Evolving Perspective on the Origin and Diversification of Cellular Life and the Virosphere

Anja Spang[1,2,*], Tara A. Mahendrarajah[1,*,†], Pierre Offre[1,*,†], and Courtney W. Stairs[3,*,†]

[1]Department of Marine Microbiology and Biogeochemistry, NIOZ, Royal Netherlands Institute for Sea Research, Utrecht University, Den Burg, The Netherlands

[2]Department of Cell and Molecular Biology, Science for Life Laboratory, Uppsala University, Uppsala, Sweden

[3]Department of Biology, Microbiology research group, Lund University, Lund, Sweden

[†]These authors contributed equally to this work.

*Corresponding authors: E-mails: anja.spang@nioz.nl; tara.mahendrarajah@nioz.nl; pierre.offre@nioz.nl; courtney.stairs@biol.lu.se.

Accepted: February 18, 2022

Abstract

The tree of life (TOL) is a powerful framework to depict the evolutionary history of cellular organisms through time, from our microbial origins to the diversification of multicellular eukaryotes that shape the visible biosphere today. During the past decades, our perception of the TOL has fundamentally changed, in part, due to profound methodological advances, which allowed a more objective approach to studying organismal and viral diversity and led to the discovery of major new branches in the TOL as well as viral lineages. Phylogenetic and comparative genomics analyses of these data have, among others, revolutionized our understanding of the deep roots and diversity of microbial life, the origin of the eukaryotic cell, eukaryotic diversity, as well as the origin, and diversification of viruses. In this review, we provide an overview of some of the recent discoveries on the evolutionary history of cellular organisms and their viruses and discuss a variety of complementary techniques that we consider crucial for making further progress in our understanding of the TOL and its interconnection with the virosphere.

Key words: tree of life, viruses, archaea, bacteria and eukaryotes, eukaryogenesis, diversity and evolution, methodological progress.

Significance

Our review provides a timely overview of how recent methodological progress has allowed an updated view on the tree of life and its connection to the virosphere. It covers topics ranging from last universal common ancestor to last eukaryotic common ancestor and the extant diversity of prokaryotic and eukaryotic life as well as viruses. Furthermore, we summarize current developments in the field that can help to make further progress in our understanding of deep evolution in the coming years.

Introduction

All cellular life forms (organisms) on Earth can be assigned to one of the major domains—the Archaea, Bacteria, or Eukaryota (hereafter referred to as eukaryotes) (Woese and Fox 1977; Woese et al. 1990). Because all organisms have evolved from a shared last universal common ancestor (LUCA) (Weiss et al. 2016), the relationship of extant organisms is often depicted within the framework of a tree of life (TOL) (Dagan et al. 2008; Puigbò et al. 2009; Blais and Archibald 2021). Upon the discovery of the Archaea, it was assumed that the TOL comprises three distinct branches that evolved vertically since LUCA, with the Bacteria on one side of the root and Archaea and eukaryotes forming sister clades on the other side of the root (Woese et al. 1990). However, recent years have witnessed an increasing body of evidence suggesting that eukaryotes, which comprise both uni- and multicellular

representatives, have emerged through a symbiosis of an archaeon and a bacterium, that is, through the merging of two branches from within the Archaea and Bacteria, respectively (fig. 1) (Guy et al. 2014; Koonin and Yutin 2014; Martin et al. 2015; Eme et al. 2017; Lopez-Garcia and Moreira 2020). In turn, Archaea and Bacteria are often referred to as primary domains of life while eukaryotes form a secondary domain of life (Williams et al. 2013, 2020). In contrast, viruses are noncellular obligate intracellular parasites that infect all cellular life forms (Koonin and Starokadomskyy 2016). Similar to other selfish genetic elements, viruses are generally not considered within the framework of the TOL (Moreira and Lopez-Garcia 2009), but are an integral part of the biosphere or biological realm (Koonin and Starokadomskyy 2016). They also impact genome evolution of cellular life not only through the exchange of genes with their hosts but also through host-parasite coevolution (Popa and Dagan 2011; Koonin 2016). In fact, the prevalence of horizontal gene transfer (HGT) via both mobile genetic elements (MGEs) and viruses but also directly between distinct organisms has to some extent questioned the concept of a TOL, which may be more correctly represented as a network including both vertical and horizontal branches (Doolittle and Bapteste 2007; Dagan et al. 2008; Puigbò et al. 2009). Yet, despite this component of horizontal genome evolution, the "statistical" TOL has remained a useful concept for understanding life's diversification (Koonin 2015b; Blais and Archibald 2021).

Recently, the application of cultivation-independent metagenomic and single-cell genomic techniques has improved our knowledge of microbial and viral diversity and, in turn, our view of the TOL (Hug et al. 2016) and its connection to the virosphere (Krupovic et al. 2020). For example, during the past decade a plethora of previously unknown archaeal and bacterial taxa (e.g., reviewed in Adam et al. [2017]; Spang et al. [2017]; Castelle and Banfield [2018]) have been described, including various lineages of high-taxonomic rank at the phylum and class-level (Hug et al. 2016; Parks et al. 2018; Rinke et al. 2021). Furthermore, progress has been made with regard to our understanding of the origin of eukaryotes (Eme et al. 2017) as well as their subsequent

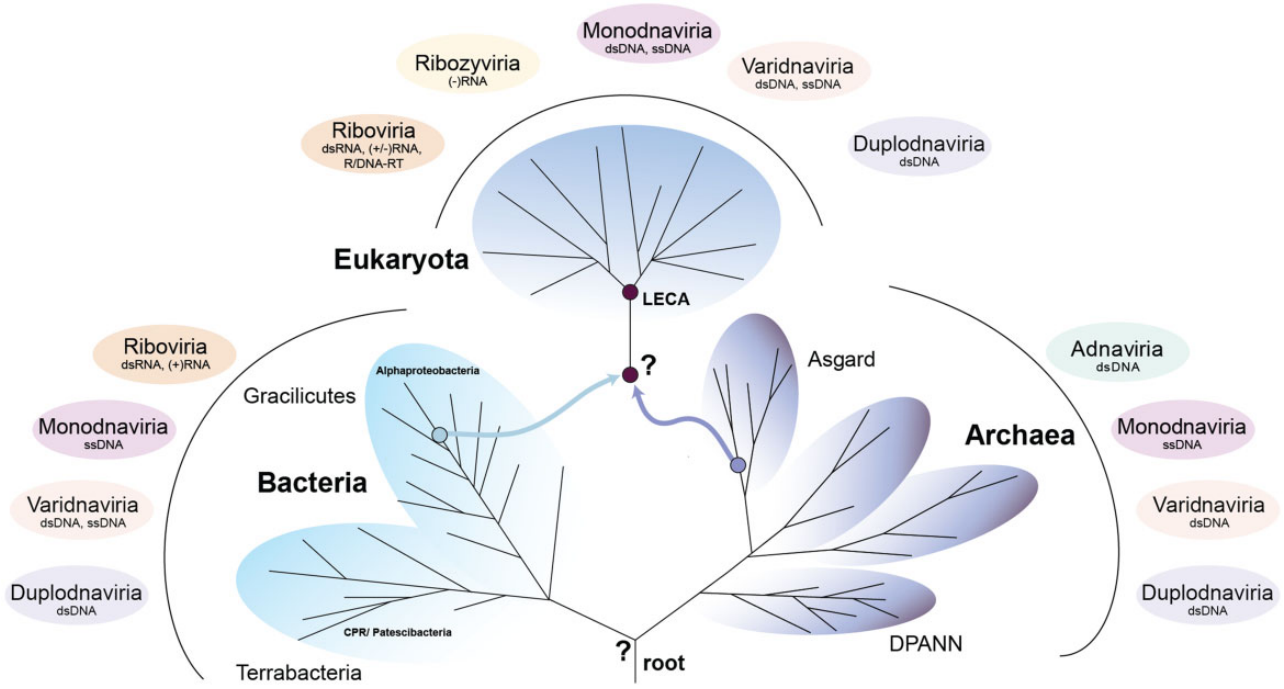

Fig. 1.—Tree of cellular life (TOL) and connection to the six major realms of viruses The tree is a schematic representation of the relationship of the major domains of life, comprised of the primary domains of Archaea and Bacteria and the secondary domain of Eukaryota. The assumption that Archaea and Bacteria form separate domains of life is dependent on the placement of the root between those domains, though this hypothesis remains to be validated. Although the node separating the DPANN (acronym referenced in text) from all other archaeal clades has been suggested to be the most ancestral split on the archaeal branch, the CPR (acronym referenced in text) most likely represents a more recently evolved sister-clade of the Chloroflexota (Coleman et al. 2021). Current data support an origin of the eukaryotic cell through a symbiosis between an ancestral member of the Asgard archaea (also Asgardarchaeota) (purple arrow) and Alphaproteobacteria (blue arrow), though the timing of the mitochondrial acquisition is debated and the events leading to LECA are poorly resolved. On the outside of the TOL, we illustrate the connection of the three cellular domains with virus representatives belonging to either of the six major viral realms, the Riboviria, Monodnaviria, Varidnaviria, Duplodnaviria, Adnaviria, and Ribozyviria (Krupovic et al. 2020; Koonin et al. 2021). The latter two realms are restricted to the Archaea or eukaryotes, respectively. The Riboviria have so far only been found associated with Bacteria and eukaryotes, whereas all other realms include members infecting cellular organisms across the TOL. LECA, last eukaryotic common ancestor.

diversification (Burki et al. 2020). Genomics approaches have also transformed our knowledge on the vast diversity of viruses (Paez-Espino et al. 2016; Martinez-Hernandez et al. 2017; Gregory et al. 2019; Beaulaurier et al. 2020; Moniruzzaman, Martinez-Gutierrez et al. 2020; Bellas and Sommaruga 2021; Edgar et al. 2022), their putative host taxa (Roux et al. 2015; Dzunkova et al. 2019; Jarett et al. 2020; Sakowski et al. 2021), and origins (Krupovic et al. 2019).

In the following, we will provide an updated perspective of the TOL and virosphere by focusing on selected key findings. Furthermore, we describe a variety of research approaches, which we consider important for making further progress on our understanding of the history of life on Earth.

The Primary Domains of Life and Deep Roots of the TOL

The nature of LUCA and the emergence of the two primary domains of life are some of the most fundamental unknowns in our understanding of life's evolution. Archaeal and bacterial cells are distinguished by major differences in their cell lipid membrane and use of contrasting molecular machinery, including for the replication, and processing of genetic information. Although a wide variety of hypotheses have been proposed to explain the distinct cell membranes of bacteria and archaea and the early evolution of their metabolism, these remain controversial and progress has been constrained by the limited availability of relevant data (Schoepp-Cothenet et al. 2013; Sousa et al. 2013; Sojo et al. 2014; Russell and Nitschke 2017). It is generally assumed that the root in the TOL separates Archaea and Bacteria as inferred based on the use of ancient paralogous gene families for rooting (Iwabe et al. 1989; Brown and Doolittle 1995; Zhaxybayeva et al. 2005; Weinheimer and Aylward 2020) and genome networks (Dagan et al. 2010) (fig. 1). Yet, the accurate placement of the root is challenging and prone to phylogenetic artifacts and alternative roots, such as within Bacteria (Cavalier-Smith 2006; Lake et al. 2009), have not been formally ruled out (Gouy et al. 2015). Further, it has recently been suggested that the branch separating the primary domains of life may be shorter than in previous estimates (Zhu et al. 2019). However, it was subsequently shown that the reduced estimate of the Archaea/Bacteria branch length most likely results from inter-domain gene transfers and, in agreement with earlier work (Koonin 2015b; Hug et al. 2016), that the longest branch in the TOL lies between Archaea and Bacteria (Martinez-Gutierrez and Aylward 2021; Moody et al. 2022) (note that these analyses did not consider extremely fast-evolving symbionts and parasites). Improved phylogenetic models, the integration of genomic data from the diversity of recently discovered taxa as well as the use of novel approaches for rooting, such as gene tree-species tree reconciliations, for example, Szöllősi et al. (2012), David and Alm (2011), and Szöllősi et al. (2013) (see below), will help to determine whether this branch indeed represents the deepest split in the TOL.

Particularly, the discovery of two previously unknown and potentially deep-branching microbial radiations in the Bacteria and Archaea, the so-called DPANN archaea (Rinke et al. 2013; Castelle et al. 2015) and the bacterial Candidate Phyla Radiation (CPR or Patescibacteria) (Brown et al. 2015), respectively, has provided important data for readdressing the deep roots of microbial life and the placement of the archaeal and bacterial roots (Williams et al. 2017; Castelle et al. 2018; Taib et al. 2020; Coleman et al. 2021; Xavier et al. 2021). The DPANN group (acronym referring to its first described member lineages, the Diapherotrites, Parv-, Aenigm-, Nano-, and Nanohaloarchaeota) now includes more than eight distinct archaeal phyla (Rinke et al. 2021) that group together with Nanoarchaeota, an archaeal clade represented by the ultrasmall and ectosymbiotic archaeon *Nanoarchaeum equitans* (Huber et al. 2002). Representatives of DPANN have small genomes and cell sizes, are characterized by restricted anabolic and catabolic capabilities, and include obligate ectosymbionts some of which have been cultivated in coculture with their hosts belonging to the Halobacteriota, Thermoproteota, and Thermoplasmatota (Huber et al. 2002; Podar et al. 2013; Munson-McGee et al. 2015; Wurch et al. 2016; Golyshina et al. 2017; Krause et al. 2017; Hamm et al. 2019; St John et al. 2019; La Cono et al. 2020; Sakai et al. 2022). Indeed, symbiotic lifestyles have been suggested to represent a common feature of genome-reduced members of the DPANN (Castelle et al. 2018). Likewise, members of the CPR, which also include various lineages of high taxonomic rank, share several genomic features with the DPANN archaea, such as small cell and genome sizes, a limited metabolic potential and potential dependency on partner organisms (Castelle et al. 2018). In line with this, two representatives of this group, that is, members of the Saccharibacteria and Absconditabacteria, have been successfully enriched as symbionts in coculture with their respective actinobacterial and gammaproteobacterial hosts (He et al. 2015; Bor et al. 2018; Moreira et al. 2021). It seems that the level of host specificity differs significantly between different representatives of the DPANN and CPR. For instance, although the most genome-reduced members of the DPANN, such as *N. equitans*, seem unable to switch between different host strains (Jahn et al. 2008), members of the Micrarchaeota infect hosts belonging to different archaeal phyla and comprise strains that can grow in coculture with hosts belonging to different genera (Golyshina et al. 2017; Krause et al. 2017; Sakai et al. 2022). Furthermore, it seems that at least DPANN may also include free-living members such as the Altiarchaeota (Probst et al. 2014) or members, which, in spite of certain

auxotrophies, do not require permanent physical contact with potentially interacting partners (Youssef et al. 2015; Beam et al. 2020).

