Rekonstruktion von Eigenschaften und Gengehalt der ersten eukaryotischen Zellen

Inaugural-Dissertation

zur Erlangung des Doktorgrades der

Mathematisch-Naturwissenschaftlichen Fakultät der

Heinrich-Heine-Universität Düsseldorf

vorgelegt von

Nico Bremer

geboren in Mülheim an der Ruhr

Düsseldorf, März 2025

Aus dem Institut für Molekulare Evolution
der Heinrich-Heine-Universität Düsseldorf

Gedruckt mit Genehmigung der
Mathematisch-Naturwissenschaftlichen Fakultät
der Heinrich-Heine-Universität Düsseldorf

Berichterstatter:
1. Prof. Dr. William F. Martin
2. Prof. Dr. Martin Lercher
Tag der mündlichen Prüfung: 09.07.2025

Eidesstattliche Erklärung

Hiermit versichere ich an Eides statt, dass diese Dissertation von mir selbstständig und ohne unzulässige fremde Hilfe unter Beachtung der „Grundsätze zur Sicherung guter wissenschaftlicher Praxis an der Heinrich-Heine-Universität Düsseldorf“ erstellt worden ist. Die Arbeit wurde bisher keiner Prüfungsbehörde vorgelegt und auch noch nicht veröffentlicht. Ich habe bisher keinen erfolglosen Promotionsversuch unternommen.

_____

Nico Bremer, Düsseldorf, 2025

Für Claudia.

Für Thomas.

Im Laufe dieser Arbeit wurden mit Zustimmung des Betreuers folgende Beiträge veröffentlicht:

Publikationen in Fachzeitschriften thematisiert in dieser Thesis

i. **Nico Bremer**, Fernando D. K. Tria, Josip Skejo, Sriram G. Garg, William F. Martin (2022). Ancestral state reconstructions trace mitochondria but not phagocytosis to the last eukaryotic common ancestor. *Genome Biology and Evolution* 14(6), evac079.
ii. **Nico Bremer**, Fernando D. K. Tria, Josip Skejo, William F. Martin (2023). The ancestral mitotic state: Closed orthomitosis with intranuclear spindles in the syncytial last eukaryotic common ancestor. *Genome Biology and Evolution* 15(3), evad016.
iii. **Nico Bremer**, William F. Martin, Mike Steel (2025). Surprising effects of differential loss in genome evolution: the last-one-out. *FEMS Microbiology Letters*, fnaf051.

Weitere Publikationen in Fachzeitschriften

iv. Josip Skejo, Sriram G. Garg, Sven B. Gould, Michael Hendriksen, Fernando D. K. Tria, **Nico Bremer**, Damjan Franjević, Neil B. Blackstone, William F. Martin (2021). Evidence for a syncytial origin of eukaryotes from ancestral state reconstruction. *Genome Biology and Evolution* 13(7), evab096.
v. **Nico Bremer**, Michael Knopp, William F. Martin, Fernando D. K. Tria (2022). Realistic gene transfer to gene duplication ratios identify different roots in the bacterial phylogeny using a tree reconciliation method. *Life* 12(7), 995.
vi. Natalia Mrnjavac, Falk S. P. Nagies, Jessica L. E. Wimmer, Nils Kapust, Michael R. Knopp, Katharina Trost, Luca Modjewski, **Nico Bremer**, Marek Mentel, Mauro Degli Esposti, Itzhak Mizrahi, John F. Allen, William F. Martin (2024). The radical impact of oxygen on prokaryotic evolution — enzyme inhibition first, uninhibited essential biosynthesis second, aerobic respiration third. *FEBS Letters* 598(14), 1692-1714.

Inhaltsverzeichnis

1 Zusammenfassung

Der Ursprung des komplexen Lebens durch die Eukaryogenese bildet einen Meilenstein in der evolutionären Entwicklungsgeschichte. Wenngleich das Wissen über heute lebende Eukaryoten kontinuierlich wächst, sind deren Ursprung, sowie frühen Merkmale weiterhin intensiv diskutierte Fragestellungen. Diverse, zum Teil gänzlich unterschiedliche Theorien über den Ursprung der Eukaryoten wurden in den letzten Jahrzehnten aufgestellt und bis heute weder einstimmig angenommen, noch verworfen. Diese Theorien können mit molekularen Daten geprüft werden.

Im Rahmen der vorliegenden Arbeit werden neun Merkmale des letzten gemeinsamen Vorfahrens aller heute lebenden Eukaryoten (LECA) rekonstruiert und in den Kontext der Eukaryogenese gebracht. Dabei wird deutlich, dass LECA bereits im Besitz von Mitochondrien war und diese nicht über Phagozytose aufgenommen wurden. Des Weiteren zeigt sich, dass eine Mehrkernigkeit von LECA nicht nur rekonstruiert wurde, sondern auch im Zusammenhang mit der rekonstruierten geschlossenen Mitose evolutionäre Vorteile geschaffen hat. Diese Rekonstruktionen geben einen neuen Blickwinkel auf mögliche Theorien der Eukaryogenese, indem sie gegen Modelle sprechen, die ein spätes Auftreten der Mitochondrien beschreiben und Phagozytose als Prozess für die Aufnahme der Mitochondrien ausschließen.

In einer weiteren Analyse beschäftigt sich diese Arbeit mit der Wahrscheinlichkeit sogenannter „letzte Hinterbliebene“-Topologien in Eukaryoten, die aufgrund von differentiellen Genverlust anstelle von lateralen Gentransfers (LGTs) von Prokaryoten zu Eukaryoten entstanden sind. Als „letzte Hinterbliebene“ werden Gene bezeichnet, die in nur einer einzigen Spezies oder monophyletischen Gruppe in Eukaryoten vorkommen und eine Vielzahl von prokaryotischen Homologen aufweisen. Es zeigt sich, dass die statistische Anzahl dieser Topologien infolge von Genverlust nicht gering ist und sich in verschiedenen Beispielen mit der Anzahl vermeintlicher LGTs in der Literatur deckt.

2 Abstract

The origin of complex life through eukaryogenesis represents a major transition in evolutionary history. Although knowledge about eukaryotes living today is constantly growing, their origin and the characteristics of the earliest eukaryotic cells are still the subject of intense debate. A number of different theories about the origin of eukaryotes have been put forward in the literature, yet no consensus view on the topic has emerged. These theories can be tested with molecular data.

In the present thesis, nine characteristics of the last common ancestor of all eukaryotes living today (LECA) are reconstructed and placed in the context of eukaryogenesis. It becomes clear that LECA was already in possession of mitochondria and that these were not taken up via phagocytosis. Furthermore, the multinucleate state of LECA was not only reconstructed, but it was also shown to provide evolutionary advantage in the context of the reconstructed closed mitosis. These reconstructions provide a new perspective on possible theories of eukaryogenesis by rejecting models that describe a late appearance of mitochondria and by excluding phagocytosis as a process for mitochondrial uptake.

In a further analysis, this work deals with the likelihood of so-called “last-one-out” topologies in eukaryotes that have arisen due to differential gene loss instead of lateral gene transfers (LGTs) from prokaryotes to eukaryotes. Genes that occur in only one species or monophyletic group in eukaryotes and have a large number of prokaryotic homologs are referred to as “last-one-out”. It turns out that the statistical number of these topologies due to gene loss is not small and in various examples corresponds to the number of putative LGTs in the literature.

3 Einleitung

3.1 Der Ursprung von einfachen Zellen (Prokaryoten)

Alle Lebensformen sind als Zellen organisiert. Biologen unterscheiden zwei grundlegend verschiedene Klassen von Zellen, gemessen an der Organisation und Komplexität: Prokaryoten und Eukaryoten. Zu den Prokaryoten werden die Bakterien und die Archaeen gezählt. Ihre Zellen sind i.d.R. klein (ca. 1-2 *µ*m Durchmesser) und sehr einfach organisiert, ihre DNA liegt frei im Zytosol, subzelluläre Kompartimente wie der Zellkern fehlen (Vellai und Vida 1999). Die Zellen der Eukaryoten sind grundlegend anders organisiert als die der Prokaryoten. Eukaryotische Zellen sind größer, (ca. 10-100 *µ*m Durchmesser), sind von einer komplexen Endomembransystem durchzogen, und ihre DNA liegt nicht frei im Zytosol, sondern ist stets im Zellkern lokalisiert (Vellai und Vida 1999). Theorien über den Ursprung der Zellen sind älter als Theorien über den Ursprung eukaryotischer Zellen, weil die Unterscheidung zwischen prokaryotischer und eukaryotischer Organisationsform ist erst in den 1930er Jahren durch Chatton erfolgt (Katscher 2004).

In seinem Werk „Über die Entstehung der Arten“ (engl. *On the origin of species by means of natural selection*) hat Charles Darwin nicht die Frage nach dem Ursprung des Lebens gestellt (Darwin 1860). Trotz seiner detaillierten Konzepte über die Abstammung und Selektion verschiedener Arten, vertrat er in seinem Buch die Ansicht, dass das Leben vermutlich einen übernatürlichen Ursprung hat (Darwin 1860). In einer Korrespondenz an seinen Freund Joseph Hooker formulierte er 1871 seine Theorie über den „kleine, warmen Tümpel“ (engl. *warm little pond*):

> *„ But if (& oh what a big if) we could conceive in some warm little pond with all sorts of ammonia & phosphoric salts, - light, heat, electricity &c present, that a protein compound was chemically formed, ready to undergo still more complex changes, at the present day such matter wd be instantly devoured, or absorbed, which would not have been the case before living creatures were formed.”*
>
> Charles Darwin – Brief an Joseph Hooker vom 1. Februar 1871 (Darwin Correspondence Project, “Letter no. 7471”)

Eine Theorie zum Ursprung der ersten Zellen stammt von Ernst Haeckel. Dieser beschrieb in seinem Werk „Generelle Morphologie der Organismen Erster Band: Allgemeine Anatomie der Organismen“ den Ursprung des Lebens als Selbstzeugung. In dieser bildeten sich die sogenannten Moneren, welche er als „organische Individuen einfachster Art“ betitelte, aus anorganischem Material (Haeckel 1866a, S. 179f).

Die nächste nennenswerte Theorie über den Ursprung des Lebens entwickelte sich erst ein halbes Jahrhundert später. Fast zeitgleich veröffentlichten der britische Biologe John B. S. Haldane und der russische Biochemiker Aleksandr I. Oparin unabhängig voneinander ihre Theorie einer Ursuppe. Haldane beschrieb eine primitive Atmosphäre bestehend aus viel Kohlenstoffdioxid und wenig bis keinen Sauerstoff. Diese Atmosphäre sorgte laut Haldane dafür, dass UV-Strahlen ungehindert auf die Ozeane treffen konnten. In seiner Theorie führte diese externe Energie im Zusammenspiel mit Wasser, Kohlenstoffdioxid und Ammoniak zur Bildung von einer Vielzahl an organischen Verbindungen (Haldane 1929). Diese Verbindungen wurden mit der Zeit komplexer, bis sich die ersten lebenden Partikel entwickelten, welche sich dann von den weiteren organischen Verbindungen in dieser Ursuppe ernährten (Haldane, 1929). Oparin beschrieb ebenfalls, dass organische Verbindungen mit der Zeit komplexer wurden und sich daraus schließlich Koazervate bildeten (Oparin 1938). Die Ursuppentheorie von Haldane und Oparin setze sich damit klar von der zuvor beschriebenen Theorie von Haeckel ab. Während Haeckel eine zufällige Selbstzeugung von Organismen aus nicht-lebender Materie beschrieb, führten Haldane und Oparin die schrittweise Bildung einfachster Organismen aus organischen Verbindungen auf. Eine zufällige Selbstzeugung „unter dem Einfluss irgendeiner physikalischen Kraft und irgendwelchen unbekannten Bedingungen“ (Oparin, 1938 S. 60) aus nicht-lebender Materie zu einem lebenden Organismus war laut Oparin nicht möglich (Oparin, 1938).

Eine Simulation dieser Ursuppentheorie gelang 1953 Stanley L. Miller. Er simulierte die Bedingungen, die sein Doktorvater Harold C. Urey ein Jahr zuvor aufstellte, in dem heute unter dem Namen Miller-Urey-Experiment bekannten Versuchsaufbau (Miller, 1953). Im Gegensatz zu den Theorien von Haldane und Oparin vermutete er keinen Kohlenstoffdioxid in der Atmosphäre, sondern ein Gemisch aus Methan, Ammoniak, Wasser und Wasserstoff (Urey, 1952). Der Versuchsaufbau

simulierte die vermutete Atmosphäre und setzte diese in einem zirkulierenden Aufbau elektrischer Entladungen aus. Dabei fungierten Lichtbögen als Energiequelle (Miller, 1953). Es gelang ihnen, geringe Mengen an Aminosäuren, sowie weitere unbekannte Verbindungen nachzuweisen (Miller, 1953).

Heutzutage gehen viele Wissenschaftler davon aus, dass das Leben in Hydrothermalquellen entstanden ist, wobei es auch andere Theorien wie etwa die Theorie der RNA-Welt gibt, welche davon ausgeht, dass das ursprüngliche Leben aus RNA hervorgegangen ist. (Noller *et al.* 1992, Bartel und Szostak 1993). Bei Hydrothermalquellen unterscheidet man häufig zwischen den sogenannten Schwarzen Rauchern und Weißen Rauchern. Beide Arten von Hydrothermalquellen findet man auf dem Meeresgrund in der Nähe von Spreizungszonen, die durch tektonische Plattenverschiebungen entstehen. Schwarze Raucher sind in unmittelbarer Nähe zu diesen Spreizungszonen zu finden. Die vorliegende Wassertemperatur innerhalb der Schwarzen Raucher liegt aufgrund der durch die Erdkruste aufsteigenden Magma bei bis zu 400 °C (Jupp und Schultz 2000). Außerdem ist das aus den Schloten der Schwarzen Raucher ausströmende Wasser reich an Kohlenstoffdioxid, Ammoniak und Wasserstoff (Corliss *et al.* 1981). Die vorherrschenden Bedingungen führten dazu, dass 1981 eine erste Theorie über den Ursprung des Lebens an diesen heißen Hydrothermalquellen aufgestellt wurde (Corliss *et al.* 1981). Der größte Kritikpunkt an dieser Theorie war die extreme Hitze innerhalb der Hydrothermalquelle. Diese würde dafür sorgen, dass die in der Theorie beschriebene Bildung von Aminosäuren aufgrund der durch die Temperatur einhergehenden Instabilität dieser Verbindungen nicht vonstattengehen könnte (Bernhardt *et al.* 1984).

Die Entdeckung von „Lost City“, einem Feld von Tiefseehydrothermalquellen hatte Einfluss auf Theorien zum Ursprung des Lebens. Diese Hydrothermalquelle konnte mit ihren spezifischen Bedingungen eine Alternative gegenüber jener Kritik am Ursprung des Lebens in Schwarzen Rauchern liefern (Kelley *et al.* 2001). Durch die Distanz von Hydrothermalquellen wie etwa „Lost City“ zur Spreizungszone sind die Temperaturen dort deutlich kühler und liegen zwischen 40–90 °C und damit in einem optimalen Bereich für den Prozess der Serpentinisierung, welcher eine große Rolle bei der Produktion von Wasserstoff spielt (Russel *et al.* 2010). Bei der Serpentinisierung wird Wasser durch eisenhaltige Mineralien zu Wasserstoff reduziert. Dabei werden

Hydroxidionen freigesetzt, die das geführte Wasser der Hydrothermalquellen alkalisch machen. Der Wasserstoff spielt eine entscheidende Rolle für den Ursprung des Lebens, da er als Reduktionsmittel zur Fixierung von Kohlenstoffdioxid vermutet wird (Martin und Russel 2003, Preiner *et al.* 2020, Wimmer *et al.* 2021c). Zudem ist die Lebenszeit dieser karbonatreichen Schornsteine aufgrund der Distanz zur Spreizungszone deutlich höher als bei Schwarzen Rauchern (Kelley *et al.* 2001). Im Gegensatz zu weißen und schwarzen Rauchern stoßen Hydrothermalquellen wie „Lost City“ keinen Rauch aus (Rona *et al.* 1986, Kelley *et al.* 2005). Es besteht heute Einvernehmen unter Biologen, dass die ersten Zellen Prokaryoten waren.

3.2 Der Ursprung von komplexen Zellen (Eukaryoten)

Die Entstehung der Eukaryoten ist erdgeschichtlich ca. 2 Milliarden Jahre vom Ursprung der Prokaryoten entfernt. Heute geht man davon aus, dass Eukaryoten aus Prokaryoten entstanden sind. Über die Mechanismen der Eukaryogenese wird noch kontrovers diskutiert. Annahmen über den Zeitpunkt der Entstehung der Eukaryoten, sowie der Lebzeit des letzten gemeinsamen eukaryotischen Vorfahrens variieren, jedoch wird deutlich, dass dieses Ereignis mindestens ca. 1,5 Milliarden Jahre zurückliegt (Parfrey *et al.* 2011, Chernikova *et al.* 2011). Andere Schätzungen gehen von einem Zeitraum von 1,8 bis 2,7 Milliarden Jahren aus, in welchem über hunderte Millionen von Jahren der Übergang von Prokaryoten zu Eukaryoten stattfand (Betts *et al.* 2018, Mahendrarajah *et al.* 2023). In einer weiteren Analyse wird anhand von Fossilienfunden beschrieben, dass vor 1,5 Milliarden Jahren bereits eukaryotische Mikroorganismen die ökologischen und zytoskelettalen Voraussetzungen für die eukaryotische Diversifizierung entwickelt hatten (Javaux *et al.* 2001). Zusätzlich zur Frage nach dem Zeitpunkt der Entstehung, gibt es aktuell keinen wissenschaftlichen Konsens über die direkten Vorfahren der Eukaryoten, sowie den Prozess ihrer Entstehung. Während manche Theorien grundsätzlich ähnlich sind und sich nur in Nuancen unterscheiden, sind Andere grundlegend verschieden und miteinander nicht vereinbar. So wurde unter anderem im Jahre 1999 diskutiert, ob nicht etwa die Prokaryoten aus den Eukaryoten entstanden sind und nicht wie der wissenschaftliche Konsens heutzutage beschreibt, dass die Eukaryoten sich aus den Prokaryoten entwickelten (Forterre und Phillippe 1999, Poole *et al.* 1999).

3.2.1 Endosymbiontentheorie

Bereits 1905 veröffentlichte Konstantin Mereschkowsky eine Theorie über den symbiontischen Ursprung von Chromatophoren (Mereschkowsky 1905). Während Mereschkowsky zwar einen endosymbiontischen Ursprung der heute unter dem Namen Chloroplasten besser bekannten Zellorganellen nachweisen konnte, war es ihm nicht möglich, die gleichen Schlüsse über die Entstehung der Mitochondrien zu ziehen (Martin *et al.* 2015). Die ersten Theorien über den endosymbiontischen Ursprung der Mitochondrien wurden von Portier und Wallin aufgestellt. Portier beschrieb in französischer Sprache seine Theorie der Verwandtschaft zwischen Bakterien und Mitochondrien (Sapp 1994) und dass diese in einer Vielzahl an Prozessen in der eukaryotischen Zelle involviert sind (Martin *et al.* 2015). Wallin erkannte ebenfalls die Verwandtschaftsbeziehungen zwischen Bakterien und Mitochondrien und formulierte, dass Mitochondrien von endosymbiontischen Bakterien abstammten (Wallin, 1927). Zudem beschrieb Wallin die Möglichkeit, dass Gentransfer zwischen dem bakteriellen Symbionten und der Wirtszelle stattfinden konnte. Dies könnte dafür sorgen, dass der bakterielle Symbiont eine vollumfängliche Symbiose mit der höher entwickelten Wirtszelle eingeht und dieser unter bestimmten Voraussetzungen einen Selektionsvorteil verschafft (Wallin, 1925)

Obwohl endosymbiontische Hypothesen über die Ursprünge von Chloroplasten und Mitochondrien zu Anfang des 20. Jahrhunderts sehr viel Anklang fanden (Sapp 1994), wurden diese in der Betrachtung lange vernachlässigt, bis sie 1967 wieder ins Zentrum der Evolutionsforschung gerückt wurden. Lynn Margulis, damals unter dem Namen „Lynn Sagan“, stellte die Hypothese auf, dass Mitochondrien, photosynthetische Plastide, sowie die Basalkörper der Flagellen früher einmal selbst freilebende (prokaryotische) Zellen waren (Sagan 1967). Die Idee eines endosymbiontischen Ursprungs von Mitochondrien und Plastiden war nicht neu, jedoch formulierte Margulis hier eine erste umfassende Theorie der Eukaryogenese.

3.2.2 Phagozytose oder nicht Phagozytose

1967 beschrieb Lynn Margulis nicht nur ihre Hypothese über den prokaryotischen Ursprung von Mitochondrien, photosynthetischen Plastiden, sowie den Basalkörpern der Flagellen, sondern auch eine erste Theorie über den Prozess, wie diese

prokaryotischen Symbionten in die späteren eukaryotischen Zellen aufgenommen wurden. Als ausschlaggebenden Grund für derartige Symbiosen nannte Margulis das Auftreten von freiem Sauerstoff in der Atmosphäre, welcher als Nebenprodukt der Photosynthese entstand (Sagan, 1967). Dies führte zu einer Krise, denn alle Zellen, die diesem freien Sauerstoff nun ausgesetzt waren, mussten sich entweder an eine Atmosphäre mit Sauerstoff anpassen oder aber in speziellen anaeroben Nischen leben. Der erste Schritt für die Entstehung der Eukaryoten war laut Margulis eine Symbiose, die dieses Problem lösen sollte. Dabei wurde eine aerobe prokaryotische Mikrobe (das Protomitochondrium) in das Zytoplasma eines heterotrophen Anaerobiers aufgenommen. Diese beschriebene Endosymbiose wurde obligat und bildete den ersten aeroben amöboiden Organismus (Sagan, 1967). In einer weiteren Symbiose nahmen diese Organismen bewegliche Prokaryoten auf, welche daraufhin die erste Form einer primitiven Geißelamöbe bildeten. Innerhalb dieser Geißelamöben entwickelte sich nach Margulis die klassische Mitose (Sagan, 1967).

Den genauen Mechanismus dieser Aufnahme von verschiedenen Symbionten beschreibt Margulis nicht, jedoch benutzt sie für ihre Beschreibung zumeist den Terminus „ingestion“, welcher übersetzt „Nahrungsaufnahme“ bedeutet und beschreibt somit Phagozytose. Dies ist der Prozess bei dem Eukaryoten über eine Einstülpung ihrer Plasmamembran Partikel mit einer Größe von über 0,4 Mikrometer erkennen und umfließen können. Phagozytose ist ein weit verbreiteter Prozess innerhalb der Eukaryoten, welcher für die Aufnahme und anschließender inneren Verdauung von Nahrungsteilchen genutzt wird (Martin *et al.* 2017, Mills 2020). Dieser Prozess wird in vielen weiteren Theorien als Mechanismus für die Aufnahme von Mitochondrien aufgegriffen. Dabei wird in fast allen Theorien der mitochondriale Vorfahre als unverdaute Mahlzeit angesehen (Doolittle 1998, Cavalier-Smith 2002, Roger *et al.* 2017, Poole und Gribaldo 2014). In einer weiteren Theorie, die Phagozytose als Ursprung der Eukaryoten sieht, beschreibt Cavalier-Smith, dass Phagozytose nicht etwa in Form einer Endosymbiose für die Entstehung der Mitochondrien zuständig war, sondern durch eine Umstrukturierung von Membranen das Organell in einem cyanobakteriellen Vorfahren der Eukaryoten erschuf (Cavalier-Smith 1975).

3.2.3 Heutige Theorien der Eukaryogenese

Heutige Theorien und Modelle der Eukaryogenese lassen sich grundsätzlich in drei Kategorien einteilen, welche nach dem zeitlichen Auftreten der Mitochondrien während der Eukaryogenese kategorisiert werden: „Mitochondrien zuerst“ (engl. *mitochondria first*), „Mitochondrien intermediär“, sowie „Mitochondrien spät“ (engl. *mitochondria late*). Die Wasserstoff-Hypothese von Martin und Müller (1998) gehört zu den Modellen, bei denen die Mitochondrien an den Anfang der Eukaryogenese gesetzt werden. Die Theorie beschreibt eine Symbiose zwischen einem anaeroben, autotrophen Archaeon, welches Wasserstoff benötigt und einem heterotrophen α-Proteobakterium, welches Wasserstoff erzeugt. Die Stoffwechsel-bedingte Symbiose dieser beiden Organismen legte den Grundstein für die Ausbildung aller weiterer eukaryotischen Merkmale (Martin und Müller 1998).

Die Modelle, die ein intermediäres Auftreten der Mitochondrien schlussfolgern, gehen davon aus, dass die Mitochondrien weder für den Beginn der Eukaryogenese nötig sind, noch dass diese als eine der letzten Schritte die Eukaryogenese abschließen. Das „Von innen nach außen“-Modell (engl. *inside-out model*) steht konträr zu vielen anderen Modellen. Hierbei bildet die Wirtszelle kleine Bläschen und gibt diese über ihre Zellwand ab. Diese Bläschen dienten zum Austausch von Stoffen mit den proto-Mitochondrien. Im Laufe der Zeit umschlossen diese Bläschen die proto-Mitochondrien und bildeten das Zytoplasma, das endoplasmatische Retikulum und eine kontinuierliche Plasmamembran (Baum und Baum 2014; Baum und Baum 2020). Da bei diesem Modell ein komplexes Zytoskelett und Endomembransystem für die Symbiose nötig ist, spricht man hier von einem intermediären Auftreten der Mitochondrien.

Das Gegenstrom-Modell (engl. *reverse-flow model*) beschreibt eine archaeale, organo-heterotrophe Wirtszelle, die α-Proteobakterien als Senke für Elektronen benötigt (Spang *et al.* 2019). Eine zeitliche Einordnung der Entstehung einzelner Zellorganelle wird in diesem Modell zwar nicht diskutiert, jedoch wird eine Umschließung der α-Proteobakterien durch Zellausläufer vorgeschlagen, die für einen effizienteren Austausch der Elektronen gesorgt haben könnte (Spang *et al.* 2019).

Eine letzte Theorie für ein intermediäres Auftreten der Mitochondrien ist das E^3-Modell. Diese Theorie wurde von Imachi *et al.* (2020) infolge der ersten erfolgreichen

Kultivierung eines Asgard-Archaeons aufgestellt. Dabei geht die Wirtszelle, ein syntrophisches, fermentatives Archaeon eine Symbiose mit einem Sulfat-reduzierenden Bakterium, sowie mit einem aeroben, organotrophen α-Proteobakterium ein. Das α-Proteobakterium wurde dabei von Membranerweiterungen der Wirtszelle umschlossen und bildete einen Vorläufer des Mitochondriums. Die Symbiose mit dem Sulfat-reduzierenden Bakterium war derweil nur kurzlebig und wurde sukzessive wieder verloren (Imachi *et al.* 2020).

