nature ecology & evolution

ARTICLES

https://doi.org/10.1038/s41559-022-01833-9

The role of mitochondrial energetics in the origin and diversification of eukaryotes

Paul E. Schavemaker [1,3] ✉ and Sergio A. Muñoz-Gómez [2,3] ✉

The origin of eukaryotic cell size and complexity is often thought to have required an energy excess supplied by mitochondria. Recent observations show energy demands to scale continuously with cell volume, suggesting that eukaryotes do not have higher energetic capacity. However, respiratory membrane area scales superlinearly with the cell surface area. Furthermore, the consequences of the contrasting genomic architectures between prokaryotes and eukaryotes have not been precisely quantified. Here, we investigated (1) the factors that affect the volumes at which prokaryotes become surface area-constrained, (2) the amount of energy divested to DNA due to contrasting genomic architectures and (3) the costs and benefits of respiring symbionts. Our analyses suggest that prokaryotes are not surface area-constrained at volumes of 10^0–10^3 μm^3, the genomic architecture of extant eukaryotes is only slightly advantageous at genomes sizes of 10^6–10^7 base pairs and a larger host cell may have derived a greater advantage (lower cost) from harbouring ATP-producing symbionts. This suggests that eukaryotes first evolved without the need for mitochondria since these ranges hypothetically encompass the last eukaryotic common ancestor and its relatives. Our analyses also show that larger and faster-dividing prokaryotes would have a shortage of respiratory membrane area and divest more energy into DNA. Thus, we argue that although mitochondria may not have been required by the first eukaryotes, eukaryote diversification was ultimately dependent on mitochondria.

The transition from prokaryotic to eukaryotic cells is often thought to be the greatest transition in the history of life[1]. This is because this is the largest gap, or discontinuity, in organismal structure or organization across the tree of life: eukaryotic cells are structurally much more complex, and on average, also larger in volume than prokaryotic cells[2]. Many authors have thus attempted to explain how eukaryotes evolved from prokaryotes[3–9]. However, much debate and speculation persist about the processes that gave rise to the first eukaryote[3,10–13].

To explain the apparent large gap or gulf in complexity between prokaryotes and eukaryotes, the energetic hypothesis for eukaryote genome complexity suggests that there is also a deep energetic divide between these two grades of organization[3] (see refs. [8,14–16] for precursors). Lane and Martin claimed that eukaryotes have, on average, approximately 200,000 times more 'energy per gene' than prokaryotes[3]. Such a drastic energetic difference is supposedly caused by two major advantages conferred by mitochondria on eukaryotes[3,17–19]. The first one is the internalization and expansion of respiratory membranes within mitochondria, which released eukaryotes from surface area constraints. The second one is the evolution of highly reduced and specialized mitochondrial genomes that conferred a genomic asymmetry on eukaryotes. Unlike prokaryotes, which have a single genome that scales up in number proportionally with cell volume, eukaryotes have a single large nuclear genome whose copy number can remain constant, as well as many much smaller mitochondrial genomes that scale up in number with cell volume. The combination of these two advantages, according to Lane and Martin, allowed a drastic increase in the energy available per gene expressed in eukaryotes relative to prokaryotes[3,17–19]. One possible interpretation of this hypothesis predicts a jump in energetic capacity that separates eukaryotes from prokaryotes (Fig. 1a). Mitochondria are the cause of these massive energetic differences, Lane and Martin argue, and were thus a prerequisite for the evolution of eukaryotic complexity[3,19].

Some authors expressed scepticism about the energetic hypothesis for the origin of eukaryotic complexity[10,13,20–25]. The notion that the evolution of cell complexity requires an increase in energy supply has been dismissed as having no evolutionary basis[23,25] and the concept of energy per gene has been criticized as evolutionarily meaningless[24,26]. Recently, Lynch and Marinov found a continuous energetic scaling across prokaryotes and unicellular eukaryotes[10]; similar results have been presented by Chiyomaru and Takemoto[20]. This suggests that there is no energetic gap (or shift in energetic capacity) between prokaryotes and eukaryotes because the amount of energy available to a cell is directly proportional to its volume regardless of whether the cell is prokaryotic or eukaryotic. Based on this, Lynch and Marinov argued that mitochondria do not provide eukaryotes with a higher energetic capacity and implied that prokaryotes are energetically unconstrained by their cell surfaces (Fig. 1b). Moreover, Lynch and Marinov showed that the number of ATP synthases scales continuously across prokaryotes and eukaryotes and argued that the increase in surface area provided by mitochondria is not particularly large when compared to that available at the cytoplasmic membrane[22]. However, their data also showed that the amount of mitochondrial membrane and the number of ATP synthases scale superlinearly with the cell surface area[22]. This suggests, in contrast to Lynch and Marinov[10,22], that prokaryotes might be constrained by their cell surfaces at larger volumes and that mitochondria may allow eukaryotes to scale up in cell volume without a shortage of respiratory membranes (Fig. 1c). Furthermore, the energetic consequences of the contrasting genomic architectures between prokaryotes and eukaryotes, first emphasized by Lane and Martin[3,17] but ignored by others, remain unaddressed.

[1]Center for Mechanisms of Evolution, The Biodesign Institute, School of Life Sciences, Arizona State University, Tempe, AZ, USA. [2]Unité d'Ecologie, Systématique et Evolution, Université Paris-Saclay, Orsay, France. [3]These authors contributed equally: Paul E. Schavemaker, Sergio A. Muñoz-Gómez. ✉e-mail: pschavem@asu.edu; sergio.munoz@universite-paris-saclay.fr

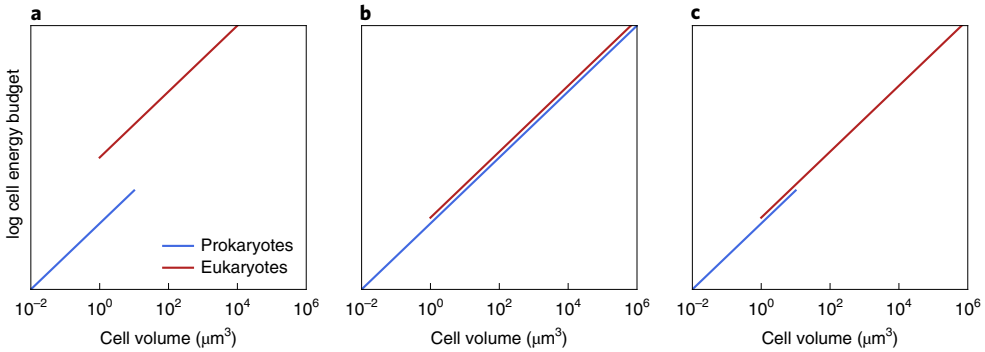


Fig. 1 | Three different possibilities for energetic scaling across cell volume for prokaryotes and eukaryotes. a, A hypothetical discontinuity in the scaling of cell energy with volume between prokaryotes and eukaryotes, where the latter exhibit a higher energetic capacity or energy density due to mitochondria. The magnitude of the energetic gap shown serves an illustrative purpose only. **b**, A hypothetical scaling in the absence of surface constraints to prokaryotic cell volume where the energetic capacity of prokaryotes accompanies that of eukaryotes over the full cell volume range. **c**, A continuous scaling of cell energy with volume over the prokaryote–eukaryote divide based on data presented by Lynch and Marinov[10] and Chiyomaru and Takemoto[20]. Unlike in **b**, the cell volume of prokaryotes is constrained. This constraint may be caused by the lack of a cytoskeleton, endomembrane system or mitochondrion-based respiration.

To explore the potential energetic benefits that mitochondria bestowed on eukaryotes, our goal in this study has been to carefully dissect major differences between mitochondrion-less and mitochondrion-bearing cells (that is, prokaryotes and eukaryotes, respectively) in light of the recent scaling laws devised by Lynch and Marinov[10,22]. To do so, we (1) explored potential factors (cell shape, cell division time and maximum fraction of respiratory membrane) that affect the volumes at which mitochondrion-less cells become surface area-constrained, (2) investigated the decreases in energy budget associated with the contrasting genomic architectures exhibited by mitochondrion-less and mitochondrion-bearing cells across a wide range of cell volumes and (3) examined the costs and benefits of a population of respiring symbionts in a host cell. We discuss our observations in the context of the prokaryote–eukaryote divide and the origin and diversification of eukaryotes.

Results

The divide between prokaryotic and eukaryotic cells. In this article, we use theoretical models to assess the respiratory membrane requirements and DNA investments of mitochondrion-less and mitochondrion-bearing cells. These models might help explain, from an energetic point of view, the differences observed between modern prokaryotes and eukaryotes and thus inform our discussions of the prokaryote–eukaryote transition. We start by presenting the distributions of cell volume, genome size and gene number from a comprehensive survey of phylogenetically disparate prokaryotes and eukaryotes (Fig. 2 and Source Data Fig. 2).

The cell volume distributions of prokaryotes and eukaryotes point at two main conclusions. First, the ranges for each grade of organization (approximately 10^{-2}–10^{2} μm^3 for prokaryotes and approximately 10^{0}–10^{8} μm^3 for eukaryotes) do not overlap for the most part: their medians (vertical dashed lines in Fig. 2, calculated over the full distribution of cell volumes) largely fall outside of each other's distributions (Fig. 2a). This is most obvious when giant bacteria like *Beggiatoa* spp. and *Thiomargarita namibiensis* are excluded (Fig. 2a and Source Data Fig. 2). Giant bacteria reach absolute volumes >10^{6} μm^3 but these are mostly inert because they contain a large central vacuole or many intracellular inclusions made of sulfur or calcium carbonate[27] (some exceptions are large cyanobacterial cells; Source Data Fig. 2). Thus, most prokaryotes are smaller than most eukaryotes. Second, prokaryotes and eukaryotes mostly overlap at cell volumes of approximately 10^{0}–10^{2} μm^3 (Fig. 2a). This overlap includes large bacteria with entirely active cytoplasms composed of energy-demanding macromolecules (for example, *Azotobacter chroococcum*, *Magnetobacterium bavaricum* and '*Candidatus* Uabimicrobium amorphum'; Source Data Fig. 2), picoeukaryotes, which are relatively reduced (for example, algae such as *Chaetoceros calcitrans*, *Micromonas pusilla*, *Nannochloris* spp. and *Nannochloropsis geditana*; Source Data Fig. 2) and phylogenetically diverse nanoeukaryotes (for example, heterotrophic flagellates such as *Andalucia godoyi*, *Mantamonas plastica*, *Bodo saltans*, *Malawimonas jakobiformis*, *Palpitomonas bilix*, *Ancyromonas mylnikovi*, *Reclinomonas americana*; Source Data Fig. 2). Thus, many small eukaryotes (both parasitic and free-living) can have sizes similar to those of many bacteria.

