

Oxygen

The Molecule that made the World

Chapter 8: Looking for LUCA

Nick Lane

Biochemist and writer

About

Dr Nick Lane is a British biochemist and writer. He was awarded the first Provost's Venture Research Prize in the Department of Genetics, Evolution and Environment at **University College London**, where he is now a Reader in Evolutionary Biochemistry. Dr Lane's research deals with evolutionary biochemistry and bioenergetics, focusing on the origin of life and the evolution of complex cells. Dr Lane was a founding member of the UCL Consortium for Mitochondrial Research, and is leading the UCL Research Frontiers Origins of Life programme. He was awarded the 2011 BMC Research Award for Genetics, Genomics, Bioinformatics and Evolution, and the 2015 Biochemical Society Award for his sustained and diverse contribution to the molecular life sciences and the public understanding of science.



Oxygen

The Molecule that made the World

Chapter 8: Looking for LUCA

WHEN DID LUCA GIVE RISE TO HER DIVERSE OFFSPRING? Cells resembling modern prokaryotes date back 3.5 billion years, to the stromatolites in south-western Australia, as we noted in Chapter 3. The first signs of eukaryotic cells, the biomarkers of membrane sterols, date to about 2.7 billion years ago. The first unequivocal eukaryotic fossils are found in rocks dating to about 2.1 billion years ago. An explosion in the number and variety of eukaryotic cells took place around 1.8 billion years ago.

Eukaryotic cells share their fundamental biochemistry with prokaryotes, but are larger and more complex. It seems likely that only the innate complexity of eukaryotes can support the added layers of organisation required for the evolution of multicellular life. Certainly, all true multicellular organisms are composed entirely of eukaryotic cells. Taken together, these bare facts suggest that the prokaryotes were the first primitive cells, and that the more advanced eukaryotes evolved from them later, gradually accruing complexity.

Many features of the eukaryotes support this conclusion. During the mid-1880s, the German biologists Schmitz, Schimper and Meyer proposed that chloroplasts (the photosynthetic organelles of plants) were derived from cyanobacteria. In 1910, the Russian biologist Konstantin Mereschkovsky took this view forwards, arguing that eukaryotic cells had evolved from a union of various different types of bacteria. With only rudimentary microscopic techniques to back his arguments, however, he failed to convince the biological establishment. His ideas stagnated for nearly 70 years until the late 1970s, when Lynn Margulis, at the University of Massachusetts at Amherst, championed the cause and marshalled evidence that organelles were once free-living bacteria, at a time when new molecular methods could prove the case.

It is now accepted, as one of the basic tenets of biology, that chloroplasts and mitochondria (the energy 'power-houses' of the cell) were once free-living bacteria. Many details betray their former status. Mitochondria, for example, retain their own genetic apparatus, including their own DNA, messenger RNA, transfer RNA and ribosomes. These bear witness to their bacterial origins. Mitochondrial DNA, like bacterial DNA, comes packaged as a single circular chromosome, and is naked (not wrapped in proteins). The sequence of letters in its genes is closely related to the equivalent genes in a class of purple bacteria called the α -proteobacteria. Mitochondrial ribosomes also resemble those of the proteobacteria in their size and detailed structure, as well as their sensitivity to antibiotics such as streptomycin. Again, like bacteria, mitochondria divide simply by splitting in half, usually at different times from each other and from the rest of the cell.

Despite these atavistic features, mitochondria have lost almost all their former independence. Two billion years of shared evolution has left the mitochondrial genome with little to call its own. The closest bacterial relatives, the α -proteobacteria, have a total of at least fifteen hundred genes, whereas most mitochondria have retained less than a hundred genes. As we saw in Chapter 3, evolution tends towards simplicity as readily as complexity. Any bacterial genes unnecessary for survival inside the eukaryotic cell would have been quickly lost, since genes in the nucleus could take over their role without competition or antagonism. Other mitochondrial genes have physically moved to the nucleus – 90 per cent of mitochondrial genes now reside in the cell nucleus. Why the remaining 10 per cent of genes stayed put in the mitochondria is something of an enigma, but their location must confer some sort of advantage (see Footnote 7, Chapter 13).

As far as our story is concerned, the horizontal movement of genes around the gene pool in this way (from free-living bacteria into eukaryotes) has a profound impact on the web of relations between living things. Clearly, the nuclei of eukaryotic cells contain bacterial genes abstracted from the mitochondria. Any attempt to trace the earliest genetic heritage of eukaryotes based on these genes would be misleading: they are a

late graft rather than an ancestral trait of the eukaryotes. But in many respects the mitochondrial genes are easy to track. At least we know their context and their function. What we don't know is how many other genes were once subsumed in this manner; or indeed, how to tell which ones they are. This is the problem posed by lateral gene transfer in general, and it is widespread and difficult. If genes circulate with the freedom of money in an economic union, it becomes virtually impossible to trace the descent of an organism – it may have inherited its genes vertically from its own ancestors, or laterally from an unrelated species. The further back we go in time, the more twisted and obscure this web becomes.