Die Modelle für ein spätes Auftreten der Mitochondrien setzen den Ursprung der Mitochondrien an das Ende der Eukaryogenese. Ein erstes Beispiel hierfür ist das Modell des phagozytierenden Archaeons (engl. *PhAT*) von Martijn und Ettema (2013). Dieses Modell beschreibt die Aufnahme eines anzestralen Mitochondriums durch ein heterotrophes, phagozytierendes Archaeon. Das Mitochondrium wurde nicht verdaut, sondern bildete mit der Wirtszelle eine Symbiose. Für eine derartige Aufnahme des Symbionten ist bereits vorher eine Ausbildung eines komplexen Zytoskeletts, sowie eines Endomembransystems in diesem Modell zwingend notwendig (Martijn und Ettema 2013). Martijn und Ettema (2013) waren zudem der Ansicht, dass sich vor der Aufnahme des Endosymbionten der Zellkern gebildet haben muss, weil dieser eine Barriere gegenüber durch Phagozytose aufgenommene DNA bildet. Diese spezifische Reihenfolge an Events ist jedoch keine zwingend erforderliche Voraussetzung für dieses Modell, da das Mitochondrium zu jedem Zeitpunkt nach der Ausbildung der Phagozytose aufgenommen werden konnte (Donoghue *et al.* 2023).

Ein weiteres Modell, welches ein spätes Auftreten der Mitochondrien beschreibt, ist die Syntrophie-Hypothese (Moreira und López-García 1998). Dabei wird in diesem Modell eine Symbiose zwischen einem Wasserstoff-produzierenden Archaeon, einem Sulfat-reduzierenden Deltaproteobakterium, sowie einem fakultativ aeroben, Sulfid-oxidierenden Alphaproteobakterium beschrieben. Das Archaeon bildet den späteren Zellkern, das Deltaproteobakterium fungiert als Wirtszelle und das Alphaproteobakterium entwickelt sich zum späteren Mitochondrium (Moreira und López-García 1998; López-García und Moreira 1999; López-García und Moreira 2020). In der zeitlichen Abfolge wird in dieser Theorie zuerst das Archaeon in das Deltaproteobakterium aufgenommen und bildet den Zellkern, bevor die Syntrophie mit dem Alphaproteobakterium eingegangen wird.

Das letzte hier zu nennende Modell über ein spätes Auftreten der Mitochondrien ist die serielle Endosymbiontentheorie (engl. *serial endosymbiotic theory, SET*). Diese Theorie wurde mit kleinen Abwandlungen von einer Vielzahl von Wissenschaftlerinnen und Wissenschaftlern publiziert. Die wahrscheinlich bekannteste, jedoch nicht erste Version stammt von Lynn Margulis und wurde in mehreren Publikationen von ihr verfeinert (Sagan 1967, Margulis 1970, Margulis und Fester 1991, Margulis *et al.* 2006). Nach dieser Theorie war die eukaryotische Zelle das Produkt von einer Reihe an langwierigen endosymbiontischen Verbindungen zwischen der anaeroben Wirtszelle und verschiedenen bakteriellen Symbionten. Die bakteriellen Symbionten sind hierbei der cyanobakterielle Vorfahre der Chloroplasten, sowie der alphaproteobakterielle Vorfahre der Mitochondrien (Margulis 2004). Obwohl auf einige wichtige Details in dieser Theorie nicht eingegangen wird, lässt sich schließen, dass die eukaryotische Zelle zuerst ein komplexes Zytoskelett, ein Endomembransystem und einen Zellkern besessen haben muss, bevor das Mitochondrium aufgenommen werden konnte (Donoghue *et al.* 2023). Die serielle Endosymbiontentheorie ist somit klar den Modellen zuzuschreiben, die ein spätes Auftreten der Mitochondrien beschreiben.

3.3 Der letzte gemeinsame eukaryotische Vorfahre

Der letzte gemeinsame eukaryotische Vorfahre (engl. *last eukaryotic common ancestor*, kurz LECA) beschreibt den Vorfahren aller heute lebenden eukaryotischen Organismen und gilt als wichtiges theoretisches Konstrukt in der Erforschung eukaryotischer Evolution. Dabei ist es möglich über die Merkmale von LECA Rückschlüsse auf die Evolutionsgeschichte und explizit der Entstehung der Eukaryoten zu ziehen. Obwohl das Konzept LECA immer mehr in das Zentrum der eukaryotischen Evolutionsforschung fällt, gibt es weiterhin viele offene Fragestellungen und strittige Themenfelder. Prinzipiell lassen sich die rekonstruierten Merkmale von LECA in zwei Gruppen unterteilen: Merkmale, die aufgrund universeller Verbreitung innerhalb heute lebender Eukaryoten auf den gemeinsamen Vorfahren zurückzuführen sind und Merkmale, die aufgrund von vielfältigen Ausprägungen innerhalb heute lebender Eukaryoten nicht eindeutig zu rekonstruieren sind.

Der Konsens attestiert LECA den Besitz von Mitochondrien (Lane und Martin 2010), einen Zellkern (Mans *et al.* 2004, Bapteste *et al.* 2005, Neumann *et al.* 2010), ein endoplasmatisches Retikulum (Kontou *et al.* 2022), lineare Chromosomen mit Zentromeren (Ishikawa und Naito 1999, van Hooff *et al.* 2017), Flagellen (Carvalho-Santos *et al.* 2011, Lindemann 2022), ein Mikrotubuli-organisierendes Zentrum (Yubuki und Leander 2013), einen Nukleolus (Gardner *et al.* 2010, Hoeppner und Poole 2012), sowie die Fähigkeit zur sexuellen Fortpflanzung mittels Meiose (Villeneuve und Hillers 2001, Loidl 2016). Viele weitere Merkmale von LECA, die innerhalb von Eukaryoten in verschiedensten Variationen auftreten, sind bisher nicht in den wissenschaftlichen Konsens aufgenommen worden und werden auch heutzutage in grundlegend gegensätzlichen Theorien beschrieben. Wie in Abschnitt 3.1 bereits beschrieben, ist die Phagozytose eines der Merkmale, die auch heute noch nicht übereinstimmend LECA zugeordnet bzw. aberkannt werden kann. Während sich in vielen Theorien der Prozess der Nahrungsaufnahme bereits in der archaealen Wirtszelle entwickelt haben muss (Doolittle 1998, Spang *et al.* 2015, Zaremba-Niedzwidzka *et al.* 2017, Vosseberg *et al.* 2021), ist diese Annahme nicht mit gefundenen Mikrofossilien (Mills 2020), sowie aus einem physiologischen Standpunkt vereinbar (Martin *et al.* 2017). Ein weiteres bislang umstrittenes Merkmal von LECA ist dessen Mehrkernigkeit. Es zeigt sich sogar, dass LECA in den meisten Fällen als einzelliger Organismus mit nur einem Kern dargestellt wird (Gould und Dring 1979, Cavalier-Smith 1987, Lake und Rivera 1994, Gupta und Golding 1996, Horiike *et al.* 2004, Martijn und Ettema 2013, Martin *et al.* 2015, Imachi *et al.* 2020). Diese Annahme wird zudem meist nicht in Frage gestellt und gilt dementsprechend bei vielen Wissenschaftlern als Konsens (Garg und Martin 2016).

Eine Mehrkernigkeit in LECA hätte jedoch große evolutionäre Vorteile mit sich gebracht. Bei dem Übergang von einer prokaryotischen zu einer eukaryotischen Zelle mussten bei vielen Prozessen innerhalb der Zelle, aber auch der Teilung der Chromosomen, sowie der Zellteilung, Anpassungen vorgenommen werden. Da diese Anpassungen nicht nach einer zuvor erstellten Anleitung vonstattengehen können, sondern das Ergebnis zufälliger Mutationen sind, kann es dazu führen, dass etwaige benötigte Mutationen auf dem Weg zu den heutigen Prozessen zuerst einen Selektionsnachteil hervorrufen, bevor sich durch weitere darauffolgende Mutationen ein Selektionsvorteil bildet. Sind nun mehrere Zellkerne in einer Zelle vorhanden,

können diese schädlichen Mutationen innerhalb eines Zellkerns im Zytosol durch mRNA anderer Zellkerne ausgeglichen werden. Dies hätte zur Folge, dass der Übergang der prokaryotischen zur eukaryotischen Chromosomenteilung abgepuffert werden würde. Zudem würde dieser Prozess dadurch von der evolutionären Hürde des Übergangs von der prokaryotischen zur eukaryotischen Zellteilung, sowie der Chromatinorganisation während des Zellzyklus entkoppelt werden (Brunk und Martin 2019). Trotz dieser Vorteile und einer nachgewiesenen Verbreitung von mehrkernigen Organismen in allen eukaryotischen Supergruppen (Archibald *et al.* 2017, Adl *et al.* 2019), ist die Mehrkernigkeit von LECA weiterhin mehr als umstritten.

3.4 Lateraler Gentransfer

Der laterale Gentransfer (LGT), auch oft als horizontaler Gentransfer (HGT) bezeichnet, spielt eine entscheidende Rolle in der Evolution von Prokaryoten. LGT ist zudem für die Rekombination in Prokaryoten hauptverantwortlich und dabei immer unidirektional. Durch die Vielzahl an LGTs innerhalb der Bakterien und Archaeen ist es bei Prokaryoten durchaus schwierig die Verwandtschaftsbeziehungen dieser in einem klaren phylogenetischen Baum darzustellen. Aufgrund dieser Problematik haben einige Wissenschaftler die Metapher eines Netzwerks für die Erklärung der Verwandtschaftsbeziehungen ins Spiel gebracht (Doolittle 1999, Blais und Archibald 2021).

Bei Prokaryoten gibt es drei prinzipielle Mechanismen die LGT hervorrufen: Transformation, Transduktion und Konjugation (Jones und Sneath 1970). Die Transformation beschreibt hierbei die Aufnahme von freier Desoxyribonukleinsäure (DNA) aus der Umgebung. Dieser Prozess begünstigt den Austausch von genetischem Material zwischen entfernt verwandten Organismen (Ochman *et al.* 2000). Als Transduktion wird der Transfer von DNA mittels Bakteriophagen gekennzeichnet. Die Bakteriophagen haben sich dabei in Spenderzellen reproduziert und Fragmente von DNA in ihr Kapsid aufgenommen. Die Menge wird dabei durch die Größe des Kapsids limitiert, kann jedoch bis zu 100 Kilobasen betragen (Ochman *et al.* 2000). Die Übertragung von DNA ist bei der Transduktion wie bereits bei der Transformation nicht davon abhängig, dass die Spenderzelle, sowie die Empfängerzelle zeitgleich am selben Ort vorhanden sein müssen. Im Gegensatz

hierzu ist bei der Konjugation ein physischer Kontakt zwischen der Spenderzelle und der Empfängerzelle notwendig. Dabei wird über eine Konjugationsbrücke ein Plasmid von der Spenderzelle in die Empfängerzelle übertragen und in das dortige Genom integriert. Zusätzlich zu diesen drei Mechanismen gibt es weitere „nicht-kanonische" Prozesse, welche LGT über Vesikel, Nanoröhrchen und Phagen-ähnliche Gentransfer-Agenten (engl. *gene transfer agents*) hervorrufen (Arnold *et al.* 2022).

Während LGT in Prokaryoten kontinuierlich nachgewiesen und sogar mit dem eigenen Auge beobachtet werden kann, ist dies von Prokaryoten zu Eukaryoten nicht möglich. Deshalb ist es von Vorteil, die Betrachtung von LGT zu Eukaryoten in zwei Bereiche zu unterteilen. Hierbei wird zwischen zwei Arten von LGT unterschieden: LGT von den Vorfahren der Mitochondrien und Chloroplasten zur Wirtszelle während der Endosymbiose, sowie kontinuierliche LGTs von Prokaryoten zu Eukaryoten. Gentransfer von den Vorfahren der Mitochondrien, sowie der Chloroplasten, in das Genom der Wirtszelle wird auch als endosymbiontischer Gentransfer (EGT) bezeichnet und ist gut dokumentiert (Martin 1999, Ku *et al.* 2015). Kontinuierlicher Gentransfer zu Eukaryoten über einen Mechanismus wie Transformation, Transduktion oder Konjugation konnte bislang noch nicht dokumentiert werden, weder von Prokaryoten zu Eukaryoten, noch zwischen verschiedenen Eukaryoten. Im Vergleich zu Prokaryoten wird die Rekombination in Eukaryoten durch Meiose und sexuelle Reproduktion erzielt und ist stets reziprok (Garg und Martin 2016).

Obgleich der fehlenden Belege über die kontinuierliche Akquisition von prokaryotischen Genen mittels LGT in Eukaryoten, wird LGT häufig als einfache Erklärung für phylogenetisch prokaryotische Signale innerhalb von Eukaryoten gewählt. Das wahrscheinlich prominenteste Beispiel hierfür liefert die erste Sequenzierung des menschlichen Genoms (International Human Genome Sequencing Consortium 2001). Die Analyse des Genoms ergab, dass hunderte menschliche Gene auf LGT von Bakterien in die Abstammungslinie der Wirbeltiere zurückzuführen sind. Diese zuerst getroffenen Annahmen wurden allerdings schnell widerlegt und die gefundenen Gene wurden auf differentiellen Genverlust sowie Artefakte in der Analyse zurückgeführt (Salzberg *et al.* 2001, Stanhope *et al.* 2001, Salzberg 2017).

Es liegt auf der Hand, dass dem Gentransfer von Prokaryoten zu Eukaryoten große Hindernisse im Wege stehen. Damit prokaryotisches genetisches Material die

Chromosomen der Eukaryoten erreicht, muss es sowohl in die eukaryotische Zelle gelangen, sowie im weiteren Verlauf in den Zellkern integriert werden (Genereux und Logsdon 2003). Für das Beispiel eines Gentransfers in die Abstammungslinie der Wirbeltiere gibt es ein weiteres Hindernis, welches die Fixierung eines Genes erheblich erschweren würde. Damit eine weitere Vererbung des erhaltenen Gens möglich wäre, müsste das Gen in eine Keimzelle gelangen, da Wirbeltiere bereits in einem frühen Stadium ihrer Entwicklung ihre Zellen spezialisieren (Genereux und Logsdon 2003).
Die Kombination dessen, dass die Prozesse von lateralem Gentransfer von Prokaryoten zu Eukaryoten bisher nicht direkt nachgewiesen werden konnten, sowie der häufigen Annahme, dass lateraler Gentransfer die Ursache für einzelne eukaryotische Homologe zu ansonsten prokaryotischen Genen ist, führt zwangsläufig zu einer weiteren Betrachtung dieser Problematik.

4 Zielsetzung

Die RefSeq-Datenbank des *National Center for Biotechnology Information* ist eine umfassende Sammlung von nicht redundanten Referenzsequenzen (O'Leary *et al.* 2016). In der 228. Veröffentlichung dieser Datenbank (8. Januar 2025) finden sich bereits 2057 komplett sequenzierte Genome von einzigartigen Eukaryoten und mit jeder weiteren Veröffentlichung erhöht sich diese Anzahl. Während das Wissen über aktuell lebende Eukaryoten demnach immer weiter zunimmt und die Analyse dieser Daten stetig voranschreitet, ist es deutlich schwieriger, auf die komplexen Entstehungsprozesse der Eukaryoten vor circa ein bis zwei Milliarden Jahre zurückzublicken und diese zu verstehen.

Fragen, die sich unmittelbar aus der Entstehungsgeschichte der Eukaryoten, sowie ihrer weiteren evolutionären Entwicklung stellen, sind vor allem: i) Welcher Prozess sorgte für die Bildung komplexer eukaryotischer Zellen aus vormals einfachen prokaryotischen Zellen, ii) welche Merkmale und zellulären Strukturen hatte der letzte gemeinsame Vorfahre aller heute lebender Eukaryoten, iii) inwieweit ist lateraler Gentransfer von Prokaryoten zu Eukaryoten vorhanden und welche Rolle spielt dieser?

Vor diesem Hintergrund ist es das Ziel dieser Arbeit, mit Hilfe von phylogenetischen Analysen Merkmale des letzten gemeinsamen eukaryotischen Vorfahrens zu rekonstruieren und Rückschlüsse auf die Entstehungsgeschichte der Eukaryoten zu ziehen. Dazu wurden die Merkmale heutiger Eukaryoten auf phylogenetische Bäume übertragen, um mit Hilfe dieser eine Rekonstruktion des letzten gemeinsamen Vorfahrens durchzuführen. Die rekonstruierten Merkmale und die zeitliche Divergenz ihres Auftretens innerhalb der Eukaryoten geben neue Perspektiven auf mögliche Theorien der Eukaryogenese, sowie der Lebensweise des letzten gemeinsamen eukaryotischen Vorfahrens. Zudem wird in einer weiteren Analyse die Wahrscheinlichkeit des Auftretens von einzelnen eukaryotischen Genen mit Homologen innerhalb der Prokaryoten aufgrund von differentiellem Genverlust anstelle von lateralen Gentransfers untersucht.

5 Publikationen

i. 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, William F. Martin.

Institut für Molekulare Evolution, Heinrich-Heine-Universität Düsseldorf, Universitätsstraße 1, 40225 Düsseldorf, Deutschland.

Dieser Artikel wurde am 01.06.2022 in *Genome Biology and Evolution* Ausgabe 14 veröffentlicht.

Beitrag von Nico Bremer (Erstautor):
Ich habe sämtliche bioinformatischen Analysen zu den bereits erstellten phylogenetischen Bäumen durchgeführt. Dazu gehörten die Rekonstruierungen anzestraler Merkmale, sowie deren Analyse mittels statistischer Tests. Die Visualisierung der Ergebnisse in Abbildungen und Tabellen wurde ebenfalls von mir durchgeführt. Des Weiteren habe ich das initiale Manuskript überarbeitet.

GBE

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

Downloaded from https://academic.oup.com/gbe/article/14/6/evac079/6596370 by Universitaets- und Landesbibliothek Duesseldorf user on 20 February 2024

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 vaculoes. 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 orgin 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,

Downloaded from https://academic.oup.com/gbe/article/14/6/evac079/6596370 by Universitaets- und Landesbibliothek Duesseldorf user on 20 February 2024

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

Downloaded from https://academic.oup.com/gbe/article/14/6/evac079/6596370 by Universitaets- und Landesbibliothek Duesseldorf user on 20 February 2024

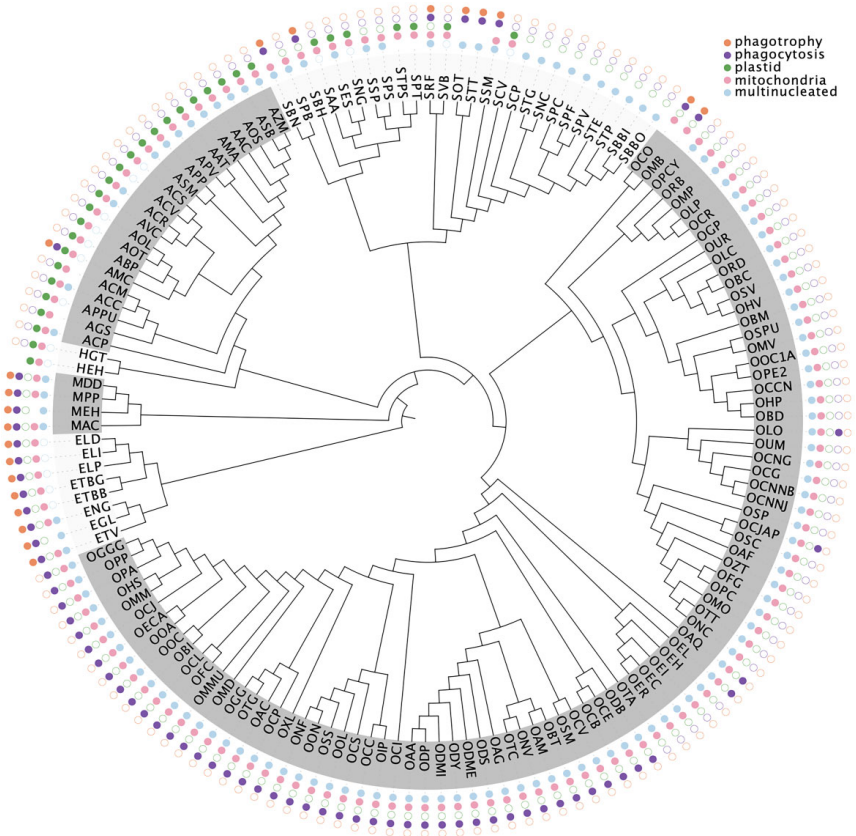

Fig. 1.—Presence (filled circle) absence (empty circle) distribution of five traits in 150 eukaryotic species. Species with no circle for a given trait indicate missing annotation. The reference tree was inferred from the alignment of 18S RNA sequences, rooted on the Excavates branch, with the sole purpose 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 1 848 936 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 in 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 lambia* and the unicellular photosynthetic species *Nannochloropsis gaditana* (supplemental table 1, Supplementary Material online). Reconstructing a reliable species tree without

Downloaded from https://academic.oup.com/gbe/article/14/6/evac079/6596370 by Universitaets- und Landesbibliothek Duesseldorf user on 20 February 2024

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 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 uncontestably 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

Downloaded from https://academic.oup.com/gbe/article/14/6/evac079/6596370 by Universitaets- und Landesbibliothek Duesseldorf user on 20 February 2024

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

Downloaded from https://academic.oup.com/gbe/article/14/6/evac079/6596370 by Universitaets- und Landesbibliothek Duesseldorf user on 20 February 2024

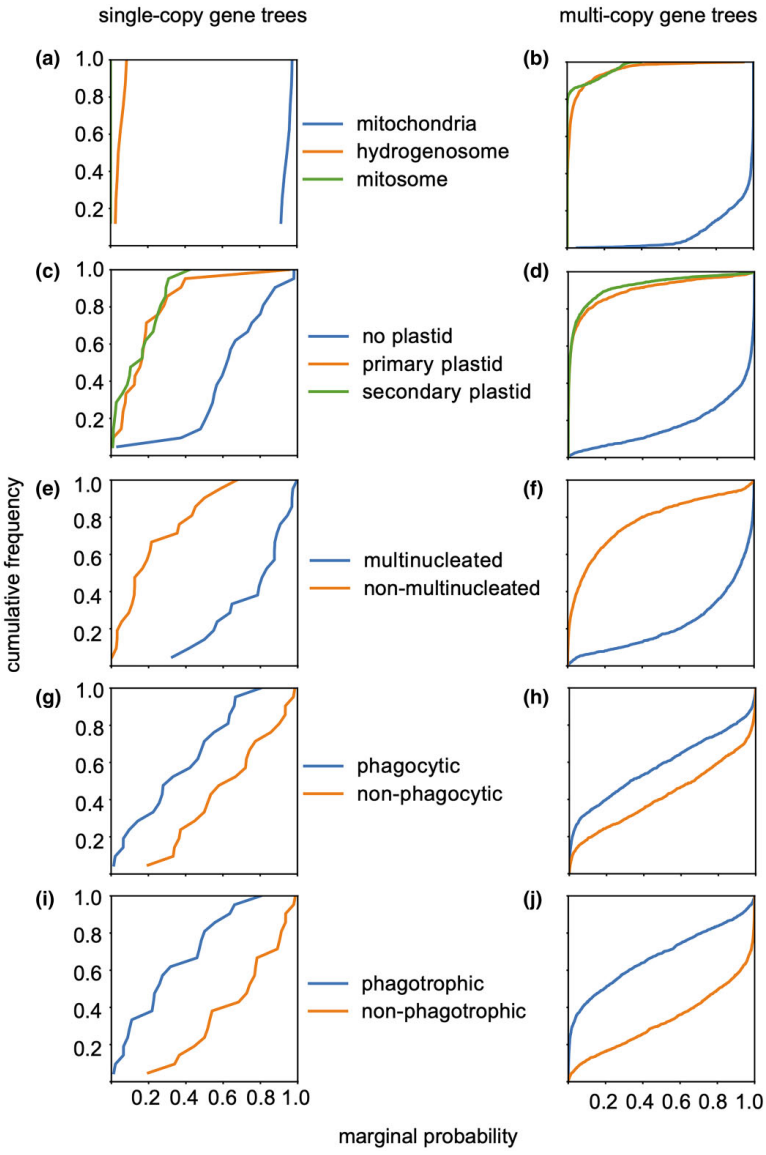

Fig. 2.—Distribution of marginal probabilities for alternative trait-states in LECA across single-copy gene trees (left; without paralogs) and multi-copy gene trees (right; with paralogs). Multinucleate, phagocytosis and phagotrophy were treated as binary traits, while plastids and mitochondria were treated as traits with three states each. For plastids the states were: absence, primary plastid or secondary plastid. For mitochondriam 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 show in table 1. Trait-states with high probabilities in the trees have distributions (colored lines) that are right-shifted in the plots.

Downloaded from https://academic.oup.com/gbe/article/14/6/evac079/6596370 by Universitaets- und Landesbibliothek Duesseldorf user on 20 February 2024

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 ($\text{rho} = -0.4$, $P < 0.01$, 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

Downloaded from https://academic.oup.com/gbe/article/14/6/evac079/6596370 by Universitaets- und Landesbibliothek Duesseldorf user on 20 February 2024

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).

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)

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

[a]**Note**: Numbers indicate mean, median, and standard deviation across trees.