The histogram of genome sizes follows a similar pattern to that of cell volumes: prokaryotes and eukaryotes have distinct but overlapping distributions (Fig. 2b). The genome size range for prokaryotes is less than 1–16 megabase pair (Mbp), whereas that of eukaryotes is approximately 8–10,000 Mbp. This suggests that there is an upper genome size constraint to prokaryotes based on the currently available data. Prokaryotes and eukaryotes also overlap at genomes sizes of approximately 8–16 Mbp if the genomes of highly reduced eukaryotic parasites are excluded (Fig. 2b and Source Data Fig. 2). Thus, many eukaryotes (protists) have genome sizes smaller than those of some prokaryotes. For example, prokaryotes such as myxobacteria, actinomycetes, cyanobacteria and planctomycetes may have genomes of up to 16 Mbp in size (Source Data Fig. 2). The smallest genomes for free-living eukaryotes are those of some small green algae, red algae and yeasts (8–13 Mbp); some parasitic eukaryotes have genome sizes of just 2 or 6 Mbp (for example, *Encephalitozoon* and *Babesia*; Source Data Fig. 2). The small heterotrophic nanoflagellate *A. godoyi* (Jakobea), which has the most ancestral-like mitochondrial genome, has a nuclear genome size of approximately 20 Mbp[28], barely larger than the largest prokaryotic genomes. For gene number, there is an even wider overlap between prokaryotes and eukaryotes (Fig. 2c). Prokaryotes with the greatest number of genes have 10,000–13,000 genes (Source Data Fig. 2), whereas eukaryotes with the lowest number of genes include intracellular parasites (approximately 2,000 genes in *Encephalitozoon*), free-living fungi (approximately 4,500 genes in *Malassezia restricta* or approximately 6,400 in *Saccharomyces cerevisiae*) and small algae (approximately 5,300 genes in *Cyanidioschyzon* and approximately 7,800 genes in *Ostreococcus tauri*). Some of the closest relatives of animals, the free-living flagellate *Monosiga brevicollis* (Choanoflagellata) and the symbiotic amoeba *Capsaspora owczarzaki* have approximately 9,200 and 8,800 genes, respectively

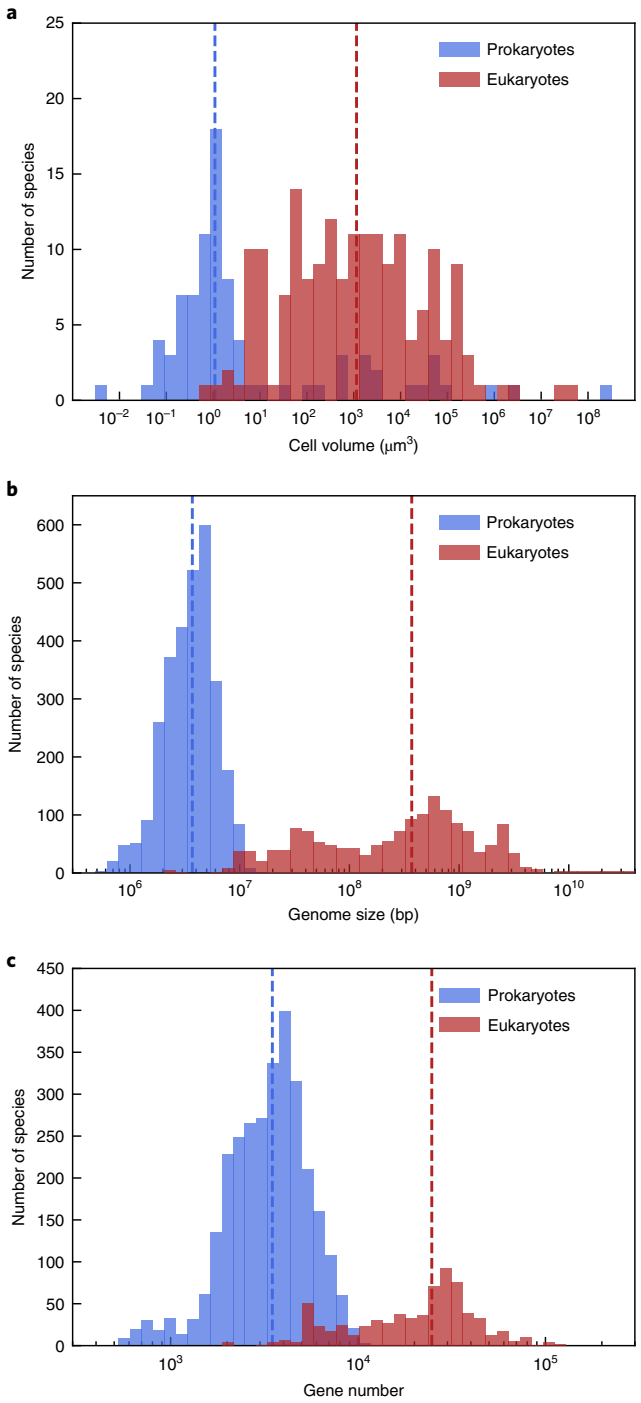


Fig. 2 | Cell volumes, genome sizes and gene numbers for prokaryotes and eukaryotes. a–c, Cell volumes (**a**) for diverse eukaryotes were obtained from Lynch and Marinov[10] and additional data were added from several sources (Source Data Fig. 2). Genome sizes (**b**) and gene numbers (**c**) were acquired from the National Center for Biotechnology Information GenBank and manually curated to remove outliers due to gene misannotations. The vertical dashed lines show the medians. Total cell volumes, instead of energy-demanding active cytoplasmic volume, were used for giant prokaryotes (>10^2 μm³).

(Source Data Fig. 2). In summary, the data suggest that, although there is some overlap between prokaryotes and eukaryotes, there are also upper constraints to cell volumes, gene numbers and genome sizes that prokaryotes can attain.

The respiratory membrane requirements of cells. The cell volume of prokaryotes is potentially constrained by respiration at the cell surface or by not having mitochondria. In terms of energy, the rate of ATP synthesis at the cell surface must meet the rate of ATP consumption by the whole cell volume. However, surface area decreases relative to volume as cells grow larger—surface area scales with the square of the linear dimension, whereas volume scales with the cube of the linear dimension[29–31]. A developmental or evolutionary increase in cell volume thus poses a challenge to cells because, if internal volumes remain active, processes that are carried out at the cell surface (for example, respiration or nutrient transport) will, at some cell volume, be unable to support processes that occur in the cytoplasm (for example, protein synthesis). Such scaling sets a maximum volume that cells cannot overcome in the absence of structural adaptations (for example, mitochondria and endomembranes in eukaryotes and intracytoplasmic membranes in prokaryotes[32]).

To determine the volumes at which cells first face a deficit in respiratory membrane, we examined the ratio between the amount of respiratory membrane needed and the maximum amount of respiratory membrane possible for a mitochondrion-less cell (that is, a prokaryote). This ratio provides a measure of respiratory deficit or the degree to which there is an excess or dearth of surface area allocated to respiration. Note that we do not assume any major structural adaptations (for example, internalized membranes, external membrane protrusions or appendages or internal inert spaces). The amount of respiratory membrane needed can be defined as the membrane area occupied by all respiratory complexes (or respiratory units) that are required to sustain the volume of a cell throughout its lifetime (A_{needed}, in μm²). The maximum amount of respiratory membrane, in turn, is defined as the largest possible membrane area that can be devoted to respiratory complexes by a cell ($A_{possible}$, in μm²). This area is necessarily only a fraction (f_{max}) of the total membrane area available (A_{total}, in μm²) because a cell also needs to allocate some of its surface area to lipids, nutrient transporters, protein translocases and flagella; thus, f_{max} represents the maximum fraction of the total surface membrane that can be used for respiration. The respiratory deficit can then be expressed as:

$$\text{Respiratory deficit} = \frac{A_{needed}}{A_{possible}} = \frac{A_{needed}}{f_{max}A_{total}} \quad (1)$$

The amount of respiratory membrane needed by a cell can be calculated by multiplying the number of respiratory units (that is, a complete set of respiratory complexes including ATP synthase) required by the membrane area that each one of them takes up (A_r). The number of respiratory units can be estimated by dividing the metabolic rate of a cell (R, in ATP h^{-1}) by the ATP production rate of a single respiratory unit (r, in ATP h^{-1}; Supplementary Table 1). The metabolic rate of a cell can be expressed as the total ATP budget of a cell throughout its lifetime (E_t) divided by the cell division time (t_d in h). The total energy budget of a cell (in ATP units) comprises both growth and maintenance costs (c_g and c_m, respectively) and is calculated as in Lynch and Marinov[10]; this is adjusted to only include direct costs using the f_d factor[33]. Growth costs are those associated with biomass accumulation, whereas maintenance costs are those associated with non-growth-related processes, such as protein turnover, metabolic reactions and intracellular transport. Opportunity costs, which deal with the potential energy stored in previously synthesized macromolecules (biomass), are excluded from growth costs because this energy is not directly available as ATPs. (Note also that GTPs used by, for example, ribosomes or tubulin are treated as equivalent to ATPs.) The metabolic rate calculations agree with those reported previously by Chiyomaru and Takemoto and are thus validated by empirical data[20] (Extended Data Fig. 1). Thus, the amount of respiratory membrane needed by a cell depends on its

energy demands, cell division time and rate of ATP synthesis and area occupied by a single respiratory unit:

$$A_{\text{needed}} = \frac{R}{r}A_r = \frac{E_t/t_d}{r}A_r = \frac{(f_d c_g + t_d c_m)/t_d}{r}A_r \quad (2)$$

If the respiratory deficit is expressed as a function of cell volume (Supplementary Information), we obtain:

$$\text{Respiratory deficit} = \frac{\frac{(f_d \alpha V^{0.97}/t_d)+\beta V^{0.88}}{r}A_r}{f_{\max} S V^{2/3}} \quad (3)$$

where S is a factor that specifies the shape of a cell (for example, a perfect sphere or differently flattened spheroids; Supplementary Information). The parameters f_d, α, β, A_r and r are constants whose values have been previously determined[10,33–36] (Supplementary Table 1). The parameters t_d, $f_{\max}$ and S are constrained within biologically plausible ranges. For example, t_d is varied between 1 and 10 h, corresponding to the lower range of prokaryotic cell division times and the geometric mean of eukaryotic cell division times, respectively[10]. The $f_{\max}$ parameter varies between 8 and 18%, which are the largest possible fraction of respiratory membrane in *Escherichia coli*[36] and the membrane fraction at which roughly half of all membrane proteins are respiratory enzymes[37]. The shape factor, S, varies between 4.8 and 12.1, which correspond to a sphere and an oblate spheroid with a cell length to width ratio of 0.1 (Extended Data Fig. 2 and Supplementary Information).

To assess the maximum possible volume that mitochondrion-less cells can achieve, we calculated the deficit in respiratory membrane (equation (3)) across a wide range of cell volumes (Fig. 3a). Values of less than 1 for the respiratory deficit indicate that the cell has an excess of respiratory membrane, whereas values of greater than 1 indicate that the cell has insufficient respiratory membrane to sustain its own volume. Thus, respiratory deficit values of 1 point to the maximum volume that a mitochondrion-less cell can achieve. Our analyses show that spherical cells with a cell division time of 1 h and a maximum respiratory membrane fraction of 8% are surface area-constrained above a cell volume of about 1 μm³ (blue line, Fig. 3a). These parameter values and the estimated cell volume limit agree with what is seen for a small and fast-growing bacterium like *E. coli*[36]. If half of membrane proteins are respiratory enzymes (that is, a maximum respiratory membrane fraction of 18%), the largest volume that a cell can achieve is about 10 μm³ (black line, Fig. 3a). This large fraction of respiratory membrane would be possible if a bacterium devotes less of its surface area to other processes (for example, flagella or chemotactic receptors), or alternatively, if a bacterium develops intracytoplasmic membranes for respiratory processes[32]. A similar cell volume limit of 10 μm³ is achieved if the cell shape is changed to that of an oblate spheroid with a cell length to width ratio of 0.1 (dashed black line, Fig. 3a); some small and flattened flagellates like the eukaryote *Petalomonas minor*[38] or the phagocytic amoeboid prokaryote *Candidatus* Uabimicrobium amorphum have such cell body shapes[39]. The cell volume limit is raised even more, to about 500 μm³, if the cell division time is increased to 10 h (dotted black line, Fig. 3a). The combination of these 3 changes raises the cell volume limit to higher than 10^5 μm³ (red line, Fig. 3a). This might correspond to giant bacteria like *Thiomargarita* and *Epulopiscium* whose active cytoplasm is restricted to a thin enveloping sheet (that is, 2% of the whole cell volume[40]), have long cell division times (1–2 weeks[41]) and develop extensive intracytoplasmic membranes (for example, *Epulopiscium*[42]).

Cells with longer cell division times have lower metabolic rates and thus require fewer respiratory units (equation (2)). This is because longer cell division times allow cells to accumulate the same amount of ATP required for growth over longer time spans. Thus, cells with longer cell division times can achieve larger cell volumes. Our model predicts that a spherical cell with a maximum respiratory membrane fraction of 8% can, potentially, reach an upper volume of about 10^5 μm³ at a cell division time of roughly 10^3 h (Fig. 3b). However, the cumulative amount of ATP required for cell maintenance continues to increase throughout the cell's lifetime[10] and this eventually limits the maximum cell volume that is possible (Fig. 3b).

