

Ancestral State Reconstructions Trace Mitochondria But Not Phagocytosis to the Last Eukaryotic Common Ancestor

Nico Bremer [†], Fernando D. K. Tria ^{*†}, Josip Skejo , Sriram G. Garg , and William F. Martin

Institute for Molecular Evolution, Heinrich Heine University Düsseldorf, Düsseldorf 40225, Germany

*Corresponding author: E-mail: tria@hhu.de.

[†]N.B. and F.D.K.T. contributed equally.

Accepted: 25 May 2022

Abstract

Two main theories have been put forward to explain the origin of mitochondria in eukaryotes: phagotrophic engulfment (undigested food) and microbial symbiosis (physiological interactions). The two theories generate mutually exclusive predictions about the order in which mitochondria and phagocytosis arose. To discriminate the alternatives, we have employed ancestral state reconstructions (ASR) for phagocytosis as a trait, phagotrophy as a feeding habit, the presence of mitochondria, the presence of plastids, and the multinucleated organization across major eukaryotic lineages. To mitigate the bias introduced by assuming a particular eukaryotic phylogeny, we reconstructed the appearance of these traits across 1789 different rooted gene trees, each having species from opisthokonts, mycetozoa, hacrobia, excavate, archeplastida, and Stramenopiles, Alveolates and Rhizaria. The trees reflect conflicting relationships and different positions of the root. We employed a novel phylogenomic test that summarizes ASR across trees which reconstructs a last eukaryotic common ancestor that possessed mitochondria, was multinucleated, lacked plastids, and was non-phagotrophic as well as non-phagocytic. This indicates that both phagocytosis and phagotrophy arose subsequent to the origin of mitochondria, consistent with findings from comparative physiology. Furthermore, our ASRs uncovered multiple origins of phagocytosis and of phagotrophy across eukaryotes, indicating that, like wings in animals, these traits are useful but neither ancestral nor homologous across groups. The data indicate that mitochondria preceded the origin of phagocytosis, such that phagocytosis cannot have been the mechanism by which mitochondria were acquired.

Significance

The origin of mitochondria within eukaryotes is often assumed to be linked with the ability of some eukaryotic species to intake organic matter from the environment via a process known as phagocytosis. Some theories invoke phagocytosis as a mechanism to explain how mitochondria entered the eukaryotic cell, by definition they assume that phagocytosis originated before mitochondria. Alternative theories for the origin of mitochondria invoke microbial symbiotic interactions that do not require phagocytosis as a mechanism of mitochondrial entry into their host cell. Here, we were able to establish that mitochondria arose before phagocytosis did; hence, phagocytosis cannot have been the mechanism by which mitochondria arose. This indicates in turn that large complex nucleated cells (eukaryotes) required mitochondria to become phagocytotic.

© The Author(s) 2022. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Key words: last eukaryote common ancestor, phagocytosis, phagotrophy, origin of mitochondria, eukaryogenesis, ancestral state reconstruction.

Introduction

Phagocytosis is the process through which eukaryotic cells specifically recognize and engulf cell-sized particles (≥ 0.4 micrometer) via cytoskeleton-dependent invagination of the plasma membrane. Phagocytosis is a trait widely distributed among and exclusive to eukaryotes, serving as a strategy for internal digestion of food particles (Martin *et al.* 2017; Mills 2020) as opposed to extracellular digestion via secreted enzymes. For as long as mitochondria have been discussed as endosymbionts, phagocytosis has been discussed in the context of mitochondrial origin. In her revitalization of the endosymbiotic theories of Mereschkowsky (Kowallik and Martin 2021, Martin and Kowallik 1999) and Wallin (1927), Margulis, then named Sagan (1967) suggested in passing that phagocytosis was the mechanism by which the ancestral mitochondrion and its host became established. Cavalier-Smith proposed that phagocytosis directly gave rise to mitochondria and chloroplasts, but not via endosymbiosis, rather by origin of the organelles via restructuring of membranes in a cyanobacterial ancestor of eukaryotes (Cavalier-Smith 1975). Many subsequent theories followed Margulis's idea and emphasized phagocytosis as a mechanism for mitochondrial acquisition, hereafter collectively referred to as phagocytic models, and in most if not all such theories, the engulfed mitochondrial ancestor is interpreted as an undigested meal (Doolittle 1998; Cavalier-Smith 2002; Roger *et al.* 2017; Poole and Gribaldo 2014).

By contrast, a number of alternative theories for the origin of mitochondria do not entail a phagocytosing host, placing an emphasis on microbial interactions. Two kinds of microbial interactions are discussed: predatory bacteria and metabolic symbioses. The predatory bacteria class of theories posits mitochondria origin via predation by bacteria upon other bacteria. These theories lean on examples of predatory bacteria that enter the periplasm of their bacterial host, multiply there, and consume the hosts's cytosolic content. Although initially proposed on the basis of *Bdellovibrio* predators from the deltaproteobacteria (Guerrero *et al.* 1986), a number of alphaproteobacterial predators have been found and discussed in the context of mitochondrial origin (Davidov *et al.*, 2006; Davidov and Jurkevitch, 2009). In models involving predatory bacteria, the mitochondrion is seen not as an undigested meal but as an attenuated predator.

Most current theories for mitochondrial origin involve metabolic symbioses among free living prokaryotes, though few take into account the low oxygen history of eukaryotic evolution, as recently reviewed by Mills *et al.* (2022). Metabolic symbioses typically have a nutritional

basis and often involve anaerobic syntrophy (Schink 1997; Stams and Plugge 2009; Imachi *et al.* 2020) and hydrogen dependence (Martin and Müller 1998; reviewed in Zimorski *et al.* 2014). Because phagotrophy is a feeding mechanism that supports day-to-day survival, its main function for cells is of physiological nature, involving the channeling of growth substrates from food vacuoles to mitochondria for ATP (Martin *et al.* 2017). In non-phagocytic eukaryotes, such as fungi, digestive enzymes are secreted into the environment rather than into food vacuoles. Despite the popularity of the idea that phagocytosis was the key to eukaryote origin (Cavalier-Smith 1975; Embley and Williams 2015), physiological and cytological evidence suggests that the host was likely non-phagocytotic (Gould *et al.* 2016; Martin *et al.* 2017) in line with fossil evidence indicating a late origin of phagocytosis (Mills 2020). The main physiological evidence against the phagocytic origin of mitochondria is 2-fold: (1) A mitochondrion-lacking phagotrophic archaeal host would have to ingest about 34 times its body weight in prokaryotic prey to obtain enough ATP to support one cell division at maximum energetic efficiency and (2) in contrast to all other archaea, it would lack ion gradients and chemiosmotic ATP synthesis at the plasma membrane, because phagocytosis and chemiosmotic ATP synthesis cannot coexist in the same membrane (Martin *et al.* 2017). Furthermore, more recent observations show that the closest archaeal relatives to the host that acquired mitochondria are very small and simply organized archaeal cells (Imachi *et al.* 2020), not phagocytotic proto-eukaryotes. Yet despite much evidence to the contrary (Speijer, 2015), the phagocytic origin of mitochondria remains a very popular theory (Dacks *et al.* 2016).

The presence of mitochondria at the base of eukaryotic evolution (Martin and Müller 1998; Embley and Martin 2006; Müller *et al.* 2012), combined with the lack of evolutionary intermediates, render the cell-morphological grade at the prokaryote-to-eukaryote transition steep and its phylogenetic reconstruction challenging. Most current theories agree that mitochondria and their related organelles—mitosomes and hydrogenosomes—descend from a proteobacterial symbiont that took up residence within its host (Gray *et al.* 2001; Fan *et al.* 2020; Betts *et al.* 2018), whereby the host was a member of an ancient archaeal lineage (Williams *et al.* 2013; Martin *et al.* 2015). A more debated issue concerns the timing of mitochondrial acquisition relative to the emergence of other eukaryotic traits, cell complexity in particular (Lane and Martin 2010). In the context of the present study, if mitochondria were acquired via phagocytosis, then the host had already evolved a phagocytic lifestyle, meaning that large cell size,

the endomembrane system, vesicle flux and cytoskeleton—the salient components of eukaryote cell complexity—had already arisen prior to mitochondrial acquisition (Poole and Gribaldo 2014; Roger et al. 2017; Cavalier-Smith 2002; De Duve 2007). That is, according to phagocytic theories, eukaryotic complexity arose independent of mitochondrial functions or mitochondrial genes. However, no modern-day archaea grown in laboratory cultures or observed in nature are known to phagocytose.