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 reappearance) 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

Downloaded from https://academic.oup.com/gbe/article/14/6/evac079/6596370 by Universitaets- und Landesbibliothek Duesseldorf user on 20 February 2024

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

Downloaded from https://academic.oup.com/gbe/article/14/6/evac079/6596370 by Universitaets- und Landesbibliothek Duesseldorf user on 20 February 2024

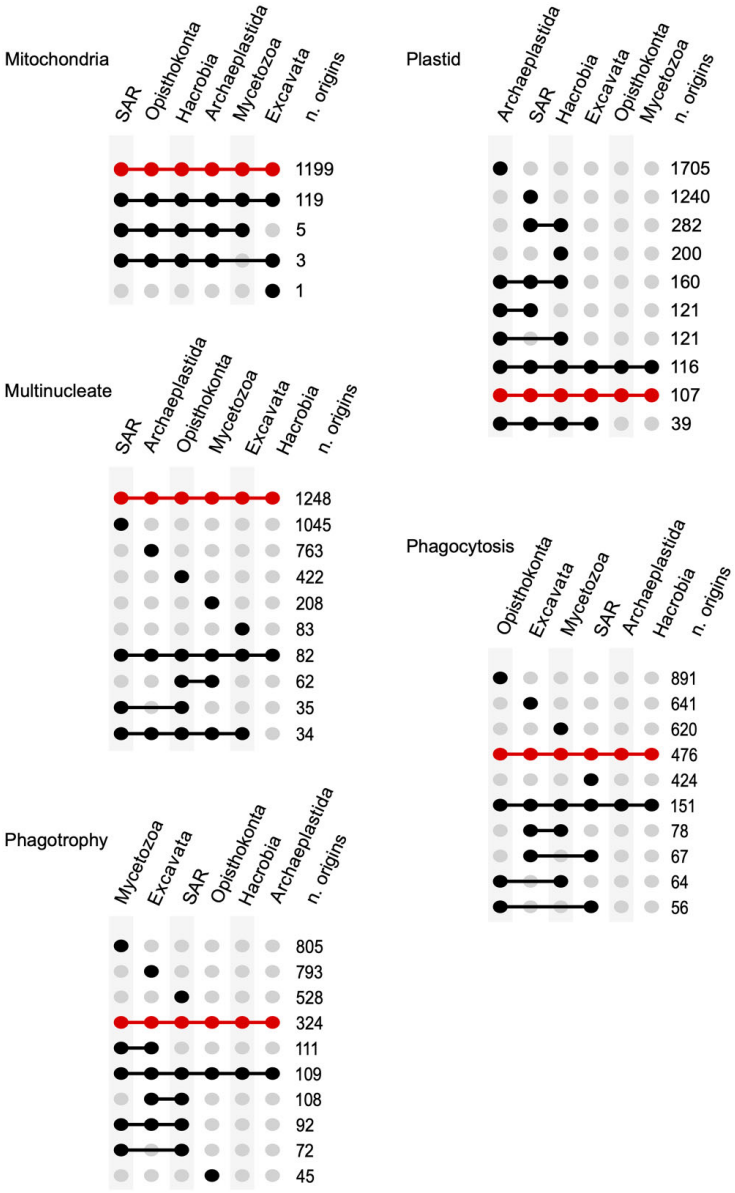

Fig. 3.—Distribution of supergroups descending from origin nodes across 1789 trees. For each internal node reconstructed as a trait origin, all the species (tips) descending from it were used to score an origin to the combination of supergroups (filled circles) to which the descending species belong.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

Downloaded from https://academic.oup.com/gbe/article/14/6/evac079/6596370 by Universitaets- und Landesbibliothek Duesseldorf user on 20 February 2024

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 Transfrom (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 PR^2 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

Downloaded from https://academic.oup.com/gbe/article/14/6/evac079/6596370 by Universitaets- und Landesbibliothek Duesseldorf user on 20 February 2024

in both databases and therefore used alternatives from the same genus. The alignment was generated using MAFFT (Katoh 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 Gα 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.

Downloaded from https://academic.oup.com/gbe/article/14/6/evac079/6596370 by Universitaets- und Landesbibliothek Duesseldorf user on 20 February 2024

Jacobs ME, DeSouza LV, Samaranayake 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 Mereschkowky'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 'Uber 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.
Skejo 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 Symbionticism 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

Downloaded from https://academic.oup.com/gbe/article/14/6/evac079/6596370 by Universitaets- und Landesbibliothek Duesseldorf user on 20 February 2024

ii. The ancestral mitotic state: Closed orthomitosis with intranuclear spindles in the syncytial last eukaryotic common ancestor

Nico Bremer, Fernando D. K. Tria, Josip Skejo, William F. Martin.

Institut für Molekulare Evolution, Heinrich-Heine-Universität Düsseldorf, Universitätsstraße 1, 40225 Düsseldorf, Deutschland.

Dieser Artikel wurde am 08.02.2023 in *Genome Biology and Evolution* Ausgabe 15 veröffentlicht.

Beitrag von Nico Bremer (Erstautor und Korrespondenz):
Ich habe sämtliche bioinformatischen Analysen in dieser Arbeit durchgeführt. Dazu gehören die Rekonstruierungen anzestraler Merkmale, sowie deren Analyse mittels statistischer Tests. Zudem habe ich die Visualisierung von Ergebnissen in Abbildungen und Tabellen durchgeführt. Das initiale Manuskript wurde von mir geschrieben. An der Überarbeitung des Manuskripts war ich ebenfalls beteiligt.

GBE

The Ancestral Mitotic State: Closed Orthomitosis With Intranuclear Spindles in the Syncytial Last Eukaryotic Common Ancestor

Nico Bremer *, Fernando D.K. Tria , Josip Skejo , and William F. Martin

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

*Corresponding author: E-mail: nico.bremer@hhu.de.

Accepted: 03 February 2023

Abstract

All eukaryotes have linear chromosomes that are distributed to daughter nuclei during mitotic division, but the ancestral state of nuclear division in the last eukaryotic common ancestor (LECA) is so far unresolved. To address this issue, we have employed ancestral state reconstructions for mitotic states that can be found across the eukaryotic tree concerning the intactness of the nuclear envelope during mitosis (open or closed), the position of spindles (intranuclear or extranuclear), and the symmetry of spindles being either axial (orthomitosis) or bilateral (pleuromitosis). The data indicate that the LECA possessed closed orthomitosis with intranuclear spindles. Our reconstruction is compatible with recent findings indicating a syncytial state of the LECA, because it decouples three main processes: chromosome division, chromosome partitioning, and cell division (cytokinesis). The possession of closed mitosis using intranuclear spindles adds to the number of cellular traits that can now be attributed to LECA, providing insights into the lifestyle of this otherwise elusive biological entity at the origin of eukaryotic cells. Closed mitosis in a syncytial eukaryotic common ancestor would buffer mutations arising at the origin of mitotic division by allowing nuclei with viable chromosome sets to complement defective nuclei via mRNA in the cytosol.

Key words: last eukaryotic common ancestor, ancestral state reconstruction, mitosis, syncytium, eukaryogenesis.

Significance

Knowledge about the ancestral state of mitosis (nucleus, chromosome, and cell division) in eukaryotes would shed light on the biology of the last eukaryotic common ancestor (LECA). To address that question, we used methods of ancestral state reconstruction to ascertain the type of mitosis present in the LECA. We found that LECA did not disintegrate its nuclear membrane at chromosome division, but instead kept the nuclear membrane intact so that it divided by a process similar to constriction. The chromosomes were pushed apart by microtubules that formed within the mother nucleus. The data indicate that nuclear division took place without cell division in LECA, giving its cells a filamentous multinucleated state. This reconstructed state sheds light on an important aspect of the prokaryote to eukaryote transition.

Introduction

The origin of eukaryotes is a classical topic of debate. There was a time, not too long ago, when the prospect was discussed that prokaryotes arose from eukaryotes (Forterre and Phillippe 1999; Poole et al. 1999). Today it is now generally agreed that eukaryotes arose from prokaryotes, that the endosymbiotic event that led to mitochondria played a role in their origin, that eukaryotes and mitochondria share a single common origin, and that eukaryotes therefore share a last eukaryotic common ancestor (LECA). Based upon the universal distribution of the traits among major eukaryotic groups, it is furthermore agreed that LECA possessed, in addition to mitochondria (Lane and Martin

© The Author(s) 2023. 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

Downloaded from https://academic.oup.com/gbe/article/15/3/evad016/7031494 by Universitaetsbibliothek Duesseldorf user on 21 February 2024

2010), a nucleus (Mans et al. 2004; Bapteste et al. 2005; Neumann et al. 2010), an endoplasmic reticulum (Kontou et al. 2022), linear chromosomes with centromeres (Ishikawa and Naito 1999; van Hooff et al. 2017), flagellae (Carvalho-Santos et al. 2011; Lindemann 2022), microtubule organizing centers (Yubuki and Leander 2013), nucleoli (Gardner et al. 2010; Hoeppner and Poole 2012), meiosis, and sex (Villeneuve and Hillers 2001; Loidl 2016). Those traits are easily traced to LECA because they are present in all eukaryotes. Yet eukaryotes exhibit almost boundless cytological and morphological diversity, leaving the biological nature of the LECA far less clearly resolved than one might tend to think (Katz 2012; Koumandou et al. 2013; Booth and Doolittle 2015; López-García and Moreira 2015; Porter 2020; Gabaldón 2021; Roger et al. 2021; Mills et al. 2022).

The traditional method of inferring information about the nature of any last common ancestor is to construct an evolutionary tree for the group and assign a root, in hope that traits that are variable across the tree might map to the rooted tree in such a manner as to reveal the state of the trait at the root, ideally, without conflicting data (Jermann et al. 1995; Kohn et al. 1996; Gold et al. 2015; Klim et al. 2018). The traditional approach is very difficult for eukaryotes, however, because there is little agreement among experts (Williams 2014; Keeling and Burki 2019; Burki et al. 2020) and little agreement across molecular data sets (Stechmann and Cavalier-Smith 2003; Richards and Cavalier-Smith 2005; Kim et al. 2006; Rodríguez-Ezpeleta et al. 2007; Cavalier-Smith 2009; Roger and Simpson 2009; Rogozin et al. 2009; Derelle and Lang 2012; Katz et al. 2012; He et al. 2014; Cerón-Romero et al. 2022) as to the position of the root in the eukaryotic tree. Two main causes are discussed for the differing pictures concerning the position of the eukaryotic root: (1) the time between the divergence of major eukaryotic supergroups may have been very short in the sense of a "radiation" rendering resolution difficult (Philippe et al. 2000; Eme et al. 2014), and (2) hundreds of gene duplications that took place in the genome that led to LECA, generating vast amounts of hidden paralogy in gene trees hence discordant placements of roots (Tria et al. 2021), or both.

For traits that are universal among eukaryotes, reconstruction to LECA is trivial. For traits that are not universal, reconstruction of the trait in LECA requires more work. One example is phagocytosis, the ability to eat and digest other cells as food. Many lineages of eukaryotes possess phagocytosis but many do not; reconstruction of the trait indicates that LECA was not phagocytotic (Bremer et al. 2022). Many lineages of eukaryotes possess multinucleated states that are distinct from those generated during meiosis, whereas many eukaryotes lack such multinucleated (syncytial) states; reconstruction of the trait indicates that LECA had a syncytial (multinucleated) habit (Skejo et al. 2021; Bremer et al. 2022). An ancestrally multinucleated state for LECA bears upon the nature of mitosis in LECA because there exist a variety of mitotic types in eukaryotes which differ in their compatibility with the syncytial habit. In the present study, we are interested in reconstructing the ancestral state of mitosis in LECA.

Though LECA possessed the molecular machinery required for mitotic chromosome division (Tromer et al. 2019), there is also no doubt that LECA possessed meiotic sex (Speijer et al. 2015), leaving open the question of whether mitosis preceded meiosis on the path to LECA or vice versa (Garg and Martin 2016). The state of mitosis in LECA is the focus of our present study. Mitotic types across the eukaryotic tree are diverse. The greatest differences are the state of the nuclear envelope and the position and symmetry of the spindles (fig. 1). Different combinations of those traits can be found across the eukaryotic tree. The nuclear envelope can either remain intact during mitosis (closed mitosis) or it can be partly or completely dispersed (semi-open or open mitosis). When the nuclear envelope remains intact, the position of the central spindle can either be intranuclear or extranuclear. The symmetry of the spindles can either be axial (orthomitosis) or bilateral (pleuromitosis). Almost all combinations of those three traits can be found in eukaryotes, with the exception of open pleuromitosis and closed extranuclear orthomitosis that are topologically self-exclusive (Raikov 1994). During open orthomitosis, the nuclear envelope dissolves completely and the spindles have an axial symmetry. This form of mitosis can be found for example in the algal species *Chilomonas paramecium* (Heywood 1988) and *Isochrysis galbana* (Hori and Green 1985). If the nuclear envelope disperses only partly and the symmetry of spindles is axial, the mitotic type is called semi-open orthomitosis. This process is not universal and can be found in several variants (reviewed in Raikov 1994). Examples for this type of mitosis have been found in *Amoeba proteus* (Gromov 1985) and the algal flagellates *Pavlova lutheri* and *Pavlova salina* (Green and Hori 1988). During semi-open pleuromitosis, the nuclear envelope disperses partly and the spindles have a bilateral symmetry. This type of mitosis is typical for species of the phylum Apicomplexa and was found for example in *Aggregata eberthi* (Bělař 1926). Species that perform closed intranuclear pleuromitosis have an intact nuclear envelope with bilaterally symmetrical intranuclear spindles. This constellation can be found in a variety of species across the eukaryotic tree, including for example fungi, Kinetoplastida, and Haplosporidia (Heath 1980). Closed intranuclear orthomitosis is characterized by an intact nuclear envelope with intranuclear spindles that have an axial symmetry. Multiple variants of this type can be found in eukaryotes (reviewed in Raikov 1994). One variant for example has been found in the testate amoebae *Arcella vulgaris* (Raikov and Mignot 1991). The remaining combination of states is a closed extranuclear pleuromitosis. The interesting

Downloaded from https://academic.oup.com/gbe/article/15/3/evad016/7031494 by Universitaetsbibliothek Duesseldorf user on 21 February 2024

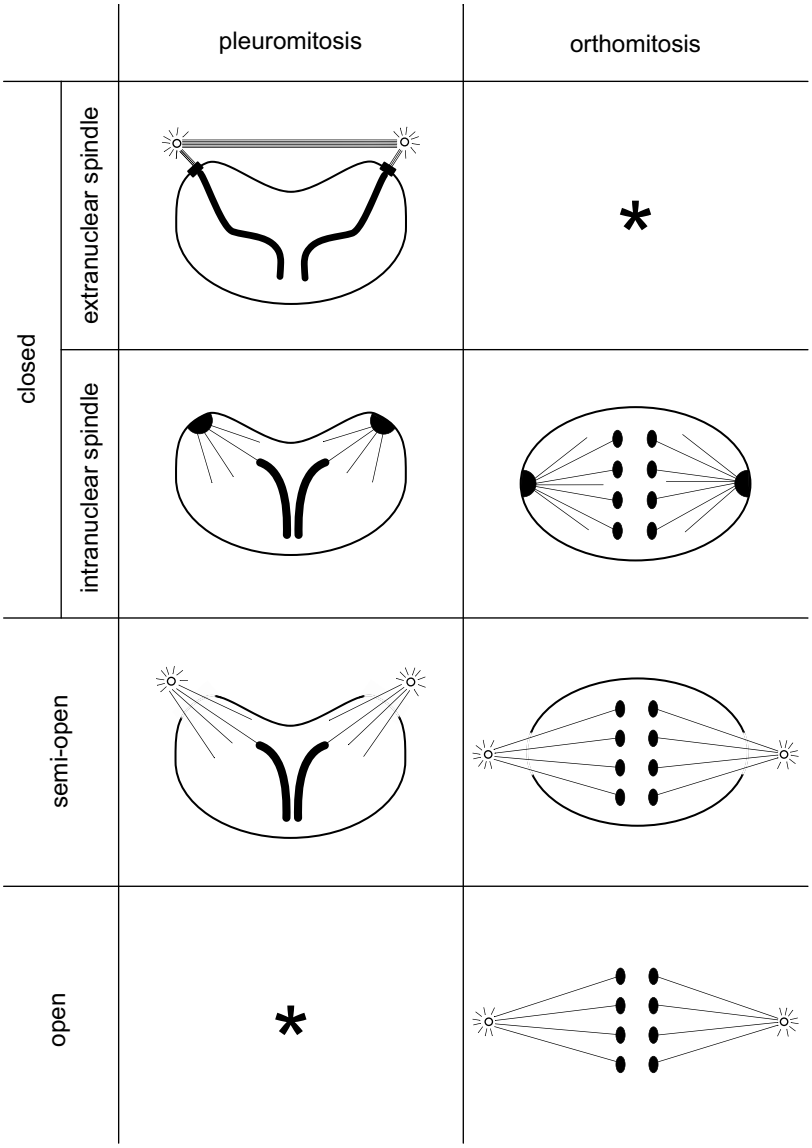

Fig. 1.—Mitotic traits and combinations studied in this manuscript (using symbolism of Raikov 1994). During mitosis, the nuclear envelope can remain intact (closed), disperse partially (semi-open), or disperse completely (open). If the nuclear envelope stays intact during mitosis, the spindles can be either intranuclear or extranuclear. The symmetry of spindles is divided into axial symmetry (orthomitosis) and bilateral symmetry (pleuromitosis). An asterisk (*) indicates that the combination is not possible according to Raikov (1994).

Downloaded from https://academic.oup.com/gbe/article/15/3/evad016/7031494 by Universitaetsbibliothek Duesseldorf user on 21 February 2024

part of this type of mitosis is that the spindles are extranuclear with a bilateral symmetry although the nuclear envelope stays intact. One prominent species with this type of mitosis is *Trichomonas vaginalis* (Ribeiro et al. 2005).

Different types of mitosis have been observed within eukaryotes, but there is no consensus concerning the type of mitosis within LECA. This is mainly due to open and closed mitosis being widespread across various groups of eukaryotes (Sazer et al. 2014). It has been suggested that closed mitosis must have been the ancestral state as it occurs among suspectedly primitive or simple eukaryotic organisms (Pickett-Heaps 1969; Leedale 1970; Pickett-Heaps 1974). Though it has been suggested that mitosis is never completely open mitosis nor completely closed (Dey and Baum 2021), the terms have standard meaning and eukaryotes studied can be classified along that spectrum. Open mitosis of animals and streptophytes has been interpreted as convergent secondary adaptions (Cavalier-Smith 2010). Another correlation concerns the fate of the nuclear envelope and the size of the eukaryotic cell. A larger cell results in a larger nucleus due to the classical "Kernplasmarelation" or karyoplasmic ratio (Hertwig 1903; Jorgensen et al. 2007; Neumann and Nurse 2007). As a consequence of this, a larger nucleus has a larger change in surface during mitosis. The amount of additional membrane that has to be produced in order to perform closed mitosis could force large cells to change their mitosis to an open mitosis (Boettcher and Barral 2013).

Some species can exhibit more than one mitotic type depending on the life cycle stage. For example, the slime mold *Physarum polycephalum* can form a syncytial plasmodium with multiple nuclei but can also exist as a uninucleate amoeba. Depending on the phase of its life cycle, its mitotic type changes. In its syncytial (multinucleated) state, it undergoes closed mitosis and switches to open mitosis during its uninucleate phase (Solnica-Krezel et al. 1991; Tanaka 1973). This mitotic polymorphism during different phases of the life cycle is also seen in *Physarum flavicomum*. Throughout its myxamoebal form the nuclear envelope disperses during prometaphase and remains absent until telophase, whereas during its plasmodial form, the nuclear envelope remains intact and slightly discontinuous at the poles in late stages (Aldrich 1969). Closed nuclear division with intranuclear spindles is typical for cells with a syncytial (multinucleated) habit. This is because open mitosis or extranuclear spindles in a syncytium would lead to microtubule attachment to chromosomes from different nuclei, hence missegregation of chromosomes and therefore a failing mitosis (De Souza and Osmani 2007). Skejo et al. (2021) recently published ancestral state reconstructions (ASRs) indicating that LECA possessed a syncytial morphology, in contrast with standard depictions of LECA as a mononucleated cell, but consistent with earlier suggestions (Garg and Martin 2016) that the syncytial habit of LECA would dramatically ease the transition from prokaryotic to eukaryotic cell division. Here we investigate the ancestral state of mitosis in LECA.

Results and Discussion

Framework and Data

We used a data set of 4 eukaryotic traits (3 mitotic traits)—nuclear envelope during mitosis, symmetry of the spindle apparatus, the position of the central spindle in the presence of an intact nuclear envelope, and sexual reproduction, as well as the distribution of these traits across 150 eukaryotic species spanning a total of 6 lineages: Opisthokonta, Archaeplastida, Hacrobia, Excavata, SAR, and Mycetozoa (fig. 2; see Methods for details). We performed ASR in order to time the origin of these traits relative to the LECA. We clustered 1,848,936 protein-coding genes from the 150 eukaryotic genomes using MCL (Enright et al. 2002) and obtained a total of 239,012 gene families as previously described (Bremer et al. 2022). Since the reconstruction of a reliable eukaryotic species tree remains challenging and the position of the root in the eukaryotic tree is still debated (Williams 2014; Keeling and Burki 2019; Burki et al. 2020), we used a total of 1,789 gene families with at least one representative species of each of the six supergroups. The reconstruction of a reliable species tree is furthermore complicated by the paucity of "universal" orthologs in the data set. The causes for this are frequent gene duplications and gene losses.

Our approach of analyzing 1,789 rooted gene trees—instead of one or a few published rooted species trees—covers a wider range of phylogenetic history recorded in genes. Each eukaryotic gene tree has its own history and therefore the root position will vary across different trees. This is important because eukaryotic evolution (and evolution in general) is, obviously, not recorded or reconstructed the same for each gene. If all eukaryotic genes tended to generate exactly the same tree, the eukaryotic tree would have been inferred, and rooted, decades ago. One could also argue that our method has some similarities with conventional methods. The analysis of widely distributed genes, in our case distributed in each supergroup, is similar to the summation of signals across a sample of gene trees in the case of building a consensus tree. In our method, we have the benefit that the individual phylogenetic signal of each gene is recorded because the analyzed gene trees were reconstructed from independent phylogenetic markers.

LECA Reconstructs With Closed Intranuclear Orthomitosis and Sexual Reproduction

We labeled the species at the tips of each tree according to their trait-state annotations and performed maximum-likelihood ASR. A gene tree was informative for ASR of a given trait if it contained representative species for both possible trait states. Trees with only one trait state across all annotated tips were not considered for ASR as they

Downloaded from https://academic.oup.com/gbe/article/15/3/evad016/7031494 by Universitaetsbibliothek Duesseldorf user on 21 February 2024

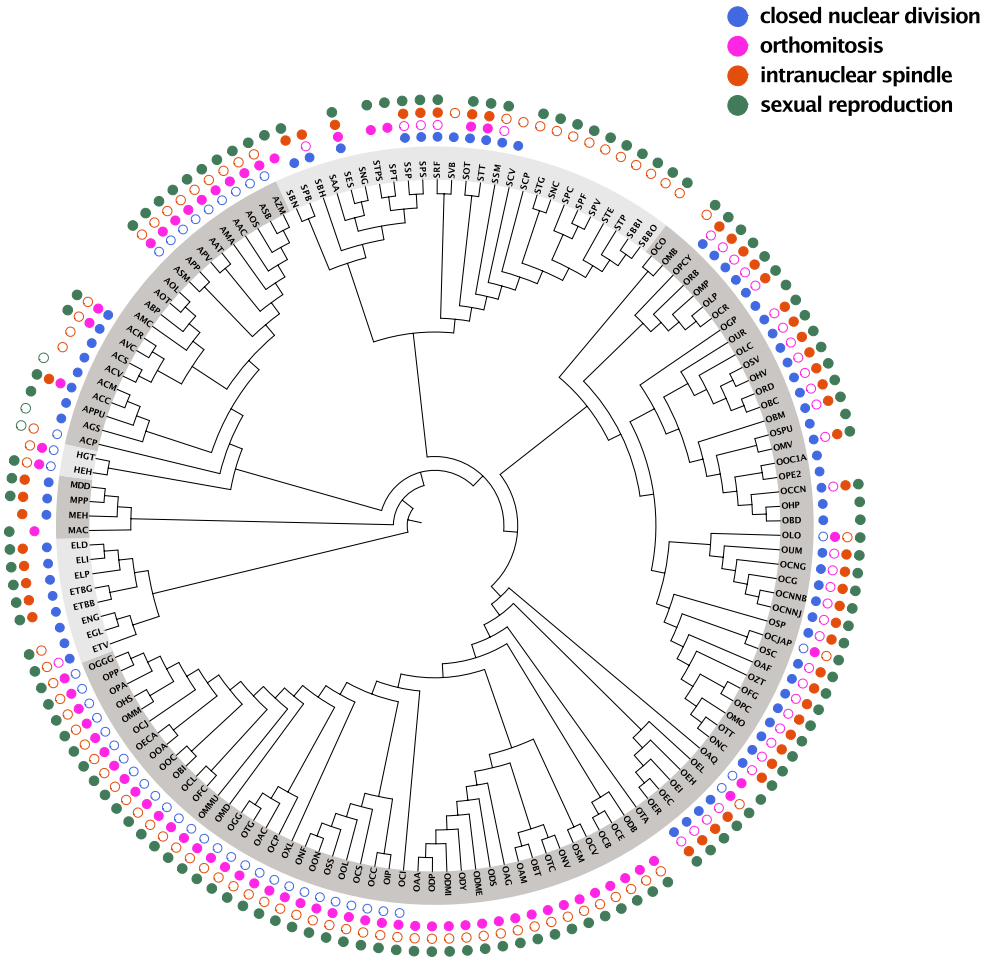

Fig. 2.—Presence (filled circle) absence (empty circle) distribution of 4 traits in 150 eukaryotic species. Species with no circle for a given trait indicate missing annotation. The reference tree was inferred from the alignment of 18*S* RNA sequences, rooted on the Excavates branch, with the sole purpose of data display (see Methods). Tip labels are species codes (see supplementary table S1, 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.

were uninformative. Due to the fact that each tree has at least one representative from each of the six eukaryotic supergroups, the root corresponds to LECA. In a first step, we summarized the ASR across all trees by counting the frequency of each trait-state appearance in LECA (table 1). High frequencies indicate the most likely state of a trait in LECA, whereas low frequencies indicate lineage specific origins of the trait or errors.

With this majority rule, we found that 90% of the trees reconstruct a closed nuclear envelope at the root node and therefore in LECA. An absence of this trait state in LECA was only found in 2% of the trees. The remaining 8% had ambiguous results at the root node. Our analysis also shows that 65% of the trees recover orthomitosis as the ancestral state, whereas only 9% result in pleuromitosis in LECA. The remaining 26% of the trees have unresolved

Downloaded from https://academic.oup.com/gbe/article/15/3/evad016/7031494 by Universitaetsbibliothek Duesseldorf user on 21 February 2024

Table 1
Maximum-likelihood Ancestral Reconstruction of 4 Traits From 150 Eukaryotic Species, Across a Broad Sample of Gene Trees as Estimates of the Underlying Phylogeny

Trait	Presence	Absence	Ambiguous	Total (*N*)
Single-copy gene trees				
Closed nuclear division	13 (62%)	0	8 (38%)	21
Orthomitosis	8 (40%)	0	12 (60%)	20
Intranuclear spindle	3 (15%)	3 (15%)	14 (70%)	20
Sexual reproduction	10 (91%)	0	1 (9%)	11
Multicopy gene trees				
Closed nuclear division	1,595 (90%)	29 (2%)	144 (8%)	1768
Orthomitosis	1,152 (66%)	154 (9%)	448 (25%)	1754
Intranuclear spindle	782 (44%)	491 (28%)	494 (28%)	1767
Sexual reproduction	1,480 (98%)	2 (0.1%)	23 (1.5%)	1505

NOTE.—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. All analyzed traits were modeled as binary traits.