Our model allows us to predict the number of respiratory units and amount of respiratory membrane area required by a cell (equation (3)). The number of respiratory units predicted by our model follows closely, in both scaling exponent and intercept, the empirical data on the number of ATP synthases of cells previously reported by Lynch and Marinov[22] (Fig. 3c). Similarly, the amount of respiratory membrane required by eukaryotic cells also follows the data on mitochondrial inner membrane areas reported by Lynch and Marinov[22] after adjusting for the cross-sectional surface areas and stoichiometries of mitochondrial respiratory complexes[43] (Extended Data Fig. 3).

We also calculated the respiratory deficit for prokaryotes and eukaryotes whose cell volumes and cell division times were determined empirically[10]. For these calculations, we assumed a spherical cell body shape (S=4.8) and a maximum respiratory membrane fraction of 8%. These analyses showed that mitochondrion-less cells may have eukaryote-like volumes of up to 10^4 μm³ without a shortage of surface area for respiration (Fig. 3d). Therefore, many eukaryotes might be able, theoretically, to support their cell volumes by respiring at their cytoplasmic membranes (that is, without the need for internalizing respiratory membranes). Overall, our analyses reveal that longer cell division times (t_d), flattened or elongated cell shapes (S) and a larger allocation of surface area to respiration ($f_{\max}$) can, together or in isolation, allow cells to obtain larger volumes without the need for expanded respiratory membranes (for example, mitochondria). On the other hand, increasingly larger, rounder and faster-dividing cells have higher respiratory deficits (that is, larger than 1) and are thus dependent on an excess of respiratory membranes that cannot be fully accommodated on their cytoplasmic membranes.

The DNA investments of contrasting genomic architectures. Another claimed advantage of mitochondria is a drastic increase in energy per gene due to the asymmetric genomic architecture (or 'bioenergetic architecture' sensu Lane and Martin) that they allow for in eukaryotes[3,17–19]. Eukaryotes have both a single nuclear genome and many small and specialized mitochondrial genomes that scale with cell volume (that is, genomic asymmetry, Fig. 4). Prokaryotes, in contrast, only have a single genome type whose copy number scales with cell volume (that is, genomic symmetry, Fig. 4). Therefore, if a prokaryote were the size of an average eukaryote, the massive increase in gene number that accompanies polyploidy would keep its amount of energy per gene roughly equal to that of an average prokaryote despite having a much larger volume and energy available[3]. On the other hand, according to Lane and Martin's logic[3], eukaryotes have much more energy available per gene expressed as their nuclear genomes and gene numbers do not scale up with cell volume[3].

The concept of 'energy per gene expressed' has been criticized as having no evolutionary relevance[10,23,24]. This concept, as used by Lane and Martin, heavily penalizes large prokaryotes as their gene numbers increase with polyploidy. However, the amount of gene expression from each gene, irrespective of how many times the gene is duplicated, is proportional to cell volume. In other words, the relative cost of a gene, a more evolutionarily meaningful concept[10,22], is constant. This is because the energetic demands of cells strictly depend on their cell volumes, that is, prokaryotes and eukaryotes of the same volume require the same amount of energy. Thus, the

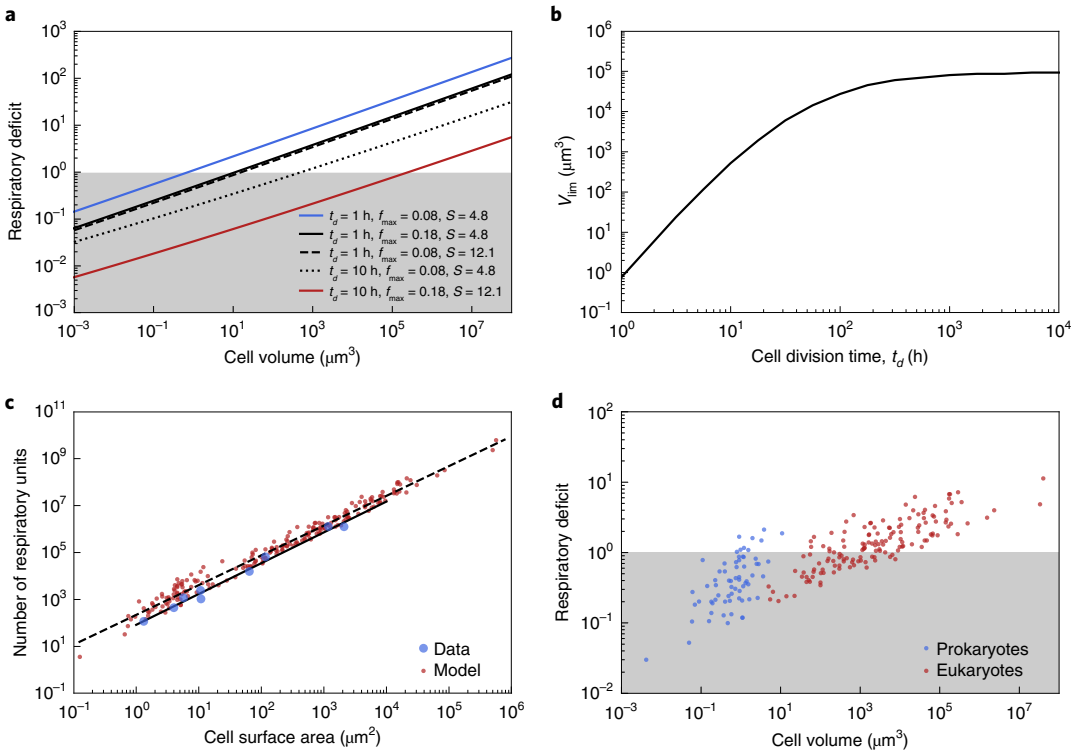


Fig. 3 | Factors that affect the volumes at which mitochondrion-less cells become constrained by their surface. a, Respiratory deficit as a function of cell volume. The blue line reflects cells that have a cell division time (t_d) of 1 h, a maximum membrane occupancy of respiratory proteins (f_{max}) of 8% and a shape factor (S) of 4.8. The black lines reflect cells for which a single parameter, either t_d, f_{max} or S, has been changed (inset). The red line reflects cells for which all parameters have been simultaneously changed. The intersection between each line (a defined set of parameters; inset) and a respiratory deficit of 1 determines the maximum volumes that cells can achieve. **b**, The surface area-limited cell volume, V_{lim}, plotted as a function of the cell division time. Fold deficit = 1, f_{max} = 8% and S = 4.8. **c**, The number of respiratory units (or ATP synthases) as a function of cell surface area. Empirically determined numbers of respiratory units (represented by ATP synthases) and cell surface areas, for prokaryotic and eukaryotic species, were obtained from Lynch and Marinov[22] (blue points). The number of respiratory units was calculated (red points) using $\left(\left(f_d \alpha V^{0.97}/t_d\right) + \beta V^{0.88}\right)/r$, with the cell volumes and cell division times for a range of prokaryotic and eukaryotic species obtained from Lynch and Marinov[10]. The solid line is a fit to the data: $y = 83x^{1.31}$. The dashed line is a fit to the model: $y = 221x^{1.27}$. **d**, Respiratory deficit calculated for individual prokaryotic and eukaryotic species whose cell volumes and cell division times were previously estimated[10]. f_{max} = 8% and S = 4.8 (spherical cells).

concept of energy per gene unfairly penalizes prokaryotes or any polyploid. Furthermore, the measurements of energy per gene previously performed by Lane and Martin[3] unfairly favour eukaryotes because gene copies due to mitochondrial genome polyploidy (which scale with cell volume) were ignored[3]. Because the concept of energy per gene is inappropriate, our approach below relies on estimating the cost of cellular features (that is, DNA synthesis) relative to the entire energy budget of a cell[10,22].

To test the hypothesis that the genomic architecture of eukaryotic cells provides an overwhelming advantage, we developed an explicit model that compares the energetic capacity of eukaryotes to prokaryotes. The goal was to isolate the genomic architecture of a cell from other confounding factors that also separate eukaryotes from prokaryotes. Because ATP demands depend on cell volume (and not complexity or gene number[10,20]), we considered the amount of ATP that remains ($1 - c_{DNA,euk}$ and $1 - c_{DNA,prok}$) after accounting for the relative cost of DNA that is associated with each genomic architecture ($c_{DNA,euk}$ and $c_{DNA,prok}$; Fig. 4 and equation (4)). This remaining amount of ATP is devoted to all cell processes other than DNA synthesis (for example, translation, transcription, lipid biosynthesis); the more ATP a cell invests in DNA, the less ATP there is to sustain other cellular processes. Thus, the ratio between the remaining ATP of a mitochondrion-bearing and mitochondrion-less cell provides a measure of the energetic advantage (>1), or disadvantage (<1) that mitochondrion-bearing cells might have. This can be expressed as:

$$\text{Energetic advantage } (\%) = \left(\frac{1 - c_{DNA,\,euk}}{1 - c_{DNA,\,prok}} - 1\right) \times 100 \quad (4)$$

To calculate the cost of DNA for a prokaryote (a mitochondrion-less cell), we considered a cell with only a single main genome type. In prokaryotes, the number of genomes increases proportionally with cell volume, as seen in *Synechococcus elongatus*[44] or in giant bacteria like *Epulopiscium*[45] (Fig. 4). The cause of this scaling might be the need to either bypass a diffusion constraint in the absence of active intracellular transport[46] or maintain genomes physically adjacent to respiratory membranes for efficient regulation[3,19]. We compiled data for several prokaryotes that showed that the cell volume per genome does not exceed 2 μm³ in several prokaryotes (V_{gserv} in equation (5); Supplementary Table 2). Our model thus assumes that if cell volume increases, the number of genomes must increase accordingly (see Supplementary Information for more details). The absolute total cost of DNA (in units of ATP) for a prokaryotic cell is the product of the amount of ATP required for synthesizing a single bp (101 ATPs; we used the cost value by Lynch and Marinov[10], which differs slightly from that of Mahmoudabadi et al.[33]), the length of a single genome in base pairs (L_{prok} in equation (5)) and the number of genomes. The number of genomes is the ratio between the total cell volume (V) and cell volume serviced by a single genome (V_{gserv}) (equation (5)). Note that, in contrast to the respiratory deficit calculations, we included the opportunity costs of synthesizing DNA

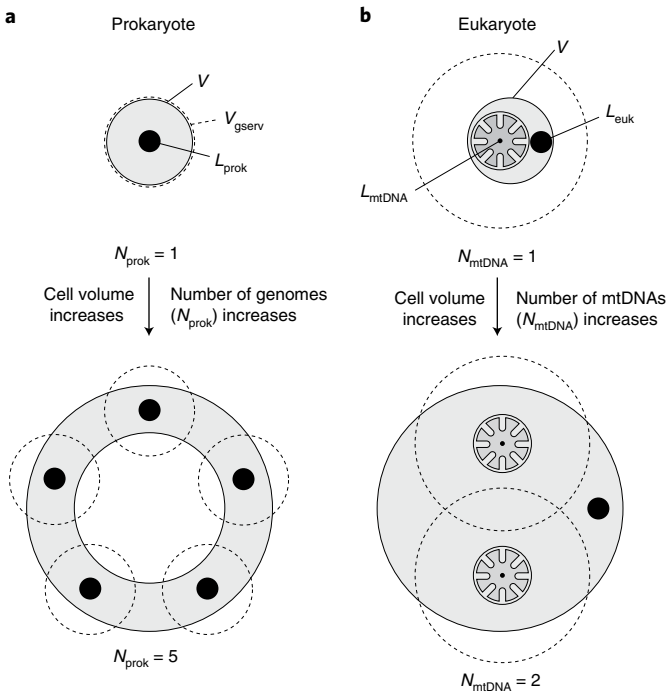


Fig. 4 | Graphical representation of contrasting genomic architectures in prokaryotes and eukaryotes. From equation (4); see the main text for an explanation of the parameters. **a**, The genomic symmetry of prokaryotes. We have represented a large prokaryotic cell as a shell of cytoplasm surrounding a large inert central space, as seen in giant bacteria like *Epulopiscium* and *Thiomargarita*. Even though this cell architecture is irrelevant for our calculations (equation (5)) since only the number of genomes is considered (filled black circles), prokaryotic cells have to scale up in cell volume with such an architecture to remain viable in the absence of an active intracellular transport network[46]. The total number of genomes N_{prok} is a function of the ratio of the cell volume and the volume controlled by a single genome (that is, V/V_{gserv}; see equation (5)). **b**, The genomic asymmetry of eukaryotes. The dashed circles hypothetically represent the amount of volume that can be energetically supported by mitochondria. Because of cristae (expanded internalized respiratory membranes), mitochondria can, in principle, energetically support large cytoplasmic volumes. The total number of mitochondrial genomes N_{mtDNA} is a function of the total volume of mitochondria and the number of mtDNA molecules per μm³ of mitochondrial volume ($n_{mtDNA}f_{mt}V$; equation (6) and main text)[47–49].

