11 The Last Common Ancestor of Modern Cells

David Moreira and Purificación López-García

11.1 The Last Common Ancestor, the Cenancestor, LUCA: What’s in a Name?

All living beings share a number of essential features pertaining to their biochemistry and fundamental processes. Some of these features are so complex (an outstanding example is the genetic code) that their probability to have appeared several times independently is almost negligible. This authorizes the concept of a common ancestor to all living beings that possessed these traits, and from which diversification occurred, leading to the emergence of the three domains of life that we recognize today, Archaea, Bacteria, and Eucarya. Nevertheless, the concept of a common ancestor was born well before these universal features were even identified, mostly on the basis of purely philosophical and theoretical considerations.

11.1.1 Some Historical Grounds

The idea that all living beings are linked by a natural process has influenced the scientific thought for a long time. For example in his 1686 *Discourse on Metaphysics*, the German philosopher Leibniz explicitly states the existence of intermediate species between those that are distant, implying that all living beings are related in a great chain, which is intrinsic to the harmony of the universe. Following this inspiration, many scientists have tried to identify the source of diversity and the nature of the process unifying the different species. Darwin’s concept of descent with modification and natural selection provided the solid grounds to explain the patterns of similarity and difference among species that have served as a general framework for biology until present days. In *The Origin of Species*, Darwin concluded that a logical outcome of the premises of descent with modification and natural selection is that probably all the organic beings, which have ever lived on this earth have descended from some one primordial form, into which life was first breathed (Darwin 1859). The concept of a common ancestor for all life forms was born.

This idea received strong support during the last century, thanks to the spectacular development of biochemistry and, especially, of molecular biology. Bacteria, eukaryotes and the recently discovered archaea (Woese and Fox 1977a)
share the basic structural constituents of the cell and the most fundamental metabolic reactions. All life is based on common biochemical themes (Kornberg 2000; Pace 2001).

11.1.2 The Hypothesis of a Cenancestor

This hypothetical ancestral organism to all living beings has been baptized with different names according to different authors. The most popular are the last common ancestor, the cenancestor (from the Greek kainos meaning recent and koinos meaning common) (Fitch and Upper 1987) and the last universal common ancestor or LUCA (Forteerre and Philippe 1999). Although all evolutionary biologists agree that some type of cenancestor existed, its nature is a matter of intense and highly speculative debates. Most authors favor the idea that the cenancestor was a single organism, an individual cell that existed at a given time and that possessed most of the features (and genes encoding them) that are common to all contemporary organisms. From this single ancestor, the different domains of life would have diverged (Fig 11.1A–D). Others, on the contrary, envisage a population of cells that, as a whole, possessed all those genes, although no single individual did (Kandler 1994; Woese 1998; Woese 2000). This implies that the level of gene exchange and spreading in this population was very high. At some point, however, a particular successful combination of genes occurred in a subpopulation that became isolated and gave rise to a whole line of descent. Kandler, for instance, proposed in his pre-cell theory that bacteria, archaea, and eukaryotes emerged sequentially in this way (Fig 11.1E) (Kandler 1994; Wächtershäuser 2003).

Single cell or population, all researchers agree that the cenancestor was already quite complex, having evolved from simpler entities, and that there was a more or less long evolutionary path from the origin of life to the cenancestor stage. Both, the origin of life and the nature of the cenancestor are different evolutionary questions. Nonetheless, despite the general agreement that the cenancestor was already quite complex, the level of complexity attributed to it varies depending on the model. For Woese, the cenancestor was a relatively primitive entity, which he called a “progenote”, that had not completely evolved the link between genotype and phenotype (Woese and Fox 1977b). For others, the progenote state occurred prior to the cenancestor, which was nearly a modern cell (see review in Doolittle 2000a).

Given their universality, a ribosome-based protein synthesis (translation), a well-developed transcription machinery for the synthesis of structural and messenger RNAs, and an energy-obtaining process based on the generation of a proton gradient across membranes (Gogarten et al. 1989), were among the features that the cenancestor certainly possessed. Other cenancestor properties are, however, much more controversial, such as the existence of a DNA-based genome or even the possession of lipid-based membranes. The occurrence in the cenancestor of these properties was initially deduced mostly from biochemistry and molecular
Fig. 11.1. Current hypotheses for the evolution of the three domains of life from a last common ancestor or cenancestor. The cenancestor stage is indicated by a blue sphere. Models A to D envisage that all the properties attributed to the cenancestor co-existed in the same cell, whereas in E, they were present collectively in a population of primitive cells.
biology studies. Obviously, many genes (and proteins) were involved in these ancestral machineries and processes. At present, the most powerful tools to infer in more detail which were the genes and proteins already present in the cenancer are comparative genomics and molecular phylogeny. This type of analysis is greatly facilitated by the increasing number of complete genome sequences available for very different organisms. Although several problems, in particular horizontal gene transfer and differential gene loss, make the reconstruction of the ancestral gene content troublesome (Koonin 2003). We briefly discuss these aspects in the following sections.