ASR for that trait. For the third trait, we reconstructed intranuclear spindle in 44% of the analyzed trees. The absence of this trait, and therefore the presence of extranuclear spindles, was recovered only in 28% of the trees and the remaining 28% of the trees had ambiguous results. The majority rule indicates that LECA had a closed nuclear envelope with axial symmetry of the spindle apparatus and intranuclear spindles.

The reconstruction of the sexual reproduction resulted in 98% of the trees with this trait being present in LECA and <2% with ambiguous results. The absence of this trait was only recovered in 0.1% of all analyzed trees. The presence of sexual reproduction in LECA is in accordance with the conservation of meiosis across all major eukaryotic groups despite a remarkable variability of this process across the eukaryotic tree (Egel and Penny 2007; Loidl 2016).

Whereas the reconstructions of a sexual reproduction and a closed nuclear envelope were robust, the analyses for orthomitosis and intranuclear spindles had a higher proportion of unresolved reconstructions, with both however still favoring presence of the trait in LECA. To further analyze these ambiguities, we tested the statistical significance of our results by matching marginal probabilities of all possible trait states for each tree (fig. 3). This was done regardless of the reconstruction results in LECA (trait-presence, trait-absence, or unresolved). The differences in distributions were assessed using the Wilcoxon-signed-rank test. This test is a refinement relative to the majority rule as it not only uses the reconstruction result at LECA, but also takes the magnitude of probabilities into account. The advantage of this analysis is that we can also look at trees with unresolved ASR in LECA. The two-tailed Wilcoxon tests indicate that closed nuclear envelope, orthomitosis, intranuclear spindle, and sexual reproduction were present in LECA at $P < 0.01$.

The 200 Best Trees for Tree Quality, Sampling, and Conflicting Evidence Recover the Same Reconstructions

Ancestral state reconstructions depend on the quality of the underlying gene trees. This quality can be influenced by the sequence alignment, the rooting or species sampling. In order to show that our reconstructions with 1,789 trees are not the result of low-quality trees, we examined the tree quality for eight independent criteria by analyzing only the top 200 trees for each criteria individually. Quality of sequence alignments was tested by performing heads or tails (HoT) analyses (Landan and Graur, 2007). For this, we compared the original alignment (heads) against the alignment that was obtained using the same sequences, but in reverse amino acid order (tails). Our analysis showed that tree quality in the majority of our 1,789 trees is high. Most trees have a mean column score above 0.6 and a mean residue pair score above 0.9. Additionally, ASRs of only the best 200 trees according to both scores individually uncover the same reconstructions (supplementary fig. S1, Supplementary Material online). The Wilcoxon tests are significant for almost all traits, except for the intranuclear spindle, yet the reconstruction of closed orthomitosis as the ancestral state inevitably leads to intranuclear spindle as this is the only possible combination of these three trait states (see fig. 1). Therefore, the alignment quality does not impact our results obtained with 1,789 gene trees.

Tree rooting can have an influence on the results of ASRs. The rooting method we used for our trees is the minimal ancestor deviation (MAD) approach (Tria et al. 2017). It has been shown in independent studies that this approach outperforms other current rooting methods (Wade et al. 2020; Lamarca et al. 2022). Additionally, MAD does not require an outgroup for rooting a gene tree. Here, we analyzed two root scores: the ancestor deviation (AD) statistic and the root ambiguity index (AI). AD scores measure the degree of deviation from the molecular clock corresponding to the

Downloaded from https://academic.oup.com/gbe/article/15/3/evad016/7031494 by Universitaetsbibliothek Duesseldorf user on 21 February 2024

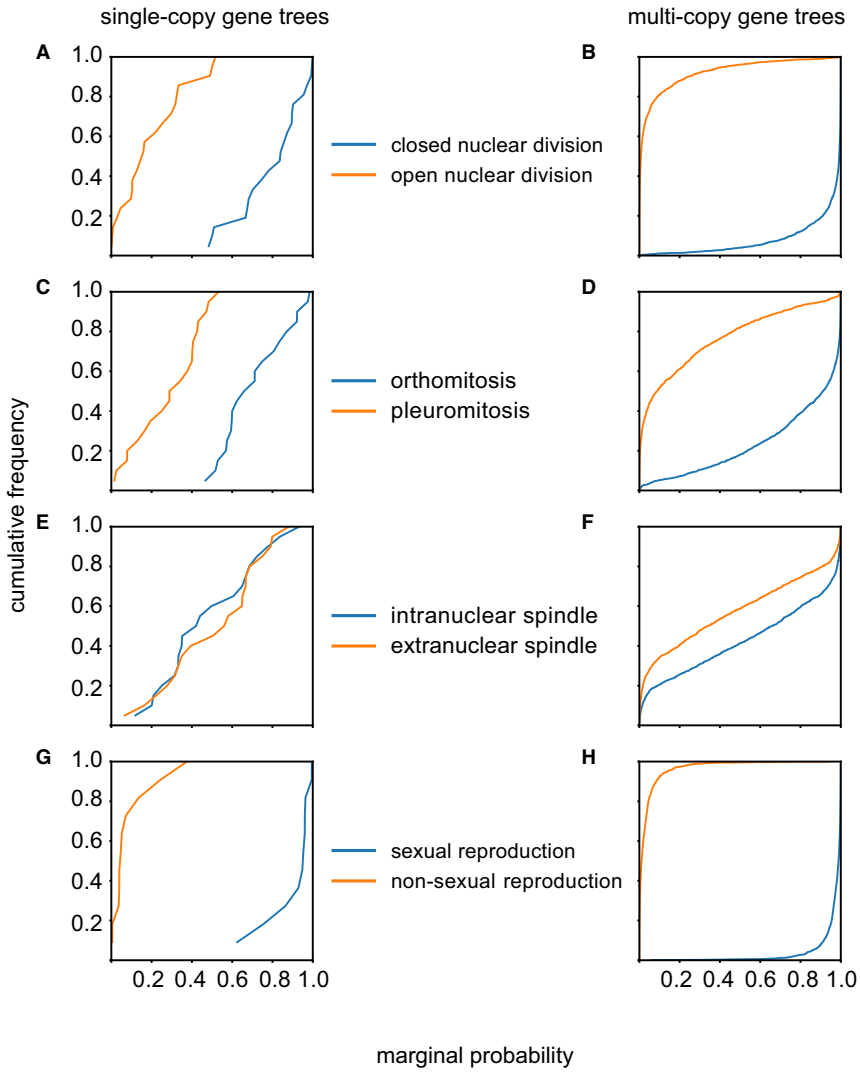

Fig. 3.—Distribution of marginal probabilities for alternative trait states in LECA across a maximum of 21 single-copy gene trees (A,C,E,G; without paralogs) and a maximum of 1,768 multicopy gene trees (B,D,F,H; with paralogs). All traits were treated as binary traits. The number of trees used in the analyses are show in table 1. Trait states with high probabilities of reconstructing to LECA in the trees have distributions (colored lines) that are right shifted in the plots, for example, sexual reproduction (*G*, *H*).

inferred root. AI scores are the ratio of AD scores for the inferred root over the second-best root. Both scores were noticeably similar for trees uncovering different trait states at the root node. This suggests that our rooting method did not cause significant bias during reconstructions. The Wilcoxon tests for the top 200 trees judged by AD and AI recovered the same reconstruction results, with the exception of intranuclear spindle being not

Downloaded from https://academic.oup.com/gbe/article/15/3/evad016/7031494 by Universitaetsbibliothek Duesseldorf user on 21 February 2024

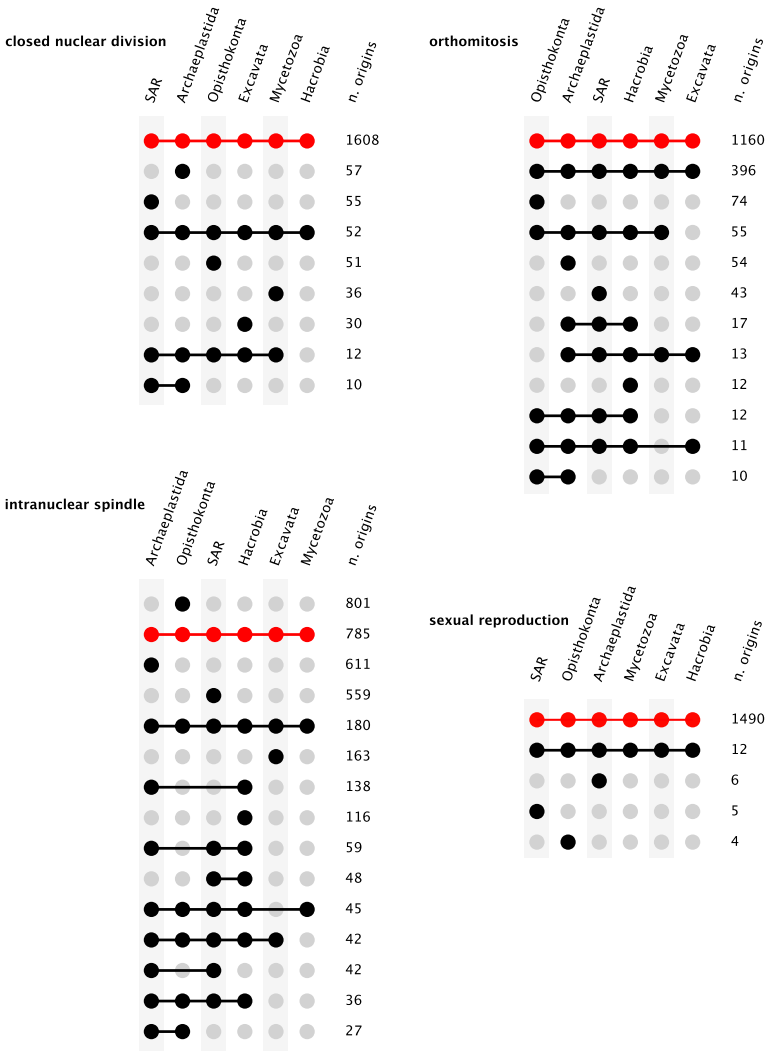

Fig. 4.—Distribution of supergroups descending from origin nodes across 1,789 trees. For each internal node reconstructed as a trait origin, all the species (tips) descending from it were used to score an origin to the combination of supergroups (filled circles) to which the descending species belong. Origins at the root node (LECA) are shown in red (number of origins: 1608; 1160; 785; 1490).

significant (supplementary fig. S1, Supplementary Material online). As explained above, the only possible combination having a closed mitosis and axial spindle (orthomitosis) requires intranuclear spindle (see fig. 1).

Another aspect that can influence the results of ASRs are the sampled species within the analysis. For the construction of gene clusters, we avoided metagenomic and transcriptomic sequences. It has been shown recently that

Downloaded from https://academic.oup.com/gbe/article/15/3/evad016/7031494 by Universitaetsbibliothek Duesseldorf user on 21 February 2024

there is more contamination, base calling and assembly errors in those sequences (Garg et al. 2021). We relied on genomic sequences mainly from RefSeq (O'Leary et al. 2016; see Supplementary file 1, Supplementary Material online). It is inevitable due to our tree selection (at least one species from each eukaryotic supergroup) that gene families are not uniformly distributed across the sampled genomes. We therefore tested four different criteria to investigate the effect of differential sampling: (1) the total number of operational taxonomic units; (2) the total number of species; (3) the fraction of basal lineages (Excavates and Mycetozoa) relative to Opisthokonts; and (4) the fraction of the least frequent trait state occurring at the tips of our trees. The Wilcoxon tests for the top 200 trees independent for all four sampling criteria significantly recovered the same reconstructions we found while using 1,789 trees. The only exception is, once again, the intranuclear spindle being only significantly recovered in two out of the four cases: total number of species and fraction of the least frequent trait state (supplementary fig. S1, Supplementary Material online). It is still the only possible state for this trait due to the other significant reconstructions (see fig. 1).

Timing Trait Origins Relative to the Emergence of Eukaryotic Supergroups

We investigated eukaryotic species that descended from internal trait origin nodes in order to time the origin of those traits relative to the divergence of the six eukaryotic supergroups that were considered here: Opisthokonta, Archaeplastida, SAR, Hacrobia, Excavata, and Mycetozoa. For this, we recorded the combination of descending species of each internal origin node for each trait across all trees. The distribution of those combinations of supergroups was plotted (fig. 4). Internal origin nodes are defined as internal nodes with a newly acquired trait state that is not present in its parent node. We distinguish between a descending combination of all six supergroups that has its trait origin at the root node (six red circles) from trait origins with descendants from each supergroup that were found at other internal nodes (six black circles). The former case identifies the presence of the trait at the root (LECA). The latter case identifies the presence of a node in the tree that subtends descendants of all six supergroups but is not the root (LECA), which would be compatible with trait origin in LECA but could also be the consequence of phylogenetic error or duplications (Bremer et al. 2022).

Combinations with a high frequency (fig. 4) likely represent a true origin node in the underlying supergroup phylogeny. Low-frequency supergroup combinations can be interpreted as likely conflicting results. In three out of the four analyzed eukaryotic traits the results are very clear. For the trait "closed nuclear division," we see that the majority of origin nodes is found in the root node of the tree ($n = 1{,}608$). The same was observed for the trait "sexual reproduction." Almost all origins of the trait were found in LECA ($n = 1{,}490$). "Orthomitosis" also has the majority of origins in the root node ($n = 1{,}160$) followed by additional origins with descendants of all six supergroups that were not at the root node ($n = 396$). Origin nodes for "intranuclear spindles" were found with a wider spectrum of combinations of descending supergroups. The highest number of origins was found in Opisthokonta ($n = 801$) followed closely by origins within LECA ($n = 785$), Archaeplastida ($n = 611$) and SAR ($n = 559$), indicating that for the present sample, intranuclear spindles are more common in opisthokonts than in other supergroups.

Number of Origins of Different Traits Across Eukaryotic Trees

When looking at the origin of a trait in the tree, it is not only of interest to investigate the ancestral state, but also how

Table 2
Summary Statistics for the Number of Trait Origins Across Trees

Trait	Number of origin[a]		
	Terminal nodes	Internal nodes	All nodes
Single-copy gene trees			
Closed nuclear division	2, 1, 2.97	1.23, 1, 0.88	3.23, 2, 3.59
Orthomitosis	1.45, 1, 1.88	1, 1, 0.56	2.45, 2, 1.90
Intranuclear spindle	1.8, 2, 1.40	0.85, 1, 0.93	2.65, 2.5, 1.27
Sexual reproduction	0.64, 0, 1.57	1.18, 1, 0.75	1.82, 1, 1.60
Multi-copy gene trees			
Closed nuclear division	0.69, 0, 1.83	1.64, 1, 1.58	2.33, 1, 3.17
Orthomitosis	0.43, 0, 2.69	1.22, 1, 0.68	1.6, 1, 2.89
Intranuclear spindle	1.94, 2, 1.90	2.35, 2, 1.38	4.29, 4, 2.88
Sexual reproduction	0.75, 0, 1.26	1.38, 1, 1.07	2.13, 1, 2.04

Note.—Trait origin in internal and terminal nodes are distinguished. Single-copy trees (without paralogs) were distinguished from multicopy trees (with paralogs).
[a]Numbers indicate mean, median, and standard deviation across trees.

Downloaded from https://academic.oup.com/gbe/article/15/3/evad016/7031494 by Universitaetsbibliothek Duesseldorf user on 21 February 2024

many times a trait arose within eukaryotes. In order to analyze this, we counted the number of origins for each analyzed gene tree. The average number of origins of all analyzed traits are shown in table 2. Despite no trait having a clear single origin, the number of average origins is still comparatively low. This makes sense as all analyzed traits have been reconstructed to being ancestral in LECA. The additional origins within the trees can be the results of turnovers of these traits. As we already highlighted above, the syncytial state of a cell favors a closed mitosis and therefore a change in the lifestyle could have also changed the traits analyzed here. We have recently shown that LECA was multinucleated, but the trait itself had a high turnover rate ranging on average from three to seven origins per tree (Bremer et al. 2022). The present data indicate that mitotic traits evolved more stable than the multinucleated state, whereby the presence of an intracellular spindle and closed mitosis are required for the syncytial state to become manifest.

Conclusions

Despite the reconstruction of LECA being syncytial with closed orthomitosis using intranuclear spindles, is it still possible that open mitosis was somehow present in LECA but escaped identification? The key difference between open and closed mitosis concerns continuity of the nuclear envelope. A complete breakdown and reassembly of the nuclear envelope at every cell division requires more in the way of membrane fragmentation and reassembly processes (Heath 1980) than closed mitosis, which is mechanistically simpler, entailing enlargement and median constriction of the nuclear envelope (reviewed in Ungricht and Kutay 2017). Eukaryotes arose from simple prokaryotic ancestors having prokaryotic chromosome and cell division processes. The presence of closed mitosis with intranuclear spindles in a syncytial LECA eases the prokaryote to eukaryote transition, because it decouples the processes of chromosome division, chromosome partitioning, and cell division (cytokinesis), allowing them to evolve in sequence as independent traits rather than simultaneously, while also buffering for the existence of defective chromosome combinations through intracellular complementation from nuclei with viable chromosome combinations via mRNA in the cytosol. Therefore, it is unlikely that open mitosis was present in LECA and escaped identification. In summary, the present findings indicate that LECA had, in addition to a nucleus (Mans et al. 2004; Bapteste et al. 2005; Neumann et al. 2010), an endoplasmic reticulum (Kontou et al. 2022), linear chromosomes with centromeres (Ishikawa and Naito 1999; van Hooff et al. 2017), flagellae (Carvalho-Santos et al. 2011; Lindemann 2022), microtubule organizing centers (Yubuki and Leander 2013), nucleoli (Gardner et al. 2010; Hoeppner and Poole 2012), meiosis and sex (Villeneuve and Hillers 2001; Loidl 2016), facultatively anaerobic mitochondria (Müller et al. 2012; Mills et al. 2022), and a syncytial habit (Skejo et al. 2021) that lacked phagocytosis (Bremer et al., 2022), closed orthomitosis with intranuclear spindles. In terms of overall physiology and lifestyle, LECA is beginning to look like a filamentous fungus (Martin et al. 2003) able to survive in anaerobic environments.

Materials and Methods

Phylogenetic Trees

We clustered protein sequences from 150 eukaryotic genomes by firstly performing an all-versus-all BLAST (Altschul et al. 1990) and selecting the best reciprocal BLAST hits with an expectation value (*e*-value) $\leq 10^{-10}$. Those hits were then globally aligned using the Needleman–Wunsch algorithm implemented in the EMBOSS needle program (Rice et al. 2000). Protein pairs with a global identity <25% were discarded before clustering with the MCL algorithm (Enright et al. 2002), version 12-068 using default parameters. A total of 1,789 protein clusters that possessed at least one species of each eukaryotic supergroup were found and selected for further analyses. Alignments of those protein clusters were generated using MAFFT (Katoh et al. 2002), using the iterative refinement method that assimilates local pairwise alignment information (L-INS-i). The alignments were not trimmed and maximum-likelihood trees were reconstructed with IQ-Tree (Nguyen et al. 2015), using the best-fit model and the following parameters: "-bb 1000" and "-alrt 1000." We differentiated between trees without paralogs (single-copy trees) and trees with paralogs (multicopy trees) for further analyses. The trees were rooted with MAD (Tria et al. 2017). All 1,789 analyzed trees showed no ambiguous root inferences.

Trait Annotation, Coding, and Definition

All four analyzed traits were coded as binary traits (presence "1" or absence "0"). While the sexual reproduction may be either present or absence, the mitotic traits have to be seen a little different. An absence for the closed nuclear division means that the nuclear envelope is open or semi-open during mitosis. For the intranuclear spindles, an absence corresponds to extranuclear spindles and the absence of orthomitosis (axial symmetry) stands for pleuromitosis (bilateral symmetry). The annotation of traits is based on literature (supplementary table S1, Supplementary Material online). Not every species in our data set is annotated for every trait in literature. We therefore applied a majority rule for each group in question. If there is data on the exact ancestral state of a trait, we annotated it to be present in the whole group. In cases where only one representative of a group is annotated in literature, this annotation was

Downloaded from https://academic.oup.com/gbe/article/15/3/evad016/7031494 by Universitaetsbibliothek Duesseldorf user on 21 February 2024

suspected to be present in the whole group. Groups with different annotations for different members were annotated by majority rule. No cases with a 50:50 distribution were found in our data set. Two species in our data set are annotated with incompatible mitotic combinations (closed orthomitosis with extranuclear spindle): *Chlamydomonas reinhardtii* and *Volvox carteri*, both members of the taxon Chlorophyceae. Although the combination of traits itself is incompatible, the majority rule resulted in this combination for the group.

Ancestral State Reconstruction

The reconstruction of ancestral states was performed using PastML version 1.9.20 (Ishikawa et al. 2019). The chosen parameters were a maximum-likelihood approach based on marginal posterior probabilities approximation and the F81 model of character evolution (Felsenstein 1981). We annotated the tips of the trees based on a trait matrix for the 150 eukaryotic species (supplementary table S1, Supplementary Material online), with the inclusion of missing data (unknown tip state). Trees with the same state of a trait at each tip of the tree were discarded from the analysis for this specific trait. Analyses of phylogenetic trees and trait origins were performed using the python toolkit Environment for Tree Exploration ETE v3 (Huerta-Cepas et al. 2016).

18S RNA Reference Tree

The reconstruction of a reference tree was performed with 18S RNA sequences for all 150 eukaryotes in our data set. Sequences were primarily searched for in the SILVA rRNA database (release 138.1 from November 2020; Quast et al. 2013). As not all eukaryotic species from our data set were found within this database, sequences were secondarily searched for within the PR^2 sequence database (version 4.12.0 from August 2019; Guillou et al. 2013). For a total of eight species, we were not able to find 18S RNA sequences in both databases and therefore used alternatives from the same genus. We generated an alignment using MAFFT (Katoh et al. 2002), using the iterative refinement method that assimilates local pairwise alignment information (L-INS-i), reconstructed a maximum-likelihood tree with IQ-tree (Nguyen et al. 2015) and rooted the resulting tree on the branch leading to Excavates. The sole purpose of the reconstruction and rooting of this tree was to display our data.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

Acknowledgments

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant 666053 and 101018894) and the Volkswagen Foundation (grant 93046) 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.20591172.

Literature Cited

Aldrich HC. 1969. The ultrastructure of mitosis in myxamoebae and plasmodia of *Physarum Flavicomum*. Am J Bot. 56:190–299.
Altschul SF, et al. 1990. Basic local alignment search tool. J Mol Biol. 215:403–410.
Bapteste E, Charlebois RL, MacLeod D, Brochier C. 2005. The two tempos of nuclear pore complex evolution: highly adapting proteins in an ancient frozen structure. Genome Biol. 6(10):R85.
Bělař K. 1926. Zur cytologie von *Aggregata eberthi*. Arch Protistenk. 53:312–325.
Boettcher B, Barral Y. 2013. The cell biology of open and closed mitosis. Nucleus 4(3):160–165.
Booth A, Doolittle WF. 2015. Eukaryogenesis, how special really? Proc Natl Acad Sci U S A. 112:10278–11028.
Bremer N, Tria FDK, Skejo J, Garg SG, Martin WF. 2022. Ancestral state reconstructions trace mitochondria but not phagocytosis to the last eukaryotic common ancestor. Genome Biol Evol. 14(6):evac079.
Burki F, Roger AJ, Brown MW, Simpson AGB. 2020. The new tree of eukaryotes. Trends Ecol Evol. 35:43–55.
Carvalho-Santos Z, Azimzadeh J, Pereira-Leal JB, Bettencourt-Dias M. 2011. Tracing the origins of centrioles, cilia, and flagella. J Cell Biol. 194(2):165–175.
Cavalier-Smith T. 2009. Kingdoms Protozoa and Chromista and the eozoan root of the eukaryotic tree. Biol Lett. 6:342–345.
Cavalier-Smith T. 2010. Origin of the cell nucleus, mitosis and sex: roles of intracellular coevolution. Biol Direct. 5:7.
Cerón-Romero MA, Fonseca MM, de Oliveira Martins L, Posada D, Katz LA. 2022. Phylogenomic analyses of 2,786 genes in 158 lineages support a root of the eukaryotic tree of life between opisthokonts and all other lineages. Genome Biol Evol. 14(8):evac119.
Derelle R, Lang BF. 2012. Rooting the eukaryotic tree with mitochondrial and bacterial proteins. Mol Biol Evol. 29(4):1277–1289.
De Souza CPC, Osmani SA. 2007. Mitosis, not just open or closed. Eukaryotic Cell 6(9):1521–1527.
Dey G, Baum B. 2021. Nuclear envelope remodelling during mitosis. Curr Opin Cell Biol. 70:67–74.
Egel R, Penny D. 2007. On the origin in eukaryotic evolution: coevolution of meiosis and mitosis from feeble beginnings. In: Egel R, Lankenau DH, editors. Recombination and meiosis. Genome dynamics and stability, Vol. 3. Berlin, Heidelberg: Springer. p. 249–288.
Eme L, Sharpe SC, Brown MW, Roger AJ. 2014. On the age of eukaryotes: evaluating evidence from fossils and molecular clocks. Cold Spring Harb Perspect Biol. 6(8):a016139.
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.