because we wanted to account for the complete, evolutionarily significant drain on cell resources. Finally, to obtain the relative cost of DNA for a prokaryotic cell, the absolute cost of the DNA was divided by the total ATP budget of the cell throughout its lifetime (equation (5)). This can be expressed as a function of cell volume:[10]

$$c_{DNA,prok} = \frac{101L_{prok}\left(\frac{V}{V_{gserv}}\right)}{\alpha V^{0.97} + t_d\beta V^{0.88}} \quad (5)$$

To calculate the cost of DNA for a eukaryote (a mitochondrion-bearing cell), we considered a cell with a single main (nuclear) genome and a variable number of mitochondrial genomes (mitochondrial DNA (mtDNA)). If the cell volume increases, the number of mitochondrial genomes increase proportionally but the main genome does not (see Supplementary Information for more details). The total number of mitochondrial genomes (N_{mtDNA} in Fig. 4) is the number of mtDNA molecules per μm³ of mitochondrial volume (n_{mtDNA})[47–49] multiplied by the total mitochondrial volume of the cell ($f_{mt}V$) (equation (6)). We compiled data that showed that the cell volume fraction occupied by mitochondria (f_{mt}) ranges from 1 to 20% across diverse eukaryotes; our calculations thus use the geometric mean of 4.4% (Fig. 5a and Supplementary Table 3). The number of mtDNA molecules per nucleoid (or per μm³ of mitochondrial volume, N_{mtDNA}), and the size of the mitochondrial genome (L_{mtDNA}) varied between 1 and 100 and 10 and 70 kilobase pairs (kbp)[50,51], respectively, with negligible effects on the results (Extended Data Fig. 4). Thus, the total cost of DNA comprises the cost of the main genome and of all mitochondrial genomes required to support the whole cell volume (equation (6)). The relative cost of DNA for a eukaryotic cell is calculated as shown above. If expressed as a function of cell volume, we have:

$$c_{DNA,euk} = \frac{101L_{euk} + 101L_{mtDNA}n_{mtDNA}f_{mt}V}{\alpha V^{0.97} + t_d\beta V^{0.88}}. \quad (6)$$

Our model (equations (4–6) and Supplementary Table 1) allows us to compare the contrasting genomic architectures of extant eukaryotes and prokaryotes across a range of cell volumes and genome sizes (Fig. 5). Note that for these calculations, we kept the (main) genome size for eukaryotes equal to that of prokaryotes (that is, $L_{prok} = L_{euk}$) because we are only interested in determining whether the genomic asymmetry of eukaryotes provides an advantage over prokaryotes. We also kept the cell division time (t_d) equal for prokaryotes and eukaryotes; varying t_d from 0 to 100 h did not have major effects on the calculated energetic advantage for eukaryotes (Extended Data Fig. 4); thus, it does not change our conclusions. The main conclusions from our calculations are as follows. First, prokaryotes invest a larger fraction of their ATP budget on DNA as their cells increase in volume; thus, they are left with less ATP for other processes such as gene expression (Fig. 5b). Second, the decrease in ATP available for other cell functions in prokaryotes is more pronounced as genome size increases (Fig. 5c). In contrast, eukaryotes suffer a negligible decline in their cellular ATP budget as their cell volume or main genome size increase (Fig. 5b,c). Third, eukaryotes have an energetic advantage (in terms of DNA cost savings) of less than approximately 200% for genome sizes of 10^6–10^8 bp and across a cell volume range of 8 orders of magnitude or 10^0–10^8 μm³ (Fig. 5d; a volume of 10^0 μm³ approximately corresponds to that of *E. coli*, whereas a cell volume of 10^8 μm³ is similar to that of a giant single-cell species like *Chaos carolinensis*[52]). At genome sizes of 10^6–10^7 bp, the energetic advantage of eukaryotes over prokaryotes is less than 10% across a similar range in cell volume (Fig. 5e). Fourth, a prokaryote with a genome size of 3×10^7 bp, which is characteristic of many single-cell eukaryotes (see below), would have an energetic disadvantage of approximately 20% relative to a eukaryote with the same genome size. Such a genome size could, in principle, accommodate 2×10^4 genes (assuming a mean gene length of 1,000 bp) while devoting about a third of its size (approximately 1×10^7 bp) to regulatory sequences. Fifth, prokaryotic genomes cannot get larger than approximately 3×10^8 because the cost of DNA would exceed the total ATP budget of the cell (at any cell volume). Eukaryotes, on the other hand, can achieve (main) genome sizes orders of magnitude larger as cell volume increases (Fig. 5f). If 2–10% of the ATP budget of the cell is devoted to DNA synthesis, prokaryotes can reach genomes of 6×10^6–3×10^7 bp in size (Fig. 5f).

To compare prokaryotes to eukaryotes at the onset of mitochondrial symbiosis, we assumed that the ancestral mitochondrial genome size was as large as that of an average prokaryote (that is, $L_{mtDNA} = L_{prok}$) and the mitochondrial volume fraction was much larger since mitochondrial function was not yet optimized for aerobic respiration (that is, $f_{mt} = 0.3$). However, this model (equation (6)) assumes that there was a dynamic cytoskeleton in place that allowed for active cytoplasmic transport and therefore

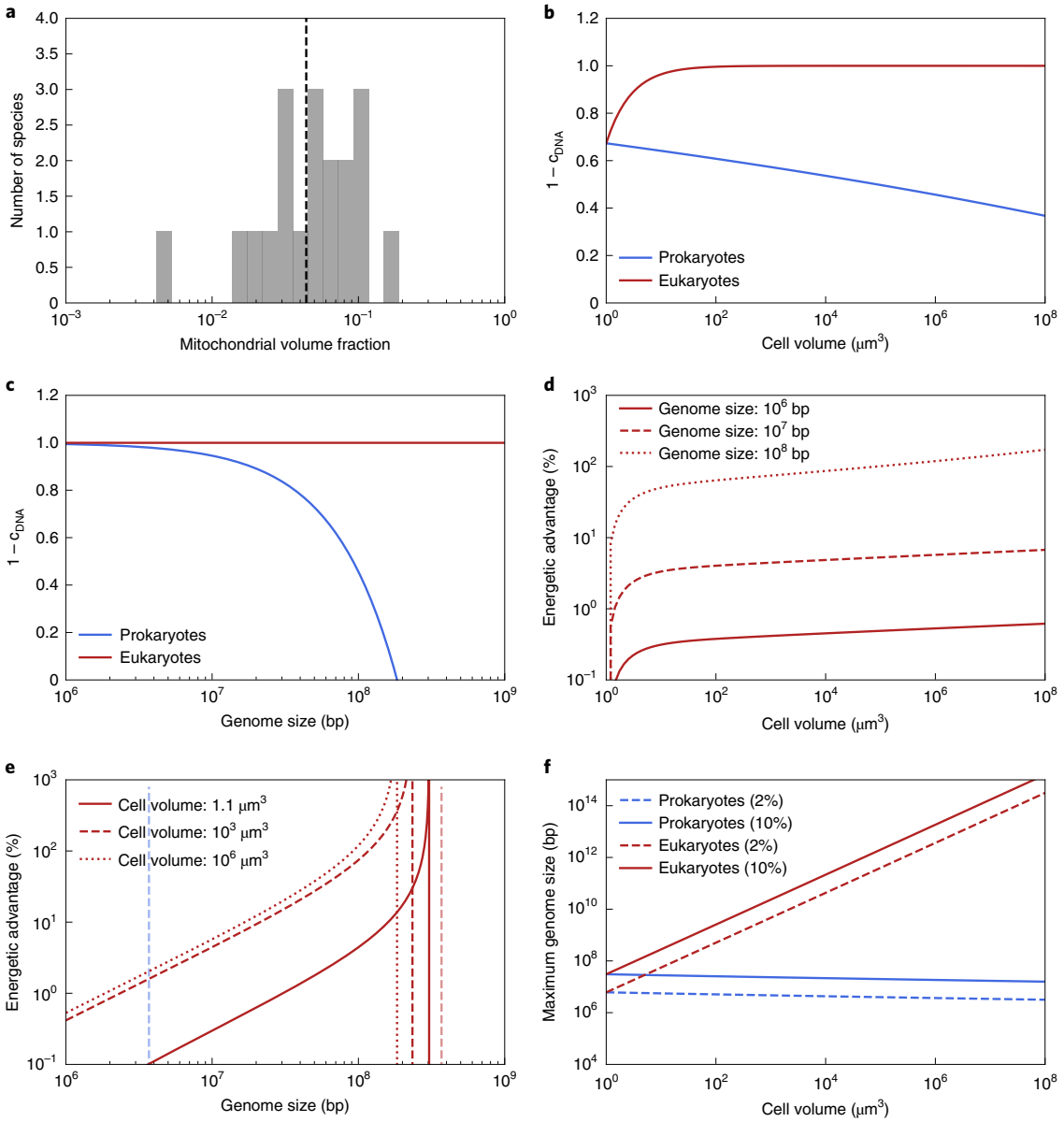


Fig. 5 | The impact of genomic architecture on energy allocation in cells. a, Distribution of mitochondrial volume fractions across a sample of phylogenetically diverse eukaryotes. The vertical dashed line indicates the geometric mean, 0.044 or 4.4% (Supplementary Table 3). **b**, Relative cell energy budget available for cellular features other than DNA as a function of cell volume. The plot was calculated with equations (5) and (6) and $L_{euk} = L_{prok} = 10^8$ bp, $L_{mtDNA} = 70$ kbp, $n_{mtDNA} = 10$, $t_d = 10$ h and $V_{gserv} = 1\,\mu m^3$ (these values were also used for **c**–**f**). **c**, Relative cell energy budget available for cellular features other than DNA as a function of genome size. The plot was calculated with equations (5) and (6) and $V = 10^6\,\mu m^3$. **d**, Energetic advantage of eukaryotes over prokaryotes (red lines) as a function of cell volume. The plot was calculated with equation (4) and three different genome sizes as shown in the inset. **e**, Energetic advantage of eukaryotes over prokaryotes (red lines) as a function of genome size. The vertical red lines denote the genome sizes at which the entire ATP budget of a prokaryote is devoted to DNA synthesis ($1 - c_{DNA,prok} = 0$). The plot was calculated with equation (4) and three different cell volumes as shown in the inset. The transparent blue and red dashed lines show the median genome sizes for prokaryotes and eukaryotes. **f**, Maximum (main) genome size as a function of cell volume for prokaryotes and eukaryotes, for maximum DNA investments of 2 and 10% of the entire cell energy budget. The plot was calculated with equations (5) and (6).

relinquished the need for genomes to scale with cell volume, as in modern eukaryotes (Discussion). This may best reflect a proto-eukaryote with a nucleus and cytoskeleton as envisioned by some mitochondrion-late scenarios. To model ancestral eukaryotes as conceived by mitochondrion-early scenarios, we assumed that the number of genomes scaled with cell volume (that is, as in modern prokaryotes) and that the ancestral mitochondrial genome was much larger and equivalent to that of an average prokaryote (that is, $L_{mtDNA} = L_{prok}$). To do this, equation (6) was modified in such a way that the number of genomes of ancestral eukaryotes is a function of the cell volume not occupied by (pre-)mitochondria:

$$c_{DNA,proto-euk} = \frac{101L_{euk}\left(\frac{(1-f_{mt})V}{V_{gserv}}\right) + 101L_{mtDNA}n_{mtDNA}f_{mt}V}{\alpha V^{0.97} + t_d\beta V^{0.88}} \quad (7)$$

The analyses show that a modern eukaryotic genomic architecture (which would also have been present in proto-eukaryotes as

postulated by mitochondrion-late scenarios) almost always gives an energetic advantage to eukaryotes over prokaryotes (Extended Data Fig. 5a,b). This is true even at much larger mitochondrial genome sizes and volume fractions that must have been present at the onset of mitochondrial symbiosis (Extended Data Fig. 5c,d). On the other hand, if the ancestral host cell that took up the (pre-)mitochondrial symbiont had a genomic architecture equivalent to that of prokaryotes (as implied by mitochondrion-early scenarios), the first eukaryotes derived no energetic advantage relative to prokaryotes (Extended Data Fig. 5e). Both the first eukaryotes and prokaryotes invested the same amount of energy in DNA synthesis. Further reductions in mitochondrial genome size would have provided larger DNA cost savings to early eukaryotes. If ancestral mitochondria were slightly more polyploid (that is, had a higher genome copy number per μm^3; $n_{mtDNA}=3$), like some modern intracellular symbionts that increase ploidy to physiologically support their host (for example, see refs. [53,54]), prokaryotes might have even had an energetic advantage and invested less ATP on DNA synthesis compared to ancestral eukaryotes (Extended Data Fig. 5f).

