HHS Public Access

Author manuscript

Nat Microbiol. Author manuscript; available in PMC 2024 June 01.

Published in final edited form as:

Nat Microbiol. 2023 June ; 8(6): 1008–1017. doi:10.1038/s41564-023-01378-y.

The virome of the last eukaryotic common ancestor and eukaryogenesis

Mart Krupovic[1,✉], **Valerian V. Dolja**[2], **Eugene V. Koonin**[3,✉]

[1]Institut Pasteur, Université Paris Cité, CNRS UMR6047, Archaeal Virology Unit, Paris, France.

[2]Department of Botany and Plant Pathology, Oregon State University, Corvallis, OR, USA.

[3]National Center for Biotechnology Information, National Library of Medicine, Bethesda, MD, USA.

Abstract

All extant eukaryotes descend from the last eukaryotic common ancestor (LECA), which is thought to have featured complex cellular organization. To gain insight into LECA biology and eukaryogenesis—the origin of the eukaryotic cell, which remains poorly understood—we reconstructed the LECA virus repertoire. We compiled an inventory of eukaryotic hosts of all major virus taxa and reconstructed the LECA virome by inferring the origins of these groups of viruses. The origin of the LECA virome can be traced back to a small set of bacterial—not archaeal—viruses. This provenance of the LECA virome is probably due to the bacterial origin of eukaryotic membranes, which is most compatible with two endosymbiosis events in a syntrophic model of eukaryogenesis. In the first endosymbiosis, a bacterial host engulfed an Asgard archaeon, preventing archaeal viruses from entry owing to a lack of archaeal virus receptors on the external membranes.

Eukaryotes differ from archaea and bacteria due to their complex cellular organization. This includes endomembranes (in particular, the nuclear compartment), a complex cytoskeleton and the mitochondrion, which itself evolved from an alphaproteobacterial endosymbiont[1–3]. All of these features seem to be traceable to the last eukaryotic common ancestor (LECA)[2,4]. Several models for the origin of eukaryotes (eukaryogenesis) have been proposed, but each differs with respect to the timing of the origin of the typical eukaryotic cellular organization[5–7]. Phylogenomic analyses indicate that eukaryotes possess a mix of genes originating from archaea (in particular, *Asgardarchaeota*) and genes of apparent

✉**Correspondence** should be addressed to Mart Krupovic or Eugene V. Koonin. mart.krupovic@pasteur.fr; koonin@ncbi.nlm.nih.gov.

Author contributions

M.K., V.V.D. and E.V.K. researched and analysed the data and wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41564-023-01378-y.

Peer review information *Nature Microbiology* thanks Susanne Erdmann, Jeremy Wideman and the other, anonymous, reviewer(s) for their contribution to the peer review of this work

bacterial origin[8–12]. This dichotomy among eukaryotic genes is thought to reflect the symbiotic origin of eukaryotes. However, whereas the origin of the mitochondria from an alphaproteobacterium appears indisputable, the nature of the host of the proto-mitochondrial endosymbiont remains uncertain. The most straightforward models suggest an Asgard archaeon as the host[2,5] (Fig. 1). However, such scenarios of eukaryogenesis are incompatible with the chemistry of cell membranes and the enzymology of membrane biosynthesis, which are unrelated in archaea and bacteria, as membranes in eukaryotes are of the bacterial type[13]. Thus, any eukaryogenesis scenario with an archaeal host would require a membrane replacement step. Alternatively, more complex models of eukaryogenesis propose that an Asgard archaeon was engulfed by a bacterium, followed by the loss of the archaeal membrane, then a second endosymbiosis that gave rise to mitochondria[6,14,15] (Fig. 1).

All life forms are hosts to viruses and other mobile genetic elements (MGEs), which can be either parasites or mutualists[16]. Recent phylogenomic efforts yielded an evolutionary taxonomy that encompasses most known viruses[16]. This taxonomic system comprises six realms: *Riboviria*, *Monodnaviria*, *Duplodnaviria*, *Varidnaviria*, *Adnaviria* and *Ribozyviria*[16,17]. Unifying features in each group include hallmark proteins involved in genome replication (such as homologous RNA-dependent RNA polymerases (RdRPs) and reverse transcriptases) or virion formation (namely, distinct varieties of major capsid proteins and enzymes involved in genome encapsidation). Notably, there are major differences between the virome compositions of bacteria, archaea and eukaryotes (Box 1 and Fig. 2).

We previously reconstructed the complex virome of the last universal cellular ancestor (LUCA) and found that *Varidnaviria* and *Duplodnaviria* (and possibly *Riboviria* and *Monodnaviria*) evolved at early stages of life predating the LUCA[18]. Given that viruses are obligate intracellular parasites that intimately interact with various components of the host cell—in particular, with cell membranes—analysis of virome composition can inform our understanding of host cell biology.

We sought to gain insight into the virome of the LECA and its evolutionary origin. We reconstructed the LECA virome and traced its origins to bacterial viruses, but did not find any links to archaeal viruses. We also address the implications of our reconstructed LECA virome for eukaryogenesis.

Reconstruction of the LECA virome

Using the reported host ranges of different virus groups across the branches of the eukaryotic phylogenetic tree[19], we reconstructed the LECA virome (Supplementary Table 1). In this reconstruction, we faced two major obstacles. First, viromes are sparsely sampled across the diversity of eukaryotes, especially among unicellular organisms that encompass most of that diversity and are themselves not uniformly sampled[19] (Fig. 2). Nevertheless, there have been considerable advances in characterizing the viromes of unicellular eukaryotes through metagenomics and metatranscriptomics[20–26]. Second, ancestral reconstruction is confounded by horizontal virus transfer among diverse host lineages, whereby viruses change hosts (for example, between animals and plants via

vectors such as insects or nematodes, or between plants and fungi via direct interaction)[27]. Given these obstacles, we did not attempt formal, maximum likelihood reconstruction approaches[28,29], but rather, applied a semi-formal, parsimony-based approach, as we did previously for the LUCA virome[18]. We assumed that a group of viruses could be assigned to the LECA virome if it was represented in three or more of the six supergroups of eukaryotes (Fig. 1). Information on virus host range was extracted from the published literature using keyword searches (Supplementary Table 1). Considering the uncertainty of the topology in the deepest branchings of the eukaryotic tree[19], we surmised that this simple majority rule approach would result in the most realistic approximation of the LECA virome.

The realm (the top rank in virus taxonomy) *Riboviria* is broadly represented in the LECA virome. Within *Riboviria*, all five phyla in the *Orhtornavirae* kingdom are widely spread across the tree and map back to the LECA. Moreover, the diversification of some ribovirus phyla seems to pre-date the LECA (Supplementary Table 1). One potential caveat is the possibility of horizontal virus transfer over long evolutionary distances, such as between plants and animals, particularly in the case of the ribovirus phylum *Negarnaviricota*[27]. In the kingdom *Pararnavirae*, two virus families, *Metaviridae* and *Pseudoviridae* (also known as Ty3/Gypsy and Ty1/Copia retrotransposons, respectively), are nearly ubiquitous among eukaryotes and confidently map back to the LECA (Fig. 1 and Supplementary Table 1). The remaining pararnaviruses, however, appear to have evolved later, in animals and plants (Supplementary Table 1).

In the realm *Monodnaviria*, two phyla, *Cressdnaviricota*[30] and *Cossaviricota*, are represented in eukaryotes. The *Cressdnaviricota* are broadly distributed and trace back to the LECA (Fig. 1 and Supplementary Table 1). In contrast, the phylum *Cossaviricota* consists of several groups of single-stranded DNA (ssDNA) viruses and viruses with small double-stranded DNA (dsDNA) genomes with relatively narrow host ranges confined to animals, and thus appears to have evolved post-LECA (Fig. 1 and Supplementary Table 1).