In the late 1960s, the web of genetic relatedness between organisms came to obsess a young researcher at the University of Illinois, a biophysicist turned evolutionist by the name of Carl Woese. Woese recognised that if entire genomes could be sequenced, the 'average' relatedness of different species might still shine through the superimposed layers of lateral gene movement. At the time, however, sequencing such a massive number of genes was not feasible. What was needed instead was a single gene that could be relied upon to stay put – a gene that would not be transmitted sideways, but only vertically to the next generation. The fate of such a gene would be linked irrevocably with individual lineages, allowing, in principle, a grand reconstruction of all evolution.

This rare gene would also need to be highly resistant to change. The problem here is that the sequence of letters in genes drifts over evolutionary time, as a result of random mutations. In terms of the function of their products, most genetic mutations are harmful, but some are 'neutral' and a few are beneficial. Since neutral or beneficial changes are not penalised by natural selection, they can accumulate over time. Where two species differ in but a detail, the assumption is that at least one of the genes mutated since divergence.

The detailed sequences of equivalent genes in different species drift apart over time. For example, the genes encoding haemoglobins have diverged at a rate of about 1 per cent every 5 million years. This means that close relatives, which diverged only recently, have

similar haemoglobin sequences, whereas distant relatives have quite different haemoglobins. Similar patterns apply to other genes, such as the respiratory protein cytochrome c. Our gene for cytochrome c is approximately 1 per cent different from chimpanzees, 13 per cent different from kangaroos, 30 per cent different from tuna fish and 65 per cent different from *Neurospora* fungus. Clearly, at this rate, genetic drift may result in the complete loss of any sequence similarity between genes over billions of years, even if they once shared a common ancestor.

Some genes drift faster than others. The fastest changes take place in junk DNA, since these sequences do not code for a product and so cannot be subjected to the restraining influences of natural selection. On the other hand, a few genes are so central to the life of the cell – as structurally important as a cantilever – that almost any tampering is detrimental. Since any cell is likely to pay with its life for such changes, the ‘cantilever genes’ are the least likely to drift. Changes are almost never passed to the next generation because almost all the affected cells die. Even so, on very rare occasions, a change might occur that is not penalised by natural selection. Such unusual changes would accumulate infinitesimally slowly over billions of years in different species, eventually producing a branching tree of relationships that preserves a record of the earliest evolutionary patterns.

Does such a gene exist? Woese reasoned that cells depend on a supply of building materials in the same way that a society depends on a supply of bricks and steel to build schools, factories and hospitals. Just as society would quickly grind to a halt if no building materials were available, Woese argued that life is unthinkable without protein building blocks, or the DNA code to ensure the subtlety and continuity of protein function. Protein synthesis must be one of the most ancient and fundamental aspects of life, so it is no surprise to find that the pathways of protein synthesis are tightly integrated into the workings of the cell. Since any alterations in the genes controlling protein synthesis are likely to be penalised by death, these genes, more than any others, are likely to stay put, rather than moving around the gene pool horizontally or drifting genetically.

We have seen that proteins are built on ribosomes. Ribosomes themselves are made from a mixture of proteins with yet another form of RNA, called ribosomal RNA. Both the proteins and the ribosomal RNA are encoded by DNA and so both are subject to the restraints of natural selection. Woese recognised that of all the components of a cell, ribosomes were the closest approximation to a cantilever – absolutely indispensable to all aspects of cellular function – and were therefore highly unlikely to drift or wander horizontally around the gene pool. Furthermore, because the sequence of letters in ribosomal RNA is an exact replica of the gene, ribosomal RNA sequences could be compared directly, without recourse to the genes themselves. In the 1960s and 1970s this was invaluable, as ribosomal RNA was much easier to isolate and sequence than the parent genes. Thus, Woese settled on ribosomal RNA as a yardstick of evolution. He set about comparing ribosomal RNA sequences from library databases and from his own lab, to produce a map of the genetic relatedness of all life. This grand objective was taken up by many labs, and the project quickly gathered momentum.

Along with everyone else working in the field, Woese expected to uncover an ancient genetic link between the bacteria and the eukaryotes – something analogous to the clear relationship between mitochondria and α -proteobacteria. Two great surprises were in store. First, the gap between the two domains continued to yawn. No microbial missing link could be found, nor indeed any continuum between bacterial and eukaryotic ribosomes, as would be expected if the eukaryotes had simply evolved from bacteria. Instead, the RNA sequences clustered obstinately into two distinct groups, as if they had nothing in common. This could only mean that the split between bacteria and eukaryotes had taken place very early indeed, perhaps not long after the first stirrings of life itself. This in turn meant that the eukaryotes could not have evolved gradually from bacteria over two billion years, as everyone had expected. It must have happened very quickly and very early.

Then came the second surprise, announced by Woese and Fox in 1977, and now seen as one of the great paradigm shifts of science. A deep divide emerged within the prokaryotic domain itself. A little known group of prokaryotes, most of which inhabited

extreme environments such as hot springs and hypersaline lakes, confounded all expectations when their ribosomal RNA was analysed. The analyses showed that they shared little more with the bacteria than the absence of a nucleus. As more of their ribosomal RNA was sequenced and compared, it became clear that the divergence was not just a new kingdom within the prokaryotes, but something much more basic – an entirely new domain, which has become known as the Archaea. Today, instead of five kingdoms, we recognise three great domains of life: the Bacteria, the Archaea and the Eukaryotes. We ourselves, as animals, occupy no more than a small corner of the Eukaryotes.

THE EXISTENCE OF THE ARCHAEA allows us to paint a far more convincing picture of LUCA. We can now compare the characteristics of three different domains. The archaea are obviously comparable to the bacteria in that they lack a cell nucleus and so are defined as prokaryotes. The organisation of their genes is also similar to bacteria: they have a single circular chromosome, they cluster groups of related genes into operons, and they carry little junk DNA. Other aspects of their organisation, such as the structure and function of proteins in the cell membranes, bear a more superficial resemblance to bacteria. Most archaea have a cell wall, but unlike bacteria a few do not. Again, unlike bacteria, the cell wall contains no peptidoglycans. The similarities quickly tail away.

In other respects, the archaea lie much closer to the eukaryotes. Although they do not have as many genes as eukaryotes, archaea have on average more than twice as many genes as bacteria. The DNA of archaea is not naked, but is wrapped in proteins similar to those used by eukaryotes. The detailed mechanism of DNA replication and protein synthesis is much closer to the eukaryotes. For example, their genes are recognised by specific proteins, which initiate and propagate the transcription of DNA into messenger RNA in a way that is closely analogous to the eukaryotic method. The protein constituents of the ribosomes also resemble those of the eukaryotes in their structure. Other details of ribosomal function, including the initiation of protein synthesis, the elongation of protein chains, and the termination steps, parallel the eukaryotic process. Finally, and most convincingly of all, genetic analyses of so-called paralogous gene pairs

– the products of gene duplications in a common ancestor, followed by divergent evolution in different groups of descendants – indicates that archaea are indeed specific relatives of eukaryotes. In essence, archaea are prokaryotes that behave like eukaryotes. They are as close to a missing link as we are ever likely to find [FIGURE 7].

What does all this say about the identity of LUCA? It seems likely that the split between the archaea and the bacteria occurred very early in the history of life, perhaps 3.8 to 4 billion years ago. We assume that both the archaea and the bacteria retain some of the original features of LUCA herself. Calculations suggest that the eukaryotes split from the archaea later, perhaps around 2.5 to 3 billion years ago, since they share far more fundamental traits with the archaea. We know that the eukaryotes acquired mitochondria and chloroplasts around 2 billion years ago by engulfing bacteria. We also know that some of these bacterial genes fused with existing genes in the nucleus of eukaryotic cells. Here we return to the problem of lateral gene transfer. If the eukaryotes are essentially a fusion of archaea and bacteria, then it is plain that lateral gene transfer has taken place across domains. If we wish to draw a portrait of LUCA by comparing the properties of the different domains, can we be sure that they are not completely mixed up?

Luckily, there is some evidence that lateral gene transfer is not common across domains. The development of the eukaryotes seems to have been a singular event, possibly propelled by the unique environmental conditions around the time of the snowball Earth of 2.4 billion years ago (Chapter 3). In general, however, the archaea have kept themselves very much to themselves, and give every appearance of having changed little since the beginnings of time. No archaea is pathogenic, which means that they do not cause infectious diseases in eukaryotes, and so do not mix their genes with eukaryotes in the course of intimate war. Nor do they compete with bacteria in other settings. Their predilection for extreme conditions isolates them from most other organisms, even bacteria. Hyperthermophiles, such as *Pyrolobus fumaris*, live at searing temperatures, well over 100°C, and high pressures in deep sea hydrothermal vents. Other archaea, such as *Sulfolobus acidocaldarius*, add acidity to the heat and live in

sulphur springs in places like Yellowstone National Park, at pH values as low as 1, the equivalent of dilute sulphuric acid. At the other end of the pH scale, some archaea thrive in the soda lakes of the Great Rift Valley in East Africa and elsewhere, at a pH of 13 and above – enough to dissolve rubber boots. The halophiles are the only organisms that can live in hypersaline salt lakes, such as the Great Salt Lake in Utah and the Dead Sea. The psychrophiles prefer the cold and grow best at 4°C in Antarctica (their growth is actually retarded at higher temperatures).