Downloaded from https://academic.oup.com/gbe/article/15/3/evad016/7031494 by Universitaetsbibliothek Duesseldorf user on 21 February 2024

Felsenstein J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J Mol Evol. 17(6):368–376.
Forterre P, Phillippe H. 1999. Where is the root of the universal tree of life? BioEssays 21(10):871–879.
Gabaldón T. 2021. Origin and early evolution of the eukaryotic cell. Annu Rev Microbiol. 75(1):631–647.
Gardner PP, Bateman A, Poole AM. 2010. Snopatrol: how many snoRNA genes are there? J Biol. 9(1):4.
Garg SG, et al. 2021. Anomalous phylogenetic behavior of ribosomal proteins in metagenome-assembled Asgard Archaea. Genome Biol Evol 13(1):439.
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.
Gold DA, Runnegar B, Gehling JG, Jacobs DK. 2015. Ancestral state reconstruction of ontogeny supports a bilaterian affinity for Dickinsonia. Evol Dev. 17:315–324.
Green JC, Hori T. 1988. The fine structure of mitosis in *Pavlova* (Prymnesiophyceae). Can J Bot. 66(8):1497–1509.
Gromov DB. 1985. Ultrastructure of mitosis in *Amoeba proteus*. Protoplasma 126:130–139.
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(Database Issue): D597–D604.
He D, et al. 2014. An alternative root for the eukaryotic tree of life. Curr Biol. 24(4):465–470.
Heath IB. 1980. Variant mitoses in lower eukaryotes: indicators of the evolution of mitosis? Int Rev Cytol. 64:1–80.
Hertwig R. 1903. Ueber Korrelation von Zell- und Kerngrösse und ihre Bedeutung für die geschlechtliche Differenzierung und die Teilung der Zelle. Biologisches Centralblatt 23:4–62.
Heywood P. 1988. Ultrastructure of *Chilomonas paramecium* and the phylogeny of the cryptoprotists. Biosystems 21(3–4):293–298.
Hoeppner MP, Poole AM. 2012. Comarative genomics of eukaryotic small nucleolar RNAs reveals deep evolutionary ancestry amidst ongoing intragenomic mobility. BMC Evol Biol. 12:183.
Hori T, Green JC. 1985. The ultrastructure of mitosis in *Isochrysis galbana* parke (Prymnesiophyceae). Protoplasma 125:140–151.
Huerta-Cepas J, Serra F, Bork P. 2016. ETE 3: reconstruction, analysis, and visualization of phylogenomic data. Mol Biol Evol. 33(6): 1635–1638.
Ishikawa SA, et al. 2019. A fast likelihood method to reconstruct and visualize ancestral scenarios. Mol Biol Evol. 36(9):2069–2085.
Ishikawa F, Naito T. 1999. Why do we have linear chromosomes? A matter of Adam and Eve. Mutat Res. 434(2):99–107.
Jermann TM, Opitz JG, Stackhouse J, Benner SA. 1995. Reconstructing the evolutionary history of the artiodactyl ribonuclease superfamily. Nature 374:57–59.
Jorgensen P, et al. 2007. The size of the nucleus increases as yeast cells grow. Mol Biol Cell. 18(9):3523–3532.
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.
Katz LA. 2012. Origin and diversification of eukaryotes. Annu Rev Microbiol. 66:411–427.
Katz LA, Grant JR, Parfrey LW, Burleigh JG. 2012. Turning the crown upside down: gene tree parsimony roots the eukaryotic tree of life. Syst Biol. 61(4):653–660.
Keeling PJ, Burki F. 2019. Progress towards the tree of eukaryotes. Curr Biol. 29(16):R808–R817.
Kim E, Simpson AGB, Graham L. 2006. Evolutionary relationships of Apusomonads inferred from taxon-rich analyses of 6 nuclear encoded genes. Mol Biol Evol. 23(12):2455–2466.
Klim J, Gładki A, Kucharczyk R, Zielenkiewicz U, Kaczanowski S. 2018. Ancestral state reconstruction of the apoptosis machinery in the common ancestor of eukaryotes. G3 (Bethesda) 8(6):2121–2134.
Kohn JR, Graham SW, Morton B, Doyle JJ, Barrett SCH. 1996. Reconstruction of the evolution of reproductive characters in Pontederiaceae using phylogenetic evidence from chloroplast DNA restriction-site variation. Evolution 50:1454–1469.
Kontou A, Herman EK, Field MC, Dacks JB, Koumandou VL. 2022. Evolution of factors shaping the endoplasmic reticulum. Traffic 1:1–12.
Koumandou VL, et al. 2013. Molecular paleontology and complexity in the last eukaryotic common ancestor. Crit Rev Biochem Mol Biol. 48:373–396.
Lamarca AP, Mello B, Schrago CG. 2022. The performance of outgroup-free rooting under evolutionary radiations. Mol Phylogenet Evol. 169: 107434.
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 467:929–934.
Leedale GF. 1970. Phylogenetic aspects of nuclear cytology in the algae. Ann N Y Acad Sci. 175(2):429–453.
Lindemann CB. 2022. The flagellar germ-line hypothesis: how flagellate and ciliate gametes significantly shaped the evolution of organismal complexity. BioEssays 44(3):2100143.
Loidl J. 2016. Conservation and variability of meiosis across the eukaryotes. Annu Rev Genet. 50:293–316.
López-García P, Moreira D. 2015. Open questions on the origin of eukaryotes. Trends Ecol Evol. 30(11):697–708.
Mans BJ, Anantharaman V, Aravind L, Koonin EV. 2004. Comparative genomics, evolution and origin of the nuclear envelope and nuclear pore complex. Cell Cycle. 3(12):1612–1637.
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.
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. 76:444–495.
Neumann N, Lundin D, Poole AM. 2010. Comparative genomic evidence for a complete nuclear pore complex in the last eukaryotic common ancestor. PLoS One. 5(10):e13241.
Neumann FR, Nurse P. 2007. Nuclear size control in fission yeast. J Cell Biol. 179(4):593–600.
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.
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.
Philippe H, Germot A, Moreira D. 2000. The new phylogeny of eukaryotes. Curr Opin Genet Dev. 10(6):596–601.
Pickett-Heaps JD. 1969. The evolution of the mitotic apparatus: an attempt at comparative ultrastructural cytology in dividing plant cells. Cytobios 1(3):257–280.
Pickett-Hepas JD. 1974. The evolution of mitosis and the eukaryotic condition. Biosystems 6(1):37–48.
Poole A, Jeffares D, Penny D. 1999. Early evolution: prokaryotes, the new kids on the block. BioEssays 21(10):880–889.
Porter M. 2020. Insights into eukaryogenesis from fossil record. Interface Focus 10:20190105.
Quast C, et al. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 41(Database Issue):D590–D596.

Downloaded from https://academic.oup.com/gbe/article/15/3/evad016/7031494 by Universitaetsbibliothek Duesseldorf user on 21 February 2024

Raikov IB. 1994. The diversity of forms of mitosis in protozoa: a comparative review. Eur J Protistol. 30(3):253–269.
Raikov IB, Mignot JP. 1991. Fine-structural study of mitosis in the testacean *Arcella vulgaris* Ehrbg. Eur J Protistol. 26:340–3349.
Ribeiro KC, Monteiro-Leal LH, Benchimol M. 2005. Contributions of the axostyle and flagella to closed mitosis in the protists *Tritrichomonas foetus* and *Trichomonas vaginalis*. J Eukaryot Microbiol. 47:481–492.
Rice P, Longden L, Bleasby A. 2000. EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet 16(6):276–277.
Richards TA, Cavalier-Smith T. 2005. Myosin domain evolution and the primary divergence of eukaryotes. Nature 436:1113–1118.
Rodríguez-Ezpeleta N, et al. 2007. Toward resolving the eukaryotic tree: the phylogenetic positions of Jacobids and Cercozoans. Curr Biol. 17(16):1420–1425.
Roger AJ, Simpson AGB. 2009. Evolution: revisiting the root of the eukaryotic tree. Curr Biol. 19(4):R165–R167.
Roger AJ, Susko E, Leger MM. 2021. Evolution: reconstructing the timeline of eukaryogenesis. Curr Biol. 31(4):R193–R196.
Rogozin IB, Basu MK, Csürös M, Koonin EV. 2009. Analysis of rare genomic changes does not support the Unikont-Bikont phylogeny and suggests cyanobacterial symbiosis as the point of primary radiation of eukaryotes. Genome Biol Evol. 1:99–113.
Sazer S, Lynch M, Needleman D. 2014. Deciphering the evolutionary history of open and closed mitosis. Curr Biol. 24(22): R1099–R1103.
Skejo J, et al. 2021. Evidence for a syncytial origin of eukaryotes from ancestral state reconstruction. Genome Biol Evol. 13(7):evab096.
Solnica-Krezel L, Burland TG, Dove WF. 1991. Variable pathways for developmental changes of mitosis and cytokinesis in Physarum polycephalum. J Cell Biol 113(3):591–604.
Speijer D, Lukeš J, Eliáš M. 2015. Sex is a ubiquitous, ancient, and inherent attribute of eukaryotic life. Proc Natl Acad Sci U S A. 112: 8827–8834.
Stechmann A, Cavalier-Smith T. 2003. The root of the eukaryote tree pinpointed. Curr Biol. 13(17):R665–R666.
Tanaka K. 1973. Intranuclear microtubule organizing center in early prophase nuclei of the plasmodium of the slime mold, Physarum Polycephalum. J Cell Biol 57(1):220–224.
Tria FDK, et al. 2021. Gene duplications trace mitochondria to the onset of eukaryote complexity. Genome Biol Evol 13(5):429.
Tria FDK, Landan G, Dagan T. 2017. Phylogenetic rooting using minimal ancestor deviation. Nat Ecol Evol. 1:193.
Tromer EC, van Hooff JJE, Kops GJPL, Snel B. 2019. Mosaic origin of the eukaryotic kinetochore. Proc Natl Acad Sci U S A. 116(26): 12873–12882.
Ungricht R, Kutay U. 2017. Mechanisms and functions of nuclear envelope remodelling. Nat Rev Mol Cell Biol. 18:229–245.
van Hooff JJ, Tromer E, van Wijk LM, Snel B, Kops GJ. 2017. Evolutionary dynamics of the kinetochore network in eukaryotes as revealed by comparative genomics. EMBO Rep. 18(9): 1559–1571.
Villeneuve AM, Hillers KJ. 2001. Whence meiosis? Cell 106(6): 647–650.
Wade T, et al. 2020. Assessing the accuracy of phylogenetic rooting methods on prokaryotic gene families. PLoS One 15:1–22.
Williams TA. 2014. Evolution: rooting the eukaryotic tree of life. Curr Biol. 24(4):R151–R152.
Yubuki N, Leander BS. 2013. Evolution of microtubule organizing centers across the tree of eukaryotes. Plant J. 75:230–244.

Associate editor: Luis Delaye

Downloaded from https://academic.oup.com/gbe/article/15/3/evad016/7031494 by Universitaetsbibliothek Duesseldorf user on 21 February 2024

iii. Surprising effects of differential loss in genome evolution: the last-one-out

Nico Bremer, William F. Martin, Mike Steel.

Institut für Molekulare Evolution, Heinrich-Heine-Universität Düsseldorf, Universitätsstraße 1, 40225 Düsseldorf, Deutschland.

Dieser Artikel wurde am 22.05.2025 in *FEMS Microbiology Letters* Ausgabe 372 veröffentlicht.

Beitrag von Nico Bremer (Erstautor und Korrespondenz):
Ich war an der Entwicklung des Studienkonzepts beteiligt. Der bioinformatische Algorithmus wurde von mir auf Grundlage der mathematischen Theoreme programmiert. Die bioinformatische Analyse, sowie die Visualisierung der Ergebnisse wurde von mir durchgeführt. Teile des initialen Manuskriptes wurden von mir geschrieben. Die Überarbeitung des Manuskripts wurde ebenfalls von mir durchgeführt.

OXFORD FEMS

FEMS Microbiology Letters, 2025, **372**, fnaf051

DOI: 10.1093/femsle/fnaf051

Advance access publication date: 22 May 2025

Research Article – Taxonomy, Systematics and Evolutionary Microbiology

Surprising effects of differential loss in genome evolution: the last-one-out

Nico Bremer[1,*], William F. Martin[1], Mike Steel[2]

[1]Faculty of Mathematics and Natural Sciences, Institute of Molecular Evolution, Heinrich Heine University Düsseldorf, 40225 Duesseldorf, Germany
[2]Biomathematics Research Centre, University of Canterbury, 8140 Christchurch, New Zealand
*Corresponding author. Faculty of Mathematics and Natural Sciences, Institute of Molecular Evolution, Heinrich Heine University of Duesseldorf, 40225 Duesseldorf, Germany. E-mail: nico.bremer@hhu.de
Editor: [Aharon Oren]

Abstract

Gene loss is an important process in genome evolution, though its power is often underestimated. If a gene is present at the root of a phylogenetic tree and can be lost in one lineage across the tree, it can potentially be lost in all, leading to gene extinction. Just before gene extinction, there will be one lineage that still retains the gene, generating a "last-one-out" distribution. Such an isolated gene presence will emulate the result of recent lateral gene acquisition, even though its distribution was generated by loss. How probable is it to observe "last-one-out" distributions in real data? Here, we mathematically derive this probability and find that it is surprisingly high, depending upon the tree and the gene loss rate. Examples from real data show that loss can readily account for observed frequencies of last-one-out gene distributions that might otherwise be attributed to lateral gene transfer.

Keywords: gene loss; lateral gene transfer; birth–death process

Introduction

Gene loss is an important and ubiquitous mechanism of genome evolution. In prokaryotes, gene loss acting on the whole genome is traditionally called reductive evolution (Andersson and Kurland 1998, van Ham et al. 2003, Oshima et al. 2004, Hosokawa et al. 2006) and can result in miniscule genome sizes in parasites and endosymbiotic bacteria, the current record being Macrosteles quadrilineatus (Moran and Bennett 2014) an endosymbiotic bacterium of leafhoppers that harbors only 137 protein-coding genes. Reductive evolution is also observed in symbiotic archaea (Waters et al. 2003) and in eukaryotes, especially among intracellular parasites (Tovar et al. 2003, Nicholson et al. 2022). Genome reduction through gene loss is also the central underlying theme of genome evolution in mitochondria and plastids, the endosymbiotic organelles of eukaryotic cells (Moore and Archibald 2009), which can sometimes lose their genomes altogether (Müller et al. 2012), because many genes lost from organelle genomes have been transferred to the nucleus (Martin et al. 1998, Timmis et al. 2004). In eukaryotes, gene loss is also very common and particularly well studied following whole-genome duplications (Blanc and Wolfe 2004, Kellis et al. 2004, Brunet et al. 2006, Scannel et al. 2006), where duplicate gene copies are rapidly lost by mutation, restoring diploid genetics in chromosome polyploids (Blanc and Wolfe 2004). Additionally, gene loss is often seen as a driving factor in genome evolution (Olson 1999, Albalat and Cañestro 2016, Guijarro-Clarke et al. 2020). In general, if a gene belonging to a clade can be lost once in one lineage during evolution, it can be lost again in other lineages as well.

In comparative genome studies, gene loss is easy to detect if losses are rare, as shown in Fig. 1. If most genomes in a sample contain a given gene of interest, but one or a few do not, there can be little doubt that gene loss has occurred in the genomes lacking the gene. But the more common loss is, the more difficult it becomes to distinguish from lateral gene transfer (LGT). If a given gene is present in about half of the genomes in a sample, the decision between loss and LGT becomes a matter of weighing the relative probabilities of LGT and gene loss, entailing an *a priori* assumption that LGT is roughly as common as loss. In eukaryotes, gene loss is much more common than LGT from prokaryotes (Ku et al. 2015, Ku and Martin 2016). But there have been a number of highly publicized claims for widespread LGT to eukaryotes, though, for example "hundreds" of LGTs in the human genome (Consortium 2001) or fully "17%" of the tardigrade (a primitive animal) genome being the result of recent LGTs (Boothby et al. 2015). Both the human genome LGT claims and the tardigrade LGT claims were reinspected and turned out to be data contaminations and data interpretation problems, not LGT (Salzberg et al. 2001, Stanhope et al. 2001, Koutsovoulos et al. 2016, Salzberg 2017). Most reports of LGT in eukaryotes are not critically reinspected, but they are highly cited (Martin 2017, Keeling 2024). Nonetheless, there are cases where contaminations can effectively be ruled out—for example when a gene in question is observed in the genome sequence of several individuals from a given species (Koutsovoulos et al. 2016). Distributions of the type seen in Fig. 1(c) are observed in real data for eukaryote genomes, where LGT might appear to be the most likely cause, because the possibility of many independent losses as the cause of the pattern would seem, at face value, extremely unlikely.

But is gene loss leading to last-one-out topologies really unlikely? Or do we just assume it is unlikely, thereby opting to sug-

Received 4 March 2025; **revised** 29 April 2025; **accepted** 20 May 2025
© The Author(s) 2025. Published by Oxford University Press on behalf of FEMS. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs licence (https://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Downloaded from https://academic.oup.com/femsle/article/doi/10.1093/femsle/fnaf051/8140845 by Universitaets- und Landesbibliothek Duesseldorf user on 04 August 2025

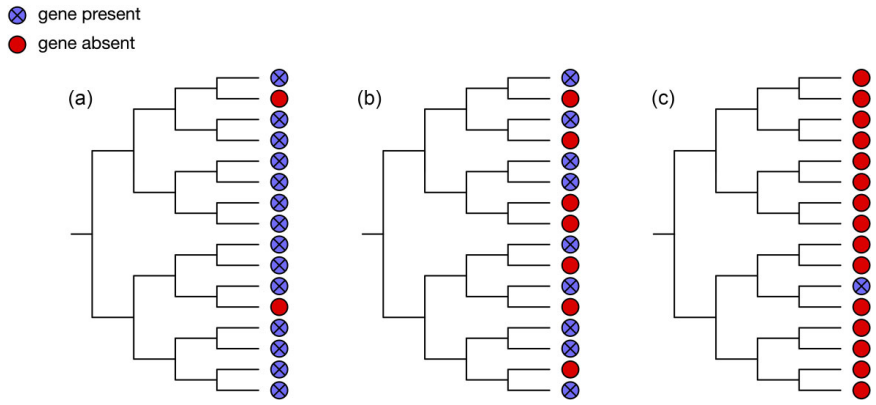

Figure 1. Hypothetical phylogenetic species trees showing the presence and absence of genes across all species in the trees. A circle with a cross indicates that the gene is present in this species, an empty circle indicates that a gene is absent. (a) A distribution where gene loss most likely appeared on the branches to two species. (b) A case where the distribution of the genes that are present and absent is almost equal across the species tree. The decision between LGT and gene loss is highly dependent on the weighing of their relative probabilities. (c) Illustrates a case where the gene is only present in one species. An easy (but not necessarily true) explanation for this would be LGT. This gene distribution across the tree can also be the result of a minimum of four gene losses if the gene was already present at the root node.

gest that LGT was at work without even testing the possibility that the distribution is actually the result of many independent losses. Are there even tools available to test such a case? This answer, until now, has been no. Many analytical tools to study prokaryotic genomes are currently in use that employ different and usually predetermined gain/loss ratios that are designed to differentiate between loss and LGT (Goodman et al. 1979, Page 1994, Bansal et al. 2012, Szöllősi et al. 2013). In many cases, the overall average ratio of gene loss to LGT ends up being close to 1 in such applications, for obvious reasons. If loss predominates, then genomes steadily decrease in size across the reference tree (that is, ancestral genomes inflate), and if LGT predominates, genomes steadily increase in size across the reference tree (that is, ancestral genomes become too small) (Dagan and Martin 2007). Some tools for estimating loss versus LGT in current use can entail differences in loss versus transfer probabilities for individual genes that differ by 20 orders of magnitude (Bremer et al. 2022).

If gene loss is the predominant mode of genome evolution for a given gene in a given group, it will become lost in many lineages, ultimately in all. Just before the gene goes extinct in the group, however, there will exist a state in which the gene is present in only a few genomes, and finally, over time, only in one genome of the group. If this gene is in a eukaryote, but has homologs in prokaryotes, gene loss in eukaryotes will produce a pattern that looks exactly like LGT: The gene is present in prokaryotes and one (or a few) eukaryotes. Under a loss-only mode of evolution, the last-one-out looks like an LGT, but the pattern was generated solely through gene loss. Here, we address the question of how likely it is to observe a last-one-out gene distribution under loss-only models.

Results

Mathematical modeling and algorithms

We now describe mathematical and computational methods to investigate the probability of last-one-out scenarios in both synthetic and real trees. We assume that each gene in a phylogeny

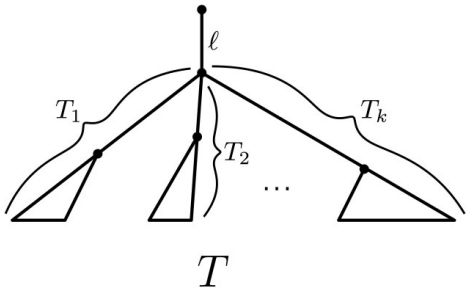

Figure 2. Hypothetical phylogenetic tree T with the subtrees $T_1, T_2 \ldots T_k$ and branch length ℓ.

can be lost along each lineage of a tree according to a continuous-time Markov process with loss rate μ, and which operates independently across genes and lineages.

Recursion for a given tree

Let T be a rooted tree with a stem edge of length ℓ, and let $T_1, T_2 \ldots T_k$ denote the subtrees of T incident with this stem edge, as shown in Fig. 2. Although the lengths of edges may correspond to time, and so be ultrametric, the algorithm described in this first section does not assume that edge lengths are ultrametric. Let π_T^+ denote the probability that a gene g that is present at start of the stem edge of T is present in *exactly one* leaf of T, and let π_i^+ denote $\pi_{T_i}^+$ (the corresponding probabilities for the subtrees $T_1 \ldots, T_k$). To calculate π_T^+ recursively, we also need to calculate the probability π_T that g is not present at any of the leaves of T, and we let π_i denote π_{T_i}.

Note that if T consists of just a single stem edge of length ℓ (the base case in the recursion), then $\pi_T = 1 - e^{-\mu\ell}$ and $\pi_T^+ = e^{-\mu\ell}$. Thus we may suppose that $k \geq 2$. The following result (proved in the Appendix) provides a polynomial-time way to compute these quantities recursively via dynamic programming (progressing from the

Downloaded from https://academic.oup.com/femsle/article/doi/10.1093/femsle/fnaf051/8140845 by Universitaets- und Landesbibliothek Duesseldorf user on 04 August 2025

leaves to the root). Note that both Parts (i) and (ii) are required for computing π_T^+.

Proposition 1. For the tree shown in Fig. 2, the following recursions hold:

$$\pi_T = \left(1 - e^{-\mu\ell}\right) + e^{-\mu\ell}\pi_1\pi_2\ldots\pi_k, \tag{1}$$

$$\pi_T^+ = e^{-\mu\ell}\pi_1\pi_2\ldots\pi_k\left(\frac{\pi_1^+}{\pi_1} + \frac{\pi_2^+}{\pi_2} + \ldots + \frac{\pi_k^+}{\pi_k}\right). \tag{2}$$

For binary trees, Equation (2) simplifies to:

$$\pi_T^+ = e^{-\mu\ell}\left(\pi_1\pi_2^+ + \pi_1^+\pi_2\right). \tag{3}$$

If there are $G \geq 1$ genes present at the top of the stem edge of T, and losses occur independently among the genes (each with rate μ), then the number of genes that appear in just one leaf of T has a binomial distribution with parameters $(G,\ \pi_T^+)$.

To illustrate Proposition 1 with a simple example, consider the tree in Fig. 2, where each of the subtrees $T_1\ldots,\ T_k$ is a single leaf at the same distance from the root, and $\ell = 0$ (the "star tree"). Under the gene-loss model, a gene that is present at the root of the tree will be present at exactly one leaf of this tree precisely if there are *exactly* $k - 1$ loss events. This might seem very unlikely for large values of k. However, the probability of this event can be as large as $e^{-1} = 0.367$ even as k becomes large, provided that μ is chosen appropriately (and dependent on k); details are provided in the "Analysis of the star tree" section of the Appendix. Nevertheless, if we consider the posterior value of this probability by taking a uniform prior on $1 - e^{-\mu}$ (setting the height of the tree to 1), then this posterior probability tends to 0 as the number of leaves of the tree (k) grows. The proof of these claims and the analysis of this star tree when we allow $\ell > 0$ are provided in the Appendix. Of course, the star tree is a highly nonbinary tree, which raises the question of whether π_T^+ can be close to e^{-1} when T is binary and the number of leaves is large. This is indeed possible: we can simply resolve the polytomy at the root by using very short interior edges to obtain a binary tree for which π_T^+ will be close to the corresponding value for a star tree and hence can be close to e^{-1} for a suitably chosen value of μ. However, for trees generated by simple phylodynamic models, this is no longer the case, as we demonstrate in the next section. The analyses in this manuscript mainly focus on binary trees.

Random trees

Suppose now that T is generated by a standard birth–death model (Kendall 1948, Lambert and Stadler 2013) with speciation rate λ and extinction rate ν, starting from a single lineage at time t in the past. The tree T is now a random variable, denoted T_t, and the number of species at the present (denoted N_t) is also a random variable and has a (modified) geometric distribution with expected value $\mathbb{E}[N_t] = e^{(\lambda-\nu)t}$. We will suppose that $\lambda > \nu$ since otherwise the tree T_t is guaranteed to die out as t grows. Let π_t^+ be the probability that a gene g that is present at start of the stem edge of T_t is present in *exactly one* leaf of T_t. The following result precisely describes the maximum value that π_t^+ can take as μ (the rate of gene loss) varies over all possible positive values. The short proof is provided in the Appendix.

Proposition 2.

$$\max_{\mu} \pi_t^+ = \frac{1}{(1+\lambda t)^2} = \frac{1}{\left(1 + \frac{\ln \mathbb{E}[N_t]}{1 - \frac{\nu}{\lambda}}\right)^2}. \tag{4}$$

Notice that although $max_\mu \pi_t^+ \to 0$ for Yule trees as they grow in their expected size, the convergence is quite slow as a function of the expected number of leaves of the tree, due to the presence of the logarithmic function on the right of Equation (4). Also, if there are $G \geq 1$ genes present at time 0, then the expected number of genes that will be present in just leaf of T_t is $G \cdot \pi_t^+$. However, in contrast to Proposition 1(iii), the number of genes present in just one leaf of T_t is no longer binomially distributed, since this number is now a compound random variable because it is dependent on the random variable T_t.

To illustrate Proposition 2, consider (pure–birth) Yule trees (i.e. $\nu = 0$) with an expected number of 150 leaves. Then $max_\mu \pi_t^+ \approx 0.028$, and so for 10 000 independent genes and this optimal rate of gene loss, the expected number of genes that would be last-one-out (i.e. present in just one leaf of these Yule trees) would be around 280. This provides some insight into the results described in the next section.

Application to real genome data

To test this algorithm on real genome data we chose the example of genes in eukaryotic genomes that have homologs in prokaryotes but that are present in only one or a few eukaryotic lineages. Such patterns are taken as evidence for the workings of differential loss, under the assumption that loss will generate such patterns (Ku et al. 2015), or as evidence for the workings of LGT (Cote-L'Heureux et al. 2022) under the assumption that LGT rather than loss generates such patterns. The calculation of the probability of a gene being present at the root node and remaining in exactly one leaf of a eukaryotic tree requires a rooted species tree and a gene loss rate μ. Reconstructing a eukaryotic species tree is challenging, and there is currently no consensus on the position of the root (Keeling and Burki 2019, Burki et al. 2020). Although the loss rates can be adjusted and averaged across a range of values, the backbone trees with all their nodes, branches and branch lengths are not that easily adjustable.