In the realm *Varidnaviria*, the phylum *Nucleocytoviricota* is exclusive to eukaryotes and widespread across the eukaryote diversity, tracing back to the LECA (Fig. 1 and Supplementary Table 1). In the phylum *Preplasmiviricota*, eukaryotic viruses are represented by a diverse group of endogenous viruses known as polintons or polintoviruses[31], virophages[32] and the currently unclassified polinton-like viruses[33]. These viruses are widespread in eukaryotes and probably belong to the LECA heritage (Fig. 1 and Supplementary Table 1). In contrast, adenoviruses that belong to the same phylum are limited in their host range to animals and seem to be a late (that is, postdating the origin of animals) derivative of polintons[31].

The realm *Duplodnaviria* is dominated by bacterial and archaeal viruses. Until recently, the phylum *Peploviricota*, which includes herpesviruses, was the only group of eukaryotic viruses in this realm and appeared to be limited to animal hosts. However, the recent discovery of mirusviruses that are expected to be assigned to *Duplodnaviria*[34], given the presence of the corresponding hallmark structural proteins, changed the picture. The host range of mirusviruses has not been directly characterized but probably includes unicellular

eukaryotes, suggesting that duplodnaviruses were represented in the LECA virome (Fig. 1 and Supplementary Table 1).

Similarly, the recent expansion of the previously tiny realm *Ribozyviria*[35,36], to include viruses probably infecting diverse protists, suggests the possibility that the diversity of this realm has been largely overlooked and brings into question its origin in animals. We cannot rule out the presence of ribozyviruses in LECA, although under the criteria adopted here they were not included (Fig. 1).

There are similarities between the LECA and LUCA viromes. The viromes of both common ancestors are complex and include representatives of most viral phyla. This finding leads us to propose that the time between the origin of eukaryotes and the advent of the LECA involved extensive diversification of the virosphere, concomitant with the evolution of distinct architecture of the eukaryotic cell[37].

Next, we discuss potential origins of the eukaryotic virome, through a process that we name eukaryovirogenesis, in the context of specific models of eukaryogenesis.

Bacterial origins of the LECA virome

The information-processing systems of the eukaryotic cell—replication, transcription and translation machineries—evolved from cognate systems in archaea[38]. Given that viruses and other MGEs are informational parasites, it seems plausible that the eukaryotic virome might have evolved from the archaeal virome. Recently, the origins of eukaryotic information systems, along with many cytoskeleton and endomembrane components, were traced to the archaeal phylum *Asgardarchaeota*[10–12], which includes the closest archaeal relatives to eukaryotes. In the best-supported phylogenies of universal genes, eukaryotes branched from within *Asgardarchaeota*. Owing to this evolutionary relationship, viruses of Asgard archaea may be possible ancestors of the viruses of eukaryotes. However, analyses of several families of viruses associated with *Asgardarchaeota* did not find support for any of these viruses being candidates for ancestors of known eukaryotic viruses[39–41]. The proposed relationship between some of the Asgard viruses and *Nucleocytoviricota*[41] stems entirely from the generic homology among proteins involved in DNA replication and nucleotide metabolism. These proteins are common in diverse viruses with large DNA genomes, particularly in other archaeal *Caudoviricetes*[42], as well as cell-based organisms. Indeed, the fraction of *Nucleocytoviricota* gene homologues in Asgard archaeal viruses is not higher than in bacteriophages or non-Asgard archaeal viruses[41]. The shared presence of these widespread genes does not reflect a common origin of the Asgard viruses and eukaryotic *Nucleocytoviricota*, and similarly, there is no specific relationship traceable between any viruses of eukaryotes and viruses of other archaea. Nevertheless, the Asgard archaeal virome is as-yet sparsely sampled, so it cannot be ruled out that uncharacterized archaeal viruses seeded some part of the eukaryotic virome; furthermore, extensive study of the Asgard virome will be needed to address this possibility.

In contrast, bacterial roots were detected for the eukaryote-infecting viruses from all four virus realms, as reported by previous studies on the evolution of each of the realms (Fig. 3). In the phylogenetic trees of the RdRPs of the kingdom *Orthornavirae*, the deepest

branch is the phylum *Lenarviricota*, which consists of bacterial leviviruses and their direct descendants infecting a broad range of eukaryotes[43–45]. The evolutionary scenario for this phylum can be readily reconstructed (Fig. 3): initially, an ancestral levivirus lost its capsid protein gene, giving rise to eukaryotic capsidless RNA replicators in the classes *Amabiliviricetes* and *Howeltoviricetes*, with the latter replicating in the mitochondria. The *Amabiliviricetes* subsequently gave rise to *Miaviricetes*, the largest group within *Orthornavirae*[45], by capturing the single jelly-roll (SJR) capsid protein gene (the most common capsid protein among viruses of eukaryotes thought to originally derive from a host sugar-binding protein[46], probably from an RNA virus of the phylum *Kitrinoviricota*). The progenitor of the remaining four phyla of *Orthornavirae* apparently originated from a common ancestor with *Lenarviricota* and followed a similar evolutionary path whereby the RdRP was inherited from a bacterial ancestor but the levivirus capsid protein was replaced with structurally unrelated proteins. In line with the general trend in virus evolution, the origin of eukaryote-infecting orthornaviruses seems to have involved preservation of the ancestral replication machinery that ultimately combined with eukaryote-specific virion proteins. It should be emphasized that, albeit with incomplete sampling, no RNA viruses of archaea have been discovered so far, leaving the bacterial origin of the eukaryotic RNA virome as the only viable scenario at the time of writing.

Pararnavirae is the only possible exception to the bacterial origin of the eukaryotic virome because archaeal origin cannot be ruled out. The reverse transcriptase of the pararnaviruses was reported to be inherited from prokaryotic group II introns (retrotransposons), which are broadly represented in both bacteria and archaea. Indeed, phylogenetic analyses do not unequivocally link pararnaviruses with either bacterial or archaeal ancestors[47]. The general pathway of evolution, though, seems to be the same, whether from a bacterial or archaeal retrotransposon, whereby the ancestral reverse-transcribing virus evolved by recruiting cellular proteins for the functions of capsid proteins, nucleocapsids and virus proteases, at a pre-LECA stage of eukaryogenesis[48,49] (Fig. 3).

The ancestral eukaryotic group in *Monodnaviria*, the cressdna-viruses, as in the case for pararnaviruses, evolved from non-viral bacterial MGEs—in this case, small bacterial (but not archaeal) plasmids replicating via the rolling circle mechanism[50]. These plasmids provided the genome scaffold and the gene encoding the endonuclease and superfamily 3 helicase (a signature of eukaryotic cressdnaviruses) required for replication initiation, whereas the capsid protein was apparently acquired via recombination with complementary DNA copies of ribovirus SJR capsid protein genes (Fig. 3).

The phylum *Preplasmiviricota* and specifically polintoviruses in the realm *Varidnaviria* appear to be direct descendants of tailless bacteriophages belonging to this phylum[31]. The origin of the phylum *Nucleocytoviricota* probably involved recombination between a mirusvirus-like duplodnavirus (from which the replication machinery of the nucleocytoviruses was inherited) and a polintovirus that donated the structural module replacing the duplodnavirus morphogenetic genes (Fig. 3).

Finally, the structural gene module of mirusviruses, which probably comprise the ancestral group from the realm *Duplodnaviria* infecting eukaryotic hosts, clearly derives from the

homologous genes of tailed bacteriophages or archaeal viruses, which comprise the class *Caudoviricetes* within this realm[34].

The above findings led us to propose two key features of eukaryovirogenesis that seem to have occurred concomitantly with eukaryogenesis itself (Fig. 2). First, all diverse groups of viruses comprising the LECA virome evolved from ancestral bacterial viruses, or non-viral MGEs, with the only possible exception being pararnaviruses. Second, the provenances of the genes encoding the components of the replication apparatus and those encoding virion components are markedly different, recapitulating the key trend of primordial virogenesis, where the replication machinery is thought to descend from the primordial pool of replicators whereas the structural proteins were apparently captured from the host at early stages of cellular evolution[18].