The costs and benefits energy-producing symbionts. Mitochondria, or respiring symbionts, expand the maximum volume that cells can achieve (Fig. 3), although they are not required for many combinations of cell volumes and division times observed among modern eukaryotes (Fig. 3d). However, symbionts and organelles also drain resources from their host cells since they have both growth and maintenance costs. We investigated the maximum volume afforded by, and costs associated with, a population of respiring symbionts in a host cell. To do this, we conceived a model that envisages an aerobically respiring host cell with a population of one or more small spherical symbionts that respire aerobically. These respiring symbionts rely on a slight excess of surface area devoted to aerobic respiration ($f_{max}=0.1$ instead of $f_{max}=0.08$ as above, which leads to approximately 13% of ATP overproduction by a symbiont of 1 μm^3 in volume) to support themselves and partially contribute to the energy demands of their host cell; this would have been true at the early stages of mitochondrial symbiosis when symbiont function was not yet evolutionarily optimized. The amount of resources (ATP units) diverted from the host cell are directly proportional to the energy requirements dictated by the volume of each symbiont. The advantage that the respiring symbionts confer to their host is quantified as an increase in cell volume and is ultimately a function of both the excess ATP they provide to the host cell and the amount of ATP consumed to support their own volume (where the former includes only direct costs and the latter both direct and opportunity costs).

To find out the amount of cell volume that is supported by the symbionts' membranes (V_{supp_sym}), we can first express the amount of respiratory membrane supplied by the symbionts as a function of the number and volume of symbionts and then as a function of the number of respiratory units harboured by the symbiont population (that is, the ratio between the lifetime metabolic rate of the symbiont-supported volume and the rate of a single ATP synthase). If these two expressions are equated, the resulting equation:

$$f_{max\_sym}N_{sym}S_{sph}V_{sym}^{2/3}=\frac{\left(f_d\alpha V_{supp\_sym}^{0.97}/t_d\right)+\beta V_{supp\_sym}^{0.88}}{r}A_r \quad (8)$$

can be solved numerically to obtain the values of V_{supp_sym} for different cell division times and number of symbionts. In this example, f_{max_sym} is the maximum respiratory membrane fraction used by the symbiont (=0.1), N_{sym} is the number of intracellular symbionts (which ranges from 1 to 4×10^4), S_{sph} is the shape factor for the volume of a sphere and V_{sym} is the volume occupied by a single symbiont (equal to 0.25–1 μm^3). As before, the parameter values for f_d, α, β, A_r and r are constants (Supplementary Table 1).

To calculate the amount of cell volume that is supported by the host membrane (V_{supp_host}), we followed a logic similar to that used above to determine V_{supp_sym}. We first expressed the amount of respiratory membrane supplied by the host as a function of the host cell volume and then as a function of the number of respiratory units harboured by the host cell (that is, the ratio between the lifetime metabolic rate of the host-supported volume and the rate of a single ATP synthase). If these two expressions are equated, we obtain the following expression:

$$f_{max}S_{sph}V_{host}^{2/3}=\frac{\left(f_d\alpha V_{supp\_host}^{0.97}/t_d\right)+\beta V_{supp\_host}^{0.88}}{r}A_r \quad (9)$$

which can be combined with $V_{host}=V_{supp\_sym}+V_{supp\_host}$ (Supplementary Information) and then rearranged to obtain V_{supp_host}, the cell volume supported by host respiration:

$$V_{supp\_sym}+V_{supp\_host}=\left(\frac{\left(f_d\alpha V_{supp\_host}^{0.97}/t_d\right)+\beta V_{supp\_host}^{0.88}}{rf_{max}S_{sph}}A_r\right)^{2/3} \quad (10)$$

Adding V_{supp_sym} and V_{supp_host} yields V_{host}, which is the largest volume for the host cell that can be supported by the population of respiring symbionts. The benefit of a much larger possible volume for the host cell can be compared to the relative cost of maintaining the symbiont population, c_{rel_sym}. This cost is approximated by:

$$c_{rel,sym}=\frac{V_{sym}N_{sym}}{V_{host}} \quad (11)$$

The analyses reveal that the larger the population of respiring symbionts, the larger the increase in cell volume that is possible (Fig. 6). The increase in volume is large for fast-dividing cells, although the cost of maintaining their symbiont population is also very high. On the other hand, slow-dividing cells gain a smaller increase in volume but the energetic cost of their symbiont populations is also much lower. For example, a small and fast-growing cell (t_d=h) that harbours respiring symbionts derives a large advantage that allows its volume to increase substantially; however, this also comes at a steep cost of >30% (Fig. 6). A larger and slower-dividing cell (for example, $t_d=8$ h) has a relative symbiont cost of up to 2 orders of magnitude lower (Fig. 6). If two cells have the same number of symbionts, the slower-dividing cell will achieve a larger volume and also has a lower symbiont cost. Thus, if a larger and slower-dividing cell took up respiring symbionts, it would have been able to maintain or increase its volume while having a comparatively low symbiont cost. A lower symbiont cost leaves more energetic resources, that is, ATP, for the rest of the cell volume, or leaves a larger volume fraction for other cell functions. These analyses may provide some insights into the nature of the ancestral host cell that took up symbionts at the origin of eukaryotes (Discussion).

Discussion

The role of mitochondrial energetics in the origin and diversification of eukaryotes is highly contested[3,10,22,26,55]. As an attempt to resolve this debate, we investigated the respiratory deficit of mitochondrion-less cells and the maximum cell volume that can be supported by respiration at the cell surface. We showed that the maximum volume a cell can attain is dependent on at least three major factors: cell body shape, cell division time and maximum respiratory membrane fraction. A combination of biologically plausible values for these factors may allow mitochondrion-less cells to achieve volumes of up to 10^3–10^5 μm^3 without a deficit in surface

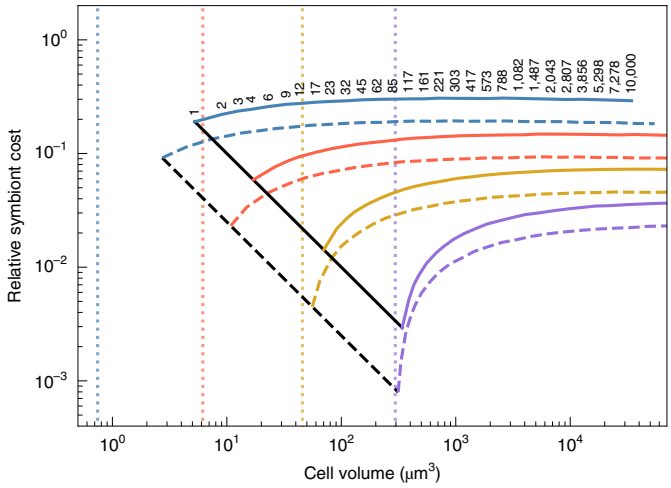


Fig. 6 | Costs and benefits of harbouring ATP-exporting respiring symbionts. Relationship between maximum cell volume and symbiont population cost for different symbiont population sizes at several cell division times. The plot was generated from equations (8–11). Solid lines, $V_{sym} = 1\,\mu m^3$; dashed lines, $V_{sym} = 0.25\,\mu m^3$. The vertical dotted lines show the host cell volumes in the absence of a symbiont population. The colours indicate the cell division time: 1 h (blue); 2 h (orange); 4 h (gold); and 8 h (purple). For all the coloured lines, the cell division time, t_d, is constant and the number of symbionts, N_{sym}, is varied. The numbers above the solid blue line indicate the number of symbionts. For the black lines, $N_{sym} = 1$ and t_d varied from 1 to 8 h.

area devoted to respiration (Fig. 3). Furthermore, we investigated the energetic consequences of the contrasting genomic architectures of mitochondrion-less and mitochondrion-bearing cells. Our results show that the asymmetrical genomic architecture of mitochondrion-bearing cells provides slight energetic savings in DNA costs relative to mitochondrion-less cells across a wide range of cell volumes and genome sizes (Fig. 5). The model further predicts that mitochondrion-less cells can achieve a genome size of 3×10^7 bp if they devote 10% of their ATP budget to DNA synthesis, at an energetic disadvantage of 20% (or 1.2-fold) (Figs. 5e,f).

The upper cell volumes and genome sizes of mitochondrion-less cells can be predicted based on energetic considerations, as done in this study. However, evolutionary success depends not only on the energetic capacity of a cell to sustain its own features but also on the selective or ecological advantages conferred by such features. For example, a cell that has an energetic disadvantage by investing a large proportion of energy in DNA (and thus less in ribosome biogenesis or growth) but has a feature that confers a large ecological advantage (for example, phagocytosis or antibiotic secretion) may otherwise outcompete cells that invest less in DNA but lack such a feature. Similarly, the reproductive disadvantage that may accompany longer cell division times in larger cells may be overcome by ecological specialization to avoid competition. This is the sentiment behind some of the criticisms of the energetic hypothesis for the origin of eukaryotic complexity previously raised by others[24,25]. Moreover, even features that do not confer selective advantages large enough to offset the selective disadvantages associated with their energetic costs can passively emerge and be maintained in evolution. This is more likely in species with larger cells that have longer cell division times and smaller effective population sizes. In such species, the relative energetic cost of cellular features is lower and the power of random genetic drift is stronger[10].

We have shown that the genomic architecture of modern eukaryotes is advantageous in comparison to that of prokaryotes. However, this energetic advantage never exceeds 200% (or threefold) across a vast cell volume range of 10^1–$10^8\,\mu m^3$ (Fig. 5b). These results stand in sharp contrast with the claim that "[an average] eukaryotic gene commands some 200,000-fold more energy than a prokaryotic gene"[3]. The discrepancy resides not only in the inappropriateness of the energy per gene concept (see above) but also in that previous analyses compared idealized averages for modern eukaryotes and prokaryotes. Such averages, however, differ drastically in cell volumes. Because the energy demands of cells (that is, ATP requirements and maximum metabolic rates[10,20]) scale continuously and nearly linearly with volume, and prokaryotes and eukaryotes overlap across this continuum (Fig. 2a), such comparisons between rough averages are misleading.

The maximum advantage of threefold for eukaryotes found in this study stems from the comparison of two considerably different types of cells: mitochondrion-less and mitochondrion-bearing cells. Whereas the former represents an average prokaryote, the latter arguably represents a derived (proto-)eukaryote with a highly reduced mitochondrial genome and dynamic cytoskeleton. This is because our model considers a mitochondrial genome that is less than 7% the size of the main genome (that is, ≤70 kbp, which is equivalent to that of jakobids and last eukaryote common ancestor) and assumes a main nuclear genome whose copy number does not increase with a larger cell volume. Such a reduced mitochondrial genome could only have evolved after the invention of a protein import machinery that sped up gene transfer to the nucleus or main genome (that is, by allowing the import of transferred gene products back into mitochondria). In addition, only the presence of active intracellular transport (that is, a dynamic cytoskeleton and motor proteins that bypassed diffusion constraints) would have allowed the nuclear or main genome not to scale up with cell volumes (unlike in prokaryotes[45,46]). Thus, a great degree of evolutionary change (and time) separates the two types of cells compared in this study. This suggests that the energetic advantages between immediate ancestor and descendant populations of (proto-)eukaryotes were necessarily much smaller than the threefold energetic advantage for mitochondrion-bearing cells found in this study. Indeed, using a model and a set of parameters that best reflect the genomic architecture of ancestral eukaryotes at the onset of mitochondrial symbiosis (as viewed by mitochondrion-early scenarios) shows that these early eukaryotes did not derive any energetic advantage relative to prokaryotes. Eukaryotes only gained an energetic advantage after their mitochondrial genomes reduced considerably and their main genome copy number was no longer required to scale with cell volume (Extended Data Fig. 5). Moreover, if the ancestral eukaryote that served as host for the pre-mitochondrial symbionts possessed a dynamic cytoskeleton (as predicted by mitochondrion-late scenarios), the energetic costs associated with DNA synthesis would have been slightly lower compared to those of a cytoskeleton-less prokaryote (Extended Data Fig. 5).

