Current Biology
Magazine

CellPress

Primers

Archaea

Laura Eme and W. Ford Doolittle*

A headline on the front page of the *New York Times* for November 3, 1977, read "Scientists Discover a Way of Life That Predates Higher Organisms". The accompanying article described a spectacular claim by Carl Woese and George Fox to have discovered a third form of life, a new 'domain' that we now call Archaea. It's not that these microbes were unknown before, nor was it the case that their peculiarities had gone completely unnoticed. Indeed, Ralph Wolfe, in the same department at the University of Illinois as Woese, had already discovered how it was that methanogens (uniquely on the planet) make methane, and the bizarre adaptations that allow extremely halophilic archaea (then called halobacteria) and thermoacidophiles to live in the extreme environments where they do were already under investigation in many labs. But what Woese and Fox had found was that these organisms were related to each other not just in their 'extremophily' but also phylogenetically. And, most surprisingly, they were only remotely related to the rest of the prokaryotes, which we now call the domain Bacteria (Figure 1).

It was Woese's new approach to doing phylogeny that produced this completely unexpected and very newsworthy result. His method, which seems impossibly tedious now that we can sequence whole genomes over lunch, involved dangerous amounts of radioisotopes, highly flammable solvents, and weeks to months of effort to produce, for any cultured microbial species, an 'oligonucleotide catalog'. This simple list of 16S rRNA-derived sequences, terminating in G and between one and twenty bases in length, was then compared to previously obtained catalogs for other species, and the number of shared oligonucleotide sequences was used to prepare a tree with computational methods laughably primitive by today's standards. With unrelenting dedication, Woese and his technicians, students and postdocs (Fox was then one of the latter) managed by 1977 to accumulate enough data to conclude that prokaryotes (then thought to be synonymous with bacteria) were actually composed of two very distinct kinds, no more similar to each other than either was to eukaryotes (as determined by cataloging their 18S rRNA). They called the two domains "Eubacteria" and "Archaebacteria" but later, to dispel the notion that the latter were just another kind of bacteria, switched to "Bacteria" and "Archaea". These names have stuck, though a battle continues over whether another word — prokaryotes, meaning Bacteria plus Archaea together — has any legitimate use. And it is no longer believed that Archaea are any older than Bacteria, as their name and the *New York Times* headline might imply.

At first, the three-domain view had to struggle against traditionalists like Ernst Mayr, who saw the prokaryote–eukaryote dichotomy as evolution's most important division, and Lynn Margulis, who never accepted the primacy of molecular over ultrastructural and physiological assessments of relatedness. But, with the striking consilience of the rRNA data, the subunit structure of archaeal RNA polymerases (much more eukaryote-like, as shown by Wolfram Zillig) and the bizarre character of archaeal membrane lipids (composed of isoprenoid rather than straight chain fatty acids in ether rather than ester-linkage to glycerol with unique stereochemistry) and cell envelopes (made of a substance generally other than murein, as stressed by Otto Kandler), the uniqueness and coherence of Archaea became widely accepted. Wolfe thus referred to an evidential "three-legged stool" on which the archaeal concept could rest secure. Now, probably all textbooks show Life as comprising the domains Bacteria, Archaea and Eukarya, with the last two the more closely related.

Early on, we believed that Archaea were all extremophiles of one sort or another (that is, thriving in high temperature, high salt, low or high pH, absolutely no oxygen or combinations thereof) and that maybe something about their physiology made them poor competitors with bacteria in more normal niches. This was in some ways a self-fulfilling prophecy — in those early years it was largely in such nasty places that we looked for them. What turned the tide was the PCR amplification and Sanger-based sequencing of 16S rRNA genes present in environmental samples. This cultivation-independent approach, introduced by Norman Pace, has revolutionized environmental microbiology. Over 85% of the more than three million rRNA sequences in the current databases were obtained as 'environmental sequences'. Many new archaeal and bacterial lineages (including phyla) have been discovered, most of which still have no cultured representatives. In 1992, PCR amplification independently led Ed Delong and Jed Fuhrman to the discovery of new archaeal lineages — marine groups I and II — from marine samples. This was the beginning of the realization that archaea also thrive in non-extreme (mesophilic) and even very cold environments, including cold terrestrial springs, coastal and deep marine waters, lakes, soil, and as commensals or symbionts to many animals. Archaea are everywhere, though curiously there seem to be no frank pathogens among them.