The LECA virome and eukaryogenesis

A bacterial origin for the LECA virome demands explanation, given the archaeal origin of the eukaryotic information-processing systems. A key feature that is probably relevant for eukaryovirogenesis and that links eukaryotes to bacteria rather than archaea is the bacterial provenance of eukaryotic cell membranes[13]. Bacterial and eukaryotic membranes are based on glycerol-3-phosphate ester linked to fatty acids, whereas archaeal membranes comprise glycerol-1-phosphate ether bound to isoprenoids[51]. Although eukaryotes inherited archaeal pathways for the biosynthesis of isoprenoids, these are not major structural components of eukaryotic membranes[52].

For any model of eukaryogenesis, the disparity between bacterial and eukaryotic membranes on the one hand and archaeal membranes on the other is a major challenge. The simplest symbiogenetic scenarios[5], which involve an archaeal host and an alphaproteobacterial proto-mitochondrial endosymbiont as the only two partners in eukaryogenesis, face the difficulty of replacing the archaeal membrane with the bacterial one, for which there is no known precedent. The bacterial provenance of the LECA virome seems more compatible with alternative eukaryogenesis models, in which the emerging protoeukaryote never had an archaeal plasma membrane[14,15,53]. Initially motivated by the plausibility of a metabolic symbiosis between a hydrogen-producing bacterium (possibly, a deltaproteobacterium) and a methanogenic archaeon, these syntrophy scenarios received a major boost with the recent demonstration of the syntrophic relationship between an Asgard archaeon and a deltaproteobacterium[54–56]. In an updated syntrophy model[15], eukaryogenesis started as a metabolic ectosymbiosis between a sulfate-reducing deltaproteobacterium and a hydrogen-producing Asgard archaeon, which was subsequently internalized and lost its membrane, probably after the emergence of bacterial endomembranes that surrounded the engulfed archaeon. The internalized archaeon became the progenitor of the eukaryotic nucleus (Figs. 1 and 4). This model implies two-stage eukaryogenesis, in which the merger between a (deltaproteo)bacterial host and an archaeal endosymbiont gave rise to an intermediate—the first eukaryotic common ancestor (FECA). This first endosymbiotic event was followed by the secondary endosymbiosis, whereby the FECA gave rise to the second eukaryotic common ancestor (SECA) by capturing a versatile sulfur-oxidizing and facultatively aerobic alphaproteobacterium that became the mitochondrion (Fig. 4). This eukaryogenesis scenario

is buttressed by the wide spread of serial endosymbiosis in the subsequent evolution of eukaryotes, of which the evolution of chloroplasts from cyanobacteria in the ancestor of Archaeplastida is only one example[57,58]. Furthermore, this scenario is compatible with phylogenomic analysis indicating that alphap roteobacterial proteins were acquired relatively late in the evolution of eukaryotes[59,60]. Notably, the enzymes of the biosynthetic pathway for steroids, which are essential components of eukaryotic membranes, have clear deltaproteobacterial origin[61].

Under the syntrophy scenario, the LECA virome was shaped by two waves of adaptation of bacterial viruses: first from the deltaproteobacterial virome and then from the virome of the alphaproteobacterial proto-mitochondrion (Fig. 4). Crucially, in this scenario, the emerging eukaryotic cell maintained the bacterial membranes through all stages of eukaryogenesis, whereas the archaeal membrane of the primary endosymbiont was lost. Thus, the viruses of the Asgard archaeon were excluded during the first stage of eukaryogenesis, primarily due to the inaccessibility of the archaeal virus receptors following the internalization of the archaeal symbiont. The escape from viruses would facilitate the endogenization of the archaeal symbiont en route to the FECA, jumpstarting eukaryogenesis. The bacterial provenance of the eukaryote virome buttresses models of eukaryogenesis that postulate the evolutionary continuity of bacterial membranes, such as the syntrophy scenario.

Possible stages in eukaryovirogenesis

Although it is difficult to assign the origins of specific virus groups in the LECA virome to one of the two bacterial partners, there are clues for this assignment (Figs. 3 and 4). Starting with *Orthornavirae*, the evolution of *Lenarviricota* from alphaproteobacterial leviviruses seems most likely, given the mitochondrial replication site of mitoviruses, which are direct eukaryotic descendants of leviviruses[62,63]. Thus, the eukaryotic members of *Lenarviricota* apparently evolved along the path from SECA to LECA (Fig. 4). The origins of the rest of the eukaryotic riboviruses are much less clear, but could be more ancient considering the topology of the phylogenetic tree of the RdRP where the first split is between *Lenarviricota* and the rest of *Orthornavirae*[45]. The common ancestor of the four phyla of *Orthornavirae*, which consist primarily of eukaryotic viruses, emerged en route from the FECA to the SECA, from an RNA virus of deltaproteobacteria—either a levivirus or an unknown ancestral virus (Fig. 4). Additionally, the origin of the eukaryotic riboviruses from a deltaproteobacterial rather than alphaproteobacterial ancestor appears likely because this scenario does not require virus escape from the endosymbiont. The virome of deltaproteobacteria has barely been sampled, suggesting that ancestors of riboviruses might not yet be discovered. The origin of eukaryotic riboviruses was precipitated by the exaptation of a cellular protein as SJR capsid protein.

The provenance of *Pararnavirae* remains uncertain, with possibilities being an origin in group II introns of the Asgard archaeal endosymbiont or any of the bacterial partners. *Pararnavirae* are the only major group of viruses of eukaryotes to which the exclusion of archaeal ancestor due to membrane incompatibility does not apply, because these viruses apparently originated from non-viral MGEs at a relatively late stage of eukaryogenesis. *Cressdnaviricota*, the dominant phylum of *Monodnaviria* in eukaryotes[30], probably evolved

from bacterial, potentially deltaproteobacterial, plasmids through the acquisition of capsid protein genes (possibly from RNA viruses) during the transition from the FECA to the SECA (Fig. 4).

The topology of the phylogenetic tree of the protein-primed DNA polymerase (the hallmark protein of *Preplasmiviricota*), where the first split in the eukaryotic portion of the tree is between mitochondrial linear dsDNA plasmids and eukaryotic viruses[31], implies an alphaproteobacterial origin of the eukaryotic members of this phylum en route from the SECA to the LECA. Notably, the host range of contemporary tectiviruses includes alphaproteobacteria[64]. Some of the descendants of an alphaproteobacterial tectivirus lost the capsid protein genes and became the mitochondrial plasmids, whereas others migrated to the proto-nucleus as polintons and then gave rise to the other eukaryotic lineages of *Preplasmiviricota* (polinton-like viruses, adenoviruses, virophages and linear cytoplasmic plasmids).

The ancestor of the mirusviruses was probably a deltaproteobacterial phage that gave rise to '*Mirusviricota*' (and eventually, to *Peploviricota* in animals) in protoeukaryotes en route from the FECA to the SECA and then to *Nucleocytoviricota* through recombination with *Preplasmiviricota* that donated the structural gene module replacing that of duplodnaviruses. Given the apparent alphaproteobacterial origin of the eukaryotic members of *Preplasmiviricota*, the origin of *Nucleocytoviricota* should be associated with the SECA to LECA stage (Fig. 4). Along a similar yet opposite route, the replication module of preplasmiviruses recombined with the morphogenetic module of parvoviruses giving rise to the *Mouviricetes* within the phylum *Cossaviricota*[65], underscoring the importance of module shuffling during diversification of the eukaryotic virome.

Outlook

The recent expansion of characterized diversity in all realms of viruses enables far more robust reconstruction of ancestral viromes than was previously possible. By examining the distributions of viruses infecting members of different eukaryotic clades across the evolutionary tree of eukaryotes, we propose that each of the four major virus realms was probably already represented by multiple groups in the LECA virome. The principal diversification of the eukaryotic virome apparently occurred during the relatively short time separating the origin of protoeukaryotes via endosymbiosis and the advent of the LECA, which already resembled extant unicellular eukaryotes. For each major virus group, with the possible exception of reverse-transcribing viruses, an origin from bacterial viruses or non-viral bacterial MGEs such as rolling circle plasmids is readily traceable. Although some of the corresponding virus groups can be traced back to the LUCA virome[18], all evidence points to eukaryotes inheriting the bacterial rather than archaeal descendants of these ancient viruses.