11.2 How Did the Cenancestor Make Proteins?

Prokaryotic (archaeal and bacterial) species contain approximately between 500 and 10 000 genes, whereas eukaryotic species contain approximately between 2000 and 30 000 genes. Nevertheless, when the gene content of all available genomes is compared, only $\sim$60 genes are found to be common to all of them. This set of genes is almost entirely integrated by genes encoding ribosomal RNA and ribosomal proteins as well as other proteins involved in translation (especially aminoacyl-tRNA synthetases and translation factors) (Koonin 2003). This implies that these genes are ancestral, and provide strong evidence that the cenancer possessed a fully developed ribosomal-based translation machinery for protein synthesis that was comparable to the one found in modern organisms. Protein synthesis by ribosomes is, therefore, the most universally conserved process. Furthermore, the level of conservation appears so high that the process of protein synthesis has remained practically unchanged for, likely, more than three billion years.

A few of the $\sim$60 genes that conform the universal core encode RNA polymerase subunits. RNA polymerase is responsible for the synthesis of messenger and other RNAs from genes (DNA templates) (Koonin 2003). As in the case of translation, this implies that the cenancer possessed at least part of the transcription machinery found in contemporary organisms. Nevertheless, the degree of conservation of the transcription machinery is not as high, since several RNA polymerase subunits and transcription factors are not universally distributed.

Transcription and translation are among the most conserved processes in cells, and can be traced back to the cenancer. The fact that both processes depend primarily on structural and catalytic RNA molecules, e.g., the ribosomal RNAs, indicates that RNA played an essential role since very early in cell evolution. This and the recent discovery of ribozymes (small catalytic RNAs) have been important elements leading to the proposal of an “RNA!world” (for review, see Joyce 2002). According to this model, there was a very early evolutionary stage when RNA carried both, information-storing and catalytic functions, which are performed today by DNA and proteins, respectively. While being broadly accepted, this model remains, however, purely hypothetical.
11.3 What Was the Nature of the Genetic Material?

Despite the fact that, in all contemporary cells, DNA is the molecule where genetic information is stored, only three out of the ∼60 universal genes are related to DNA replication and/or repair. These are one DNA polymerase subunit, one exonuclease and one topoisomerase (Koonin 2003). From all the genes known to be involved in DNA replication in organisms of the three domains of life, many are shared by archaea and eukaryotes but are absent from bacteria. The latter apparently possesses completely unrelated genes encoding the proteins that perform the equivalent functions. Various hypotheses have been proposed to explain the profound differences between the bacterial and archaeal/eukaryotic replication machineries. One of them even postulates that the cenancestor did not possess a DNA genome at all. According to this model, the cenancestor would have had an RNA genome, and DNA replication would have evolved twice independently, once in the bacterial line of descent and other in the lineage leading to archaea and eukaryotes (Forterre 2002; Leipe et al. 1999). However, this model is contested by different lines of evidence that, put together, suggest that the cenancestor already possessed a DNA genome. First, although few, universally conserved proteins and protein domains involved in DNA metabolism indeed exist (Giraldo 2003). Second, RNA is much more error-prone than DNA due to its higher mutation rate, so that single RNA molecules cannot exceed a certain size (Eigen limit) without falling into replicative catastrophe (Eigen 1971, 2002). This limit is so small, ∼30–50kb, that an RNA molecule could contain only a few dozen genes. Recent estimates suggest that the, already quite complex, cenancestor may have had probably more than 600 genes (Koonin 2003). If its genome was made of RNA, many RNA molecules would have been required to contain all those genes, which poses a serious problem of stability during replication and partition among daughter cells. The existence of this problem is attested by the characteristics of DNA and RNA viruses today. Whereas DNA viruses can have genomes that reach very large sizes, up to ∼1 Mbp (Raoult et al. 2004), RNA viruses’ genomes do not exceed ∼30kb (Domingo and Holland 1997).