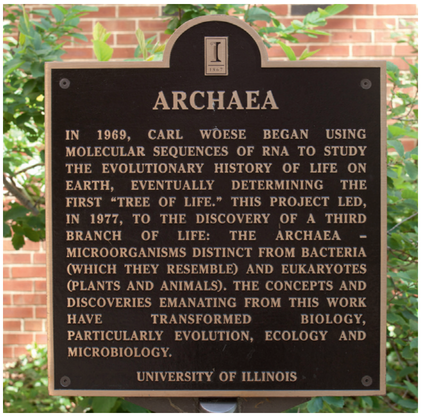
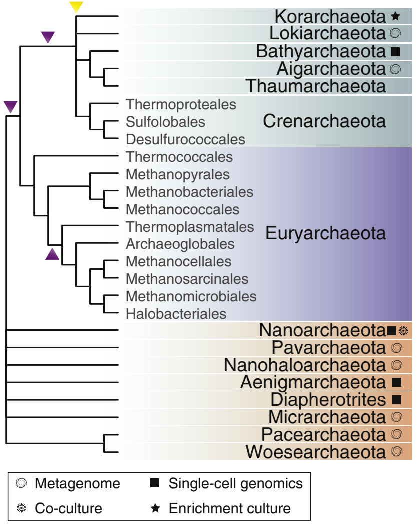


Figure 1. Commemorating the Archaea.
Plaque in front of Burrill Hall at the University of Illinois at Urbana-Champaign, where Carl Woese began oligonucleotide cataloguing of 16S rRNA (photo by W.F. Doolittle).

The first archaeal genome to be sequenced was that of *Methanococcus jannaschii* (published in 1996). Nearly 400 complete or semi-complete archaeal genomes are now publicly available, 75% of which have been obtained in the past 4 years. Earlier sequences were obtained from archaea in pure culture, co-culture or enriched



Current Biology

Figure 2. Schematic phylogenetic tree of Archaea.
Lineages belonging to the TACK supergroup are shaded in teal, Euryarchaeota are in purple, while the orange shadings indicate lineages of nano-sized archaea proposed to form the ‘DPANN’ supergroup. Polytomies (mutiple apparently simultaneous branchings) highlight uncertainties about phylogenetic relationships between lineages. The tree is unrooted and purple triangles indicate favored hypotheses regarding the position of the root; between Euryarchaeota and the rest, within Euryarchaeota, and at the base of, or within, DPANN. The yellow triangle points at a possible branching point for the eukaryotic lineage (either sister to the TACK supergroup, or within it — possibly sister to the Lokiarchaeota) with the other favored one being at the base of all Archaea (not shown). Symbols indicate the type of DNA collection method used to determine genome sequence (metagenome, co-culture, single-cell genomics or enrichment culture), unless pure culture was used for at least one representative organism (no symbol). Additional lineages uncovered via 16S rRNA-based diversity surveys alone, without extensive genomic data available, are not depicted.

culture. Recent methodological and computational advances have driven the discovery of archaeal lineages at a faster pace, mainly through two new methods that allow sequencing of genomes from uncultivated (and even unseen) organisms. The first is metagenomics, and uses various binning and assembly methods to put together individual genome sequences from complex, environmental DNA sequence reads. The second method, single-cell genomics, uses DNA that is amplified from individual cells that have been isolated by fluorescence-activated cell sorting, for instance.

These methods have led to the identification of numerous lineages that escaped detection in 16S rRNA surveys because of their extremely low relative abundance or mismatches to the designed archaeal 16S rRNA gene primer sequences.