In current formulations, phagocytic models rely on inferences from metagenome-assembled genomes (MAGs) from uncultured asgard archaea, which are reported to encode homologs of phagocytosis-related genes in eukaryotes (Zaremba-Niedzwiedzka et al. 2017; Spang et al. 2015), such as actin and tubulins. However, the purity of these MAGs has been questioned (Garg et al. 2021), and the few phagocytosis-related genes found in asgard MAGs are arguably insufficient to confer full phagocytic capability as observed in eukaryotes today. This has been demonstrated with enriched cultures of *Candidatus Prometheoarchaeum syntrophicum* MK-D1 the only asgard archaeon that has been cultivated in the laboratory to date. MK-D1, the closest archaeon to the eukaryotic host, showed no evidence of phagocytic ability under the microscope, although it was able to generate membrane protrusions (Imachi et al. 2020) which are feeding appendages that increase surface area for its fermentative lifestyle, similar to the function of hyphae in filamentous fungi (Scannell et al. 2006).

Alternatives to phagocytic models for the origin of mitochondria, the symbiotic models, have it that the archaeal host was not phagocytotic and that the mitochondrial ancestor established a symbiotic relationship living in close physical contact with its archaeal host (Martin et al. 2001, 2015). Over the course of time, the symbiosis of prokaryotes stabilized, the host became strictly dependent upon its symbiont (anaerobic syntrophy), leading to entry of the bacterial symbiont into the host's cytosol (endosymbiosis). Several examples of prokaryotes that have taken up symbiotic relationships within the cytosol of another—nonphagocytotic—prokaryote are known (Martin et al. 2017). In addition, modern-day archaea can undergo membrane fusions and cell fusions (Naor and Gophna 2013), such that symbiogenic models do not require an origin of phagocytosis within archaea prior to mitochondrial origin. At face value, both phagocytotic and symbiogenic theories would predict the origin of the eukaryotic plasma membrane to be of archaeal origin, but the eukaryotic outer membrane is chemically more similar to that of bacteria. To account for this, symbiotic models have a corollary in which the lipids of the eukaryotic plasma membrane arose via secretion of membrane vesicles by the bacterial endosymbiont, which ultimately replaced the original host outer membrane (Gould et al. 2016).

Both phagocytic and symbiotic models for the origin of mitochondria are currently discussed and debated, whereby the role of environmental oxygen levels roughly 1% that of current oxygen levels during eukaryotic and mitochondrial origin as well as during the first billion years of eukaryotic evolution bear heavily upon these issues. For a balanced and comprehensive review, see Mills et al. (2022). Discrimination between the theories requires more data and analyses, not more debate. One largely unexplored issue concerns the premise underlying phagocytic theories, namely that phagocytosis evolved prior to mitochondrial acquisition and hence was present in the last eukaryotic ancestor (LECA). An earlier study focused on the identification of phagocytosis-related genes in eukaryotic genomes followed by reconstruction of phylogenetic trees and used the gene trees as proxies to speculate about the origin of phagocytosis as a process (Yutin et al. 2009). However, due to the multiplicity of functions a gene can have, identifying phagocytosis-related genes can lead to many false positives (Gotthardt et al. 2006; Okada et al. 2006; Marion et al. 2005; Jacobs et al. 2006). Furthermore, genes that precipitated phagocytosis may have been lost or replaced, and eukaryotic genes that are currently known to be involved in phagocytosis may have originated prior to phagocytosis. Hence, inferences indicating that phagocytosis-related genes originated in LECA cannot be readily equated to an early-origin of phagocytosis. For example, both archaea and bacteria are known to possess tubulin homologues (Erickson et al. 2010), but neither archaea nor bacteria are phagocytotic.

Here, we address the origin of phagocytosis in eukaryotes within the framework of ancestral state reconstruction (ASR) analyses. By examining the presence of phagocytosis as a process, rather than the presence of a few phagocytosis-related genes, across a diverse sample of eukaryotic species we readdress the phagocytosis-origin problem from a novel empirical perspective. We specifically examine the timing of phagocytosis and phagotrophy in eukaryote evolution, in addition to the antiquity of the multinucleated (syncytial) state (Skejo et al. 2021) and, as controls, the origin of mitochondria and plastids.

Results and Discussion

Framework and Data

Our dataset consists of five eukaryotic traits—mitochondria, phagocytosis (the ability to engulf bacterial cells), phagotrophy (phagocytosis as a feeding habit), multinucleate organization, and plastids, as well as the distribution of these traits across 150 eukaryotic species that span six lineages: Opisthokonta, Archaeplastida, Hacrobia, Excavata, Stramenopiles, Alveolates and Rhizaria (SAR), and

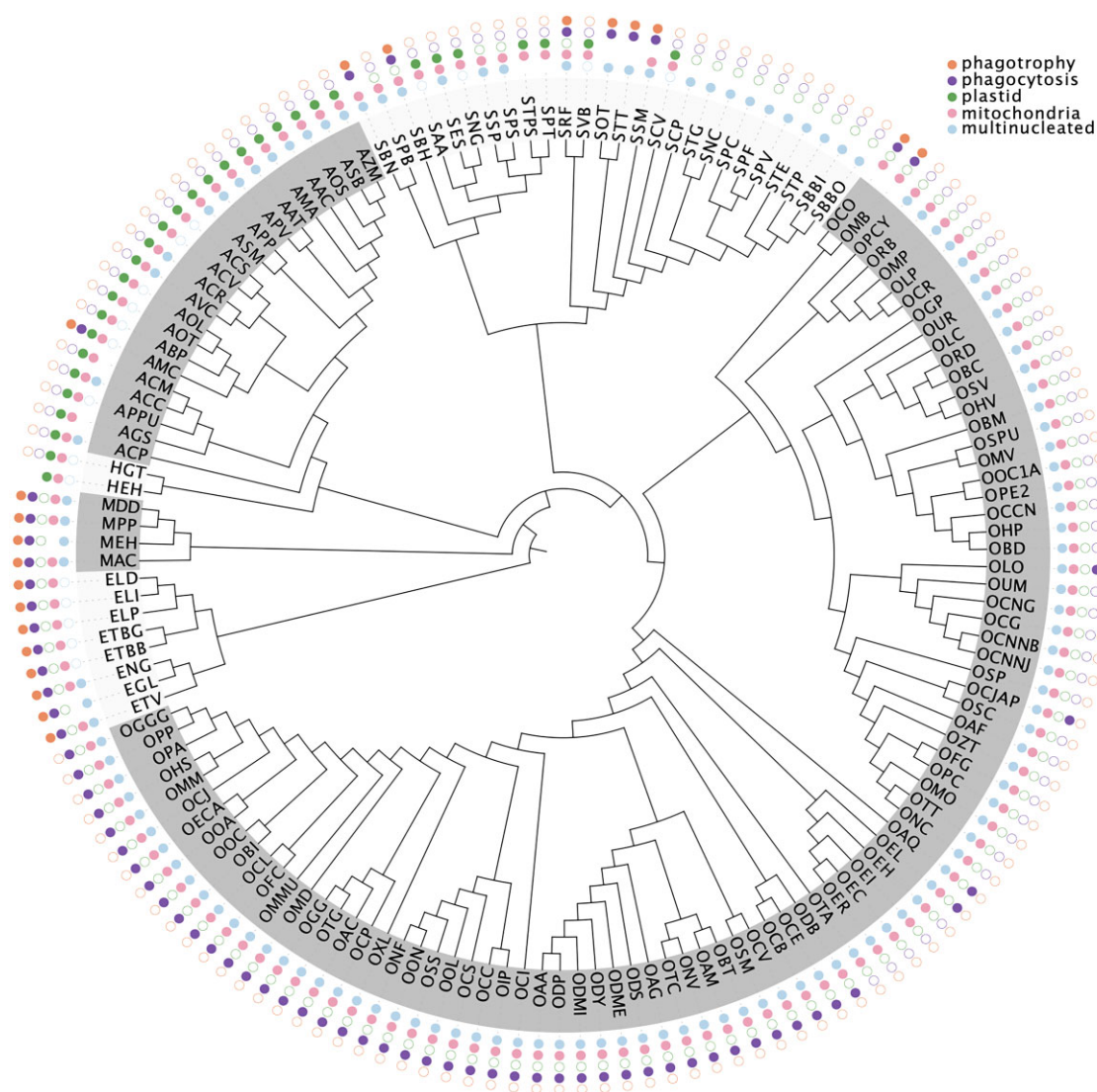


FIG. 1.—Presence (filled circle) absence (empty circle) distribution of five traits in 150 eukaryotic species. Species with no circle for a given trait indicate missing annotation. The reference tree was inferred from the alignment of 18S RNA sequences, rooted on the Excavates branch, with the sole purpose of data display (see Materials and Methods). Tip labels are species codes (see [supplemental table 1, Supplementary Material](#) online, for complete species names and detailed trait annotations). The first character of the species codes indicates supergroup affiliation of the species: Excavates (E), Mycetozoa (M), Hacrobia (H), Archaeplastida (A), SAR (S) and Opisthokonta (O). The shades of gray show the clades of the six eukaryotic supergroups.