We propose that eukaryovirogenesis involved extensive diversification of virus genomes: in particular, the replacement of structural gene modules. One caveat to this proposal is that the archaeal virosphere, and particularly the Asgard viruses, remain undersampled. When

new groups of archaeal viruses are discovered, we will need to incorporate them into our analyses.

The bacterial origin of the LECA virome is most compatible with a model of eukaryogenesis in which the emerging protoeukaryote maintained the bacterial membrane through two stages of symbiogenesis. The first stage probably involved a (deltaproteo)bacterium engulfing an Asgard archaeal endosymbiont, which gave rise to the nucleus. The second stage probably featured the capture of an alphaproteobacterium by the archaeo-bacterial chimera to form the mitochondrion. This evolutionary continuity of bacterial membranes during eukaryogenesis would have caused exclusion of viruses specific for the archaeal endosymbiont from the evolving eukaryotic cell due to lack of archaeal virus receptors on the bacterial membranes. We further propose that the origins of the major groups of eukaryotic viruses can be tentatively assigned to one of the two steps in this endosymbiotic eukaryogenesis scenario. It seems that viruses of eukaryotes have either deltaproteobacterial or alphaproteobacterial origins and the diversification of virus genomes during eukaryovirogenesis involved multiple recombination events between these two groups of viruses. Viruses of both deltaproteobacteria and alphaproteobacteria have been poorly sampled, so in-depth study of these viromes should shed new light on eukaryovirogenesis.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank P. López-García for invaluable, inspiring discussions and critical reading of the manuscript and S. Roux for helpful advice. E.V.K. is supported by funds from the Intramural Research Program of the National Institutes of Health (National Library of Medicine). V.V.D. was partially supported by a National Institutes of Health/National Library of Medicine/National Center for Biotechnology Information Visiting Scientist Fellowship. M.K. was supported by l'Agence Nationale de la Recherche grant ANR-21-CE11–0001-01.

References

1. Alberts B et al. Molecular Biology of the Cell 6th edn (Garland Science, 2022).
2. Gabaldon T Origin and early evolution of the eukaryotic cell. Annu. Rev. Microbiol. 75, 631–647 (2021). [PubMed: 34343017]
3. Munoz-Gomez SA et al. Site-and-branch-heterogeneous analyses of an expanded dataset favour mitochondria as sister to known alphaproteobacteria. Nat. Ecol. Evol. 6, 253–262 (2022). [PubMed: 35027725]
4. Nobs SJ, MacLeod FI, Wong HL & Burns BP Eukarya the chimera: eukaryotes, a secondary innovation of the two domains of life? Trends Microbiol. 30, 421–431 (2022). [PubMed: 34863611]
5. Embley TM & Martin W Eukaryotic evolution, changes and challenges. Nature 440, 623–630 (2006). [PubMed: 16572163]
6. Lopez-Garcia P & Moreira D Open questions on the origin of eukaryotes. Trends Ecol. Evol. 30, 697–708 (2015). [PubMed: 26455774]
7. Lopez-Garcia P, Eme L & Moreira D Symbiosis in eukaryotic evolution. J. Theor. Biol. 434, 20–33 (2017). [PubMed: 28254477]
8. Koonin EV et al. A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. Genome Biol. 5, R7 (2004). [PubMed: 14759257]

9. Spang A, Spang EF & Ettema TJG Genomic exploration of the diversity, ecology, and evolution of the archaeal domain of life. Science 357, eaaf3883 (2017). [PubMed: 28798101]
10. Spang A et al. Complex archaea that bridge the gap between prokaryotes and eukaryotes. Nature 521, 173–179 (2015). [PubMed: 25945739]
11. Zaremba-Niedzwiedzka K et al. Asgard archaea illuminate the origin of eukaryotic cellular complexity. Nature 541, 353–358 (2017). [PubMed: 28077874]
12. Liu Y et al. Expanded diversity of Asgard archaea and their relationships with eukaryotes. Nature 593, 553–557 (2021). [PubMed: 33911286]
13. Lombard J, Lopez-Garcia P & Moreira D The early evolution of lipid membranes and the three domains of life. Nat. Rev. Microbiol. 10, 507–515 (2012). [PubMed: 22683881]
14. Lopez-Garcia P & Moreira D Metabolic symbiosis at the origin of eukaryotes. Trends Biochem. Sci. 24, 88–93 (1999). [PubMed: 10203753]
15. Lopez-Garcia P & Moreira D The syntrophy hypothesis for the origin of eukaryotes revisited. Nat. Microbiol. 5, 655–667 (2020). [PubMed: 32341569]
16. Koonin EV et al. Global organization and proposed megataxonomy of the virus world. Microbiol. Mol. Biol. Rev. 84, e00061–19 (2020). [PubMed: 32132243]
17. International Committee on Taxonomy of Viruses Executive Committee The new scope of virus taxonomy: partitioning the virosphere into 15 hierarchical ranks. Nat. Microbiol. 5, 668–674 (2020). [PubMed: 32341570]
18. Krupovic M, Dolja VV & Koonin EV The LUCA and its complex virome. Nat. Rev. Microbiol. 18, 661–670 (2020). [PubMed: 32665595]
19. Burki F, Roger AJ, Brown MW & Simpson AGB The new tree of eukaryotes. Trends Ecol. Evol. 35, 43–55 (2020). [PubMed: 31606140]
20. Lachnit T, Thomas T & Steinberg P Expanding our understanding of the seaweed holobiont: RNA viruses of the red alga *Delisea pulchra*. Front. Microbiol. 6, 1489 (2015). [PubMed: 26779145]
21. Paez-Espino D et al. Uncovering Earth's virome. Nature 536, 425–430 (2016). [PubMed: 27533034]
22. Roux S et al. Ecogenomics of virophages and their giant virus hosts assessed through time series metagenomics. Nat. Commun. 8, 858 (2017). [PubMed: 29021524]
23. Paez-Espino D et al. Diversity, evolution, and classification of virophages uncovered through global metagenomics. Microbiome 7, 157 (2019). [PubMed: 31823797]
24. Kinsella CM et al. *Entamoeba* and *Giardia* parasites implicated as hosts of CRESS viruses. Nat. Commun. 11, 4620 (2020). [PubMed: 32934242]
25. Charon J, Murray S & Holmes EC Revealing RNA virus diversity and evolution in unicellular algae transcriptomes. Virus Evol. 7, veab070 (2021). [PubMed: 36819971]
26. Schulz F, Abergel C & Woyke T Giant virus biology and diversity in the era of genome-resolved metagenomics. Nat. Rev. Microbiol. 20, 721–736 (2022). [PubMed: 35902763]
27. Dolja VV & Koonin EV Metagenomics reshapes the concepts of RNA virus evolution by revealing extensive horizontal virus transfer. Virus Res. 244, 36–52 (2018). [PubMed: 29103997]
28. Csuros M & Miklos I Streamlining and large ancestral genomes in archaea inferred with a phylogenetic birth-and-death model. Mol. Biol. Evol. 26, 2087–2095 (2009). [PubMed: 19570746]
29. Cohen O, Ashkenazy H, Belinky F, Huchon D & Pupko T GLOOME: gain loss mapping engine. Bioinformatics 26, 2914–2915 (2010). [PubMed: 20876605]
30. Krupovic M et al. *Cressdnaviricota*: a virus phylum unifying seven families of Rep-encoding viruses with single-stranded, circular DNA genomes. J. Virol. 94, 00582–20 (2020).
31. Krupovic M & Koonin EV Polintons: a hotbed of eukaryotic virus, transposon and plasmid evolution. Nat. Rev. Microbiol. 13, 105–115 (2015). [PubMed: 25534808]
32. Fischer MG The virophage family *Lavidaviridae*. Curr. Issues Mol. Biol. 40, 1–24 (2021). [PubMed: 32089519]
33. Yutin N, Shevchenko S, Kapitonov V, Krupovic M & Koonin EV A novel group of diverse polinton-like viruses discovered by metagenome analysis. BMC Biol. 13, 95 (2015). [PubMed: 26560305]

