ANNUAL REVIEWS

Annual Review of Earth and Planetary Sciences

Toward a Natural History of Microbial Life

Cara Magnabosco,[1] Fatima Husain,[2] Madeline M. Paoletti,[2] Chris Parsons,[2] Jack G. Payette,[2] Sarah L. Schwartz,[3] Erik Tamre,[2] and Gregory P. Fournier[2]

[1]Geological Institute, Department of Earth Sciences, ETH Zurich, Zurich, Switzerland

[2]Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA; email: g4nier@mit.edu

[3]Department of Civil and Environmental Engineering, University of California, Berkeley, California, USA

Annu. Rev. Earth Planet. Sci. 2024. 52:85–108

The *Annual Review of Earth and Planetary Sciences* is online at earth.annualreviews.org

https://doi.org/10.1146/annurev-earth-031621-070542

Keywords

natural history, molecular phylogenetics, microbial evolution, biomarkers, geochemistry, extinction

Abstract

For most of Earth's history life was microbial, with archaeal and bacterial cells mediating biogeochemical cycles through their metabolisms and ecologies. This diversity was sufficient to maintain a habitable planet across dramatic environmental transitions during the Archean and Proterozoic Eons. However, our knowledge of the first 3 billion years of the biosphere pales in comparison to the rich narrative of complex life documented through the Phanerozoic geological record. In this review, we attempt to lay out a microbial natural history framework that highlights recent and ongoing research unifying microbiology, geochemistry, and traditional organismal evolutionary biology, and we propose six broadly applicable principles to aid in these endeavors. In this way, the evolutionary history of microbial life—once considered only a prelude to the much more storied history of complex metazoan life in the Phanerozoic—is finally coming into its own.

- The outlines of microbial natural history are now starting to appear through the integration of genomic and geological records.

- Microorganisms drive Earth's biogeochemical cycles, and their natural history reflects a coevolution with the planet.
- Past environmental changes have induced microbial biotic transitions, marked by extinction, taxonomic shifts, and new metabolisms and ecologies.
- Microbial evolution can benefit from a historical perspective of processes and successions as established by macropaleontology.

1. INTRODUCTION

For over two centuries, it has been recognized that the diversity of fossil forms preserved in the geological record reveals that life on Earth has changed over time, constituting a natural history. However, until very recently microorganisms have been largely absent from this account. Although they were discovered in the late 1600s (for a review, see Gest 2004), it would be another 200 years before bacteria were firmly established as living things akin to macroscopic organisms (for a review, see Brock 1975), surreptitiously around the same time as Darwin's work establishing evolutionary adaptation through natural selection, in which they were largely omitted. Even today, the otherness of single-celled organisms—devoid of tissues, anatomical development, and clear sets of inherited morphological characters, but often easy to grow in culture and manipulate experimentally—has subjected them to entirely different kinds of scientific inquiry than what evolutionary biology has traditionally relied upon (for a review and discussion on the topic of Darwin and microbiology, see O'Malley 2009, Wainwright 2009). While some early attempts were made to establish a natural history of microorganisms using physiological and morphological characteristics (for a review, see Woese 1987), in the absence of modern molecular methods, these efforts had limited power, for reasons further discussed in Section 2.

In more recent decades, the development of genomics and molecular phylogenetics raised the possibility of reconstructing a comprehensive natural history of life on Earth that includes microbial diversity and evolution. It is now understood that there is a universality of genetic inheritance that provides evolutionary continuity between microbial and complex multicellular life. Sequence-based phylogenetic reconstructions of the relationships between different groups of microorganisms also firmly establish the existence of microbial organismal lineages, analogous to those within more complex life (Woese 1987). Additionally, the phylogeny of individual gene families can now be reconstructed, tracing the history of microbial metabolic processes that are biogeochemically significant in planetary history. Molecular phylogenetics thereby not only reveals the record of microbial lineages and metabolisms preserved within genomes but also empowers other scientific tools for reconstructing a more comprehensive natural history of the Earth-life system. Phylogenetics also enables many additional powerful methodologies, including estimating divergence times of genes and lineages using molecular clocks, ancestral sequence reconstruction (ASR), and tracing the complex genomic histories of horizontal gene transfer (HGT).

With these tools, the broad outlines of microbial natural history have begun to emerge, but many questions remain. Has microbial diversity been shaped by the same selective processes as complex life, including mass extinctions and adaptive radiations? Have there been major successions within microbial ecological niches, with extant groups in roles previously held by extinct diversity? Does it make sense to consider such concepts as "Archean microbiota" or "Proterozoic microbiota," analogous to the major biotic transitions observed in the Phanerozoic? And how do these processes relate to the history of planetary change itself, including climate, atmosphere,

tectonics, and even impacts? Here, we provide a cursory framework that describes how interdisciplinary geobiological approaches can address these questions, propose six broadly applicable principles to aid in their investigation, and include key examples of investigations of microbial systems relating to major planetary/biogeochemical processes and events in Earth history. In doing so, we have necessarily limited our scope to microbial and molecular evolution operating within the context of a biosphere via processes observed today. Therefore, we have not extended this framework to questions of the origin of life, inclusive of prebiotic chemistry and the emergence of the earliest cells. Nevertheless, an improved understanding of microbial evolutionary history may also provide clues that can offer insight into the origin and early evolution of life on this planet.

2. RECORDS OF PAST MICROBIAL LIFE

Principle 1: Compared to complex life, microbial natural history is obscured by the near-absence of a diagnostic fossil record that preserves lineage-specific traits. A partial history of microbial lineages, their metabolisms, and their ecological roles is nevertheless preserved within two independent and complementary records: genomic and geological.

2.1. Evolutionary Guiding Principles

The understanding that evolutionary processes of speciation and extinction generate a branching tree of life that relates all extant groups is derived from anatomical and paleontological studies of complex organisms, beginning in the nineteenth century, and further supported by modern genomic approaches (for a review, see Telford et al. 2015). These processes constitute several types of selection acting upon variation within populations, so that some traits become lost while others are fixed—passed down to descendants. These accumulated changes eventually lead to speciation and, over even longer timescales, the large genotypic and phenotypic differences that are observed between major groups of organisms such as the different metazoan (animal) phyla (e.g., Arthropoda versus Chordata) or different kingdoms of eukaryotes (e.g., Fungi versus Metazoa). If we look backward from any arbitrary point of time, every species that has existed up to that point is part of a continuity between an ancestor and extant forms. This is because of extinction: The vast majority of species in Earth's history no longer exist, and with their loss, diversity becomes discrete, separated by wide valleys of evolutionary change with only the fossil record left to traverse them. While extinction is always happening at some level, rates of extinction vary between groups and environments, and across time. Occasionally, major planetary changes induce extinction rates so high that they dramatically impact the fossil record, showing a clear demarcation between the diversity of life before and after the event. These mass extinctions and the biotic successions that followed are a major component of the historicism of complex life on Earth (for reviews, see Hull 2015, Whiteside & Grice 2016). Such macroevolutionary processes are far too complex to be detailed here, but this cursory overview serves as a backdrop for understanding how applying a microbial perspective will lead to a more comprehensive natural history of life on Earth.

2.2. The Microbial Fossil Record

The traditional paleontological approach to life history and organismal evolution describes and compares fossil evidence to characterize and classify past life. However, this record is largely limited to the Phanerozoic and rapidly diminishes during the Neoproterozoic, with scant fossil evidence of single-celled eukaryotic and prokaryotic life before this time (Knoll 2015, Knoll et al. 2016). In the case of bacteria, this morphological evidence consists of both body fossils and trace fossils. Fossils enable the determination of two types of phylogenetic information: evolutionary relationships and timing. Evolutionary relationships may be ascertained from identification of

shared, derived characters in a fossil, whereas timing is established by either stratigraphic context or radiometric dating. Assigning taxonomy to fossils is often challenging, even for well-described types of complex life. While precise dates do exist for many instances of microbial fossil evidence, in most cases the lack of any diagnostic morphology prevents any taxonomic characterization, and even the biotic origin of the structure may be in question (Javaux 2019).

The assignment of taxonomy to particular fossils is complicated by common taphonomic processes, which erase fine details and delicate peripheral structures. Unfortunately, these same features tend also to be the most derived and therefore the most useful diagnostically, resulting in a bias whereby fossils appear more ancestral than they truly are (Sansom et al. 2010). This bias is especially problematic in metazoans but can generally be overcome by finding specimens with exceptional preservation. In the case of microorganisms, however, for most groups identifiable preservation is probably impossible, and in any case the preserved body fossils would be taxonomically ambiguous. Furthermore, the limited morphological information that may be present, such as overall cell shape, is often associated with many different groups of microorganisms that may not be closely related. Nevertheless, some microbial body fossils possess distinctive features that permit general taxonomic identifications, which have been substantially improved with the development of advanced microscopic imaging technologies (Demoulin et al. 2019, Pang et al. 2018).

Some microorganisms can undergo differentiation into specialized cell forms with diagnostic morphological traits with high preservation potential. For example, members of the cyanobacterial group including Nostocales and Stigonematales can form dormant akinete cells to endure harsh environmental conditions or heterocyst cells specialized for nitrogen fixation (Buikema & Haselkorn 1993). The oldest described heterocyst fossils are dated to 2.1 Ga, indicating that the total group inclusive of extant heterocyst-forming cyanobacteria must be at least this old (Tomitani et al. 2006). Fortunately, not all preserved microbial fossils are at the scale of a single cell; multicellular forms are also preserved, and these particular cell arrangements are occasionally diagnostic for specific taxa. For example, the tight helical coils of the filamentous cyanobacterial fossil *Obruchevella* from 1.56 Ga (Shi et al. 2017) have been used to constrain the appearance of this morphology in a clade including modern *Spirulina* and its relatives, while the pseudofilaments of endolithic *Eohyella* from 0.8 Ga (Knoll et al. 1986) share a specific branching pattern and morphology with the modern rock-dwelling *Chroococcidiopsis* (Fournier et al. 2021). Even older, the specific division and compartmentalization patterns of *Eoentophysalis* from 2.0 Ga pustular mats closely and uniquely resemble putative descendants from similar mats in modern hypersaline and peritidal environments (Golubic & Seong-Joo 1999). Interpretation of microfossil evidence continues to be a subject of debate; more details regarding the paleontology of cyanobacterial evolutionary history are described by Tomitani et al. (2006) and Demoulin et al. (2019).