Mycetozoa (fig. 1; see Materials and Methods for details). To evaluate the potential contribution of each of the traits to eukaryogenesis, we first set out to time their origin relative to the last eukaryotic common ancestor (LECA) using ASR. To perform ASR, two types of data are required: a table with the distribution of traits in some species and a phylogenetic tree upon which ASR is calculated for all internal nodes in the tree. Typically, the tree used for ASR is a species tree which is commonly reconstructed from sequences of single-copy genes common to all species under scope (universal orthologs). By clustering 1848936 protein-coding genes from the 150 eukaryotic genomes

using a markov clustering algorithm (MCL) (Enright et al. 2002), we obtained 239 012 gene families in total. Of the total, 313 gene families are present in at least 140 genomes, 130 gene families are present in at least 145 genomes, and 15 gene families are present in 149 genomes. However, no gene family in our data is strictly universal, that is, with gene-copies present in all 150 eukaryotic genomes, because our species set includes species with highly reduced genomes including the parasite *Giardia lamblia* and the unicellular photosynthetic species *Nannochloropsis gaditana* ([supplemental table 1, Supplementary Material](#) online). Reconstructing a reliable species tree without

universal orthologs is a challenging task, which is further complicated by the abundance of paralogues in the present data (Tria et al. 2021), as a result of frequent gene duplications in eukaryotic evolution.

To harness phylogenetic information contained in genomes and bypass the reliance on a backbone species tree, we used 1789 gene families to reconstruct maximum-likelihood trees for each individual gene family. The 1789 gene families are distributed variably across eukaryotic genomes (mean = 105.3, median = 114, SD = 34.8) but are present in at least one representative species from the six eukaryotic supergroups. The presence of the genes in six supergroups indicates that these gene families likely trace to LECA or prior to it. We then used each resulting gene tree to perform an independent ASR experiment, under the principle that each gene family is an independent data-sample. Gene trees are informative for ASR as long as the underlying gene families originated in LECA and not after it. Yet, the accuracy of the ASR may vary across trees due to tree errors and sampling effect generated by gene duplication and gene loss. We will address these issues along the paper. Notably, the gene trees used here served only as phylogenetic markers. We neither assumed nor expected the functions of the genes to be either directly or indirectly involved in the establishment of the eukaryotic traits investigated.

LECA Had Mitochondria and Was Multinucleated, But It Was Neither Phagocytic nor Phagotrophic

For each tree and trait, we labeled the species at the tips of the tree according to their trait-state annotations (supplemental table 1, Supplementary Material online) and performed maximum-likelihood ASR (see Materials and Methods for details). In each tree, we could identify the trait-state (for example, presence or absence) that traced to LECA. A tree was only used for ASR of a given trait if the tree contained representative species for at least two trait-states. Trees displaying only one state for a given trait (for example, all taxa having mitochondria) were uninformative and not considered for ASR of that trait. A maximum-likelihood ASR yields probabilities for each possible trait-state at the root of the tree, where the result may be resolved or ambiguous when alternative trait-states are tied with equal probabilities. Because each tree spans all major eukaryotic lineages, its root corresponds to LECA. One way of summarizing ASR across trees is by counting the frequency in which each trait-state appeared in LECA across trees (the majority-rule). A trait-state occurring in LECA at a high frequency across trees likely reflects the true state in LECA, whereas trait-states occurring in low frequencies in LECA are the result of lineage specific origins for the trait or errors. It is important to note that the majority-rule method does not utilize trees with

unresolved trait-states in LECA, and the magnitude of the difference in probabilities for alternative trait-states is not considered at all.

Using the majority-rule, we found that the ASRs traced the presence of canonical mitochondria to LECA in 90% of the trees, recovering the (now) well-accepted notion of mitochondria being present in the LECA (table 1) as posited by most current theories that address the origin of mitochondria. Alone, the presence of mitochondria in LECA has no weight in distinguishing current alternative theories for the origin of mitochondria in eukaryotes, because all current theories have mitochondria in LECA—a radical change from 20 years ago (Martin et al. 2001)—but it serves as a first validation of our approach. Another validation was obtained with the analyses of photosynthetic plastids, a trait uncontestedly thought to have originated after the eukaryotic supergroups diverged, at the base of photosynthetic lineages (Archaeplastida). Our analyses indicated a late origin of plastids relative to LECA in 78% of the trees, in accordance to the expectation. The ASR placed the origin of photosynthetic plastids in LECA in only 6% of trees, with the remaining 16% trees having unresolved ASR. The 10% of trees that trace the origin of mitochondria after LECA and the 6% trees that traced plastids into LECA are clear deviations from the expected results, indicating a roughly 10% error rate underlying the majority-rule analyses. Our ASRs also show LECA as a multinucleate (syncytial) organism in 69% of the trees, in accordance with an independent study (Skejo et al. 2021). Multinucleate species and stages, in which different nuclei divide independently both of each other and of cell division, are surprisingly common among eukaryotes (Skejo et al. 2021). The results we obtained for mitochondria, plastids and the multinucleated state are in accordance with commonly accepted notions of eukaryotic trait evolution, serving as an internal control and validation for our analyses.

The most relevant traits for investigating the question of how mitochondria entered the eukaryotic lineage are phagocytosis—the process of engulfing cells, like macrophages—and phagotrophy—engulfing cells as a feeding habit as opposed to osmotrophy, whereby enzymes are excreted outside the cell to digest and uptake digestion products. We analyzed each trait independently. Phagocytosis was defined as species harboring cells with the ability to actively internalize particles larger than 400 nm (the size of a small bacterium), whereas phagotrophy was defined as the special case of using phagocytosis as a feeding habit. For example, humans are phagocytic because of macrophage activity during infection but not phagotrophic, because we digest food in the intestine and uptake breakdown products via plasma membrane importers. Despite a wide distribution of phagocytosis and a moderate distribution of phagotrophy in the 150 eukaryotic species in our dataset (fig. 1), the majority-rule across

Table 1

Maximum-likelihood ancestral reconstruction of five traits from 150 eukaryotic species, across a broad sample of gene trees as estimates of the underlying phylogeny. Absolute values indicate the number of trees with a trait state (presence/absence) tracing to LECA. The total number of trees used (*N*) as well as the number of trees with ambiguous reconstructions in LECA are indicated. Multinucleate, phagocytosis, and phagotrophy were modeled as binary traits, while mitochondria and plastids were modeled as traits with three states each (see [supplemental table 1, Supplementary Material](#) online and Materials and Methods for details). For mitochondria, “presence” indicates that canonical mitochondrion is the reconstructed ancestral state, while “absence” indicates that the reconstruction is either mitosome or hydrogenosome.

Single-copy gene trees				
Trait	Presence	Absence	Ambiguous	Total (<i>N</i>)
<i>mitochondria</i>	8 (100%)	0	0	8
<i>plastid</i>	1 (5%)	7 (33%)	13 (62%)	21
<i>multinucleate</i>	14 (67%)	0	7 (33%)	21
<i>phagocytosis</i>	1 (5%)	7 (33%)	13 (62%)	21
<i>phagotrophy</i>	1 (5%)	10 (48%)	10 (48%)	21
Multi-copy gene trees				
Trait	Presence	Absence	Ambiguous	Total (<i>N</i>)
<i>mitochondria</i>	1191 (90%)	4 (0,3%)	123 (9%)	1318
<i>plastid</i>	106 (6%)	1372 (78%)	290 (16%)	1768
<i>multinucleate</i>	1234 (70%)	162 (9%)	372 (21%)	1768
<i>phagocytosis</i>	475 (27%)	779 (44%)	514 (29%)	1768
<i>phagotrophy</i>	323 (18%)	963 (54%)	482 (27%)	1768

trees indicates that LECA was neither phagotrophic nor phagocytic. That is, the origin of phagocytosis was reconstructed after LECA in 44% of the trees, in LECA in 27% of the trees, whereas 29% trees had unresolved ASRs ([table 1](#)). Phagotrophy, the trait that phagotrophic models for the origin of mitochondria require, appeared in LECA in only 18% of the trees, with 54% of trees placing the origin of phagotrophy after LECA. For 28% of the trees, the ASRs of phagotrophy were unresolved.

The analyses for phagocytosis and phagotrophy yielded a higher proportion of unresolved ASRs in comparison to the traits mitochondria, multinucleate organization and plastids ([table 1](#)). To assess the statistical significance of our results, we performed a test by matching the probabilities of the alternative trait-states for each tree regardless of outcome in LECA (trait-presence, trait-absence or tie) and assessed the differences in distributions using the Wilcoxon signed-rank test ([fig. 2](#)). The test can be seen as a refinement of the majority-rule as it considers the magnitude of probabilities for all possible trait-states in LECA, which are directly obtained from the ASR, and integrate information from trees with unresolved ASR in LECA. The results of the two-tailed Wilcoxon tests indicate that the traits phagotrophy and phagocytosis were not present in LECA at $P < 0.01$.