34. Gaïa M et al. Mirusviruses link herpesviruses to giant viruses. Nature 10.1038/s41586-023-05962-4 (2023).
35. Forgia M et al. Extant hybrids of RNA viruses and viroid-like elements. Preprint at *bioRxiv* 10.1101/2022.08.21.504695 (2022).
36. Lee BD et al. Mining metatranscriptomes reveals a vast world of viroid-like circular RNAs. Cell 186, 646–661.e4 (2023). [PubMed: 36696902]
37. Koonin EV, Senkevich TG & Dolja VV The ancient Virus World and evolution of cells. Biol. Direct 1, 29 (2006). [PubMed: 16984643]
38. Brown JR & Doolittle WF Archaea and the prokaryote-to-eukaryote transition. Microbiol. Mol. Biol. Rev. 61, 456–502 (1997). [PubMed: 9409149]
39. Medvedeva S et al. Three families of Asgard archaeal viruses identified in metagenome-assembled genomes. Nat. Microbiol. 7, 962–973 (2022). [PubMed: 35760839]
40. Tamarit D et al. A closed Candidatus Odinarchaeum chromosome exposes Asgard archaeal viruses. Nat. Microbiol. 7, 948–952 (2022). [PubMed: 35760836]
41. Rambo IM, Langwig MV, Leao P, De Anda V & Baker BJ Genomes of six viruses that infect Asgard archaea from deep-sea sediments. Nat. Microbiol. 7, 953–961 (2022). [PubMed: 35760837]
42. Liu Y et al. Diversity, taxonomy, and evolution of archaeal viruses of the class *Caudoviricetes*. PLoS Biol. 19, e3001442 (2021). [PubMed: 34752450]
43. Wolf YI et al. Origins and evolution of the global RNA virome. mBio 9, e02329–18 (2018). [PubMed: 30482837]
44. Wolf YI et al. Doubling of the known set of RNA viruses by metagenomic analysis of an aquatic virome. Nat. Microbiol. 5, 1262–1270 (2020). [PubMed: 32690954]
45. Neri U et al. Expansion of the global RNA virome reveals diverse clades of bacteriophages. Cell 185, 4023–4037.e18 (2022). [PubMed: 36174579]
46. Krupovic M & Koonin EV Multiple origins of viral capsid proteins from cellular ancestors. Proc. Natl Acad. Sci. USA 114, E2401–E2410 (2017). [PubMed: 28265094]
47. Gladyshev EA & Arkhipova IR A widespread class of reverse transcriptase-related cellular genes. Proc. Natl Acad. Sci. USA 108, 20311–20316 (2011). [PubMed: 21876125]
48. Krupovic M & Koonin EV Homologous capsid proteins testify to the common ancestry of retroviruses, caulimoviruses, pseudoviruses and metaviruses. J. Virol. 91, e00210–17 (2017). [PubMed: 28356531]
49. Koonin EV, Dolja VV & Krupovic M The logic of virus evolution. Cell Host Microbe 30, 917–929 (2022). [PubMed: 35834963]
50. Kazlauskas D, Varsani A, Koonin EV & Krupovic M Multiple origins of prokaryotic and eukaryotic single-stranded DNA viruses from bacterial and archaeal plasmids. Nat. Commun. 10, 3425 (2019). [PubMed: 31366885]
51. Villanueva L, Damste JS & Schouten S A re-evaluation of the archaeal membrane lipid biosynthetic pathway. Nat. Rev. Microbiol. 12, 438–448 (2014). [PubMed: 24801941]
52. Hoshino Y & Gaucher EA On the origin of isoprenoid biosynthesis. Mol. Biol. Evol. 35, 2185–2197 (2018). [PubMed: 29905874]
53. Moreira D & Lopez-Garcia P Symbiosis between methanogenic archaea and δ-proteobacteria as the origin of eukaryotes: the syntrophic hypothesis. J. Mol. Evol. 47, 517–530 (1998). [PubMed: 9797402]
54. Lopez-Garcia P & Moreira D Eukaryogenesis, a syntrophy affair. Nat. Microbiol. 4, 1068–1070 (2019). [PubMed: 31222170]
55. Imachi H et al. Isolation of an archaeon at the prokaryote–eukaryote interface. Nature 577, 519–525 (2020). [PubMed: 31942073]
56. Rodrigues-Oliveira T et al. Actin cytoskeleton and complex cell architecture in an Asgard archaeon. Nature 613, 332–339 (2023). [PubMed: 36544020]
57. Keeling PJ The number, speed, and impact of plastid endosymbioses in eukaryotic evolution. Annu. Rev. Plant Biol. 64, 583–607 (2013). [PubMed: 23451781]

58. Husnik F et al. Bacterial and archaeal symbioses with protists. Curr. Biol. 31, R862–R877 (2021). [PubMed: 34256922]
59. Pittis AA & Gabaldon T Late acquisition of mitochondria by a host with chimaeric prokaryotic ancestry. Nature 531, 101–104 (2016). [PubMed: 26840490]
60. Gabaldon T Relative timing of mitochondrial endosymbiosis and the 'pre-mitochondrial symbioses' hypothesis. IUBMB Life 70, 1188–1196 (2018). [PubMed: 30358047]
61. Hoshino Y & Gaucher EA Evolution of bacterial steroid biosynthesis and its impact on eukaryogenesis. Proc. Natl Acad. Sci. USA 118, e2101276118 (2021). [PubMed: 34131078]
62. Hillman BI & Cai G The family narnaviridae: simplest of RNA viruses. Adv. Virus Res. 86, 149–176 (2013). [PubMed: 23498906]
63. Nibert ML, Vong M, Fugate KK & Debat HJ Evidence for contemporary plant mitoviruses. Virology 518, 14–24 (2018). [PubMed: 29438872]
64. Philippe C et al. Bacteriophage GC1, a novel tectivirus infecting Gluconobacter cerinus, an acetic acid bacterium associated with wine-making. Viruses 10, 39 (2018). [PubMed: 29337868]
65. Krupovic M & Koonin EV Evolution of eukaryotic single-stranded DNA viruses of the *Bidnaviridae* family from genes of four other groups of widely different viruses. Sci. Rep. 4, 5347 (2014). [PubMed: 24939392]
66. Walker PJ et al. Recent changes to virus taxonomy ratified by the International Committee on Taxonomy of Viruses (2022). Arch. Virol. 167, 2429–2440 (2022). [PubMed: 35999326]
67. Edgar RC et al. Petabase-scale sequence alignment catalyses viral discovery. Nature 602, 142–147 (2022). [PubMed: 35082445]
68. Zayed AA et al. Cryptic and abundant marine viruses at the evolutionary origins of Earth's RNA virome. Science 376, 156–162 (2022). [PubMed: 35389782]
69. Koonin EV, Krupovic M & Dolja VV The global virome: how much diversity and how many independent origins? Environ. Microbiol. 25, 40–44 (2023). [PubMed: 36097140]
70. Krupovic M et al. *Adnaviria*: a new realm for archaeal filamentous viruses with linear A-form double-stranded DNA genomes. J. Virol. 95, e0067321 (2021). [PubMed: 34011550]
71. Chang WS et al. Novel hepatitis D-like agents in vertebrates and invertebrates. Virus Evol. 5, vez021 (2019). [PubMed: 31321078]
72. Gong Z & Han G-Z Insect retroelements provide novel insights into the origin of hepatitis B viruses. Mol. Biol. Evol. 35, 2254–2259 (2018). [PubMed: 29924338]

Box 1

Viromes of bacteria, archaea and eukaryotes

Three of the four major virus realms, *Duplodnaviria*, *Varidnaviria* and *Monodnaviria*, are represented in each of the domain-specific viromes (*Riboviria* are so far missing in archaea), but with major differences in abundance, diversity and representation of kingdoms and phyla, where the lower taxa are confined to individual domains, as are the two smaller realms, *Adnaviria* and *Ribozyviria*. The RNA viruses in the kingdom *Orthornavirae* (*Riboviria*) are far more broadly represented in eukaryotes than they are in bacteria. Although the latest findings indicate that riboviruses are more prominent contributors to the bacterial virome than previously suspected[45,67,68], *Riboviria* remains dominated by viruses of eukaryotes. Even more strikingly, the kingdom *Pararnavirae*, which consists of reverse-transcribing viruses, is exclusively associated with eukaryotes, although bacteria and archaea harbour many non-viral retroelements.

The ssDNA viruses of the realm *Monodnaviria* are abundantly represented in both prokaryotic and eukaryotic viromes (Fig. 2), but the host ranges of the ssDNA viruses do not overlap already at the kingdom level[69].

In the vast realm *Varidnaviria*, the small kingdom *Helvetiavirae* is restricted to archaea and bacteria, whereas the expansive kingdom *Bamfordvirae* includes viruses from all three domains of life. Within *Bamfordvirae*, the phylum *Nucleocytoviricota* encompasses diverse large and giant viruses that are fully eukaryote specific. The second phylum, *Preplasmiviricota*, is a rare case of viruses infecting each of the three domains of life mixing at this taxonomy level.

The realm *Duplodnaviria* consists mostly of tailed bacteriophages and the related viruses of archaea (both within the class *Caudoviricetes*). Until recently, herpesviruses (order *Herpesvirales*) that are presently confined to animals were the only group of eukaryotic viruses within *Duplodnaviria*. However, the recent discovery of protist-infecting mirusviruses[34] suggests that duplodnaviruses could be far more widespread among eukaryotes than was previously suspected.

The small realm *Adnaviria* is widespread in archaea[70], but there is no detectable connection to viruses of bacteria or eukaryotes. The realm *Ribozyviria* includes hepatitis delta virus and hepatitis delta virus-like viruses discovered in other vertebrates, as well as insects[71].

All in all, comparison of the viromes of prokaryotes and eukaryotes reveals distinct compositions, with all classes of viruses, most phyla and even some kingdoms and realms being domain specific. A major distinction between the viromes of prokaryotes and eukaryotes is the dominance of dsDNA viruses (both *Duplodnaviria* and *Varidnaviria*) in prokaryotes and the contrasting preponderance of *Riboviria* in eukaryotes (Fig. 2a). The elaborate endomembrane system of the eukaryotic cell apparently provides a fertile ground for RNA virus reproduction, whereas the nucleus presents a barrier for DNA viruses that few of them managed to clear or circumvent.

An orthogonal view from the vantage point of the diversity of the virus realms (Fig. 2b) shows that *Duplodnaviria* are heavily dominated by bacterial viruses, with small fractions of viruses infecting archaea and eukaryotes; among the *Riboviria*, the representation of viruses of bacteria and eukaryotes is comparable, with a slight excess of eukaryotic ones; *Adnaviria* is an exclusively archaeal realm; and the remaining three realms are heavily (or completely, in the case of *Ribozyviria*) dominated by viruses of eukaryotes (Fig. 2b).

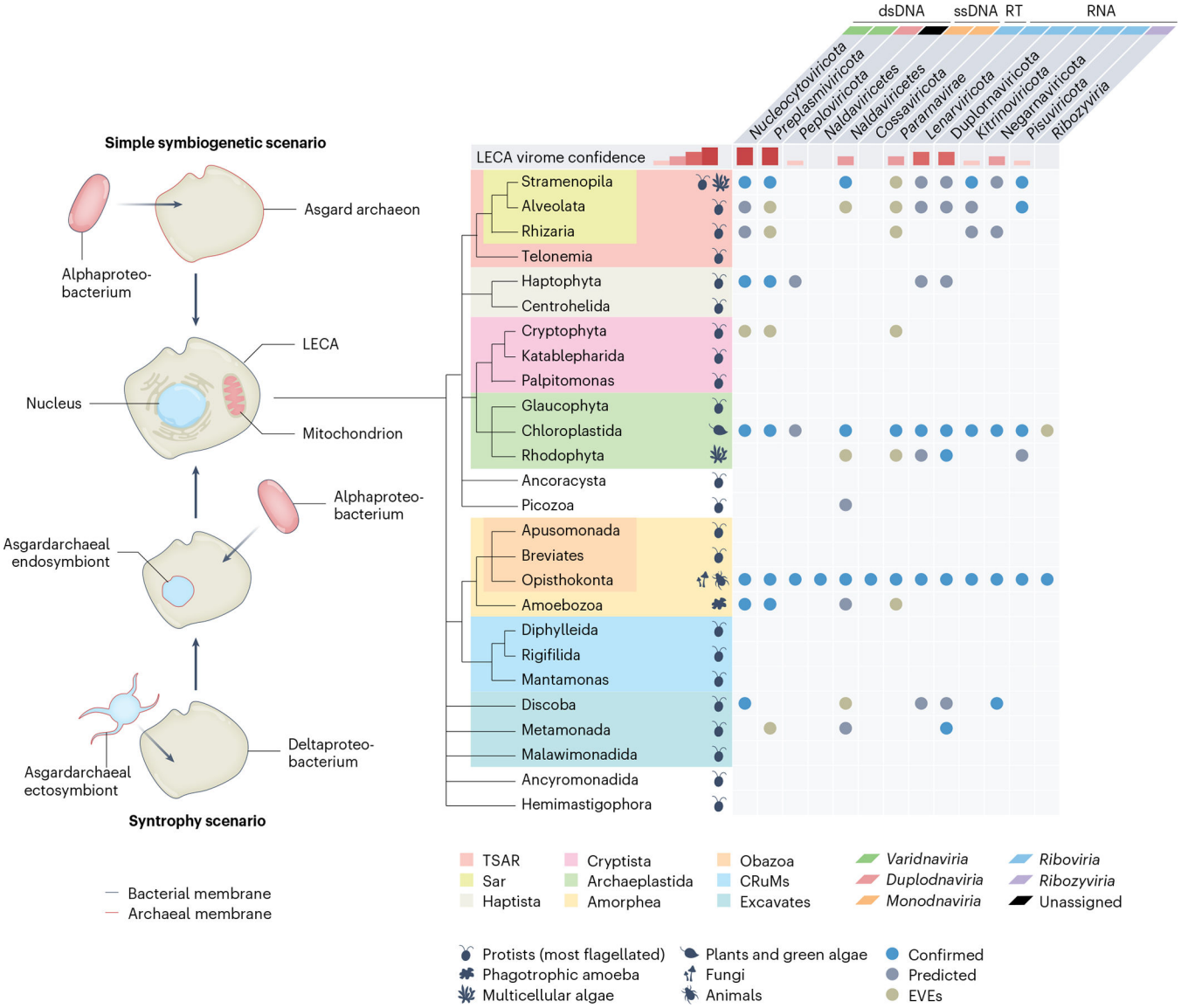

Fig. 1 |. The LECA virome.

Left, schematic of two alternative scenarios of eukaryogenesis that include either one endosymbiotic event (with an Asgard archaeon engulfing an alphaproteobacterium) or two such events (with a deltaproteobacterium engulfing an Asgard archaeon first and the resulting chimera then engulfing an alphaproteobacterium). Right, schematic phylogenetic tree of eukaryotes[19], with major clades of eukaryotes indicated at the tree leaves and the broadly used names of the informal supergroups shown at the bottom of the figure. The predominant types of organisms in each clade are depicted with pictograms. Only Chloroplastida (green plants), Stramenopila (brown algae), Rhodophyta (red algae) and Opisthokonta (animals and some fungi) include multicellular eukaryotes, whereas the rest consist of unicellular forms. The phyla of eukaryote viruses are shown as a grid next to the corresponding cellular taxa. Genome types of the corresponding viruses are indicated above the taxon names, which are also colour coded according to the virus realm. The known virus–host associations are shown with coloured circles for cultivated viruses

(blue), associations predicted from metagenomics and metatranscriptomics studies (grey) and endogenous viruses integrated in the host genomes (green). References exemplifying the depicted associations are provided in Supplementary Table 1. The composition of the LECA virome was inferred from the distribution via an informal parsimony approach whereby a group was assigned to the LECA if it was represented in at least three of the six supergroups of eukaryotes. Virus phyla mapped to the LECA are indicated by the coloured bars shown at the top of the grid. The height and intensity of the colour of the bars indicate the confidence of the inference. CRuMs, collodictyonids (syn. diphylleids) + Rigifilida + *Mantamonas*; EVE, endogenous viral element; RT, reverse-transcribing viruses; TSAR, telonemids, stramenopiles, alveolates and Rhizaria.