Phylogenetic diversity

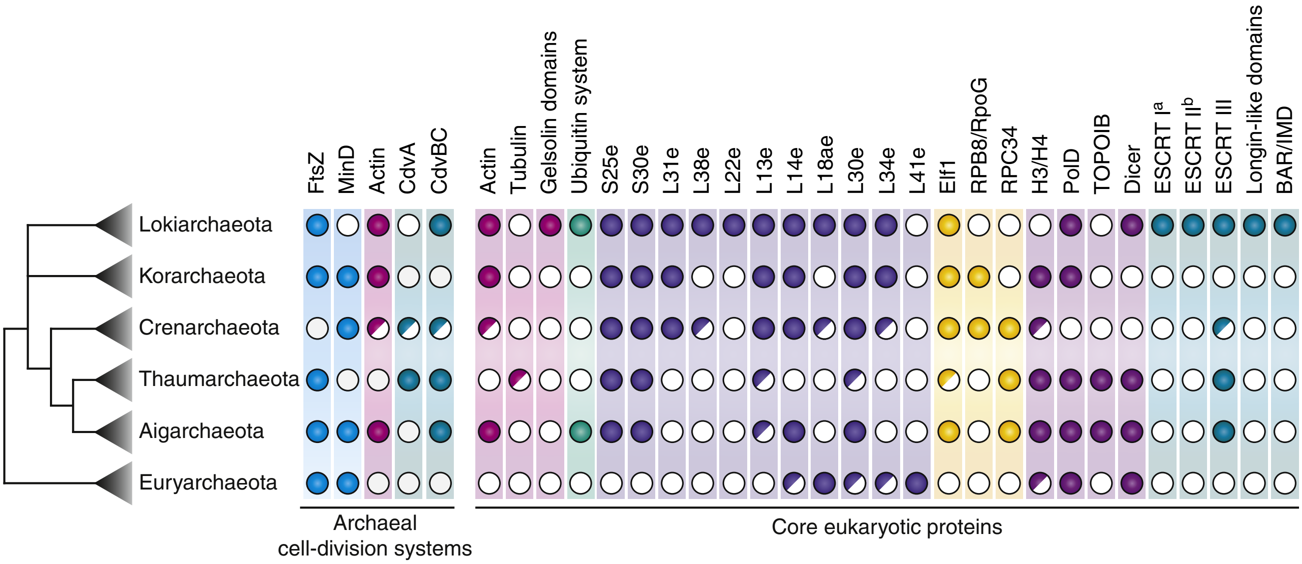
Initially, and for some time, 16S rRNA-based phylogenies of cultivated Archaea showed two groups, the Euryarchaeota and the Crenarchaeota, separated by a large evolutionary distance. Since these early trees, phylogenetic methods have grown enormously more sophisticated, and the confident phylogenetic placement of new lineages is no longer inferred from the analysis of the small subunit rRNA genes alone. Indeed, a single gene does not bear sufficient phylogenetic signal to fully resolve all nodes of the tree of life, especially the most basal ones, which correspond to ancient evolutionary events. This can be overcome by the use of phylogenomic approaches that involve simultaneously analyzing several gene or protein sequences (up to a few hundred) conserved in most or all genomes.

Concomitantly, culture-independent detection and phylogenomic analyses have produced a bonanza of new major lineages only distantly related to the Crenarchaeota and Euryarchaeota. The years 2006–2008 saw the publication of genome data from two other phyla — the Korarchaeota, a poorly characterized group of uncultured hyperthermophilic anaerobic archaea, and the Thaumarchaeota, a widespread microbial group of mesophilic marine archaea previously included in the Crenarchaeota. Three years later, a genome assembled purely from metagenomic data was proposed to represent a novel phylum — Aigarchaeota, a sister-group of the Thaumarchaeota. Together, the Thaumarchaeota, Aigarchaeota, Crenarchaeota, and Korarchaeota were found to form a monophyletic group, generally referred to as the ‘TACK’ superphylum or Proteoarchaeota. However, other new lineages have subsequently been suggested to branch within this group, such as the proposed phylum Bathyarchaeota, frequently detected in deep subsurface sediments, and the very recently described Lokiarchaeota, identified from hemipelagic-glaciomarine sediments (Figure 2).