Initial phylogenetic analyses have recovered both the CPR (Brown et al. 2015) and DPANN (Rinke et al. 2013; Castelle et al. 2015) as monophyletic and early diverging branches in the TOL (fig. 1), but these findings are being debated (Dombrowski et al. 2019; Meheust et al. 2019). In particular, several authors have raised the concern, that the deep and monophyletic placement of DPANN and CPR lineages may be the result of phylogenetic artifacts (Brochier-Armanet et al. 2011; Petitjean et al. 2014; Raymann et al. 2014; Aouad et al. 2018; Feng et al. 2021) such as long-branch attraction, that leads to the erroneous grouping of fast-evolving taxa in a monophyletic clade as well as their attraction to a distant outgroup (Bergsten 2005; Philippe et al. 2005). For example, previous studies have revealed that genomes of other symbionts (e.g., obligate intracellular bacterial endosymbionts) indeed experience faster evolutionary rates, have compositional biases and form long branches in phylogenetic trees (Moran 1996; Rodriguez-Brito et al. 2006). In turn, elucidating the phylogenetic placement of the symbiotic CPR and DPANN has proven challenging and requires careful phylogenetic approaches implementing, among others, careful marker gene and taxon selection approaches and/or the use of complex models of evolution that account for differences in evolutionary rates across sites and lineages (Dombrowski et al. 2020; Coleman et al. 2021; Martinez-Gutierrez and Aylward 2021). Furthermore, such analyses benefit from taking into account potentially increased rates of HGT between symbionts and their hosts (Dombrowski et al. 2020).

Recently, outgroup-free rooting methods have been applied to assess the placement of CPR and DPANN in the TOL. For instance, Coleman et al. (2021) have used a gene tree—species tree reconciliation approach (Szöllősi et al. 2012; David and Alm 2011; Szöllősi et al. 2013) to root the bacterial tree and reconstruct the proteome of the last bacterial common ancestor. Interestingly, and in contrast to several earlier studies, this has revealed that the CPR most likely represents a more recently evolved monophyletic sister-lineage of the Chloroflexota (Coleman et al. 2021) rather than an early diverged bacterial clade (Brown et al. 2015) (fig. 1). Thus, CPR members seem to be derived from more complex ancestors with their small genomes being a result of genome-streamlining processes (Coleman et al. 2021). In agreement with this, a recent analysis aiming to resolve the evolution of cell envelopes in Bacteria not only indicated the ancestry of didermy with several independent transitions to monoderm phenotypes but also supported a sisterhood relationship of Chloroflexota and CPR nested within Terrabacteria (Taib et al. 2020). Finally, the careful assessment of marker genes for multidomain phylogenies has further confirmed this derived placement of the CPR (Martinez-Gutierrez and Aylward 2021).

In contrast, several recent studies have provided support for the "clanhood" of DPANN in unrooted phylogenies, their characteristic set of genes and their placement as an early radiation on the archaeal branch of the TOL raising the possibility that DPANN clades may have evolved in parallel with their host lineages over much of evolutionary time, see for example, Williams et al. (2017), Dombrowski et al. (2020), Castelle et al. (2021), Martinez-Gutierrez and Aylward (2021), and Aouad et al. (2022) (fig. 1). However, conflicting results regarding the placement of certain putative DPANN clades remain (Feng et al. 2021). Furthermore, it is important to note that the exact placement of the root in the archaeal tree is not yet fully resolved and could be located between two distinct DPANN clades, thus leaving open the possibility that DPANN are paraphyletic (Dombrowski et al. 2020; Aouad et al. 2022). Further analyses, such as the application of gene tree-species tree reconciliations applied to a larger set of representative archaeal genomes will help to test current hypotheses on the early divergence of DPANN. Finally, a reliable interpretation of the early evolution of cellular life, the features of the last universal common ancestor, and the relationship of DPANN and CPR, hinges on the accurate placement of the universal root (Gouy et al. 2015).

Origin of the Eukaryotic Cell from Prokaryotic Ancestors

The origin of the eukaryotic cell represents one of the most significant and at the same time debated events in life's evolution. Over the years, a variety of eukaryogenesis models have been put forth, which can be broadly categorized into symbiogenetic and autogenous models, discussed in several comprehensive reviews (Guy et al. 2014; Lopez-Garcia and Moreira 2015; Martin et al. 2015; Koonin 2015a). Although autogenous models assume the vertical evolution of a proto-eukaryotic lineage from a root shared with the archaeal and bacterial line of descent, symbiogenetic models suggest that the origin of the eukaryotic cell is a result of a merger of members of at least two distinct microbial lineages belonging to the Archaea and Alphaproteobacteria (Roger et al. 2017; Lopez-Garcia and Moreira 2020) (fig. 1).

Recently, the genomics-based discovery of the Asgard archaea (Spang et al. 2015; Zaremba-Niedzwiedzka et al. 2017) (also referred to as the phylum Asgardarchaeota [Rinke et al. 2021]), has provided important data shedding new light on the origin of the eukaryotic cell. Asgard archaea were originally described to comprise the Loki-, Thor-, Odin, and Heimdallarchaea (Spang et al. 2015; Seitz et al. 2016; Zaremba-Niedzwiedzka et al. 2017), but are now known to include a variety of additional clades (Seitz et al. 2019; Cai et al. 2020; Farag et al. 2021; Liu et al. 2021; Zhang et al. 2021; Wu et al. 2022). Notably, phylogenetic analyses have revealed that the Asgard archaea comprise the closest archaeal sister lineage of eukaryotes (Zaremba-Niedzwiedzka

et al. 2017; Liu et al. 2021; Wu et al. 2022) and thereby provided increasing evidence for the evolution of eukaryotes from within the Archaea (Williams et al. 2013, 2020) (fig. 1). But although there is strong support for the monophyly of Asgard archaea and eukaryotes, the exact placement of the eukaryotic branch relative to the various Asgard lineages varies depending on data set composition and evolutionary models used (Zaremba-Niedzwiedzka et al. 2017; Williams et al. 2020; Liu et al. 2021). Expanded sampling of Asgard diversity combined with careful phylogenetic analyses, is likely to provide improved resolution of branching orders and will allow to pinpoint the closest sister-lineage of eukaryotes more precisely.

In agreement with phylogenetic evidence, comparative analyses of the Asgard archaeal genomes have revealed the presence of so-called eukaryotic signature proteins (ESPs) (reviewed in Hartman and Fedorov [2002]; Eme et al. [2017]; Spang et al. [2017]), that is, proteins that were previously thought to be absent from prokaryotic genomes. Notably, these ESPs are homologous to proteins integral to the functioning of complex eukaryotic cells and comprise essential building blocks of the ESCRT (endosomal sorting complex required for transport) system, ubiquitin, trafficking, and informational processing machineries as well as the cytoskeleton (Spang et al. 2015; Zaremba-Niedzwiedzka et al. 2017; Liu et al. 2021). Although the function of these proteins in Asgard archaea remains to be elucidated, the heterologous expression and structural analyses of some of these proteins such as profilins and gelsolins have revealed that they are functionally equivalent to their eukaryotic homologs and suggests that a regulated actin cytoskeleton precedes eukaryogenesis (Akil and Robinson 2018; Akil et al. 2020; Survery et al. 2021).

Because even high quality metagenome assembled genomes (MAGs) (i.e., completeness >90% and contamination <5%, according to Bowers et al. [2017]) usually do not assemble into complete genomes and may contain a low amount of contamination from genomes of other community members or closely related strains, some studies have questioned the reliability of the Asgard archaeal MAGs and in particular raised concerns as to whether ESPs may represent contamination rather than being genuine genomic signatures (Da Cunha et al. 2017, 2018; Garg et al. 2021). However, various lines of evidence during the past years have supported the existence of Asgard archaea, the emergence of the archaeal ancestor of eukaryotes from within this group as well as the presence of ESPs as part of their coding potential: among others, ESPs are encoded within a prokaryotic genomic context, lack introns characteristic of many eukaryotic genes, and are significantly divergent from eukaryotic homologs to exclude contamination (Spang et al. 2015; Zaremba-Niedzwiedzka et al. 2017; Spang et al. 2018). Furthermore, Asgard MAGs have now been reconstructed from a large variety of metagenomes from different environmental samples all over the world and by many different research groups, yet show consistent genomic signatures across the various member clades (Manoharan et al. 2019; Cai et al. 2020; Chen, Wong, et al. 2020; Farag et al. 2020, 2021; Liu et al. 2021; Zhang et al. 2021). Even though the presence/absence pattern of ESPs across Asgard archaea is variable and indicates a complex history of ESP evolution involving duplications, differential loss, and transfers, the shared set of ESPs within specific taxon-level (e.g., class-level) lineages is very consistent and provides strong evidence for ESPs representing genuine signatures of Asgard proteomes (Liu et al. 2021). In line with this, the successful enrichment of the first representative of the Asgard archaea, *Candidatus* Prometheoarchaeum syntrophicum has not only proven the viability of members of this group but also allowed the reconstruction of the first complete genome of a Lokiarchaeote with a characteristic and consistent set of ESPs (Imachi et al. 2020). Finally, initial microscopy analyses have provided insights into the cellular features of extant members of the Asgard archaea including cellular protrusions (Imachi et al. 2020; Avci et al. 2022) and revealed the spatial separation of genomic DNA and ribosomes in certain representatives (Avci et al. 2022).

The analysis of the genomic repertoire of the Asgard archaea has not only enabled predictions of their extant metabolic characteristics but also provided a first baseline to refine symbiogenetic eukaryogenesis models, which predict a syntrophic interaction as an important initial driver for cell–cell interactions (Spang et al. 2019; Imachi et al. 2020; Lopez-Garcia and Moreira 2020; Liu et al. 2021), and represent an extension of the Hydrogen (Martin and Muller 1998) and Syntrophy (Moreira and Lopez-Garcia 1998) Hypotheses. However, more detailed models hinge on resolving the exact placement of the eukaryotic and mitochondrial branches relative to the Asgard archaea (Zaremba-Niedzwiedzka et al. 2017; Liu et al. 2021) and Alphaproteobacteria (Roger et al. 2017; Martijn et al. 2018; Fan et al. 2020; Munoz-Gomez et al. 2022), respectively, as well as the cellular and metabolic features of these ancestors. Additionally, controversies remain with regard to the timing of the events during eukaryogenesis, that is, the timing of the mitochondrial acquisition, the evolution of an endomembrane system as well as the establishment of a nucleus, for example, Baum and Baum (2014), Poole and Gribaldo (2014), Gould et al. (2016), Pittis and Gabaldon (2016), Eme et al. (2017), Tria et al. (2021), Vosseberg et al. (2021) (fig. 1). Finally, the extent to which additional microbial lineages and/or viruses (see below) have contributed to the eukaryotic proteome are still to be determined. Phylogenomics analyses have for example provided support for the hypothesis that the genomic repertoire of eukaryotes was shaped through genetic input from Bacteria other than Alphproteobacteria (Koonin 2010; Rochette et al. 2014; Santana-Molina et al. 2020; Stairs et al. 2020; Hoshino and Gaucher 2021) as well as by viruses, for example,

Cermakian et al. (1997), Filée and Forterre (2005), Shutt and Gray (2006), and Harada and Inagaki (2021). Furthermore, a recently proposed updated symbiogenetic model on the origin of the eukaryotic cell has implicated the potential involvement of an additional bacterial lineage (i.e., a Deltaproteobacterium) during eukaryogenesis (Lopez-Garcia and Moreira 2020).

The combination of novel techniques in phylogenetics with cell biological and cultivation approaches (see below) will help to address those conflicting hypotheses of the origin of the complex eukaryotic cell from its prokaryotic ancestors and continue to illuminate the timing of the events during eukaryogenesis (Eme et al. 2017; Roger et al. 2021).

Eukaryotic Diversity and the Last Eukaryotic Common Ancestor

Even though various aspects of eukaryogenesis remain enigmatic, our knowledge of the last eukaryotic common ancestor (LECA) (reviewed in Eme et al. [2017]) and its subsequent diversification has grown substantially in recent years, enabled by a tremendous increase in our sampling of extant eukaryotic diversity. Indeed, although the majority of formally described eukaryotes are multicellular and fall into two phylogenetic groups: Archaeplastida (plants and algae) and Opisthokonta (animals and fungi), it is now clear that the bulk of phylogenetic diversity of eukaryotes is composed of unicellular representatives including "protists" and algae (fig. 2). Major advances in cultivation-dependent (Burki et al. 2020) and cultivation-independent (Burki et al. 2021) methods including symbiosis-aware strategies (Alacid and Richards 2021) for generating sequence data combined with sophisticated bioinformatic tools for genome assembly, gene annotation, and phylogenomic inference have been critical for the genomics-driven exploration of eukaryotic biodiversity. In particular, the last decade has witnessed the discovery of numerous kingdom- and phylum-level lineages and confidently placed those in the eukaryotic TOL (fig. 2), for example, Rhodelphia (Gawryluk et al. 2019), Picozoa (Schön et al. 2021), Anaeramoebae (Stairs et al. 2021), and "CruMs" (Brown et al. 2018) (Collodictyonids, Rigifilids, Mantamonads). Sequence data has also been collected from lineages that have no clear phylogenetic position including *Ancoracysta twista* (Janouskovec et al. 2017), Hemimastigophora (Lax et al. 2018), Ancyromonadida (Torruella et al. 2015), and Malawimonadida (Heiss et al. 2018) that might each represent phylum- (or higher-) level taxonomic ranks.