We started by investigating a set of ten eukaryotic gene trees with 150 leaves each. These gene trees need not be representative of the true phylogeny of eukaryotes, nor need they show a pattern of gene distribution that could be indicated as LGT. They are just used to test the algorithm, whereby the different trees were selected merely to show that different phylogenies can have an influence on the calculated probability of a gene being present at the root node and remaining only in one leaf of a eukaryotic tree. Furthermore, the different gene trees with 150 leaves provide an opportunity to estimate the overall probability of observing a last-one-out pattern if we consider thousands of eukaryotic genes with prokaryotic homologs (Fig. 3). In Fig. 3, we assume that 10 000 genes were present in the last common ancestor of 150 eukaryotes. For those trees, the mean probabilities across a range of different loss rates μ, where $1 - e^{-\mu}$ ranges from 0 to 1, result in 232 (lowest mean) to 790 (highest mean) last-one-out cases that would look like LGT but actually are the result of differential loss in a loss-only mode of evolution for a 10 000 gene ancestral genome. Looking at the median, we would find 11 (lowest median) to 755 (highest median) cases, depending on the tree itself. Since loss rates are not constant over time, we cannot assume that these percentages resemble the "real" amount of those cases due to differential loss. This first look into the data with our new tool does show, however, that last-one-out cases are by no means so rare that they can be excluded *a priori*. If the loss rate is ideal, meaning that the maximum probability of last-one-out cases for the given tree is achieved, we would see between 532 (lowest maximum) and 2502 (highest maximum) out of the 10 000 genes resulting in a last-one-out scenario, which is a substantial frequency. That is, in a study of 10 000 gene families present in the eukaryotic common

Downloaded from https://academic.oup.com/femsle/article/doi/10.1093/femsle/fnaf051/8140845 by Universitaets- und Landesbibliothek Duesseldorf user on 04 August 2025

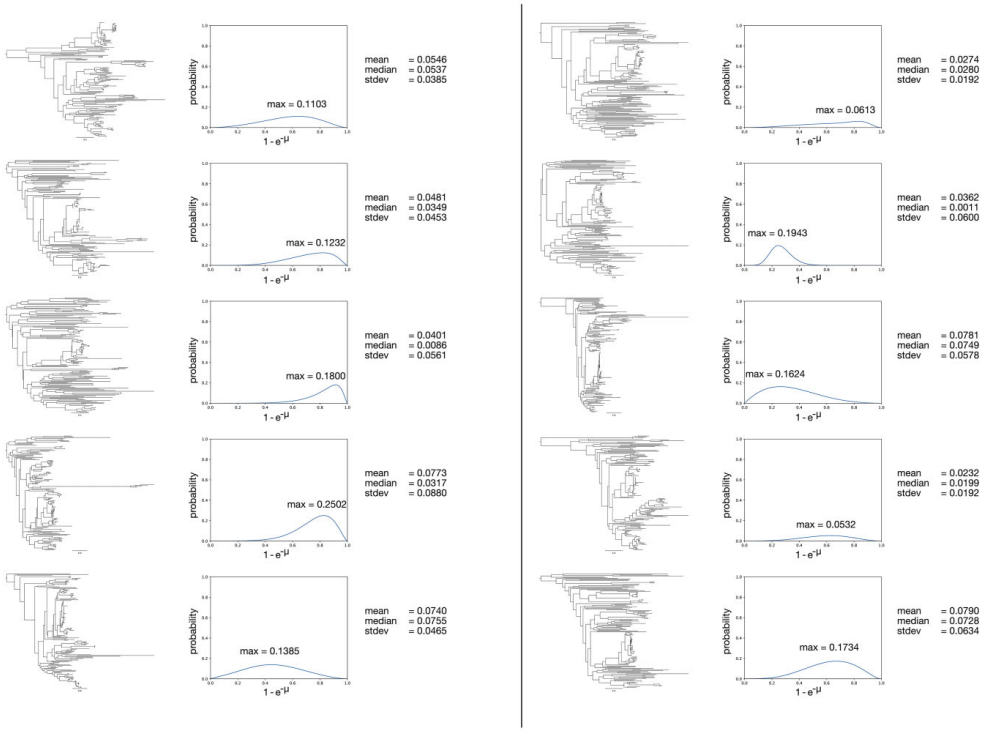

Figure 3. Ten eukaryotic gene tree phylogenies with 150 leaves each and the corresponding probabilities for a last-one-out scenario against $1 - e^{-\mu}$ (μ = gene loss rate). The trees show various possibilities of species trees without assuming that those trees represent a real eukaryotic backbone tree. They show that the phylogeny itself has an influence on the probability of a last-one-out scenario, but that the overall probability is comparably high.

ancestor, one would expect to observe dozens, hundreds, or even thousands of last-one-out patterns in trees sampling 150 genomes obtained solely as the result of differential loss. Put another way, we would expect to observe last-one- out distributions at a frequency that is not far off from the number of genomes in the tree. These cases would appear, in a gene phylogeny, as a single eukaryote (or group thereof) branching within prokaryotic homologs. Such cases are observed with real data.

Three test cases

The surprisingly high probability to observe a gene that is present in the root node and only in one species or clade and lost in all other leaves of a tree offers a new approach to investigate data that looks like evidence for LGT based on a rare or sparse gene distribution. Differential loss can—and will—produce last-one-out patterns that look just like lineage specific LGT. It is therefore possible, if not probable, that some reports suggesting evidence for LGT are perhaps last-one-out cases attributable to differential loss. We asked whether we could identify such cases in real data. In the following, we examine two studies that propose LGT as the cause of last-one-out topologies. Our aim is not to challenge these specific papers, but simply to see if the model proposed here (requiring only a tree, gene loss, and a specific case of gene distribution) can account for the data directly, without recourse to LGT. The aim of our study was to get an exact method to calculate the probability of observing such an event under a loss-only process.

One recent study is very helpful. Cote-L'Heureux et al. (2022) looked for lineage-specific presence of prokaryotic genes in eukaryotes that would provide the strongest possible evidence, in their view, for the workings of LGT from prokaryotes to eukaryotes. They sampled 13 600 gene families, 189 eukaryotic genomes and 540 eukaryotic transcriptomes, looking for recent lineage-specific LGT (topologies that we designate as last-one-out patterns). Among the 13 600 eukaryotic gene families sampled, they found ∼94 putative cases of LGT that represent a last-one-out pattern, that is, a restricted single-tip distribution of a prokaryotic gene in a eukaryotic genome or group, which they interpreted as strong evidence for LGT. Our present findings (Fig. 3) indicate that in Cote-L'Heureux et al. (2022) the number of cases identified in their study (94) is very close to the lower bound of the expectations for last-one-out topologies of similarly sized data sets, in which all the last-one-out topologies can be accounted for by differential loss alone, with no need to invoke LGT.

One clear prediction of lineage-specific LGT versus loss for last-one-out cases is this: If lineage-specific acquisition is the mechanism behind the observed rare presence pattern for a eukaryotic gene, then the acquisition would need to be evolutionarily late (that is, a tip acquisition). That is, the prokaryotic donor and the eukaryotic gene should share a higher degree of sequence

Downloaded from https://academic.oup.com/femsle/article/doi/10.1093/femsle/fnaf051/8140845 by Universitaets- und Landesbibliothek Duesseldorf user on 04 August 2025

similarity, on average, in comparison to genes that trace back to the eukaryotic common ancestor. This is the reasoning behind the analysis of Ku et al. (2015) and Ku and Martin (2016), who looked for evidence of recent acquisitions of prokaryotic genes in sequenced eukaryotic genomes. Ku et al. (2015) found that, in eukaryotic genomes, rare genes that have prokaryotic homologs were not more recently acquired (that is, they were not more similar to prokaryotic homologs) than genes that trace back to the eukaryotic common ancestor, suggesting that their rare occurrence is the result of differential loss rather than lineage-specific acquisition (Ku et al. 2015, Ku and Martin 2016) (Fig. 4a and b).

Cote-L'Heureux et al. (2022) employed the same test, making the same kind of comparison that Ku et al. (2015) performed, namely, they looked for cases in which the prokaryotic gene was acquired recently by the eukaryotic lineage, using the criterion of sequence similarity. What they found was the distribution shown in Fig. 4(c), namely that the cases they suspected to be LGTs were just as old, in terms of sequence divergence, as genes that were acquired from the mitochondrion. In other words, there were no obviously recent acquisitions, as all of the prokaryotic genes that they interpreted as recent LGTs had the hallmark of ancient acquisition, just as Ku et al. (2015) suggested. Cote-L'Heureux et al. (2022) offered no explanation for the finding that genes they interpreted as recent acquisitions via LGT were just as ancient, in terms of sequence identity, as genes acquired from mitochondria (Fig. 4c). One interpretation is that the genes in their LGT class were not LGTs after all but were the result of differential loss instead. Differential loss directly explains why such genes show just as much sequence divergence to prokaryotic homologues (Ku et al. 2015) (Fig. 4c) as genes present in the eukaryotic common ancestor. LGT models would need to invoke an *ad hoc* corollary assumption of substitution rate acceleration for every gene with a last-one-out pattern to account for the absence (Fig. 4c) of eukaryotic LGTs having high (> 70%) sequence similarity to prokaryotic homologs. Differential loss requires no rate acceleration corollary. Furthermore, the model presented here closely predicts the frequency of observing last-one-out patterns under a variety of topologies and loss rates, which becomes increasingly relevant as new data point to a gene rich mitochondrial ancestor (Leger and Gawryluk 2024).

A second recent study provide an additional opportunity to test the method. We investigated a dataset of 332 budding yeast species published by Shen et al. (2018). They reported 365 distinct events of horizontal gene transfer to yeast lineages. Of those, 230 appeared to be species-specific. Could those 230 cases also be the result of differential loss, unsuspected "last-one-out" cases? We analyzed the species tree provided in Shen et al. (2018) for "last-one-out" probabilities. The authors reported that the last common ancestor of budding yeasts was similar to an archaetypal member of its sister subphylum Pezizomycetes with ∼10 000–13 000 genes. We were therefore able to calculate the mean and median probabilities for a "last-one-out" and therefore the expected number of these events (Fig. 5). The mean probability across all possible loss rates in the interval between 0 and 1 for $1 - e^{-\mu}$ is roughly 0.0287 and the median probability for this interval is 0.0229. If the last common ancestor of budding yeasts had 10 000–13 000 genes, this results in 287–373 genes being statistically the "last-one-out" for the mean probabilities and 229–298 "last-one-out" cases for the median probability. In comparison to the 230 analyzed cases that are supposed to be the result of LGT according to Shen et al. (2018), the statistical probabilities of observing these "last-one-out" cases though differential loss (rather than LGT) is not unlikely at all, it is in agreement with the expectation, and our simple model is surprisingly accurate.

As a third example, we examined a dataset where differential loss rather than LGT was reported to be the cause of last-one-out topologies. Ku et al. (2015) clustered 956 053 protein sequences from 55 eukaryotic genomes across six supergroups and compared them to a total of 6 103 025 protein sequences from prokaryotes across 1847 bacterial and 134 archaeal genomes. They found a total of 2585 eukaryote–prokaryote clusters and 101 eukaryotic singletons [Supplemental Table 9 in Ku et al. (2015)] with prokaryotic homologs. The ancestral genome size of the last eukaryotic common ancestor of this dataset comprises 2686 genes. The calculation of last-one-out probabilities yielded a mean probability of 0.0967 and a median probability of 0.1057 (Fig. 6). Since the ancestral genome size was 2686 genes for this dataset, one would expect on average 260 last-one-out cases and 284 cases using the median probability using our method; the 101 last-one-out cases observed are fewer than expected. In this example, our simple algorithm again works on real world data. Considering that this algorithm is based on a loss-only model of evolution, the lower number of observed cases compared to expected cases will likely be the result of gene duplications, which play a significant role in eukaryotic evolution (Scannel et al. 2006, Hittinger and Carrol 2007, van de Peer et al. 2009), within this eukaryotic data set.

The role of selection

How do gene duplications and selection figure into this issue? Gene duplications, genome duplications, and recurrent duplications leading to gene family expansions lead to growth of gene families during evolution. Members of such families can, and do, undergo differential loss in different lineages. For a gene that was present in the eukaryote common ancestor, and that underwent loss across eukaryotic lineages in such a manner as to generate a last-one-out topology, the question arises about the role of selection in that process. Clearly, if the gene in question was essential during eukaryote evolution, selection would have maintained its presence in all lineages. Losses indicate phases of evolution in which the gene was required under some conditions for some lineages, with relaxed or absent selective pressures in others, allowing loss in some lineages but retention in others, possibly as a result of persistent selection or the gene having acquired a novel function. Such an example can be found in the evolution of Fe–Fe hydrogenases, where the gene is present and functional among green algal lineages, which often experience anaerobiosis (Happe and Kaminski 2002). Yet, during the transition to life on land in an atmosphere of 21% O_2, the gene lost its function in anaerobic energy metabolism, whereby a duplicate of the Fe–Fe hydrogenase gene acquired a new function in O_2-sensing in the land plant lineage instead (Gould et al. 2019). Changing environments or developmental contexts can alter the selective pressures that act upon gene retention, gene loss, or gain-of-function (Li et al. 2019).

Of course, it is also possible that selective pressures could, in principle, lead to gene gain via LGT. What form of selection is strong enough to cause LGT on time frames, where we can directly observe the effects? The experiment has already been done, we just need to tally the result—growth inhibitors. The best known, and best studied, example of selection for LGT is the spread of antibiotic resistance genes across bacteria in hospitals starting in the 1950s, which led to the discovery of both plasmids and LGT among bacteria. A literature search returns, for example, over 97 000 papers on bacteria∗ and antibiotic∗ and resistance∗ with over 14 000 of those papers containing the search terms plasmid∗ or transfer

Downloaded from https://academic.oup.com/femsle/article/doi/10.1093/femsle/fnaf051/8140845 by Universitaets- und Landesbibliothek Duesseldorf user on 04 August 2025

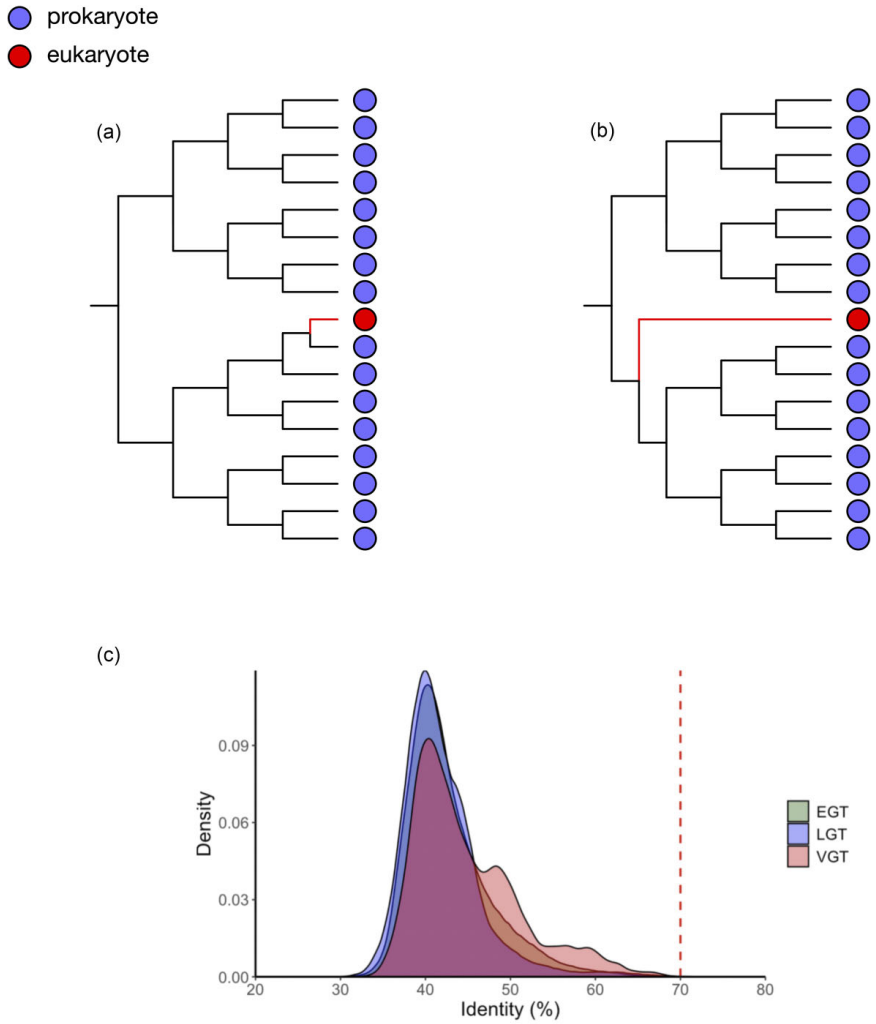

Figure 4. Similarity of eukaryotic last-one-out cases to prokaryotic homologs. (a) Phylogenetic distribution of genes, where the eukaryotic gene is considered to be the result of LGT due to its high similarity to one prokaryotic homolog. (b) The eukaryotic gene does not have a substantially high similarity to its prokaryotic homologs. It can therefore not be the result of recent LGT and is more likely the result of differential gene loss. (c) Supplementary Figure 9 from Cote-L'Heureux et al. (2022) showing that genes assumed to be the result of LGT are at most 70% similar to their prokaryotic homologs. This finding supports the "70% rule" of Ku and Martin (2016) and furthermore shows that these cases are more likely to be the result of differential loss instead of LGT. EGT: endosymbiotic gene transfer (genes acquired from chloroplasts or mitochondria); LGT: lateral gene transfer; and VGT: vertical gene transmission.

(The "∗" within the literature search are wildcards that stand for any amount of possible characters). Though a small sample, this underscores a point well-known among all microbiologists: bacteria immediately respond to antibiotic selective pressure by acquiring resistance genes via LGT. Antibiotic resistance represents a clear case of gene gain bringing selective benefit, in bacteria. Is the same true for eukaryotes, where reports for LGT have recently been reviewed (Keeling 2024)?

The closest eukaryotic equivalent to antibiotics in hospitals would be the use of fungicides in agriculture, which have been in use for over a century (Russell 2005). Does the strong selective pressure exerted by fungicides also lead to LGT among fungi, a group for which reports of LGT have also been reviewed (Richards and Talbot 2013)? A literature search returns over 13 000 papers on fung∗ and fungicid∗ and resistance∗, yet only 12 of those papers contained the search terms "horizontal gene transf∗" or "lat-

Downloaded from https://academic.oup.com/femsle/article/doi/10.1093/femsle/fnaf051/8140845 by Universitaets- und Landesbibliothek Duesseldorf user on 04 August 2025

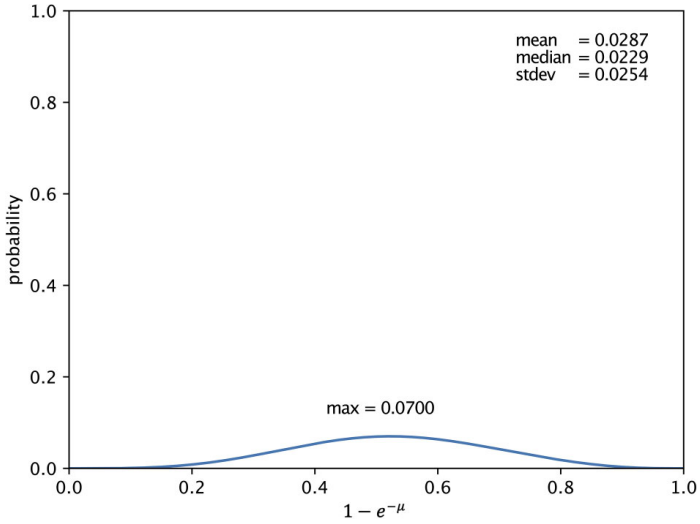

Figure 5. Probabilities of last-one-out cases across a spectrum of loss rates.

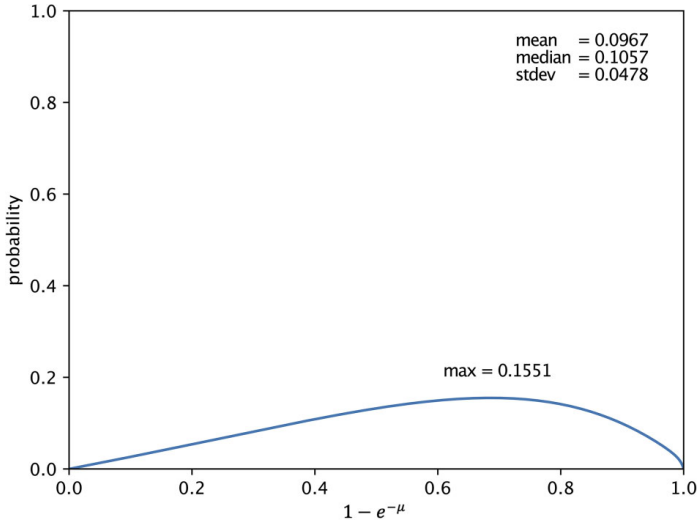

Figure 6. Probabilities of last-one-out phylogenies across a spectrum of loss rates for a eukaryotic species tree of 55 genomes from Ku et al. (2015) with forced monophyly for eukaryotic groups using a concatenated alignment of five genes universally present in those 55 genomes. For species that were originally used in Ku et al. (2015) but do not have a representative in the RefSeq dataset, we chose alternative genomes that are taxonomically near the original genomes. 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 tree was constructed with IQ-Tree (Nguyen et al. 2014), using the best-fit model and forcing monophyly of eukaryotic groups described in Ku et al. (2015) and the tree was rooted with MAD (Tria et al. 2017).

Downloaded from https://academic.oup.com/femsle/article/doi/10.1093/femsle/fnaf051/8140845 by Universitaets- und Landesbibliothek Duesseldorf user on 04 August 2025

eral gene transf∗"and only one of those 12 papers reported a case of phylogenetic evidence for LGT among fungi for a putative (not documented) resistance gene against cyanate (Elmore et al. 2015). The other 11 papers were mainly about fungicide-induced mobilization of plasmids in bacteria. Fungicide resistance represents a clear case where gene gain via LGT could bring (life-saving) benefit against a lethal selective pressure, but LGT is not observed. The relative frequency of LGT events conferring resistance to growth inhibitors (14 000 reported cases for bacteria, one possible candidate case reported for fungi) suggests that bacteria and fungi respond very differently to selection pressure generated by growth inhibitors. How do fungi respond?

Of course, resistance to fungicides is widespread and well-known in agriculture, as are resistance to insecticides and herbicides. Yet, the many known cases of resistance against fungicides, insecticides and herbicides do not entail LGT; instead, they involve *de novo* point mutations in the target-site encoding genes (fungicides), selection of polygenic metabolic resistance from standing variation (herbicides), and a combination of standing variation and *de novo* mutations in the target site or major metabolic resistance genes (insecticides) (Hawkins et al. 2019). That is, humans have already performed the experiment involving the application of strong selection pressure to prokaryotes (antibiotics in hospitals) and to eukaryotes (agricultural pests), and the result is generally clear. Bacteria respond by LGT of preexisting resistance genes, while eukaryotes respond by *de novo* point mutations and sexual recombination of standing variation, at least in cases reported so far. The present comparative example from the application of strong selective pressure indicates that there is a fundamental difference between prokaryotes and eukaryotes with regard to their tendency to acquire genes via LGT in response to strong (lethal) selection. However, both prokaryotes and eukaryotes do undergo gene loss at high rates.

As a caveat, strong selection for resistance is just one ecological context. A literature search for the search terms "fung∗" and "horizontal gene transf∗" or "lateral gene transf∗" returns over 1000 papers on fungal LGT (though none for resistance against fungicides), and we are by no means suggesting that those 1000 papers, usually founded in genome comparisons and gene phylogenies, and many invoking trait selection, are in error, collectively or otherwise. However, a recent study reexamined the strength of phylogenetic claims for LGT among fungi and found that only about 1.5% of trees that have been published as evidence for LGT among ascomycetes (the group of fungi that includes yeast) withstand critical inspection (Aguirre-Carvajal et al. 2025). If one accepts phylogenetic evidence for LGT among eukaryotes, the lack of abundant reports indicating eukaryotes to respond to strong selective pressure with LGT of resistance genes, while prokaryotes obviously do respond to selection with LGT, presents a puzzling observation. The ability of gene loss to generate last-one-out topologies at surprisingly high frequencies, as we have demonstrated here, might help to reconcile some discrepancies and contribute to solving the puzzle.

Conclusion

Sparse gene distributions in eukaryotes are often interpreted as evidence for gene acquisition via LGT from prokaryotes. However, gene loss can generate the same patterns, but estimates for the probability of observing a single gene at the tip of a phylogenetic tree as the result of differential loss within a given clade, as opposed to LGT, have been lacking, because methods were not even available. Here, we have derived the probability of observing such cases, which we call last-one-out patterns, because under a loss-only model, the last gene to be lost looks like an instance of LGT. The probability depends on the size and shape of the tree, and the loss rate μ. We find that the probability of observing a last-one-out topology can be (surprisingly) high.

This is not to say that there is no LGT to eukaryotes at all. But if LGT to eukaryotes were as common as many reviews would have us think, there have to be visible cumulative effects would have to accrue. That is, if we find a "new" gene in a eukaryotic lineage, and if we assume that LGT is going on all the time during evolution, then eukaryotic genomes should become increasingly patchwork over evolutionary time, which is exactly what we see in prokaryotes: A typical bacterial or archaeal genome contains genes from all sorts of different donors (Nagies et al. 2020), and in prokaryotes, the accessory genome (that component of the genome that is constantly in flux) typically comprises about 20%–30% of an average genome, and always has, ever since the first cells roamed the ocean floor 4 billion years ago (Trost et al. 2024). In eukaryotes we see cumulative effects for differential loss, for example in the case of the microsporidian *Encephalitozoon cuniculi* (Katinka et al. 2001), reduced parasitic fungi with a 2.9 Mb genome (smaller than *Escherichia coli*) or nucleomorph genomes, eukaryotic genomes that have shrunk to <700 kb in size (Gilson 2001). But in eukaryotes, we do not see cumulative effects for LGT. How so? If eukaryotic lineages acquired just one new gene from prokaryotes per million years, on average, then after 1.5 billion years of eukaryote evolution (Mills et al. 2022) separate eukaryotic supergroups would each harbor roughly 1500 different prokaryotic genes each. If that were the case, genomes would have told us so by now. But that is not what we see. Eukaryotes have different subsets of the same ancestral collection of genes (Müller et al. 2012, Ku et al. 2015, Brueckner and Martin 2020). Reviews of eukaryote LGT (Martin 2017, Keeling 2024) tend to cover case studies of single eukaryote genes or single eukaryote genomes, that is, an odd gene here or an odd genome there. Comparative studies involving many eukaryotic lineages are still rare. Now that we have a method to estimate the frequency of last-one-out topologies, we can compare the expectation for observing such "LGT-like" topologies as a result of differential loss. The cases we tested here are fully consistent with the expectations for differential loss, alleviating the need to assume LGT involving curious mechanisms, such as gene transfer via meteorites as vectors (Bergthorsson et al. 2003) as one prominent study suggested.

A simple algorithm applied to simulated eukaryotic trees provides estimates for the frequency of last-one-out patterns resulting from a loss only model that are slightly higher than, but generally in good agreement with, observations from a recent study in which all last-one-out topologies were interpreted as evidence for LGT. Gene loss is a prevalent process in eukaryotic genome evolution. If one lineage can lose a given gene, others can as well. Gene loss can, and does, generate patterns that look just like LGT. Even for large data sets, the probability of last-one-out topologies can be surprisingly large, because, depending upon the tree, the number of losses required to account for a last-one-out topology can be small.

Author contributions

Conceptualization, W.F.M. and M.S.; methodology, N.B., W.F.M., and M.S.; investigation, N.B., W.F.M., and M.S.; writing—original draft, N.B., W.F.M., and M.S.; writing—review & editing, N.B., W.F.M., and M.S.; funding acquisition, W.F.M. and M.S.; resources, W.F.M. and M.S.; supervision, W.F.M. and M.S.

Downloaded from https://academic.oup.com/femsle/article/doi/10.1093/femsle/fnaf051/8140845 by Universitaets- und Landesbibliothek Duesseldorf user on 04 August 2025

Supplementary data

Supplementary data are available at *FEMSLE Journal* online.

Conflict of interest: The authors declare that they have no competing interests.

Funding

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement number 101018894). M.S. thanks the Alexander von Humboldt Foundation for supporting his visit to Germany in 2023.

Data availability

Data files and/or online-only appendices can be found in the Dryad data repository: https://doi.org/10.5061/dryad.612jm649v. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

References

Aguirre-Carvajal K, Cárdenas S, Munteanu CR *et al.* Rampant interkingdom horizontal gene transfer in *Pezizomycotina*? An updated inspection of anomalous phylogenies. *Int J Mol Sci* 2025;**26**:1795.

Albalat R, Cañestro C. Evolution by gene loss. *Nat Rev Genet* 2016;**17**:379–91.

Andersson SGE, Kurland CG. Reductive evolution of resident genomes. *Trends Microbiol* 1998;**6**:263–8.

Bansal MS, Alm EJ, Kellis M. Efficient algorithms for the reconciliation problem with gene duplication, horizontal transfer and loss. *Bioinformatics* 2012;**28**:i283–i291.

Bergthorsson U, Adams KL, Thomason B *et al.* Widespread horizontal transfer of mitochondrial genes in flowering plants. *Nature* 2003;**424**:197–201.

Blanc G, Wolfe KH. Functional divergence of duplicated genes formed by polyploidy during *Arabidopsis* evolution. *Plant Cell* 2004;**16**:1679–91.

Boothby TC, Tenlen JR, Smith FW *et al.* Evidence for extensive horizontal gene transfer from the draft genome of a tardigrade. *Proc Natl Acad Sci USA* 2015;**112**:15976–81.

Bremer N, Knopp M, Martin WF *et al.* Realistic gene transfer to gene duplication ratios identify different roots in the bacterial phylogeny using a tree reconciliation method. *Life* 2022;**12**:995.

Brueckner J, Martin WF. Bacterial genes outnumber archaeal genes in eukaryotic genomes. *Genome Biol Evol* 2020;**12**:282–92.

Brunet FG, Crollius HR, Paris M *et al.* Gene loss and evolutionary rates following whole-genome duplication in teleost fishes. *Mol Biol Evol* 2006;**23**:1808–16.

Burki F, Roger AJ, Brown MW *et al.* The new tree of eukaryotes. *Trends Ecol Evol* 2020;**35**:43–55.

Consortium IHGS. Initial sequencing and analysis of the human genome. *Nature* 2001;**409**:860–921.

Cote-L'Heureux A, Maurer-Alcal´a XX, Katz LA. Old genes in new places: a taxon-rich analysis of interdomain lateral gene transfer events. *PLoS Genet* 2022;**18**:e1010239.

Dagan T, Martin WF. Ancestral genome sizes specify the minimum rate of lateral gene transfer during prokaryote evolution. *Proc Natl Acad Sci USA* 2007;**104**:870–5.

Elmore MH, McGary KL, Wisecaver JH *et al.*. Clustering of two genes putatively involved in cyanate detoxification evolved recently and independently in multiple fungal lineages. *Genome Biol Evol* 2015;**7**:789–800.

Gilson PR. Nucleomorph genomes: much ado about practically nothing. *Genome Biol* 2001;**2**:reviews1022.1.

Goodman M, Czelusniak J, Moore GW *et al.* Fitting the gene lineage into its species lineage, a parsimony strategy illustrated by cladograms constructed from globin sequences. *Syst Zool* 1979;**28**:132–63.

Gould SB, Garg SG, Handrich M *et al.* Adaptation to life on land at high O_2 via transition from ferredoxin-to NADH-dependent redox balance. *Proc Biol Sci* 2019;**286**:20191491.

Guijarro-Clarke C, Holland PWH, Paps J. Widespread patterns of gene loss in the evolution of the animal kingdom. *Nat Ecol Evol* 2020;**4**:519–23.

Happe T, Kaminski A. Differential regulation of the Fe-hydrogenase during anaerobic adaptation in the green alga *Chlamydomonas reinhardtii*. *Eur J Biochem* 2002;**269**:1022–32.

Hawkins NJ, Bass C, Dixon A *et al.* The evolutionary origins of pesticide resistance. *Biol Rev* 2019;**94**:135–55.

Hittinger C, Carrol S. Gene duplication and the adaptive evolution of a classic genetic switch. *Nature* 2007;**449**:677–81.

Hosokawa T, Kikuchi Y, Nikoh N *et al.* Strict host–symbiont cospeciation and reductive genome evolution in insect gut bacteria. *PLoS Biol* 2006;**4**:e337.

Katinka MD, Duprat S, Cornillot E *et al.* Genome sequence and gene compaction of the eukaryote parasite *Encephalitozoon cuniculi*. *Nature* 2001;**414**:450–3.

Katoh K, Misawa K, Kuma K *et al.* MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* 2002;**30**:3059–66.

Keeling PJ, Burki F. Progress towards the tree of eukaryotes. *Curr Biol* 2019;**29**:R808–17.

Keeling PJ. Horizontal gene transfer in eukaryotes: aligning theory with data. *Nat Rev Genet* 2024;**25**:416–30.

Kellis M, Birren BW, Lander ES. Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature* 2004;**428**:617–24.

Kendall DG. On the generalized 'birth-and-death' process. *Ann Math Stat* 1948;**19**:1–15.

Koutsovoulos G, Kumar S, Laetsch DR *et al.* No evidence for extensive horizontal gene transfer in the genome of the tardigrade *Hypsibius dujardini*. *Proc Natl Acad Sci USA* 2016;**113**:5053–8.

Ku C, Martin WF. A natural barrier to lateral gene transfer from prokaryotes to eukaryotes revealed from genomes: the 70% rule. *BMC Biol* 2016;**14**:89.

Ku C, Nelson-Sathi S, Roettger M *et al.* Endosymbiotic origin and differential loss of eukaryotic genes. *Nature* 2015;**524**:427–32.

Lambert A, Stadler T. Birth–death models and coalescent point processes: the shape and probability of reconstructed phylogenies. *Theor Popul Biol* 2013;**90**:113–28.

Leger MM, Gawryluk RMR. Evolution: a gene-rich mitochondrial genome sheds light on the last eukaryotic common ancestor. *Curr Biol* 2024;**34**:R776–9.

Li Y, Zhang Y, Li X *et al.* Gain-of-function mutations: an emerging advantage for cancer biology. *Trends Biochem Sci* 2019;**44**:659–74.

Martin WF, Stoebe B, Goremykin V *et al.* Gene transfer to the nucleus and the evolution of chloroplasts. *Nature* 1998;**393**:162–5.

Martin WF. Too much eukaryote LGT. *Bioessays* 2017;**39**:1700115.

Mills DB, Boyle RA, Daines SJ *et al.* Eukaryogenesis and oxygen in Earth history. *Nat Ecol Evol* 2022;**6**:520–32.

Moore CE, Archibald JM. Nucleomorph genomes. *Annu Rev Genet* 2009;**43**:251–64.

Downloaded from https://academic.oup.com/femsle/article/doi/10.1093/femsle/fnaf051/8140845 by Universitaets- und Landesbibliothek Duesseldorf user on 04 August 2025

Moran NA, Bennett GM. The tiniest tiny genomes. *Annu Rev Microbiol* 2014;**68**:195–215.
Müller M, Mentel M, van Hellemond JJ *et al.* Biochemistry and evolution of anaerobic energy metabolism in eukaryotes. *Microbiol Mol Biol Rev* 2012;**76**:444–95.
Nagies FSP, Brueckner J, Tria FDK *et al.* A spectrum of verticality across genes. *PLoS Genet* 2020;**16**:1–28.
Nguyen L, Schmidt HA, von Haeseler A *et al.* IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 2015;**32**:268–74.
Nicholson D, Salamina M, Panek J *et al.* Adaptation to genome decay in the structure of the smallest eukaryotic ribosome. *Nat Commun* 2022;**13**:591.
Olson MV. When less is more: gene loss as an engine of evolutionary change. *Am Hum Genet* 1999;**64**:18–23.
Oshima K, Kakizawa S, Nishigawa H *et al.* Reductive evolution suggested from the complete genome sequence of a plant-pathogenic phytoplasma. *Nat Genet* 2004;**36**:27–29.
Page RDM. Maps between trees and cladistic analysis of historical associations among genes, organisms, and areas. *Syst Biol* 1994;**43**:58–77.
Richards TA, Talbot NJ. Horizontal gene transfer in osmotrophs: playing with public goods. *Nat Rev Microbiol* 2013;**11**:720–7.
Russel PE. A century of fungicide evolution. *J Agric Sci* 2005;**143**: 11–25.
Salzberg SL, White O, Peterson J *et al.* Microbial genes in the Human genome: lateral transfer or gene loss?. *Science* 2001;**292**: 1903–6.
Salzberg SL. Horizontal gene transfer is not a hallmark of the human genome. *Genome Biol* 2017;**18**:85.
Scannel DR, Byrne KP, Gordon JL *et al.* Multiple rounds of speciation associated with reciprocal gene loss in polyploid yeasts. *Nature* 2006;**440**:341–5.
Shen X, Opulente DA, Kominek J *et al.* Tempo and mode of genome evolution in the budding yeast subphylum. *Cell* 2018;**175**:1533–1545.e20.
Stanhope MJ, Lupas A, Italia MJ *et al.* Phylogenetic analyses do not support horizontal gene transfers from bacteria to vertebrates. *Nature* 2001;**411**:940–4.
Szöllősi GJ, Rosikiewicz W, Boussau B *et al.* Efficient exploration of the space of reconciled gene trees. *Syst Biol* 2013;**62**:901–12.
Timmis JN, Ayliffe MA, Huang CY *et al.* Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nat Rev Genet* 2004;**5**:123–35.
Tovar J, Le´on-Avila G, S´anchez LB *et al.* Mitochondrial remnant organelles of Giardia function in iron-sulphur protein maturation. *Nature* 2003;**426**:172–6.
Tria F, Landan G, Dagan T. Phylogenetic rooting using minimal ancestor deviation. *Nat Ecol Evol* 2017;**1**:0193.
Trost K, Knopp MR, Wimmer JLE *et al.* A universal and constant rate of gene content change traces pangenome flux to LUCA. *FEMS Microbiol Lett* 2024;**371**:fnae068.
van de Peer Y, Maere S, Meyer A. The evolutionary significance of ancient genome duplications. *Nat Rev Genet* 2009;**10**:725–32.
van Ham RCHJ, Kamerbeek J, Palacios C *et al.* Reductive genome evolution in *Buchnera aphidicola*. *Proc Natl Acad Sci USA* 2003;**100**:581–6.
Waters E, Hohn MJ, Ahel I *et al.* The genome of *Nanoarchaeum equitans*: insights into early archaeal evolution and derived parasitism. *Proc Natl Acad Sci USA* 2003;**100**:12984–8.

Downloaded from https://academic.oup.com/femsle/article/doi/10.1093/femsle/fnaf051/8140845 by Universitaets- und Landesbibliothek Duesseldorf user on 04 August 2025

Received 4 March 2025; **revised** 29 April 2025; **accepted** 20 May 2025
© The Author(s) 2025. Published by Oxford University Press on behalf of FEMS. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs licence (https://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

6 Zusammenfassung der Ergebnisse

Die Entstehung der ersten eukaryotischen Zellen vor etwa 1,5 Milliarden Jahren (Parfrey *et al.* 2011, Chernikova *et al.* 2011, Eme *et al.* 2014, Betts *et al.* 2018, Mahendrarajah *et al.* 2023) markiert den Anfang der Entwicklungsgeschichte des komplexen Lebens. Eindeutige fossile Eukaryoten vor dieser Zeit fehlen. Dieser Umstand erschwert eine Rekonstruktion der Einzelschritte und deren zeitlichen Abfolge während der Eukaryogenese. Folglich ist man in der Frage nach dem Ursprung der Eukaryoten auf Ableitung aus molekularen Daten angewiesen (Eme *et al.* 2014). Einen möglichen Einblick in die frühe Entwicklungsgeschichte der Eukaryoten bildet LECA, der letzte gemeinsame Vorfahre aller heute lebenden Eukaryoten. LECA fungiert als wissenschaftliches Konstrukt und ist ebenfalls nicht durch Fossilienfunde gestützt (Eme *et al.* 2014). Eine Rekonstruktion von LECA ermöglicht jedoch Rückschlüsse auf die Merkmale der ersten eukaryotischen Zellen, sowie der Eukaryogenese zu ziehen. Infolgedessen hat sich LECA in den Fokus der Forschung zum Thema gestellt (z.B. Margulis *et al.* 2006, Katz 2012, Koumandou *et al.* 2013, O'Malley *et al.* 2019, Gabaldón 2021, Skejo *et al.* 2021, Bremer *et al.* 2022, Bremer *et al.* 2023, Krupovic *et al.* 2023, Vosseberg *et al.* 2024).

In **Publikation I** werden ursprüngliche, eukaryotische Merkmale aufgrund ihrer Verteilung innerhalb des eukaryotischen Stammbaums auf LECA rekonstruiert. Durch die Rekonstruktion anhand von 1789 verschiedenen, gewurzelten Gen-Bäumen, mit einem breiten Spektrum an eukaryotischen Vertretern aus sechs verschiedenen eukaryotischen Supergruppen, wird ein neuer phylogenetischer Ansatz gewählt, der die Problematik der verschiedenen publizierten Wurzelpositionen im eukaryotischen Baum umgeht (Burki *et al.* 2020, Cerón-Romero *et al.* 2022, Al Jewari und Baldauf 2023, Eme und Tamarit 2024). Hierfür wurde zuvor die Verteilung der folgenden fünf Merkmale in heute lebenden Eukaryoten zusammengefasst: (I) Phagozytose als Prozess der Aufnahme von Partikeln über Einstülpungen der Plasmamembran (II) Phagotrophie als eine Ernährungsweise, (III) das Auftreten von Mitochondrien beziehungsweise ihrer verwandten Zellorganellen, (IV) das Auftreten von Plastiden, sowie (V) die Mehrkernigkeit der eukaryotischen Zelle. Die Rekonstruktion dieser Merkmale ergibt, dass LECA im Besitz von Mitochondrien ist, keine Plastiden besitzt

und eine Mehrkernigkeit vorliegt. Des Weiteren wird festgestellt, dass LECA weder über Phagozytose verfügt, noch Phagotrophie als Ernährungsweise ausübt und beide Merkmale mehrfach innerhalb der Eukaryoten entstanden sind. Die Existenz von Mitochondrien in LECA, gekoppelt mit der Abwesenheit der Phagozytose, lässt zudem darauf schließen, dass der Mechanismus, über den die Mitochondrien während der Eukaryogenese in die Zelle gelangten, nicht Phagozytose war. Dadurch können Modelle der Eukaryogenese, die einen Ursprung der Mitochondrien durch Phagozytose beschreiben, wie etwa das PhAT-Modell (Martijn und Ettema 2013), abgelehnt werden. Vielmehr kann der Ursprung der Mitochondrien mit einer mikrobiellen Symbiose erklärt werden (Bremer *et al.* 2022).

In einer weiteren Analyse wird in **Publikation II** der anzestrale Zustand der Mitose, sowie die Fähigkeit der sexuellen Fortpflanzung in LECA rekonstruiert. Dabei wird die Mitose auf drei spezifische Ausprägungen aufgeteilt: (I) die Intaktheit der Zellkernhülle, wobei dabei zwischen offen, halb-offen und geschlossen unterschieden wird (Raikov 1994), (II) die intranukleare oder extranukleare Positionierung der Spindel (Raikov 1994), sowie (III) der symmetrischen Ausbildung der Spindel, wobei hier zwischen einer axialen und einer bilateralen Symmetrie unterschieden wird (Raikov 1994). Anhand derselben Gen-Bäume wird auch in dieser Analyse auf eine bestimmte Wurzel des eukaryotischen Baumes verzichtet und vielmehr eine Mittelung vieler verschiedener phylogenetischer Anordnungen der einzelnen Supergruppen, sowie Wurzeln des eukaryotischen Baumes verwendet. Die Rekonstruktion dieser untersuchten Merkmale zeigt, dass LECA sich sexuell fortgepflanzt hat und die in LECA verwendete Mitose eine geschlossene Zellkernhülle mit einer intranuklearen Positionierung der Spindel, die wiederum eine axiale Symmetrie aufweisen, hatte.

Dieses Ergebnis ist kompatibel zu der bereits in **Publikation I** rekonstruierten Mehrkernigkeit (Skejo *et al.* 2021, Bremer *et al.* 2022), da eine geschlossene Mitose drei wichtige Prozesse voneinander abkoppelt: (I) Chromosomentrennung, (II) Chromosomenaufteilung und (III) Zellteilung. Dies führt dazu, dass sich zum einen diese drei Prozesse unabhängig voneinander entwickeln können und zum anderen können schädliche Kombinationen von Chromosomen durch andere Zellkerne innerhalb desselben Zytosols durch mRNA ausgeglichen werden (Bremer *et al.* 2023). Mutationen, die zuerst einen Selektionsnachteil schaffen, bevor durch weitere Mutationen ein Selektionsvorteil erzeugt wird, könnten so ermöglicht werden (Orr

2009). Insbesondere während sich schnell ändernder Umweltfaktoren ist eine Mehrkernigkeit für die Anpassung an stetig wechselnde Bedingungen von Vorteil. So kann dies die Effekte der „Muller-Ratsche“ in asexuellen Eukaryoten hinauszögern (Muller 1964, Kondrashov 1994). Die „Muller-Ratsche“ beschreibt die Akkumulation von nachteiligen Mutationen in einer begrenzten Population in Abwesenheit von Rekombination. Da sich LECA während der Eukaryogenese zwangsläufig bis zu einem gewissen Punkt asexuell fortpflanzen musste, wird die Mehrkernigkeit hier von Vorteil gewesen sein.

Schlussfolgernd aus den Analysen aus **Publikation I** und **Publikation II**, sowie weiterer Rekonstruktionen von LECA in der Literatur, ergibt sich ein immer genaueres Bild des letzten gemeinsamen eukaryotischen Vorfahren. So wird deutlich, dass LECA folgende Merkmale besitzt: einen Zellkern (Mans *et al.* 2004, Bapteste *et al.* 2005, Neumann *et al.* 2010), ein endoplasmatisches Retikulum (Kontou *et al.* 2022), lineare Chromosomen mit Zentromeren (Ishikawa und Naito 1999, van Hooff et al, 2017), Flagellen (Carvalho-Santos *et al.* 2011, Lindemann 2022), ein Mikrotubuli-organisierendes Zentrum (Yubuki und Leander 2013), einen Nukleolus (Gardner *et al.* 2010, Hoeppner und Poole 2012), sowie die Fähigkeit zur sexuellen Fortpflanzung mittels Meiose (Villeneuve und Hillers 2001, Loidl 2016). Des Weiteren wurde hier gezeigt, dass LECA zudem Mitochondrien (Bremer *et al.* 2022), eine Mehrkernigkeit (Skejo *et al.* 2021, Bremer *et al.* 2022), sowie eine geschlossene Mitose mit intranuklearen, axial angeordneten Spindeln besitzt (Bremer *et al.* 2023). Zudem wurde nachgewiesen, dass keine Plastiden in LECA vorhanden sind (Bremer *et al.* 2022) und weder Phagozytose, noch Phagotrophie in LECA existiert (Bremer *et al.* 2022).

Gene, die nur in einer einzigen Klade des eukaryotischen Baumes vorkommen und zugleich homologe Sequenzen in Prokaryoten aufweisen, werden häufig auf lateralen Gentransfer zurückgeführt (International Human Genome Sequencing Consortium 2001, Boothby *et al.* 2015, Martin 2017). **Publikation III** hingegen zeigt, dass sich derartige Verteilungen auch mit einem Auftreten des Gens in LECA und sukzessiver Genverluste im eukaryotischen Stammbaum erklären lassen. Diese Gene werden auch als „letzte Hinterbliebene“ (engl. last-one-out) bezeichnet. Für diese Analyse wurde ein einfacher Algorithmus implementiert, der die Wahrscheinlichkeit errechnet, dass in einem phylogenetischen Baum, mit einer Präsenz des Gens im

letzten gemeinsamen Vorfahren, genau eine Spezies das Gen erhält, während es in allen anderen Abstammungslinien früher oder später verloren geht. Dabei wurde die Wahrscheinlichkeit über ein Spektrum an Genverlustraten gemittelt, um den variierenden Genverlustraten für unterschiedliche Gene, sowie in verschiedenen Gruppen des eukaryotischen Baumes, zu entsprechen (Albalat und Cañestro 2016). Dies ist notwendig, da der Verlust bzw. Erhalt von Genen stark davon abhängig ist, welche Rolle diese Gene für die jeweilige Population spielen. Ein Gen, welches für die Fortpflanzung innerhalb einer Population essentiell ist, wird beispielsweise weniger häufig verloren, als ein Gen, welches keinen oder nur einen sehr geringen selektiven Vorteil verschafft.

Um die Anwendbarkeit dieses Algorithmus zu testen, wurden zehn zufällig ausgewählte eukaryotische Gen-Bäume mit jeweils 150 Spezies als Testdatensatz verwendet. Diese Bäume spiegeln hierbei weder die echte Phylogenie der Eukaryoten ab, noch müssen diese Bäume eine Verteilung von Genen aufweisen, die auf lateralen Gentransfer hinweisen würden. Die Bäume dienen lediglich der Veranschaulichung von unterschiedlichen Wahrscheinlichkeiten für „letzte Hinterbliebene“ bei verschiedenen Baumtopologien. Unter der Annahme, dass 10000 Gene in LECA vorhanden waren, ergibt sich ein Mittelwert von 232 für die geringste Wahrscheinlichkeit bis 790 für die höchste Wahrscheinlichkeit innerhalb dieser zehn Bäume, bei denen das Gen in LECA, sowie einer einzigen Spezies des Baumes vorhanden ist und in allen weiteren Abstammungslinien verloren wurde. Dies zeigt, dass in diesem Testdatensatz davon ausgegangen werden muss, dass hunderte Gene rein statistisch einen „letzten Hinterbliebenen“ bilden und nicht das Ergebnis von lateralem Gentransfer sind.

In einer Analyse von Cote-L’Heureux *et al.* (2022) wurden in 13630 Genfamilien, welche 189 eukaryotische Genome und 540 eukaryotischen Transkriptome beinhalteten, 306 laterale Gentransfers von Prokaryoten zu Eukaryoten festgestellt, wobei 189 als endosymbiontische Gentransfers charakterisiert wurden, 52 LGTs in der eukaryotische Supergruppe der Opisthokonta und 42 LGTs in anaerobe Eukaryoten gefunden wurden. Im Vergleich zu den ermittelten Wahrscheinlichkeiten für zufällig gewählte eukaryotische Bäume ähnlicher Größe, ist die gefundene Anzahl an „letzte Hinterbliebene“-Topologien von Cote-L’Heureux *et al.* (2022) verhältnismäßig nah an dem unteren Rand unserer berechneten Wahrscheinlichkeiten. Eine Erklärung dieser

von Cote-L'Heureux *et al.* (2022) gefundenen Verteilung ist also ebenso mit differentiellem Genverlust zu erklären, anstelle von lateralem Gentransfer.

In einer weiteren Analyse wird in **Publikation III** der Algorithmus an einem Datensatz von Shen *et al.* (2018) angewandt. Hier wurden 365 LGTs, wovon 230 spezies-spezifisch waren, in der Abstammungslinie der Hefen gefunden. Diese 230 LGTs spiegeln somit eine „letzte Hinterbliebene"-Topologie wider und könnten ebenfalls durch differentiellem Genverlust erklärt werden. Der letzte gemeinsame Vorfahre dieser Hefen soll zwischen 10000 und 13000 Genen gehabt haben. Eine Berechnung der mittleren Wahrscheinlichkeit für den vorliegenden Spezies-Baum ergibt 287–373 Gene, die eine „letzte Hinterbliebene"-Topologie ergeben und das Ergebnis von Genverlust sind. Der Median der Wahrscheinlichkeiten liegt mit 229–298 derartiger Topologien ziemlich genau im Bereich der von Shen *et al.* (2018) gefundenen LGTs. Eine Erklärung dieser Topologien ist somit auch mithilfe von Genverlust anstelle von lateralem Gentransfer möglich.

In einer dritten Analyse wurde ein Datensatz von Ku *et al.* (2015) analysiert. Hierbei wurden in der ursprünglichen Analyse 101 eukaryotische „letzte Hinterbliebene"-Topologien gefunden. Die Anwendung des Algorithmus ergab bei einer angenommenen anzestralen Genomgröße von 2686 Genen eine durchschnittliche Wahrscheinlichkeit von 260 „letzte Hinterbliebene"-Topologien, sowie 284 dieser Topologien bei Betrachtung des Medians. Die in dieser Analyse höhere Anzahl zu erwartenden derartiger Topologien im Gegensatz zu den beobachteten Topologien lässt sich dadurch erklären, dass der Algorithmus ausschließlich Genverluste betrachtet und Genduplikationen, die eine erhebliche Rolle in der eukaryotischen Entwicklungsgeschichte spielen, zur Vereinfachung nicht in die Berechnungen mit einbezieht.

Was entscheidend bei derartigen Beobachtungen hinzukommt, ist die 70%-Regel (Ku und Martin 2016). Sollte lateraler Gentransfer von Prokaryoten zu Eukaryoten möglich sein, muss dies ein fortwährender Prozess sein und laterale Gentransfers, die erst kürzlich erfolgt sind, sollten eine hohe Ähnlichkeit zu ihren prokaryotischen Homologen besitzen. Es wurde jedoch bewiesen, dass diese Gene maximal 70% Ähnlichkeit zu ihren prokaryotischen Homologen besitzen und damit, was die Sequenzunterschiede angeht, sich nicht von Genen unterscheiden lassen, die durch das Mitochondrium aufgenommen wurden. Die hohe Wahrscheinlichkeit eben

dieser hier vorgestellten „letzte Hinterbliebene“-Topologien durch Genverlust, anstelle von lateralem Gentransfer, unterstützen zuvor publizierte Thesen bezüglich fehlendem lateralem Gentransfer von Prokaryoten in die eukaryotische Domäne.