In the past few years, metagenomics and single-cell genomics have also turned up many intriguing tiny (in terms of cell and/or genome size) archaea, including Parvarchaeota, Aenigmarchaeota, Diapherotrites, Nanohaloarchaeota, Pacearchaeota, Woesearchaeota, and Micrarchaeota (Figure 2). These ‘nano’ organisms (including the previously isolated Nanoarchaeota) are found in diverse environments, such as acid mine drainage biofilms or high-salt environments, and are characterized by their very small cells (~ 400–500 nm) and genomes (~ 550–1,200 genes, versus 1,500–3,000 for most other archaea, and even > 5,000 for some euryarchaeal lineages), as well as rapidly evolving gene sequences. Simultaneous with their discovery, it was proposed that these ultrasmall archaea (sometimes referred to as the ‘DPANN’ clade) represent early diverging lineages among Archaea, possibly pointing at a small and ‘simple’ archaeal ancestor. However, their phylogenetic placement is still hotly debated: their deep-branching position could be the result of tree reconstruction artefacts caused by unusually small gene content combined with a fast rate of evolution. Some authors propose that these organisms are in fact representatives of various other major archaeal lineages that evolved by convergence toward a similarly small gene content.

Cellular, biochemical and ecological diversity

The features initially found to define Archaea presumably still hold for the majority of organisms identified by 16S rRNA sequences, though this needs to be carefully confirmed: Wolfe’s three-legged stool remains sturdy as far as anyone currently knows. Furthermore, there are now many other systems or structures found in a variety of Archaea that distinguish them from most bacteria. For example, archaeal flagella are not homologous to bacterial flagella despite striking structural and functional similarity. Eukaryote-like proteasomes are found in most Archaea but in only a few



Current Biology

Figure 3. Taxonomic distribution of archaeal orthologs of cell-division system components and of eukaryotic signature proteins.
Archeal cell-division components are shown on the left: FtsZ-based systems are depicted in blue, actin-based in pink and Cdv-based in teal. Eukaryotic signature proteins are shown on the right: proteins involved in cell-shape determination (pink); protein recycling (green); translation (purple); transcription (yellow); DNA packaging, replication and repair (pink-violet); and cell division, vesicle formation and membrane remodelling (teal). Some components appear in both categories. Phylogenetic relationships are depicted as a consensus of recent phylogenomic analyses. Color-filled circles indicate the presence of homologues in all members of a lineage, whereas half-filled and white circles denote patchy distribution and absence of homologues, respectively. ^{a}Vps28-like, ^{b}EAP30 domain and Vps25.

bacterial phyla. Moreover, an analysis of gene family content in diverse archaeal and bacterial genomes showed a non-random distribution, some families being predominantly archaeal and some predominantly bacterial. That said, there may be extensive Bacteria-to-Archaea lateral gene transfer, especially in some newly discovered planktonic Thaumarchaeota and Euryarchaeota, and there are very few gene families that, by being found in all archaea and no bacteria (or vice versa), might be said to define domain ‘essences’. Within the Archaea there is diversity that touches essential cellular features, such as the proteins involved in DNA maintenance, protein ubiquitinylation, transcription and translation. For instance, archaea possess one or more of three distinct cell division systems (Figure 3). The bacterial-like division system involving FtsZ (a distant homolog of eukaryotic tubulins) can be found in most TACK archaea, although MinD, one of the main known interactors of FtsZ, is absent in Thaumarchaeota. Some archaea encode a CdvABC system, CdvB and CdvC being homologues of the eukaryotic ESCRT-III and Vps4 proteins. Cdv proteins are broadly, although patchily, distributed among Thaumarchaeota, Lokiarchaeota, Euryarchaeota, Crenarchaeota and Aigarchaeota, the latter lacking CdvA. Finally, some crenarchaeotal lineages lack both of these systems, suggesting the existence of a third mode of division, putatively actin-based or as yet unidentified.

Ecological diversity surveys using 16S rRNA amplification document the occurrence of Archaea in both extreme and temperate habitats, making them truly ubiquitous in the biosphere. Archaea represent more than 20% of oceanic prokaryotes and likely are the most abundant organisms in marine subsurface sediments and geothermal habitats. In addition, environmental genomic approaches have yielded valuable insights into their metabolic capabilities and roles in major geochemical cycles. Archaea display a wide range of lifestyles, including anaerobic and aerobic respiration, fermentation, chemoautotrophy, heterotrophy, and photoheterotrophy. Through these various energy metabolisms, many archaea are able to fix carbon from inorganic sources, making them major ecological actors in the Earth’s global geochemical cycles. They play crucial roles in the carbon, nitrogen and sulfur cycles, and significantly influence greenhouse gas emissions.