Microbial communities also preserve at the macroscopic level, namely through the lithification of microbial mats and associated sedimentary structures. In fact, the oldest uncontested traces of life on Earth are stromatolites—laminated structures constructed through the precipitation and binding of sedimentary material with the growth of microbial mat layers (for a review, see Bosak et al. 2013). The biogenicity of some similar structures is debated, but the canonical examples—reaching back 3.45 billion years to Western Australia's Strelley Pool Formation (Allwood et al. 2009)—are regarded as evidence of cellular, phototrophic life accreting upward toward light. However, the taxonomy and metabolisms of the organisms that constructed these early stromatolites are not obvious. In the modern, cyanobacteria dominate (now uncommon) stromatolite-growing communities, but other groups could have formed earlier stromatolite communities—especially the oldest examples, which substantially predate any geochemical evidence of oxygenic photosynthesis or age estimates for cyanobacteria (Fournier et al. 2021). While many stromatolite

morphologies in the fossil record are not represented in the modern, the differences may have as much to do with environmental changes as taxonomic composition. As such, stromatolites remain crucial in establishing the early presence of bacterial life and possibly phototrophy, but at present, their history in deep time cannot be linked to any specific group of bacteria.

The prevalence of stromatolites and other microbialites in the rock record nevertheless reflects changes in the ecological dominance of prokaryotic versus eukaryotic organisms over time. Stromatolites lose their prominent position in life's sedimentary record toward the end of the Proterozoic, likely owing to predation by early metazoans or even Foraminifera (Bernhard et al. 2013). In the Phanerozoic, the microbial precipitation of carbonate has largely been eclipsed by the more tightly controlled biomineralization of metazoan skeletal reefs, but it remains a dynamic ecological balance: Microbial reefs return as the main component of the carbonate record multiple times, especially in the aftermath of mass extinctions devastating metazoan calcifiers (Wood 2011).

Aside from morphological fossils, microorganisms also leave behind chemical traces in the form of specific organic compounds synthesized during their lives. Organic molecules show widely varying stabilities dependent upon their chemical properties and preservational environment. DNA and polypeptides are generally short lived, as their phosphodiester and peptide bonds readily hydrolyze. Nevertheless, under the right conditions, they can persist on geological timescales, possibly for millions of years (Callaway 2022). In such cases, fossil DNA and peptide sequences of microorganisms can be related to extant lineages through phylogenetic reconstructions (for reviews, see Arning & Wilson 2020, Hendy 2021). However, the reduced preservability of these molecules and the difficulties in ruling out contamination from modern microbial material limit their utility in studying microbial natural history over planetary timescales (e.g., Duchêne et al. 2016, Hebsgaard et al. 2005, Hendy 2021, Saitta et al. 2019; for ancient bacterial DNA see Arriola et al. 2020, Fish et al. 2002).

For this reason, the molecular fossil record is almost entirely composed of other compounds, specifically, the chemically stable remnants of lipid membrane components that can potentially persist for billions of years. These biomarkers are often similar or identical to those in specific groups of modern living organisms, enabling reliable source assignment. Biomarkers are frequently used to establish the presence and activity of specific microbial groups and their metabolisms in deep time. They may vary in their taxonomic specificity due to shared lipid biosynthetic pathways across major groups of organisms. For example, fossil steranes are primarily attributed to eukaryotic sources. Ether-linked lipids, depending on their stereochemistry, may be attributed to archeal or bacterial sources, such as in isoprenoidal versus branched glycerol dialkyl glycerol tetraether lipids, respectively (Weijers et al. 2006). Some biomarkers show much greater taxonomic specificity and can be linked to metabolically distinct groups of microorganisms that more directly inform the reconstruction of past ecologies. Well-documented examples include crenarchaeol, which is diagnostic for members of the archaeal phylum Thaumarchaeota involved in marine ammonia oxidation (Damsté et al. 2002), and the aromatic carotenoids okenane and chlorobactane, which are diagnostic for photosynthetic purple sulfur bacteria (PSB) and green sulfur bacteria (GSB), respectively (Brocks & Schaeffer 2008, Summons & Powell 1986).

Major ecological and environmental changes in Earth history have been characterized through investigations of the lipid biomarker record (Naeher et al. 2022). Brocks et al.'s (2017) examination of the eukaryotic steroid record ranging from the Tonian to the Phanerozoic revealed the emergence of marked sterol diversity between the Sturtian and Marinoan glaciations, signifying the rise of marine planktonic algae. Similarly, the recovery of carotenoid biomarkers in the 1.64 Ga Barney Creek Formation (BCF) established the prominence of GSB and PSB in Proterozoic aquatic settings (Brocks et al. 2005), although recent studies have proposed a potential contribution from cyanobacterial sources for some of these compounds (Cui et al. 2020).

Interpretation of the biomarker record is often complicated by geologic processes that overwrite or otherwise alter evidence of ancient life. Few rocks preserve taxonomically and ecologically diagnostic traces of life free from contamination by extant microorganisms or drilling fluids. High temperatures can also chemically alter this record. While thermal alteration brought about by normal burial conditions structurally modifies biomarkers in a predictable manner, the temperatures associated with metamorphism crack organic molecules and destroy their biomarker potential (Peters et al. 2004). Furthermore, analyses of lipid biomarkers via traditional chromatographic-mass spectrometric techniques are destructive, making repeated analyses, replication, and reexaminations of past records increasingly difficult. Recent developments in nondestructive spectroscopic techniques are a major advance toward overcoming this limitation (Ferralis et al. 2016). Ultimately, although the microbial fossil record is, and always will be, incomplete, it is essential for a comprehensive understanding of microbial natural history, and its power is greatly extended when combined with genomic approaches, as described below.

2.3. The Genomic Record

The genomes of contemporary microorganisms are the product of over 3 billion years of evolution, driven in part by accumulated adaptations to planetary change. As such, they preserve important information about Earth's history, which we refer to as the genomic record. The genomic record is investigated using molecular phylogenetics—the study of genetic differences to infer evolutionary relationships—and is especially useful in cases where traditional paleontology lacks a sufficiently resolved fossil record to reconstruct a lineage history (Enk et al. 2009). However, and unlike the fossil record, molecular phylogenies are limited to the ancestor lineages of extant organisms with genomes that can be sampled.

Microbial genomes are densely packed with discrete regions (genes) that encode functional protein or RNA sequences. The first step in extracting evolutionary signals from genomes is to establish homology (shared ancestry) between specific genes across a broad taxonomic range via multiple sequence alignment (Phillips et al. 2000). Nucleotide or amino acid identities within aligned sites preserve evidence of the changes (substitutions) that have generated the observed diversity within homologous sequences. These alignments are analogous to morphological character matrices, which group organisms (including fossils) based on the presence of specific derived traits; as such, site substitutions are the basic building blocks of molecular phylogenetic inference (Felsenstein 2004, Holder & Lewis 2003). The final product of this inference, a phylogenetic tree, is a series of bifurcating branches tracing the likely divergences that led to the observed diversity of sequences across organisms. The essential components of each tree are crowns, grouping taxa that share a common ancestor (inclusive of all extinct relatives not sampled in the tree), and stems, the branches leading from crown groups to their divergences from other crown groups (inclusive of all extinct relatives more closely related to one crown group than another). Several methods have been developed for establishing evolutionary relationships from this information (for a review, see Holder & Lewis 2003). Specific histories recovered within phylogenetic trees are often dependent upon sequence sampling and the particular methods used for alignment and tree construction. As such, the impacts of these specific methods on the resulting phylogenies should be carefully tested.

In order to generate trees relating different lineages (species trees) in a standardized manner, sampled sequences must exhibit wide taxonomic range and evolve at sufficient rates to provide informative sites for reconstruction (Zuckerkandl 1962, Zuckerkandl & Pauling 1965). Based on these considerations, the 16S ribosomal RNA (rRNA) molecule of the small (30S) ribosomal subunit is the standard for resolving deep evolutionary relationships between microbial lineages

(Woese & Fox 1977). Translated protein sequences are also frequently used for species tree inference, using concatenations of highly conserved, single-copy genes that provide many more phylogenetically informative sites; typically, this set largely consists of ribosomal subunits and associated elongation factor proteins (Baldauf et al. 1996, Gutell et al. 1985). While a detailed discussion of the reconstruction of ancient evolutionary relationships through phylogenetic methods is outside of the scope of this review, others have provided comprehensive overviews on the topic (e.g., Gribaldo & Philippe 2002, Philippe & Laurent 1998).

Phylogenies of individual gene families can also provide information about the evolution of phenotypes. Within a given gene family, the catalytic active site of the encoded protein is usually the most conserved region, and in many cases it can be assigned to a specific enzymatic activity (Ribeiro et al. 2018). Biogeochemically relevant microbial activity typically involves the coordination of multiple enzymes that catalyze a series of reactions in a metabolic pathway. These pathways are often conserved across taxa, and therefore they enable microbiologists to determine the functional potential of microorganisms purely from their genome content. However, it is frequently the case that different enzymatic reactions are performed by similar enzymes, which may be indistinguishable by primary sequence data alone (e.g., as is the case for cytochromes, described in Paquete et al. 2019). Experimental validation is, therefore, needed to underpin many of these functional annotations.

Another tool that attempts to leverage the genomic record for a more direct investigation of past microbial life is ASR. In the course of generating any maximum-likelihood tree, the probabilities of ancestral states for each site in a sequence are calculated; therefore, the most likely sequence at any ancestor node in a gene tree can also be inferred. While ancestral states have inherent uncertainties due to the stochastic nature of substitutions in evolutionary history and the biases that influence phylogenetic tree reconstruction (Selberg et al. 2021), best-guess ancestor candidates generated through ASR are still valuable for investigating how microbial protein functions and metabolisms have changed across deep time. In silico analyses of ASR can constrain the geometry and composition of enzyme binding sites, which can inform substrate selectivity (Schwartz et al. 2022) or physical properties such as the spectral absorbance range of ancient light-harvesting proteins (Sephus et al. 2022). Ancestral proteins can also be synthesized/expressed in vitro or in vivo, so that their properties can be experimentally determined by biochemical and genetic assays (Garcia & Kaçar 2019), such as in reconstructing ancient ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) enzymes to study Proterozoic primary productivity (Kędzior et al. 2022).

As the majority of life history consists of extinct groups known only by their fossil traces, the genomic record can never replace the fossil record. Additionally, these extinct forms often preserve morphologies that permit the ordering of derived traits, allowing for the macroevolutionary process to be reconstructed in much greater detail than is possible working from extant lineages alone. For example, crocodilians are the sister group to birds, but this provides very little information about how flight evolved. Only the fossil record preserves this information, through a succession of nonavian archosaur fossils that document the accumulation of many derived traits over ~200 million years (Brocklehurst et al. 2020, Xu et al. 2014). Nevertheless, as described in Section 3, integrating these records enhances the power of both.