GTPases, tubulins, and actins are common among eukaryotes and play key roles in phagocytosis ([Rougerie et al. 2013](#); [Hall 2012](#); [Lancaster et al. 2018](#)); the likely presence of these genes in LECA has been interpreted as evidence for an early origin of phagocytosis relative to mitochondria. However, the origin of phagocytosis-related genes is not

guaranteed to coincide with the origin of phagocytosis because the genes that precipitated the origin of phagocytosis may have been lost or replaced over the course of 1.5 billion years of evolution since eukaryotes emerged ([Betts et al. 2018](#)). Contrary to phagotrophic theories for the origin of mitochondria, but in line with some earlier views ([Martin et al. 2003](#)), our results show that LECA was neither phagotrophic nor phagocytic, obviating the requirement of these traits for the origin of mitochondria in eukaryotes. A previous study based on the comparative analyses of gene expression data for phagocytic-related genes ([Yutin et al. 2009](#)) also suggested a late origin of phagocytosis, as did a study of microfossil evidence for the late origin of phagocytosis ([Mills 2020](#)).

Tree Quality, Sampling, and Conflicting Evidence in Phylogenomic Analyses

The accuracy of ASR depends on the quality of the individual gene trees. Because of gene duplications and gene losses, topological discordance cannot be equated to the ever-present problem of tree reconstruction errors. Tree reconstruction strongly depends on the quality of the sequence alignments, which can be assessed using the heads or tails (HoT) analyses ([Landan and Graur, 2007](#)). We investigated the grade of HoT scores across the 1789 trees by comparing the positional consistency of the original alignments (heads) against the alignments obtained from the sequences in their reversed amino-acid order (tails). Higher HoT values indicate higher positional consistency between the original and reversed alignments, which is

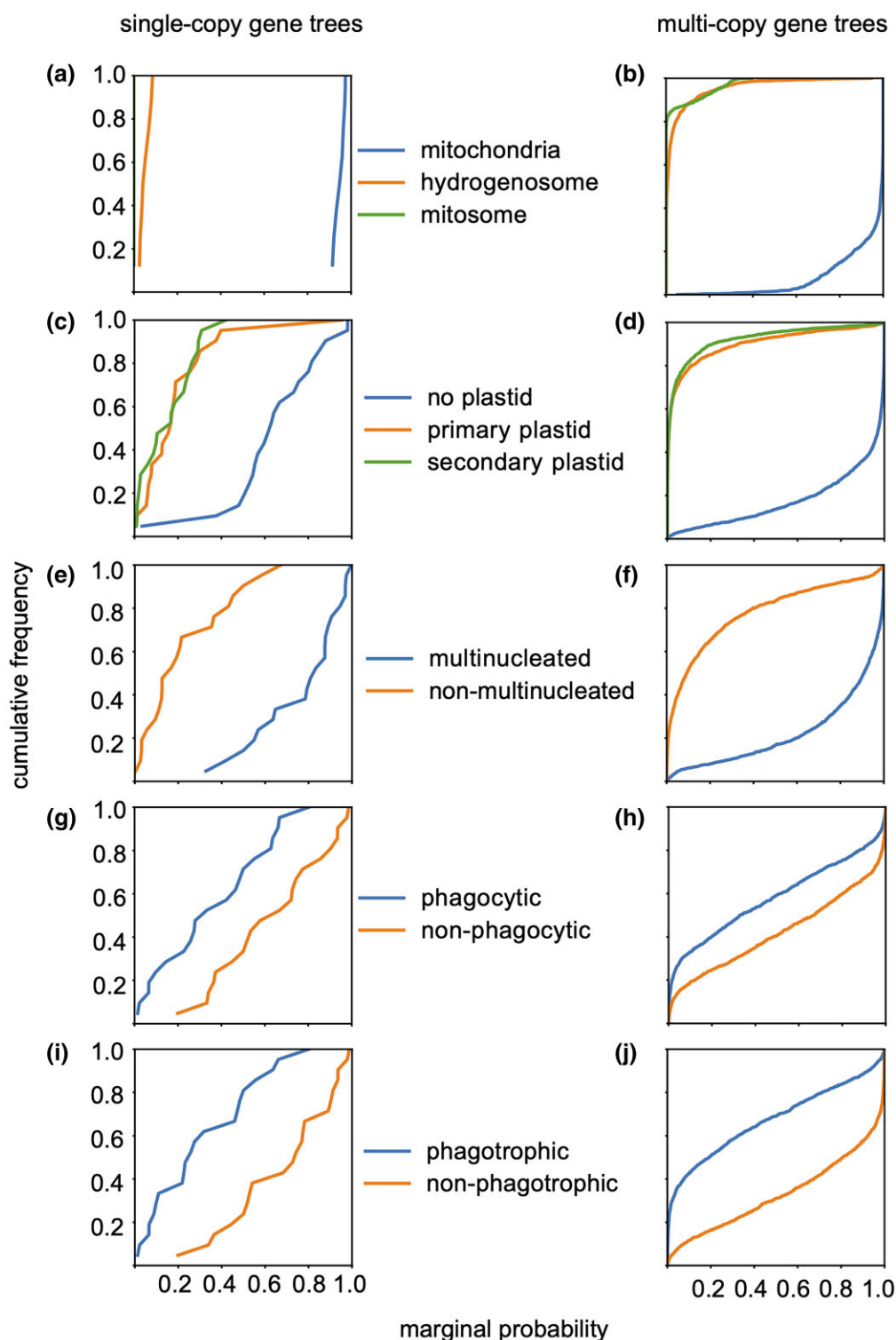


Fig. 2.—Distribution of marginal probabilities for alternative trait-states in LECA across single-copy gene trees (left; without paralogs) and multi-copy gene trees (right; with paralogs). Multinucleate, phagocytosis and phagotrophy were treated as binary traits, while plastids and mitochondria were treated as traits with three states each. For plastids the states were: absence, primary plastid or secondary plastid. For mitochondria the states were as follows: mitosome, hydrogenosome, or canonical mitochondria (see Methods and [supplemental table 1, Supplementary Material](#) online for details). The number of trees used in the analyses are shown in [table 1](#). Trait-states with high probabilities in the trees have distributions (colored lines) that are right-shifted in the plots.

indicative of well-aligned sequences. The distribution of HoT scores for all the 1789 gene trees, grouping trees according to trait-state outcome in LECA, for each trait separately, are shown in [supplemental fig. 1, Supplementary Material](#) online. The HoT scores indicate little difference between forward and reverse alignments. We found that the overall tree quality is high, with the majority of trees having scores above 0.6 according to the mean column score (MCS), which indicates the proportion of identically aligned site columns, and above 0.9 for the mean residue pair score (MRPS, identically aligned pairwise site comparisons). Furthermore, the distributions of alignment scores underlying trees that recovered different trait-states in LECA had no clear difference, suggesting that tree reconstruction errors are unlikely to explain different ASR results. Alignment quality does not impact our current results because the Wilcoxon-tests, using only the top 200 trees according to HoT scores, recovered the same ASR for all five traits ([supplemental fig. 3, Supplementary Material](#) online; [fig. 2](#)).

Another factor that may influence ASR is the position of the root within the trees. We used the minimal ancestor deviation (MAD) approach to root the trees (Tria et al., 2017), which outperformed alternative approaches in independent studies (Wade et al. 2020; Lamarca et al. 2022) and has the advantage of not requiring outgroups. Yet, MAD rooting is expected to fail for trees with high levels of molecular-clock departure, which may vary across trees. Indeed, we found microsporidians, a highly specialized group of fungal pathogens with highly relaxed functional constraints (high rates) for many genes, at the base of 10% of our gene trees, which is indicative of errors due to long branch attraction (Brinkmann et al. 2005). To account for the effect of the quality of inferences, we analyzed the distribution of two root scores calculated by MAD: the ancestor deviation (AD) statistic for the inferred root position, which measures the degree of deviation from the molecular-clock associated to the inferred root, and the root ambiguity index (AI), defined as the ratio of AD scores for the inferred root over the second-best root. We found the distribution of AD and AI to be remarkably similar for trees that obtained a different trait-state in LECA ([supplemental fig. 1, Supplementary Material](#) online), suggesting that no significant bias of ASR was caused by variable levels of root inference accuracy. Furthermore, by repeating the Wilcoxon tests with the top 200 trees with best root quality, as judged independently for AD and AI, we recovered the same ASR as obtained with all trees in the sample ([supplemental fig. 3, Supplementary Material](#) online).