The remaining models propose that the cenancestor had a DNA-based genome. One of these explains the dichotomy of DNA replication in bacteria and archaea/eukaryotes by stating that, whereas transcription and translation were already well developed in the cenancestor, DNA replication was still very primitive. DNA replication would have been improved and refined after, or simultaneously to, the separation of the two lines of descent leading to the bacteria and to the archaea/eukaryotes (Olsen and Woese 1997). A contrasting model could be that the cenancestor had a highly complex DNA-based metabolism and contained the ancestral versions of the proteins found today in both the bacterial and the archaeal/eukaryotic replication machineries. One set of proteins would be involved in replication, whereas the other would be specialized in DNA repair. During the speciation of the two lines, only one set of proteins would have been retained in each line of descent. Another proposal suggests that the replication machinery was already well developed in the cenancestor, but that it evolved
very fast in one or the two lineages descending from the cenancestor to the point
that the similarity between the homologous genes in both lines is no longer
recognizable (Moreira 2000; Olsen and Woese 1997). Finally, an additional hy-
pothesis suggests that archaea and eukaryotes have retained the ancestral DNA
replication system, whereas the bacteria have seen their machinery replaced by
genes imported from viruses (Forterre 1999). However, there is no evidence for
this hypothesis but, on the contrary, many examples of the opposite, i.e., numer-
ous acquisitions of replication-related genes by viruses from the genome of their
cellular hosts (Moreira 2000). In conclusion, although still a matter of debate,
it appears likely that the cenancestor had a DNA genome. However, the mecha-
nism for its replication and the early divergence of the replication machineries
in the bacteria and the archaea/eukaryotes remain a mystery.

11.4 What Did the Cellular Metabolism Look Like?

The question of how metabolism looked in the cenancestor is a difficult and
complex one, and hence, is generally put aside in descriptions of a putative
ancestor. The theoretical proposal of a cenancestor is fundamentally based on
conserved genes related to the storage, expression and transmission of the ge-
netic information present in all contemporary organisms (for recent reviews, see
Doolittle 2000a, 2000b; Koonin 2003). However, genes involved in energy and
carbon metabolism not only display a patchy distribution in organisms of the
three domains of life, but they very often belong to large multigenic families
whose members have been recruited for different functions in various metabolic
pathways. Furthermore, horizontal gene transfer is known to affect preferentially
metabolic genes, since they may confer an immediate adaptive advantage. There-
fore, the reconstruction of ancestral metabolic pathways is frequently masked by
a complex history of gene duplication and differential enzyme recruitment (Cas-
tresana and Moreira 1999).

Discussions and controversies about early metabolism do take place, but they
concern the very first metabolism at the time when life arose rather than the
metabolic traits in a more evolved cenancestor. These discussions are model-
dependent. If the first living beings were heterotrophs feeding upon an organic
prebiotic soup as proposed classically (Oparin 1938), then fermentation, which
is mechanistically simple, may have been the first way of gaining energy (Broda
1970), while cell building blocks were directly uptaken from the soup. If the first
living beings were chemolithoautotrophs initially developed on pyritic (FeS₂)
surfaces (Wächtershäusser 1988) or in iron monosulfide (FeS) tridimensional
compartments (Russell and Hall 1997), energy was derived from redox reactions
involving inorganic molecules such as H₂S, H₂ and FeS. In this case, organic
molecules were synthesized de novo from CO₂ or CO by an ancestral, yet un-
determined, metabolic pathway. Today, four different pathways of autotrophic
carbon fixation are known (revised in Peretó et al. 1999). The Calvin–Benson
cycle (reductive pentose-phosphate pathway) is found in oxygenic photosynthesizers such as cyanobacteria and plants, but also in other autotrophic prokaryotes. The Arnon cycle (reductive citric acid pathway) is found in several bacteria and archaea. The Wood–Ljundahl cycle (reductive acetyl-CoA pathway) is found in acetogenic and sulfate-reducing bacteria and methanogenic archaea. Finally, the hydroxypropionate pathway is found in the green non-sulfur bacterium *Chloroflexus* sp. and a few thermophilic archaea. There is general agreement to consider that the Calvin–Benson cycle appeared relatively late during bacterial evolution. However, which of the other three is older is a matter of debate. The Arnon cycle is proposed to be older by Wächtershäuser and other adherents of the pyrite-based chemoautotrophic origin of life because of its wide distribution in bacteria and archaea (Wächtershäuser 1990). Nevertheless, a reductive acetyl-CoA pathway is proposed to be the primordial autotrophic pathway because of its higher simplicity (acetate, a two-carbon molecule is synthesized) and its exergonic (energy-releasing) nature under certain hydrothermal conditions (Russell and Martin 2004). Finally, although less known, the hydroxypropionate pathway is also very simple and has been proposed to be the first autotrophic pathway at least in phototrophic bacteria (Peretó et al. 1999).