Archaea are uniquely responsible for two metabolic pathways of major importance for the circulation of chemical nutrients. First, methanogenesis, thought to be one of the oldest metabolisms on Earth, is carried out by a subset of euryarchaeota and generates ~85% of the methane on the planet. The second, anaerobic methane oxidation, is achieved through a syntrophic association between sulphate-reducing bacteria and anaerobic methane oxidizing archaea (ANME), the latter being closely related to methanogens and seeming to use their methanogenic machinery in reverse. A third key metabolic contribution of Archaea is to the nitrogen cycle, through aerobic ammonia oxidation carried out by the extraordinarily prolific ammonia-oxidizing archaea (AOA), which are members of Thaumarchaeota and are found in virtually all aerobic habitats. Their discovery actually points to

Archaea, and not Bacteria, as the major nitrifiers at a global level.

LACA and the evolution of Archaea

It is now clear that Archaea are eminently versatile, with a whole range of genes, metabolisms, cellular systems and lifestyles, which often appear to be broadly yet patchily distributed across this domain of life. Given this diversity, it has been difficult to infer the nature of the last archaeal common ancestor (LACA) and how it gave rise to these diverse descendant lineages. Essentially, two scenarios can be contemplated. The LACA could have possessed a fairly large genome — larger than most extant archaea — encoding a collection of features far more complex than the ones found in modern cells, with its descendant lineages mainly evolving through the loss of many of these features via genomic streamlining. Alternatively, the LACA was a simpler organism, with a genome just as small, or smaller, than today's typical archaea. In such a case, the patchy presence of certain features in the various archaeal lineages would have arisen from a combination of convergent evolution and lateral gene transfer.

Although lateral gene transfer between distant mesophilic lineages is the favored hypothesis to explain how these have independently adapted to cooler temperatures from a LACA believed to be thermophilic, the possibility of an archaeal ancestor more complex than its present-day descendants has been generally well received among researchers. In particular, comparative genomic analyses relying on a sophisticated evolutionary model to infer gene gains and losses during archaeal evolution suggested a general tendency for these genomes to evolve through massive reduction interspersed with periods of explosive genome expansion.

A crucial aspect of these inferences about LACA's features is that they heavily depend on the placement of the root of the Archaea (that is, the placement of LACA itself in the universal tree). While it has long been assumed that this root falls between Euryarchaeota and all other lineages, the recent windfall of genomic data from novel lineages, accompanied by ever-improving phylogenetic methods, has raised some doubt. Some alternative results suggest that LACA could fall within Euryarchaeota (Figure 2). Knowing the true position of the root will greatly impact inferences regarding the archaeal ancestor, as well as the origins of major archaeal lineages. For instance, a root between Euryarchaeota and other archaea leads to a reconstruction of the gene repertoire of LACA that does not support the ability of this organism to carry out methanogenesis (Figure 2). In contrast, a root falling within Euryarchaeota, and more specifically between two clades of methanogenic archaea, suggests that this metabolism could be much older than previously thought.

An archaeal ancestor for eukaryotes?

Few questions have fascinated biologists like that of the origin of the eukaryotes, and it is generally accepted that they share a specific evolutionary link with Archaea. In particular, many 'informational genes' (those involved in the transmission and expression of genetic information, such as translation, replication and transcription) are more similar between Archaea and Eukarya than to their bacterial homologs, when these even exist. It is these genes, rather than the bacterially derived 'operational' genes (those participating in housekeeping functions and metabolism), that are thought to track the evolutionary history of the eukaryotic nucleus and the cytoplasm under its control. Concerning that history, two opposing ideas have been alternatively supported by phylogenetic analyses.

The first model, proposed in the 1990s by Woese and colleagues, advocates for Eukarya and Archaea being two distinct domains, each monophyletic, that descended from a common ancestor to the exclusion of Bacteria (the third domain). According to this scenario, features shared between Archaea and eukaryotes represent ancestral characters that were inherited from their common ancestor, which was technically (and possibly phenotypically) neither archaeal nor eukaryotic. While phylogenetic analyses have frequently supported this model, they have also suggested alternative scenarios, in which eukaryotes evolved from *within* the Archaea, though the nature of the archaeal lineage ancestral to eukaryotes differs depending on the analysis. Such scenarios would be more difficult to reconcile with the presence of bacterial-like (and not archaeal-like) lipids in eukaryotic membranes: they would require two radical alterations in membrane composition rather than just one (from one glycerol stereoisomer to the other in the branch leading to LACA) implicated in the traditional three-domain tree. Membrane biochemistry of any archaea branching more deeply than eukaryotes becomes a topic of great interest.