2.4. Gene Flow in Microorganisms: Horizontal Gene Transfer

Along with highly conserved sequences such as the 16S rRNA gene, microbial genomes largely consist of more dynamic gene content that varies greatly between lineages (Lapierre & Gogarten 2009). Moreover, individual gene families often produce phylogenies that disagree with inferred

species trees. This directly implies that microorganisms can obtain genetic information from a source other than their parent cell, and at least occasionally, this acquired material becomes fixed in a population and passed onto descendant lineages. This mechanism, known as HGT, is a major evolutionary process within genomes (for a review, see Arnold et al. 2022). Together with gene losses and duplication events, HGT generates discordances between species tree evolutionary relationships and individual gene histories, called reticulations (e.g., Baldauf et al. 1996, Deckert et al. 1998, Wetmur et al. 1994). In particular, genes encoding metabolic pathways for energy metabolism and nutrient utilization are among the most frequently transferred within microbial genomes and play a key role in microbial niche adaptation (Lapierre & Gogarten 2009). HGT of metabolic genes therefore provides one of the most important biological links to the geochemical record, as described in Section 3.

HGT may be facilitated by viral intermediaries, conjugation, or even direct uptake of detrital DNA from the environment—however, for natural history purposes, the particular mechanism is less important than the observed phylogenetic signals generated (Soucy et al. 2015). Importantly, any given single HGT event within a gene tree introduces only a single reticulate branch with respect to a hypothetical species tree; subsequent diversifications in the recipient HGT lineage may continue to trace the species tree. Therefore, even gene families experiencing extensive HGT often retain species tree evolutionary signals that can be used to anchor the history of a gene (**Figure 1**). HGT's decoupling of gene inheritance from vertical ancestry also introduces a distinctly paleontological signal into gene trees: HGT donors are never the direct ancestors of modern taxa, in the same way that fossils of so-called ancestors are more accurately described as relatives of true ancestral lineages (Szöllősi et al. 2013). On geological timescales, genes having undergone HGT therefore uniquely preserve genetic information from extinct microbial lineages, similar to how the paleontological record uniquely preserves a history of extinct groups with no modern descendants. In these ways, HGTs carry important, unique information not otherwise preserved in species tree phylogenies (Doolittle 1999). As described in the following sections, when integrated with the geochemical and fossil records, phylogenomic inferences can enable reconstructions of ecological/environmental processes across a comprehensive microbial natural history.

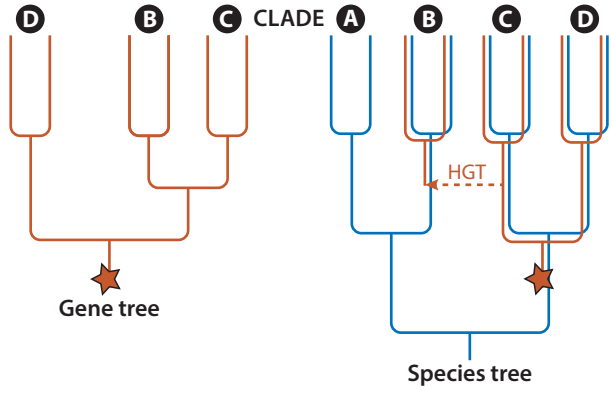

Figure 1

Inferring horizontal gene transfer (HGT) from phylogenies. Four groups of organisms constituting a clade (A, B, C, D) are represented in a hypothetical rooted gene tree (*red*) and species tree (*blue*). The topology of the two trees is incongruent and can be reconciled by inferring an HGT event from the ancestor lineage of C to the ancestor lineage of B, with the gene originating in the common ancestor lineage of C and D (*red star*). In real phylogenies these inferences are often far more complex and include multiple transfers, losses, and duplications.

2.5. Geochemical Record of Microbial Life

Chemical traces of past life are a powerful complement to paleontological and phylogenomic lines of evidence, preserving important information about past environments and the metabolic processes of life inhabiting them. While the biomarker record described in Section 2.2 is also chemical, it is akin to the body fossil record, in that the primary information is within the uniquely biogenic structures preserved. In contrast, the geochemical record of life is preserved in both organic and inorganic materials that have interacted with living systems. While this record underreports past environmental complexity and is strongly impacted by preservational bias, globally distributed, ecologically important processes are far more likely to be resistant to this bias and leave detectable geochemical traces.

Body fossils often retain stable isotopic fractionations of elements that preserve information about metabolism and ecology. For example, stable nitrogen isotopes ($\delta^{15}N$) can be used as natural tracers of trophic levels, while stable carbon isotopes ($\delta^{13}C$) can trace the carbon fixation pathway involved in primary productivity (for reviews, see Peterson & Fry 1987, Post 2002). The fixation of biomass from inorganic sources selectively enriches for isotopically light elements in differing and well-characterized ranges, dependent on the specific enzymatic pathway used (Peterson & Fry 1987, Post 2002). The most extensive biological impact on carbon isotope fractionations can be traced to the most abundant enzyme on Earth, RuBisCO. RuBisCO is a key enzyme in the Calvin-Benson-Bassham (CBB) cycle and is used by all plants, cyanobacteria, and many other microorganisms for carbon fixation (Andersson & Backlund 2008). The CBB cycle imparts a characteristic isotopic offset between the $\delta^{13}C$ of fixed biomass and $\delta^{13}C$ of the inorganic carbon source, ranging between ~10 and 35‰ $\Delta\delta^{13}C$ (Garcia et al. 2021). Alternative carbon fixation pathways used by other microbial groups imprint variable but diagnostic fractionation ranges, such as the reductive tricarboxylic acid (rTCA) cycle, which imprints a less pronounced fractionation from ~2 to 23‰ $\Delta\delta^{13}C$ (for a review on carbon isotope fractionation across metabolisms, see Garcia et al. 2021). Subsequent biosynthetic pathways can also impart isotope fractionations preservable in biomarkers, which can be detected by compound-specific isotope analysis and used to infer a more detailed metabolic evolutionary history of life (Hayes et al. 1990).

The aforementioned geochemical records track assimilatory processes, in that they consist of material that organisms have incorporated from the environment into their bodies; however, far greater biogeochemical fluxes are generated through dissimilatory processes related to the energy metabolisms of organisms that directly impact the chemistry of their environments. This is particularly true for microorganisms, which are less likely to leave body fossils and, instead, more likely to leave a metabolically driven isotopic footprint in the sedimentary record. For example, stable isotope fractionations of sulfur preserved in the geological record indicate dissimilatory sulfur metabolisms operating in past environments and have been extensively studied to reconstruct past biological redox processes (for reviews, see Fike et al. 2015, Johnston 2011). The isotopic record of sulfate and sulfide minerals such as pyrite serves as an indicator of the oxidation state of the ocean over time, while specific isotopic signatures remain diagnostic for the presence of sulfur metabolic pathways (e.g., Canfield & Teske 1996, Johnston et al. 2005). Microbial sulfate reduction preferentially consumes sulfate composed of isotopically light ^{32}S, resulting in fractionation between sulfide products and sulfate reactants (Bradley et al. 2016). The isotopic evidence of sulfate reduction in Earth history is closely dependent upon the availability of sulfate itself, as fractionations can take place only when a substrate is not limiting. If sulfate concentrations are sufficiently low, sulfate reducers in the environment may consume all available sulfate without preference for the lighter isotope, leaving behind no record at all (Peterson & Fry 1987). This would result in a geochemical false negative, a concept further discussed in Section 4.1.

The increasing availability of sulfate therefore marks another major transition in microbial natural history. Without the contribution of significant atmospheric oxygen to mineral weathering processes, the primary inventory of sulfate available for microbial reduction during the Archean would have been photochemically oxidized volcanic sulfur gas, leaving the biosphere sulfur limited (Farquhar et al. 2000, Walker & Brimblecombe 1985). Increased continental weathering in the Proterozoic from the rise of atmospheric oxygen likely facilitated an increased flux of sulfate into marine systems, yielding a stronger signal of biological sulfate reduction fractionation within sedimentary sulfides (Walker & Brimblecombe 1985). As with other geochemical records, the interpretation of these isotopic signatures requires the consideration of environmental contributions to sulfur fractionation, as several abiotic sources also fractionate sulfide, including evaporitic mineral gypsum or the thermochemical sulfate reduction into pyrite (Machel et al. 1995).

2.6. Microbial Diversity and Ecology

Observations of extant microbial diversity and the roles of microbes in modern ecosystems provide context for the interpretation of the fossil and genomic record. These observations are traditionally made through cultivation and experimentation; however, since the turn of the twenty-first century, the field of microbiology has undergone a genomics revolution, revealing over 500,000 new microbial genomes in the past decade (**Figure 2**). The continued expansion of genetic databases through environmental sequencing demonstrates that the majority of microorganisms have not been cultured or experimentally observed (Rappé & Giovannoni 2003) and has identified several new groups of organisms that help inform deep evolutionary relationships (**Figure 2**, points A and G–K). Environmental sequencing has also revealed new metabolic pathways that provide a deeper understanding of microbial ecosystems across a wide set of environmental conditions, including those that may have prevailed on the early Earth (**Figure 2**, points B–F and L).

Today, microorganisms inhabit a wide temperature, salinity, pressure, and pH range where their existence, function, and abundances are largely determined by the amount of energy and essential elements supplied by their habitat (for a review, see Takai 2019). As such, environmental gradients or large changes in the energy or nutrients supplied by an environment can result in changes in microbial community composition and function; however, community function and bulk biochemical flux rates can, alternatively, be insensitive to environmentally driven taxonomic changes due to the decoupling of taxonomic composition and metabolic function across the microbial tree of life (for a review, see Louca et al. 2018). Rare members of microbial communities and members of metabolic syntrophies are also proving to play an outsized role in driving biogeochemical cycles and ecosystem function in modern environments (Jousset et al. 2017, Lau et al. 2016). These observations indicate that, although major redox transitions have taken place throughout the course of Earth history, there are likely to be multiple factors that control the emergence, prevalence and diversity of new metabolisms and ecologies that arise during these periods. Section 3 discusses how the integration of genomic and geological records with modern ecology is informing a microbial natural history.

3. INTEGRATING GENOMIC AND GEOLOGICAL RECORDS

3.1. Molecular Clocks

Principle 2: Dated phylogenies can provide younger bounds on the origin of microbial lineages and their metabolisms in Earth history, as well as time estimates for their subsequent ecological dispersals and diversifications.

One of the first ways that phylogenies and geological data were rigorously integrated was through the use of fossils to constrain the relative and absolute divergence times between

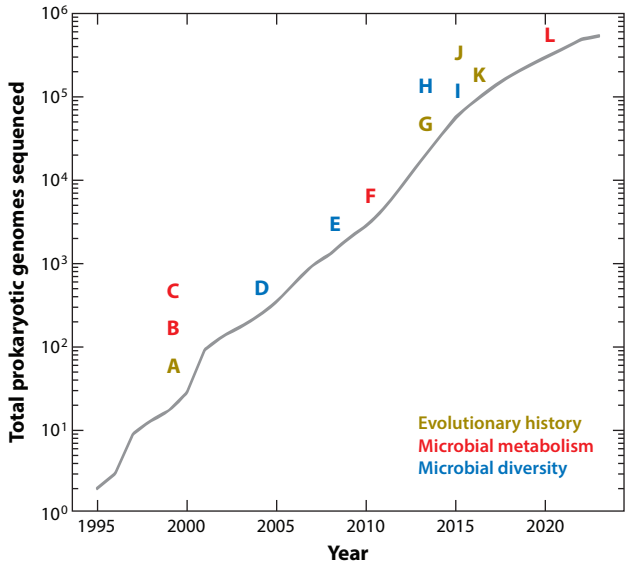

Figure 2

The rapid growth of genomics databases. Advances in sequencing and computing technologies have resulted in the rapid growth of genome databases such as the National Center for Biotechnology Information's prokaryotic genome database (Sayers et al. 2021). A selection of genome-informed discoveries that have improved our interpretation of microbial natural history are indicated on this timeline, color coded by their primary significance. A. First report of bacteria-archaea horizontal gene transfer (HGT) (Nelson et al. 1999). B. Lipid- and DNA-based identification of archaea performing anaerobic methane oxidation and controlling atmospheric methane efflux from the ocean (Hinrichs et al. 1999). C. The identification and genomic characterization of a missing lithotroph (Broda 1977) anaerobically oxidizing ammonia to dinitrogen gas (anammox) (Strous et al. 1999). D. Revelation of over 1 million new genes from Sargasso Sea metagenomic sequencing (Venter et al. 2004). E. Identification of a microbial community dominated (>99.9%) by sulfate-reducing bacteria in the South African subsurface (Chivian et al. 2008). F. Discovery of a nonphotosynthetic archaeon capable of generating intracellular oxygen via the reduction of nitrate to oxidize methane (Ettwig et al. 2010). G. Identification of Melainabacteria (Vampirovibrionia), the nonphotosynthetic sister group of Cyanobacteria, in human gut and groundwater metagenomes (Di Rienzi et al. 2013). H. An archaeal superphylum, DPANN, that is unified by small cell and genome size revealed by single-cell sequencing of microbial dark matter from a variety of environments (Rinke et al. 2013). I. Identification of a candidate phyla radiation of bacteria comprising greater than 15% of bacterial diversity and exhibiting small genome size and reduced metabolic capacity (Brown et al. 2015). J. Lokiarchaeota, the archaeal bridge to eukaryotes, identified in marine sediments (Spang et al. 2015). K. First attempt to construct a tree of life including over 1,000 newly discovered, uncultivated organisms (Hug et al. 2016). L. The first chemolithoautotrophic organism performing manganese oxidation identified and sequenced, completing a biogeochemical cycle for manganese (Yu & Leadbetter 2020).

different groups of organisms based on the first appearances of taxa containing diagnostic characters (traits). Molecular phylogenetics greatly expanded the power of these approaches, as it provides a large set of changes between character states (substitutions between different amino acids or nucleotides) that can be quantitatively modeled based on observed relative frequencies of site-specific changes. The rates of these changes along branches in a tree can therefore be used as a molecular clock that estimates the relative amount of time each branch in the tree traverses (Dos Reis et al. 2016). As rates vary and many processes can impact site-specific substitutions, the clock is said to be relaxed, producing distributions of expected relative divergence times (Bromham et al. 2018). If groups have diagnostic fossils, their absolute ages can be used as

temporal constraints on divergences in the tree, also known as calibrations. In its simplest usage, a dated fossil possessing shared derived traits acquired in the ancestor lineage of a group must be younger than the split in the tree between the ancestor and its closest relatives; in this way, the fossil provides a younger bound on this node in the phylogenetic tree. In practice, applying fossil calibrations is often far more complex, especially for groups of organisms with a rich fossil record that can also permit older bounds to be estimated and applied (Marshall 2019).

One of the important strengths of molecular clock analysis is that it permits estimating the age of groups that lack a fossil record, so long as they can be placed in a tree with other lineages that provide fossil calibrations. Molecular clocks can therefore be used to infer divergence times of microbial phylogenies if reliable fossil calibrations can be found. While microbial body fossils are extremely rare, biomarkers can also be used as calibrations for some microbial groups, as described in Section 2.2.

Even with informative calibrations, many sources of uncertainty impact molecular clock analyses. These uncertainties generally grow as one moves further into the past, and so the ages of most microbial groups remain very poorly constrained. While these approaches continue to improve, current work shows how even age estimates with large uncertainties can discriminate between specific biogeochemical hypotheses. For example, many studies have attempted to estimate the origin and diversification time of Cyanobacteria, which considerably expanded the habitability of the biosphere by producing oxygen and increasing primary productivity (e.g., Betts et al. 2018, Cardona et al. 2019, Fournier et al. 2021, Magnabosco et al. 2018, Sánchez-Baracaldo & Cardona 2020, Schirrmeister et al. 2015). The onset of oxygenic photosynthesis is constrained by an older bound on the divergence time of stem Cyanobacteria from other lineages and by a younger bound constrained by the age of crown Cyanobacteria. Varying age estimates have been obtained for stem and crown Cyanobacteria using different sets of calibrations derived from cyanobacterial microfossils, plastid-containing eukaryote fossils, and geochemical evidence of oxygenation. Furthermore, a wide variety of taxon samplings, outgroups, and tree and molecular clock model parameters have been shown to also impact these divergence time estimates. Different combinations of these priors generate mean crown age estimates for Cyanobacteria from as early as the Mesoarchean (~3.2 Ga) to as young as the Mesoproterozoic (1.5–1.0 Ga), with general agreement that stem Cyanobacteria diverged sometime during the Archean (Betts et al. 2018, Schirrmeister et al. 2013). In general, inclusion of cyanobacterial microfossil evidence as younger-bound calibrations tends to favor older Archean estimates (Fournier et al. 2021; Schirrmeister et al. 2015, 2016). As such, even with these uncertainties, one can infer that the interpreted diagnostic cyanobacterial fossil record is most consistent with an Archean origin of oxygenic photosynthesis before the detected atmospheric accumulation of oxygen during the Great Oxygenation Event (GOE). While many of these uncertainties are inherent in the modeling process, variances in ages arising from different taxonomic and fossil data sets should converge over time, as genomic and paleontological evidence accumulates.

3.2. Interpreting Metabolic Gene Histories

Principle 3: Microbial metabolic gene histories provide the most direct links between the genomic and geological records, including organic biomarkers as well as isotopic fractionations of biogeochemically processed elements. Integrating these records is necessary for accurately reconstructing microbial planetary history and mitigating sources of bias such as geological preservation and lineage extinction. These gene histories likely track major changes in planetary systems and ecologies more closely than the histories of microbial lineages themselves, as they are more likely to persist through these events.

Phylogenies of metabolic genes having undergone HGT are one of the most important parts of the genomic record, as these trees potentially traverse both extinct and extant lineages. In such

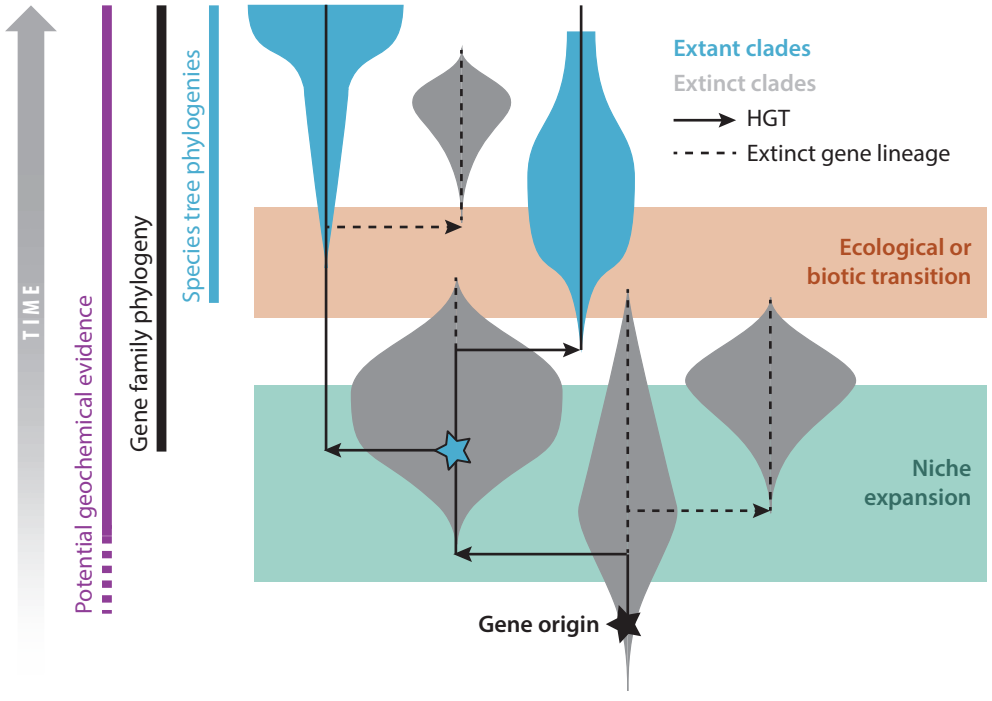

Figure 3

Records of microbial natural history as shaped by molecular, organismal, and ecological evolutionary processes. The history of a hypothetical gene encoding a biogeochemically significant metabolic process is depicted over time (y-axis) as a gene tree (*black lines*) embedded in the histories of extinct and extant species tree lineages (*gray* and *blue violin plots*, respectively, with thickness indicating diversity/ecological prevalence). Following its origin within a specific microbial lineage, the gene spreads to other lineages via horizontal gene transfer (HGT) as the favorable niche expands (*teal*). Subsequently, a major ecological or biotic transition (*orange*) drives extinction of many of these lineages, to be replaced by other groups. The recoverable genomic record, therefore, traces back earlier than associated species tree histories (*blue star*) but not necessarily as far back as the geochemical record (*purple line*).

cases, any associated products and processes can be inferred to have existed long before the extant groups possessing these traits arose (**Figure 3**). This is especially important for ecologically/biogeochemically significant processes that impact the geochemical record. Genes involved in dissimilatory sulfate reduction, the nitrogen cycle, the carbon cycle, and photosynthetic machinery have all been extensively studied in detailed phylogenetic analyses, each showing evidence of extensive HGT (Cardona et al. 2019, Müller et al. 2015, Wagner et al. 1998).