Supported by these new data, numerous lines of evidence suggest that LECA dated to the Proterozoic (ca. 1.9–1.6 billion years ago) (Parfrey et al. 2011; Eme et al. 2014; Betts et al. 2018) and was characterized by a nucleus and nuclear pores, linear chromosomes with telomeres, genes with spliceosomal introns, complex RNA processing, and regulatory mechanisms, an elaborate endomembrane system (including a Golgi apparatus, endosomes, lysosomes, and peroxisomes), mitochondria, bacterial-type lipids as well as a complex cell cycle (extensively reviewed in Koumandou et al. [2013] and Eme et al. [2017]). Some analyses predict that the LECA proteome was already quite complex with many orthologs (~10,000) tracing their origin to LECA (Deutekom et al. 2021), though many details regarding components of the various cellular and molecular machineries remain to be further illuminated. One current limitation lies in the unresolved placement of the root in the eukaryotic tree. Depending on

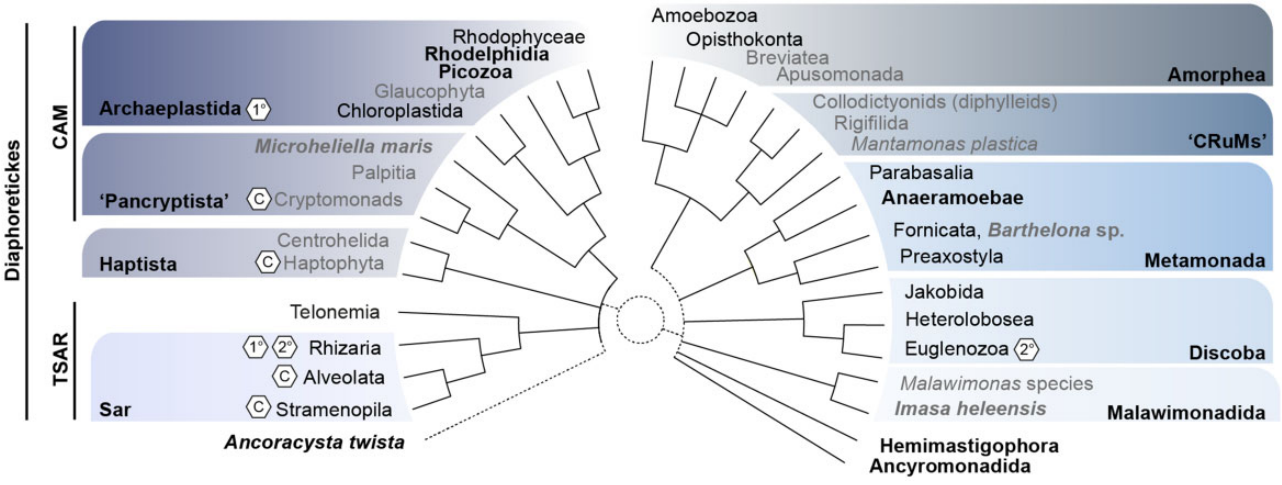

FIG. 2.—Schematic representation of the phylogenetic diversity of eukaryotes Groups with taxonomic rankings of phylum level or higher are shown in black (according to Adl et al. 2019 and references in text). Select lineages or organisms that have been recently discovered and placed in the eukaryotic TOL are shown in bold. Eukaryotic supergroups are colored for clarity. Lineages with one or more representative with a primary (1°) secondary (2°) or complex red (C) plastids are indicated with hexagons based on Sibbald and Archibald (2020). Sar, Stramenopila-Alveolata-Rhizaria; TSAR, Telonemia+SAR (Strassert et al. 2019), CAM (Yazaki et al. 2021), Cryptista-Archaeplastida-*Microheiliella maris*; "CRuMs," Collodictyonids, Rigifilida, *Mantamonas plastica*.

gene set and methodology used, the root of the eukaryotic tree has been inferred between Discoba and other eukaryotes (He et al. 2014), between Diaphoretickes + Discoba and Amorphea + CruMs + Malawimonads (Derelle et al. 2015) or between Opisthokonta and all other eukaryotes (Katz et al. 2012; Cerón-Romero et al. 2021). Therefore, the best-studied eukaryotes on which various previous LECA inferences are based, represent derived clades on either side of the putative root: the Archaeplastida within Diaphoretikes and Opisthokonta within Amorphea. It is conceivable that genes conserved in either of these lineages may not necessarily trace their origins back to LECA. For example, a recent review by More et al. (2020) put forth a new term defining hidden ancient homologs as "jotnarlogs" that are shared across eukaryotic biodiversity exclusive of the "model system" lineages. They show that these jotnarlogs are highly relevant for our understanding of the earliest steps in eukaryotic evolution and, among others, comprise proteins mediating fundamentally eukaryotic processes including mitochondrial division (Leger et al. 2015) and membrane trafficking (More et al. 2020). In turn, prospective analyses that make use of the increased sampling of eukaryotic genomic diversity will be crucial to further improve our knowledge on the nature of LECA as well as the root placement in the eukaryotic TOL.

Although most modern eukaryotes share key cellular features, the recent discovery of novel eukaryotic representatives forming distinct branches in the eukaryotic tree have revealed interesting insights into eukaryotic metabolic and cellular diversity. For example, although the alphaproteobacteria-derived mitochondria in extant aerobic eukaryotes house the respiratory chain that couples ATP biosynthesis to the reduction of oxygen, in some anaerobic animals and fungi, the respiratory chain uses alternative electron acceptors to oxygen in order to synthesize ATP, often by "tinkering" with existing cellular systems to synthesize anaerobiosis-specific cofactors or by encoding anaerobiosis-specific proteins (Müller et al. 2012; Gawryluk and Stairs 2021). Further, many anaerobic protists have lost most, if not all, respiratory capabilities and instead couple ATP biosynthesis to fermentative H_2 production within so-called mitochondria-related organelles (MROs) (Müller et al. 2012; Stairs et al. 2015; Gawryluk and Stairs 2021). Some representatives, such as *Monocercomonoides,* have lost their MROs (Karnkowska et al. 2016), and/or mitochondrial genomes (Stairs et al. 2015) entirely. The genetic origins of the anaerobic metabolism of MROs remains a widely debated topic (see, e.g., Katz 2015; Martin 2017; Leger et al. 2018; Sibbald et al. 2020; Stairs et al. 2020; Tria et al. 2021).

Photosynthesis is a widespread trait across the tree of eukaryotes with representatives in Stramenopila, Alveolata, Rhizaria, Haptista, Pancryptista, Archaeplastida, and Discoba. Primary plastids, derived from the engulfment of an ancestral photosynthetic cyanobacterium with the closest present day relative likely being *Gloeomargarita lithophora* (Ponce-Toledo et al. 2017; Moore et al. 2019), have evolved at least once on the tree of eukaryotes in the Archaeplastida (Sibbald and Archibald 2020) between 2.1 and 1.6 (Sanchez-Baracaldo et al. 2017; Strassert et al. 2021) or 1.8 and 1.1 billion years ago (Betts et al. 2018). There is at least one additional candidate of a primary photosynthetic organelle in eukaryotes in the Rhizarian *Paulinella chromatophora* (Nowack et al. 2008; Nakayama and Ishida 2009). This amoeba houses a specialized organelle called the chromatophore that has its own genome and is thought to have evolved from an ancestral endosymbiont of the *Synechococcus/Prochlorococcus* clade (Marin et al. 2005) roughly 90–140 Ma (Delaye et al. 2016). The chromatophore provides a rare opportunity to study the early stages of endosymbiosis having occurring nearly 1 billion years more recently than the primary plastids of Archaeplastida. Other eukaryotes, that is, heterotrophic protists, have acquired secondary or higher order plastids through serial endosymbiosis events, reviewed in Sibbald and Archibald (2020). These higher-order plastids are often surrounded by three or four membranes and, in at least three separate lineages, retain the nuclei (dubbed the nucleomorph) from the engulfed endosymbiotic algae (Sibbald and Archibald 2020). In these cells, there can be as many as four distinct genomes derived from the host nucleus, host mitochondrion, plastid, and nucleomorph. Continued investigations comparing the origin of the gene content and cell biology of these diverse and complex algal lineages as well as phylogenetic and molecular dating approaches will help in identifying the mechanisms necessary for enabling endosymbiosis events and help to further improve our understanding of their timing throughout eukaryotic diversification (Strassert et al. 2021).

Viruses and the Tree of Life

MGEs are semiautonomous replicative genomic entities that are ubiquitous in the natural environment and believed to be an intrinsic part of cellular evolution (Koonin et al. 2021). They include viruses which may encode one or more proteins comprising the viral particle (virion) encasing the genome of the respective MGE (Koonin et al. 2021). Categorically, viruses are believed to be the most abundant biological entities on the planet, shaping ecological and evolutionary components of the biosphere (Krupovic et al. 2019). The diverse characteristics of MGEs stratify the semiautonomous replicative genomic entities or replicator groups, blurring the boundaries between the major categories within the replicator space, with the Virosphere defined at its core by the *Orthovirosphere*, followed by the *Perivirosphere*, and the remaining replicators falling within the periphery (Koonin et al. 2021).

Recent evolutionary insight has classified the core of the virosphere, that is, the *Orthovirosphere*, into six major realms, the *Riboviria*, *Varidnaviria*, *Duplodnaviria*, *Monodnaviria*,

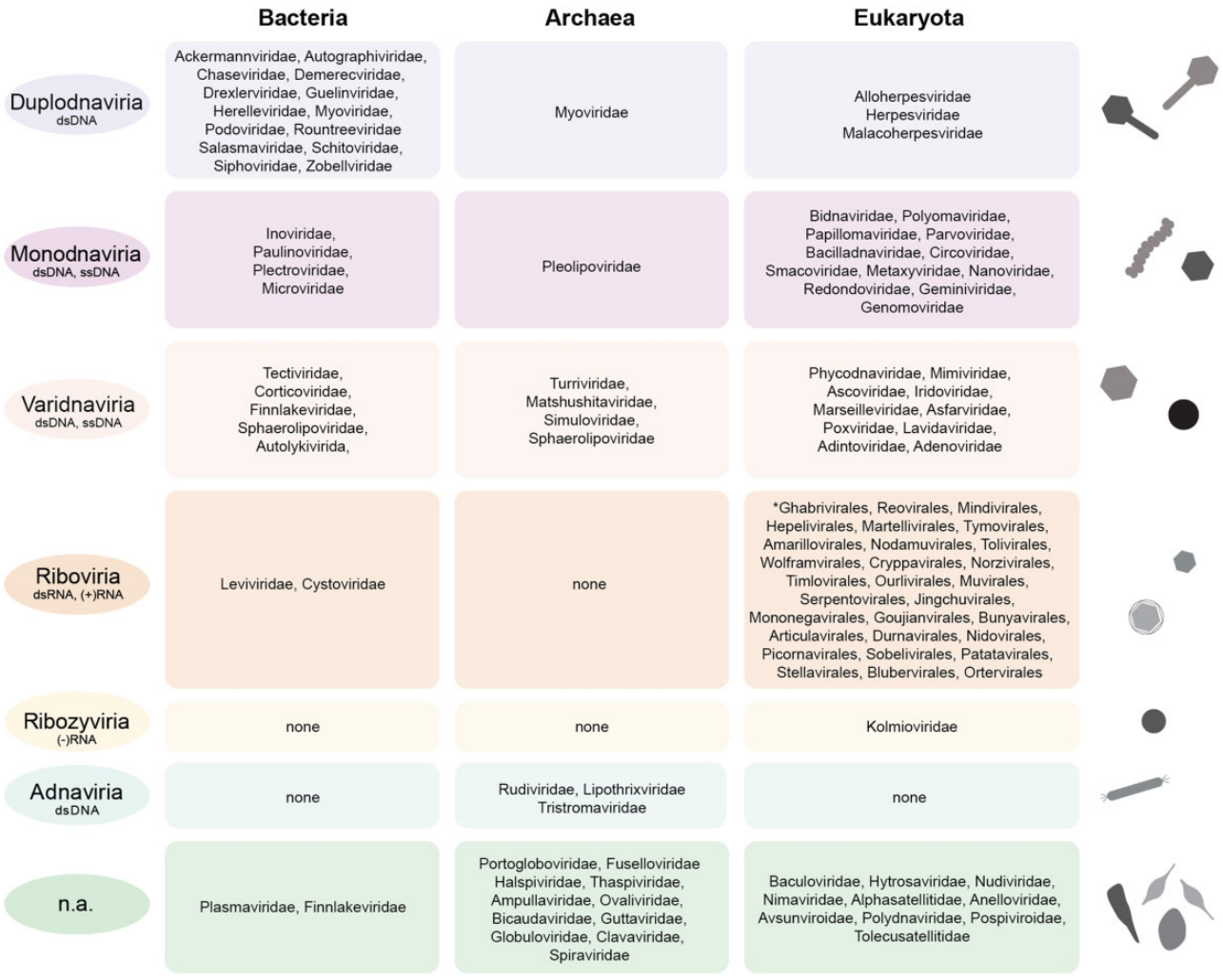

Fig. 3.—The diversity of the core virosphere and its links to bacterial, archaeal, and eukaryotic hosts For each viral realm, we depict the diversity of viral families that have representatives infecting members either the Bacteria, Archaea, or Eukaryota, respectively. Asterisk: for eukaryotic viruses assigned to the Riboviria, we report orders instead of families. The shapes represent a small selection of characteristic morphologies seen within certain viral realms. The information on viral families comprising the various realms is derived from the ICTV database (https://talk.ictvonline.org/files/master-species-lists/), that is, ICTV Master Species List 2020.v1.xlsx. (Krupovic et al. 2020; Koonin et al. 2021).

Adnaviria, and *Ribozyviria* (Koonin et al. 2021), comprising many but not all viral families (figs. 1 and 3). Apart from the *Ribozyviria*, which has been identified in specific vertebrates, all realms are believed to have emerged before or near the origination of the last universal cellular ancestor (LUCA) (Krupovic et al. 2020; Koonin et al. 2021). To fully understand the roles viruses played during the earliest stages of the evolution of cellular life, studies have sought to understand the origins of key viral components. Generally, viral genomes are unified by two core modules: a module that encodes the proteins responsible for genome replication (the replication module) and a module that encodes the proteins that form the virion particle that encapsulates the genome (the morphogenetic module) (Krupovic et al. 2019). Despite great viral diversity, most replication modules can be captured by four hallmark replication protein families: the RNA-dependent RNA polymerase, the reverse transcriptase, the protein-primed family B DNA polymerase, and the rolling-circle endonuclease (Krupovic et al. 2019). All of these share the common ancient RNA-recognition fold and importantly, have minimal to no close sequence identity with replication proteins from cellular organisms. Conversely, investigation into the origins of the capsid proteins that comprise the virion suggests descent from protein families from cellular ancestors, specifically those involved in carbohydrate- or nucleic acid binding (Krupovic et al. 2019). These findings are the foundation of the proposed chimeric model of viral evolution which describes the emergence of the replication module from the primordial replicon pool, with the morphogenetic module evolving on several different occasions

through life's history by acquisitions of structural proteins from hosts (Krupovic et al. 2019). Notably, recent structural and genomics studies into the diversity of archaeal viruses have revealed an abundance of archaea-specific viruses that share no genetic or structural similarity to bacterial and eukaryotic counterparts (Prangishvili et al. 2017; Krupovic et al. 2018) and cannot currently be assigned to any of the viral realms (fig. 3). Beyond unique morphologies across the archaeal viruses, the archaea-specific *Adnaviria* possess a morphogenetic module composed of a capsid protein with a distinct fold not captured by viruses in the other two domains (Koonin et al. 2021). These findings underscore the need for further exploration into the diversity, structure, and function of archaeal viruses.

