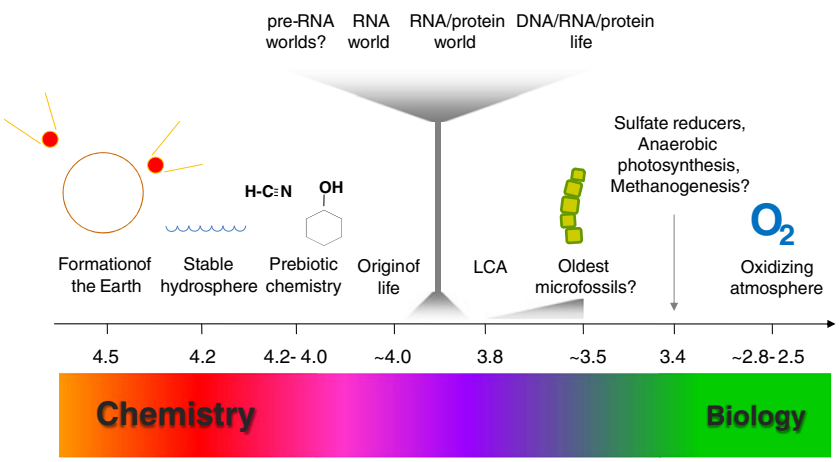

Fig. 4 Early life, time arrow. Starting from the origin of Earth, and ending in the origin of an oxidizing atmosphere, several million years of chemical and biological evolution are compressed in this figure. The *bottom line* shows that there is a continuum that goes from the prebiotic chemistry to the first cells

Although there have been considerable advances in the understanding of chemical processes that may have taken place before the emergence of the first living systems, life's beginnings are still shrouded in mystery. A phylogenetic approach to this problem is not feasible, since all possible intermediates that may have once existed have long disappeared. The temptation to do otherwise is best resisted. Given the huge gap existing in current descriptions of the evolutionary transition between the prebiotic synthesis of biochemical compounds and the cenancestor, it may be naive to attempt to describe the origin of life and the nature of the first living systems from the available rooted phylogenetic trees.

Remarks and Outlooks

Darwin suggested that species diverge from one another, generating a tree-like pattern of common ancestry. The existence of a universal ancestor is logically a derived from this mode of evolution. Modern biology has shown that Darwin's insights were correct. All living beings are very alike in their basic biochemistry and molecular biology; the existence of a common genetic code is one of the most prominent evidences of our common ancestry. However, reconstructing the biology of the cenancestor is not an easy task. Although the logic to recognize which genes have been inherited from the last common ancestor is straightforward (i.e., a gene that is present in Archaea, Eukaryaand Bacteria because of vertical inheritance was present in the last common ancestor), the accumulation of more than 3.5 billions of years of evolution from the cenancestor to extant biology makes the inference of the properties of this biological entity a formidable intellectual challenge.

However, it is clear that in spite of the qualitative and quantitative differences in the methodological approaches used to identify the gene complement of the cenancestor, the inventories show an overlap which reflects an impressive level of conservation of a significant number of sequences involved in basic biological processes. It is enough to assume that the cenancestor: (a) was not a progenote or a protocell, but an entity similar to extant prokaryotes; (b) was preceded by earlier entities in which RNA molecules played a more conspicuous role in cellular processes and in which ribosome-mediated protein synthesis had already evolved; (c) had a genome of DNA, originated prior to the evolutionary divergence of the three main cell domains; and (d) maybe was not an extremophile (Becerra et al. 2007).

Acknowledgment Financial support from CONACYT (Mexico) project 100199 to AB, is acknowledged. L.D. wishes to thank CINVESTAV Unidad Irapuato for all facilities provided.

References

Altermann W, Kazmierczak J. Archean microfossils: a reappraisal of early life on Earth. Res Microbiol. 2003;154:611–7.

Anantharaman V, Koonin EV, Aravind L. Comparative genomics and evolution of proteins involved in RNA metabolism. Nucleic Acid Res. 2002;30:1427–64.

Becerra A, Islas S, Leguina JI, Silva E, Lazcano A. Polyphyletic gene losses can bias backtrack characterizations of the cenancestor. J Mol Evol. 1997;45:115–8.

Becerra A, Delaye L, Islas A, Lazcano A. Very early stages of biological evolution related to the nature of the last common ancestor of the three major cell domains. Annu Rev Ecol Evol Sys. 2007;38:361–79.

Brasier M, Green O, Lindsay J, Steele A. Earth's oldest approximately 35 Ga fossils the Early Eden hypothesis: questioning the evidence. Orig Life Evol Biosph. 2004;341–2:257–69.

Brasier M, McLoughlin N, Green O, Wacey D. A fresh look at the fossil evidence for early Archaean cellular life. Philos Trans R SocLond B Biol Sci. 2006;29:887–902.

Brown JR, Douady CJ, Italia MJ, Marshall WE, Stanhope MJ. Universal trees based on large combined protein sequence datasets. Nat Genet. 2001;28:281–5.

Chatton E. Titre et travaux scientifique (1906–1937) de Edouard Chatton. Sottano: Sette; 1938.

Delaye L, Lazcano A. RNA-binding peptides as molecular fossils. In: Chela-Flores J, Lemerch G, Oró J, editors. Origins from the Big-Bang to Biology: Proceedings of the First Ibero-American School of Astrobiology. Dordrecht: Klüwer Academic Publishers; 2000. p. 285–8.

