Trends in
Ecology & Evolution

CellPress

Review

The earliest history of eukaryotic life: uncovering an evolutionary story through the integration of biological and geological data

Phoebe A. Cohen [1,*,@] and Robin B. Kodner[2,*]

While there is significant data on eukaryogenesis and the early development of the eukaryotic lineage, major uncertainties regarding their origins and evolution remain, including questions of taxonomy, timing, and paleoecology. Here we examine the origin and diversification of the eukaryotes in the Proterozoic Eon as viewed through fossils, organic biomarkers, molecular clocks, phylogenies, and redox proxies. Our interpretation of the integration of these data suggest that eukaryotes were likely aerobic and established in Proterozoic ecosystems. We argue that we must closely examine and integrate both biological and geological evidence and examine points of agreement and contention to gain new insights into the true origin and early evolutionary history of this vastly important group.

Highlights

We have made great strides in understanding the origin and early evolution of eukaryotes, both from a biological perspective using genomic data and molecular clocks, and from a geological perspective using fossils, biomarkers, and geochemical proxies.

Often these lines of evidence have been developed in isolation and sometimes come to different conclusions about the deep history of eukaryotes; however, recently the biological and geological perspectives have become more integrated.

These integrations have helped sharpen our understanding of origin and early evolution of eukaryotes while highlighting important conflicts in the data that leave many outstanding questions about the taxonomy of early eukaryotes, the role of ecology in their early evolution, and influences of the environment, especially oxygen, during the Proterozoic Eon.

Early eukaryotic history

The early evolution of eukaryotes is one of the most important processes in the history of life on Earth (Box 1). Eukaryotes, as a **clade** (see Glossary) or **lineage**, are defined in terms of cellular complexity [1]. This complexity enables morphological diversity in both unicellular and multicellular lineages, including animals, fungi, algae, and plants. Eukaryotes have also played a role in shaping the physical landscape and biogeochemical cycles on Earth, ever since their origin in the early **Proterozoic**. Because the origin and early evolution of eukaryotes is so ancient, and many of the features that define the group are unlikely to be preserved in the geologic record, it has been challenging to uncover their early evolutionary history. Traditionally, our knowledge of the history of a major lineage such as the eukaryotes would rely heavily on the fossil record. While the fossil record of early eukaryotes is indeed vital to telling their story, it is also a relatively sparse and patchy source of information. Many fields, including genomics, phylogenetics, organic geochemistry, and redox geochemistry, have now added new layers to our understanding of this ancient history. There are now convincing data for the origin of the eukaryote clade in the early Proterozoic and compelling data for the diversification and increased ecological importance in the late Proterozoic, but much less consensus on events happening during the vast middle of this important interval. Here, we discuss the new interdisciplinary approaches that have made major contributions to our understanding of the origin and early evolutionary history of eukaryotes, focusing on areas of uncertainty to reveal new perspectives.

Lines of evidence used to investigate Proterozoic eukaryotes

The fossil record

The majority of the early eukaryotic fossil record consists of **acritarchs**: closed, spherical single-celled organic-walled microfossils of uncertain taxonomic affinity (Figure 1). As soon as one

[1]Williams College Department of Geosciences, Williamstown, MA, USA
[2]Western Washington University Department of Environmental Sciences, Bellingham, WA, USA

*Correspondence:
pac3@williams.edu (P.A. Cohen) and
kodnerr@wwu.edu (R.B. Kodner).
@Twitter: @PhoebeFossil (P.A. Cohen).

figures out what an acritarch actually is, it is no longer an acritarch! Much more detail on the history of these fossils and their morphology and categorization exists elsewhere (e.g., [2]), but critical for this discussion is that acritarchs are a polyphyletic group that likely consists of algae, **metazoans**, and non-algal microeukaryotes, if not also fungi; some may also represent bacteria or archaea, although their size and structure often precludes these affinities. Despite this broad array of possible affinities, starting ca. 1.6 Ga, some groups of acritarchs appear in the fossil record that can confidently be assigned to **total group** eukaryotes and show evidence of diagnostic morphological features, such as a dynamic cytoskeleton [3,4].

Other major fossil groups in the Proterozoic include complex branching filaments (often assigned to algal or fungal groups [5]), tests (essentially single-celled shells [6,7]), and biomineralized structures, most notably apatitic scale microfossils [8] (Figure 1). Many of these early fossils, while unequivocally eukaryotic, cannot be assigned to a modern clade. Thus, while there are many eukaryotic fossils present in this time period, the limited information on their taxonomic affinity makes it difficult to infer more about their biology, ecology, and evolution.

Molecular phylogenetics and phylogenomics

One of the major recent innovations in reconstructing the evolutionary history of life uses the rapidly growing databases of genomic data of key lineages in large-scale **phylogenomic** analyses. Recent reconstructions of the eukaryotic tree of life using phylogenomics have higher confidence in the topology of super-kingdom level structure: of the tree as well as in definitions of the major supergroups [9] (Figure 2). **Molecular clock** analyses, models that date major divergences on phylogenetic trees, have greatly contributed to our understanding of the timing of origin of eukaryotes and their early lineages [10,11]. Molecular clocks use sophisticated statistical and evolutionary models to predict the timing of the divergence points in lineages based on gene sequence data and are often calibrated with fossils [10]. The divergence times predicted using these methods are broad but provide critical context for integrating biological information with the fossil record, especially when fossils that are often used to root the models are well-dated and can be confidently assigned to a specific taxonomic group. Generally, molecular clock estimates of divergence times agree with the fossil record (in part by design, since they are often calibrated with fossils), especially with the important caveat that the first appearance of a taxon in the fossil record will always be later than its true evolutionary origination date; this lag is likely even more pronounced in the Proterozoic given the relative incompleteness of the Proterozoic rock record as compared with the Phanerozoic [12].

Organic biomarkers

Biolipid molecules can be preserved in deep time, retain structures that can be affiliated with specific taxa, and act as molecular fossils known as organic biomarkers (Box 2). **Steranes**, the diagenetic byproduct of **sterol** molecules, are the most commonly used biomarker for reconstructing early eukaryotic evolutionary history. Sterols are required by all modern eukaryotes (with the exception of a few derived taxa that have replaced them with other molecules, e.g., [13]) and most biosynthesize them. While sterols are also produced by some bacteria [14], these differ in structure from those produced by most eukaryotes [15], thus steranes are still considered reliable eukaryotic biomarkers. In addition to their utility as a taxonomic biomarker for eukaryotes, sterol biosynthesis requires free O_2 [15,16], thus their presence reinforces the relationship between eukaryotes and oxygen when they are found in the rock record. In addition, new evidence suggests that oxygen-dependent sterol biosynthesis may have evolved in bacteria in response to the initial rise in O_2 at the **great oxidation event**

Glossary

Acritarch: an organic spherical microfossil of uncertain taxonomic affinity.
Aerobic: referring to a metabolism that does require free oxygen.
Anaerobic: referring to a metabolism that does not function in the presence of free oxygen.
Anoxic: environment without any detectable free oxygen.
Crown group: a lineage of organisms still extant today. Includes the last common ancestor of a clade and all of its direct descendants, including extant lineages (can include extinct organisms, as long as they have the same common ancestor as all living members of the clade).
Clade: a natural group of organisms descended from a common ancestor (aka monophyletic). See 'Lineage'.
Cryogenian: geological period from 720 to 635 Ma.
Ediacaran: geological period from 635 to 541 Ma.
Eukaryogenesis: the events that led to the formation of the eukaryotic clade.
Eukaryvory: predation by one organism on a eukaryote.
Euxinic: environments that are anoxic and sulfidic.
Great oxidation event (GOE): the initial rise of oxygen ca. 2.5 billion years ago (Ga).
Lineage: a group of related organisms descended from a common ancestor. Lineages make up parts of a phylogenetic tree and lineages can include multiple clades.
Metazoan: taxonomic term for the animal lineage.
Molecular clock: a phylogenetic algorithm that uses the rate at which a species' genome accumulates mutations to measure and predict divergence times of lineages.
Neoproterozoic: geological era from 1000 to 541 Ma.
Oxic: environment with detectable levels of free oxygen.
Proterozoic: geological eon lasting from 2500 to 541 Ma.
Phylogenomics: large-scale phylogenetic analysis using genomic data. Analyses include hundreds of genes, large sections of genomes, or even whole genomes.
Snowball Earths: two events during the Cryogenian Period where the globe was fully glaciated all the way down to the poles; the first event is called the

(GOE) and was transferred to eukaryotes, suggesting that features of eukaryotic membranes existed prior to **eukaryogenesis** [17].

Redox proxies

While **crown group** eukaryotes as a clade are not obligate aerobes, the vast majority of them perform respiration with oxygen as the electron acceptor and produce oxygen-requiring sterols. In addition, metazoan eukaryotes require even higher levels of oxygen to attain their large body size [18]. Thus, the history of eukaryotes and the history of oxygen are inexorably intertwined. Reconstructing past atmospheric and oceanic oxygen levels is a notoriously difficult process because there is no known method to directly measure paleo-oxygen. A myriad array of redox-sensitive geochemical proxies has been used over the past decades to try to reveal deep time oxygen levels, including the geochemical speciation of Fe (Box 3) and the abundance of and isotopic fractionation of Mo, Cr, S, Mo, V, and U [19–21]. The specific trends that each proxy reveals vary based on their sensitivities and the element-specific processes that impart fractionation and/or changes in abundance. While there are many different views about the actual levels of oxygen present in the atmosphere/surface ocean and the deep ocean during the Proterozoic, the broad brush picture is relatively agreed upon: oxygen increased at the GOE approximately 2.4 Ga, was low during the Proterozoic, and rose through the late **Ediacaran**-early Paleozoic to reach modern levels by the Devonian Period ca. 400 million years ago [22–25].

Origin of the eukaryotic clade

It is highly unlikely that we could ever identify the origins of eukaryotes via the fossil record itself due to the incompleteness of the fossil record and the lack of preservable defining morphological characteristics in many eukaryotes [1]. The first fossils unequivocally identified as total group eukaryotes are acritarchs from ca. 1.7–1.5 Ga [4,26]. These fossils are considered eukaryotes due to a combination of morphological features, including overall size, complex cell wall ultrastructure, external processes (arms), and double-walled enveloping membranes. Meanwhile, multiple molecular clock analyses indicate that crown group eukaryotes evolved in the middle to late Mesoproterozoic, or even in the earliest **Neoproterozoic** (ca. 2.0–1.0 Ga) [27–29]. While many clocks focus on the crown, Betts *et al.* estimate total-group eukaryotes [i.e., first eukaryotic common ancestor (FECA)] evolved by 1.6 Ga at a minimum and more likely closer to 3.0 Ga [27]; these authors also suggest that last eukaryotic common ancestor (LECA) most likely emerged ca. 2.2 Ga. As there are no conclusive sterane biomarkers until the **Tonian** Period, the current biomarker record cannot shed light on the origin of the eukaryotic lineage (Box 2 and Figure 2).

The early diversification and ecology of the eukaryotic clade

Biodiversity: fossil, biomarkers, and clocks

The first fossil confidently attributed to a known taxonomic clade of crown group eukaryotes is the ca. 1.0 Ga fossil *Bangiomorpha pubescens*, which is interpreted as a member of the Bangiales within the Rhodophyta [30,31], a member of the Archaeplastida supergroup (Figure 2). Other fossils of equivalent age have been assigned to the Chlorophyta, a sister lineage to the Rhodophyta within the Archaeplastida, including *Proterocladus antiquus* [32] and an unnamed macroscopic filamentous fossil from northwest Canada [33]. Thus, by ca. 950 Ma, at least two major clades within Archaeplastida are present in the fossil record (Figure 2).

Many molecular clock analyses use *B. pubescens* to calibrate their clocks; however, there has been much discussion about how to categorize *B. pubescens* [28,34]. Some authors interpret it as a member of the red algae **stem group**, which has significant implications for the clock predictions [29]. Some use the 1.6 Ga Changcheng Formation acritarchs as the oldest fossil evidence of total-group eukaryotes [27], whereas others do not include this datum in their

Sturtian Glaciation and the second the Marinoan Glaciation.
Stem group: phylogenetic category that includes exclusively extinct groups on basal branches of phylogenetic trees that have some, but not all, traits of the crown group and thus diverged prior to the last common ancestor.
Sterane: the carbon skeleton of a sterol molecule produced by diagenesis, the fossilized form of a sterol.
Sterol: a triterpenoid lipid molecule made by most eukaryotes as a major membrane component.
Synapomorphy: a trait present in an ancestral species and shared exclusively by its evolutionary descendants.
Syntrophy: one species living off of the metabolic products of another species.
Tonian: geological period lasting from 1000 to 720 Ma.
Total group: a phylogenetic category that includes both the stem and crown groups or clades of a lineage.

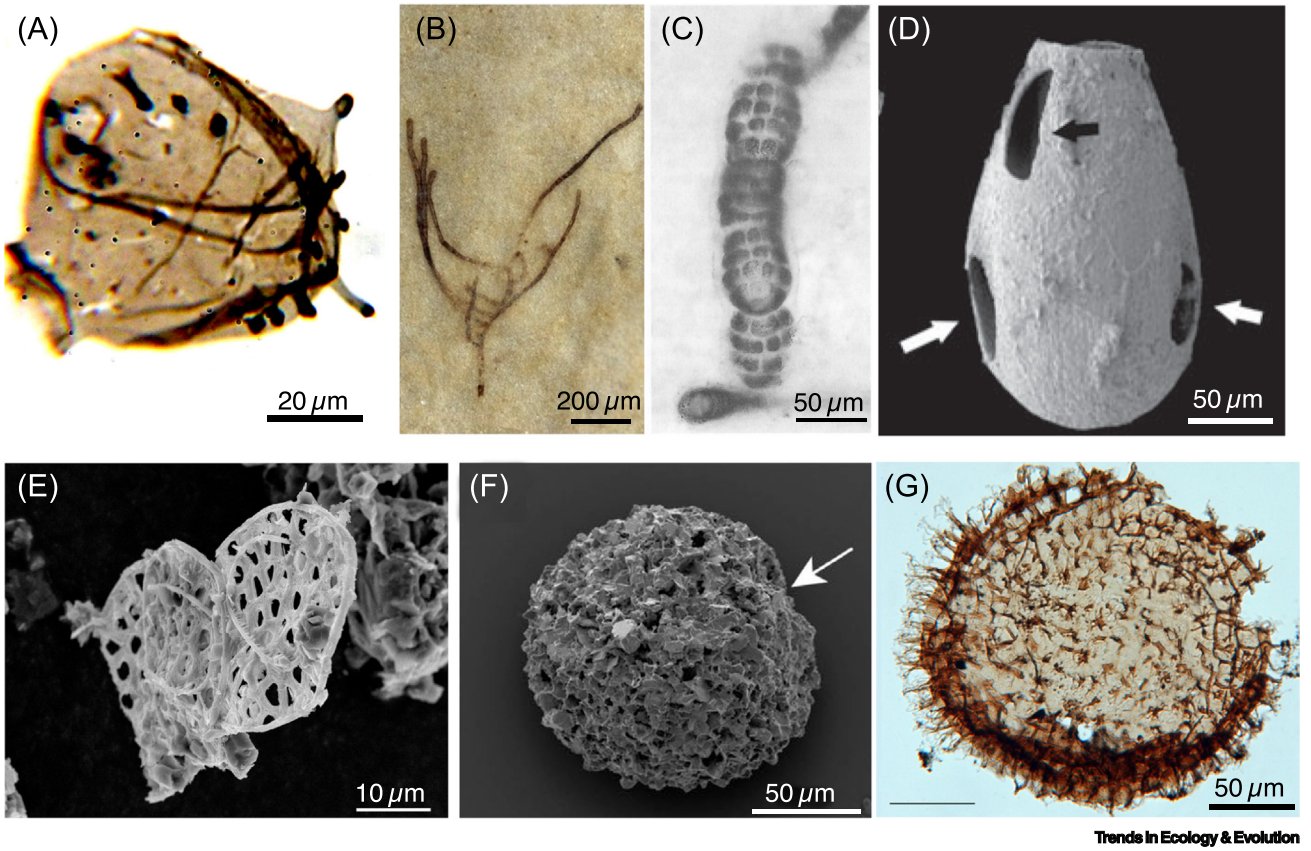

Figure 1. Representative Proterozoic eukaryotic fossils. (A) *Tappania plana* from the Mesoproterozoic Ruyang Group showing complex morphology, including processes [4] (B) Green alga *Proterocladus antiquus* from the ca. 1000 Ma Nanfen Formation, North China [32]. (C) Red alga *Bangiomorpha pubescens* from the ca. 1000 Ma Hunting Formation, Canada [30]. (D) Vase-shaped microfossil with predation marks, ca. 735 Ma Chuar Group, USA [49]. (E) Unnamed taxon apatitic scale microfossil from the 810 Ma Fifteenmile Group, Yukon [8]. (F) Agglutinated eukaryotic fossil from the Cryogenian Kakontwe Formation of Zambia [7]. (G) *Archaeotunisphaeridium fimbriatu* from the Ediacaran Giles 1 core, Australia [63].

analyses [28]. No published molecular clocks include the new green algal fossil *P. antiquus* in their analyses. Regardless if *B. pubescens* or *P. antiquus* are members of their crown groups, the fact that both are members of their total groups indicates that the split between red and green algae had already taken place by ca. 950 Ma (Figure 2). Thus, their presence in the fossil record indicates that crown group Archaeplastida, and thus crown group Eukaryota, were also present (Figure 2).

The oldest well-accepted indigenous steranes are from the middle Tonian (Box 2 and Figure 2) [35]. These biomarkers are unusual because the sterane signature is dominated by cholestane [36], a pattern not seen in younger biomarker records. This is interpreted as either a sign of red algal dominance, or of heterotrophic input, as both groups have a higher proportion of cholestane as compared with other groups such as green algae [37]. As the Neoproterozoic progresses, fossil and biomarker records show the appearance of additional crown group eukaryotes in the rock record of the late Tonian (vase-shaped microfossils interpreted as amoebozoans, [38]) and **Cryogenian** (biomarkers indicative of metazoans, Box 2 [39]), both members of the Amorphea supergroup. While there is only direct evidence for two of the major eukaryotic supergroups during this period, the Archeoplastida and Amorphea are somewhat derived in the most

Box 1. What is a eukaryote and where did they come from?

Eukaryotes compartmentalize their genetic material into a nucleus, are contained within a structurally complex cytoskeleton, and all have mitochondria or an organelle of mitochondrial origin [65]. These cellular innovations work together to enable compartmentalization of biochemistry, the evolution of complex cell morphologies, and, ultimately, multicellularity in multiple lineages. Eukaryotes are chimeras in both structure and function: mitochondria and plastids (when present) originated from ancient endosymbiotic events with bacteria (alphaproteobacteria and cyanobacteria, respectively), and recent discoveries of new taxa confidently suggest the host cell (i.e., nucleocytoplasm) derived from an Asgard archaeal lineage [64,66,67]. There are a number of phylogenetic, cellular, and biochemical features uncovered in recent years that link Asgard archaea to eukaryotes [64,66–69] and have led to new hypotheses for the role of **syntrophy** in the development of oxygen tolerance during eukaryogenesis [70].

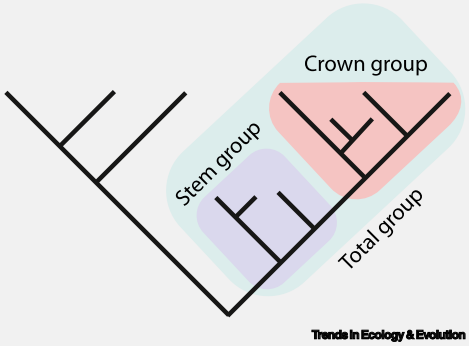

Figure I. Overview of phylogenetic stem versus crown versus total group concept. Note that stem groups are by definition all extinct and diverge before the last common ancestor of all members of the crown group. See Glossary for definitions of stem, crown, and total groups.

LECA, once established, must have had at least a cytoskeleton, nucleus, and mitochondria to be a eukaryote. The vast majority of eukaryotes perform respiration with oxygen as the electron acceptor and, while some can perform fermentation and other forms of anaerobic respiration, their metabolic pathways for generating energy are depauperate when compared with bacteria and archaea [65,71]. The original plastids (i.e., chloroplasts) derived from a subsequent endosymbiotic event with a cyanobacteria and became the center of oxygenic photosynthesis in eukaryotes [72]. However, most photosynthetic eukaryotes are best described as mixotrophs because both respiration and photosynthesis play a major role in their metabolisms. Through their complex evolutionary history, plastids also became a driver for evolution and diversification in many other clades within the eukaryotes, such as stramenopiles, haptophytes, alveolates, excavates, and others, via subsequent endosymbiotic events [73]. Despite agreement on the overall story, there is still controversy over the actual partners involved as well as the timing and mechanisms of the primary symbiotic events [64]. The first agreed upon eukaryote fossils have a complex cell wall but no other preserved features, thus it is unclear if they represent stem or crown groups (Figure I). The first crown group eukaryote in the fossil record is an Archaeplastida, and thus a product of two endosymbiotic events, and a few evolutionary steps from LECA. It is possible, or even likely, that unclassifiable stem groups and recognizable crown groups coexisted in Proterozoic ecosystems.

recent eukaryote tree of life, thus other major lineages must have been present by the Tonian/Cryogenian as well (Figure 2).

Based on the relatively late appearance of steranes, Brocks *et al.* interpret the overall Proterozoic biomarker record to indicate an almost total lack of eukaryotic phytoplankton before 800 Ma [35]. Using the same data set, these authors locate a 'rise of algae' between 659 and 645 Ma, defined by a distinct biomarker signature in the Cryogenian and Ediacaran, interpreted as a chlorophyte signal [36]. Thus, the biomarker record, when it appears, is consistent with other lines of evidence for mid to late Proterozoic Archaeplastida.

Eukaryotic fossil diversity took an (understandable) hit during the 55-Ma long Sturtian **Snowball Earth** event [40,41]. After the end of the second Snowball Earth event at ~635 Ma, within-assemblage diversity bounces back, including a high diversity of large, spinose acritarchs that have been interpreted as possible animal egg cases [42]. As the Ediacaran progresses, large, diverse macroscopic organisms appear in the fossil record, both putative metazoan Ediacarans, as well as macrophytes (seaweeds) [40,43–45]. With the close of the Proterozoic, numerous fossils

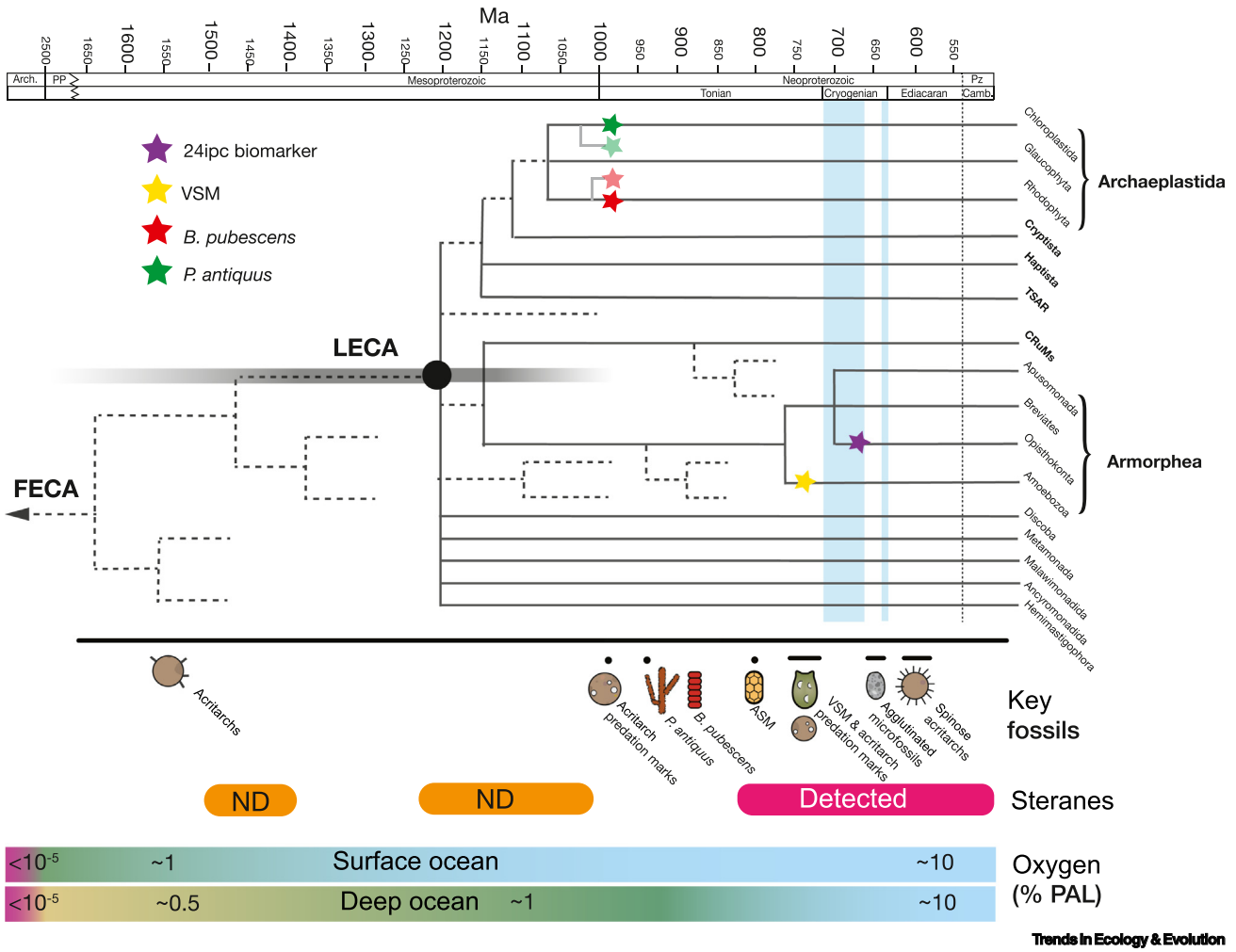

Figure 2. Overview of major trends and proxies in the Proterozoic evolution of eukaryotes. Shaded blue areas indicate Snowball Earth events, filled stars represent crown group fossils, faded stars represent stem group fossils. Phylogeny adapted and simplified from [9]; bolded names are supergroups. See text for discussion of biomarker and oxygen data. Shading around the LECA point indicates uncertainty about the timing of LECA based on molecular clock analyses [27,64]. Abbreviations: Arch, archaean; ASM, apatitic scale microfossils; FECA, first eukaryotic common ancestor (base of total group Eukaryota; LECA, last eukaryotic common ancestor (base of crown group Eukaryota); PAL, present atmospheric levels; PP, Paleoproterozoic; Pz, Paleozoic; VSM, vase-shaped microfossil.

assigned to the metazoan clade are present in the fossil record [46], ushering in the Phanerozoic Eon and the rapid diversification of animal life known as the Cambrian Radiation [46,47].

Ecology: interactions, predations, and biomineralization

The fossil record can inform us not only about the origination of specific clades, but also about paleoecology and organismal interactions. Eukaryotic fossils have been identified with what look like predation marks [48,49] starting at 1.0 Ga and then again at ca. 760 Ma, indicating the origin of **eukaryvory**. Biomineralization in eukaryotes first appears ca. 800 Ma [8], indicating that organisms were under selective pressure to put energy and materials into the construction of hard parts, perhaps as a response to predation pressure [50,51]. Resistant fossil structures such as agglutinated tests continue to appear throughout the Tonian and Cryogenian (e.g., [7]), though these are not primarily biomineralized structures (Figure 1). Nutrient dynamics may have also played a role in driving ecological evolution in Proterozoic oceans. Brocks *et al.* argue that eukaryote

Box 2. The bumpy road of organic biomarkers

The use of organic biomarkers to reconstruct evolutionary events only came into common practice in the 1960s [74]. As such, the last few decades have seen major changes in how this work is done and interpreted. Contamination is a major issue because organic biomarkers in ancient rocks are often found at very low concentrations and sources of these molecules are ubiquitous in modern petroleum-based materials. For example, evidence of eukaryotic steranes from Archean rocks from Australia [75], once heralded as the earliest evidence of eukaryotes, have now been shown to be contamination, perhaps from drilling fluids [76]. However, new analytical techniques have greatly reduced this risk, including continuous-flow hydropyrolysis, which examines biomarkers from kerogen (insoluble organic matter) instead of bitumen (soluble organic matter) and is therefore much less likely to be impacted by contamination. In addition, increased attention to contamination and specialized drilling using nonorganic drilling fluids has led to more reliable results [76,77]. While the authenticity of data derived using best practices is now less of an issue, interpretation of ancient steranes can still be controversial. The best examples of this are questions about the source of 24-ipc steranes, which may represent the oldest evidence of metazoans. These molecules, with a unique preservable side chain, have been interpreted as diagnostic biomarkers for the demosponge clade and are found in Cryogenian-aged rocks from Oman [39]. Some argue that this datum is a robust record of the oldest direct evidence for the metazoan clade [78], while others argue that the molecule could have been produced by other groups, such as algae or protists [79–81]. Phylogenetic investigation of the sterol biosynthetic pathways of algae, fungi, and metazoans does not rule out the possibility that organisms other than metazoans could have produced this molecule [82]. Despite the ongoing arguments about the affinity of 24-ipc [39,76–79], other lines of evidence, including molecular clock estimates on the origin of metazoa and branches co-equal to the metazoa in the eukaryotic tree, present as body fossils in older rocks (Figure 2), agree with this timing of the origin of the animal lineage [47].

evolution was limited by phosphate until the latest Neoproterozoic and that increases in phosphate availability, driven by the melting Snowball Earth, set off a cascade of diversification and increases in abundance of eukaryote primary producers, which culminated in the origin of metazoa [36]. However, this argument is highly dependent on the accuracy of the biomarker record and the true lack of steranes in the Mesoproterozoic.

Conflict is inevitable: the biomarker gap

Interpretations of the multiple lines of evidence used to reconstruct early evolutionary history of eukaryotes are not always aligned. The biggest issue is that the body fossil record and molecular clock predictions suggest a Paleo- to- Mesoproterozoic origin of total group eukaryotes, yet there is a lack of eukaryotic biomarkers until the Neoproterozoic (Figure 2). The early eukaryote fossil record is relatively robust: there are 32 stratigraphic units with mean ages within the 1700–800 Ma range that contain eukaryotic fossils, the same period when the biomarker record seems devoid of steranes. It is not until the late Tonian when the first well-established steranes appear in the biomarker record, which we refer to as the ‘sterane gap’. Proterozoic molecular paleontology is a relatively new field and, as such, the community is still putting bounds on

Box 3. How does the iron speciation proxy for molecular oxygen work?

The iron speciation redox proxy tracks the proportion of iron in a sample that is highly reactive. This includes pyrite, iron carbonates, magnetite, and iron (oxyhydr)oxides. Generally, two ratios are analyzed. The first is the amount of reactive iron, or Fe_{HR}, relative to the total amount of iron, or Fe_T. This ratio can distinguish between **oxic** and anoxic water columns. Importantly, this proxy is essentially a binary: anything below a value of ~0.2 is oxic, but could be dysoxic (i.e., still have very low amounts of oxygen) [83]. For the purposes of eukaryotic evolution, this proxy works well because we know eukaryotes likely only needed trace amounts of oxygen to evolve and synthesize sterols (e.g., [16]). The other ratio that is analyzed is the amount of pyrite, Fe_{Py}, as compared with Fe_{HR}. This ratio can determine whether a water column experienced ferruginous (anoxic + iron rich) or euxinic (anoxic + sulfidic) conditions [83]. This also has implications for eukaryotic evolution as hydrogen sulfide is toxic to most eukaryotes because it inhibits the mitochondrial respiratory chain (although many eukaryotes have evolved adaptations to deal with this issue [71]). While trace metal enrichment data sets (e.g., [84,85]) generally respond to the total size of anoxic sinks in the global ocean, iron speciation data generally tracks the percentage of samples in a given time bin that interacted with anoxic waters. Thus, changes in time-binned iron speciation data sets require a larger shift in overall global ocean oxygenation than trace metal enrichment proxies [56]. Importantly, most redox proxies, including iron speciation, record water column oxygen levels on a millennial or greater time scale [86]. Thus, short-term oxygenation, or anoxia, may not be recorded by redox proxies, but may still have major impacts on eukaryotic organisms and communities. These discrepancies between geological time and biological and organismally relevant time are inherent challenges in this interdisciplinary field.

confidence in these data (Box 2). Some have argued that the Proterozoic biomarker record shows a late appearance of steranes because eukaryotic organisms were a minor part of global ecosystems relative to bacteria, thus their signal was swamped by bacterial biomarkers [35]. But why should we trust biomarkers as a proxy for true organismal relative abundance over the fossil record itself? One potential reason is that bacteria and archaea do not fossilize at the same rate as eukaryotes because of their generally smaller sizes and lack of recalcitrant structures, so the fossil record is biased against their preservation. However, there may be other biases in favor of their biomarker preservation relative to eukaryotes [52,53].

Modern taphonomic studies indicate preferential degradation of sterols over hopanoids within seafloor microbial mat communities that could bias the biomarker record [53]. This is known as the 'mat seal effect' and has been implicated in the sterane gap [52]. The presence of abundant microbial mats may select for hopane preservation sourced from abundant mat bacterial communities, while also contributing to physical and biogeochemical degradation of eukaryotic sterols to the point that they are no longer present when the organic material reaches sediments [52,53]. The mat seal effect research indicates that we should potentially view the sterane biomarker record as a one-sided constraint (i.e., if reliable steranes are detected then we can trust that there were indeed eukaryotes living in that area). But if we do not see sterane biomarkers, then **aerobic** eukaryotes could still have been living in those environments. In addition to taphonomic arguments, there is some evidence that Mesoproterozoic steranes actually do exist [54], though this study was unable to recover steranes from bitumen in the same samples and thus may represent contamination or some alternative preservation pathway.

It is also possible that eukaryotic communities truly were marginal pre-800 Ma. The within-assemblage diversity of eukaryotes and ecosystem complexity (i.e., evidence of eukaryvory) does increase during the late Tonian at around the same time that the biomarker record starts to include abundant steranes, suggesting that some abundance threshold might have been met to lead to detectable steranes in the geologic record [50]. Thus, it is possible that a shift in community dynamics and abundance, possibly in concert with a change in taphonomic barriers, could help to reconcile the biomarker with other lines of evidence for eukaryotic diversification.

Detangling the relationship between eukaryotes and molecular oxygen

Availability of molecular oxygen is often considered in the origin and evolution of eukaryotes, because of the fundamental role that heterotrophic metabolism plays in this clade. The GOE marks the time in Earth's history when oxygenation of Earth's ecosystems first began on a global scale. The disappearance of mass independent sulfur fractionation ca. 2.4–2.5 billion years ago is the best agreed upon marker of the GOE, though this signature may not reflect individual habitat oxygenation [21,22]. While the absolute magnitude of oxygen increase at the GOE transition was not particularly great, it was biogeochemically significant, and represents a permanent shift to a world where atmospheric oxygen levels are above 10^{-5} present atmospheric levels (PAL) [21,55]. After the GOE, the oxygen proxy record provides broad upper and lower bounds, but no quantitative constraints, until the middle Paleozoic, where multiple lines of evidence indicate full ocean oxygenation [24]. Unfortunately, this leaves oxygen levels during the exact 2-billion-year time period that hosts the origin and early evolution of eukaryotes poorly constrained.

Despite the lack of consensus on Proterozoic oxygen levels, there are some general patterns that we can decipher. Sperling *et al.* performed statistical analyses of iron speciation data from 4700 samples (proxy discussed in Box 3) [56]; their data indicate that there is not a significant overall change in oxygen levels until the early mid-Paleozoic. This, however, does not mean that there were no oxygenated water columns in the Proterozoic (e.g., [57,58]), but it does indicate that

oceans were not consistently and globally oxygenated at depth until the Paleozoic. These same data also indicate that **euxinic** conditions were not pervasive after the Mesoproterozoic; in addition, and critically for our story here, surface oceans were likely oxygenated (to at least 10^{-5} PAL) throughout the entire Proterozoic.

There are no biochemical or geochemical limitations on eukaryote sterol biosynthesis in the Mesoproterozoic. Modern yeasts only require 7 nM of free oxygen to synthesize sterols and the concentration at which facultative **anaerobic** eukaryotes start to respire is about 2 μM [16]. Current estimates indicate that Mesoproterozoic surface ocean oxygen levels would roughly correspond to 2–5.5 μM, depending on temperature and other variables [59,60]. Thus, the amount of oxygen required for sterol synthesis and eukaryotic evolution was present in at least some environments by the Mesoproterozoic. Supporting this idea, molecular clock analyses of genes responsible for sterol biosynthesis predate LECA, thus the capacity to produce sterols was present in LECA itself and in coeval bacterial lineages [15]. In addition, another major **synapomorphy** for crown group eukaryotes, the mitochondrion, also requires oxygen, further indicating that the most parsimonious hypothesis for the origin of this lineage was in a partially oxygenated world. All available lines of evidence indicate that total group eukaryotes evolved after the GOE, so the fact that oxygen played a role in the early evolution of eukaryotes should not come as a surprise but instead as a logical consequence of the coevolution of Earth and life.

However, Porter presents an alternative explanation: crown group eukaryotes might not have evolved sterols until very late, and thus LECA itself evolved very late, and/or stem group eukaryotes were exclusively living in **anoxic** environments, utilizing anoxygenic metabolisms, and not producing diagnostic sterols [29]. Though these ideas are difficult to test, phylogenetic arguments make it unlikely that LECA evolved only in the latest Mesoproterozoic/earliest Neoproterozoic because, as noted earlier, crown group Archaeplastida already existed by this time. Given the evidence presented here, it is most parsimonious to suggest that the eukaryotic capability to produce or use sterols evolved in the Mesoproterozoic and that the sterane gap is thus a preservation and/or detection issue.

Concluding remarks: what drove eukaryotic diversification in the Proterozoic?

If oxygen did not change significantly during the Proterozoic, then what other factors may have led to eukaryotic evolution, and emergence and diversification of crown groups, in the Neoproterozoic Era? Some call on intrinsic ecological changes, such as the rise of eukaryvory and predation, as drivers for Neoproterozoic diversification [49,50]. Others have called on changes in nutrient fluxes, subtle changes in oxygenation levels that impacted nutrient levels, and even the breakup of the supercontinent Rodinia [61,62] (see Outstanding questions).

Whatever the drivers, it is clear that total group eukaryotes are present early in the Proterozoic but do not leave a taxonomically definitive and biodiverse record until its end. Our analysis shows that sometimes biomarkers, molecular clocks, and fossils agree and sometimes they do not. Yet it is only by delving into these apparent disagreements that we can refine our interpretations of the many proxies used in studying the early evolutionary history of eukaryotes. Further integration and synthesis will lead towards a more nuanced understanding of the early history of this overwhelmingly successful clade of life.

Acknowledgments

This work was improved by conversations with Christopher Junium, Heda Agić, Susannah Porter, and Dave Johnston. P.C. acknowledges the NCFDD Faculty Success Bootcamp and funding from NSF EAR 1855014. This paper was written on the unceded lands of the Abenaki and Coast Salish peoples. R.K. and P.C. dedicate this manuscript to Mary Kodner.

Outstanding questions

When did FECA and, subsequently, crown group eukaryotes evolve? The best evidence we have for crown group eukaryotes are photosynthetic clades, which are more derived than the most basal eukaryotic lineages. Could basal heterotrophic crown group lineages be preserved in the geologic record? These ideas can be tested in part by recalibrating clocks using updated geochronology and collecting and incorporating new well-dated fossils. The transition between FECA and crown group eukaryotes is an important evolutionary process that is likely rooted in major environmental transitions.

What external environmental and/or paleoecological events may have triggered the origin of eukaryotes and subsequent diversification within the clade? There is compelling evidence for connections between abiotic factors and biological events recorded in the geologic record, however, it is difficult to infer causation between external environmental and evolutionary events. This can be explored via continued collection and integration of phylogenetic, geochemical, and fossil data.

Can we confirm that the sterane record is biased/incomplete? The mat seal effect, and/or low eukaryotic biomass, could mean that the sterane signal is masked by bacterial biolipids and thus not detected for much of the Proterozoic. This can be tested by further experimental work on sterol preservation coupled with increased biomarker sampling of Proterozoic strata.

Declaration of interests

No interests are declared.

References

1. Penny, D. *et al.* (2014) The relative ages of eukaryotes and akaryotes. *J. Mol. Evol.* 79, 228–239
2. Agić, H. and Cohen, P.A. (2021) Non-pollen palynomorphs in deep time: unravelling the evolution of early eukaryotes. *Geol. Soc. Spec. Publ.* 511, 321
3. Javaux, E.J. and Knoll, A.H. (2016) Micropaleontology of the lower Mesoproterozoic Roper Group, Australia, and implications for early eukaryotic evolution. *J. Paleontol.* 91, 199–229
4. Agić, H. *et al.* (2017) Diversity of organic-walled microfossils from the early Mesoproterozoic Ruyang Group, North China Craton–a window into the early eukaryote evolution. *Precambrian Res.* 297, 101–130
5. Butterfield, N.J. *et al.* (1994) Paleobiology of the Neoproterozoic Svanbergfjellet Formation, Spitsbergen. *Lethaia* Vol. 27
6. Porter, S.M. *et al.* (2003) Vase-shaped microfossils from the Neoproterozoic Chuar Group, Grand Canyon: a classification guided by modern testate amoebae. *J. Paleontol.* 77, 409–429
7. Moore, K.R. *et al.* (2017) Biologically agglutinated eukaryotic microfossil from Cryogenian cap carbonates. *Geobiology* 15, 499–515
8. Cohen, P.A. *et al.* (2017) Controlled hydroxyapatite biomineralization in an~ 810 million-year-old unicellular eukaryote. *Sci. Adv.* 3, e1700095
9. Burki, F. *et al.* (2020) The new tree of eukaryotes. *Trends Ecol. Evol.* 35, 43–55
10. Kumar, S. (2005) Molecular clocks: four decades of evolution. *Nat. Rev. Genet.* 6, 654–662
11. Tiley, G.P. *et al.* (2020) Molecular clocks without rocks: new solutions for old problems. *Trends Genet.* 36, 845–856
12. Husson, J.M. and Peters, S.E. (2018) Nature of the sedimentary rock record and its implications for Earth system evolution. *Emerg. Topics Life Sci.* 2, 125–136
13. Bouwknegt, J. *et al.* (2021) A squalene–hopene cyclase in *Schizosaccharomyces japonicus* represents a eukaryotic adaptation to sterol-limited anaerobic environments. *Proc. Natl. Acad. Sci. U. S. A.* 118, e2105225118
14. Wei, J.H. *et al.* (2016) Sterol synthesis in diverse bacteria. *Front. Microbiol.* 7, 990
15. Gold, D.A. *et al.* (2017) Paleoproterozoic sterol biosynthesis and the rise of oxygen. *Nature* 543, 420–423
16. Waldbauer, J.R. *et al.* (2011) Microaerobic steroid biosynthesis and the molecular fossil record of Archean life. *Proc. Natl. Acad. Sci. U. S. A.* 108, 13409–13414
17. Hoshino, Y. and Gaucher, E.A. (2021) Evolution of bacterial steroid biosynthesis and its impact on eukaryogenesis. *Proc. Natl. Acad. Sci. U. S. A.* 118, e210276118
18. Sperling, E.A. *et al.* (2013) Oxygen, ecology, and the Cambrian radiation of animals. *Proc. Natl Acad. Sci. U. S. A.* 110, 13446–13451
19. Diamond, C.W. and Lyons, T.W. (2018) Mid-Proterozoic redox evolution and the possibility of transient oxygenation events. *Emerg. Topics Life Sci.* 41 ETLS20170146–11
20. Scott, C. *et al.* (2008) Tracing the stepwise oxygenation of the Proterozoic ocean. *Nature* 452, 456–459
21. Poulton, S.W. *et al.* (2021) A 200-million-year delay in permanent atmospheric oxygenation. *Nature* 592, 232–236
22. Chi Fru, E. *et al.* (2016) Cu isotopes in marine black shales record the Great Oxidation Event. *Proc. Natl. Acad. Sci. U. S. A.* 113, 4941–4946
23. Lyons, T.W. *et al.* (2014) The rise of oxygen in Earth's early ocean and atmosphere. *Nature* 506, 307–315
24. Dahl, T.W. *et al.* (2010) Devonian rise in atmospheric oxygen correlated to the radiations of terrestrial plants and large predatory fish. *Proc. Natl. Acad. Sci. U. S. A.* 107, 17911–17915
25. Wallace, M.W. *et al.* (2017) Oxygenation history of the Neoproterozoic to early Phanerozoic and the rise of land plants. *Earth Planet. Sci. Lett.* 466, 12–19
26. Lamb, D.M. *et al.* (2009) Evidence for eukaryotic diversification in the ~1800 million-year-old Changzhougou Formation, North China. *Precambrian Res.* 173, 93–104
27. Betts, H.C. *et al.* (2018) Integrated genomic and fossil evidence illuminates life's early evolution and eukaryote origin. *Nat. Ecol. Evol.* 17 71–11
28. Eme, L. *et al.* (2014) On the age of eukaryotes: evaluating evidence from fossils and molecular clocks. *Cold Spring Harb. Perspect. Biol.* 6, a016139
29. Porter, S.M. (2020) Insights into eukaryogenesis from the fossil record. *Interface Focus* 10, 20190105
30. Butterfield, N.J. (2011) *Bangiomorpha pubescens* n. gen., n. sp.: implications for the evolution of sex, multicellularity, and the Mesoproterozoic/Neoproterozoic radiation of eukaryotes. *J. Paleontol.* 26, 1–20
31. Gibson, T.M. *et al.* (2017) Precise age of *Bangiomorpha pubescens* dates the origin of eukaryotic photosynthesis. *Geology* 46, 135–138
32. Tang, Q. *et al.* (2020) A one-billion-year-old multicellular chlorophyte. *Nat. Ecol. Evol.* 4, 543–549
33. Maloney, K.M. *et al.* (2021) New multicellular marine macroalgae from the early Tonian of northwestern Canada. *Geology* 49, 743–747
34. Butterfield, N.J. (2014) Early evolution of the Eukaryota. *Palaeontology* 58, 5–17
35. Brocks, J.J. (2018) The transition from a cyanobacterial to algal world and the emergence of animals. *Emerg. Topics Life Sci.* 2, 181–190
36. Brocks, J.J. *et al.* (2017) The rise of algae in Cryogenian oceans and the emergence of animals. *Nature* 506, 307–318
37. Kodner, R.B. *et al.* (2008) Sterols in red and green algae: quantification, phylogeny, and relevance for the interpretation of geologic steranes. *Geobiology* 6, 411–420
38. Porter, S.M. and Knoll, A.H. (2000) Testate amoebae in the Neoproterozoic Era: evidence from vase-shaped microfossils in the Chuar Group, Grand Canyon. *Paleobiology* 26, 360–385
39. Love, G.D. *et al.* (2008) Fossil steroids record the appearance of Demospongiae during the Cryogenian period. *Nature* 457, 718–721
40. Cohen, P.A. and Macdonald, F.A. (2015) The Proterozoic record of eukaryotes. *Paleobiology* 41, 610–632
41. Rooney, A.D. *et al.* (2015) A Cryogenian chronology: two long-lasting synchronous Neoproterozoic glaciations. *Geology* 48, 459–462
42. Cohen, P.A. *et al.* (2009) Large spinose microfossils in Ediacaran rocks as resting stages of early animals. *Proc. Natl. Acad. Sci. U. S. A.* 106, 6519–6524
43. Bykova, N. *et al.* (2020) Seaweeds through time: morphological and ecological analysis of Proterozoic and early Paleozoic benthic macroalgae. *Precambrian Res.* 350, 105875
44. Droser, M.L. *et al.* (2017) The rise of animals in a changing environment: global ecological innovation in the late Ediacaran. *Annu. Rev. Earth Planet. Sci.* 45, 593–617
45. Bobrovskiy, I. *et al.* (2018) Ancient steroids establish the Ediacaran fossil *Dickinsonia* as one of the earliest animals. *Science* 361, 1246–1249
46. Wood, R. *et al.* (2019) Integrated records of environmental change and evolution challenge the Cambrian explosion. *Nat. Ecol. Evol.* 3, 1–11
47. Erwin, D.H. *et al.* (2011) The Cambrian conundrum: early divergence and later ecological success in the early history of animals. *Science* 334, 1091–1097
48. Loron, C.C. *et al.* (2018) Implications of selective predation on the macroevolution of eukaryotes: evidence from Arctic Canada. *Emerg. Topics Life Sci.* 2, 247–255
49. Porter, S.M. (2016) Tiny vampires in ancient seas: evidence for predation via perforation in fossils from the 780–740 million-year-old Chuar Group, Grand Canyon, USA. *Proc. R. Soc. B Biol. Sci.* 283, 20160221–20160226
50. Cohen, P.A. and Riedman, L.A. (2018) It's a protist-eat-protist world: recalcitrance, predation, and evolution in the Tonian-Cryogenian ocean. *Emerg. Top Life Sci* 2, 173–180