Viruses and other MGEs are generally not considered part of the TOL (Lopez-Garcia 2012), however the nature of their replication and propagation mechanisms have linked them to critical components of cellular genome dynamics and evolution. Recent efforts have tried to connect the deep origins and diversification of viruses to the earliest transitions in the TOL and diversification of cellular life (Koonin et al. 2020; Weinheimer and Aylward 2020; Irwin et al. 2022). Parasitic replicators play important roles in host-parasite coevolutionary dynamics and the evolution of host genomes (Koonin and Krupovic 2018) and have been placed at the centre of debates regarding eukaryotic evolution and diversification (Koonin et al. 2015; Forterre 2016; Guglielmini et al. 2019; Moniruzzaman, Weinheimer et al. 2020; Collens and Katz 2021; Irwin et al. 2022). Particularly the discovery of eukaryotic NucleoCytoplasmic Large DNA viruses (NCLDVs), also referred to as giant viruses (Raoult et al. 2004), has sparked debates on the boundaries between viruses and cellular organisms as well as raised questions regarding their origins, relationship to cellular life and role in the origin of the eukaryotic cell. NCLDVs comprise members with unique features among viruses including genome sizes that resemble those of some free-living microorganisms, the presence of genes for DNA maintenance including repair, replication, transcription, and translation, complex metabolic capabilities, cytoskeleton components, as well as other signature proteins of complex eukaryotic cells, all of which were originally thought to be confined to cellular life (Schulz et al. 2017; Abrahao et al. 2018; Schvarcz and Steward 2018; Koonin and Yutin 2019; Yoshikawa et al. 2019; Da Cunha et al. 2022; Moniruzzaman, Martinez-Gutierrez et al. 2020; Kijima et al. 2021). Some representatives replicate within viral factories, that is, intracellular compartments in which viral components are localized and that may be enclosed by membranes (Novoa et al. 2005; Suzan-Monti et al. 2007), and can be parasitized by their own virophages (Krupovic et al. 2016). But although those characteristics have originally been suggested to indicate that NCLDVs may form a separate branch within the TOL (Raoult et al. 2004), careful phylogenetic analyses have subsequently shown that NCLDVs have acquired hallmark cellular genes through HGT by their hosts and evolved gigantism multiple times (Williams et al. 2011; Moreira and Lopez-Garcia 2015; Koonin and Yutin 2018; Backstrom et al. 2019), validating the distinction of viruses and cellular life (Moreira and Lopez-Garcia 2009; Lopez-Garcia 2012; Forterre et al. 2014; Koonin and Starokadomskyy 2016). Viruses and in particular NCLDVs have also been hypothesized to have played a role in the origin of the nucleus due to the ability of some representatives to assemble viral factories reminiscent of eukaryotic nuclei (Takemura 2020). However, the direct involvement of a virus in the origin of eukaryotic organellar complexity remains debated (Lopez-Garcia et al. 2017) and viral factories, including those established by certain *Pseudomonas* phages enclosed by a proteinaceous shell (Chaikeeratisak et al. 2017), likely represent analogous structures to eukaryotic nuclei. Nevertheless, viruses and/or MGEs have been found to have shaped the eukaryotic proteome early on including through virus-to-host HGT (Guglielmini et al. 2019; Irwin et al. 2022). For example, the mitochondrial single-subunit RNA polymerase (ssRNAP) has been suggested to be derived from T-odd phages (Cermakian et al. 1997; Filée and Forterre 2005; Shutt and Gray 2006) and eukaryotic telomerases, that ensure the replication of linear chromosomes, are likely derived from a Penelope-like retroelement reverse transcriptase (Koonin et al. 2015). The finding of widespread endogenization of viral genomes, including those of NCLDVs, into eukaryotic host genomes highlights a potentially important strategy underlying virus-to-host HGTs (Feschotte and Gilbert 2012; Moniruzzaman, Weinheimer et al. 2020). Thus, to further disentangle the sources of the eukaryotic proteome and cellular features, prospective phylogenetic analyses benefit from taking into account the wide diversity of viral in addition to prokaryotic genome data (Irwin et al. 2022). In this regard, it is particularly noteworthy that recent metagenomics approaches (some only available as preprints so far) have identified a suite of viruses likely infecting Asgard archaea and belonging to different viral realms (Medvedeva et al. 2021; Rambo et al. 2021; Tamarit et al. 2021; Wu et al. 2022). The genomic and experimental analysis of these and other novel viruses may help to test hypotheses on the features and impact of MGEs in the earliest transitions and diversification of eukaryotic cells.

Taken together, a better understanding of the TOL and major evolutionary transitions hinges on the continued exploration of the virosphere combined with improved phylogenomics and network analyses that allow illuminating the impact of viruses and other MGEs on cellular evolution.

How to Make Further Progress

Making further progress in our understanding of the TOL and resolving the phylogenetic placement of taxa near key

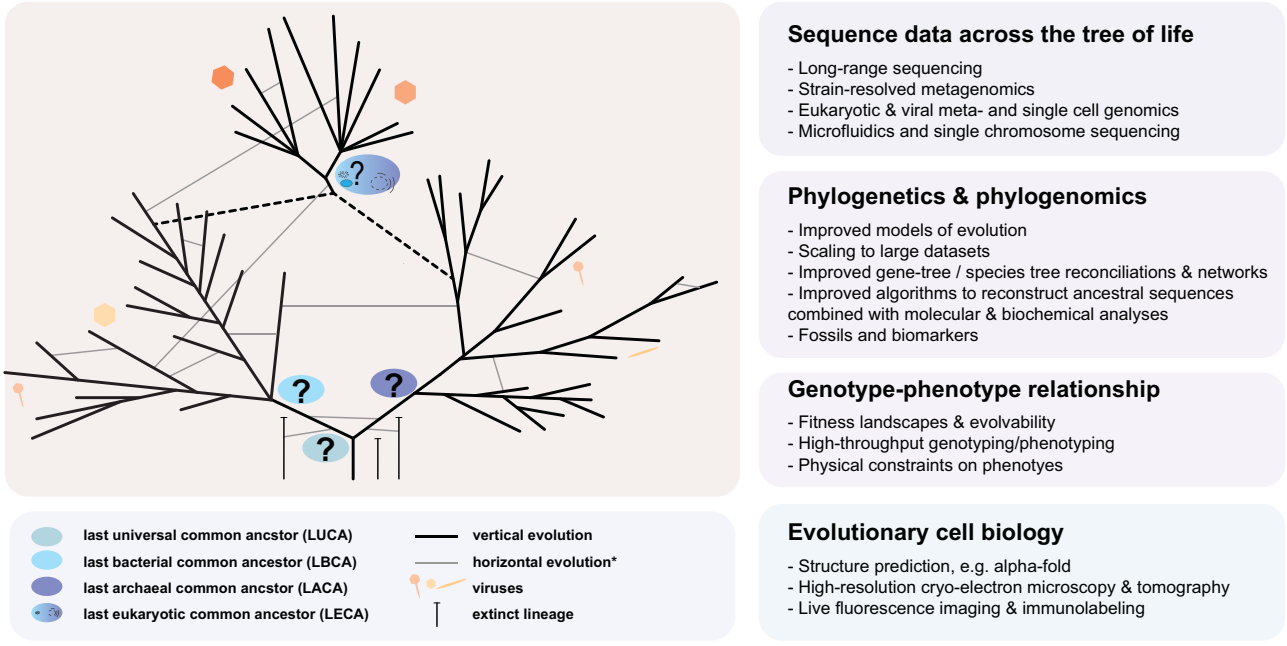

Fig. 4.—Schematic representation of TOL highlighting key questions and approaches to further illuminate cellular evolution and its connection to viral evolution See text for more details. Asterisks: please note that horizontal evolution has been estimated to be much more prevalent than indicated in the schematic tree.

evolutionary branching points requires advances within a wide range of research topics, which we summarize below (Liberles et al. 2020, fig. 4).

Sequence Data across the TOL

The availability of molecular sequence data for appropriate and extensive taxa sets is a key factor for the reconstruction of congruent phylogenies and understanding life's evolutionary history in general (Som 2015). Advances in sequencing and data processing techniques have considerably expanded the set of genomes from uncultivated organisms across the TOL and led to a large set of single-cell and metagenome-assembled genomes (SAGs, MAGs) (Eloe-Fadrosh, Ivanova, et al. 2016; Eloe-Fadrosh, Paez-Espino, et al. 2016; Kyrpides et al. 2016; Parks et al. 2017; Gregory et al. 2019). However, the quality of these SAGs and MAGs differs widely (Bowers et al. 2017) and, thus far, rarely provide resolution on single strain level. Current developments of hybrid metagenome assembly methodologies combining both short and long DNA sequence reads (Liao et al. 2019; Wang et al. 2021), innovative genome scaffolding approaches using chromosome conformation capture techniques (Yildirir et al. 2022), and sophisticated (meta)genome assembly computer software (e.g., Bertrand et al. 2019; Kolmogorov et al. 2020; Wang et al. 2021 for review) are promising avenues to obtain high quality strain-resolved MAGs (Chen, Anantharaman, et al. 2020; Olm et al. 2021; Quince et al. 2021) including their CRISPR loci as well as ribosomal RNA operon(s). Such improved metagenomics-driven analyses are also valuable not only for expanding the known diversity of DNA viruses (Paez-Espino et al. 2016; Martinez-Hernandez et al. 2017; Gregory et al. 2019; Moniruzzaman, Martinez-Gutierrez et al. 2020; Bellas and Sommaruga 2021; Edgar et al. 2022), but also to link putative viral genomes to their potential hosts through matching CRISPR spacers (Al-Shayeb et al. 2020); an approach recently used for the identification of viruses infecting Asgard archaea (Medvedeva et al. 2021; Rambo et al. 2021; Tamarit et al. 2021). Considering the complexity of viral populations, a perhaps even more promising approach relies on improved long-range sequencing technologies and was recently used to obtain complete viral genomes without the need for assembly and binning (Beaulaurier, 2020).

In contrast to prokaryotes and viruses, many lineages of eukaryotes, and especially microbial representatives, remain only sparsely sampled, which considerably limits our understanding of the early evolution and diversification of these organisms (Sibbald and Archibald 2017). Only a small number of protists have been enriched in culture and metagenomic approaches targeting uncultivated protists directly are difficult to implement due to the unique and complex genomic features of many representatives (McGrath and Katz 2004), which poses challenges for genome assembly and metagenomic procedures. Further, it should be emphasized that establishing methods for cultivation (or single-cell isolation), nucleic acid isolation, and sequencing from understudied eukaryotes in and of itself is not trivial and requires years of

optimization before data analysis can begin (Burki et al. 2020). Many protists harbor symbionts and/or can only be cultivated with other microbes thereby making most protist sequencing projects mini-metagenomics initiatives. Assuming high-quality genomic or transcriptomic data sets can be obtained, the next major obstacle is gene prediction. For genome projects, the nonuniform sequence composition across the genome and the complex architecture of eukaryotic genomes (i.e., large intergenic regions, introns) is a challenge for metagenomic "binning" and gene prediction tools, respectively. Although recent advances in assembling eukaryotic genomes and predicting gene content from complex samples (e.g., nonaxenic cultures or environmental samples) will help in overcoming these obstacles, e.g., West et al. (2018) and Yildirir et al. (2022). Finally, the lack of high-quality reference annotations from diverse eukaryotic representatives, large number of paralogues, and high proportions of lineage or organism-specific putative protein-coding genes in eukaryotic genomes (up to 60% [Karnkowska et al. 2019]) can impede clustering of orthologous groups and poses challenges for the accurate inference of gene history evolution.

Phylogenetics and Phylogenomics

Ways to resolve incongruences and uncertainties in phylogenies inferred with state-of-the-art phylogenetic and phylogenomic approaches have been reviewed recently (Som 2015; Williams et al. 2021) and will not be extensively discussed. These strategies include, among various others, the development of models of DNA and protein sequence evolution that better capture the processes by which molecular sequences evolve and adequately deal with sources of systematic error (i.e., nonphylogenetic signal) in sequence data: for example, see the recent development of heterotachy mixture models (Crotty et al. 2020). Much of our understanding of the evolutionary history of life mainly derives from analyses of multigene concatenations based on a limited set of universally conserved single-copy marker genes (see, e.g., Martinez-Gutierrez and Aylward 2021; Moody et al. 2022). Elucidating ancient divergences is challenging and requires the use of metrices to assess confidence in tree topologies and bipartitions. However, classical metrices such as the bootstrap, originally designed for single gene trees, have the tendency to overestimate confidence in bipartitions when the analyses are based on long alignments from multigene concatenations (Salichos and Rokas 2013). In turn, it is valuable to explore improved measures to assess confidence in tree and branching patterns (Thomson and Brown 2022), such as, for example, the recently developed internode and tree certainty metrices (Kobert et al. 2016; Martinez-Gutierrez and Aylward 2021). Furthermore, although key to inferring phylogenetic relationships of taxa, multigene concatenations are insufficient to reconstruct the evolution of genomes, which not only results from substitutions but also from gene and genome rearrangements, duplications and the loss and gain of new genes (Long et al. 2013; Andersson et al. 2015). Novel methodologies, capable of capturing simultaneously the vertical and horizontal components of genome evolution such as phylogenetic networks (Dagan 2011), topological data analyses (Chan et al. 2013; Cámara 2017), as well as gene tree-species tree reconciliation methods (Szöllõsi et al. 2012; David and Alm 2011; Szöllosi et al. 2013; Morel et al. 2022), open up new perspectives toward integrating data from viruses, and other genetic elements as well as providing a deeper understanding of gene family evolution including both vertical and horizontal components, across the TOL. For instance, reconciliation methods rely on a model to describe gene tree evolution involving originations, duplications, transfers, and losses under a given species tree and allow to determine the probability of any protein family at any given node in a tree (Williams et al. 2017; Coleman et al. 2021). Furthermore, such approaches can be used to determine the likelihood of certain root positions in the absence of a remote outgroup (Williams et al. 2017; Coleman et al. 2021), which, if available, can cause phylogenetic artifacts such as long branch attraction (Bergsten 2005; Philippe et al. 2005). The modeling of reticulate evolution has recently also been shown to allow dating the TOL (Davin et al. 2018; Wolfe and Fournier 2018), which previously solely relied on the scarce fossil and biomarker record available for the early steps of microbial evolution. Together, this can greatly enhance the understanding and timing of the evolutionary trajectories of life.

Reconstruction of Ancestral Sequences and Genomes

Progress in the sequencing and assembly of ancient DNA has been successfully applied to reconstruct the genome sequence of organisms (Orlando et al. 2015; Leonardi et al. 2017; Cappellini et al. 2018; Pont et al. 2019) including microorganisms (Arriola et al. 2020; Lammers et al. 2021; Liang et al. 2021) that existed up to hundreds of thousands years ago (i.e., allochronic reconstruction). However, such data is scarce; thus genes, proteins, and genomes of ancestral organisms are predominantly inferred from the sequence of extant taxa using so-called ancestral state reconstruction methodologies (i.e., synchronic reconstruction) (Omland 1999). This includes both ancestral (gene) sequence (Joy et al. 2016; Merkl and Sterner 2016; Gumulya and Gillam 2017; Selberg et al. 2021) and genome reconstruction approaches such as gene tree-species tree reconciliations (see above) (Szöllõsi et al. 2012; David and Alm 2011; Szöllosi et al. 2013; Williams et al. 2017; Coleman et al. 2021; Morel et al. 2022). In turn, features of ancestral organisms and the direction of evolutionary change can be investigated simultaneously.

Progressing further in our knowledge of the features of ancestral organisms involves "resurrecting" those life forms or, at least, some of their proteins (Thornton 2004; Hochberg

and Thornton 2017; Mascotti 2022) before characterizing them using molecular, biochemical, and biophysical approaches. Although this has been successfully undertaken for several types of proteins and protein complexes (Finnigan et al. 2012; Shih et al. 2016; Siddiq et al. 2017; Pillai et al. 2020), features of ancestral proteins and protein complexes thought to have played roles in major evolutionary transitions remain largely unknown. In contrast, the "de novo synthesis" of minimal, ancestral cells, still poses significant challenges (Schwille et al. 2018).