It is noteworthy that the results of our ASR analyses depend upon the eukaryotic species sampled, which were limited to the species with genomic sequences in RefSeq (O'Leary et al. 2016). We deliberately avoided the inclusion of

metagenomic and transcriptomic sequences, because they are notoriously more susceptible to contamination (false taxon label), base-calling, and assembly errors, which bias phylogenetic reconstructions (Garg et al. 2021). Nevertheless, sampling is an important factor in ASR analyses. Since the gene families we used to reconstruct the trees are not uniformly distributed across the eukaryotic genomes sampled here, we could investigate the effect of differential sampling upon our results, using the natural distribution of the genes as reference. We analyzed four sampling parameters calculated for each tree: (1) the fraction of the least frequent trait-state occurring at the tips of the trees; (2) the fraction basal lineages measured as the number of Excavates and Mycetozoa relative to Opisthokonts; (3) the total number of species; and (4) the total number of OTUs (operational taxonomic units). For each of the four sampling parameters, we ranked the trees in decreasing order, selected the top 200 trees, and repeated the Wilcoxon tests ([supplemental fig. 3, Supplementary Material](#) online). With only one exception, these tree subsamples corroborated the results shown in [fig. 1](#), albeit with variable P -values due to decreased sample size. The only exception occurred for the subsample of trees with highest fraction of basal lineages, where the ASR for phagocytosis in LECA could not be resolved (P -value > 0.05).

To find out which species were enriched in the subsample of 200 trees with enriched basal lineages, we calculated the frequency of appearance for each species across the tree subsample and compared to that of the entire tree sample. We found that the three microsporidia species present in our genome set and one SAR species had the highest degree of sampling improvement, when comparing how frequently these species appeared in the tree subsample relative how frequent they appeared in all trees ([supplemental table 2, Supplementary Material](#) online). By restricting the analyses to trees with high sampling of basal species resulted in a subsample of trees that are also rich in species with highly reduced genomes. This is a noteworthy result because the Microsporidia and SAR species, enriched in the subsample of 200 trees, are fast-evolving lineages known to introduce bias in phylogenetic analyses (Brinkmann et al. 2005). Unrestricted species sampling, although theoretically desirable to cover grades of biological diversity, can hinder phylogenetic analyses by increasing heterogeneity in the data. Indeed, we found a significant negative correlation of HoT scores with the total number of species in the sequence alignments underlying the trees ($\rho = -0.4$, $P < 0.01$, two-tailed Spearman-rank correlation).

As in all molecular phylogenetic studies, there is conflicting evidence in the form of conflicting signals in the present data. Conflicting signals can arise as a result of fragmented or contaminated data and therefore lead to falsely constructed clades in tree topologies (Wägele et al. 2009), which we avoided by excluding metagenomic and

transcriptomic data. An important source of conflict in eukaryotic gene families is gene duplication and the presence of paralogs. An earlier independent study found that at least 475 genes were duplicated in LECA (Tria et al. 2021). Although these duplications complicate the analysis of eukaryotic phylogenies, it is important to keep in mind that duplications are the hallmark of eukaryotic genes such that phylogeny-based analyses of eukaryote evolution have to take this into account. Eliminating genes with duplications or paralogs would eliminate almost all gene families from this or any other study of eukaryote gene or genome evolution, as nearly half of all eukaryotic protein-coding genes exist as multiple copies in at least one genome (Tria et al. 2021). The “manual” removal of paralogs from individual trees would also introduce biases of effectively arbitrary nature. Nonetheless, we could rule out paralogues as a potential bias because independent analyses of multi-copy and single-copy trees rendered the same ASR for the five eukaryotic traits we analyzed (table 1 and fig. 2). Whether or not paralogues actually hinder phylogenetic reconstructions is still unanswered, possibly case-dependent, and our analyses will motivate further investigations.

Phagocytosis and Phagotrophy Evolved Multiple Times within the Eukaryotic Lineage

A late origin of phagocytosis and phagotrophy, together with the wide distribution of these traits across eukaryotic species, raises the question of how many times these traits evolved within eukaryotes. One possibility is that phagocytosis and phagotrophy evolved only once prior to the

divergence of some eukaryotic supergroups or multiple independent times within supergroups. To test the multiple origin hypothesis, we counted for each tree the number of trait-origins. The average number of trait-origins across trees is shown in table 2, for each of the five traits investigated here. Only mitochondria showed up as a clear single origin trait, with an average of one origin per tree. By counting the number of origins for plastids, regardless of its type (that is, primary or secondary), rendered an average of four to six origins which is in line with one primary acquisition of plastids in the Archaeplastida ancestor followed by subsequent acquisitions via secondary and tertiary plastids in Hacrobia and SAR (Gould et al. 2015).

Our analyses show that even though LECA was multinucleated, the trait had on average three to seven origins across trees, indicating a high turnover rate (loss with re-appearance) for this trait in eukaryote evolution. Instances of multiple origins for the multinucleate state may reflect the selective trade-offs imposed by the co-existence of multiple nuclei within the same cell. It has been suggested that the existence of multiple nuclei in LECA permitted mutations, chromosomal rearrangements, and aneuploidies to occur freely during chromosomal segregation, because the eventual loss of gene function in one nucleus, arising from defective mutations, can be compensated by the proper functioning of the same gene in another nucleus (Garg and Martin 2016; Skejo et al. 2021). While stable environmental conditions may favor individuals with few nuclei per cell, the multinucleate state offers important adaptive capacity for populations inhabiting rapidly changing environments. In that sense, the multinucleated state is a special case of polyploidy, which can postpone the effects of Muller’s ratchet in asexually reproducing eukaryotes (Kondrashov 1994), which LECA was at some point during the transition from a symbiosis of prokaryotes to a nucleated cell with mitochondria.

Phagocytosis originated as a trait two to five times on average in the trees. Even though some key genes for these processes were already present in LECA such as GTPases, tubulins, and actins, which also exist in prokaryotes (Shih and Rothfield 2006; Verstraeten et al. 2011; Fletcher and Mullins 2010), the presence of these genes alone does not imply in the capacity to perform phagocytosis. The multiple independent origins of phagocytosis supported by our data align very well with previous observations that phagocytosis-related genes are rarely shared among distantly related eukaryotes (Yutin et al. 2009). Gene expression analyses have shown thousands of genes being differentially expressed during phagocytosis (Gotthardt et al. 2006; Okada et al. 2006; Marion et al. 2005; Jacobs et al. 2006). Among these, only about a dozen are common to phagocytic eukaryotic genomes, with the vast majority of differentially expressed genes being supergroup exclusive (Yutin et al. 2009). Overall, both ASR and comparative

Table 2

Summary statistics for the number of trait origins across trees (see note below the table). Trait origin in internal and terminal nodes are distinguished. Single-copy trees (without paralogs) were distinguished from multi-copy trees (with paralogs)

<i>n. origin^a</i> <i>trait</i>	Single-copy gene trees		
	Terminal nodes	Internal nodes	All nodes
<i>mitochondria</i>	0, 0, 0	1, 1, 0	1, 1, 0
<i>plastid</i>	3, 3, 1.6	1.1, 1, 0.8	4.1, 4, 1.4
<i>multinucleate</i>	2, 1, 2.7	1, 1, 0.6	3, 2, 2.8
<i>phagocytosis</i>	4.4, 4, 1.8	0.7, 1, 0.6	5.1, 5, 1.65
<i>phagotrophy</i>	4.1, 4, 1.7	0.6, 1, 0.6	4.7, 4, 1.4
<i>n. origin^a</i> <i>trait</i>	Multi-copy gene trees		
	Terminal nodes	Internal nodes	All nodes
<i>mitochondria</i>	0, 0, 0.2	1, 1, 0.2	1, 1, 0.4
<i>plastid</i>	2.9, 3, 1	2.9, 3, 1.3	5.8, 6, 2.4
<i>multinucleate</i>	3.5, 1, 4.6	3.6, 3, 2.9	7.1, 4, 7.1
<i>phagocytosis</i>	4.6, 4, 3.3	2.5, 2, 1.5	7.1, 7, 4.3
<i>phagotrophy</i>	5.8, 6, 3.3	2, 2, 1.3	7.8, 8, 3.9

^aNote: Numbers indicate mean, median, and standard deviation across trees.

genome analyses point to multiple origins of phagocytosis in eukaryotes. One important implication of our finding is that phagocytosis, as a process, is not homologous among eukaryotic species capable of phagocytosis. Hence, comparative analyses targeting a better understanding of phagocytosis as a process need to take process homology among species, or lack thereof, into account. One possibility is to restrict comparative genome analyses to species suspected to share phagocytic homology, which might assist the identification of currently unknown phagocytic-related genes. In a broader context, assessing trait homology using ASR as done here has the potential to improve studies aimed towards a better understanding of trait evolution across the tree of life. It also allows us to address the relative order of appearance of the eukaryotic traits investigated here, as outlined in the following.