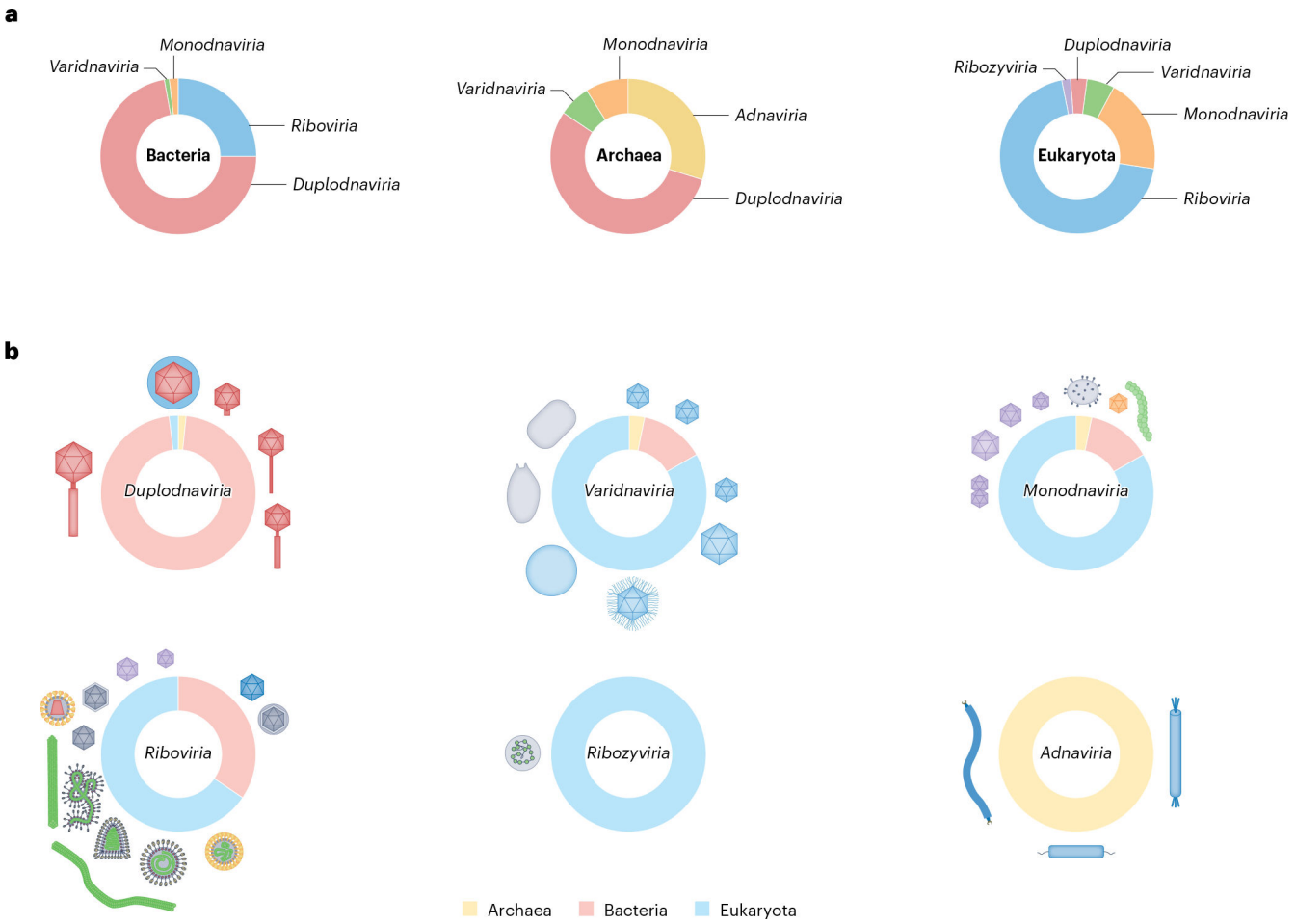

Fig. 2 |. Viromes of prokaryotes and eukaryotes.

a, Representation of the six virus realms in bacteria, archaea and eukaryotes. **b**, The host ranges, at the domains of life level, of the six realms of viruses. Virus diversity in each realm is illustrated by images of the corresponding virions. The fractions of each realm were calculated as the fractions of virus genera recognized by the International Committee on Taxonomy of Viruses[66]. Virions constructed from structural proteins with distinct folds are coloured differently.

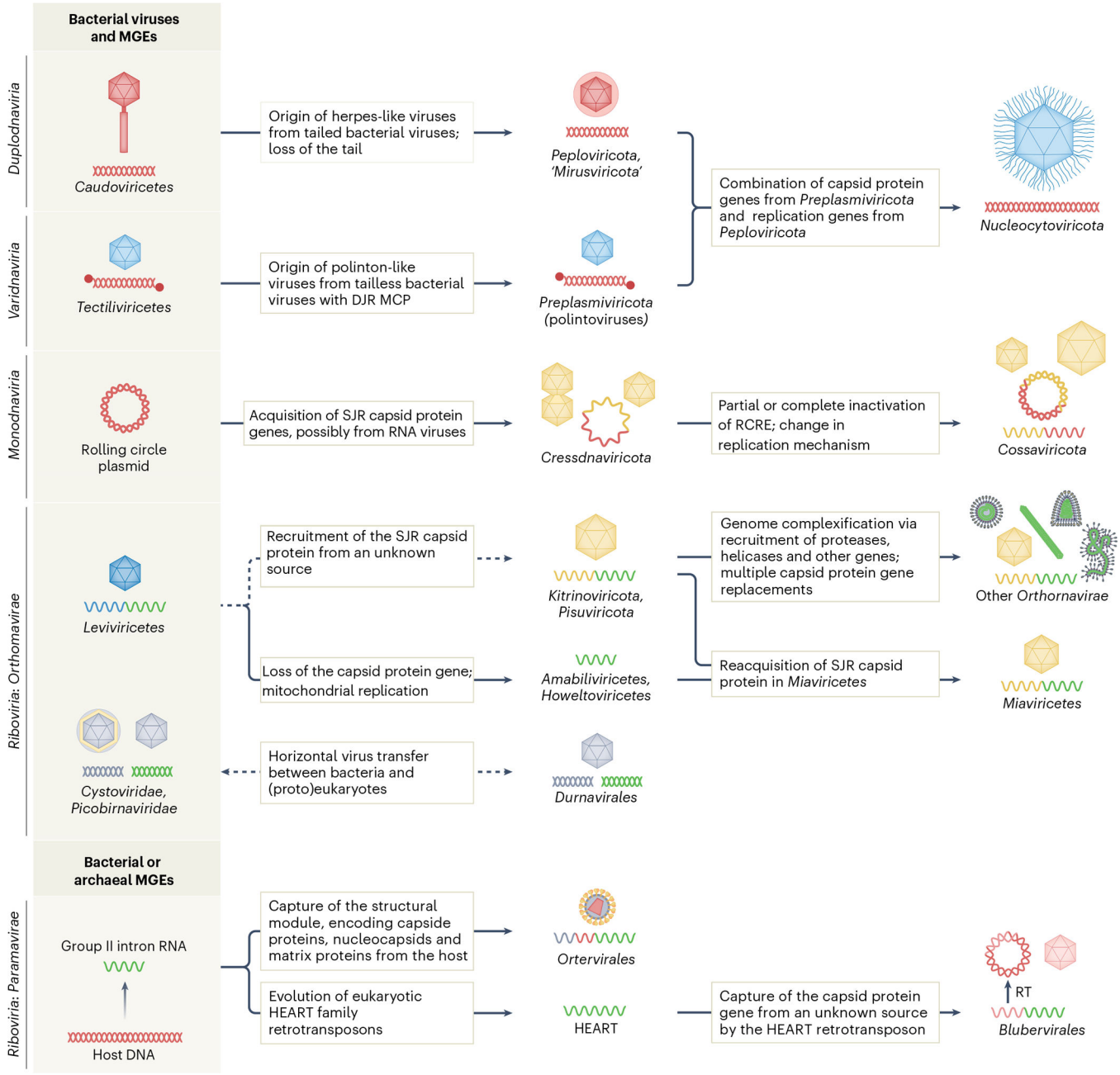

Fig. 3 |. Bacterial roots of the LECA virome.

The hypothetical scenarios of the origin of the major components of the LECA virome from bacterial viruses and non-viral MGEs. The major changes accompanying the evolutionary transitions between the corresponding bacterial and eukaryotic viruses and MGEs and subsequent evolution in eukaryotes are explained within text boxes over the arrows. Viruses and MGEs are depicted with the schematics of the corresponding genomes and virions. DNA and RNA genomes are indicated with red and green wavy lines, respectively. The capsid protein genes are shown in blue, yellow, pink and grey, with the corresponding capsids depicted with the matching colours. DJR, double jelly-roll; MCP,

major capsid protein; RCRE, rolling circle replication endonuclease; HEART, hepadnavirus-like retroelement[72].

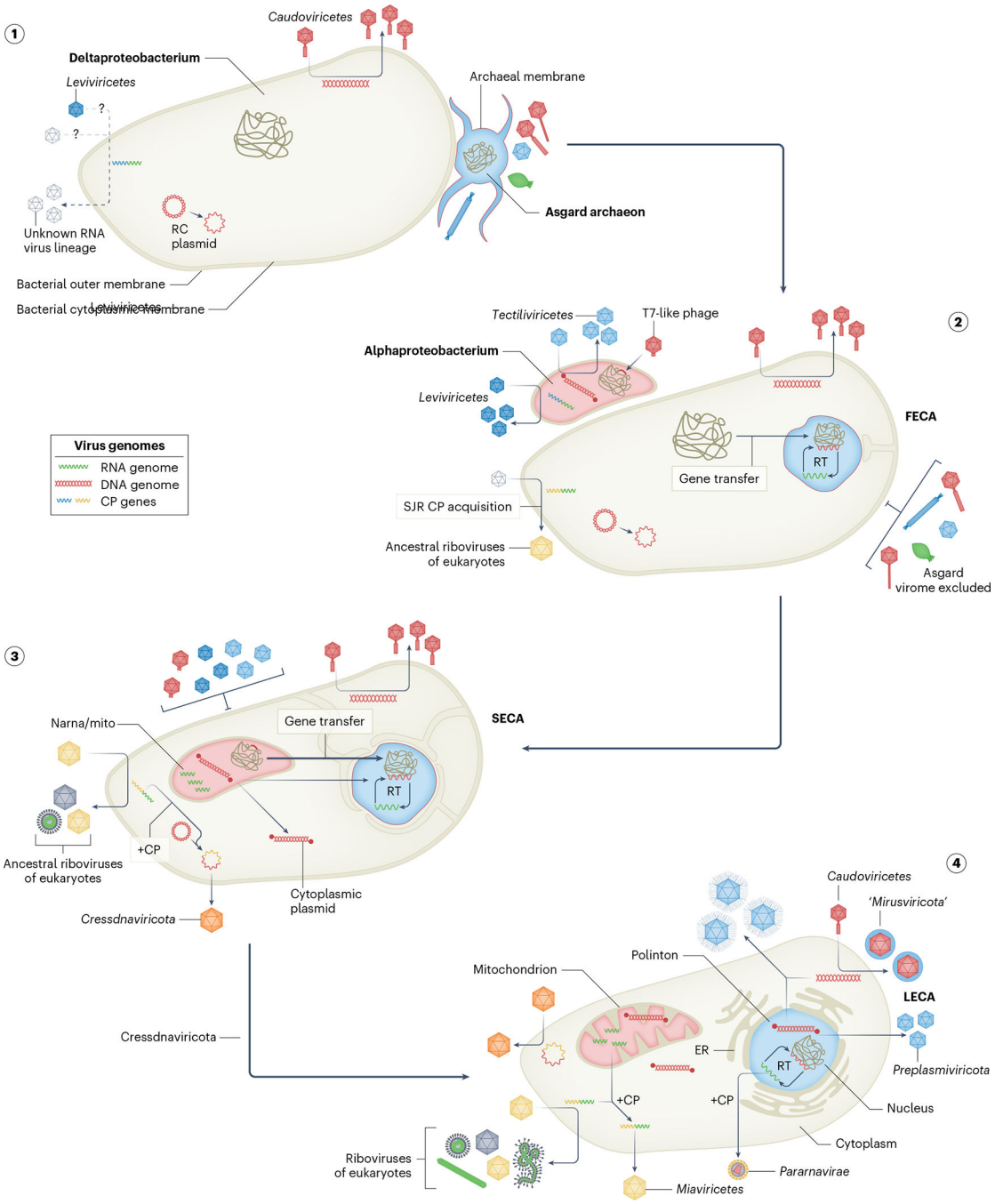

Fig. 4 |. Eukaryogenesis and eukaryovirogenesis.

The scenario of eukaryovirogenesis is based on the updated syntrophy model of the eukaryogenesis scenario with a two-stage endosymbiosis[15]. The main stages of eukaryogenesis and eukaryovirogenesis are indicated with numbers. (1) Formation of a syntrophic metabolic consortium consisting of a deltaproteobacterium and an Asgard archaeon, where each organism is associated with a specific virome. (2) Internalization of the Asgard archaeal symbiont by the deltaproteobacterium results in the emergence of the FECA and in exclusion of the archaeal virome by the bacterial membrane of

the FECA. At this stage, the Asgard archaeon is still bound by the archaeal-type membrane, but periplasmic space starts to develop around it. The symbiosis is stabilized through fusion of the bacterial and archaeal genomes, entailing horizontal exchange of genes and chromosomal retroelements such as group II introns. RNA viruses of the deltaproteobacterium undergo diversification accompanied by replacement of the ancestral capsid protein gene with the SJR capsid protein. The deltaproteobacterium also carries rolling circle plasmids and is infected with tailed dsDNA bacteriophages. A distinct consortium between the FECA and an alphaproteobacterium carrying its own specific virome is formed. The alphaproteobacterial virome includes a T7-like virus that integrates into the genome and persists as a prophage, as well as distinct RNA viruses (leviviruses). (3) Internalization of the alphaproteobacterium by the FECA results in shedding of most of the alphaproteobacterial viruses, except for tectiviruses, leviviruses and the T7-like prophage. An endomembrane system develops from the cytoplasmic membrane of the deltaproteobacterium in proximity of the Asgard archaeal membrane. The alphaproteobacterial genome migrates to the shaping nucleus, leading to the emergence of the SECA. Loss of the levivirus capsid protein gene leads to the emergence of capsidless narnaviruses and mitoviruses (narna/mito) replicating within the alphaproteobacterium. Escape of the tectivirus genome into the cytoplasm of the deltaproteobacterium by losing the capsid protein genes yields linear cytoplasmic plasmids, whereas integration of the tectivirus into the emerging nucleus gives rise to polintons and polintoviruses. Capture of the SJR capsid protein gene by rolling circle plasmids gives rise to *Cressdnaviricota*. (4) The alphaproteobacterial endosymbiont undergoes final transformation into the mitochondrion, retaining capsidless narnaviruses and linear tectivirus-derived plasmids. The archaeal membrane of the Asgard endosymbiont is replaced with the endomembrane-derived nuclear envelope with nuclear pores, surrounded by the endoplasmic reticulum network, yielding the LECA. Re-acquisition of the capsid protein gene by narna-like viruses yields members of *Miaviricetes*; other members of the *Orthornavirae* undergo extensive diversification. Polintons in the nuclear genome give rise to polinton-like viruses, virophages and other groups of viruses with double jelly-roll capsid proteins. Recombination between polintons and duplodnaviruses of the mirusvirus group gives rise to *Nucleocytoviricota*. Reverse-transcribing pararnaviruses emerge from retroelements originating either from Asgard archaea or bacterial group II introns. Question marks signify uncertainty regarding the nature of the ancestral deltaproteobacterial RNA virus. CP, capsid protein; ER, endoplasmic reticulum; RC, rolling circle.