7 Literaturverzeichnis

Adl, S.M., Bass, D., Lane, C.E., Lukeš, J., Schoch, C.L., Smirnov, A., Agatha, S., Berney, C., Brown, M.W., Burki, F., Cárdenas, P., Čepička, I., Chistyakova, L., del Campo, J., Dunthorn, M., Edvardsen, B., Eglit, Y., Guillou, L., Hampl, V., Heiss, A.A., Hoppenrath, M., James, T.Y., Karnkowska, A., Karpov, S., Kim, E., Kolisko, M., Kudryavtsev, A., Lahr, D.J.G., Lara, E., Le Gall, L., Lynn, D.H., Mann, D.G., Massana, R., Mitchell, E.A.D., Morrow, C., Park, J.S., Pawlowski, J.W., Powell, M.J., Richter, D.J., Rueckert, S., Shadwick, L., Shimano, S., Spiegel, F.W., Torruella, G., Youssef, N., Zlatogursky, V. und Zhang, Q. (2019). Revisions to the classification, nomenclature, and diversity of eukaryotes. *The Journal of Eukaryotic Microbiology* 66:4–119.

Al Jewari, C. und Baldauf, S. L. (2023). An excavate root for the eukaryote tree of life. *Science Advances* 9:eade4973.

Albalat, R. und Cañestro, C. (2016). Evolution by gene loss. *Nature Reviews Genetics* 17:379–391.

Archibald, J. M., Simpson, A. G. B. und Slamovits, C. H. (2017). Handbook of the protists. Cham: Springer International Publishing.

Arnold, B. J., Huang, I-T., Hanage, W. P. (2022). Horizontal gene transfer and adaptive evolution in bacteria. *Nat Rev Microbiol* 20:206-218.

Bapteste, E., Charlebois, R. L., MacLeod, D. und Brochier, C. (2005). The two tempos of nuclear pore complex evolution: highly adapting proteins in an ancient frozen structure. *Genome Biology* 6:R85.

Bartel, D. P. und Szostak, J. W. (1993). Isolation of new ribozymes from a large pool of random sequences. *Science* 261:1411–1418.

Baum, D. A. und Baum, B. (2014). An inside-out origin for the eukaryotic cell. *BMC Biology* 12:76.

Baum, B. und Baum, D. A. (2020). The merger that made us. *BMC Biology* 18:72.

Bernhardt, G., Lüdemann, H. D., Jaenicke, R., König, H. und Stetter, K. O. (1984). Biomolecules are unstable under “black smoker” conditions. *Naturwissenschaften* 71:583–6.

Betts, H. C., Puttick, M. N., Clark, J. W., Williams, T. A., Donoghue, P. C. J. und Pisani, D. (2018). Integrated genomic and fossil evidence illuminates life´s early evolution and eukaryote origin. *Nature Ecology & Evolution* 2:1556–1562.

Blais, C. und Archibald, J. M. (2021). The past, present and future of the tree of life. *Current Biology* 31:PR314–R321.

Boothby, T. C., Tenlen, J. R., Smith, F. W., Wang, J. R., Patanella, K. A., Nishimura, E. O., Tintori, S. C., Li, Q., Jones, C. D., Yandell, M., Messina, D. N., Glasscock, J. und Goldstein, B. (2015). Evidence for extensive horizontal gene transfer from the draft genome of a tardigrade. *Proceedings of the National Academy of Sciences of the United States of America* 112:15976–15981.

Bremer, N., Tria, F. D. K., Skejo, J., Garg, S. G. und Martin, W. F. (2022). Ancestral state reconstructions trace mitochondria but not phagocytosis to the last eukaryotic common ancestor. *Genome Biology and Evolution* 14:evac079.

Bremer, N., Tria, F. D. K., Skejo, J. und Martin, W. F. (2023). The ancestral mitotic state: Closed orthomitosis with intranuclear spindles in the syncytial last eukaryotic common ancestor. *Genome Biology and Evolution* 15:evad016.

Brunk, C. F. und Martin, W. F. (2019). Archaeal histone contributions to the origin of eukaryotes. *Trends in Microbiology* 27:703–714.

Burki, F., Roger, A. J., Brown, M. W. und Simpson, A. G. B. (2020). The new tree of eukaryotes. *Trends in Ecology and Evolution* 35:P43–55.

Carvalho-Santos, Z., Azimzadeh, J., Pereira-Leal, J. B. und Bettencourt-Dias, M. (2011). Tracing the origins of centrioles, cilia, and flagella. *Journal of Cell Biology* 194:165–175.

Cavalier-Smith, T. (1975). The origin of nuclei and of eukaryotic cells. *Nature* 256:463–468.

Cavalier-Smith, T. (1987). The origin of cells: A symbiosis between genes, catalysts, and membranes. *Cold Spring Harbor Symposia on Quantitative Biology* 52:804–824.

Cavalier-Smith, T. (2002). The phagotrophic origin of eukaryotes and phylogenetic classification of protozoa. *International Journal of Systematic and Evolutionary Microbiology* 52:297–354.

Cerón-Romero, M. A., Fonseca, M. M., de Oliveira Martins, L., Posada, D. und Katz, L. A. (2022). Phylogenomic analyses of 2,786 genes in 158 lineages support a

root of the eukaryotic tree of life between Opisthokonts and all other lineages. *Genome Biology and Evolution* 14:evac119.

Chernikova, D., Motamedi, S., Csürös, M., Koonin, E. V. und Rogozin, I. B. (2011). A late origin of the extant eukaryotic diversity: divergence time estimates using rare genomic changes. *Biology Direct* 6:26.

Corliss, J. B., Baross, J. A. und Hoffmann, S. (1981). An hypothesis concerning the relationship between submarine hot springs and the origin of life on Earth. *Oceanologica Acta* 1980:59–69.

Cote-L'Heureux, A., Maurer-Alcalá, X. X. und Katz, L. A. (2022). Old genes in new places: A taxon-rich analysis of interdomain lateral gene transfer events. *PLoS Genetics* 18:e1010239.

Darwin, C. (1860). Über die Entstehung der Arten im Thier- und Pflanzen-Reich durch natürliche Züchtung. Stuttgart: Schweizerbart.

Darwin Correspondence Project, "Letter no. 7471," aufgerufen am 17 Februar 2025, https://www.darwinproject.ac.uk/letter/?docId=letters/DCP-LETT-7471.xml

Donoghue, P. C. J., Kay, C., Spang, A., Szöllősi, G., Nenarokova, A., Moody, E. R. R., Pisani, D. und Williams, T. (2023). Defining eukaryotes to dissect eukaryogenesis. *Current Biology* 33:PR919–R929.

Doolittle, W. F. (1998). You are what you eat: a gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes. *Trends in Genetics* 14:P307–311.

Doolittle, W. F. (1999). Lateral genomics. *Trends in Biochemical Sciences* 24:PM5–M8.

Eme, L., Sharpe, S. C., Brown, M. W. und Roger, A. J. (2014). On the age of eukaryotes: Evaluating evidence from fossils and molecular clocks. *Cold Spring Harbor Perspectives in Biology* 6:a016139.

Eme, L. und Tamarit, D. (2024). Microbial diversity and open questions about the deep tree of life. *Genome Biology and Evolution* 16:evae053.

Forterre, P. und Philippe, H. (1999). Where is the root of the universal tree of life? *BioEssays* 21:871–879.

Gabaldon, T. (2021). Origin and early evolution of the eukaryotic cell. *Annual Review of Microbiology* 75:631–647.

Gardner, P. P., Bateman, A. und Poole, A. M. (2010). SnoPatrol: how many snoRNA genes are there? *Journal of Biology* 9:4.

Garg, S. G. und Martin, W. F. (2016). Mitochondria, the cell cycle, and the origin of sex via a syncytial eukaryote common ancestor. *Genome Biology and Evolution* 8:1950–1970.

Genereux, D. P. und Logsdon, J. M. (2003). Much ado about bacteria-to-vertebrate lateral gene transfer. *Trends in Genetics* 19:P191–195.

Gould, G. W. und Dring, G. J. (1979). On a possible relationship between bacterial endospore formation and the origin of eukaryotic cells. *Journal of Theoretical Biology* 81:47–53.

Gupta, R. S. und Golding, G. B. (1996). The origin of the eukaryotic cell. *Trends in Biochemical Sciences* 21:166–171.

Haeckel, E. (1866a). Generelle Morphologie der Organismen Erster Band: Allgemeine Anatomie der Organismen. Berlin: Verlag von Georg Reimer.

Haldane, J. B. S. (1929). The origin of life. *The Rationalist Annual* 148:3–10.

Hoeppner, M. P. und Poole, A. M. (2012). Comparative genomics of eukaryotic small nucleolar RNAs reveals deep evolutionary ancestry amidst ongoing intragenomic mobility. *BMC Evolutionary Biology* 12:183.

Horiike, T., Hamada, K., Miyata, D. und Shinozawa, T. (2004). The origin of eukaryotes is suggested as the symbiosis of Pyrococcus into γ-Proteobacteria by phylogenetic tree based on gene content. *Journal of Molecular Evolution* 59:606–619.

Imachi, H., Nobu, M. K., Nakahara, N., Morono, Y., Ogawara, M., Takaki, Y., Takano, Y., Uematsu, K., Ikuta, T., Ito, M., Matsui, Y., Miyazaki, M., Murata, K., Saito, Y., Sakai, S., Song, C., Tasumi, E., Yamanaka, Y., Yamaguchi, T, Kamagata, Y., Tamaki, H. und Takai, K. (2020). Isolation of an archaeon at the prokaryote-eukaryote interface. *Nature* 577:519–525.

International Human Genome Sequencing Consortium (2001). Initial sequencing and analysis of the human genome. *Nature* 409:860–921.

Ishikawa, F. und Naito, T. (1999). Why do we have linear chromosomes? A matter of Adam and Eve. *Mutation Research/DNA Repair* 434:99–107.

Javaux, E. J., Knoll, A. H. und Walter, M. R. (2001). Morphological and ecological complexity in early eukaryotic ecosystems. *Nature* 412:66-69.

Jones, D. und Sneath, P. H. A. (1970). Genetic transfer and bacterial taxonomy. *Bacteriological Reviews* 34:40–81.

Jupp, T. und Schultz, A. (2000). A thermodynamic explanation for black smoker temperatures. *Nature* 403:880–883.

Katscher, F. (2004). The history of the terms prokaryotes and eukaryotes. *Protist* 155:257-263.

Katz, L. A. (2012). Origin and diversification of eukaryotes. *Annual Review of Microbiology* 66:411–427.

Kelley, D. S., Karson, J. A., Blackman, D. K., Früh-Green, G. L., Butterfield, D. A., Lilley, M. D., Olson, E. J., Schrenk, M. O., Roe, K. K., Lebon, G. T., Rivizzigno, P. und AT3-60 Shipboard Party (2001). An off-axis hydrothermal vent field near the mid-atlantic ridge at 30° n. *Nature* 412:145–9.

Kelley, D. S., Karson, J. A., Früh-Green, G. L., Yoerger, D. R., Shank, T. M., Butterfield, D. A., Hayes, J. M., Schrenk, M. O., Olson, E. J., Proskurowski, G., Jakuba, M., Bradley, A., Larson, B., Ludwig, K., Glickson, D., Buckman, K., Bradley, A. S., Brazelton, W. J., Roe, K., Elend, M. J., Delacour, A., Bernasconi, S. M., Lilley, M. D., Baross, J. A., Summons, R. E. und Sylva, S. P. (2005). A serpentinite-hosted ecosystem: The lost city Hydrothermal field. *Science* 307:1428–1434.

Kondrashov, A. S. (1994). Muller's ratchet under epistatic selection. *Genetics* 136:1469–1473.

Kontou, A., Herman, E. K., Field, M. C., Dacks, J. B. und Koumandou, V. L. (2022). Evolution of factors shaping the endoplasmic reticulum. *Traffic* 1:1–12.

Koumandou, V. L., Wickstead, B., Ginger, M. L., van der Giezen, M., Dacks, J. B. und Field, M. C. (2013). Molecular paleontology and complexity in the last eukaryotic common ancestor. *Critical Reviews in Biochemistry and Molecular Biology* 48:373–396.

Krupovic, M., Dolja, V. V. und Koonin, E. V. (2023). The virome of the last eukaryotic common ancestor and eukaryogenesis. *Nature Microbiology* 8:1008–1017.

Ku, C., Nelson-Sathi, S., Roettger, M., Garg, S., Hazkani-Covo, E. und Martin, W. F. (2015). Endosymbiotic gene transfer from prokaryotic pangenomes: Inherited chimerism in eukaryotes. *Proceedings of the National Academy of Sciences of the United States of America* 112:10139–10146.

Ku, C. und Martin, W. F. (2017). A natural barrier to lateral gene transfer from prokaryotes to eukaryotes revealed from genomes: the 70 % rule. *BMC Biology* 14:89.

Lake, J. A. und Rivera, M. C. (1994). Was the nucleus the first endosymbiont? *Proceedings of the National Academy of Sciences of the United States of America* 91:2880–2881.

Lane, N. und Martin, W. F. (2010). The energetics of genome complexity. *Nature* 467:929–934.

Lindemann, C. B. (2022). The flagellar germ-line hypothesis: how flagellate and ciliate gametes significantly shaped the evolution of organismal complexity. *BioEssays* 44:2100143.

Loidl, J. (2016). Conservation and variability of meiosis across the eukaryotes. *Annual Review of Genetics* 50:293–316.

Lopéz-García, P. und Moreira, D. (1999). Metabolic symbiosis at the origin of eukaryotes. *Trends in Biochemical Sciences* 24:88–93.

Lopéz-García, P. und Moreira, D. (2020). The syntrophy hypothesis for the origin of eukaryotes revisited. *Nature Microbiology* 5:655–667.

Mahendrarajah, T. A., Moody, E. R. R., Schrempf, D., Szánthó, L. L., Dombrowski, N., Davín, A. A., Pisani, D., Donoghue, P. C. J., Szöllősi, G. J., Williams, T. A. und Spang, A. (2023). ATP synthase evolution on a cross-braced dated tree of life. *Nature Communications* 14:7456.

Mans, B. J., Anantharaman, V., Aravind, L. und Koonin, E. V. (2004). Comparative genomics, evolution and origin of the nuclear envelope and nuclear pore complex. *Cell Cycle* 3:1625–1650.

Margulis, L. (1970). Origin of eukaryotic cells. New Haven: Yale University Press.

Margulis, L. und Fester, R. (1991). Symbiosis as a source of evolutionary innovation. Boston: MIT Press.

Margulis, L. (2004). Serial endosymbiotic theory (SET) and composite individuality. *Microbiology Today* 31:172–175.

Margulis, L., Chapman, M., Guerrero, R. und Hall, J. (2006). The last eukaryotic common ancestor (LECA): Acquisition of cytoskeletal motility from aerotolerant spirochetes in the Proterozoic Eon. *Proceedings of the National Academy of Sciences of the United States of America* 103:13080–13085.

Martijn, J. und Ettema, T. J. G. (2013). From archaeon to eukaryote: the evolutionary dark ages of the eukaryotic cell. *Biochemical Society Transactions* 41:451-457

Martin, W. F. und Müller, M. (1998). The hydrogen hypothesis for the first eukaryote. *Nature* 393:162–165.

Martin, W. F. (1999). Mosaic bacterial chromosomes: a challenge en route to a tree of genomes. *BioEssays* 21:99–104.

Martin, W. F. und Russel, M. J. (2003). On the origin of cells: A hypothesis for the evolutionary transitions from abiotic geochemistry to chemoautotrophic prokaryotes, and from prokaryotes to nucleated cells. *Philosophical Transactions of the Royal Society B* 358:59–85.

Martin, W. F., Garg, S. und Zimorski, V. (2015). Endosymbiotic theories for eukaryote origin. *Philosophical Transactions of the Royal Society B* 370:20140330.

Martin, W. F., Tielens, A. G. M., Mentel, M., Garg, S. G. und Gould, S. B. (2017). The physiology of phagocytosis in the context of mitochondrial origin. *Microbiology and Molecular Biology Reviews* 81:e00008–17.

Martin, W. F. (2017). Too much eukaryote LGT. *BioEssays* 39:1700115.

Mereschkowsky, C. (1905). Über Natur und Ursprung der Chromatophoren im Pflanzenreiche. *Biologisches Centralblatt* 25:593–604.

Miller, S. L. (1953). A production of amino acids under possible primitive Earth conditions. *Science* 117:528–9.

Mills, D. B. (2020). The origin of phagocytosis in earth history. *Interface Focus* 10:20200019.

Moreira, D. und López-García, P. (1998). Symbiosis between methanogenic archaea and δ-Proteobacteria as the origin of eukaryotes: The syntrophic hypothesis. *Journal of Molecular Evolution* 47:517–530.

Muller, H. J. (1964). The relation of recombination to mutational advance. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* 1:2–9.

Neumann, N, Lundin, D. und Poole, A. M. (2010). Comparative genomic evidence for a complete nuclear pore complex in the last eukaryotic common ancestor. *PLoS One* 5:e13241.

Noller, H. F., Hoffarth, V. und Zimniak, L. (1992). Unusual resistance of peptidyl transferase to protein extraction procedures. *Science* 256:1416–1419.

Ochman, H., Lawrence, J. G. und Groisman, E. A. (2000). Lateral gene transfer and the nature of bacterial innovation. *Nature* 405:299–304.

O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D., Astashyn, A., Badretdin, A., Bao, Y., Blinkova, O., Brover, V., Chetvernin, V., Choi, J., Cox, E., Ermolaeva, O., Farrell, C. M., Goldfarb, T., Gupta, T., Haft, D., Hatcher, E., Hlavina, W., Joardar, V. S., Kodali, V. K., Li, W., Maglott, D., Masterson, P., McGarvey, K. M., Murphy, M. R., O'Neill, K., Pujar, S., Rangwala, S. H., Rausch, D., Riddick, L. D., Schoch, C., Shkeda, A., Storz, S. S., Sun, H., Thibaud-Nissen, F., Tolstoy, I., Tully, R. E., Vatsan, A. R., Wallin, C., Webb, D., Wu, W., Landrum, M. J., Kimchi, A., Tatusova, T., DiCuccio, M., Kitts, P., Murphy, T. D. und Pruitt, K. D. (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research* 44:D733–D745.

O'Malley, M. A., Leger, M. M., Wideman, J. G. und Ruiz-Trillo, I. (2019). Concepts of the last eukaryotic common ancestor. *Nature* 3:338–344.

Oparin, A. I. (1938). The origin of life. Translation with annotations by Sergius Morgulis. New York: The Macmillan Company.

Orr, H. A. (2009). Fitness and its role in evolutionary genetics. *Nature Reviews Genetics* 10:531–539.

Parfrey, L. W., Lahr, D. J. G., Knoll, A. H. und Katz, L. A. (2011). Estimating the timing of early eukaryotic diversification with multigene molecular clocks. *Proceedings of the National Academy of Sciences of the United States of America* 108:13624–13629.

Poole, A., Jeffares, D. und Penny, D. (1999). Early evolution: prokaryotes, the new kids on the block. *BioEssays* 21:880–889.

Poole, A. M. und Gribaldo, S. (2014). Eukaryotic origins: How and when was the mitochondrion acquired? *Cold Spring Harbor Perspectives in Biology* 6:a015990.

Preiner, M., Igarashi, K., Muchowska, K. B., Yu, M., Varma, S. J., Kleinermanns, K., Nobu, M. K., Kamagata, Y., Tüysüz, H., Moran, J. und Martin, W. F. (2020). A hydrogen-dependent geochemical analogue of primordial carbon and energy metabolism. *Nature Ecology and Evolution* 4:534–42.

Raikov, I. B. (1994). The diversity of forms of mitosis in protozoa: a comparative review. *European Journal of Protistology* 30:253–269.

Roger, A. J., Muñoz-Gómez, S. A. und Kamikawa, R. (2017). The origin and diversification of mitochondria. *Current Biology* 27:R1177–R1192.

Rona, P. A., Klinkhammer, G., Nelsen, T. A., Trefry, J. H. und Elderfield, H. (1986). Black smokers, massive sulphides and vent biota at the Mid-Atlantic ridge. *Nature* 321:33–37.

Russel, M. J., Hall, A. J. und Martin, W. (2010). Serpentinization as a source of energy at the origin of life. *Geobiology* 8:355–371.

Sagan, L. (1967). On the origin of mitosing cells. *Journal of Theoretical Biology* 14:225–274.

Salzberg, S. L., White, O., Peterson, J. und Eisen, J. A. (2001). Microbial genes in the human genome: Lateral transfer or gene loss? *Science* 292:1903–1906.

Salzberg, S. L. (2017). Horizontal gene transfer is not a hallmark of the human genome. *Genome Biology* 18:85.

Sapp, J. (1994). Evolution by association: A history of symbiosis. New York: Oxford University Press.

Shen, X. X., Opulente, D. A., Kominek, J., Zhou, X., Steenwyk, J. L., Buh, K. V., Haase, M. A. B., Wisecaver, J. H., Wang, M., Doering, D. T., Boudouris, J. T., Schneider, R. M., Langdon, Q. K., Ohkuma, M., Endoh, R., Takashima, M., Manabe, R., Čadež, N., Libkind, D., Rosa, C. A., DeVirgilio, J., Hulfachor, A. B., Groenewald, M., Kurtzman, C. P., Hittinger, C. T. und Rokas, A. (2018). Tempo and mode of genome evolution in the budding yeast subphylum. *Cell* 175:1533–1545.

Skejo, J., Garg, S. G., Gould, S. B., Hendriksen, M., Tria, F. D. K., Bremer, N., Franjević, D., Blackstone, N. W. und Martin, W. F. (2021). Evidence for a syncytial origin of eukaryotes from ancestral state reconstruction. *Genome Biology and Evolution* 13:evab096.

Spang, A., Saw, J. H., Jørgensen, S. L., Zaremba-Niedzwiedzka, K., Martijn, J., Lind, A. E., van Eijk, R., Schleper, C., Guy, L. und Ettema, T. J. G. (2015). Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature* 521:173–179.

Spang, A., Stairs, C. W., Dombrowski, N., Eme, L., Lombard, J. Caceres, E. F., Greening, C., Baker, B. J. und Ettema, T. J. G. (2019). Proposal of the reverse flow model for the origin of the eukaryotic cell based on comparative analyses of Asgard archaeal metabolism. *Nature Microbiology* 4:1138–1148.

Stanhope, M. J., Lupas, A., Italia, M. J., Koretke, K. K., Volker, C. und Brown, J. R. (2001). Phylogenetic analyses do not support horizontal gene transfers from bacteria to vertebrates. *Nature* 411:940–944.

Urey, H. C. (1952). On the early chemical history of the Earth and the origin of life. *Proceedings of the National Academy of Sciences of the United States of America* 38:351–63.

Van Hooff, J. J., Tromer, E., van Wijk, L. M., Snel, B. und Kops, G. J. (2017). Evolutionary dynamics of the kinetochore network in eukaryotes as revealed by comparative genomics. *EMBO reports* 18:1559–1571.

Vellai, T. und Vida, G. (1999). The origin of eukaryotes: the difference between prokaryotic and eukaryotic cells. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 266:1571-1577.

Villeneuve, A. M. und Hillers, K. J. (2001). Whence meiosis? *Cell* 106:647-650.

Vosseberg, J., van Hooff, J. J., Marcet-Houben, M., van Vlimmeren, A., van Wijk, L. M., Gabaldón, T und Snel, B. (2021). Timing the origin of eukaryotic cellular complexity with ancient duplications. *Nature Ecology & Evolution* 5:92–100.

Vosseberg, J., van Hooff, J. J. E., Köstlbacher, S., Panagiotou, K., Tamarit, D., Ettema, T. J. G. (2024). The emerging view on the origin and early evolution of eukaryotic cells. *Nature* 633:295–305.

Wallin, I. E. (1925). On the nature of mitochondria. IX. Demonstration of the bacterial nature of mitochondria. *American Journal of Anatomy* 36:131–149.

Wallin, I. E. (1927). Symbionticism and the origin of species. Baltimore: The Williams & Wilkins Company.

Wimmer, J. L. E., Xavier, J. C., do Nascimento Vieira, A., Pereira, D. P. H., Leidner, J., Sousa, F. L., Kleinermanns, K., Preiner, M. und Martin, W. F. (2021c). Energy at origins: Favorable thermodynamics of biosynthetic reactions in the Last Universal Common Ancestor (LUCA). *Frontiers in Microbiology* 12:793664.

Yubuki, N. und Leander, B. S. (2013). Evolution of microtubule organizing centers across the tree of eukaryotes. *The Plant Journal* 75:230–244.

Zaremba-Niedzwiedzka, K., Caceres, E. F., Saw, J. H., Bäckström, D., Juzokaite, L., Vancaester, E., Seitz, K. W., Anantharaman, K., Starnawski, P., Kjeldsen, K. U., Stott, M. B., Nunoura, T., Banfield, J. F., Schramm, A., Baker, B. J., Spang, A. und Ettema, T. J. G. (2017). Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* 541:353–358.

Danksagung

Zuerst möchte ich mich bei Prof. Dr. William F. Martin bedanken, der mir in unsicheren Covid-Zeiten eine Doktorandenstelle angeboten hat und mich in meinem Thema stets unterstütze.
Des Weiteren bedanke ich mich bei Prof. Dr. Martin Lercher für die Übernahme des Zweitgutachtens dieser Arbeit.

I would like to thank my former supervisor Dr. Fernando Tria who helped me a ton during my Master's degree and the first two years of my PhD. Your advices and critical questions improved my scientific thinking tremendously and your knowledge about eukaryotes and phylogenetic analyses strongly inspired me in my research.
I would also like to thank Dr. Josip Skejo for his enthusiasm about early eukaryotic evolution. You helped me a lot understanding this complex topic.

Ein weiterer Dank gilt meinen Korrekturlesern Andrea Alexa, Katharina Trost, Loraine Schwander, Natalia Mrnjavac und Luca Modjewski, die sowohl inhaltlich, als auch grammatikalisch diese Arbeit deutlich verbessert haben.

Zudem möchte ich mich bei allen aus dem Molevol-Team und speziell dem Spaßbüro bedanken. Ihr seid der Grund, warum ich trotz des alltäglichen Stresses immer wieder gerne ins Büro gekommen bin.

Zu guter Letzt danke ich meinen Eltern für ihre umfassende Unterstützung auf meinem Weg durch Schule, Studium und Promotion. Ihr habt immer an mich geglaubt und mir in schwierigen Phasen die Hilfe an die Hand gegeben, die mich zum Weitermachen motivierte. Ohne euch wäre dies alles nicht möglich gewesen.