Timing the Origin of Eukaryotic Traits Relative to the Emergence of Eukaryotic Supergroups

To time the origin of traits relative to the divergence of six well-known eukaryotic supergroups considered here, we identified the eukaryotic species that descend from the origin node and recorded the corresponding supergroup affiliation of descending species. We repeated this process for each trait and trait-origin independently using all origin nodes as inferred by the ASR, across all trees, and plotted the distribution supergroups descending from the origin nodes (fig. 3). For each reconstructed origin, all the species (tips) descending from it in the tree were used to score an origin as a combination of supergroups so identified. In this way, we were able to estimate the approximate origin of the traits relative to the supergroups without committing to any particular eukaryotic supergroup phylogeny, which is a recognized challenge and hotly debated topic (Burki *et al.* 2020). Furthermore, the possibility that some of the supergroups used here might not be monophyletic has no influence on our results because the species were allowed to assume any relationship in the trees, without topological constraints. The supergroups only serve the purpose of displaying the results, as higher order leaf labels regardless of underlying backbone species tree, and some traits map well to supergroup assignments used here. As it concerns the traits that originated in LECA, we only considered the ASRs that placed an origin at the root of the trees (red circles in fig. 3).

We distinguish origins that occurred after the root node, for which the descending species represent all six supergroups (black circles) which could also be indicative of trait origin at LECA but with some level of uncertainty since they could alternatively be the result of phylogenetic errors. The combination of supergroups with high frequency of origins across trees are likely to coincide with a true trait-origin in the underlying supergroup phylogeny, while low-frequency supergroup combinations are more likely spurious results.

For mitochondria the result is very clear, for 1326 gene trees a high number of origin nodes ($n = 1199$) occurred in LECA. For plastids, the highest number of origins occurred in an Archaeplastida ancestor ($n = 1705$) for 1789 gene trees, followed closely by the number of origins in SAR ($n = 1240$). A moderate number of plastids origins was also observed in the SAR + Hacrobia ancestor ($n = 282$) and the Hacrobia exclusive ancestor ($n = 200$). The multinucleate trait had the highest number of origins in LECA ($n = 1234$) for 1789 gene trees, albeit a high to moderate number of origins was also observed in the ancestor of each supergroup (fig. 3), indicating presence in LECA in addition to multiple lineage specific (secondary) origins for the multinucleate form. That is, the multinucleate state was likely lost several times subsequent to LECA's divergence but recurrently reemerged within each supergroup. For phagocytosis, the highest number of origins occurred in Opisthokonta ($n = 891$) for 1789 gene trees, followed by Excavata ($n = 641$) and Mycetozoa ($n = 620$). The natural diversity of the processes usually classified as phagocytosis across eukaryotic supergroups, together with our results, indicates that the phagocytic processes evolved independently in Opisthokonta, Mycetozoa, and Excavata. For phagotrophy, the highest numbers of origins for 1789 gene trees were found within three supergroups: Mycetozoa ($n = 805$), Excavata ($n = 793$), and SAR ($n = 528$). For clarity, 805 origins of phagocytosis refer to the sum of origins scored across 1789 separate trees having on average 105 species, each tree containing representatives from all six eukaryotic supergroups sampled here.

Conclusions

In the context of eukaryogenesis, our findings reject phagocytic models because the results indicate that the underlying premise of an ancestral phagocytic state for eukaryotes (in LECA) is unlikely to be true. Furthermore, our results indicate multiple independent origins of four of the five traits studied here, namely, plastids (including secondary plastids), the multinucleated state, phagocytosis, and phagotrophy. By contrast, mitochondria appeared with a clear single-origin in our analyses, tracing to LECA or prior. While recurrent acquisitions of photosynthetic plastids were previously described, multiple origins of multinucleate state, phagocytosis, and phagotrophy in eukaryotes are under-investigated issues. As such, our study here provides new insights into early eukaryote history and new methods for ASR that do not require the use of an agreed or accepted backbone species tree. All we require for this approach to work is codable information about traits, a sufficient number of genes present across members of the group in question, and taxonomic assignments regardless of phylogenetic relationship. Our results have implications for understanding the mechanism

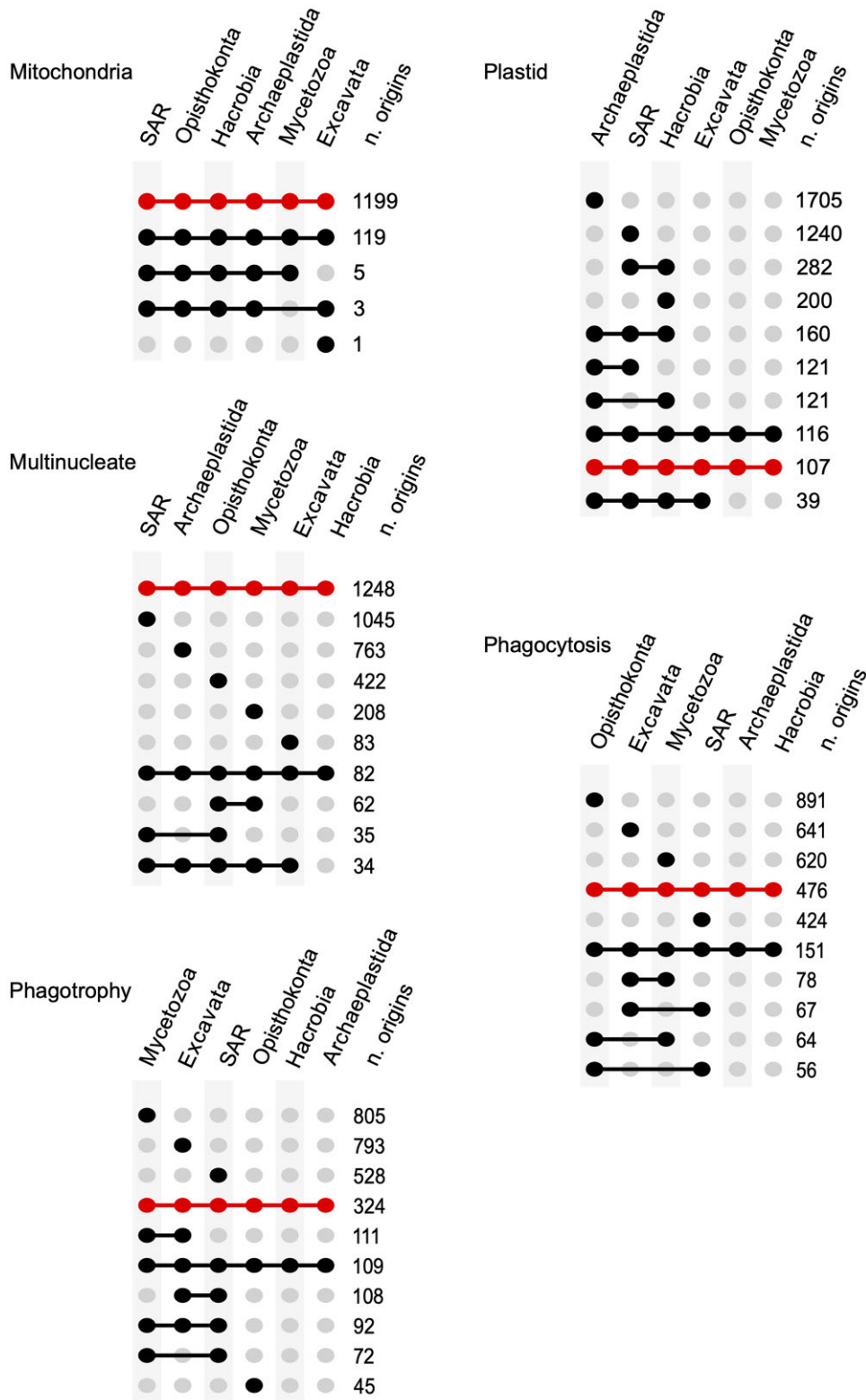


FIG. 3.—Distribution of supergroups descending from origin nodes across 1789 trees. For each internal node reconstructed as a trait origin, all the species (tips) descending from it were used to score an origin to the combination of supergroups (filled circles) to which the descending species belong. Origins at the root node (LECA) are shown in red.

underlying the acquisition of mitochondria, a feature exclusive to eukaryotic cells. The broader significance of these findings is that the origin of mitochondria can be attributed to a fateful case of microbial symbiosis but cannot be attributed to a fateful case of indigestion.