Evolutionary Cell Biology

Reconstructing and understanding the evolution of the ultrastructural complexity of cells and their components throughout the TOL and, most notably, during eukaryogenesis, requires linking gene and genome sequences to protein structures and cellular features. Although the intracellular organization of bacterial and archaeal cells has long been thought to be relatively simple, tremendous advances of microscopy techniques and image analyses now allow probing the cells of these organisms with sufficient resolution to reveal their cytological features in unprecedented detail (Surovtsev and Jacobs-Wagner 2018). Cryoelectron microscopy (Milne et al. 2013) and cryoelectron tomography (Beck and Baumeister 2016; Oikonomou and Jensen 2017) have notably revealed that the ultrastructure of bacterial and archaeal cells is far more complex and diverse than assumed previously (Dobro et al. 2017; Surovtsev and Jacobs-Wagner 2018; Greening and Lithgow 2020; Seeger et al. 2021). Microorganisms are now known to have a wide variety of intracellular organelles (Greening and Lithgow 2020), as well as other intracellular compartments of unknown function including nanospheres and both intracellular and periplasmic vesicles (Dobro et al. 2017). Further, bacterial and archaeal cells often include various types of intracellular filaments, bundles, arrays, and tubes in addition to varied cell appendages (Dobro et al. 2017). The extent to which the cytological features of certain bacteria and archaea, such as *Ca.* P. syntrophicum (Imachi et al. 2020), are related to one another and to those of eukaryotes, remains for now largely unknown considering that genes and proteins involved in their formation have not been identified in many cases. Current advances in the computational prediction of the structure of individual proteins (Baek et al. 2021; Jumper et al. 2021) and both the composition and structure of protein complexes (Baek et al. 2021; Humphreys et al. 2021) have the potential to accelerate the identification of genes involved in protein complexes forming cytological features. Indeed, the accuracy of the protein structures predicted by the neural-network models AlphaFold2 (Jumper et al. 2021) and RoseTTA fold (Baek et al. 2021) rivals that of experimentally determined structures (Baek et al. 2021; Kryshtafovych et al. 2021). Predicted protein structures can help interpreting Coulomb potential maps obtained by cryoelectron microscopy and cellular cryoelectron tomography for the experimental determination of protein structures (Gupta et al. 2021). Furthermore, the development of standards to adequately evaluate the fit of computationally predicted protein models to the Coulomb potential maps of protein complexes may allow to refine protein complex structures and identify genes coding for protein complex components (Masrati et al. 2021). We envision that progress in the computational predictions of protein structures may also allow for the identification of proteins, which share similar folds but little to no amino acid sequence similarity to known components of well-characterized cellular features. Once candidate protein components of a cellular feature of interest have been identified by, for instance, immunogold labeling (Mayhew 2011), the localization, dynamics, and function of the proteins, and corresponding cytological features can be investigated using antibodies conjugated with fluorescent labels and superresolution microscopy (Tuson and Biteen 2015; Möckl and Moerner 2020) as performed, for example, for the analysis of the cytokinesis machinery of bacteria (Holden 2018) and archaea (Pende et al. 2021). Altogether, these protein structure-based approaches combined with high-end microscopy now allow us to bridge the gap between bioinformatic analyses and cell biology and to reconstruct major steps in the evolution of cellular complexity.

Genotype–Phenotype Relationship

Moving from the reconstruction of the evolutionary history of life to understanding the evolutionary trajectories taken by life forms through time requires clarifying their evolvability (Kirschner and Gerhart 1998; Pigliucci 2008; Payne and Wagner 2019). This includes elucidating the physical constraints on the phenotypes that organisms or their cellular components may take (Alexander 1985; Smith et al. 1985; Arnold 1992; Furusawa and Irie 2020) but also identifying features of biological systems opening opportunities for the emergence of phenotypic variation, innovation, and diversification (Sharov 2014). This emphasizes the need to study fundamental attributes of microbial cells including for example, trade-offs (Garland 2014; Acerenza 2016), allometric scaling laws (West et al. 1997, 2002; Giometto et al. 2013) and robustness (de Visser et al. 2003; Kitano 2007; Masel and Trotter 2010) and their respective underlying causes at the molecular level. Progress in this research area will allow for a better understanding of the relation between genotype and phenotype (i.e., genotype-phenotype map [Pigliucci 2010; Wagner and Zhang 2011; Ahnert 2017]) thereby clarifying the landscape of possible genetic changes. Advances in high-throughput phenotyping and genotyping, targeted genome editing, and single cell approaches (Prakadan et al. 2017; Adli 2018; Ohan et al. 2019; Zahir et al. 2019; Acin-Albiac et al. 2020; Kaster and Sobol 2020; McCarty et al. 2020; Arroyo-Olarte et al. 2021; Rubin et al. 2022),

evolutionary synthetic biology (Peisajovich 2012; Baier and Schaerli 2021; Ijäs and Koskinen 2021), and experimental evolution (Van den Bergh et al. 2018), are currently driving progress in the exploration of the genotype–phenotype map. Yet, conceptual, and theoretical developments need to follow technological advances to derive the principles determining the evolution of (micro)organisms. Although such studies are typically conducted on model organisms, a focus on microbial groups placed near key evolutionary branching points would be beneficial for understanding major transitions in the early evolution of life on Earth. This emphasizes the need to isolate and develop laboratory cultivation systems to study members of these microbial groups, most of which remain currently uncultivated (Lewis et al. 2021).

Conclusion

The TOL is a constantly changing and evolving concept in evolutionary biology, which has helped to depict the vast biodiversity on Earth, including both vertical and horizontal relations of organisms as well as connections to MGEs including viruses. Of course, it will always constitute a simplified illustration of the diversification of life on Earth and can only account for the evolutionary path of extant organisms even though extinct organisms may have contributed to the genetic repertoire of extant genomes. For example, all organisms today are derived from LUCA, yet the early diversification of LUCA was likely shaped by gene influx from now extinct organisms living at the time of LUCA.

Nevertheless, the TOL provides a useful concept for describing and classifying the diversity of organismal life on Earth today (Rinke et al. 2021; Parks et al. 2018) and for improving our understanding of events leading to major evolutionary changes that have dramatically impacted our biosphere. The continuous improvement of analytical, experimental and computational approaches to the study of life's biodiversity and integration of geological records will further improve our insights into the evolutionary past and allow linking diversification to Earth history. Further, this will help to refine our understanding of evolutionary principles underlying biodiversification, which is crucial for predicting evolution and may help efforts to preserve biodiversity in an ever-changing world.

Acknowledgments

A.S. was supported by the European Research Council (ERC STG ASymbEL: 947317), the Swedish Research Council (VR starting grant 2016-03559 to A.S.) and the NWO-I foundation of the Netherlands Organisation for Scientific Research (WISE fellowship). C.W.S. was supported by the Swedish Research Council (VR starting grant 2020-05071). We apologize to colleagues whose work we were unable to cite due to length constraints.

Author Contributions

A.S. conceptualized and all authors have contributed to the writing of this review.

Literature Cited

Abrahao J, et al. 2018. Tailed giant Tupanvirus possesses the most complete translational apparatus of the known virosphere. Nat Commun. 9(1):749.
Acerenza L. 2016. Constraints, trade-offs and the currency of fitness. J Mol Evol. 82(2–3):117–127.
Acin-Albiac M, Filannino P, Gobbetti M, Di Cagno R. 2020. Microbial high throughput phenomics: the potential of an irreplaceable omics. Comput Struct Biotechnol J. 18:2290–2299.
Adam PS, Borrel G, Brochier-Armanet C, Gribaldo S. 2017. The growing tree of Archaea: new perspectives on their diversity, evolution and ecology. ISME J. 11(11):2407–2425.
Adl SM, et al. 2019. Revisions to the classification, nomenclature, and diversity of eukaryotes. J Eukaryot Microbiol. 66(1):4–119.
Adli M. 2018. The CRISPR tool kit for genome editing and beyond. Nat Commun. 9(1):1911.
Ahnert SE. 2017. Structural properties of genotype–phenotype maps. J R Soc Interface. 14(132):20170275.
Akil C, Robinson RC. 2018. Genomes of Asgard archaea encode profilins that regulate actin. Nature 562(7727):439–443.
Akil C, et al. 2020. Insights into the evolution of regulated actin dynamics via characterization of primitive gelsolin/cofilin proteins from Asgard archaea. Proc Natl Acad Sci USA. 117(33):19904–19913.
Al-Shayeb B, et al. 2020. Clades of huge phages from across Earth's ecosystems. Nature 578(7795):425–431.
Alacid E, Richards TA. 2021. A cell–cell atlas approach for understanding symbiotic interactions between microbes. Curr Opin Microbiol. 64:47–59.
Alexander RM. 1985. The ideal and the feasible: physical constraints on evolution. Biolog J Linn Soc. 26(4):345–358.
Andersson DI, Jerlström-Hultqvist J, Näsvall J. 2015. Evolution of new functions de novo and from preexisting genes. Cold Spring Harb Perspect Biol. 7(6):a017996.
Aouad M, et al. 2018. Extreme halophilic archaea derive from two distinct methanogen Class II lineages. Mol Phylogenet Evol. 127:46–54.
Aouad M, et al. 2022. A divide-and-conquer phylogenomic approach based on character supermatrices resolves early steps in the evolution of the Archaea. BMC Ecol Evol. 22(1):1.
Arnold SJ. 1992. Constraints on phenotypic evolution. Am Nat. 140:S85–S107.
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.
Arroyo-Olarte RD, Bravo Rodríguez R, Morales-Ríos E. 2021. Genome editing in bacteria: CRISPR-Cas and beyond. Microorganisms 9(4):844.
Avci B, et al. 2022. Spatial separation of ribosomes and DNA in Asgard archaeal cells. ISME J. 16(2):606–610.
Backstrom D, et al. 2019. Virus genomes from deep sea sediments expand the ocean megavirome and support independent origins of viral gigantism. mBio 10(2), e02497-18. doi:10.1128/mBio.02497-18.
Baek M, et al. 2021. Accurate prediction of protein structures and interactions using a three-track neural network. Science 373(6557):871–876.
Baier F, Schaerli Y. 2021. Addressing evolutionary questions with synthetic biology. In: Crombach A, editor. Evolutionary systems biology: advances, questions, and opportunities. Springer, Cham: Springer International Publishing. p. 135–157. https://doi.org/10.1007/978-3-030-71737-7_7

Baum DA, Baum B. 2014. An inside-out origin for the eukaryotic cell. BMC Biol. 12:76.
Beam JP, et al. 2020. Ancestral absence of electron transport chains in patescibacteria and DPANN. Front Microbiol. 11:1848.
Beaulaurier J, et al. 2020. Assembly-free single-molecule sequencing recovers complete virus genomes from natural microbial communities. Genome Res. 30(3):437–446.
Beck M, Baumeister W. 2016. Cryo-electron tomography: can it reveal the molecular sociology of cells in atomic detail? Trends Cell Biol. 26(11):825–837.
Bellas CM, Sommaruga R. 2021. Polinton-like viruses are abundant in aquatic ecosystems. Microbiome 9(1):13.
Bergsten J. 2005. A review of long-branch attraction. Cladistics 21(2):163–193.
Bertrand D, et al. 2019. Hybrid metagenomic assembly enables high-resolution analysis of resistance determinants and mobile elements in human microbiomes. Nat Biotechnol. 37(8):937–944.
Betts HC, et al. 2018. Integrated genomic and fossil evidence illuminates life's early evolution and eukaryote origin. Nat Ecol Evol. 2(10):1556–1562.
Blais C, Archibald JM. 2021. The past, present and future of the tree of life. Curr Biol. 31(7):R314–R321.
Bor B, et al. 2018. Rapid evolution of decreased host susceptibility drives a stable relationship between ultrasmall parasite TM7x and its bacterial host. Proc Natl Acad Sci USA. 115(48):12277–12282.
Bowers RM, et al. 2017. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. Nat Biotechnol. 35(8):725–731.
Brochier-Armanet C, Forterre P, Gribaldo S. 2011. Phylogeny and evolution of the Archaea: one hundred genomes later. Curr Opin Microbiol. 14(3):274–281.
Brown CT, et al. 2015. Unusual biology across a group comprising more than 15% of domain bacteria. Nature 523(7559):208–211.
Brown JR, Doolittle WF. 1995. Root of the universal tree of life based on ancient aminoacyl-tRNA synthetase gene duplications. Proc Natl Acad Sci USA. 92(7):2441–2445.
Brown MW, et al. 2018. Phylogenomics places orphan protistan lineages in a novel eukaryotic super-group. Genome Biol Evol. 10(2):427–433.
Burki F, Roger AJ, Brown MW, Simpson AGB. 2020. The new tree of eukaryotes. Trends Ecol Evol. 35(1):43–55.
Burki F, Sandin MM, Jamy M. 2021. Diversity and ecology of protists revealed by metabarcoding. Curr Biol. 31(19):R1267–R1280.
Cai M, et al. 2020. Diverse Asgard archaea including the novel phylum Gerdarchaeota participate in organic matter degradation. Sci China Life Sci. 63(6):886–897.
Cappellini E, et al. 2018. Ancient biomolecules and evolutionary inference. Annu Rev Biochem. 87:1029–1060.
Castelle CJ, Banfield JF. 2018. Major new microbial groups expand diversity and alter our understanding of the tree of life. Cell 172(6):1181–1197.
Castelle CJ, et al. 2015. Genomic expansion of domain archaea highlights roles for organisms from new phyla in anaerobic carbon cycling. Curr Biol. 25(6):690–701.
Castelle CJ, et al. 2018. Biosynthetic capacity, metabolic variety and unusual biology in the CPR and DPANN radiations. Nat Rev Microbiol. 16(10):629–645.
Castelle CJ, et al. 2021. Protein family content uncovers lineage relationships and bacterial pathway maintenance mechanisms in DPANN Archaea. Front Microbiol. 12:660052.
Cavalier-Smith T. 2006. Rooting the tree of life by transition analyses. Biol Direct. 1(1):19.
Cermakian N, et al. 1997. On the evolution of the single-subunit RNA polymerases. J Mol Evol. 45(6):671–681.
Cerón-Romero MA, Fonseca MM, de Oliveira Martins L, Posada D, Katz LA. 2021. Phylogenomic analyses of 2,786 genes in 158 lineages support a root of the eukaryotic tree of life between opisthokonts (animals, fungi and their microbial relatives) and all other lineages. bioRxiv. 2021.02.26.433005; doi:10.1101/2021.02.26.433005.
Chaikeeratisak V, et al. 2017. Assembly of a nucleus-like structure during viral replication in bacteria. Science 355(6321):194–197.
Chan JM, Carlsson G, Rabadan R. 2013. Topology of viral evolution. Proc Natl Acad Sci USA. 110(46):18566–18571.
Chen L-X, Anantharaman K, Shaiber A, Eren AM, Banfield JF. 2020. Accurate and complete genomes from metagenomes. Genome Res. 30(3):315–333.
Chen R, Wong HL, et al. 2020. Discovery of an abundance of biosynthetic gene clusters in shark bay microbial mats. Front Microbiol. 11:1950.
Coleman GA, et al. 2021. A rooted phylogeny resolves early bacterial evolution. Science 372(6542):eabe0511.
Collens AB, Katz LA. 2021. Opinion: genetic conflict with mobile elements drives eukaryotic genome evolution, and perhaps also eukaryogenesis. J Hered. 112(1):140–144.
Crotty SM, et al. 2020. GHOST: recovering historical signal from heterotachously evolved sequence alignments. Syst Biol. 69(2):249–264.
Cámara PG. 2017. Topological methods for genomics: present and future directions. Curr Opin Syst Biol. 1:95–101.
Da Cunha V, Gaia M, Gadelle D, Nasir A, Forterre P. 2017. Lokiarchaea are close relatives of Euryarchaeota, not bridging the gap between prokaryotes and eukaryotes. PLoS Genet. 13(6):e1006810.
Da Cunha V, Gaia M, Nasir A, Forterre P. 2018. Asgard archaea do not close the debate about the universal tree of life topology. PLoS Genet. 14(3):e1007215.
Da Cunha V, et al. 2022. Giant viruses encode actin-related proteins. Mol Biol Evol. 39(2):msac022. doi:10.1093/molbev/msac022.
Dagan T. 2011. Phylogenomic networks. Trends Microbiol. 19(10):483–491.
Dagan T, Artzy-Randrup Y, Martin W. 2008. Modular networks and cumulative impact of lateral transfer in prokaryote genome evolution. Proc Natl Acad Sci USA. 105(29):10039–10044.
Dagan T, Roettger M, Bryant D, Martin W. 2010. Genome networks root the tree of life between prokaryotic domains. Genome Biol Evol. 2:379–392.
David LA, Alm EJ. 2011. Rapid evolutionary innovation during an Archaean genetic expansion. Nature 469(7328):93–96.
Davin AA, et al. 2018. Gene transfers can date the tree of life. Nat Ecol Evol. 2(5):904–909.
de Visser JAGM, et al. 2003. Perspective: evolution and detection of genetic robustness. Evol 57(9):1959–1972.
Delaye L, Valadez-Cano C, Pérez-Zamorano B. 2016. How really ancient is Paulinella Chromatophora? PLoS Curr. 8. doi:10.1371/currents.tol.e68a099364bb1a1e129a17b4e06b0c6b.
Derelle R, et al. 2015. Bacterial proteins pinpoint a single eukaryotic root. Proc Natl Acad Sci USA. 112(7):E693–E699.
Deutekom ES, Snel B, van Dam TJP. 2021. Benchmarking orthology methods using phylogenetic patterns defined at the base of Eukaryotes. Brief Bioinform. 22:bbaa206.
Dobro MJ, et al. 2017. Uncharacterized bacterial structures revealed by electron cryotomography. J Bacteriol. 199:00100–00117.
Dombrowski N, Lee JH, Williams TA, Offre P, Spang A. 2019. Genomic diversity, lifestyles and evolutionary origins of DPANN archaea. FEMS Microbiol Lett. 366:fnz008.
Dombrowski N, et al. 2020. Undinarchaeota illuminate DPANN phylogeny and the impact of gene transfer on archaeal evolution. Nat Commun. 11(1):3939.
Doolittle WF, Bapteste E. 2007. Pattern pluralism and the tree of life hypothesis. Proc Natl Acad Sci USA. 104(7):2043–2049.

Dzunkova M, et al. 2019. Defining the human gut host-phage network through single-cell viral tagging. Nat Microbiol. 4(12):2192–2203.
Edgar RC, et al. 2022. Petabase-scale sequence alignment catalyses viral discovery. Nature 602(7895):142–147.
Eloe-Fadrosh EA, Ivanova NN, Woyke T, Kyrpides NC. 2016. Metagenomics uncovers gaps in amplicon-based detection of microbial diversity. Nat Microbiol. 1:15032.
Eloe-Fadrosh EA, Paez-Espino D, et al. 2016. Global metagenomic survey reveals a new bacterial candidate phylum in geothermal springs. Nat Commun. 7:10476.
Eme L, Sharpe SC, Brown MW, Roger AJ. 2014. On the age of eukaryotes: evaluating evidence from fossils and molecular clocks. Cold Spring Harb Perspect Biol. 6:a016139.
Eme L, Spang A, Lombard J, Stairs CW, Ettema TJG. 2017. Archaea and the origin of eukaryotes. Nat Rev Microbiol. 15(12):711–723.
Fan L, et al. 2020. Phylogenetic analyses with systematic taxon sampling show that mitochondria branch within Alphaproteobacteria. Nat Ecol Evol. 4(9):1213–1219.
Farag IF, et al. 2020. Metabolic potentials of archaeal lineages resolved from metagenomes of deep Costa Rica sediments. ISME J. 14(6):1345–1358.
Farag IF, Zhao R, Biddle JF. 2021. "Sifarchaeota," a novel Asgard phylum from Costa Rican sediment capable of polysaccharide degradation and anaerobic methylotrophy. Appl Environ Microbiol. 87(9):e02584-20.
Feng Y, et al. 2021. The evolutionary origins of extreme halophilic archaeal lineages. Genome Biol Evol. 13:evab166.
Feschotte C, Gilbert C. 2012. Endogenous viruses: insights into viral evolution and impact on host biology. Nat Rev Genet. 13(4):283–296.
Filée J, Forterre P. 2005. Viral proteins functioning in organelles: a cryptic origin?. Trends Microbiol. 13(11):510–513.
Finnigan GC, Hanson-Smith V, Stevens TH, Thornton JW. 2012. Evolution of increased complexity in a molecular machine. Nature 481(7381):360–364.
Forterre P. 2016. To be or not to be alive: how recent discoveries challenge the traditional definitions of viruses and life. Stud Hist Philos Biol Biomed Sci. 59:100–108.
Forterre P, Krupovic M, Prangishvili D. 2014. Cellular domains and viral lineages. Trends Microbiol. 22(10):554–558.
Furusawa C, Irie N. 2020. Toward understanding of evolutionary constraints: experimental and theoretical approaches. Biophys Rev. 12(5):1155–1161.
Garg SG, et al. 2021. Anomalous phylogenetic behavior of ribosomal proteins in metagenome-assembled Asgard archaea. Genome Biol Evol. 13:evaa238.
Garland T. 2014. Trade-offs. Curr Biol. 24(2):R60–R61.
Gawryluk RMR, et al. 2019. Non-photosynthetic predators are sister to red algae. Nature 572(7768):240–243.
Gawryluk RMR, Stairs CW. 2021. Diversity of electron transport chains in anaerobic protists. Biochim Biophys Acta Bioenerg. 1862(1):148334.
Giometto A, Altermatt F, Carrara F, Maritan A, Rinaldo A. 2013. Scaling body size fluctuations. Proc Natl Acad Sci USA. 110(12):4646–4650.
Golyshina OV, et al. 2017. 'ARMAN' archaea depend on association with euryarchaeal host in culture and in situ. Nat Commun. 8(1):60.
Gould SB, Garg SG, Martin WF. 2016. Bacterial vesicle secretion and the evolutionary origin of the eukaryotic endomembrane system. Trends Microbiol. 24(7):525–534.
Gouy R, Baurain D, Philippe H. 2015. Rooting the tree of life: the phylogenetic jury is still out. Philos Trans R Soc Lond B Biol Sci. 370(1678):20140329.
Greening C, Lithgow T. 2020. Formation and function of bacterial organelles. Nat Rev Microbiol. 18(12):677–689.
Gregory AC, et al. 2019. Marine DNA viral macro- and microdiversity from pole to pole. Cell 177(5):1109–1123.e1114.
Guglielmini J, Woo AC, Krupovic M, Forterre P, Gaia M. 2019. Diversification of giant and large eukaryotic dsDNA viruses predated the origin of modern eukaryotes. Proc Natl Acad Sci USA. 116(39):19585–19592.
Gumulya Y, Gillam EMJ. 2017. Exploring the past and the future of protein evolution with ancestral sequence reconstruction: the 'retro' approach to protein engineering. Biochem J. 474(1):1–19.
Gupta M, et al. 2021. CryoEM and AI reveal a structure of SARS-CoV-2 Nsp2, a multifunctional protein involved in key host processes. bioRxiv. 2021.05.10.443524; doi:10.1101/2021.05.10.443524.
Guy L, Saw JH, Ettema TJ. 2014. The archaeal legacy of eukaryotes: a phylogenomic perspective. Cold Spring Harb Perspect Biol. 6(10):a016022.
Hamm JN, et al. 2019. Unexpected host dependency of Antarctic Nanohaloarchaeota. Proc Natl Acad Sci USA. 116(29):14661–14670.
Harada R, Inagaki Y. 2021. Phage origin of mitochondrion-localized family a DNA polymerases in kinetoplastids and diplonemids. Genome Biol Evol. 13:evab003.
Hartman H, Fedorov A. 2002. The origin of the eukaryotic cell: a genomic investigation. Proc Natl Acad Sci USA. 99(3):1420–1425.
He D, et al. 2014. An alternative root for the eukaryote tree of life. Curr Biol. 24(4):465–470.
He X, et al. 2015. Cultivation of a human-associated TM7 phylotype reveals a reduced genome and epibiotic parasitic lifestyle. Proc Natl Acad Sci USA. 112(1):244–249.
Heiss AA, et al. 2018. Combined morphological and phylogenomic re-examination of malawimonads, a critical taxon for inferring the evolutionary history of eukaryotes. R Soc Open Sci. 5(4):171707.
Hochberg GKA, Thornton JW. 2017. Reconstructing ancient proteins to understand the causes of structure and function. Annu Rev Biophys. 46:247–269.
Holden S. 2018. Probing the mechanistic principles of bacterial cell division with super-resolution microscopy. Curr Opin Microbiol. 43:84–91.
Hoshino Y, Gaucher EA. 2021. Evolution of bacterial steroid biosynthesis and its impact on eukaryogenesis. Proc Natl Acad Sci USA. 118:e2101276118.
Huber H, et al. 2002. A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont. Nature 417(6884):63–67.
Hug LA, et al. 2016. A new view of the tree of life. Nat Microbiol. 1(5):Article number: 16048.
Humphreys IR, et al. 2021. Computed structures of core eukaryotic protein complexes. Science 374(6573):eabm4805.
Ijäs T, Koskinen R. 2021. Exploring biological possibility through synthetic biology. Eur J Philos Sci. 11(2):39.
Imachi H, et al. 2020. Isolation of an archaeon at the prokaryote–eukaryote interface. Nature 577(7791):519–525.
Irwin NAT, Pittis AA, Richards TA, Keeling PJ. 2022. Systematic evaluation of horizontal gene transfer between eukaryotes and viruses. Nat Microbiol. 7(2):327–336.
Iwabe N, Kuma K, Hasegawa M, Osawa S, Miyata T. 1989. Evolutionary relationship of archaebacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes. Proc Natl Acad Sci USA. 86(23):9355–9359.
Jahn U, et al. 2008. *Nanoarchaeum equitans* and *Ignicoccus hospitalis*: new insights into a unique, intimate association of two archaea. J Bacteriol. 190(5):1743–1750.
Janouskovec J, et al. 2017. A new lineage of eukaryotes illuminates early mitochondrial genome reduction. Curr Biol. 27(23):3717–3724.e3715.
Jarett JK, et al. 2020. Insights into the dynamics between viruses and their hosts in a hot spring microbial mat. ISME J. 14(10):2527–2541.
Joy JB, Liang RH, McCloskey RM, Nguyen T, Poon AFY. 2016. Ancestral reconstruction. PLoS Comput Biol. 12(7):e1004763.