In fact, obtaining a definitive answer to this question has been challenged by the uneven distribution of eukaryotic characters among the archaeal lineages, with no single organism possessing them all (Figure 3). However, most recent phylogenomic analyses seem to support an origin of eukaryotes from within Archaea, as they have uncovered a specific phylogenetic affinity between the TACK group and eukaryotes, although the relationships among these lineages are still ambiguous. The very recently discovered Lokiarchaeota, a novel candidate archaeal phylum belonging to the TACK supergroup, seems to harbor more eukaryotic features than any other known archaeal lineage, and apparently forms a monophyletic group with Eukarya in phylogenomic analyses (Figure 2). If further data confirm these results, this will represent a major breakthrough in tackling the question of the origin of eukaryotes.

Final thoughts

Woese and Fox's discovery of the Archaea is, without doubt, the signal event in 20th century microbial systematics. Indeed it affects all of biology, given that Archaea comprise such a large part of Life's diversity and are so important to Earth's biogeochemistry. Whether anything unites Archaea with Bacteria as 'prokaryotes', and whether or not any such shared features are primitive (present in the common ancestor of all extant life) or derived (arrived at convergently through streamlining, or by lateral gene transfer between the two domains) remain open questions. Similarly, whether Archaea are a sister group to the 'host' (or nucleocytoplasmic) component of eukaryotes, or instead gave rise to the latter from a *bona fide* archaeal lineage (making them paraphyletic) will not be settled for some time. If indeed eukaryotes arose *within* Archaea, then adherence to cladistic principles would require renaming. Eukaryotes and the archaeal group to which they are sisters would be one kingdom-level clade (‘eocytes’ has precedence as a name), and the rest of the Archaea would be one or more other kingdoms, depending on the structure of the tree. As to a phenotypic definition, we would need to do a lot more physiological, biochemical and ultrastructural work to make sure that Wolfe’s three-legged stool has not at last begun to wobble.

FURTHER READING

Brochier-Armanet, C., Forterre, P., and Gribaldo, S. (2011). Phylogeny and evolution of the Archaea: one hundred genomes later. Curr. Opin. Microbiol. *14*, 274–281.

Csűrös, M., and Miklós, I. (2009). Streamlining and large ancestral genomes in Archaea inferred with a phylogenetic birth-and-death model. Mol. Biol. Evol. *26*, 2087–2095.

Deschamps, P., Zivanovic, Y., Moreira, D., Rodriguez-Valera, F., and López-García, P. (2014). Pangenome evidence for extensive inter-domain horizontal transfer affecting lineage-core and shell genes in uncultured planktonic Thaumarchaeota and Euryarchaeota. Genome Biol. Evol. *6*, 1549–1563.

Gribaldo, S., Poole, A.M., Daubin, V., Forterre, P., and Brochier-Armanet, C. (2010). The origin of eukaryotes and their relationship with the Archaea: are we at a phylogenomic impasse? Nat. Rev. Microbiol. *8*, 743–752.

Koonin, E.V., and Yutin, N. (2014). The dispersed archaeal eukaryome and the complex archaeal ancestor of eukaryotes. Cold Spring Harb. Perspect. Biol. *6*, a016188.

Offre, P., Spang, A., and Schleper, C. (2013). Archaea in biogeochemical cycles. Annu. Rev. Microbiol. *67*, 437–457.

Spang, A., Martijn, J., Saw, J.H., Lind, A.E., Guy, L., and Ettema, T.J.G. (2013). Close encounters of the third domain: The emerging genomic view of archaeal diversity and evolution. Archaea *2013,* Article ID 202358, 12 pages, 2013. http://dx.doi.org/10.1155/2013/202358.