Materials and Methods

Phylogenetic Trees

Protein sequences from all 150 eukaryotic genomes were clustered as follows: all-vs.-all Basic Local Alignment Search Tool (BLAST) (Altschul *et al.* 1990) of the protein sequences was performed. The reciprocal best BLAST hits, with an expectation value (e-value) $\leq 10^{-10}$ were selected and globally aligned with the Needleman–Wunsch algorithm, as implemented in the European Molecular Biology Open Software Suite (EMBOSS) needle program (Rice *et al.* 2000). Protein pairs with a global identity <25% were discarded. The remaining pairs were then used for clustering with MCL algorithm (Enright *et al.* 2002), version 12-068 using default parameters. One-thousand seven-hundred eighty-nine protein clusters for proteins distributed in at least one species of each eukaryotic supergroup were selected to derive estimates for underlying eukaryotic phylogeny. Protein alignments were generated using Multiple Alignment using Fast Fourier Transform (MAFFT) (Katoh *et al.* 2002), using the iterative refinement method that assimilates local pairwise alignment information (L-INS-i). The nontrimmed alignments were used to reconstruct maximum likelihood trees with IQ-TREE (Nguyen *et al.* 2015), using the best-fit model. The applied parameters were “-bb 1000” and “-alrt 1000.” Trees without paralogs, here termed single-copy gene trees, were distinguished from trees with paralogs, termed here multi-copy gene trees, for the purpose of evaluating the effect of paralog inclusion in ancestral reconstructions. All trees were rooted with MAD (Tria *et al.* 2017), and none of the 1789 trees showed ambiguous root inferences.

Trait Annotation, Coding, and Definition

In the field of eukaryogenesis phagocytosis is often used as an overarching term encompassing all forms of membrane engulfment, while ignoring specific cell biological differences between the various processes. Phagocytosis here was defined as internalization of particles typically larger than 400 nanometers. The main function of phagocytosis in unicellular organisms is feeding on prokaryotes, while for the immune system of multicellular animals phagocytosis serves other functions like apoptotic cell removal. The use of phagocytosis to feed on bacteria for energy is distinct from its use by the immune system and therefore we distinguished feeding phagocytosis using the term phagotrophy which refers to unicellular eukaryotes that ingest bacteria for feeding. Both phagotrophy and phagocytosis were treated here as binary traits (presence

“1” or absence “0” in the [supplemental table 1, Supplementary Material](#) online), as was the multinucleate trait. Photosynthetic plastids and mitochondria were treated as multi-state traits. For plastids, we distinguished no plastids (0), primary plastids (1) and secondary plastids (2). While for mitochondria we distinguished canonical mitochondria (1), mitosome (2) and hydrogenosome (3).

For the distribution of traits across the species, see [supplemental table 1, Supplementary Material](#) online.

ASR

The reconstruction of ancestral states was performed using PastML version 1.9.20 (Ishikawa *et al.* 2019). PastML is an algorithm that requires a rooted phylogenetic tree and annotated tips for the tree. The analyses were conducted with using a maximum likelihood approach based on marginal posterior probabilities approximation with the F81 model of character evolution (Felsenstein 1981). The annotation of the tips of the trees was based on the trait matrix for the 150 eukaryotic species ([supplemental table 1, Supplementary Material](#) online), with the inclusion of missing data (unknown tip state). For a given trait, trees with the same state of a trait for every tip of the tree were discarded from the analysis.

The analysis of the constructed phylogenetic trees and trait origins was conducted with the python toolkit Environment for Tree Exploration ETE v3 (Huerta-Cepas *et al.* 2016).

Statistical Tests

For testing the significance of ASR across a sample of trees we collected from each tree (ASR) the marginal probability for the trait being present in LECA and the marginal probability for the trait being absent in LECA, as given by PastML. Differences in the distribution of marginal probabilities for alternative trait-states were assessed with the two-tailed paired Wilcoxon test and considered significant at $P \leq 0.05$. The test assesses whether the distribution of probabilities across all trees are significant larger for one of the trait-states. The test permits the resolution of ASR for which a simple count of trait-states (majority-rule) is not sufficient, such as for phagocytosis.

18S RNA Reference Tree

For the construction of a reference tree for our 150 eukaryotes, we collected 18S RNA sequences for each species. We therefore searched primarily in the SILVA rRNA database (release 138.1 from November 2020) (Quast *et al.* 2013). As we were not able to find sequences for all 150 eukaryotic species in the SILVA rRNA database, we secondarily searched for sequences in the PR² sequence database (version 4.12.0 from August 2019) (Guillou *et al.* 2013). For eight species, we were not able to find a 18S RNA sequence

in both databases and therefore used alternatives from the same genus. The alignment was generated using MAFFT (Kato et al. 2002), using the iterative refinement method that assimilates L-INS-i. The alignment was then used to reconstruct a maximum likelihood tree with IQ-TREE (Nguyen et al. 2015). The resulting tree was rooted on the branch leading to Excavates. The tree was constructed and rooted for the sole purpose of data display.

HoT Analyses

For each gene family from which gene trees were reconstructed, the original protein alignments (heads) were compared with the alignments for the sequences in their reversed amino-acid order (tails). The positional consistency between the “heads” and “tails” alignments was assessed using two scores: the mean column score and the mean residue pair score. The analyses were performed with the HoT program (Landan and Graur 2007).

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

This project has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant 666053 and 101018894), the Volkswagen Foundation (grant 93 046), and the Moore Simons Initiative on the Origin of the Eukaryotic Cell (grant 9743).

Data Availability

Sequence alignments, phylogenetic trees, and ASR are available as **Supplemental Data** under <https://doi.org/10.6084/m9.figshare.18520409>.

Literature Cited

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol.* 215:403–410.
- Betts HC, Puttick MN, Clark JW, Williams TA, Donoghue PCJ, Pisani D. 2018. Integrated genomic and fossil evidence illuminates life’s early evolution and eukaryote origin. *Nat Ecol Evol.* 2:1556–1562.
- Brinkmann H et al. 2005. An empirical assessment of long-branch attraction artefacts in deep eukaryotic phylogenomics. *Syst. Biol.* 54:743–757.
- Burki F, Roger AJ, Brown MW, Simpson AGB. 2020. The new tree of eukaryotes. *Trends Ecol Evol.* 35(1):43–55.
- Cavalier-Smith T. 1975. The origin of nuclei and of eukaryotic cells. *Nature.* 256:463–468.
- Cavalier-Smith T. 2002. The phagotrophic origin of eukaryotes and phylogenetic classification on protozoa. *Int J Syst Evol Microbiol.* 52:297–354.
- Dacks JB et al. 2016. The changing view of eukaryogenesis - fossils, cells, lineages and how they all come together. *J Cell Sci.* 129(20):3695–3703.
- Davidov Y, Huchon D, Koval SF, Jurkevitch E. 2006. A new alphaproteobacterial clade of Bdellovibrio-like predators: implications for the mitochondrial endosymbiotic theory. *Environ Microbiol.* 8:2179–2188.
- Davidov Y, Jurkevitch E. 2009. Predation between prokaryotes and the origin of eukaryotes. *BioEssays* 31:748–757.
- de Duve C. 2007. The origin of eukaryotes: a reappraisal. *Nat Rev Genet.* 8(5):395–403.
- Doolittle WF. 1998. You are what you eat: a gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes. *Trends Genet.* 14(8):307–311.
- Embley TM, Martin W. 2006. Eukaryotic evolution, changes and challenges. *Nature.* 440(7084):623–630.
- Embley TM, Williams TA. 2015. Steps on the road to eukaryotes. *Nature.* 521(7551):169–170.
- Enright AJ, Van Dongen S, Ouzounis CA. 2002. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.* 30(7):1575–1584.
- Erickson HP, Anderson DE, Osawa M. 2010. FtsZ in bacterial cytokinesis: cytoskeleton and force generator all in one. *Microbiol Mol Biol Rev.* 74(4):504–528.
- Fan L et al. 2020. Phylogenetic analyses with systematic taxon sampling show that mitochondria branch within alphaproteobacteria. *Nat Ecol Evol.* 4(9):1213–1219.
- Felsenstein J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. *J Mol Evol.* 17(6):368–376.
- Fletcher DA, Mullins RD. 2010. Cell mechanics and the cytoskeleton. *Nature.* 463:485–492.
- Garg SG et al. 2021. Anomalous phylogenetic behavior of ribosomal proteins in metagenome-assembled *Asgard archaea*. *Genome Biol Evol.* 13(1): evaa238.
- Garg SG, Martin WF. 2016. Mitochondria, the cell cycle and the origin of sex via a syncytial eukaryote common ancestor. *Genome Biol Evol.* 8:1950–1970.
- Gotthardt D, Blancheteau V, Bosserhoff A, Ruppert T, Delorenzi M, Soldati T. 2006. Proteomics fingerprinting of phagosome maturation and evidence for the role of a $G\alpha$ during uptake. *Mol Cell Proteomics.* 5(12):2228–2243.
- 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.
- Gould SB, Maier UG, Martin WF. 2015. Protein import and the origin of red complex plastids. *Curr. Biol.* 25:R515–R521.
- Gray MW, Gertraud B, Franz L. 2001. The origin and early evolution of mitochondria. *Genome Biol.* 2(6):1018.1–1018.5.
- Guerrero R, Pedrós-Alió C, Esteve I, Mas J, Chase D, Margulis L. 1986. Predatory prokaryotes: predation and primary consumption evolved in bacteria. *Proc Natl Acad Sci USA.* 83:2138–2142.
- Guillou L et al. 2013. The protist ribosomal reference database (PR2): a catalog of unicellular eukaryote small sub-unit rRNA sequences with curated taxonomy. *Nucleic Acids Res.* 41:D597–D604.
- Hall A. 2012. Rho family GTPases. *Biochem Soc Trans.* 40(6):1378–1382.
- Huerta-Cepas J, Serra F, Lang BF. 2016. ETE 3: Reconstruction, analysis, and visualization of phylogenomic data. *Mol Biol Evol.* 33(6):1635–1638.
- Imachi H et al. 2020. Isolation of an archaeon at the prokaryote–eukaryote interface. *Nature.* 577(7791):519–525.
- Ishikawa SA, Zhukova A, Iwasaki W, Gascuel O, Pupko T. 2019. A fast likelihood method to reconstruct and visualize ancestral scenarios. *Mol Biol Evol.* 36(9):2069–2085.