Jumper J, et al. 2021. Highly accurate protein structure prediction with AlphaFold. Nature 596(7873):583–589.
Karnkowska A, et al. 2016. A eukaryote without a mitochondrial organelle. Curr Biol. 26(10):1274–1284.
Karnkowska A, et al. 2019. The oxymonad genome displays canonical eukaryotic complexity in the absence of a mitochondrion. Mol Biol Evol. 36(10):2292–2312.
Kaster A-K, Sobol MS. 2020. Microbial single-cell omics: the crux of the matter. Appl Microbiol Biotechnol. 104(19):8209–8220.
Katz LA. 2015. Recent events dominate interdomain lateral gene transfers between prokaryotes and eukaryotes and, with the exception of endosymbiotic gene transfers, few ancient transfer events persist. Philos Trans R Soc Lond B Biol Sci. 370(1678):20140324.
Katz LA, Grant JR, Parfrey LW, Burleigh JG. 2012. Turning the crown upside down: gene tree parsimony roots the eukaryotic tree of life. Syst Biol. 61(4):653–660.
Kijima S, et al. 2021. Discovery of viral myosin genes with complex evolutionary history within plankton. Front Microbiol. 12:683294.
Kirschner M, Gerhart J. 1998. Evolvability. Proc Natl Acad Sci USA. 95(15):8420–8427.
Kitano H. 2007. Towards a theory of biological robustness. Mol Syst Biol. 3:137.
Kobert K, Salichos L, Rokas A, Stamatakis A. 2016. Computing the internode certainty and related measures from partial gene trees. Mol Biol Evol. 33(6):1606–1617.
Kolmogorov M, et al. 2020. metaFlye: scalable long-read metagenome assembly using repeat graphs. Nat Methods. 17(11):1103–1110.
Koonin EV. 2010. The origin and early evolution of eukaryotes in the light of phylogenomics. Genome Biol. 11(5):209.
Koonin EV. 2016. Horizontal gene transfer: essentiality and evolvability in prokaryotes, and roles in evolutionary transitions. F1000Res 5. doi:10.12688/f1000research.8737.1.
Koonin EV. 2015a. Origin of eukaryotes from within archaea, archaeal eukaryome and bursts of gene gain: eukaryogenesis just made easier? Philos Trans R Soc Lond B Biol Sci. 370(1678):20140333.
Koonin EV. 2015b. The turbulent network dynamics of microbial evolution and the statistical tree of life. J Mol Evol. 80(5–6):244–250.
Koonin EV, Dolja VV, Krupovic M. 2015. Origins and evolution of viruses of eukaryotes: the ultimate modularity. Virology 479–480:2–25.
Koonin EV, Krupovic M. 2018. The depths of virus exaptation. Curr Opin Virol. 31:1–8.
Koonin EV, Krupovic M, Agol VI. 2021. The Baltimore classification of viruses 50 years later: how does it stand in the light of virus evolution? Microbiol Mol Biol Rev. 85(3):e0005321.
Koonin EV, Krupovic M, Ishino S, Ishino Y. 2020. The replication machinery of LUCA: common origin of DNA replication and transcription. BMC Biol. 18(1):61.
Koonin EV, Starokadomskyy P. 2016. Are viruses alive? The replicator paradigm sheds decisive light on an old but misguided question. Stud Hist Philos Biol Biomed Sci. 59:125–134.
Koonin EV, Yutin N. 2014. The dispersed archaeal eukaryome and the complex archaeal ancestor of eukaryotes. Cold Spring Harb Perspect Biol. 6(4):a016188.
Koonin EV, Yutin N. 2018. Multiple evolutionary origins of giant viruses. F1000Res 7:1840.
Koonin EV, Yutin N. 2019. Evolution of the large nucleocytoplasmic DNA viruses of eukaryotes and convergent origins of viral gigantism. Adv Virus Res. 103:167–202.
Koumandou VL, et al. 2013. Molecular paleontology and complexity in the last eukaryotic common ancestor. Crit Rev Biochem Mol Biol. 48(4):373–396.
Krause S, Bremges A, Münch PC, McHardy AC, Gescher J. 2017. Characterisation of a stable laboratory co-culture of acidophilic nanoorganisms. Sci Rep. 7(1):3289–3289.
Krupovic M, Cvirkaite-Krupovic V, Iranzo J, Prangishvili D, Koonin EV. 2018. Viruses of archaea: structural, functional, environmental and evolutionary genomics. Virus Res. 244:181–193.
Krupovic M, Dolja VV, Koonin EV. 2019. Origin of viruses: primordial replicators recruiting capsids from hosts. Nat Rev Microbiol. 17(7):449–458.
Krupovic M, Dolja VV, Koonin EV. 2020. The LUCA and its complex virome. Nat Rev Microbiol. 18(11):661–670.
Krupovic M, Kuhn JH, Fischer MG. 2016. A classification system for virophages and satellite viruses. Arch Virol. 161(1):233–247.
Kryshtafovych A, Schwede T, Topf M, Fidelis K, Moult J. 2021. Critical assessment of methods of protein structure prediction (CASP)-Round XIV. Proteins 89(12):1607–1617.
Kyrpides NC, Eloe-Fadrosh EA, Ivanova NN. 2016. Microbiome data science: understanding our microbial planet. Trends Microbiol. 24(6):425–427.
La Cono V, et al. 2020. Symbiosis between nanohaloarchaeon and haloarchaeon is based on utilization of different polysaccharides. Proc Natl Acad Sci USA. 117(33):20223–20234.
Lake JA, Skophammer RG, Herbold CW, Servin JA. 2009. Genome beginnings: rooting the tree of life. Philos Trans R Soc Lond B Biol Sci. 364(1527):2177–2185.
Lammers Y, Heintzman PD, Alsos IG. 2021. Environmental palaeogenomic reconstruction of an Ice Age algal population. Commun Biol. 4(1):220.
Lax G, et al. 2018. Hemimastigophora is a novel supra-kingdom-level lineage of eukaryotes. Nature 564(7736):410–414.
Leger MM, Eme L, Stairs CW, Roger AJ. 2018. Demystifying eukaryote lateral gene transfer. Bioessays 40(5):e1700242.
Leger MM, et al. 2015. An ancestral bacterial division system is widespread in eukaryotic mitochondria. Proc Natl Acad Sci USA. 112(33):10239–10246.
Leonardi M, et al. 2017. Evolutionary patterns and processes: lessons from ancient DNA. Syst Biol. 66(1):e1–e29.
Lewis WH, Tahon G, Geesink P, Sousa DZ, Ettema TJG. 2021. Innovations to culturing the uncultured microbial majority. Nat Rev Microbiol. 19(4):225–240.
Liang R, et al. 2021. Genomic reconstruction of fossil and living microorganisms in ancient Siberian permafrost. Microbiome 9(1):110.
Liao X, et al. 2019. Current challenges and solutions of de novo assembly. Quantit Biol. 7:90–109.
Liberles DA, et al. 2020. Emerging frontiers in the study of molecular evolution. J Mol Evol. 88(3):211–226.
Liu Y, et al. 2021. Expanded diversity of Asgard archaea and their relationships with eukaryotes. Nature 593(7860):553–557.
Long M, VanKuren NW, Chen S, Vibranovski MD. 2013. New gene evolution: little did we know. Annu Rev Genet. 47:307–333.
Lopez-Garcia P. 2012. The place of viruses in biology in light of the metabolism- versus-replication-first debate. Hist Philos Life Sci. 34:391–406.
Lopez-Garcia P, Eme L, Moreira D. 2017. Symbiosis in eukaryotic evolution. J Theor Biol. 434:20–33.
Lopez-Garcia P, Moreira D. 2015. Open questions on the origin of eukaryotes. Trends Ecol Evol. 30(11):697–708.
Lopez-Garcia P, Moreira D. 2020. The Syntrophy hypothesis for the origin of eukaryotes revisited. Nat Microbiol. 5(5):655–667.
Manoharan L, et al. 2019. Metagenomes from coastal marine sediments give insights into the ecological role and cellular features of *Loki-* and *Thorarchaeota*. mBio 10(5):e02039-19.
Marin B, Nowack EC, Melkonian M. 2005. A plastid in the making: evidence for a second primary endosymbiosis. Protist 156(4):425–432.
Martijn J, Vosseberg J, Guy L, Offre P, Ettema TJG. 2018. Deep mitochondrial origin outside the sampled alphaproteobacteria. Nature 557(7703):101–105.
Martin W, Muller M. 1998. The hydrogen hypothesis for the first eukaryote. Nature 392(6671):37–41.

Martin WF. 2017. Too much eukaryote LGT. Bioessays 39(12):1700115.
Martin WF, Garg S, Zimorski V. 2015. Endosymbiotic theories for eukaryote origin. Philos Trans R Soc Lond B Biol Sci. 370(1678):20140330.
Martinez-Gutierrez CA, Aylward FO. 2021. Phylogenetic signal, congruence, and uncertainty across bacteria and archaea. Mol Biol Evol. 38(12):5514–5527.
Martinez-Hernandez F, et al. 2017. Single-virus genomics reveals hidden cosmopolitan and abundant viruses. Nat Commun. 8:15892.
Mascotti ML. 2022. Resurrecting enzymes by ancestral sequence reconstruction. Methods Molecular Biol. 2397:111–136.
Masel J, Trotter MV. 2010. Robustness and evolvability. Trends Genet. 26(9):406–414.
Masrati G, et al. 2021. Integrative structural biology in the era of accurate structure prediction. J Mol Biol. 433(20):167127.
Mayhew TM. 2011. Mapping the distributions and quantifying the labelling intensities of cell compartments by immunoelectron microscopy: progress towards a coherent set of methods. J Anat. 219(6):647–660.
McCarty NS, Graham AE, Studena L, Ledesma-Amaro R. 2020. Multiplexed CRISPR technologies for gene editing and transcriptional regulation. Nat Commun. 11(1):1281.
McGrath CL, Katz LA. 2004. Genome diversity in microbial eukaryotes. Trends Ecol Evol. 19(1):32–38.
Medvedeva S, et al. 2021. Viruses of Asgard archaea. bioRxiv. 2021.07.29.453957; doi:10.1101/2021.07.29.453957.
Meheust R, Burstein D, Castelle CJ, Banfield JF. 2019. The distinction of CPR bacteria from other bacteria based on protein family content. Nat Commun. 10(1):4173.
Merkl R, Sterner R. 2016. Ancestral protein reconstruction: techniques and applications. Biol Chem. 397(1):1–21.
Milne JLS, et al. 2013. Cryo-electron microscopy–a primer for the non-microscopist. FEBS J. 280(1):28–45.
Moniruzzaman M, Martinez-Gutierrez CA, Weinheimer AR, Aylward FO. 2020. Dynamic genome evolution and complex virocell metabolism of globally-distributed giant viruses. Nat Commun. 11(1):1710.
Moniruzzaman M, Weinheimer AR, Martinez-Gutierrez CA, Aylward FO. 2020. Widespread endogenization of giant viruses shapes genomes of green algae. Nature 588(7836):141–145.
Moody ERR, et al. 2022. An estimate of the deepest branches of the tree of life from ancient vertically evolving genes. Elife 11:e66695.
Moore KR, et al. 2019. An expanded ribosomal phylogeny of cyanobacteria supports a deep placement of plastids. Front Microbiol. 10:1612.
Moran NA. 1996. Accelerated evolution and Muller's rachet in endosymbiotic bacteria. Proc Natl Acad Sci USA. 93(7):2873–2878.
More K, Klinger CM, Barlow LD, Dacks JB. 2020. Evolution and natural history of membrane trafficking in eukaryotes. Curr Biol. 30(10):R553–R564.
Moreira D, Lopez-Garcia P. 1998. Symbiosis between methanogenic archaea and delta-proteobacteria as the origin of eukaryotes: the syntrophic hypothesis. J Mol Evol. 47(5):517–530.
Moreira D, Lopez-Garcia P. 2009. Ten reasons to exclude viruses from the tree of life. Nat Rev Microbiol. 7(4):306–311.
Moreira D, Lopez-Garcia P. 2015. Evolution of viruses and cells: do we need a fourth domain of life to explain the origin of eukaryotes? Philos Trans R Soc Lond B Biol Sci. 370(1678):20140327.
Moreira D, Zivanovic Y, López-Archilla AI, Iniesto M, López-García P. 2021. Reductive evolution and unique predatory mode in the CPR bacterium Vampirococcus lugosii. Nat Commun. 12(1):2454.
Morel B, et al. 2022. SpeciesRax: a tool for maximum likelihood species tree inference from gene family trees under duplication, transfer, and loss. Mol Biol Evol. 39(2):msab365. doi: 10.1093/molbev/msab365.
Munoz-Gomez SA, et al. 2022. Site-and-branch-heterogeneous analyses of an expanded dataset favour mitochondria as sister to known Alphaproteobacteria. Nat Ecol Evol. doi: 10.1038/s41559-021-01638-2.
Munson-McGee JH, et al. 2015. Nanoarchaeota, their sulfolobales host, and nanoarchaeota virus distribution across Yellowstone National Park hot springs. Appl Environ Microbiol. 81(22):7860–7868.
Möckl L, Moerner WE. 2020. Super-resolution microscopy with single molecules in biology and beyond-essentials, current trends, and future challenges. J Am Chem Soc. 142(42):17828–17844.
Müller M, et al. 2012. Biochemistry and evolution of anaerobic energy metabolism in eukaryotes. Microbiol Mol Biol Rev. 76(2):444–495.
Nakayama T, Ishida K. 2009. Another acquisition of a primary photosynthetic organelle is underway in Paulinella chromatophora. Curr Biol. 19(7):R284–R285.
Novoa RR, et al. 2005. Virus factories: associations of cell organelles for viral replication and morphogenesis. Biol Cell. 97(2):147–172.
Nowack EC, Melkonian M, Glockner G. 2008. Chromatophore genome sequence of Paulinella sheds light on acquisition of photosynthesis by eukaryotes. Curr Biol. 18(6):410–418.
Ohan J, et al. 2019. High-throughput phenotyping of cell-to-cell interactions in gel microdroplet pico-cultures. Biotechniques 66(5):218–224.
Oikonomou CM, Jensen GJ. 2017. Cellular electron cryotomography: toward structural biology in situ. Annu Rev Biochem. 86:873–896.
Olm MR, et al. 2021. inStrain profiles population microdiversity from metagenomic data and sensitively detects shared microbial strains. Nat Biotechnol. 39(6):727–736.
Omland KE. 1999. The assumptions and challenges of ancestral state reconstructions. System Biol. 48(3):604–611.
Orlando L, Gilbert MTP, Willerslev E. 2015. Reconstructing ancient genomes and epigenomes. Nat Rev Genet. 16(7):395–408.
Paez-Espino D, et al. 2016. Uncovering Earth's virome. Nature 536(7617):425–430.
Parfrey LW, Lahr DJG, Knoll AH, Katz LA. 2011. Estimating the timing of early eukaryotic diversification with multigene molecular clocks. Proc Natl Acad Sci USA. 108(33):13624–13629.
Parks DH, et al. 2018. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. Nat Biotechnol. 36(10):996–1004.
Parks DH, et al. 2017. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. Nat Microbiol. 2(11):1533–1542.
Payne JL, Wagner A. 2019. The causes of evolvability and their evolution. Nat Rev Genet. 20(1):24–38.
Peisajovich SG. 2012. Evolutionary synthetic biology. ACS Synth Biol. 1(6):199–210.
Pende N, et al. 2021. SepF is the FtsZ anchor in archaea, with features of an ancestral cell division system. Nat Commun. 12(1):3214.
Petitjean C, Deschamps P, López-García P, Moreira D. 2014. Rooting the domain archaea by phylogenomic analysis supports the foundation of the new kingdom Proteoarchaeota. Genome Biol Evol. 7(1):191–204.
Philippe H, Zhou Y, Brinkmann H, Rodrigue N, Delsuc F. 2005. Heterotachy and long-branch attraction in phylogenetics. BMC Evol Biol. 5:50.
Pigliucci M. 2008. Is evolvability evolvable? Nat Rev Genet. 9(1):75–82.
Pigliucci M. 2010. Genotype-phenotype mapping and the end of the 'genes as blueprint' metaphor. Philos Trans R Soc Lond B Biol Sci. 365(1540):557–566.
Pillai AS, et al. 2020. Origin of complexity in haemoglobin evolution. Nature 581(7809):480–485.
Pittis AA, Gabaldon T. 2016. Late acquisition of mitochondria by a host with chimaeric prokaryotic ancestry. Nature 531(7592):101–104.
Podar M, et al. 2013. Insights into archaeal evolution and symbiosis from the genomes of a nanoarchaeon and its inferred crenarchaeal host from Obsidian Pool, Yellowstone National Park. Biol Direct. 8(1):9.
Ponce-Toledo RI, et al. 2017. An early-branching freshwater cyanobacterium at the origin of plastids. Curr Biol. 27(3):386–391.