Although the metabolic properties of the cenancestor are not generally discussed, if any or several of these pathways had developed in earlier times, the cenancestor must have possessed them. In any case, the presence of a universal highly conserved membrane-bound ATPase indicates that the cenancestor was able to produce energy in the form of ATP by generating a proton gradient across the cell membrane. What is unclear is the type of electron donors and acceptors required to generate this proton gradient, although it might have been able to use a variety of oxidized inorganic molecules as electron acceptors. It is also likely that the cenancestor was able to carry out a simple heterotrophic metabolism, at least some type of fermentation, but whether it was a full autotroph or not remains an open question.

### 11.5 Was the Cenancestor Membrane-Bounded?

All contemporary cells are surrounded by a plasma membrane that is made out of phospholipids generally organized in bilayers. However, as in the case of replication, there exist profound differences between, this time, the membrane lipids of archaea and the membrane lipids of bacteria and eukaryotes. In archaea, phospholipids are made out of generally isoprenoid lateral chains that are bounded by ether linkages to glycerol-1-phosphate, whereas in bacteria and eukaryotes phospholipids are made out of fatty acids bounded by ester linkages to glycerol-3-phosphate. From these differences, the most fundamental is the opposite stereochemistry of the glycerol-phosphate, since some bacteria and eukaryotes make ether linkages under certain circumstances, and since archaea do make fatty acid ether phospholipids as well (Peretó et al. 2004). The
enzymes responsible for the synthesis of glycerol-1-phosphate and glycerol-3-phosphate are not homologous in archaea and bacteria/eukaryotes, but belong to two different enzymatic families. This difference is so hard to explain that some authors have proposed that the cenancestor had not yet reached a cellular stage. Membrane lipids, and cells simultaneously, would have evolved independently from a non-enzymatic lipid synthesis pathway to generate bacteria and archaea (Koga et al. 1998). Other authors propose that the cenancestor did not possess a lipid membrane at all, and that membrane lipids (and their biosynthetic pathways) were invented twice, in each one of the lineages that diverged from the cenancestor leading to the bacteria and to the archaea. Instead of lipids, the cenancestor would have been endowed with a mineral membrane made out of iron monosulfide; cells would be mineral compartments in a particular kind of hydrothermal chimneys (Martin and Russell 2003). Finally, another, less radical option, is that the cenancestor had a lipid membrane, but that it was heterochiral, i.e., it was composed by a mixture of lipids built upon glycerol-1-phosphate and glycerol-3-phosphate (Wächtershäuser 2003). A subsequent specialization of the biosynthetic pathways to yield the two types of homochiral membranes would have accompanied the speciation of archaea and bacteria. Recent phylogenetic analyses of the enzymes involved in the synthesis of glycerol phosphate strongly suggest that the cenancestor possessed a non-stereospecific pathway of phospholipid biosynthesis and was therefore endowed with heterochiral membranes (Peretó et al. 2004). A cenancestor endowed with a lipid membrane would be in agreement with the occurrence of several membrane-bound proteins extremely well conserved, notably the proton-pump ATPases (Gogarten et al. 1989) and the signal recognition particle, SRP (Gribaldo and Cammarano 1998).

11.6 Other Unresolved Questions

As we have seen above, some properties of the hypothetical cenancestor appear more or less well-defined, such as the possession of a quite modern transcription and translation machinery and, most likely, the existence of phospholipid membranes. Other features remain obscure, such as the type of carbon and energy metabolism or the replication of the genetic material. However, there are additional questions that remain open and that have been the subject of lively debates. Among these, whether the cenancestor was hyperthermophilic or not, and whether it was simple or complex.