A relatively simple case is provided by the dissimilatory sulfite reductase DsrAB, an enzyme that is frequently used as a marker of dissimilatory sulfur metabolism. Its phylogeny shows that this metabolism substantially predates the emergence of the known extant bacterial and archaeal sulfur reducers (Müller et al. 2015). The phylogenetic history of DsrAB therefore provides important context for interpreting the sulfur geochemical record. In particular, geochemical evidence for microbial sulfate reduction via isotopic fractionation should not be taken at face value as evidence for the presence of known groups of sulfate-reducing microorganisms. This caveat informs both phylogenomic and geochemical investigations, indicating that (*a*) dated geological evidence of sulfate reduction should not be used as a calibration on the age of extant groups of sulfate-reducing microorganisms and (*b*) geological evidence of sulfate reduction older than the inferred ages of extant clades of sulfate reducers is not necessarily abiotic in origin. This is important, given the richness of the isotopic fractionation record within S-containing sedimentary rocks. The principles applied

in this example are broadly applicable to the integration of the genomic and geochemical records, and they indicate the caution required with interpreting either in isolation.

3.3. Extinction, Microbial Diversity, and Major Transitions

Principle 4: Significant microbial biotic transitions occur during times of major environmental changes, marked by the appearance of new metabolisms and ecologies together with profound shifts in taxonomic distributions. These transitions show that microorganisms are not just the architects of planetary change but also are subject to its selective forces, and their macroevolutionary turnover is sometimes gradual and sometimes abrupt.

As genomic and metagenomic sequencing has continued to expand known microbial diversity, some conspicuous gaps in microbial phylogenetic trees have persisted, akin to ghost lineages in the fossil record (Cavin & Forey 2007). These gaps are highly suggestive of extinction shaping patterns of microbial evolution, with major lineages of microorganisms not surviving to the present, leaving long empty branches between extant groups. This is the first hint that microbial evolution on long timescales may follow patterns similar to traditional organismal evolution, as the following examples illustrate and as is generalized in **Figure 3**. One of the most significant of these gaps involves cyanobacteria. Extant cyanobacteria all share several derived traits for oxygenic photosynthesis, including chlorophyll, the oxygen evolving complex, and the coupling of two photosystem reaction centers (Sánchez-Baracaldo & Cardona 2020). These must have been acquired in some particular order along the Cyanobacteria stem lineage, with many intermediate forms possessing a subset of traits. For example, it seems essential that at some point, stem Cyanobacteria used only one photosystem and performed anoxygenic photosynthesis. One would therefore predict that continued sequencing efforts would eventually discover an anoxygenic phototrophic sister group to Cyanobacteria that preserves this ancestral state. However, to date, nothing like this has been found, and the branch leading to crown Cyanobacteria remains long. The best explanation for this discontinuity of physiological forms is extinction of diversity in stem Cyanobacteria due to some environmental challenge, perhaps the rise of oxygen itself. A good analogy is the evolution of bird flight, as has been previously alluded to; without a fossil record, the ordering and context of the numerous evolutionary adaptations leading to crown birds are bewildering. Their closest extant relatives—crocodilians—possess none of these adaptations. However, the rich fossil record of nonavian dinosaurs and stem birds permits the ordering of traits related to the evolution of flight, including those that preceded it (e.g., a wishbone) and those that came afterward (e.g., a pygostyle tail) (Brocklehurst et al. 2020, Rashid et al. 2014, Xu et al. 2014). Without such a fossil record providing direct evidence, our ability to reconstruct the evolutionary transitions leading to complex traits in microorganisms is inherently limited to logical inference, an admittedly poor substitute.

Another observation suggesting a cryptic history of Archean microbial phototrophs is the surprisingly recent ancestry of extant bacterial anoxygenic phototrophs such as GSB and PSB, both of which are inferred to have originated independently during the Proterozoic (Paoletti & Fournier 2022, Wang & Luo 2023). Many other groups of anoxygenic phototrophs show even shallower taxonomies (Woese et al. 1984), suggesting recent acquisitions of photosystem genes, e.g., in purple nonsulfur bacteria. Moreover, the deep phylogeny of photosystem reaction center types shows that they share a common ancestor that predates the crown diversification of Cyanobacteria (Cardona 2015). This further implies that there was a major turnover of phototrophic microbial lineages between the Archean and Proterozoic Eons, resulting in the eventual extinction of the "Archean flora." Specifically, the increased surface oxygenation during the GOE may have caused one of the largest mass extinctions in Earth history, although, to date, no direct evidence to support such a cataclysm has been provided. Nevertheless, the genes for phototrophy persisted, eventually

becoming present within the ancestors of the major extant microbial phototroph lineages. In fact, complete sets of photosystem genes are carried by marine phages today, suggesting that this persistence via HGT may be a longstanding feature of the ocean ecosystem (Sharon et al. 2009) and an important feature of microbial evolution in general (**Figure 3**). As marine and freshwater systems have continued to oxygenate since the Proterozoic, surviving anoxygenic phototrophs (there may have been other clades that were not so lucky) have likely experienced substantial niche loss and are relegated to more limited anoxic aquatic microenvironments where abundant electron donors still remain. This is similar to the natural history of many groups of complex organisms, which, while not being lost in mass extinctions, often experience shrinking niches accompanied by a waning of diversity [e.g., amphibians in the Mesozoic (Ruta & Benton 2008)].

Another feature of mass extinctions is recovery, often via diversification of different groups of organisms replacing those that went extinct, filling vacant niches. Cyanobacterial evolution provides evidence for this process occurring in the microbial world. Today, a large portion of primary production in marine systems is performed by the Synechococcus-Prochlorococcus (SynPro) clade of Cyanobacteria, and proposed cyanobacterial biomarkers have been found in rocks of deep marine origin predating the Cryogenian (Brocks et al. 2017). However, molecular clock studies show that SynPro is relatively young, diversifying shortly after the Cryogenian (720–635 Ma), during the latest Neoproterozoic or early Paleozoic (Sánchez-Baracaldo 2015). The timing of this diversification suggests that there was a major turnover and/or extinction of marine microbial populations associated with the Cryogenian glaciations, likely due to a long-lasting, large-scale perturbation of the global pelagic environment. Such environmental perturbations would have also impacted complex life and, beginning in the Ediacaran (635–541 Ma), the fossil record begins to show a wealth of apparently metazoan life forms. The appearance of complex animal life radically changed microbial ecology, as animal behaviors, bodies, and tissues greatly increased the richness of environmental conditions encountered by microorganisms. For example, the Cambrian substrate revolution reorganized microbial marine sediment communities, altering redox conditions and greatly increasing the complexity of these systems (Bottjer 2010) and, in the water column, the cores of sinking fecal pellets and other large particular aggregates can even accommodate reductive microbial metabolisms, a useful source of ammonium and sulfide in otherwise oxygenated waters (Bianchi et al. 2018).

The evolution of multicellular plants and animals also prompted an abundance of coevolutionary innovations. These include innumerable instances of symbioses, such as cyanobacterial endosymbionts in various corals and sponges that provide both oxygen and organic carbon to their hosts (Usher et al. 2007), and chemoautotrophic bacteria that permit animals to survive in deep hydrothermal vent systems (Stewart et al. 2005). The establishment of gut microbiomes is another major type of symbiotic relationship that has been the subject of much recent attention. In every sense, the animal gut is a microbial biome, with a totally unique set of environmental constraints not found elsewhere (Shapira 2016), and in turn, animal hosts greatly benefit from having their own personal bioreactors, which provide them with microbial metabolic functionality.

As multicellular life continued to diversify and expand, it began to effect changes in the Earth system itself, presenting a number of opportunities for microbial innovation. Chief among these revolutions is land colonization by animals and plants. As terrestrial primary production increased, so did the availability of terrestrial carbon and fixed nitrogen, kickstarting soil ecosystems, which today accommodate a large fraction of known microbial diversity (Kenrick et al. 2012). This terrestrialization also provided a large flux of nutrients to marine microbial systems through runoff and increased weathering (Retallack 2001, Vandevenne et al. 2013).

Standing as a tempting geological scaffold for those interested in microbial evolution, the history of animal evolution is beautifully preserved in the fossil record and provides a remarkable

progression of increasing ecological complexity and continual niche invasion. If specific microbial divergence times can be tied to well-constrained events in animal history, those constraints can be propagated to the microbial tree of life. For instance, certain recalcitrant animal-specific molecules, such as collagen, chitin, and keratin, may have posed unique genetic challenges to heterotrophic microbes, requiring the development (or acquisition via HGT) of new enzymes, which necessarily must be younger than their substrate and can therefore be used as absolute constraints in molecular clock analyses (Gruen et al. 2019).

4. CHALLENGES IN RECONSTRUCTING MICROBIAL NATURAL HISTORY: FALSE NEGATIVES AND FALSE POSITIVES

4.1. Geological False Negatives and False Positives

Principle 5: Analogous to the paleontological record of more complex life, the evolutionary origin of any microbial process likely substantially predates its first evidence in the geological record, making first appearances of metabolisms and evolutionary innovations difficult to infer. This is the false negative in the geological record.

Conversely, geological evidence of a given group or metabolism is sometimes spurious: A fossil can be assigned to a clade in error, the group responsible for a biomarker misidentified, or an altogether abiotic geochemical signature mistaken for a biogenic one. This is the false positive in the geological record.

In modern paleontology, molecular clocks often produce divergence time estimates that substantially predate the first appearance of fossil evidence for a particular group (first appearance datum). For example, the divergences of major bilaterian phyla are estimated to have occurred during the Ediacaran or even earlier, long before the observed Cambrian explosion in the fossil record (Dos Reis et al. 2015). While such discrepancies can be points of controversy, this is not evidence of disagreement between the genomic and geological records; rather, they are preserving evidence of two different but important defined events in a group's history (Marshall 2019). True disagreement arises only when a fossil predates the inferred divergence time of its clade, a much rarer problem that can be caused by either misidentifying a fossil or errors in making phylogenetic trees and molecular clocks.

Similarly, the first geochemical evidence of a microbial metabolism will necessarily postdate its genetic origin. In microbial natural history, this phenomenon is most clearly exemplified by the GOE, as it preserves evidence of a major biogeochemical transition that is a direct consequence of a specific evolutionary innovation (oxygenic photosynthesis) that must have occurred beforehand. Therefore, the GOE can serve as only a younger bound on the origin of oxygenic phototrophy in stem Cyanobacteria (as described in Section 3.3). How much older could oxygenic photosynthesis be? While different published molecular clock studies permit widely varying ranges for the possible duration of this pre-GOE interval (Betts et al. 2018, Fournier et al. 2021), the associated inferred false negatives can be explained by a variety of biological and geological hypotheses. Biological factors that may have prevented oxygen accumulation in the Archean include limited ecological prevalence of early cyanobacteria and/or the existence of aerobic microorganisms that locally consumed any produced oxygen; geological factors may also have played a role, such as reducing environments providing large sinks for any oxygen produced [e.g., ferrous iron in the oceans (Walker & Brimblecombe 1985)], or a lack of efficient burial of organic carbon in sediments (Bjerrum & Canfield 2004). Reconciling these inferred false negative intervals with biogeochemically consistent narratives is an important means by which the integrative microbial natural history record is constructed.