Delaye L, Vázquez H, Lazcano A. The cenancestor its contemporary biological relics: the case of nucleic acid polymerases. In: Chela-Flores J, Owen T, Raulin F, editors. First steps in the origin of life in the universe. Dordrecht: Kluwer; 2001. p. 223–30.

Delaye L, Becerra A, Lazcano A. The last common ancestor: what's in a name? Orig Life Evol Biosph. 2005;35:537–54.

Doolittle WF. The nature of the universal ancestor the evolution of the proteome. Curr Opinion Struct Biol. 2000;10:355–8.

Edgell DR, Doolittle WF. Archaea and the origin(s) of DNA replication proteins. Cell. 1997;89:995–8.

Fitch WM, Upper K. The phylogeny of tRNA sequences provides evidence of ambiguity reduction in the origin of the genetic code. Cold Spring Harbor Symp Quant Biol. 1987;52:759–67.

Forterre P. Three RNA cells for ribosomal lineages and three DNA viruses to replicate their genomes: a hypothesis for the origin of cellular domain. Proc Natl Acad Sci USA. 2006;103:3669–74.

Glansdorff N. About the last common ancestor the universal life-tree and lateral gene transfer: a reappraisal. Mol Microbiol. 2000;38:177–85.

Gogarten JP, Townsend JP. Horizontal gene transfer genome innovation evolution. Nature Rev Microbiol. 2005;3:679–87.

Harris JK, Kelley ST, Spiegelman GB, Pace NR. The genetic core of the universal ancestor. Genome Res. 2003;13:407–12.

Joyce GF. The antiquity of RNA-based evolution. Nature. 2002;418:214–21.

Koonin EV, Martin W. On the origin of genomes cells within in organic compartments. Trends Genet. 2005;21:647–54.

Kyrpides N, Overbeek R, Ouzonis C. Universal protein families and the functional content of the last universal common ancestor. J Mol Evol. 1999;49:413–23.

Lazcano A, Guerrero R, Margulis L, Oró J. The evolutionary transition from RNA to DNA in early cells. J Mol Evol. 1988;27:283–90.

Lazcano A, Fox GE, Oró J. Life before DNA: the origin early evolution of early Archean cells. In: Mortlock RP, editor. The evolution of metabolic function. Boca Raton: CRC Press; 1992.

Leipe DD, Aravind L, Koonin EV. Did DNA replication evolve twice independently? Nucleic Acid Res. 1999;27:3389–401.

Line MA. The enigma of the origin of life and its timing. Microbiology. 2002;148:21–7.

Mirkin BG, Fenner TI, Galperin MY, Koonin EV. Algorithms for computing parsimonious evolutionary scenarios for genome evolution, the last common ancestor and dominance of horizontal gene transfer in the evolution of prokaryotes. BMC Evol Biol. 2003;3:2.

Morange M. Articulating different modes of explanation: the present boundary in biological research. In: Barberousse A, Morange M, Pradeu T, editors. Mapping the future of biology. Spring: Berlin; 2009. p. 15–26.

Morange M. Some considerations on the nature of LUCA, and the nature of life. Res Microbiol. 2011;162:5–9.

Moreira D, Lopez-Garcia P. The last common ancestor. Earth, Moon and Planets. 2006;98:187–93.

Olsen G, Woese CR. Archaeal genomics: an overview. Cell. 1997;89:991–4.

Ouzonis AC, Kumin V, Darzentas N, Goldovsky L. A minimal estimate for gene content of the last universal common ancestor exobiology from a terrestrial perspective. Res Microbiol. 2006;157:57–68.

Penny D, Poole A. The nature of the last common ancestor. Curr Opin Gen Dev. 1999;9:672–7.

Philippe H, Forterre P. The rooting of the universal tree of life is not reliable. J Mol Evol. 1999;49:509–23.

Ranea AG, Sillero A, Thorton MJ, Orengo AC. Protein superfamily evolution and the last universal common ancestor LUCA. J Mol Evol. 2006;63:513–25.

Reanney DC. Genetic error and genome design. Cold Spring Harbor Symp Quant Biol. 1987;52:751–7.

Schopf JW. Microfossils of the early Archaean Apex Chert: new evidence for the antiquity of life. Science. 1993;260:640–6.

Tekaia F, Dujon B, Lazcano A. Comparative genomics: products of the most conserved protein-encoding genes synthesize degrade or interact with RNA. Abstracts of the 9th ISSOL Meeting, San Diego California USA. 1999; 46:53.

Whittaker RH. New concepts of kingdoms or organisms. Evolutionary relations are better represented by new classifications than by the traditional two kingdoms. Science. 1969;163:150–60.

Woese CR, Fox GE. The concept of cellular evolution. J Mol Evol. 1977;10:1–6.

Woese CR, Kler O, Wheelis ML. Towards a natural system of organisms proposal for the domains Archaea, Bacteria, and Eucarya. Proc Natl Acad Sci USA. 1990;87:4576–9.

Yang S, Doolittle RF, Bourne PE. Phylogeny determined by protein domain content. Proc Natl Acad Sci USA. 2005;102:373–8.

Zhaxybayeva O, Doolittle WF. Lateral gene transfer. Curr Biol. 2011;21(7):R242–6.