The proposal of a hyperthermophilic cenancestor is linked to the discovery of hyperthermophilic bacteria and archaea growing optimally at >80°C, and to the first proposals of a hot, autotrophic origin of life in a warmer early Earth with extended hydrothermal activity (Achenbach-Richter et al. 1987; Pace 1991). The most important argument used was that hyperthermophilic prokaryotes branched at the most basal positions in phylogenetic trees (Stetter 1996). The first criticisms to a hot origin of life derived from the fact that RNA and
other important biomolecules have relatively short life-times at high temperatures (Lazcano and Miller 1996). However, as mentioned before, the origin of life and the cenancestor were separated in time, and might have occurred in different environmental conditions (Arrhenius et al. 1999). Thus, some authors propose that the cenancestor was hyperthermophilic, and that the origin of life might have taken place at much lower temperatures, but only hyperthermophiles could survive the late heavy meteorite bombardment ∼3.9 Ga ago (Gogarten-Boekels et al. 1995). However, the most important criticism to a hypothetical hyperthermophilic ancestor derives from recent phylogenetic analyses. On the one hand, computer reconstruction of ancestral ribosomal RNA sequences suggests that the content of guanine and cytosine in the cenancestor’s RNA was incompatible with life at >80°C (Galtier et al. 1999). However, the analysis of the same datasets using other methods questions this conclusion favoring, on the contrary, a hyperthermophilic cenancestor (Di Giulio 2000). On the other hand, refined phylogenetic analyses of ribosomal RNA sequences suggest that the basal emergence of hyperthermophilic bacteria was an artifact of phylogenetic tree reconstruction and favor the idea that they adapted secondarily to hyperthermophily (Brochier and Philippe 2002). However, as in the previous case, the use of other phylogenetic methods comes up with opposite results (Di Giulio 2003). At any rate, although the situation in bacteria is highly controversial, there is general agreement that the ancestor of the archaea was indeed hyperthermophilic (Forterre et al. 2002). Therefore, if the bacterial ancestor was also a hyperthermophile, then the most parsimonious conclusion is that the cenancestor was hyperthermophilic. If the bacterial ancestor was not hyperthermophilic, then it will be very difficult to infer the type of environmental conditions in which the cenancestor thrived. At any rate, all the current analyses and proposals appear compatible with the occurrence of a thermophilic (60–80°C) cenancestor (López-García 1999).

Another controversial issue concerns the level of complexity that the cenancestor possessed. As stated above, it is clear that it was already very complex. Even the authors that call it simple, envisage that its genome contained several hundred genes (∼600–1000), which is in the range of the simplest present-day prokaryotes (Koonin 2003). Most authors imagine an ancestor that was also structurally simple, i.e., with a cellular organization resembling that of today’s prokaryotes. An ancestor of this type would have the genetic material directly immersed in the cytoplasm, where the replication, transcription and translation would take place. This type of ancestor is supported by the widely accepted bacterial rooting of the tree of life, i.e., that the root of the tree of life lies between the bacteria and the archaea/eukaryotes (Fig. 11.1A) (Woese et al. 1990), but would be also compatible with two alternative tree topologies (Fig. 11.1B, C). However, this has been challenged by authors that propose a eukaryotic rooting of the tree of life, i.e., that the root lies in between the eukaryotes and a branch leading to the two prokaryotic
domains (Fig. 11.1D) (Brinkmann and Philippe 1999; Philippe and Forterre 1999). This eukaryotic rooting would still be compatible with a prokaryote-like cenancetor, but it opens the possibility that the cenancetor had some of the traits that characterize modern eukaryotes, in particular the presence of a membrane-bound nucleus and of many small RNA molecules claimed to be relics of a hypothetical RNA world (Poole et al. 1999). In this model, prokaryotes would be derived by a reductive process from more complex eukaryotic-like ancestors. However, although the position of the root is indeed an open question, models proposing a eukaryotic-like cenancetor do not explain how such a complex entity was built from the prebiotic world. In this sense, a simpler, prokaryotic-like cenancetor appears much more parsimonious in evolutionary terms.

11.7 Perspectives

Obviously many questions have to be answered before achieving an accurate picture of the putative cenancetor. Fortunately, it seems likely that a number of these questions can be addressed thanks to the increasing amount of data derived from comparative genomics. Nonetheless, a crucial issue will be to determine the impact of horizontal gene transfer in evolution because, if horizontal gene transfer has been rampant all along life history as some authors suggest (Doolittle 2000b), it would then be very difficult or even impossible to reconstruct any ancestor. In a scenario of frequent or even massive horizontal gene transfer, a single cenancetor that contained all the genes ancestral to those shared among the three domains of life did not likely exist. Rather, these ancestral genes were probably present in different organisms and at different times (Zhaxybayeva and Gogarten 2004). The application of population genetics methods to the study of gene emergence and extinction over long timescales may shed some light into this problem.

Another open question concerns the origin and evolution of viruses, and how these have affected cellular evolution. For instance, they may have contributed significantly to horizontal gene transfer serving as vehicles of gene exchange. They may have also helped to increase the evolutionary rate of many genes thus contributing to the development of novel gene functions.

An alternative source of information to reconstruct a model portrait of the cenancetor may come from the generation of more accurate models of the early Earth. This could help to delineate the environmental conditions where life arose and first evolved, giving clues, for instance, as to the putative hyperthermophilic nature of the cenancetor, and/or to the most likely metabolic pathways it utilized. Similarly, resolving the controversies that currently exist on the earliest microfossils and increasing the number of unambiguous fossil data will be most helpful to establish a likely chronology of early life evolution.
References