- Jacobs ME, DeSouza LV, Samaranyake H, Pearlman RE, Siu KWM, Klobutcher LA. 2006. The tetrahymena thermophila phagosome proteome. *Eukaryot Cell*. 5(12):1990–2000.
- Katoh K, Misawa K, Kuma KI, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res*. 30(14):3059–3066.
- Kondrashov AS. 1994. The asexual ploidy cycle and the origin of sex. *Nature*. 370:213–216
- Kowallik KV, Martin WF. 2021. The origin of symbiogenesis: an annotated English translation of Mereschkowsky's 1910 paper on the theory of two plasma lineages. *BioSystems*. 199:104281.
- Lamarca AP, Mello B, Schrago CG. 2022. The performance of outgroup-free rooting under evolutionary radiations. *Mol Phylogenet Evol*. 107434
- Lancaster CE, Ho CY, Hipolito VEB, Botelho RJ, Terebiznik MR. 2018. Phagocytosis: what's on the menu? *Biochem Cell Biol*. 97(1): 21–29.
- Landan G, Graur D. 2007. Heads or tails: a simple reliability check for multiple sequence alignments. *Mol Biol Evol*. 24(6):1380–1383
- Lane N, Martin WF. 2010. The energetics of genome complexity. *Nature*. 467(7318):929–934.
- Marion S, Laurent C, Guillén N. 2005. Signalization and cytoskeleton activity through myosin IB during the early steps of phagocytosis in *Entamoeba histolytica*: a proteomic approach. *Cell Microbiol*. 7(10):1504–1518.
- Martin WF et al. 2003. Early cell evolution, eukaryotes, anoxia, sulfide, oxygen, fungi first (?), and a tree of genomes revisited. *IUBMB Life*. 55:193–204.
- Martin WF, Garg S, Zimorski V. 2015. Endosymbiotic theories for eukaryote origin. *Philos Trans R Soc Lond B*. 370(1678):20140330.
- Martin WF, Hoffmeister M, Rotte C, Henze K. 2001. An overview of endosymbiotic models for the origins of eukaryotes, their ATP-producing organelles (mitochondria and hydrogenosomes), and their heterotrophic lifestyle. *Biol Chem*. 382(11):1521–1539
- Martin WF, Kowallik KV. 1999. Annotated english translation of mereschkowsky's 1905 paper 'Über natur und ursprung der chromatophoren im Pflanzenreiche. *Eur. J. Phycol*. 34:287–295.
- Martin WF, Müller M. 1998. The hydrogen hypothesis for the first eukaryote. *Nature*. 392(6671):37–41.
- Martin WF, Tielens AGM, Mentel M, Garg SG, Gould SB. 2017. The physiology of phagocytosis in the context of mitochondrial origin. *Microbiol Mol Biol Rev*. 81(3):e00008–e00017.
- Mills DB. 2020. The origin of phagocytosis in Earth history. *Interface Focus*. 10(4):20200019.
- Mills DB et al. 2022. Eukaryogenesis and oxygen in Earth history. *Nature Ecol. Evol*. 6:520–532.
- Müller M et al. 2012. Biochemistry and evolution of anaerobic energy metabolism in eukaryotes. *Microbiol Mol Biol Rev*. 76(2):444–495.
- Naor A, Gophna U. 2013. Cell fusion and hybrids in archaea: prospects for genome shuffling and accelerated strain development for biotechnology. *Bioengineered*. 4(3):126–129.
- Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*. 32(1):268–274.
- Okada M et al. 2006. Kinetics and strain variation of phagosome proteins of *Entamoeba histolytica* by proteomic analysis. *Mol Biochem Parasitol*. 145(2):171–183.
- O'Leary NA et al. 2016. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res*. 44:D733–D745.
- Poole AM, Gribaldo S. 2014. Eukaryotic origins: how and when was the mitochondrion acquired? *Cold Spring Harb Perspect Biol*. 6(12):a015990.
- Quast C et al. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res*. 41:D590–D596.
- Rice P, Longden L, Bleasby A. 2000. EMBOSS: the European molecular biology open software suite. *Trends Genet*. 16(6):276–277.
- Roger AJ, Muñoz-Gómez SA, Kamikawa R. 2017. The origin and diversification of mitochondria. *Curr Biol*. 27(21):R1177–R1192.
- Rougerie P, Miskolci V, Cox D. 2013. Generation of membrane structures during phagocytosis and chemotaxis of macrophages: role and regulation of the actin cytoskeleton. *Immunol Rev*. 256(1):222–239.
- Sagan L. 1967. On the origin of mitosing cells. *J Theor Biol*. 14(3): 255–274.
- Scannell DR, Byrne KP, Gordon JL, Wong S, Wolfe KH. 2006. Multiple rounds of speciation associated with reciprocal gene loss in polyploid yeasts. *Nature*. 440(7082):341–345.
- Schink B. 1997. Energetics of syntrophic cooperation in methanogenic degradation. *Microbiol Mol Biol Rev*. 61(2):262–280.
- Shih Y-L, Rothfield L. 2006. The bacterial cytoskeleton. *Microbiol Mol Biol Rev*. 70(3):729–754.
- Skeje J et al. 2021. Evidence for a syncytial origin of eukaryotes from ancestral state reconstruction. *Genome Biol Evol*. 13(7):evab096.
- Spang A et al. 2015. Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature*. 521(7551):173–179.
- Speijer D. 2015. Birth of the eukaryotes by a set of reactive innovations: new insights force us to relinquish gradual models. *BioEssays*. 37: 1268–1276.
- Stams AJM, Plugge CM. 2009. Electron transfer in syntrophic communities of anaerobic bacteria and archaea. *Nat Rev Microbiol*. 7(8): 568–577.
- Tria FDK et al. 2021. Gene duplications trace mitochondria to the onset of eukaryote complexity. *Genome Biol Evol*. 13(5):evab055.
- Tria FDK, Landan G, Dagan T. 2017. Phylogenetic rooting using minimal ancestor deviation. *Nat Ecol Evol*. 1:193.
- Verstraeten N, Fauvart M, Versées W, Michiels J. 2011. The universally conserved prokaryotic GTPases. *Microbiol Mol Biol Rev*. 75(3): 507–542.
- Wade T, Rangel LT, Kundu S, Fournier GP, Bansal MS. 2020. Assessing the accuracy of phylogenetic rooting methods on prokaryotic gene families. *PLoS One*. 15:e0232950
- Wägele JW, Letsch H, Klussmann-Kolb A, Mayer C, Misof B, Wägele H. 2009. Phylogenetic support values are not necessarily informative: the case of the Serialia hypothesis (a mollusk phylogeny). *Front Zool*. 6:12.
- Wallin IE. 1927. *Symbiogenesis and the origin of species*. London: Bailliere, Tindall and Cox, p. 171.
- 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.
- Yutin N, Wolf MY, Wolf YI, Koonin EV. 2009. The origins of phagocytosis and eukaryogenesis. *Biol Direct*. 4:9.
- Zaremba-Niedzwiedzka K et al. 2017. *Asgard archaea* illuminate the origin of eukaryotic cellular complexity. *Nature*. 541(7637): 353–358.
- Zimorski V, Ku C, Martin WF, Gould SB. 2014. Endosymbiotic theory for organelle origins. *Curr Opin Microbiol*. 22:38–48.

Associate editor: Dr. Rebecca Zufall