Many cellular features other than mitochondria separate extant eukaryotes from prokaryotes. Among these, the cytoskeleton and endomembranes provided major evolutionary advantages to early (proto-)eukaryotes by synergistically enabling phagocytosis and intracellular digestion. These adaptations massively increase the surface area by which nutrients can be taken up by cells (that is, expanded nutritional membranes such as food and digestive vacuoles) and might compensate for the absence of higher efficiency in energy harnessing (that is, aerobic respiration in mitochondria). Phagotrophs also have an ecological advantage by preying on their bacterial competitors. Furthermore, a dynamic cytoskeleton with motors allowed for active intracellular transport, thereby overcoming the diffusion constraints that burden prokaryotes. These advantages may have allowed early (proto-)eukaryotes to achieve larger cell volumes entirely composed of active cytoplasm (as opposed to most larger prokaryotes) in the absence of mitochondria. The recent

discovery of a phagocytic prokaryote with a dynamic cytoskeleton and sizes of up to 10 µm in length[39] suggest that the evolution of a larger cell volume is possible in the absence of mitochondria. As shown above, these larger cell volumes and their accompanying longer cell division times are also those at which a host cell would benefit the most from harbouring a population of small respiring symbionts (Fig. 6).

Comparative genomic analyses have estimated that the last eukaryote common ancestor had 4,431 gene domains[56], approximately 4,137–5,938 gene families[57–59] or 7,447–21,840 genes (mean = 12,753)[11]. This inferred number of genes can be accommodated by a genome of approximately 20–50 Mbp in size that also devotes more than a third of its size to regulatory and other non-coding DNA. Because nuclear DNA amounts scale strongly with cell volume in eukaryotes as the power law $V = 1025.4 \times \mathrm{DNA}^{0.97}$ (where V is in µm³ and DNA in picograms)[60–62], a haploid last eukaryote common ancestor with such a genome size may have had a cell volume of approximately 23–57 µm³ (that is, the volume of a spherical cell of 3.5–4.8 µm in diameter). Indeed, these genome and cell sizes are similar to those of small heterotrophic nanoflagellates such as jakobids and malawimonads, which also have the most ancestral-like mitochondrial genomes known[50,63]. The gene number, genome size and cell volume inferred for the last eukaryote common ancestor fall within or are close to the modern prokaryote–eukaryote overlap (that is, 10^0–10^2 µm³, 10^6–10^7 bp and 4,000–13,000 genes) and also encompass the cell volumes and genome sizes at which prokaryotes may not face a shortage of surface area (Fig. 3) or a considerable energetic disadvantage due to increasing DNA costs (Fig. 5). Thus, the prokaryote–eukaryote transition may have happened under these conditions.

Even though our analyses suggest that mitochondrion-less cells may achieve relatively large volumes and genome sizes under certain conditions, they also point at constraints that these simpler cells inevitably face at even larger volumes or genome sizes. Because the amount of respiratory membrane needed (that is, the number of ATP synthases) scales superlinearly with total surface area[22] (Fig. 3c and Supplementary Table 3), prokaryotes may experience a shortage of respiratory membrane area at larger cell volumes (as long as their internal volumes are active unlike in giant bacteria). Eukaryotes, on the other hand, can maintain such a superlinear scaling and reach much larger cell volumes by internalizing respiratory membranes in mitochondria. In other words, mitochondria allow energy supply to continuously match energy demand at increasingly larger volumes. Barring non-energetic constraints (for example, DNA replication times), mitochondria may also allow eukaryotes to have shorter cell division times and rounder (or less flattened) cell body shapes than mitochondrion-less cells (for example, prokaryotes) at comparable cell volumes. Furthermore, as genome size increases, prokaryotes divest more and more of their ATP budgets to DNA synthesis due to their genomic symmetry. Therefore, the energetic advantage of eukaryotes over prokaryotes increases with larger genome sizes. The maximum genome size that prokaryotes can theoretically achieve is 3×10^8 bp if the entire ATP budget were devoted to DNA synthesis or up to 3×10^7 bp at 10% of the ATP budget. In contrast, eukaryotes can drastically expand their genomes as their cell volumes (and ATP budgets) grow larger because of their genomic asymmetry. These theoretical predictions are consistent with the constraints on prokaryotes suggested by the cell volume and genome size distributions (Fig. 2) and are at odds with the conclusions of Lynch and Marinov[10,22].

Conclusions

It has been claimed that an energy gap underlies the large differences in size and complexity between eukaryotic and prokaryotic cells. The proponents of this view further hold that the origin of mitochondria was a prerequisite for simple prokaryotic cells to bridge such a gap and evolve into complex eukaryotic cells. Based on energetic considerations, we have shown that prokaryotes can theoretically achieve eukaryote-like cell volumes and genome sizes. These findings are consistent with the modern prokaryote–eukaryote overlap in cell volumes and genome sizes. Because the last eukaryotic common ancestor was probably a small heterotrophic flagellate similar to a modern jakobid or malawimonad eukaryote, we suggest that the prokaryote–eukaryote transition did not necessarily require an expansion of respiratory membranes or the savings in DNA costs that mitochondria can provide. We also argue that the selective advantages conferred by mitochondria did not represent a quantum leap in energy supply (or 'bioenergetic jump'[3]) at the origin of eukaryotes and were, in principle, not different from those provided by other eukaryotic innovations, such as a dynamic cytoskeleton or an endomembrane system. Mitochondria, however, were much more important for increasingly larger and faster-dividing eukaryotic cells and may have thus allowed eukaryotes to successfully diversify and occupy new adaptive zones throughout their evolutionary history.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

All data are available in the Supplementary Information and Source Data Fig. 2. Source data are provided with this paper.

Received: 29 October 2021; Accepted: 20 June 2022;
Published online: 1 August 2022

References

1. Cavalier-Smith, T. The neomuran revolution and phagotrophic origin of eukaryotes and cilia in the light of intracellular coevolution and a revised tree of life. *Cold Spring Harb. Perspect. Biol.* **6**, a016006 (2014).
2. Stanier, R. Y., Douderoff, M. & Adelberg, E. *The Microbial World* (Prentice-Hall, 1963).
3. Lane, N. & Martin, W. The energetics of genome complexity. *Nature* **467**, 929–934 (2010).
4. Martin, W. & Müller, M. The hydrogen hypothesis for the first eukaryote. *Nature* **392**, 37–41 (1998).
5. Cavalier-Smith, T. Predation and eukaryote cell origins: a coevolutionary perspective. *Int. J. Biochem. Cell Biol.* **41**, 307–322 (2009).
6. López-García, P. & Moreira, D. The Syntrophy hypothesis for the origin of eukaryotes revisited. *Nat. Microbiol.* **5**, 655–667 (2020).
7. Baum, D. A. & Baum, B. An inside-out origin for the eukaryotic cell. *BMC Biol.* **12**, 76 (2014).
8. Sagan, L. On the origin of mitosing cells. *J. Theor. Biol.* **14**, 255–274 (1967).
9. Stanier, R. Y. in Charles HP & Knight BCJG. *Proc. Symposium of the Society for General Microbiology* 20 1–38 (Microbiology Society, 1970).
10. Lynch, M. & Marinov, G. K. The bioenergetic costs of a gene. *Proc. Natl Acad. Sci. USA* **112**, 15690–15695 (2015).
11. Vosseberg, J. et al. Timing the origin of eukaryotic cellular complexity with ancient duplications. *Nat. Ecol. Evol.* **5**, 92–100 (2021).
12. Pittis, A. A. & Gabaldón, T. Late acquisition of mitochondria by a host with chimaeric prokaryotic ancestry. *Nature* **531**, 101–104 (2016).
13. Zachar, I. & Szathmáry, E. Breath-giving cooperation: critical review of origin of mitochondria hypotheses. *Biol. Direct* **12**, 19 (2017).
14. Vellai, T., Takács, K. & Vida, G. A new aspect to the origin and evolution of eukaryotes. *J. Mol. Evol.* **46**, 499–507 (1998).
15. Vellai, T. & Vida, G. The origin of eukaryotes: the difference between prokaryotic and eukaryotic cells. *Proc. Biol. Sci.* **266**, 1571–1577 (1999).
16. Lane, N. *Power, Sex, Suicide: Mitochondria and the Meaning of Life* (Oxford Univ. Press, 2006).
17. Lane, N. Energetics and genetics across the prokaryote-eukaryote divide. *Biol. Direct* **6**, 35 (2011).
18. Lane, N. Bioenergetic constraints on the evolution of complex life. *Cold Spring Harb. Perspect. Biol.* **6**, a015982 (2014).
19. Lane, N. How energy flow shapes cell evolution. *Curr. Biol.* **30**, R471–R476 (2020).
20. Chiyomaru, K. & Takemoto, K. Revisiting the hypothesis of an energetic barrier to genome complexity between eukaryotes and prokaryotes. *R. Soc. Open Sci.* **7**, 191859 (2020).