Many of these favoured environments have barely changed for billions of years. Without calamity or competition, the selection pressure for innovation and change must have been negligible. While it is true that some archaea do live in more normal environments – among plankton in the surface oceans, for example, and in swamps, sewage and the rumen of cattle – the genes of their extreme and reclusive cousins have surely had little traffic with the rest of life.

The extraordinary properties of archaea quickly stimulated scientific and commercial interest, and the field blossomed as a distinct discipline during the 1990s. Enzymes that function normally at high temperatures and pressures are an answer begging for an application. Already enzymes extracted from archaea have been added to detergents and used for cleaning up contaminated sites, such as oil spills. To enlist the skills of a microbe on an industrial scale, however, requires a working knowledge of its genes. Complete genome sequences have now been reported for representatives of all the known groups of archaea. These sequences at once confirm the great antiquity of archaea, and their splendid isolation over the aeons. But the greatest surprise is how many genes the archaea do have in common with bacteria.

Considering the different forms of aerobic and anaerobic respiration alone (by which I mean energy production at a cellular level), at least 16 genes have been found in both archaea and bacteria. From the close similarities in their sequences, it seems likely that these genes were present in LUCA, and were later inherited by both archaea and bacteria, as they diverged from each other to occupy their distinct evolutionary niches.

This conclusion – that the 16 respiratory genes were present in LUCA herself – is supported by two independent lines of evidence, as argued by Jose Castresana and Matti Saraste of the European Molecular Biology Laboratory in Heidelberg.

The first line of evidence relates to evolutionary trees. The genetic similarities between the 16 respiratory genes can be used to construct a tree of relatedness, as if we were to construct a family tree based on a single trait like brown eyes. The family tree constructed from the respiratory genes is then superimposed over the evolutionary tree based on ribosomal RNA sequences. If the respiratory genes had been passed horizontally by lateral gene transfer, then closely related respiratory genes would be found in organisms that were otherwise only distantly related to each other. Put another way, the evolutionary histories of the respiratory genes would differ from the true evolutionary roots of their host organisms, just as the mitochondrial genes differ from the nuclear genes of eukaryotes. On the other hand, if the respiratory genes had stayed put in their respective organisms, then the evolutionary trees constructed from ribosomal RNA and the respiratory genes should correspond to each other. This is in fact the case: the evolutionary trees of the respiratory genes that have been analysed so far do broadly correspond to the reference tree constructed from ribosomal RNA, implying that lateral gene transfer did not occur between bacteria and archaea.

The second line of evidence relates to more recent metabolic innovations like photosynthesis. LUCA, it seems, could not photosynthesise. No form of photosynthesis based on chlorophyll is found in any archaeal group. A completely different form of photosynthesis, based on a pigment similar to the photoreceptors in our eye, called bacteriorhodopsin, is practiced by the archaea Halobacteriaceae but is not found in any bacterial group. These disparate forms of photosynthesis presumably evolved independently in bacterial and archaeal lineages some time after the age of LUCA, and subsequently remained tied to their respective domains. If a metabolic innovation as important as photosynthesis did not cross from one domain to another, there is no reason to think that other forms of respiration would have done. We should certainly be wary of postulating that respiratory genes crossed domains unless we have evidence

that they did so; and the evidence from evolutionary trees suggests that they did not.

If we accept that lateral gene transfer between archaea and bacteria has been extremely rare, then the 16 respiratory genes must have been present in LUCA, and were later vertically inherited by diverse lines of bacteria and archaea. Since these genes code for proteins involved in generating energy from a variety of compounds, including nitrate, nitrite, sulphate and sulphite, LUCA must have been a metabolically sophisticated organism. One gene in particular, however, shares a striking sequence similarity in archaea and bacteria, and it is this that Castresana and Saraste have used to paint an unexpected portrait of LUCA.

The gene codes for a metabolic enzyme called cytochrome oxidase. This enzyme couples electrons onto oxygen, to produce water, in the final step of aerobic respiration. If cytochrome oxidase was present in LUCA, then the logical, if ostensibly nonsensical, conclusion is that aerobic respiration evolved before photosynthesis. LUCA could breathe before there was any free oxygen. As Castresana and Saraste put it, no doubt relishing every word, 'This evidence, that aerobic respiration may have evolved before oxygen was released to the atmosphere by photosynthetic organisms, is contrary to the textbook viewpoint.'

Nick Lane

Biochemist and writer

www.nick-lane.net