Geological false positives also pose difficulties and can arise from the fossil record, suggesting the presence of a group of organisms before their actual appearance, as well as the geochemical

record, suggesting a biological process is in operation before its emergence. Commonly, this type of error takes the form of identifying an abiogenic phenomenon as a biological one, especially in the challenge of identifying extremely early evidence of life. For instance, the purported microfossils from the 3.46 Ga Apex Chert, once seen as the oldest fossils of cellular organisms, are now thought to be silicate grains shaped by hydrothermal alteration (Brasier et al. 2015), and the 3.8 Ga isotopically light graphite inclusions in the Isua supracrustal belt have been reinterpreted as originating from crustal processes rather than biological carbon fixation (Van Zuilen et al. 2002).

4.2. Genomic False Negatives and False Positives

Principle 6: The persistence of specific metabolisms within microbial lineages is highly variable, as revealed by both deep time phylogeny and environmental metagenomic diversity. Currently observed biogeochemical processes may have been performed by different clades of microbes in the distant past, by either extinct lineages or extant lineages that subsequently acquired different metabolisms. This is the false negative in the genomic record.

Conversely, related groups of organisms may have acquired the genetic machinery for a shared trait independently, and therefore it is erroneous to infer that this trait arose in their common ancestor lineage and predates the age of the crown group. This is the false positive in the genomic record.

Compared to complex life, microbial lineages are generally only weakly associated with specific derived traits. Without the complex developmental dependencies underpinning most morphological traits in metazoans and other multicellular organisms, microbes experience much lower barriers for gene acquisition and loss (further described in Sections 2.4 and 2.6). As metabolic traits are far more evolutionarily resilient than individual lineages, caution must be exercised when reconstructing past microbial ecologies so as to not confuse a metabolism with a taxonomic group. Thus follows Ford Doolittle's adage, "It's the song, not the singer" (Doolittle & Booth 2017).

This should not be unfamiliar to paleontologists, the difference being that for complex life, convergence, rather than HGT, is the primary mechanism through which new singers are enlisted. Through convergence, distantly related groups independently evolve analogous traits and often compete with or even replace one another. Where microbes acquire foreign metabolic machinery, metazoans converge upon niches through adaptations that are frequently morphological/physiological rather than metabolic. For instance, the calcareous reef-builder niche has experienced remarkable turnover in the past 550 Ma, going through a variety of sponges, corals, algae, and mollusks (Wood 2011).

In the microbial realm, the photoferrotrophy niche seems to be an example of a similar story: Reduced iron was far more abundant in the photic zone of early Earth than it is today, but only a few groups of microorganisms within a shallow taxonomic distribution of GSB are known to perform photoferrotrophy (Tsuji et al. 2020). Additionally, it has even been hypothesized that photoferrotrophy may have driven the primary productivity of early Earth (Thompson et al. 2019). This apparent paradox makes photoferrotrophs another candidate for microbial extinction followed by a reappearance via convergent evolution or for photoferrotrophy being a persistent microbial trait maintained by HGT. Regardless of the evolutionary processes that drive this pattern, the ancestry of photoferrotrophy based on the modern genomic record is a false negative.

Complex histories of convergence combined with HGT in microbial groups also give rise to genomic false positives—the spurious inference that a pathway or process was present in the common ancestor of a lineage. Even if individual enzymes share a common evolutionary origin, a complete pathway for a metabolic function can be assembled via multiple independent HGT events from different donor groups, so that the presence of the pathway in different lineages is the product of both HGT and convergence. A prime example of such a metabolism is the rTCA cycle for autotrophy in GSB: Phylogenomic reconstructions show that nearly all rTCA

constituent enzymes were acquired by this group via HGT (Becerra & Rivas 2014, Ward & Shih 2022). Moreover, inferred HGT donors differ across these events, and some enzymatic steps, such as isocitrate dehydrogenase, were apparently not acquired in the common ancestor but via multiple independent HGTs into different subgroups of GSB (Paoletti & Fournier 2022). Intriguingly, this piecemeal construction of the pathway is traceable in the geochemical record, as a recent study of GSB-derived chlorobactane from the BCF recovered an isotopic fractionation inconsistent with carbon fixation via rTCA (Zhang et al. 2023). This finding demonstrates how inferences from shared traits within microbial groups can lead to overestimating the ages of associated metabolic processes: The histories of genes and pathways are often more complex than the most parsimonious scenario, but the genomic record also allows the reconstruction of that complexity.

5. LOOKING FORWARD: QUESTIONS AND CHALLENGES

Through the integration of genomic and geological records and methods, the broad outlines of microbial natural history are starting to come into focus. It is unsurprising that this narrative remains disjointed and often contentious, given the relatively short time these investigations have been possible and the scant evidence that is often available for interpretation. Nevertheless, this work collectively shows that, as is the case for complex life, there are indeed major microbial evolutionary transitions in Earth history, where new metabolisms and ecologies develop amid profound taxonomic changes in microbial diversity. These transitions may coincide with known planetary changes preserved in the geological record or may even represent largely biotic events without a clear causal relationship to other types of change. Underlying all of this work is an important caveat: Even as the biogeochemically dominant microbial lineages and processes make themselves known through these records, the vast majority of past microbial diversity is not represented and is perhaps unknowable.

A more comprehensive narrative of microbial natural history will be made possible by continued explorations of both the modern microbial world and the past preserved within the geological record. Continued sequencing efforts, especially environmental and metagenomic studies, promise to provide an ever-more-representative picture of extant microbial diversity—and, by extension, an improved phylogenetic view into past microbial life and patterns of diversification. Organisms capable of performing a variety of energetically feasible metabolisms are still missing from our knowledge of microbial diversity (Broda 1977, LaRowe et al. 2021), and their discovery would substantially inform these efforts. Improved methodologies for detecting and characterizing lipid biomarkers will similarly extend the utility of the microbial fossil record, as will dedicated paleontological investigations of rocks most likely to preserve microfossil evidence. Perhaps most important to this effort is building connections between these disciplines through both scientific training and collaborative research, an exciting synthesis that will continue to shape the future of the field of geobiology.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

LITERATURE CITED

Allwood AC, Grotzinger JP, Knoll AH, Burch IW, Anderson MS, et al. 2009. Controls on development and diversity of Early Archean stromatolites. *PNAS* 106(24):9548–55

Andersson I, Backlund A. 2008. Structure and function of Rubisco. *Plant Physiol. Biochem.* 46(3):275–91

Arning N, Wilson DJ. 2020. The past, present and future of ancient bacterial DNA. *Microb. Genom.* 6(7):mgen000384
Arnold BJ, Huang IT, Hanage WP. 2022. Horizontal gene transfer and adaptive evolution in bacteria. *Nat. Rev. Microbiol.* 20(4):206–18
Arriola LA, Cooper A, Weyrich LS. 2020. Palaeomicrobiology: application of ancient DNA sequencing to better understand bacterial genome evolution and adaptation. *Front. Ecol. Evol.* 8:40
Baldauf SL, Palmer JD, Doolittle WF. 1996. The root of the universal tree and the origin of eukaryotes based on elongation factor phylogeny. *PNAS* 93(15):7749–54
Becerra A, Rivas M. 2014. A phylogenetic approach to the early evolution of autotrophy: the case of the reverse TCA and the reductive acetyl-CoA pathways. *Int. Microbiol.* 17(2):91–97
Bernhard JM, Edgcomb VP, Visscher PT, McIntyre-Wressnig A, Summons RE, et al. 2013. Insights into foraminiferal influences on microfabrics of microbialites at Highborne Cay, Bahamas. *PNAS* 110(24):9830–34
Betts HC, Puttick MN, Clark JW, Williams TA, Donoghue PCJ, Pisani D. 2018. Integrated genomic and fossil evidence illuminates life's early evolution and eukaryote origin. *Nat. Ecol. Evol.* 2(10):1556–62
Bianchi D, Weber TS, Kiko R, Deutsch C. 2018. Global niche of marine anaerobic metabolisms expanded by particle microenvironments. *Nat. Geosci.* 11(4):263–68
Bjerrum CJ, Canfield DE. 2004. New insights into the burial history of organic carbon on the early Earth. *Geochem. Geophys. Geosyst.* 5(8):Q08001
Bosak T, Knoll AH, Petroff AP. 2013. The meaning of stromatolites. *Annu. Rev. Earth Planet. Sci.* 41:21–44
Bottjer DJ. 2010. The Cambrian substrate revolution and early evolution of the phyla. *J. Earth Sci.* 21(Suppl. 1):21–24
Bradley AS, Leavitt WD, Schmidt M, Knoll AH, Girguis PR, Johnston DT. 2016. Patterns of sulfur isotope fractionation during microbial sulfate reduction. *Geobiology* 14(1):91–101
Brasier MD, Antcliffe J, Saunders M, Wacey D. 2015. Changing the picture of Earth's earliest fossils (3.5–1.9 Ga) with new approaches and new discoveries. *PNAS* 112(16):4859–64
Brock TD. 1975. *Milestones in Microbiology*. Washington, DC: Am. Soc. Microbiol.
Brocklehurst RJ, Schachner ER, Codd JR, Sellers WI. 2020. Respiratory evolution in archosaurs. *Philos. Trans. R. Soc. B* 375(1793):20190140
Brocks JJ, Jarrett AJM, Sirantoine E, Hallmann C, Hoshino Y, Liyanage T. 2017. The rise of algae in Cryogenian oceans and the emergence of animals. *Nature* 548(7669):578–81
Brocks JJ, Love GD, Summons RE, Knoll AH, Logan GA, Bowden SA. 2005. Biomarker evidence for green and purple sulphur bacteria in a stratified Palaeoproterozoic sea. *Nature* 437(7060):866–70
Brocks JJ, Schaeffer P. 2008. Okenane, a biomarker for purple sulfur bacteria (Chromatiaceae), and other new carotenoid derivatives from the 1640 Ma Barney Creek Formation. *Geochim. Cosmochim. Acta* 72(5):1396–414
Broda E. 1977. Two kinds of lithotrophs missing in nature. *Z. Allg. Mikrobiol.* 17(6):491–93
Bromham L, Duchêne S, Hua X, Ritchie AM, Duchêne DA, Ho SYW. 2018. Bayesian molecular dating: opening up the black box. *Biol. Rev.* 93(2):1165–91
Brown CT, Hug LA, Thomas BC, Sharon I, Castelle CJ, et al. 2015. Unusual biology across a group comprising more than 15% of domain Bacteria. *Nature* 523(7559):208–11
Buikema WJ, Haselkorn R. 1993. Molecular genetics of cyanobacterial development. *Annu. Rev. Plant Biol.* 44:33–52
Callaway E. 2022. Oldest-ever DNA shows mastodons roamed Greenland 2 million years ago. *Nature* 612(7940):384
Canfield DE, Teske A. 1996. Late Proterozoic rise in atmospheric oxygen concentration inferred from phylogenetic and sulphur-isotope studies. *Nature* 382(6587):127–32
Cardona T. 2015. A fresh look at the evolution and diversification of photochemical reaction centers. *Photosynth. Res.* 126(1):111–34
Cardona T, Sánchez-Baracaldo P, Rutherford AW, Larkum AW. 2019. Early Archean origin of photosystem II. *Geobiology* 17(2):127–50
Cavin L, Forey PL. 2007. Using ghost lineages to identify diversification events in the fossil record. *Biol. Lett.* 3(2):201–4

Chivian D, Brodie EL, Alm EJ, Culley DE, Dehal PS, et al. 2008. Environmental genomics reveals a single-species ecosystem deep within Earth. *Science* 322(5899):275–78

Cui X, Liu X-L, Shen G, Ma J, Husain F, et al. 2020. Niche expansion for phototrophic sulfur bacteria at the Proterozoic–Phanerozoic transition. *PNAS* 117(30):17599–606

Damsté JSS, Schouten S, Hopmans EC, Van Duin ACT, Geenevasen JAJ. 2002. Crenarchaeol. *J. Lipid Res.* 43(10):1641–51

Deckert G, Warren PV, Gaasterland T, Young WG, Lenox AL, et al. 1998. The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*. *Nature* 392(6674):353–58

Demoulin CF, Lara YJ, Cornet L, François C, Baurain D, et al. 2019. Cyanobacteria evolution: insight from the fossil record. *Free Radic. Biol. Med.* 140:206–23

Di Rienzi SC, Sharon I, Wrighton KC, Koren O, Hug LA, et al. 2013. The human gut and groundwater harbor non-photosynthetic bacteria belonging to a new candidate phylum sibling to Cyanobacteria. *eLife* 2:e01102

Doolittle WF. 1999. Phylogenetic classification and the universal tree. *Science* 284(5423):2124–28

Doolittle WF, Booth A. 2017. It's the song, not the singer: an exploration of holobiosis and evolutionary theory. *Biol. Philos.* 32(1):5–24

Dos Reis M, Donoghue PCJ, Yang Z. 2016. Bayesian molecular clock dating of species divergences in the genomics era. *Nat. Rev. Genet.* 17(2):71–80

Dos Reis M, Thawornwattana Y, Angelis K, Telford MJ, Donoghue PCJ, Yang Z. 2015. Uncertainty in the timing of origin of animals and the limits of precision in molecular timescales. *Curr. Biol.* 25(22):2939–50

Duchêne S, Holt KE, Weill F-X, Le Hello S, Hawkey J, et al. 2016. Genome-scale rates of evolutionary change in bacteria. *Microb. Genom.* 2(11):e000094

Enk JM, Yesner DR, Crossen KJ, Veltre DW, O'Rourke DH. 2009. Phylogeographic analysis of the mid-Holocene mammoth from Qagnax Cave, St. Paul Island, Alaska. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 273(1–2):184–90

Ettwig KF, Butler MK, Le Paslier D, Pelletier E, Mangenot S, et al. 2010. Nitrite-driven anaerobic methane oxidation by oxygenic bacteria. *Nature* 464(7288):543–48

Farquhar J, Bao H, Thiemens M. 2000. Atmospheric influence of Earth's earliest sulfur cycle. *Science* 289(5480):756–58

Felsenstein J. 2004. *Inferring Phylogenies*. Sunderland, MA: Sinauer Assoc.

Ferralis N, Matys ED, Knoll AH, Hallmann C, Summons RE. 2016. Rapid, direct and non-destructive assessment of fossil organic matter via microRaman spectroscopy. *Carbon* 108:440–49

Fike DA, Bradley AS, Rose CV. 2015. Rethinking the ancient sulfur cycle. *Annu. Rev. Earth Planet. Sci.* 43:593–622

Fish SA, Shepherd TJ, McGenity TJ, Grant WD. 2002. Recovery of 16S ribosomal RNA gene fragments from ancient halite. *Nature* 417(6887):432–36

Fournier GP, Moore KR, Rangel LT, Payette JG, Momper L, Bosak T. 2021. The Archean origin of oxygenic photosynthesis and extant cyanobacterial lineages. *Proc. R. Soc. B* 288(1959):20210675

Garcia AK, Cavanaugh CM, Kaçar B. 2021. The curious consistency of carbon biosignatures over billions of years of Earth-life coevolution. *ISME J.* 15(8):2183–94

Garcia AK, Kaçar B. 2019. How to resurrect ancestral proteins as proxies for ancient biogeochemistry. *Free Radic. Biol. Med.* 140:260–69

Gest H. 2004. The discovery of microorganisms by Robert Hooke and Antoni van Leeuwenhoek, fellows of the Royal Society. *Notes Rec. R. Soc. Lond.* 58(2):187–201

Golubic S, Seong-Joo L. 1999. Early cyanobacterial fossil record: preservation, palaeoenvironments and identification. *Eur. J. Phycol.* 34(4):339–48

Gribaldo S, Philippe H. 2002. Ancient phylogenetic relationships. *Theor. Popul. Biol.* 61(4):391–408

Gruen DS, Wolfe JM, Fournier GP. 2019. Paleozoic diversification of terrestrial chitin-degrading bacterial lineages. *BMC Evol. Biol.* 19(1):34

Gutell RR, Weiser B, Woese CR, Noller HF. 1985. Comparative anatomy of 16-S-like ribosomal RNA. *Progr. Nucleic Acid Res. Mol. Biol.* 32:155–216

Hayes JM, Freeman KH, Popp BN, Hoham CH. 1990. Compound-specific isotopic analyses: a novel tool for reconstruction of ancient biogeochemical processes. *Org. Geochem.* 16(4–6):1115–28
Hebsgaard MB, Phillips MJ, Willerslev E. 2005. Geologically ancient DNA: fact or artefact? *Trends Microbiol.* 13(5):212–20
Hendy J. 2021. Ancient protein analysis in archaeology. *Sci. Adv.* 7(3):eabb9314
Hinrichs KU, Hayes JM, Sylva SP, Brewer PG, DeLong EF. 1999. Methane-consuming archaebacteria in marine sediments. *Nature* 398(6730):802–5
Holder M, Lewis PO. 2003. Phylogeny estimation: traditional and Bayesian approaches. *Nat. Rev. Genet.* 4(4):275–84
Hug LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, et al. 2016. A new view of the tree of life. *Nat. Microbiol.* 1(5):16048
Hull P. 2015. Life in the aftermath of mass extinctions. *Curr. Biol.* 25(19):R941–52
Javaux EJ. 2019. Challenges in evidencing the earliest traces of life. *Nature* 572(7770):451–60
Johnston DT. 2011. Multiple sulfur isotopes and the evolution of Earth's surface sulfur cycle. *Earth-Sci. Rev.* 106(1–2):161–83
Johnston DT, Wing BA, Farquhar J, Kaufman AJ, Strauss H, et al. 2005. Active microbial sulfur disproportionation in the Mesoproterozoic. *Science* 310(5753):1477–79
Jousset A, Bienhold C, Chatzinotas A, Gallien L, Gobet A, et al. 2017. Where less may be more: how the rare biosphere pulls ecosystems strings. *ISME J.* 11(4):853–62
Kędzior M, Garcia AK, Li M, Taton A, Adam ZR, et al. 2022. Resurrected Rubisco suggests uniform carbon isotope signatures over geologic time. *Cell Rep.* 39(4):110726
Kenrick P, Wellman CH, Schneider H, Edgecombe GD. 2012. A timeline for terrestrialization: consequences for the carbon cycle in the Palaeozoic. *Philos. Trans. R. Soc. B* 367(1588):519–36
Knoll AH. 2015. Paleobiological perspectives on early microbial evolution. *Cold Spring Harb. Perspect. Biol.* 7(7):a018093
Knoll AH, Bergmann KD, Strauss JV. 2016. Life: the first two billion years. *Philos. Trans. R. Soc. B* 371(1707):20150493
Knoll AH, Golubic S, Green J, Swett K. 1986. Organically preserved microbial endoliths from the late Proterozoic of East Greenland. *Nature* 321(6073):856–57
Lapierre P, Gogarten JP. 2009. Estimating the size of the bacterial pan-genome. *Trends Genet.* 25(3):107–10
LaRowe DE, Carlson HK, Amend JP. 2021. The energetic potential for undiscovered manganese metabolisms in nature. *Front. Microbiol.* 12:636145
Lau MCY, Kieft TL, Kuloyo O, Linage-Alvarez B, Van Heerden E, et al. 2016. An oligotrophic deep-subsurface community dependent on syntrophy is dominated by sulfur-driven autotrophic denitrifiers. *PNAS* 113(49):E7927–36
Louca S, Polz MF, Mazel F, Albright MBN, Huber JA, et al. 2018. Function and functional redundancy in microbial systems. *Nat. Ecol. Evol.* 2(6):936–43
Machel HG, Krouse HR, Sassen R. 1995. Products and distinguishing criteria of bacterial and thermochemical sulfate reduction. *Appl. Geochem.* 10(4):373–89
Magnabosco C, Moore KR, Wolfe JM, Fournier GP. 2018. Dating phototrophic microbial lineages with reticulate gene histories. *Geobiology* 16:179–89
Marshall CR. 2019. Using the fossil record to evaluate timetree timescales. *Front. Genet.* 10:1049
Müller AL, Kjeldsen KU, Rattei T, Pester M, Loy A. 2015. Phylogenetic and environmental diversity of DsrAB-type dissimilatory (bi)sulfite reductases. *ISME J.* 9(5):1152–65
Naeher S, Cui X, Summons RE. 2022. Biomarkers: molecular tools to study life, environment, and climate. *Elements* 18(2):79–85
Nelson KE, Clayton RA, Gill SR, Gwinn ML, Dodson RJ, et al. 1999. Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of *Thermotoga maritima*. *Nature* 399(6734):323–29
O'Malley MA. 2009. What did Darwin say about microbes, and how did microbiology respond? *Trends Microbiol.* 17(8):341–47
Pang K, Tang Q, Chen L, Wan B, Niu C, et al. 2018. Nitrogen-fixing heterocystous cyanobacteria in the Tonian Period. *Curr. Biol.* 28(4):616–22.e1

Paoletti MM, Fournier GP. 2022. Chimeric inheritance and crown-group acquisitions of carbon fixation genes within Chlorobiales: origins of autotrophy in Chlorobiales and implication for geological biomarkers. *PLOS ONE* 17(10):e0275539

Paquete CM, Rusconi G, Silva AV, Soares R, Louro RO. 2019. A brief survey of the "cytochromome." *Adv. Microb. Physiol.* 75:69–135

Peters KE, Walters CC, Moldowan JM. 2004. *The Biomarker Guide*. Cambridge, UK: Cambridge Univ. Press. 2nd ed.