51. Porter, S. (2011) The rise of predators. *Geology* 39, 607–608
52. Pawlowska, M.M. *et al.* (2013) Lipid taphonomy in the Proterozoic and the effect of microbial mats on biomarker preservation. *Geology* 41, 103–106
53. Shen, Y. *et al.* (2020) Sterol preservation in hypersaline microbial mats. *Biogeosciences* 17, 649–666
54. Zhang, S. *et al.* (2021) Eukaryotic red and green algae populated the tropical ocean 1400 million years ago. *Precambrian Res.* 357, 106166
55. Farquhar, J. *et al.* (2000) Atmospheric influence of Earth's earliest sulfur cycle. *Science* 289, 756–759
56. Sperling, E.A. *et al.* (2015) Statistical analysis of iron geochemical data suggests limited late Proterozoic oxygenation. *Nature* 523, 451–454
57. Slotznick, S.P. *et al.* (2019) Mid-Proterozoic ferruginous conditions reflect postdepositional processes. *Geophys. Res. Lett.* 46, 3114–3123
58. Sperling, E.A. *et al.* (2014) Redox heterogeneity of subsurface waters in the Mesoproterozoic ocean. *Geobiology* 12, 373–386
59. Zhang, S. *et al.* (2016) Sufficient oxygen for animal respiration 1,400 million years ago. *Proc. Natl. Acad. Sci. U. S. A.* 113, 1731–1736
60. Kipp, M.A. *et al.* (2017) Selenium isotopes record extensive marine suboxia during the Great Oxidation Event. *Proc. Natl. Acad. Sci. U. S. A.* 114, 875–880
61. Guilbaud, R. *et al.* (2015) A global transition to ferruginous conditions in the early Neoproterozoic oceans. *Nat. Geosci.* 8, 466–470
62. Reinhard, C.T. *et al.* (2020) The impact of marine nutrient abundance on early eukaryotic ecosystems. *Geobiology* 18, 139–151
63. Willman, S. and Moczydłowska, M. (2008) Ediacaran acritarch biota from the Giles 1 drillhole, Officer Basin, Australia, and its potential for biostratigraphic correlation. *Precambrian Res.* 162, 498–530
64. Eme, L. *et al.* (2017) Archaea and the origin of eukaryotes. *Nat. Rev. Microbiol.* 15, 711–723
65. Müller, M. *et al.* (2012) Biochemistry and evolution of anaerobic energy metabolism in eukaryotes. *Microbiol. Mol. Biol. Rev.* 76, 444–495
66. Williams, T.A. *et al.* (2013) An archaeal origin of eukaryotes supports only two primary domains of life. *Nature* 504, 231–236
67. Raymann, K. *et al.* (2015) The two-domain tree of life is linked to a new root for the Archaea. *Proc. Nat. Acad. Sci. U. S. A.* 112, 6670–6675
68. Zaremba-Niedzwiedzka, K. *et al.* (2017) Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* 541, 353–358
69. Imachi, H. *et al.* (2020) Isolation of an archaeon at the prokaryote–eukaryote interface. *Nature* 577, 519–525
70. López-García, P. and Moreira, D. (2020) The syntrophy hypothesis for the origin of eukaryotes revisited. *Nat. Microbiol.* 5, 655–667
71. Mentel, M. and Martin, W. (2008) Energy metabolism among eukaryotic anaerobes in light of Proterozoic ocean chemistry. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 363, 2717–2729
72. López-García, P. *et al.* (2017) Symbiosis in eukaryotic evolution. *J. Theor. Biol.* 434, 20–33
73. Delwiche, C.F. (1999) Tracing the thread of plastid diversity through the tapestry of life. *Am. Nat.* 154, S164–S177
74. Gaines, S.M. *et al.* (2009) *Echoes of Life: What Fossil Molecules Reveal about Earth History*, Oxford University Press
75. Brocks, J.J. (1999) Archean molecular fossils and the early rise of eukaryotes. *Science* 285, 1033–1036
76. French, K.L. *et al.* (2015) Reappraisal of hydrocarbon biomarkers in Archean rocks. *Proc. Natl. Acad. Sci. U. S. A.* 112, 5915–5920
77. Zumberge, J.A. *et al.* (2019) Free and kerogen-bound biomarkers from late Tonian sedimentary rocks record abundant eukaryotes in mid-Neoproterozoic marine communities. *Geobiology* 115 246–222
78. Gold, D.A. *et al.* (2016) Sterol and genomic analyses validate the sponge biomarker hypothesis. *Proc. Natl. Acad. Sci. U. S. A.* 113, 2684–2689
79. Bobrovskiy, I. *et al.* (2020) Algal origin of sponge sterane biomarkers negates the oldest evidence for animals in the rock record. *Nat. Ecol. Evol.* 5, 165–168
80. Antcliffe, J.B. (2013) Questioning the evidence of organic compounds called sponge biomarkers. *Palaeontology* 56, 917–925
81. Nettersheim, B.J. *et al.* (2019) Putative sponge biomarkers in unicellular Rhizaria question an early rise of animals. *Nat. Ecol. Evol.* 3, 577–581
82. Kodner, R.B. *et al.* (2008) Sterols in a unicellular relative of the metazoans. *Proc. Natl. Acad. Sci. U. S. A.* 105, 9897–9902
83. Poulton, S.W. (2021) The iron speciation paleoredox proxy. In *Elements in Geochemical Tracers in Earth System Science*, Cambridge University Press
84. Gordon, G.W. *et al.* (2009) When do black shales tell molybdenum isotope tales? *Geology* 37, 535–538
85. Robbins, L.J. *et al.* (2016) Trace elements at the intersection of marine biological and geochemical evolution. *Earth-Sci. Rev.* 163, 323–348
86. Algeo, T.J. (2004) Can marine anoxic events draw down the trace element inventory of seawater? *Geology* 32, 1057–1060