Spang, A., Saw, J.H., Jørgensen, S.L., Zaremba-niedzwiedzka, K., Martijn, J., Lind, A.E., Eijk, R. Van, Schleper, C., Guy, L., and Ettema, T.J.G. (2015). Complex archaea that bridge the gap between prokaryotes and eukaryotes. Nature *521,* 173–179.

Williams, T.A., Foster, P.G., Cox, C.J., and Embley, T.M. (2013). An archaeal origin of eukaryotes supports only two primary domains of life. Nature *504*, 231–236.

Woese, C.R. (2007). The birth of the Archaea: a personal retrospective. In Archaea, Evolution, Physiology & Molecular Biology. R.A. Garrett and H.-P. Klenk, Eds. (Blackwell Publishing.)

Department of Biochemistry and Molecular Biology, Dalhousie University, P.O. Box 15000, Halifax, Nova Scotia B3H 4R2, Canada.
*E-mail: W.Ford.Doolittle@Dal.Ca

Photosynthesis and early Earth

Patrick M. Shih[1,2]

Life has been built on the evolution and innovation of microbial metabolisms. Even with our scant understanding of the full diversity of microbial life, it is clear that microbes have become integral components of the biogeochemical cycles that drive our planet. The antiquity of life further suggests that various microbial metabolisms have been core and essential to global elemental cycling for a majority of Earth’s history.

Just as geological determinants have shaped the course of biology, life has undoubtedly changed the geological course of Earth. One key biological innovation that has contributed to these geochemical cycles is the ability to use light energy to drive the metabolic process of generating chemical energy. Photosynthesis has been instrumental in the success of life on Earth, and life has had to adapt and evolve in the face of dramatic geological perturbations over billions of years. But, perhaps equally interesting to the origins and evolution of microbial metabolisms is the uncertainty concerning the implications of these metabolisms. As so much of the debate surrounding early life revolves around the origin and role of phototrophic life, this primer will focus on how the evolution of these bacterial metabolisms forever changed the paleoclimate and continues to sustain current biogeochemical cycles.

Life before oxygen

The Archean Earth (4 to 2.5 billion years ago) would be unrecognizable to most of us in the present day: an unbreathable atmosphere devoid of oxygen, the temperature warmer from an abundance of greenhouse gases, dramatically different ocean chemistries, and overall looking more like a science fiction landscape. Although all these characteristics would ultimately change on the geological timescale of billions of years, one constant that has stood throughout this unfathomable amount of time is the Sun. Fittingly, evidence for the occurrence of photosynthesis has been interpreted as signifying some of the earliest forms of life.

The Sun sustains the vast majority of life on Earth. We associate the oxygen in our atmosphere with primary producers performing oxygenic photosynthesis, but there was once a time when this life-changing metabolism had not yet evolved, and the world was devoid of oxygen. As one would expect, the predominant microbial metabolisms would have been significantly different during this period of Earth’s history.

Like all things in life, time takes a toll on rocks and fossils — especially when the timescales are over billions of years. Moreover, all Precambrian life was microbial, making identification and interpretation of these fossils incredibly difficult. Because of these harsh facts, scientists must do a bit of creative detective work to try to determine what life may have looked like, given that all we have is indirect evidence from ancient sediments to make hypotheses about early microbial metabolisms. It is important to be aware that these guesses are limited to our understanding of the diversity of life in the present day. There are still many microbes and metabolisms that we have not yet discovered. Moreover, many bacterial lineages or metabolisms may have simply gone extinct without leaving any record, thus providing us no way to understand large evolutionary gaps and adding uncertainty to our understanding of Archean life.

A good place to begin looking when trying to understand ancient microbial life is the bedrock of all metabolisms: carbon fixation. Autotrophic organisms sustain all ecosystems through the process of fixing inorganic carbon (i.e., CO_2) into organic matter. Thus, understanding the many ways Archean life gathered energy and reducing power to drive carbon fixation provides a key perspective to our fundamental understanding of early life and primordial ecosystems. Many unique scenarios have been proposed as models for how early life may have arisen, reflecting the dramatic and extreme Archean landscape. One popular example is the discovery of life near seafloor hydrothermal vents, which spew hydrogen sulfide into the environment. Bacteria have evolved to exploit sulfide oxidation to drive carbon fixation, thus providing a model for