Peterson BJ, Fry B. 1987. Stable isotopes in ecosystem studies. *Annu. Rev. Ecol. Syst.* 18:293–320

Philippe H, Laurent J. 1998. How good are deep phylogenetic trees? *Curr. Opin. Genet. Dev.* 8(6):616–23

Phillips A, Janies D, Wheeler W. 2000. Multiple sequence alignment in phylogenetic analysis. *Mol. Phylogenet. Evol.* 16(3):317–30

Post DM. 2002. Using stable isotopes to estimate trophic position: models, methods, and assumptions. *Ecology* 83(3):703–18

Rappé MS, Giovannoni SJ. 2003. The uncultured microbial majority. *Annu. Rev. Microbiol.* 57:369–94

Rashid DJ, Chapman SC, Larsson HC, Organ CL, Bebin A-G, et al. 2014. From dinosaurs to birds: a tail of evolution. *EvoDevo* 5(1):25

Retallack GJ. 2001. Cenozoic expansion of grasslands and climatic cooling. *J. Geol.* 109(4):407–26

Ribeiro AJM, Holliday GL, Furnham N, Tyzack JD, Ferris K, Thornton JM. 2018. Mechanism and Catalytic Site Atlas (M-CSA): a database of enzyme reaction mechanisms and active sites. *Nucleic Acids Res.* 46(D1):D618–23

Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, et al. 2013. Insights into the phylogeny and coding potential of microbial dark matter. *Nature* 499(7459):431–37

Ruta M, Benton MJ. 2008. Calibrated diversity, tree topology and the mother of mass extinctions: the lesson of temnospondyls. *Palaeontology* 51(6):1261–88

Saitta ET, Liang R, Lau MC, Brown CM, Longrich NR, et al. 2019. Cretaceous dinosaur bone contains recent organic material and provides an environment conducive to microbial communities. *eLife* 8:e46205

Sánchez-Baracaldo P. 2015. Origin of marine planktonic cyanobacteria. *Sci. Rep.* 5(1):17418

Sánchez-Baracaldo P, Cardona T. 2020. On the origin of oxygenic photosynthesis and Cyanobacteria. *New Phytol.* 225(4):1440–46

Sansom RS, Gabbott SE, Purnell MA. 2010. Non-random decay of chordate characters causes bias in fossil interpretation. *Nature* 463(7282):797–800

Sayers EW, Beck J, Bolton EE, Bourexis D, Brister JR, et al. 2021. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* 49(D1):D10–17

Schirrmeister BE, De Vos JM, Antonelli A, Bagheri HC. 2013. Evolution of multicellularity coincided with increased diversification of cyanobacteria and the Great Oxidation Event. *PNAS* 110(5):1791–96

Schirrmeister BE, Gugger M, Donoghue PCJ. 2015. Cyanobacteria and the Great Oxidation Event: evidence from genes and fossils. *Palaeontology* 58(5):769–85

Schirrmeister BE, Sánchez-Baracaldo P, Wacey D. 2016. Cyanobacterial evolution during the Precambrian. *Int. J. Astrobiol.* 15(3):187–204

Schwartz SL, Garcia AK, Kaçar B, Fournier GP. 2022. Early nitrogenase ancestors encompassed novel active site diversity. *Mol. Biol. Evol.* 39(11):msac226

Selberg AGA, Gaucher EA, Liberles DA. 2021. Ancestral sequence reconstruction: from chemical paleogenetics to maximum likelihood algorithms and beyond. *J. Mol. Evol.* 89(3):157–64

Sephus CD, Fer E, Garcia AK, Adam ZR, Schwieterman EW, Kaçar B. 2022. Earliest photic zone niches probed by ancestral microbial rhodopsins. *Mol. Biol. Evol.* 39(5):msac100

Shapira M. 2016. Gut microbiotas and host evolution: scaling up symbiosis. *Trends Ecol. Evol.* 31(7):539–49

Sharon I, Alperovitch A, Rohwer F, Haynes M, Glaser F, et al. 2009. Photosystem I gene cassettes are present in marine virus genomes. *Nature* 461(7261):258–62

Shi M, Feng Q, Khan MZ, Zhu S. 2017. An eukaryote-bearing microbiota from the early mesoproterozoic Gaoyuzhuang Formation, Tianjin, China and its significance. *Precambrian Res.* 303:709–26

Soucy SM, Huang J, Gogarten JP. 2015. Horizontal gene transfer: building the web of life. *Nat. Rev. Genet.* 16(8):472–82

Spang A, Saw JH, Jørgensen SL, Zaremba-Niedzwiedzka K, Martijn J, et al. 2015. Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature* 521(7551):173–79
Stewart FJ, Newton ILG, Cavanaugh CM. 2005. Chemosynthetic endosymbioses: adaptations to oxic–anoxic interfaces. *Trends Microbiol.* 13(9):439–48
Strous M, Fuerst JA, Kramer EH, Logemann S, Muyzer G, et al. 1999. Missing lithotroph identified as new planctomycete. *Nature* 400(6743):446–49
Summons RE, Powell TG. 1986. Chlorobiaceae in Palaeozoic seas revealed by biological markers, isotopes and geology. *Nature* 319(6056):763–65
Szöllősi GJ, Tannier E, Lartillot N, Daubin V. 2013. Lateral gene transfer from the dead. *Syst. Biol.* 62(3):386–97
Takai K. 2019. Limits of terrestrial life and biosphere. In *Astrobiology*, ed. A Yamagishi, T Kakegawa, T Usui, pp. 323–44. Singapore: Springer Singapore
Telford MJ, Budd GE, Philippe H. 2015. Phylogenomic insights into animal evolution. *Curr. Biol.* 25(19):R876–87
Thompson KJ, Kenward PA, Bauer KW, Warchola T, Gauger T, et al. 2019. Photoferrotrophy, deposition of banded iron formations, and methane production in Archean oceans. *Sci. Adv.* 5(11):eaav2869
Tomitani A, Knoll AH, Cavanaugh CM, Ohno T. 2006. The evolutionary diversification of cyanobacteria: molecular–phylogenetic and paleontological perspectives. *PNAS* 103(14):5442–47
Tsuji JM, Tran N, Schiff SL, Venkiteswaran JJ, Molot LA, et al. 2020. Anoxygenic photosynthesis and iron–sulfur metabolic potential of *Chlorobia* populations from seasonally anoxic Boreal Shield lakes. *ISME J.* 14(11):2732–47
Usher KM, Bergman B, Raven JA. 2007. Exploring cyanobacterial mutualisms. *Annu. Rev. Ecol. Evol. Syst.* 38:255–73
Van Zuilen MA, Lepland A, Arrhenius G. 2002. Reassessing the evidence for the earliest traces of life. *Nature* 418(6898):627–30
Vandevenne FI, Barão AL, Schoelynck J, Smis A, Ryken N, et al. 2013. Grazers: biocatalysts of terrestrial silica cycling. *Proc. R. Soc. B* 280(1772):20132083
Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, et al. 2004. Environmental genome shotgun sequencing of the Sargasso Sea. *Science* 304(5667):66–74
Wagner M, Roger AJ, Flax JL, Brusseau GA, Stahl DA. 1998. Phylogeny of dissimilatory sulfite reductases supports an early origin of sulfate respiration. *J. Bacteriol.* 180(11):2975–82
Wainwright M. 2009. Darwin and microbiology. *Microbiologist* 10:32–35
Walker JCG, Brimblecombe P. 1985. Iron and sulfur in the pre-biologic ocean. *Precambrian Res.* 28(3–4):205–22
Wang S, Luo H. 2023. Dating the bacterial tree of life based on ancient symbiosis. bioRxiv 2023.06.18.545440. **https://doi.org/10.1101/2023.06.18.545440**
Ward LM, Shih PM. 2022. Phototrophy and carbon fixation in Chlorobi postdate the rise of oxygen. *PLOS ONE* 17(8):e0270187
Weijers JWH, Schouten S, Hopmans EC, Geenevasen JAJ, David ORP, et al. 2006. Membrane lipids of mesophilic anaerobic bacteria thriving in peats have typical archaeal traits. *Environ. Microbiol.* 8(4):648–57
Wetmur JG, Wong DM, Ortiz B, Tong J, Reichert F, Gelfand DH. 1994. Cloning, sequencing, and expression of RecA proteins from three distantly related thermophilic eubacteria. *J. Biol. Chem.* 269(41):25928–35
Whiteside JH, Grice K. 2016. Biomarker records associated with mass extinction events. *Annu. Rev. Earth Planet. Sci.* 44:581–612
Woese CR. 1987. Bacterial evolution. *Microbiol. Rev.* 51(2):221–71
Woese CR, Fox GE. 1977. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *PNAS* 74(11):5088–90
Woese CR, Stackebrandt E, Weisburg WG, Paster BJ, Madigan MT, et al. 1984. The phylogeny of purple bacteria: the alpha subdivision. *Syst. Appl. Microbiol.* 5(3):315–26
Wood R. 2011. General evolution of carbonate reefs. In *Encyclopedia of Modern Coral Reefs*, ed. D Hopley, pp. 452–69. Dordrecht, Neth.: Springer Netherlands

Xu X, Zhou Z, Dudley R, Mackem S, Chuong C-M, et al. 2014. An integrative approach to understanding bird origins. *Science* 346(6215):1253293

Yu H, Leadbetter JR. 2020. Bacterial chemolithoautotrophy via manganese oxidation. *Nature* 583(7816):453–58

Zhang X, Paoletti MM, Izon G, Fournier GP, Summons RE. 2023. Late acquisition of the rTCA carbon fixation pathway by Chlorobi. *Nat. Ecol. Evol.* 7:1398–407

Zuckerkandl E. 1962. Molecular disease, evolution, and genic heterogeneity. *Horiz. Biochem.* 1962:189–225

Zuckerkandl E, Pauling L. 1965. Evolutionary divergence and convergence in proteins. In *Evolving Genes and Proteins*, ed. V Bryson, HJ Vogel, pp. 97–166. New York: Academic