21. Booth, A. & Doolittle, W. F. Eukaryogenesis, how special really? *Proc. Natl Acad. Sci. USA* **112**, 10278–10285 (2015).
22. Lynch, M. & Marinov, G. K. Membranes, energetics, and evolution across the prokaryote-eukaryote divide. *eLife* **6**, e20437 (2017).
23. Szathmáry, E. Toward major evolutionary transitions theory 2.0. *Proc. Natl Acad. Sci. USA* **112**, 10104–10111 (2015).
24. Cavalier-Smith, T. & Chao, E. E.-Y. Multidomain ribosomal protein trees and the planctobacterial origin of neomura (eukaryotes, archaebacteria). *Protoplasma* **257**, 621–753 (2020).
25. Hampl, V., Čepička, I. & Eliáš, M. Was the mitochondrion necessary to start eukaryogenesis? *Trends Microbiol.* **27**, 96–104 (2019).
26. Lynch, M. & Marinov, G. K. Reply to Lane and Martin: mitochondria do not boost the bioenergetic capacity of eukaryotic cells. *Proc. Natl Acad. Sci. USA* **113**, E667–E668 (2016).
27. Volland, J.-M. et al. A centimeter-long bacterium with DNA contained in metabolically active, membrane-bound organelles. *Science* **376**, 1453–1458 (2022).
28. Gray, M. W. et al. The draft nuclear genome sequence and predicted mitochondrial proteome of *Andalucia godoyi*, a protist with the most gene-rich and bacteria-like mitochondrial genome. *BMC Biol.* **18**, 22 (2020).
29. Huxley, J. S. *Problems of Relative Growth* (Methuen & Co. Ltd., 1935).
30. Thompson, D. W. *On Growth and Form* (Cambridge Univ. Press, 1992).
31. Snell, O. Die Abhängigkeit des Hirngewichtes von dem Körpergewicht und den geistigen Fähigkeiten. *Arch. Psychiatr. Nervenkr.* **23**, 436–446 (1892).
32. Muñoz-Gómez, S. A., Wideman, J. G., Roger, A. J. & Slamovits, C. H. The origin of mitochondrial cristae from Alphaproteobacteria. *Mol. Biol. Evol.* **34**, 943–956 (2017).
33. Mahmoudabadi, G., Phillips, R., Lynch, M. & Milo, R. Defining the energetic costs of cellular structures. Preprint at *bioRxiv* https://doi.org/10.1101/666040 (2019).
34. Etzold, C., Deckers-Hebestreit, G. & Altendorf, K. Turnover number of *Escherichia coli* F0F1 ATP synthase for ATP synthesis in membrane vesicles. *Eur. J. Biochem.* **243**, 336–343 (1997).
35. Valgepea, K., Adamberg, K., Seiman, A. & Vilu, R. *Escherichia coli* achieves faster growth by increasing catalytic and translation rates of proteins. *Mol. Biosyst.* **9**, 2344–2358 (2013).
36. Szenk, M., Dill, K. A. & de Graff, A. M. R. Why do fast-growing bacteria enter overflow metabolism? Testing the membrane real estate hypothesis. *Cell Syst.* **5**, 95–104 (2017).
37. Lindén, M., Sens, P. & Phillips, R. Entropic tension in crowded membranes. *PLoS Comput. Biol.* **8**, e1002431 (2012).
38. Larsen, J. & Patterson, D. J. Some flagellates (Protista) from tropical marine sediments. *J. Nat. Hist.* **24**, 801–937 (1990).
39. Shiratori, T., Suzuki, S., Kakizawa, Y. & Ishida, K. Phagocytosis-like cell engulfment by a planctomycete bacterium. *Nat. Commun.* **10**, 5529 (2019).
40. Schulz, H. N. & Jorgensen, B. B. Big bacteria. *Annu. Rev. Microbiol.* **55**, 105–137 (2001).
41. Schulz, H. N. et al. Dense populations of a giant sulfur bacterium in Namibian shelf sediments. *Science* **284**, 493–495 (1999).
42. Clements, K. D. & Bullivant, S. An unusual symbiont from the gut of surgeonfishes may be the largest known prokaryote. *J. Bacteriol.* **173**, 5359–5362 (1991).
43. Schlame, M. Protein crowding in the inner mitochondrial membrane. *Biochim. Biophys. Acta Bioenerg.* **1862**, 148305 (2021).
44. Ohbayashi, R. et al. Coordination of polyploid chromosome replication with cell size and growth in a cyanobacterium. *mBio* **10**, e00510-19 (2019).
45. Mendell, J. E., Clements, K. D., Choat, J. H. & Angert, E. R. Extreme polyploidy in a large bacterium. *Proc. Natl Acad. Sci. USA* **105**, 6730–6734 (2008).
46. Ionescu, D., & Bizic, M. (2020). Giant bacteria in *eLS*. Chichester (United Kingdom): John Wiley & Sons, Ltd, 1-10. https://doi.org/10.1002/9780470015902.a0020371.pub2
47. Jajoo, R. et al. Accurate concentration control of mitochondria and nucleoids. *Science* **351**, 169–172 (2016).
48. Kukat, C. et al. Super-resolution microscopy reveals that mammalian mitochondrial nucleoids have a uniform size and frequently contain a single copy of mtDNA. *Proc. Natl Acad. Sci. USA* **108**, 13534–13539 (2011).
49. Ilamathi, H. S. et al. Mitochondrial fission is required for proper nucleoid distribution within mitochondrial networks. Preprint at *bioRxiv* https://doi.org/10.1101/2021.03.17.435804 (2021).
50. Roger, A. J., Muñoz-Gómez, S. A. & Kamikawa, R. The origin and diversification of mitochondria. *Curr. Biol.* **27**, R1177–R1192 (2017).
51. Janouškovec, J. et al. A new lineage of eukaryotes illuminates early mitochondrial genome reduction. *Curr. Biol.* **27**, 3717–3724.e5 (2017).
52. Fenchel, T. & Finlay, B. J. Respiration rates in heterotrophic, free-living protozoa. *Microb. Ecol.* **9**, 99–122 (1983).
53. Komaki, K. & Ishikawa, H. Genomic copy number of intracellular bacterial symbionts of aphids varies in response to developmental stage and morph of their host. *Insect Biochem. Mol. Biol.* **30**, 253–258 (2000).
54. Mergaert, P. et al. Eukaryotic control on bacterial cell cycle and differentiation in the *Rhizobium*-legume symbiosis. *Proc. Natl Acad. Sci. USA* **103**, 5230–5235 (2006).
55. Lane, N. & Martin, W. F. Mitochondria, complexity, and evolutionary deficit spending. *Proc. Natl Acad. Sci. USA* **113**, E666 (2016).
56. Zmasek, C. M. & Godzik, A. Strong functional patterns in the evolution of eukaryotic genomes revealed by the reconstruction of ancestral protein domain repertoires. *Genome Biol.* **12**, R4 (2011).
57. Makarova, K. S., Wolf, Y. I., Mekhedov, S. L., Mirkin, B. G. & Koonin, E. V. Ancestral paralogs and pseudoparalogs and their role in the emergence of the eukaryotic cell. *Nucleic Acids Res.* **33**, 4626–4638 (2005).
58. Fritz-Laylin, L. K. et al. The genome of *Naegleria gruberi* illuminates early eukaryotic versatility. *Cell* **140**, 631–642 (2010).
59. Newman, D., Whelan, F. J., Moore, M., Rusilowicz, M. & McInerney, J. O. Reconstructing and analysing the genome of the last eukaryote common ancestor to better understand the transition from FECA to LECA. Preprint at *bioRxiv* https://doi.org/10.1101/538264 (2019).
60. Shuter, B. J., Thomas, J. E., Taylor, W. D. & Zimmerman, A. M. Phenotypic correlates of genomic DNA content in unicellular eukaryotes and other cells. *Am. Nat.* **122**, 26–44 (1983).
61. Cavalier-Smith, T. & Beaton, M. J. The skeletal function of non-genic nuclear DNA: new evidence from ancient cell chimaeras. *Genetica* **106**, 3–13 (1999).
62. Cavalier-Smith, T. Economy, speed and size matter: evolutionary forces driving nuclear genome miniaturization and expansion. *Ann. Bot.* **95**, 147–175 (2005).
63. Burger, G., Gray, M. W., Forget, L. & Lang, B. F. Strikingly bacteria-like and gene-rich mitochondrial genomes throughout jakobid protists. *Genome Biol. Evol.* **5**, 418–438 (2013).

Acknowledgements

We thank M. Lynch for comments on an early draft of this manuscript. S.A.M.-G. is supported by an EMBO Postdoctoral Fellowship (ALTF 21-2020). P.E.S. is supported by the Moore–Simons Project on the Origin of the Eukaryotic Cell, Simons Foundation 735927 (https://doi.org/10.46714/735927), the National Institutes of Health, R35-GM122566-01 and the National Science Foundation, DBI-2119963.

Author contributions

P.S. and S.A.M.-G. conceptualized the study and devised the methodology. They carried out the validation, formal analysis and investigation, curated the data, wrote the original manuscript draft, reviewed and edited it, and visualized the data.

Competing interests

The authors declare no competing interests.

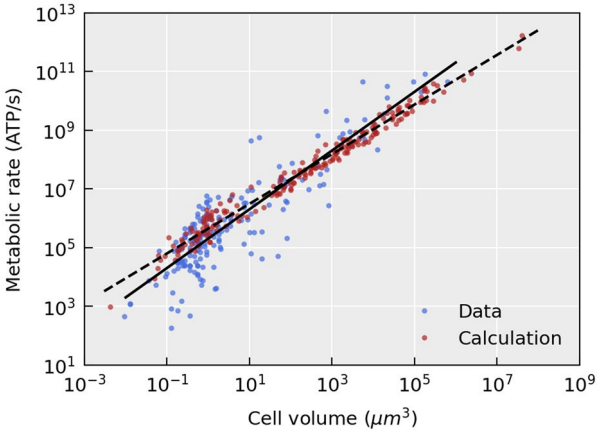
Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41559-022-01833-9.

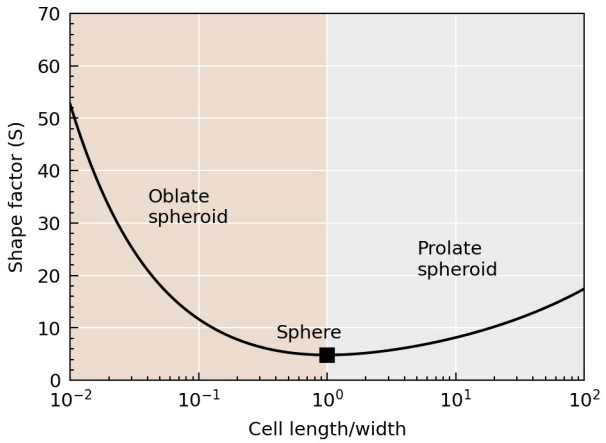
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41559-022-01833-9.

Correspondence and requests for materials should be addressed to Paul E. Schavemaker or Sergio A. Muñoz-Gómez.

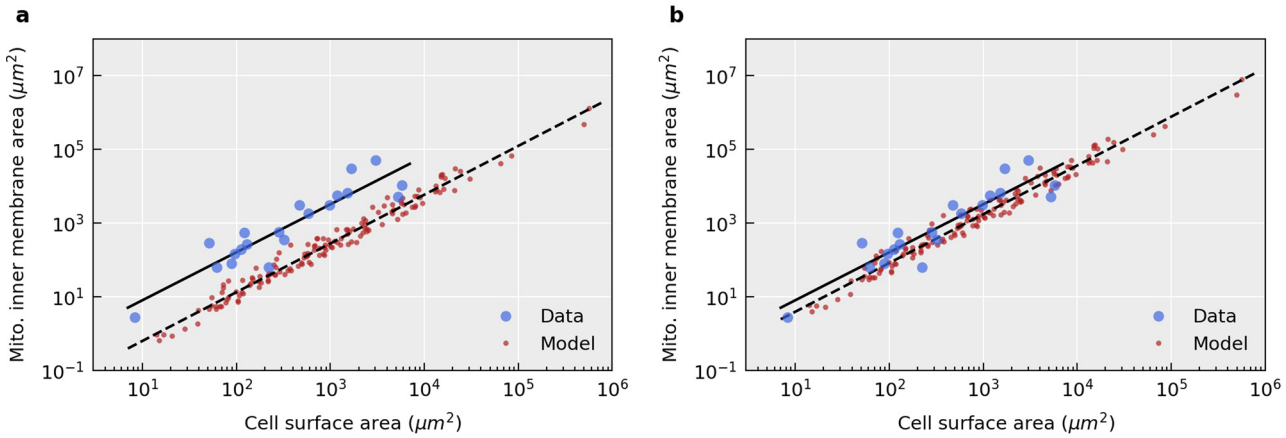
Peer review information *Nature Ecology and Evolution* thanks István Zachar and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.