Pont C, et al. 2019. Paleogenomics: reconstruction of plant evolutionary trajectories from modern and ancient DNA. Genome Biol. 20(1):29.
Poole AM, Gribaldo S. 2014. Eukaryotic origins: how and when was the mitochondrion acquired? Cold Spring Harb Perspect Biol. 6(12):a015990.
Popa O, Dagan T. 2011. Trends and barriers to lateral gene transfer in prokaryotes. Curr Opin Microbiol. 14(5):615–623.
Prakadan SM, Shalek AK, Weitz DA. 2017. Scaling by shrinking: empowering single-cell 'omics' with microfluidic devices. Nat Rev Genet. 18(6):345–361.
Prangishvili D, et al. 2017. The enigmatic archaeal virosphere. Nat Rev Microbiol. 15(12):724–739.
Probst AJ, et al. 2014. Biology of a widespread uncultivated archaeon that contributes to carbon fixation in the subsurface. Nat Commun. 5:5497–5497.
Puigbò P, Wolf YI, Koonin EV. 2009. Search for a 'Tree of Life' in the thicket of the phylogenetic forest. J Biol. 8(6):59.
Quince C, et al. 2021. STRONG: metagenomics strain resolution on assembly graphs. Genome Biol. 22(1):214.
Rambo IM, de Anda V, Langwig MV, Baker BJ. 2021. Unique viruses that infect Archaea related to eukaryotes. bioRxiv. 2021.07.29.454249; doi:10.1101/2021.07.29.454249.
Raoult D, et al. 2004. The 1.2-megabase genome sequence of Mimivirus. Science 306(5700):1344–1350.
Raymann K, Forterre P, Brochier-Armanet C, Gribaldo S. 2014. Global phylogenomic analysis disentangles the complex evolutionary history of DNA replication in archaea. Genome Biol Evol. 6(1):192–212.
Rinke C, et al. 2013. Insights into the phylogeny and coding potential of microbial dark matter. Nature 499(7459):431–437.
Rinke C, et al. 2021. A standardized archaeal taxonomy for the genome taxonomy database. Nat Microbiol. 6(7):946–959.
Rochette NC, Brochier-Armanet C, Gouy M. 2014. Phylogenomic test of the hypotheses for the evolutionary origin of eukaryotes. Mol Biol Evol. 31(4):832–845.
Rodriguez-Brito B, Rohwer F, Edwards RA. 2006. An application of statistics to comparative metagenomics. BMC Bioinform. 7:162.
Roger AJ, Munoz-Gomez SA, Kamikawa R. 2017. The origin and diversification of mitochondria. Curr Biol. 27(21):R1177–R1192.
Roger AJ, Susko E, Leger MM. 2021. Evolution: reconstructing the timeline of eukaryogenesis. Curr Biol. 31(4):R193–R196.
Roux S, Hallam SJ, Woyke T, Sullivan MB. 2015. Viral dark matter and virus-host interactions resolved from publicly available microbial genomes. Elife 4:e08490.
Rubin BE, et al. 2022. Species- and site-specific genome editing in complex bacterial communities. Nat Microbiol. 7(1):34–47.
Russell MJ, Nitschke W. 2017. Methane: fuel or exhaust at the emergence of life? Astrobiology 17(10):1053–1066.
Sakai HD, et al. 2022. Insight into the symbiotic lifestyle of DPANN archaea revealed by cultivation and genome analyses. Proc Natl Acad Sci USA. 119:e2115449119.
Sakowski EG, et al. 2021. Interaction dynamics and virus-host range for estuarine actinophages captured by epicPCR. Nat Microbiol. 6(5):630–642.
Salichos L, Rokas A. 2013. Inferring ancient divergences requires genes with strong phylogenetic signals. Nature 497(7449):327–331.
Sanchez-Baracaldo P, Raven JA, Pisani D, Knoll AH. 2017. Early photosynthetic eukaryotes inhabited low-salinity habitats. Proc Natl Acad Sci USA. 114(37):E7737–E7745.
Santana-Molina C, Rivas-Marin E, Rojas AM, Devos DP. 2020. Origin and evolution of polycyclic triterpene synthesis. Mol Biol Evol. 37(7):1925–1941.
Schoepp-Cothenet B, et al. 2013. On the universal core of bioenergetics. Biochim Biophys Acta. 1827(2):79–93.
Schön ME, et al. 2021. Single cell genomics reveals plastid-lacking Picozoa are close relatives of red algae. Nat Commun. 12(1):6651.
Schulz F, et al. 2017. Giant viruses with an expanded complement of translation system components. Science 356(6333):82–85.
Schvarcz CR, Steward GF. 2018. A giant virus infecting green algae encodes key fermentation genes. Virology 518:423–433.
Schwille P, et al. 2018. MaxSynBio: avenues towards creating cells from the bottom up. Angew Chem Int Ed Engl. 57(41):13382–13392.
Seeger C, et al. 2021. The subcellular proteome of a planctomycetes bacterium shows that newly evolved proteins have distinct fractionation patterns. Front Microbiol. 12:643045.
Seitz KW, et al. 2019. Asgard archaea capable of anaerobic hydrocarbon cycling. Nat Commun. 10(1):1822.
Seitz KW, Lazar CS, Hinrichs KU, Teske AP, Baker BJ. 2016. Genomic reconstruction of a novel, deeply branched sediment archaeal phylum with pathways for acetogenesis and sulfur reduction. ISME J. 10(7):1696–1705.
Selberg AGA, Gaucher EA, Liberles DA. 2021. Ancestral sequence reconstruction: from chemical paleogenetics to maximum likelihood algorithms and beyond. J Mol Evol. 89(3):157–164.
Sharov AA. 2014. Evolutionary constraints or opportunities? Bio Systems. 120C:21–30.
Shih PM, et al. 2016. Biochemical characterization of predicted Precambrian RuBisCO. Nat Commun. 7:10382.
Shutt TE, Gray MW. 2006. Bacteriophage origins of mitochondrial replication and transcription proteins. Trends Genet. 22(2):90–95.
Sibbald SJ, Archibald JM. 2017. More protist genomes needed. Nat Ecol Evol. 1(5):145.
Sibbald SJ, Archibald JM. 2020. Genomic insights into plastid evolution. Genome Biol Evol. 12(7):978–990.
Sibbald SJ, Eme L, Archibald JM, Roger AJ. 2020. Lateral gene transfer mechanisms and pan-genomes in eukaryotes. Trends Parasitol. 36(11):927–941.
Siddiq MA, Hochberg GKA, Thornton JW. 2017. Evolution of protein specificity: insights from ancestral protein reconstruction. Curr Opin Struct Biol. 47:113–122.
Smith JM, et al. 1985. Developmental constraints and evolution: a perspective from the mountain lake conference on development and evolution. Q Rev Biol. 60(3):265–287.
Sojo V, Pomiankowski A, Lane N. 2014. A bioenergetic basis for membrane divergence in archaea and bacteria. PLoS Biol. 12(8):e1001926.
Som A. 2015. Causes, consequences and solutions of phylogenetic incongruence. Brief Bioinform. 16(3):536–548.
Sousa FL, et al. 2013. Early bioenergetic evolution. Philos Trans R Soc Lond B Biol Sci. 368(1622):20130088.
Spang A, Caceres EF, Ettema TJG. 2017. Genomic exploration of the diversity, ecology, and evolution of the archaeal domain of life. Science 357:eaaf3883.
Spang A, et al. 2015. Complex archaea that bridge the gap between prokaryotes and eukaryotes. Nature 521(7551):173–179.
Spang A, et al. 2018. Asgard archaea are the closest prokaryotic relatives of eukaryotes. PLoS Genet. 14(3):e1007080.
Spang A, et al. 2019. Proposal of the reverse flow model for the origin of the eukaryotic cell based on comparative analyses of Asgard archaeal metabolism. Nat Microbiol. 4(7):1138–1148.
St John E, et al. 2019. A new symbiotic nanoarchaeote (Candidatus *Nanoclepta minutus*) and its host (*Zestosphaera tikiterensis* gen. nov., sp. nov.) from a New Zealand hot spring. Syst Appl Microbiol. 42(1):94–106.
Stairs CW, Leger MM, Roger AJ. 2015. Diversity and origins of anaerobic metabolism in mitochondria and related organelles. Philos Trans R Soc Lond B Biol Sci. 370(1678):20140326.
Stairs CW, et al. 2020. Chlamydial contribution to anaerobic metabolism during eukaryotic evolution. Sci Adv. 6(35):eabb7258.

Stairs CW, et al. 2021. Anaeramoebae are a divergent lineage of eukaryotes that shed light on the transition from anaerobic mitochondria to hydrogenosomes. Curr Biol. 31(24):5605–5612.e5.
Strassert JFH, Irisarri I, Williams TA, Burki F. 2021. A molecular timescale for eukaryote evolution with implications for the origin of red algal-derived plastids. Nat Commun. 12(1):1879.
Strassert JFH, Jamy M, Mylnikov AP, Tikhonenkov DV, Burki F. 2019. New phylogenomic analysis of the enigmatic phylum telonemia further resolves the eukaryote tree of life. Mol Biol Evol. 36(4):757–765.
Surovtsev IV, Jacobs-Wagner C. 2018. Subcellular organization: a critical feature of bacterial cell replication. Cell 172(6):1271–1293.
Survery S, et al. 2021. Heimdallarchaea encodes profilin with eukaryotic-like actin regulation and polyproline binding. Commun Biol. 4(1):1024.
Suzan-Monti M, La Scola B, Barrassi L, Espinosa L, Raoult D. 2007. Ultrastructural characterization of the giant volcano-like virus factory of Acanthamoeba polyphaga Mimivirus. PLoS One. 2(3):e328.
Szöllosi GJ, Boussau B, Abby SS, Tannier E, Daubin V. 2012. Phylogenetic modeling of lateral gene transfer reconstructs the pattern and relative timing of speciations. Proc Natl Acad Sci USA. 109(43):17513–17518.
Szöllõsi GJ, Rosikiewicz W, Boussau B, Tannier E, Daubin V. 2013. Efficient exploration of the space of reconciled gene trees. Syst Biol. 62(6):901–912.
Taib N, et al. 2020. Genome-wide analysis of the Firmicutes illuminates the diderm/monoderm transition. Nat Ecol Evol. 4(12):1661–1672.
Takemura M. 2020. Medusavirus ancestor in a proto-eukaryotic cell: updating the hypothesis for the viral origin of the Nucleus. Front Microbiol. 11:571831.
Tamarit D, et al. 2021. A closed Candidatus Odinarchaeum genome exposes Asgard archaeal viruses. bioRxiv. 2021.09.01.458545; doi: https://doi.org/10.1101/2021.09.01.458545.
Thomson RC. Brown JM 2022. On the need for new measures of phylogenomic support. Syst Biol.
Thornton JW. 2004. Resurrecting ancient genes: experimental analysis of extinct molecules. Nat Rev Genet. 5(5):366–375.
Torruella G, et al. 2015. Phylogenomics reveals convergent evolution of lifestyles in close relatives of animals and fungi. Curr Biol. 25(18):2404–2410.
Tria FDK, et al. 2021. Gene duplications trace mitochondria to the onset of eukaryote complexity. Genome Biol Evol. 13:evab055.
Tuson HH, Biteen JS. 2015. Unveiling the inner workings of live bacteria using super-resolution microscopy. Anal Chem. 87(1):42–63.
Van den Bergh B, Swings T, Fauvart M, Michiels J. 2018. Experimental design, population dynamics, and diversity in microbial experimental evolution. Microbiol Mol Biol Rev. 82(3):e00008-18.
Vosseberg J, et al. 2021. Timing the origin of eukaryotic cellular complexity with ancient duplications. Nat Ecol Evol. 5(1):92–100.
Wagner GP, Zhang J. 2011. The pleiotropic structure of the genotype-phenotype map: the evolvability of complex organisms. Nat Rev Genet. 12(3):204–213.
Wang Y, Zhao Y, Bollas A, Wang Y, Au KF. 2021. Nanopore sequencing technology, bioinformatics and applications. Nat Biotechnol. 39(11):1348–1365.
Weinheimer AR, Aylward FO. 2020. A distinct lineage of Caudovirales that encodes a deeply branching multi-subunit RNA polymerase. Nat Commun. 11(1):4506.
Weiss MC, et al. 2016. The physiology and habitat of the last universal common ancestor. Nat Microbiol. 1(9):16116.
West GB, Brown JH, Enquist BJ. 1997. A general model for the origin of allometric scaling laws in biology. Science 276(5309): 122–126.
West GB, Woodruff WH, Brown JH. 2002. Allometric scaling of metabolic rate from molecules and mitochondria to cells and mammals. Proc Natl Acad Sci USA. 99(Suppl 1):2473–2478.
West PT, Probst AJ, Grigoriev IV, Thomas BC, Banfield JF. 2018. Genome-reconstruction for eukaryotes from complex natural microbial communities. Genome Res. 28(4):569–580.
Williams TA, Cox CJ, Foster PG, Szöllősi GJ, Embley TM. 2020. Phylogenomics provides robust support for a two-domains tree of life. Nat Ecol Evol. 4(1):138–147.
Williams TA, Embley TM, Heinz E. 2011. Informational gene phylogenies do not support a fourth domain of life for nucleocytoplasmic large DNA viruses. PLoS One. 6(6):e21080.
Williams TA, Foster PG, Cox CJ, Embley TM. 2013. An archaeal origin of eukaryotes supports only two primary domains of life. Nature 504(7479):231–236.
Williams TA, et al. 2017. Integrative modeling of gene and genome evolution roots the archaeal tree of life. Proc Natl Acad Sci USA. 114(23):E4602–E4611.
Williams TA, et al. 2021. Inferring the deep past from molecular data. Genome Biol Evol. 13(5):evab067.
Woese CR, Fox GE. 1977. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proc Natl Acad Sci USA. 74(11):5088–5090.
Woese CR, Kandler O, Wheelis ML. 1990. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proc Natl Acad Sci USA. 87(12):4576–4579.
Wolfe JM, Fournier GP. 2018. Horizontal gene transfer constrains the timing of methanogen evolution. Nat Ecol Evol. 2(5):897–903.
Wu F, et al. 2022. Unique mobile elements and scalable gene flow at the prokaryote-eukaryote boundary revealed by circularized Asgard archaea genomes. Nat Microbiol. 7(2):200–212.
Wurch L, et al. 2016. Genomics-informed isolation and characterization of a symbiotic Nanoarchaeota system from a terrestrial geothermal environment. Nat Commun. 7:12115.
Xavier JC, et al. 2021. The metabolic network of the last bacterial common ancestor. Commun Biol. 4(1):413.
Yazaki E, et al. 2021. Phylogenomics invokes the clade housing Cryptista, Archaeplastida, and *Microheliella maris*. bioRxiv. 2021.08.29.458128; doi:10.1101/2021.08.29.458128.
Yildirir G, et al. 2022. Long reads and Hi-C sequencing illuminate the two compartment genome of the model arbuscular mycorrhizal symbiont *Rhizophagus irregularis*. New Phytol. 233(3):1097–1107.
Yoshikawa G, et al. 2019. Medusavirus, a novel large DNA virus discovered from Hot Spring water. J Virol. 93(8). doi:10.1128/JVI.02130-18.
Youssef NH, et al. 2015. Insights into the metabolism, lifestyle and putative evolutionary history of the novel archaeal phylum 'Diapherotrites'. ISME J. 9(2):447–460.
Zahir T, et al. 2019. High-throughput time-resolved morphology screening in bacteria reveals phenotypic responses to antibiotics. Commun Biol. 2:269.
Zaremba-Niedzwiedzka K, et al. 2017. Asgard archaea illuminate the origin of eukaryotic cellular complexity. Nature 541(7637):353–358.
Zhang JW, et al. 2021. Newly discovered Asgard archaea Hermodarchaeota potentially degrade alkanes and aromatics via alkyl/benzyl-succinate synthase and benzoyl-CoA pathway. Isme J. 15(6):1826–1843.
Zhaxybayeva O, Lapierre P, Gogarten JP. 2005. Ancient gene duplications and the root(s) of the tree of life. Protoplasma 227(1):53–64.
Zhu Q, et al. 2019. Phylogenomics of 10,575 genomes reveals evolutionary proximity between domains Bacteria and Archaea. Nat Commun. 10(1):5477.

Associate editor: Tom Williams