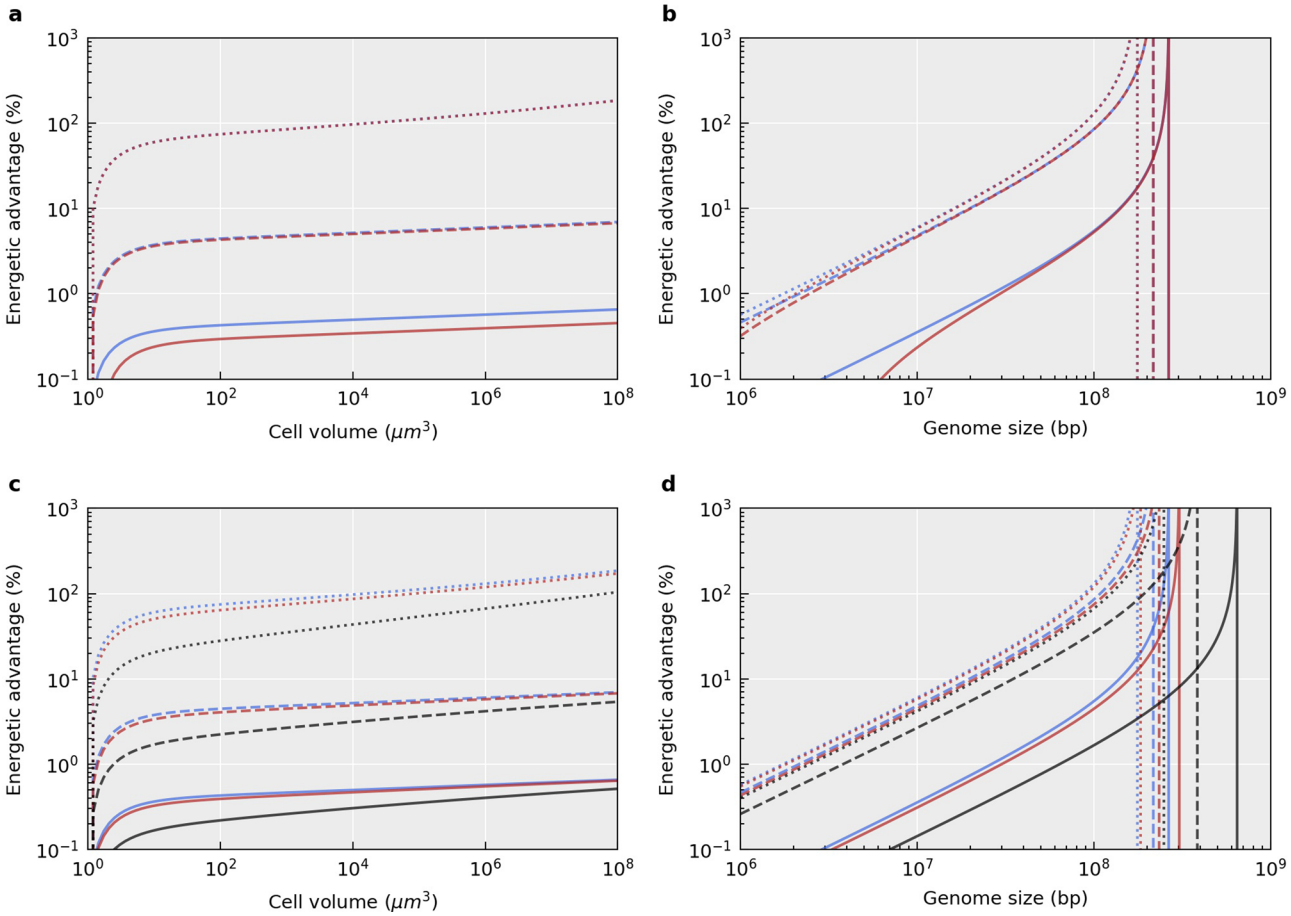
Extended Data Fig. 1 | Comparing the results of metabolic rate calculations to data. The blue points are empirically determined metabolic rates for various prokaryotic and eukaryotic species, obtained from Chiyomaru and Takemoto (2020) (units were converted by assuming that 1 mol ATP releases 50 kJ of energy). The red points are metabolic rates calculated with: $R = \left(f_d \alpha V^{0.97}/t_d\right) + \beta V^{0.88}$, with the values for cell volumes and cell division times, for both prokaryotes and eukaryotes, obtained from Lynch and Marinov (2015). The solid line is a fit to the data: $y = 2.0 \times 10^5\ x^1$, and the dashed line is a fit to the calculated points: $y = 4.4 \times 10^5\ x^{0.85}$.



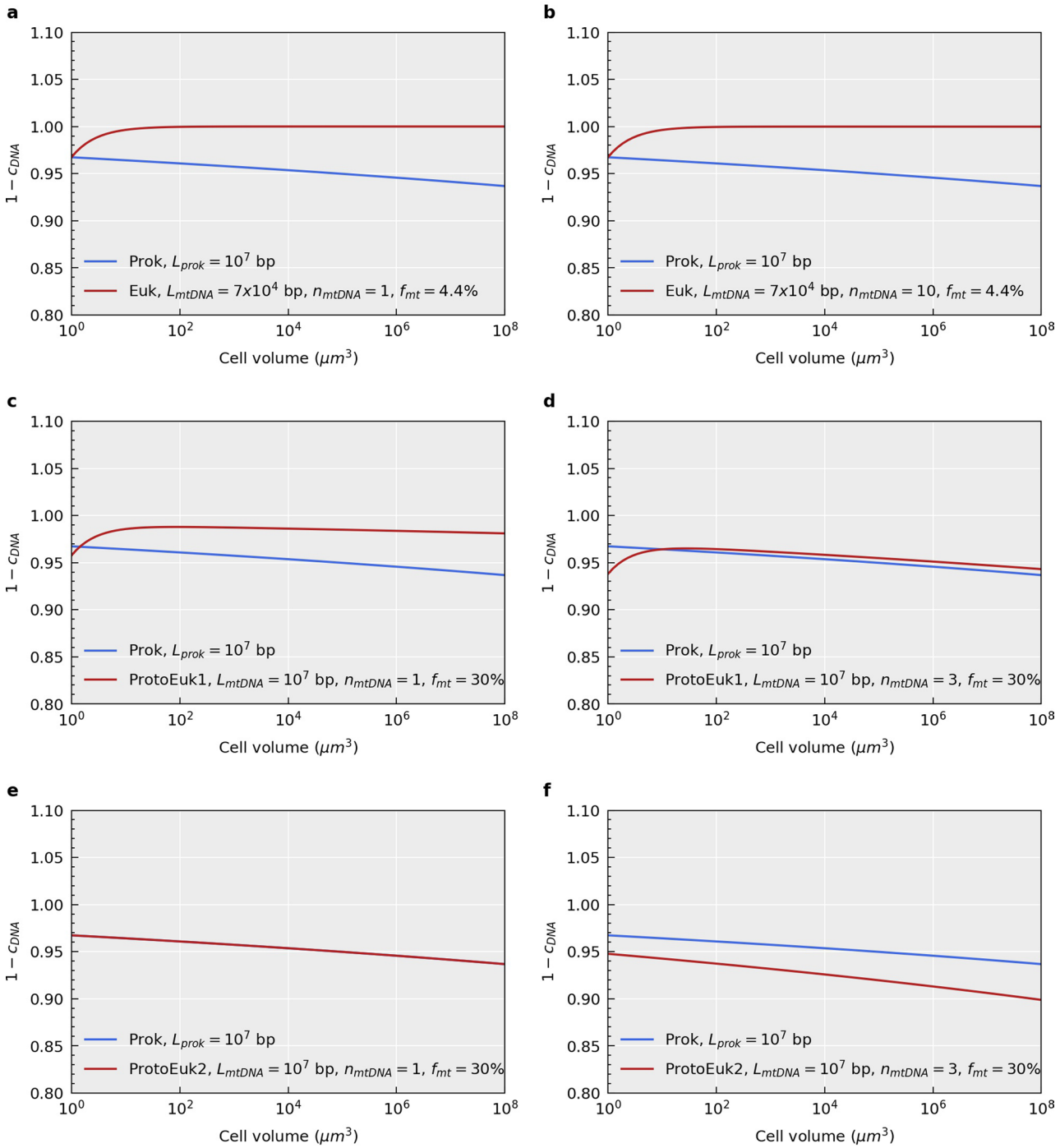
Extended Data Fig. 2 | The shape factor (*S*) as a function of the ratio between cell length and width. When this ratio is one, the cell is a sphere, and when this ratio is < 1 or >1, the cell is flattened into an oblate or prolate spheroid, respectively. The shape factor is calculated from Eq. S17.



Extended Data Fig. 3 | Prediction of the mitochondrial inner membrane surface area. a. The inner mitochondrial surface area as a function of cell surface area. Empirically determined inner mitochondrial membrane areas were obtained from Lynch and Marinov (2017) (blue points). The inner mitochondrial membrane area was calculated (red points) with: $(\left(f_d \alpha V^{0.97}/t_d\right) + \beta V^{0.88})/r \times A_r \times 2.5$, using cell volumes and cell division times for eukaryotic species obtained from Lynch and Marinov (2015). The factor 2.5 was included to account for the lipids that support the membrane (Lindén et al., 2012). Note that for the calculation it is assumed that the inner mitochondrial membrane only houses respiratory proteins. The solid line is a fit to the data: $y = 0.40\ x^{1.30}$. The dashed line is a fit to the model: $y = 0.030\ x^{1.32}$. Here, the value of A_r is the one for *E. coli*, which is listed in Table S1. **b**. As for A except that the value of A_r used for the model calculations, which is dependent on both cross-sectional surface areas and stoichiometries of respiratory enzymes, is taken from a eukaryote (bovine) (Schlame, 2021), yielding a closer correspondence between data and model.



Extended Data Fig. 4 | The effect of varying mitochondrial genome copy number, mitochondrial genome size, and cell division times on the eukaryotic advantage over prokaryotes. Plots are generated from Eqs. 4–6 with $V_{gserv} = 1\ \mu m^3$ and $f_{mt} = 0.044$. **a, b**. Varying mitochondrial genome number and size. For the blue lines, $L_{mtDNA} = 10^4$ bp and $n_{mtDNA} = 1$ per μm^3 of mitochondrial volume. For the red lines $L_{mtDNA} = 7{\times}10^4$ bp and $n_{mtDNA} = 100$ per μm^3 of mitochondrial volume. Cell division time, $t_d = 0$. In some cases, red and blue overlap. **a**. For the dotted lines $L_{prok} = L_{euk} = 10^8$, for the dashed lines $L_{prok} = L_{euk} = 10^7$, and for the solid lines $L_{prok} = L_{euk} = 10^6$. **b.** For the dotted lines $V = 10^6\ \mu m^3$, for the dashed lines $V = 10^3\ \mu m^3$, and for the solid lines $V = 1.1\ \mu m^3$. **c, d**. Varying cell division time, t_d. For all lines $L_{mtDNA} = 10^4$ bp and $n_{mtDNA} = 1$ per μm^3. For the blue lines $t_d = 0$, for the red lines $t_d = 10$ h, and for the black lines $t_d = 100$ h. **c**. For the dotted lines $L_{prok} = L_{euk} = 10^8$, for the dashed lines $L_{prok} = L_{euk} = 10^7$, and for the solid lines $L_{prok} = L_{euk} = 10^6$. **d**. For the dotted lines $V = 10^6\ \mu m^3$, for the dashed lines $V = 10^3\ \mu m^3$, and for the solid lines $V = 1.1\ \mu m^3$.



Extended Data Fig. 5 | The amount of cellular ATP that remains after DNA synthesis in prokaryotes and either modern or ancestral eukaryotes. Plots are generated from Eqs. 5 and 7 with $V_{gserv} = 1\,\mu m^3$, $L_{prok} = 10^7$bp, and $f_{mt} = 0.3$. **a**. Amount of ATP left after DNA synthesis for prokaryotes and modern eukaryotes with a small mitochondrial genome size ($L_{mtDNA} = 7{\times}10^4$bp) and volume fraction ($f_{mt} = 0.044$), a main (nuclear) genome that does not scale with cell volume, and a low mitochondrial genome copy number per unit volume ($n_{mtDNA} = 1$). **b**. As above but with $n_{mtDNA} = 10$. **c**. Amount of ATP left after DNA synthesis for prokaryotes and ancestral proto-eukaryotes with a large mitochondrial genome size ($L_{mtDNA} = 10^7$bp) and volume fraction ($f_{mt} = 0.3$), a main (nuclear) genome that does not scale with cell volume, and a low mitochondrial genome copy number per unit volume ($n_{mtDNA} = 1$); this model and parameter set best reflect an ancestral eukaryote as predicted by some mitochondrion-late scenarios. **d**. As above but with $n_{mtDNA} = 3$. **e**. Amount of ATP left after DNA synthesis for prokaryotes and ancestral eukaryotes with a large mitochondrial genome size ($L_{mtDNA} = 10^7$ bp) and volume fraction ($f_{mt} = 0.3$), a main genome size that scales with cell volume, and a low mitochondrial genome copy number per unit volume ($n_{mtDNA} = 1$); this model and parameter set best reflect an ancestral eukaryotes as predicted by mitochondrion-early scenarios. **f**. As above but with $n_{mtDNA} = 3$.

nature portfolio

Corresponding author(s): Paul E. Schavemaker and Sergio A. Muñoz-Gómez

Last updated by author(s): Jun 14, 2022

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☒	☐	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☒	☐	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No specific software was used.
Data analysis	No specific software was used.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

All data used are available in the Supplementary Information and Source Data Fig 1 files.

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	N/A
Population characteristics	N/A
Recruitment	N/A
Ethics oversight	N/A

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	N/A. This was a theoretical study that used mathematical models.
Research sample	N/A
Sampling strategy	N/A
Data collection	N/A
Timing and spatial scale	N/A
Data exclusions	N/A
Reproducibility	N/A
Randomization	N/A
Blinding	N/A

Did the study involve field work? ☐ Yes ☒ No

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging