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

2   Paul E. Schavemaker[1,2,*], Sergio A. Muñoz-Gómez[1,2,*+]

3   [1]Center for Mechanisms of Evolution, The Biodesign Institute, School of Life Sciences, Arizona
4   State University, 727 E. Tyler St. Tempe, AZ 85281-5001, U.S.A.
5   [+]Current address: Unité d'Ecologie, Systématique et Evolution, Université Paris-Saclay, Orsay,
6   France.
7   [2]Equal contribution.

8   *Correspondence to: pschavem@asu.edu and sergio.munoz@universite-paris-saclay.fr

9   **Abstract**

10   The origin of eukaryotic cell size and complexity is thought by some to have required an energy
11   excess provided by mitochondria, whereas others claim that mitochondria provide no energetic
12   boost to eukaryotes. Recent observations show that energy demand scales continuously and
13   linearly with cell volume across both prokaryotes and eukaryotes, and thus suggest that eukaryotes
14   do not have an increased energetic capacity over prokaryotes. However, amounts of respiratory
15   membranes and ATP synthases scale super-linearly with cell surface area. Furthermore, the
16   energetic consequences of the contrasting genomic architectures between prokaryotes and
17   eukaryotes have yet to be precisely quantified. Here, we investigated (1) potential factors that affect
18   the cell volumes at which prokaryotes become surface area-constrained, (2) the amount of energy
19   that is divested to increasing amounts of DNA due to the contrasting genomic architectures of
20   prokaryotes and eukaryotes, and (3) the costs and benefits of respiring symbionts for a host cell.
21   Our analyses suggest that prokaryotes are not necessarily constrained by their cell surfaces at cell
22   volumes of 10^0–10^3 µm^3, and that the genomic architecture of eukaryotes is only slightly
23   advantageous at genomes sizes of 10^6–10^7 bp. This suggests that eukaryotes may have first
24   evolved without the need for mitochondria as these ranges hypothetically encompass the Last
25   Eukaryote Common Ancestor and its proto-eukaryotic ancestors. At these slightly larger cell
26   volumes, a host cell may also have derived a greater advantage (lower cost) from harboring a
27   population of ATP-producing symbionts. However, our analyses also show that increasingly larger
28   and fast-dividing prokaryotes would have a shortage of surface area devoted to respiration and
29   would disproportionally divest more energy to DNA synthesis at larger genome sizes. We thus
30   argue that, even though mitochondria may not have been required by the first eukaryotes, the
31   successful diversification of eukaryotes into larger and faster-growing cells was ultimately
32   contingent upon the origin of mitochondria.

33   **Keywords:** Eukaryogenesis, energy, complexity, genome size, cell volume.

34   **Significance**

35   There has been a lot of theorizing about the evolution of eukaryotes from prokaryotes, but no
36   consensus seems to be on the horizon. Our quantitative analyses on the required amount of
37   respiratory membrane, and the amount of energy diverted to DNA synthesis, by both prokaryotes
38   and eukaryotes, suggest that mitochondria provided rather small advantages to the first eukaryotes,
39   but were advantageous for the macro-evolutionary diversification of eukaryotes. This conclusion
40   provides a middle road in the debate between those that claim that the origin of eukaryotes required
41   a massive energy boost provided by mitochondria, and those that argue that the origin of
42   mitochondria did not represent a quantum leap in energetic advantages to eukaryotes.

43   **Introduction**

44   The transition from prokaryotic to eukaryotic cells is often thought to be the greatest transition in
45   the history of life (1). This is because this is the largest gap, or discontinuity, in organismal structure
46   or organization across the tree of life: eukaryotic cells are structurally much more complex, and on

1   average, also larger in volume than prokaryotic cells (2). Many authors have thus attempted to
2   explain how eukaryotes evolved from prokaryotes (3–9). However, much debate and speculation
3   persist about the processes that gave rise to the first eukaryote (10, 11, 3, 12, 13).

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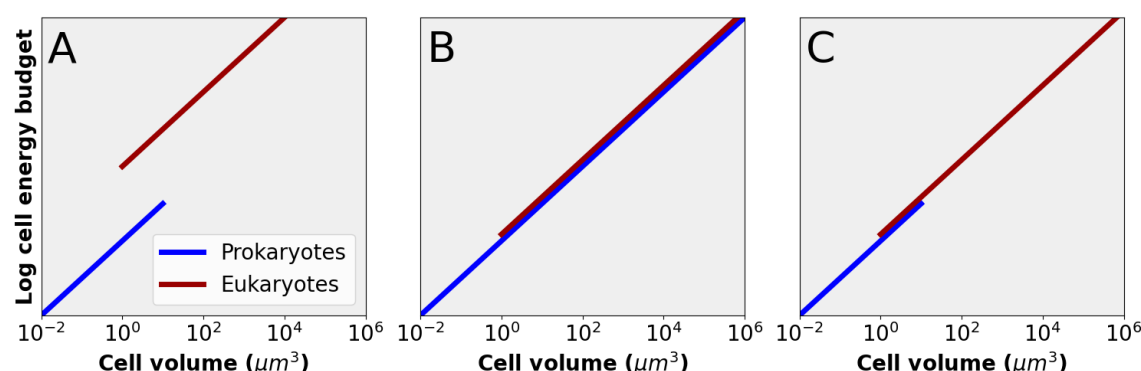20   and were thus a pre-requisite for the evolution of the eukaryotic complexity (3, 19).

21

Figure 1. Three different possibilities for the energetic scaling across cell volume for prokaryotes and eukaryotes. A. A hypothetical discontinuity in the scaling of cell energy with cell 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 cell 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.

31   Some authors have expressed skepticism about the energetic hypothesis for the origin of
32   eukaryotic complexity (10, 13, 20–25). The notion that the evolution of cell complexity requires an
33   increased energy supply has been dismissed as having no evolutionary basis (23, 25), and the
34   concept of 'energy per gene' has been criticized as evolutionarily meaningless (24, 26). Recently,
35   Lynch and Marinov found a continuous energetic scaling across prokaryotes and unicellular
36   eukaryotes (10), and similar results have been presented by Chiyomaru and Takemoto (20). This
37   suggests that there is no energetic gap (or shift in energetic capacity) between prokaryotes and
38   eukaryotes, as the amount of energy available to a cell is directly proportional to its volume
39   regardless of whether the cell is prokaryotic or eukaryotic. Based on this, Lynch and Marinov argue
40   that mitochondria do not provide eukaryotes with a higher energetic capacity and imply that
41   prokaryotes are energetically unconstrained by their cell surfaces (Fig. 1B). Moreover, Lynch and
42   Marinov showed that the number of ATP synthases scales continuously across prokaryotes and

1 eukaryotes, and argued that the increase in surface area provided by mitochondria is not
2 particularly large when compared to that available at the cytoplasmic membrane (22). However,
3 their data also show that the amount of mitochondrial membrane and the number of ATP synthases
4 scale super-linearly with the cell surface area (22). This suggests, in contrast to Lynch and Marinov
5 (10, 22), that prokaryotes might be constrained by their cell surfaces at larger volumes, and that
6 mitochondria may allow eukaryotes to scale up in cell volume without a shortage of respiratory
7 membranes (Fig. 1C). Furthermore, the energetic consequence of the contrasting genomic
8 architectures between prokaryotes and eukaryotes, first emphasized by Lane and Martin (3, 17)
9 but ignored by others, remains to be explored.

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

21 **Results**

22 <u>The cell volume, genome size, and gene number distributions of prokaryotic and eukaryotic cells</u>

23 In this manuscript, we use theoretical models to assess the respiratory membrane requirements
24 and DNA investments of mitochondrion-less and mitochondrion-bearing cells. These models might
25 help explain, from an energetic point of view, the differences observed between modern
26 prokaryotes and eukaryotes and thus inform our discussions of the prokaryote-eukaryote transition.
27 We start by presenting the distributions of cell volume, genome size and gene number from a
28 comprehensive survey of phylogenetically disparate prokaryotes and eukaryotes (Fig. 2; Dataset
29 S1).

30 The cell volume distributions of prokaryotes and eukaryotes point at two main conclusions. First,
31 the ranges for each grade of organization ($\sim 10^{-2}$–10^2 μm^3 for prokaryotes and $\sim 10^0$–10^8 μm^3 for
32 eukaryotes) do not overlap for the most part: their medians (vertical dashed lines in Fig. 2;
33 calculated over the full distribution of cell volumes) largely fall outside of each other's distributions
34 (Fig. 2A). This is most obvious when giant bacteria like *Beggiatoa* spp. and *Thiomargarita*
35 *namibiensis* are excluded (Fig. 2A; Dataset S1). Giant bacteria reach absolute volumes >10^6 μm^3
36 but these are mostly inert as they contain a large central vacuole or numerous intracellular
37 inclusions made of sulfur or calcium carbonate (some exceptions are large cyanobacterial cells;
38 Dataset S1). Thus, most prokaryotes are smaller than most eukaryotes. Second, prokaryotes and
39 eukaryotes mostly overlap at cell volumes of $\sim 10^0$–10^2 μm^3 (Fig. 2A). This overlap includes large
40 bacteria with entirely active cytoplasms composed of energy-demanding macromolecules (e.g.,
41 *Azotobacter chroococcum*, *Magnetobacterium bavaricum*, '*Candidatus* Uab amorphum', and
42 *Chromatium okenii*; Dataset S1), picoeukaryotes which are relatively reduced (e.g., algae such as
43 *Chaetoceros calcitrans*, *Micromonas pusilla*, *Nannochloris* sp., and *Nannochloropsis geditana*;
44 Dataset S1), and phylogenetically diverse nanoeukaryotes (e.g., heterotrophic flagellates such as
45 *Andalucia godoyi*, *Mantamonas plastica*, *Bodo saltans*, *Malawimonas jakobiformis*, *Palpitomonas*
46 *bilix*, *Ancyromonas mylnikovi*, *Reclinomonas americana*; Dataset S1). Many small eukaryotes (both
47 parasitic and free-living) can thus have sizes similar to those of many bacteria.
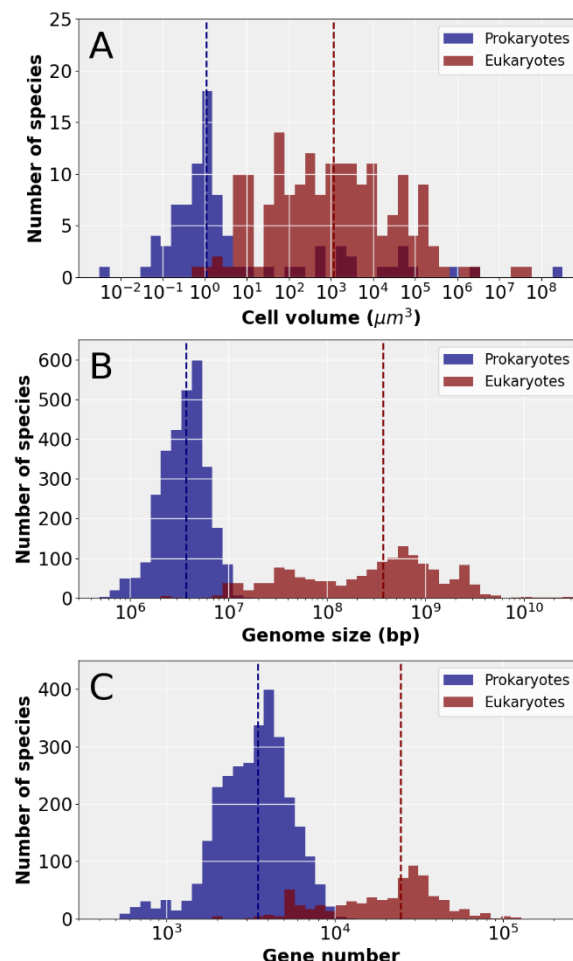
1
2 **Figure 2. Cell volumes, genome sizes, and gene numbers for prokaryotes and eukaryotes.** Cell volumes
3 for diverse eukaryotes were obtained from (10) and additional data were added from several sources (see
4 Dataset S1). Genome sizes and gene numbers were acquired from NCBI GenBank and manually curated to
5 remove outliers due to gene mis annotations. The vertical dashed lines show medians. Total cell volumes,
6 instead of energy-demanding active cytoplasmic volume, were used for giant prokaryotes (>10^2 μm^3).

7 The histogram for genome size follows a similar pattern to that of cell volume: prokaryotes and
8 eukaryotes have distinct but overlapping distributions (Fig. 2B). The genome size range for
9 prokaryotes is <1–16 Mbp, whereas that of eukaryotes is ~8–10,000 Mbp. This suggests that there
10 is an upper genome size constraint to prokaryotes based on the currently available data.
11 Prokaryotes and eukaryotes also overlap at genomes sizes of ~8–16 Mbp if the genomes of highly
12 reduced eukaryotic parasites are excluded (Fig. 2B; Dataset S1). Many eukaryotes (protists) thus
13 have genome sizes smaller than those of some prokaryotes. For example, prokaryotes such as
14 myxobacteria, actinomycetes, cyanobacteria, and planctomycetes may have genomes of up to 16
15 Mbp in size (Dataset S1). The smallest genomes for free-living eukaryotes are those of some small
16 green algae, red algae, and yeasts (8–13 Mbp); some parasitic eukaryotes have genome sizes of
17 just 2 or 6 Mbp (e.g., *Encephalitozoon* and *Babesia*; Dataset S1). The small heterotrophic
18 nanoflagellate *Andalucia godoyi* (Jakobea), which have the most ancestral-like mitochondrial
19 genomes, has a nuclear genome size of ~20 Mbp (27), barely larger than the largest prokaryotic
20 genomes. For gene number, there is an even wider overlap between prokaryotes and eukaryotes
21 (Fig. 2C). Prokaryotes with the greatest number of genes have 10,000–13,000 genes (Dataset S1),
22 whereas eukaryotes with the lowest number of genes include intracellular parasites (~2,000 genes
23 in *Encephalitozoon*), free-living fungi (~4,500 genes in *Malassezia restricta* or ~6,400 in
24 *Saccharomyces cerevisiae*), and small algae (~5,300 genes in *Cyanidioschyzon* and ~7,800 genes
25 in *Ostreococcus tauri*). Some of the closest relatives of animals, the free-living flagellate *Monosiga*

1   *brevicollis* (Choanoflagellata) and the symbiotic amoeba *Capsaspora owczarzaki* have ~9,200 and
2   8,800 genes, respectively (Dataset S1). In summary, the data suggest that, though there is some
3   overlap between prokaryotes and eukaryotes, there also appears to be upper constraints to the cell
4   volumes, gene numbers, and genome sizes that prokaryotes can attain.

5   The respiratory membrane requirement and maximum possible volume of cells

6   The cell volume of prokaryotes is potentially constrained by respiration at the cell surface (or by not
7   having mitochondria; see introduction). In terms of energy, the rate of ATP synthesis at the cell
8   surface must meet the rate of ATP consumption by the whole cell volume; GTPs used by ribosomes
9   or tubulin, for example, are treated as equivalent to ATPs. However, surface area decreases
10   relative to volume as cells grow larger—surface area scales with the square of the linear dimension,
11   whereas volume scales with the cube of the linear dimension (28–30). A developmental or
12   evolutionary increase in cell volume thus poses a challenge to cells because, if internal volumes
13   remain active, processes that are carried out at the cell surface (e.g., respiration or nutrient
14   transport) will, at some cell volume, be unable to support processes that occur in the cytoplasm
15   (e.g., protein synthesis). Such scaling sets a maximum volume that cells cannot overcome in the
16   absence of structural adaptations (e.g., mitochondria and endomembranes in eukaryotes, and
17   intracytoplasmic membranes in prokaryotes (31)).

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

33
$$Respiratory\ deficit = \frac{A_{needed}}{A_{possible}} = \frac{A_{needed}}{f_{max}A_{total}}. \quad (1)$$

34   The amount of respiratory membrane needed by a cell can be calculated by multiplying the number
35   of respiratory units (i.e., a complete set of respiratory complexes including the ATP synthase)
36   required and the membrane area that each one of them takes up (A_r). The number of respiratory
37   units can be estimated by dividing the metabolic rate of a cell (R, in ATP h⁻¹) by the ATP production
38   rate of a single respiratory unit (r, in ATP h⁻¹; Table S1). The metabolic rate of a cell can be
39   expressed as the total ATP budget of a cell throughout its lifetime (E_t) divided by the cell division
40   time (t_d in h). The total energy budget of a cell (in ATPs) comprises both growth and maintenance
41   costs (c_g and c_m, respectively) and is calculated as in (10); this is adjusted to only include direct
42   costs using the f_d factor (32). Opportunity costs, which deal with the potential energy stored in
43   previously synthesized macromolecules, are excluded because this energy is not directly available
44   as physical ATPs. The metabolic rate calculations agree with those reported previously by
45   Chiyomaru and Takemoto, and are thus validated by empirical data (20) (Fig. S1). The amount of
46   respiratory membrane needed by a cell thus depends on its energy demands, cell division time,
47   and rate of ATP synthesis and area occupied by a single respiratory unit:

1
$$A_{needed} = \frac{R}{r} A_r = \frac{E_t/t_d}{r} A_r = \frac{\left(f_d c_g + t_d c_m\right)\big/t_d}{r} A_r. \quad (2)$$

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

4
$$Respiratory\ deficit = \frac{\dfrac{f_d \alpha V^{0.97}\big/t_d + \beta V^{0.88}}{r} A_r}{f_{max} S V^{2/3}}, \quad (3)$$

5 where S is a factor that specifies the shape of a cell (e.g., a perfect sphere or differently flattened
6 spheroids; Supplementary Information). The parameters f_d, α, β, A_r, and r are constants whose
7 values have been previously determined (10, 32–35) (Table S1). The parameters t_d, f_{max}, and S
8 are constrained within biologically plausible ranges. For example, t_d is varied between 1 and 10 h,
9 corresponding to the lower range of prokaryotic cell division times and the geometric mean of
10 eukaryotic cell division times, respectively (10). The f_{max} parameter is varied between 8 and 18%,
11 which are the largest possible fraction of respiratory membrane in *E. coli* (35) and the membrane
12 fraction at which roughly half of all membrane proteins are respiratory enzymes (36). The shape
13 factor, S, is varied between 4.8 and 12.1, which correspond to a sphere and an oblate spheroid
14 with a cell length-width ratio of 0.1 (Fig. S2, Supplementary Information).

15 To assess the maximum possible volume that mitochondrion-less cells can achieve, we calculated
16 the deficit in respiratory membrane (Eq. 3) across a wide range of cell volumes (Fig. 3A). Values
17 of < 1 for the respiratory deficit indicate that the cell has an excess of respiratory membrane,
18 whereas values of > 1 indicate that the cell has insufficient respiratory membrane to sustain its own
19 volume. Respiratory deficit values of 1 thus point at the maximum volume that a simple
20 mitochondrion-less cell can achieve. Our analyses show that spherical cells with a cell division time
21 of 1 h and a maximum respiratory membrane fraction of 8% are surface area-constrained above a
22 cell volume of about 1 µm^3 (blue line; Fig. 3A). These parameter values and the estimated cell
23 volume limit agrees with what is seen for a small and fast-growing bacterium like *Escherichia coli*
24 (35). If half of membrane proteins are respiratory enzymes (i.e., a maximum respiratory membrane
25 fraction of 18%), the largest volume that a cell can achieve is about 10 µm^3 (black line; Fig. 3A).
26 This large fraction of respiratory membrane would be possible if a bacterium devotes less of its
27 surface area to other processes (e.g., flagella or chemotactic receptors), or alternatively, if a
28 bacterium develops intracytoplasmic membranes for respiratory processes (31). A similar cell
29 volume limit of 10 µm^3 is achieved if the cell shape is changed to that of an oblate spheroid with a
30 cell length-width ratio of 0.1 (dashed black line; Fig. 3A); some small and flattened flagellates like
31 the eukaryote *Petalomonas minor* (37), or the phagocytic amoeboid prokaryote *Uab amorphum*
32 have such cell body shapes (38). The cell volume limit is raised even more, to about 500 µm^3, if
33 the cell division time is increased to 10 h (dotted black line; Fig. 3A). The combination of these
34 three changes raises the cell volume limit to higher than 10^5 µm^3 (red line; Fig. 3A). This might
35 correspond to giant bacteria like *Thiomargarita* and *Epulopiscium* whose active cytoplasm is
36 restricted to a thin enveloping sheet (i.e., 2% of the whole cell volume (39)), have long cell division
37 times (1–2 weeks (40)), and develop extensive intracytoplasmic membranes (e.g., *Epulopiscium*
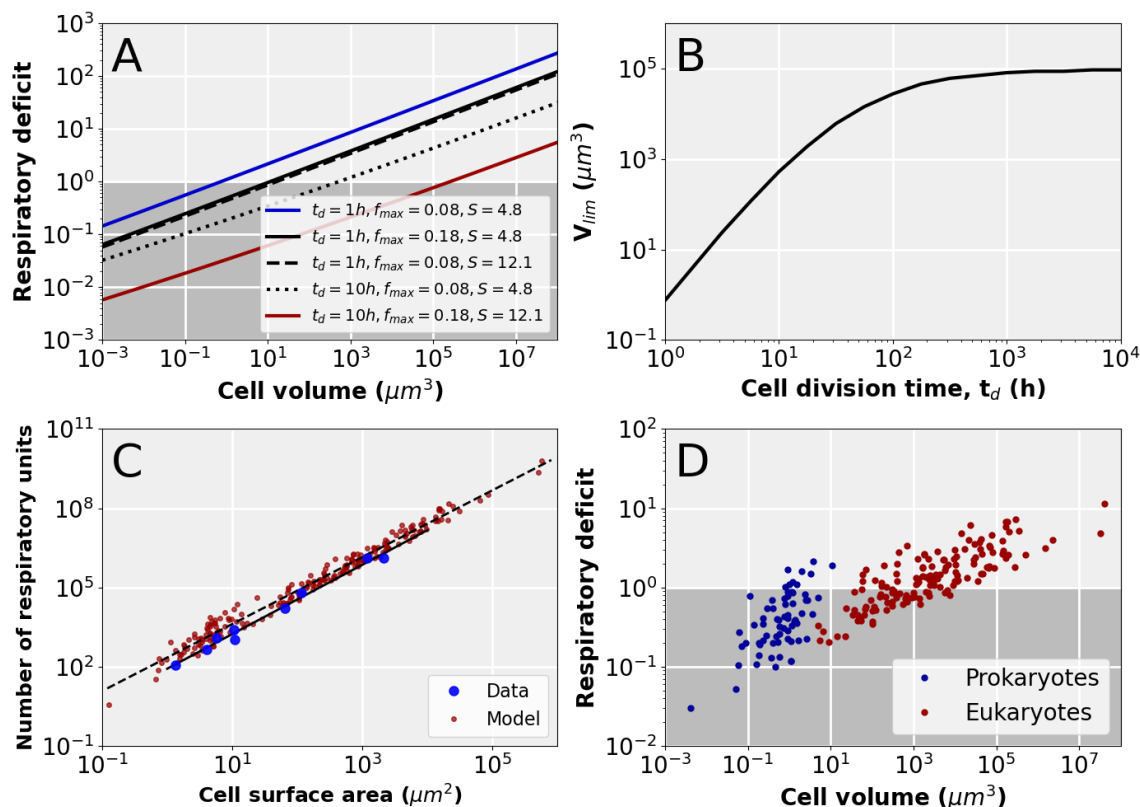38 (41)).
39

Figure 3. Factors that affect the volumes at which simple cells become constrained by their surface.
A. The 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 (see inset). The red line reflects cells for which all parameters have been simultaneously changed. The dark grey area indicates the domain in which there is enough surface area for respiration to support cell volumes. The intersection between each line (a defined set of parameters; see inset) and a respiratory deficit of one 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. Here, 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 (22) (blue points). The number of respiratory units was calculated (red points) using: $((f_d \alpha V^{0.97}/t_d) + \beta V^{0.88})/r$, with the cell volumes and cell division times, for a range of prokaryotic and eukaryotic species, obtained from (10). The solid line is a fit to the data: y = 83 x$^{1.31}$. The dashed line is a fit to the model: y = 221 x$^{1.27}$. **D**. Respiratory deficit calculated for individual prokaryotic and eukaryotic species whose cell volumes and cell division times have been previously estimated (10). Here, f_{max} = 8% and S = 4.8 (spherical cells).

Cells with longer cell division times have lower metabolic rates and thus require fewer respiratory units (Eq. 2). This is because longer cell division times allow cells to accumulate the same amount of ATP required for growth over longer times 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 µm^3 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 (Eq. 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

7

1 of cells previously reported by Lynch and Marinov (22) (Fig. 3C). Similarly, the amount of respiratory
2 membrane required by eukaryotic cells also follows the data on mitochondrial inner membrane
3 areas reported by Lynch and Marinov (22) after adjusting for the cross-sectional surface areas and
4 stoichiometries of mitochondrial respiratory complexes (42) (Fig. S3).

5 We also calculated the respiratory deficit for prokaryotes and eukaryotes whose cell volumes and
6 cell division times have been determined empirically (10). For these calculations, we assumed a
7 spherical cell body shape ($S = 4.8$) and a maximum respiratory membrane fraction of 8%. These
8 analyses showed that simple cells may have eukaryote-like volumes of up to 10^4 µm^3 without a
9 shortage of surface area for respiration (dark area; Fig. 3D). Therefore, many eukaryotes might be
10 able, theoretically, to support their cell volumes by respiring at their cytoplasmic membranes (i.e.,
11 without the need for internalizing respiratory membranes). Overall, our analyses thus reveal that
12 longer cell division times (t_d), flattened or elongated cell shapes (S), and a larger allocation of
13 surface area to respiration (f_{max}) can, together or in isolation, allow cells to obtain larger volumes
14 without the need for expanded respiratory membranes (e.g., mitochondria). On the other hand,
15 increasingly larger, rounder and faster-dividing cells have higher respiratory deficits (i.e., larger
16 than one) and are thus dependent on an excess of respiratory membranes that cannot be fully
17 accommodated on their cytoplasmic membranes.

18 The energetic investments in DNA of cells with contrasting genomic architectures

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

30 The concept of 'energy per gene expressed' has been criticized as having no evolutionary
31 relevance (see (10, 23, 24)). This concept, as used by Lane and Martin, heavily penalizes large
32 prokaryotes as their gene numbers increase with polyploidy. However, the amount of gene
33 expression from each gene, irrespective of how many times the gene is duplicated, is proportional
34 to cell volume. In other words, the relative cost of a gene, a more evolutionary meaningful concept
35 (10, 22), remains constant. This is because the energetic demands of cells strictly depend on their
36 cell volumes, i.e., prokaryotes and eukaryotes of the same volume require the same amount of
37 energy. The concept of 'energy per gene' thus unfairly penalizes prokaryotes, or any polyploid.
38 Furthermore, measurements of 'energy per gene' previously performed by Lane and Martin (3)
39 unfairly favor eukaryotes because gene copies due to mitochondrial genome polyploidy (which
40 scale with cell volume) were ignored (3). Because the concept of 'energy per gene' is inappropriate,
41 our approach below thus relies on estimating the cost of cellular features (i.e., DNA synthesis)
42 relative to the entire energy budget of a cell (10, 22).

43 To test the hypothesis that the genomic architecture of eukaryotic cells provides an overwhelming
44 advantage, we developed an explicit model that compares the energetic capacity of eukaryotes to
45 prokaryotes. The goal here is to isolate the genomic architecture of a cell from other confounding
46 factors that also separate eukaryotes from prokaryotes. Because ATP demands depend on cell
47 volume (and not complexity or gene number (10, 20)), we considered the amount of ATP that
48 remains ($1 - c_{DNA,euk}$ and $1 - c_{DNA,prok}$) after accounting for the relative cost of DNA that is
49 associated with each genomic architecture ($c_{DNA,euk}$ and $c_{DNA,prok}$; Fig. 4 and Eq. 4). This remaining
50 amount of ATP is devoted to all cell processes other than DNA synthesis (e.g., translation,

1 transcription, lipid biosynthesis, etc.); the more ATP a cell invests in DNA, the less ATP there is to
2 sustain other cellular processes. The ratio between the remaining ATP of a mitochondrion-bearing
3 and mitochondrion-less cell thus provides a measure of the energetic advantage (>1), or
4 disadvantage (<1), that mitochondrion-bearing cells might have. This can be expressed as:

5
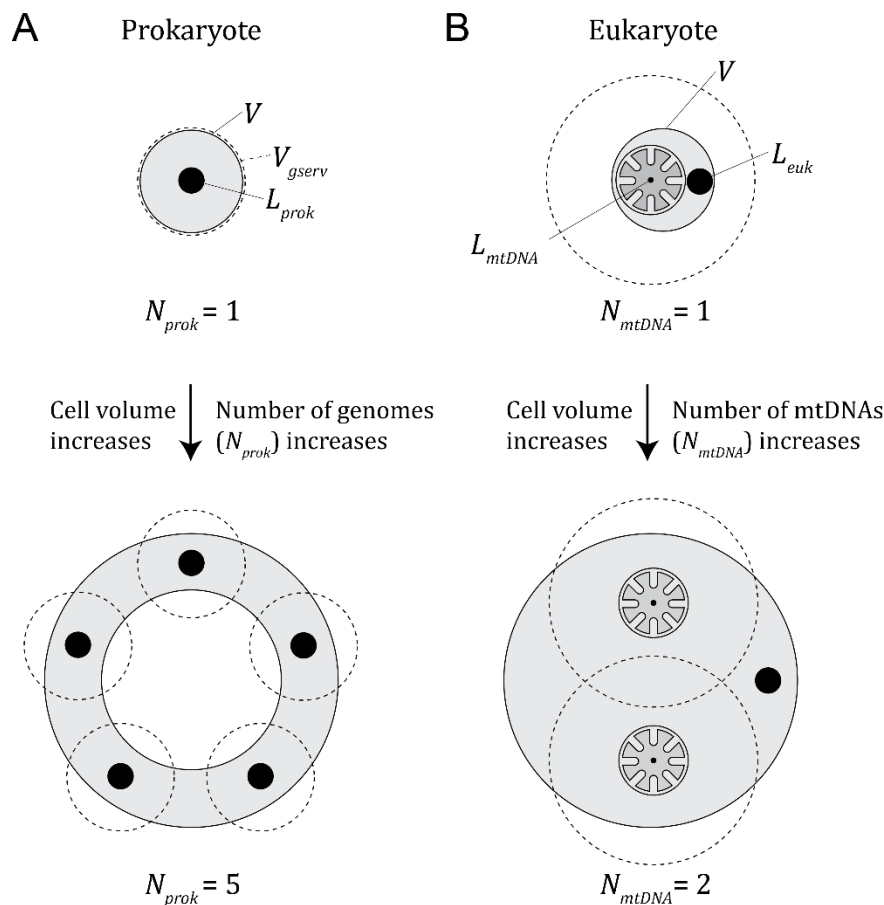$$Energetic\ advantage\ (\%) = \left(\frac{1 - c_{DNA,euk}}{1 - c_{DNA,prok}} - 1 \right) \times 100, \quad (4)$$

6

7

8
9 **Figure 4. Graphical representation of contrasting genomic architectures in prokaryotes and**
10 **eukaryotes (Eq. 4 and see main text for an explanation of parameters). A**. The genomic symmetry of
11 prokaryotes. We have represented a large prokaryotic cell as a shell of cytoplasm surrounding a large inert
12 central space, as seen in giant bacteria like *Epulopiscium* and *Thiomargarita*. Even though this cell architecture
13 is irrelevant for our calculations (Eq. 5) as only the number of genomes is considered (filled black circles),
14 prokaryotic cells have to scale up in cell volume with such an architecture to remain viable in the absence of
15 an active intracellular transport network (43). The total number of genomes N_{prok} is a function of the ratio of
16 the cell volume and the volume controlled by a single genome (i.e., V/V_{gserv}; see Eq. 5). **B**. The genomic
17 asymmetry of eukaryotes. The dashed circles hypothetically represent the amount of volume that can be
18 energetically supported by mitochondria. Because of cristae (expanded internalized respiratory membranes),
19 mitochondria can, in principle, energetically support large cytoplasmic volumes. The total number of

9

1   mitochondrial genomes N_{mtDNA} is a function of the total volume of mitochondria and the number of mtDNA
2   molecules per µm³ of mitochondrial volume ($n_{mtDNA}f_{mt}V$; see Eq. 6 and main text) (44–46).
3
4   To calculate the cost of DNA for a prokaryote (a mitochondrion-less cell), we consider a cell with
5   only a single main genome type. In prokaryotes, the number of genomes increases proportionally
6   with cell volume, as seen in *Synechococcus elongatus* (47) or in giant bacteria like *Epulopiscium*
7   (48) (see Fig. 4). The cause of this scaling might be the need to either bypass a diffusion constraint
8   in the absence of active intracellular transport (43) or maintain genomes physically adjacent to
9   respiratory membranes for efficient regulation (3, 19). We compiled data for several prokaryotes
10   that show that the cell volume per genome does not exceed 2 µm³ in several prokaryotes (V_{gserv} in
11   Eq. 5; see Table S2). Our model thus assumes that if cell volume increases, the number of
12   genomes must increase accordingly (see Supplementary Information for more details). The
13   absolute total cost of DNA (in units of ATP) for a prokaryotic cell is the product of the amount of
14   ATP required for synthesizing a single base pair (101 ATPs; we used the cost value by (10), which
15   differ slightly from that of (32)), the length of a single genome in base pairs (L_{prok} in Eq. 5) and the
16   number of genomes. The number of genomes is the ratio between the total cell volume (V) and cell
17   volume serviced by a single genome (V_{gserv}) (Eq. 5). Note that, in contrast to the respiratory deficit
18   calculations, we include here the opportunity costs of synthesizing DNA because we want to
19   account for the complete, evolutionarily significant, drain on the cell resources. Finally, to obtain
20   the relative cost of DNA for a prokaryotic cell, the absolute cost of the DNA is divided by the total
21   ATP budget of the cell throughout its lifetime (Eq. 5). This can be expressed as a function of cell
22   volume (10):

23

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

24   To calculate the cost of DNA for a eukaryote (a mitochondrion-bearing cell), we consider a cell with
25   a single main (nuclear) genome and a variable number of mitochondrial genomes (mtDNA). If the
26   cell volume increases, the number of mitochondrial genomes increase proportionally, but the main
27   genome does not (see Supplementary Information for more details). The total number of
28   mitochondrial genomes (N_{mtDNA} in Fig. 4) is the number of mtDNA molecules per µm³ of
29   mitochondrial volume (n_{mtDNA}) (44–46) multiplied by the total mitochondrial volume of the cell
30   ($f_{mt}V$) (Eq. 6). We compiled data that show that the cell volume fraction occupied by mitochondria
31   (f_{mt}) ranges from 1–20% across diverse eukaryotes; our calculations thus use the geometric mean
32   of 4.4% (Fig. 5A; see Table S3). The number of mtDNA molecules per nucleoid (or per µm³ of
33   mitochondrial volume; N_{mtDNA}), and the size of the mitochondrial genome (L_{mtDNA}), were varied
34   between 1-100 and 10 and 70 kbp (49, 50), respectively, with negligible effects on the results (Fig.
35   S4). The total cost of DNA thus comprises the cost of the main genome and of all mitochondrial
36   genomes required to support the whole cell volume (Eq. 6). The relative cost of DNA for a
37   eukaryotic cell is calculated as above. If expressed as a function of cell volume, we have:

38

$$c_{DNA,euk} = \frac{101 \cdot L_{euk} + 101 \cdot L_{mtDNA} \cdot n_{mtDNA} \cdot f_{mt}V}{\alpha V^{0.97} + t_d \beta V^{0.88}}. \quad (6)$$
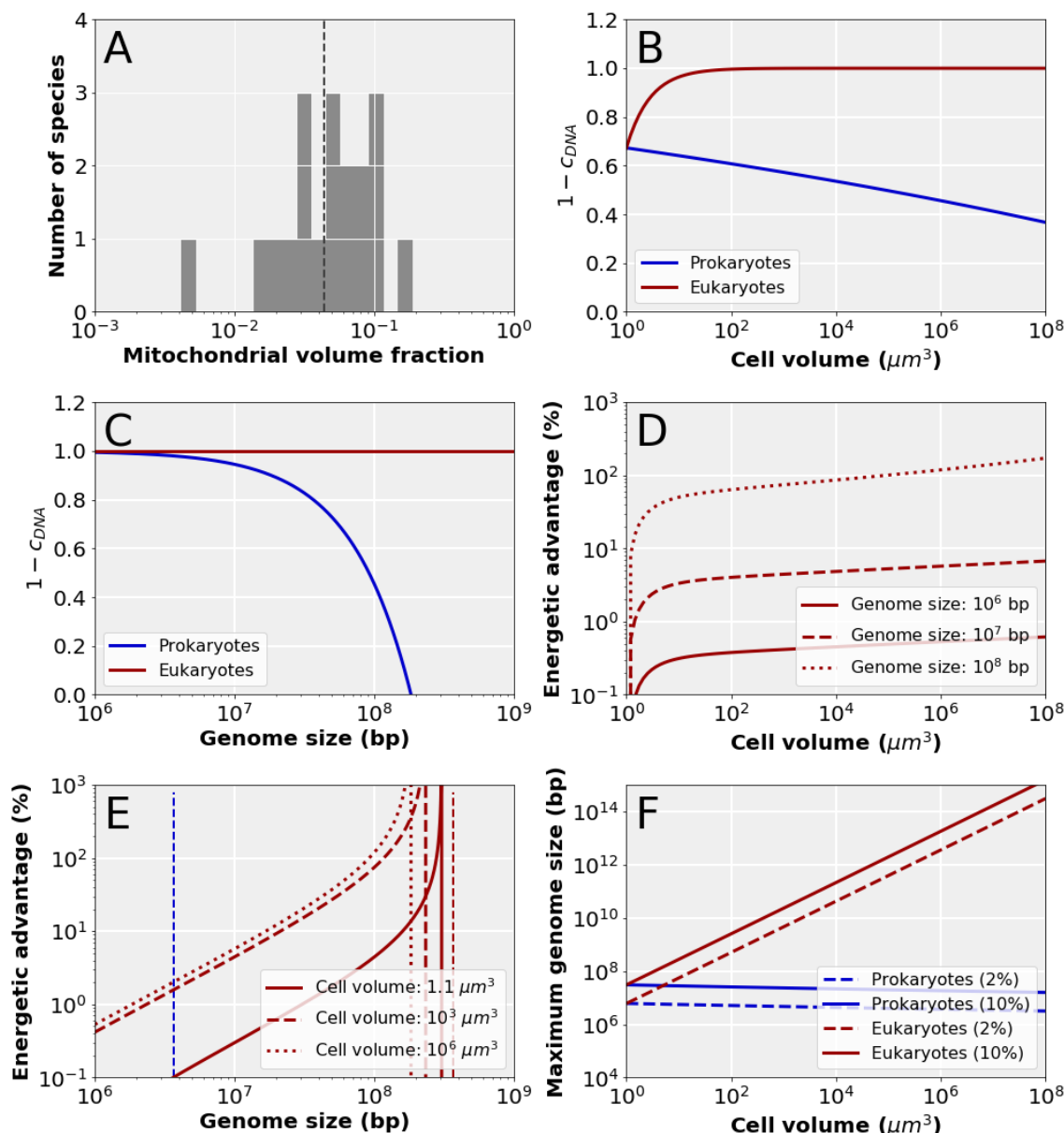
39

1

Figure 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 44% (Table S3). **B**. The relative cell energy budget available for cellular features other than DNA as a function of cell volume. The plot was calculated with Eq. 5 and 6 and $L_{euk} = L_{prok} = 10^8$ bp, $L_{mtDNA} = 70$ Kbp, $n_{mtDNA} = 10$, $t_d = 10$ h, and $V_{gserv} = 1$ μm^3 (these values were also used for C-F). **C**. The relative cell energy budget available for cellular features other than DNA as a function of genome size. The plot was calculated with Eq. 5 and 6 and $V = 10^6$ μm^3. **D**. The energetic advantage of eukaryotes over prokaryotes (red lines) as a function of cell volume. The plot was calculated with Eq. 4 and three different genome sizes as shown in inset. **E**. The 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 Eq. 4 and three different cell volumes as shown in inset. The blue and red dashed lines show the median genome sizes for prokaryotes and eukaryotes. **F**. The 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 Eq. 5 and 6.

11

1 Our model (Eq. 4-6; see also Table S1) allows us to compare the contrasting genomic architectures
2 of extant eukaryotes and prokaryotes across a range of cell volumes and genome sizes (Fig. 5).
3 Note that for these calculations, we kept the (main) genome size for eukaryotes equal to that of
4 prokaryotes (i.e., $L_{prok} = L_{euk}$) as we are only interested in determining whether the genomic
5 asymmetry of eukaryotes provides an advantage over prokaryotes. We also kept the cell division
6 time (t_d) equal for prokaryotes and eukaryotes; varying t_d from 0–100 h did not have major effects
7 on the calculated energetic advantage for eukaryotes (Fig. S4), and thus does not change our
8 conclusions. The main conclusions from our calculations are as follows. First, prokaryotes invest a
9 larger fraction of their ATP budget on DNA as their cells increase in volume and, as a result, are
10 left with less ATP for other processes such as gene expression (Fig. 5B). Second, the decrease in
11 ATP available for other cell functions in prokaryotes is more pronounced as genome size increases
12 (Fig. 5C). In contrast, eukaryotes suffer a negligible decline in their cellular ATP budget as their cell
13 volume or main genome size increase (Fig. 5B, 5C). Third, eukaryotes have an energetic
14 advantage (in terms of DNA cost savings) of less than ~200% for genome sizes of 10^6-10^8 bp and
15 across a cell volume range of eight orders of magnitude or 10^0-10^8 μm^3 (Fig. 5D; a volume of 10^0
16 μm^3 approximately corresponds to that of *Escherichia coli*, whereas a cell volume of 10^8 μm^3 is
17 similar to that of a giant single-celled species like *Chaos carolinensis* (51)). At genome sizes of 10^6-
18 10^7 bp, the energetic advantage of eukaryotes over prokaryotes is less than 10% across a similar
19 range in cell volume (Fig. 5E). Fourth, a prokaryote with a genome size of 3×10^7 bp, which is
20 characteristic of many single-celled eukaryotes (see below), would have an energetic disadvantage
21 of ~20% relative to a eukaryote with the same genome size. Such a genome size could, in principle,
22 accommodate 2×10^4 genes (assuming a mean gene length of 1,000 bp) while devoting about a
23 third of its size (~1×10^7 bp) to regulatory sequences. Fifth, prokaryotic genomes cannot get larger
24 than ~3×10^8 because the cost of DNA would exceed the total ATP budget of the cell (at any cell
25 volume). Eukaryotes, on the other hand, can achieve (main) genome sizes orders of magnitude
26 larger as cell volume increases (Fig. 5F). If 2–10% of the ATP budget of the cell is devoted to DNA
27 synthesis, prokaryotes can reach genomes of 6×10^6–3×10^7 bp in size (Fig. 5F).

28 To compare prokaryotes to eukaryotes at the onset of mitochondrial symbiosis, we assumed that
29 the ancestral mitochondrial genome size was as large as that of an average prokaryote (i.e.,
30 $L_{mtDNA} = L_{prok}$), and the mitochondrial volume fraction were much larger as mitochondrial function
31 was not yet optimized to aerobic respiration (i.e., $f_{mt} = 0.3$). However, this model (Eq. 6) assumes
32 that there was a dynamic cytoskeleton in place that allowed for active cytoplasmic transport and
33 therefore relinquished the need for genomes to scale with cell volume, as in modern eukaryotes
34 (see Discussion). This may best reflect a proto-eukaryote with a nucleus and cytoskeleton as
35 envisioned by some mitochondrion-late scenarios. To model ancestral eukaryotes as envisioned
36 by mitochondrion-early scenarios, we assumed that the number of genomes scaled with cell
37 volume (i.e., like that in prokaryotes), and also that the ancestral mitochondrial genome was much
38 larger and equivalent to that of an average prokaryote (i.e., $L_{mtDNA} = L_{prok}$). To do this, Eq. 6 was
39 modified in such a way that the number of genomes of ancestral eukaryotes is a function of the cell
40 volume not occupied by (pre-)mitochondria:

41
$$c_{DNA,proto-euk} = \frac{101 \cdot L_{euk} \cdot \left(\frac{(1-f_{mt})V}{V_{gserv}}\right) + 101 \cdot L_{mtDNA} \cdot n_{mtDNA} \cdot f_{mt}V}{\alpha V^{0.97} + t_d \beta V^{0.88}}. \quad (7)$$

42 The analyses show that a modern eukaryotic genomic architecture (which would also have been
43 present in proto-eukaryotes as postulated by mitochondrion-late scenarios) almost always gives an
44 energetic advantage to eukaryotes over prokaryotes (Fig. S5A, B). This is true even at much larger
45 mitochondrial genome sizes and volume fractions that must have been present at the onset of
46 mitochondrial symbiosis (Fig. S5C, D). On the other hand, if the ancestral host cell that took up the
47 (pre-)mitochondrial symbiont had a genomic architecture equivalent to that of prokaryotes (as
48 implied by mitochondrion-early scenarios), the first eukaryotes derived no energetic advantage
49 relative to prokaryotes (Fig. S5E). Both the first eukaryotes and prokaryotes invested the same
50 amount of energy in DNA synthesis. Further reductions in mitochondrial genome size would have

1   provided larger DNA cost savings to early eukaryotes. If ancestral mitochondria were slightly more
2   polyploid (i.e., had a higher genome copy number per µm³; $n_{mtDNA} = 3$), like some modern
3   intracellular symbionts that increase ploidy to physiologically support their host (e.g., see (52, 53)),
4   prokaryotes might have even had an energetic advantage and invested less ATP on DNA synthesis
5   compared to ancestral eukaryotes (Fig. S5F).

6   <u>The costs and benefits provided by a population of energy-producing symbionts</u>

7   Mitochondria, or respiring symbionts, expand the maximum volume that cells can achieve (Fig. 3),
8   though they are not required for many combinations of cell volumes and division times observed
9   among modern eukaryotes (Fig. 3D). However, symbionts and organelles also drain resources
10 from their host cells as they have both growth and maintenance costs. We investigated the
11 maximum volume afforded by, and costs associated with, a population of respiring symbionts in a
12 host cell. To do this, we conceived of a model that envisages an aerobically respiring host cell with
13 a population of one or more small spherical symbionts that respire aerobically. These respiring
14 symbionts rely on a slight excess of surface area devoted to aerobic respiration ($f_{max} = 0.1$ instead
15 of $f_{max} = 0.08$ as above, which leads to ~13% of ATP overproduction by a symbiont of 1 µm³ in
16 volume) to support themselves and partially contribute to the energy demands of their host cell;
17 this would have been true at early stages of mitochondrial symbiosis when symbiont function was
18 not yet evolutionarily optimized. The amount of resources (ATPs) diverted from the host cell are
19 directly proportional to the energy requirements dictated by the volume of each symbiont. The
20 advantage that the respiring symbionts confer to their host is here quantified as an increase in cell
21 volume and is ultimately a function of both the excess ATP they provide to the host cell and the
22 amount of ATP consumed to support their own volume (where the former includes only direct costs
23 and the latter both direct and opportunity costs).

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

30
$$f_{max\_sym} N_{sym} S_{sph} V_{sym}{}^{2/3} = \frac{f_d \alpha V_{supp\_sym}{}^{0.97}/t_d + \beta V_{supp\_sym}{}^{0.88}}{r} A_r, \quad (8)$$

31 can be solved numerically to obtain values of V_{supp_sym} for different cell division times and number
32 of symbionts. Here, f_{max_sym} is the maximum respiratory membrane fraction used by the symbiont
33 (=0.1), N_{sym} is the number of intracellular symbionts (which ranges from 1 and 4x10⁴), S_{sph} is the
34 shape factor for the volume of a sphere, and V_{sym} is the volume occupied by a single symbiont
35 (=0.25-1 µm³). As before, the parameter values for f_d, α, β, A_r, and r are constants (see Table S1).

36 To calculate the amount of cell volume that is supported by the host membrane (V_{supp_host}), we
37 follow a logic similar to that used above to determine V_{supp_sym}. We first express the amount of
38 respiratory membrane supplied by the host as a function of the host cell volume, and then as a
39 function of the number of respiratory units harbored by the host cell (i.e., the ratio between the

1 lifetime metabolic rate of the host-supported volume and the rate of a single ATP synthase). If these
2 two expressions are equated, we obtain the following expression:

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

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

6
$$V_{supp\_sym} + V_{supp\_host} = \left(\frac{f_d \alpha V_{supp\_host}{}^{0.97}/t_d + \beta V_{supp\_host}{}^{0.88}}{r f_{max} S_{sph}} A_r \right)^{2/3} \quad (10)$$

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

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

12 The analyses reveal that the larger the population of respiring symbionts, the larger the increase in
13 cell volume that is possible (Fig. 6). The increase in volume is large for fast-dividing cells, though
14 the cost of maintaining their symbiont population is also very high. On the other hand, slow-dividing
15 cells gain a smaller increase in volume, but the energetic cost of their symbiont populations is also
16 much lower. For example, a small and fast-growing cell ($t_d = 1$ h) that harbors respiring symbionts
17 derives a large advantage that allows its volume to increase substantially, but this also comes at a
18 steep cost of >30% (Fig. 6). A larger and slower-dividing cell (e.g., $t_d = 8$ h) has a relative symbiont
19 cost of up to two orders of magnitude lower (Fig. 6). If two cells have the same number of symbionts,
20 the slower-dividing cell will achieve a larger volume, and also has a lower symbiont cost. Thus, if a
21 larger and slower-dividing cell took up respiring symbionts, it would have been able to maintain or
22 increase its volume while having a comparatively low symbiont cost. A lower symbiont cost leaves
23 more energetic resources, i.e., ATP, for the rest of the cell volume, or leaves a larger volume

1   fraction for other cell functions. These analyses may provide some insights into the nature of the
2   ancestral host cell that took up symbionts at the origin of eukaryotes (see Discussion).
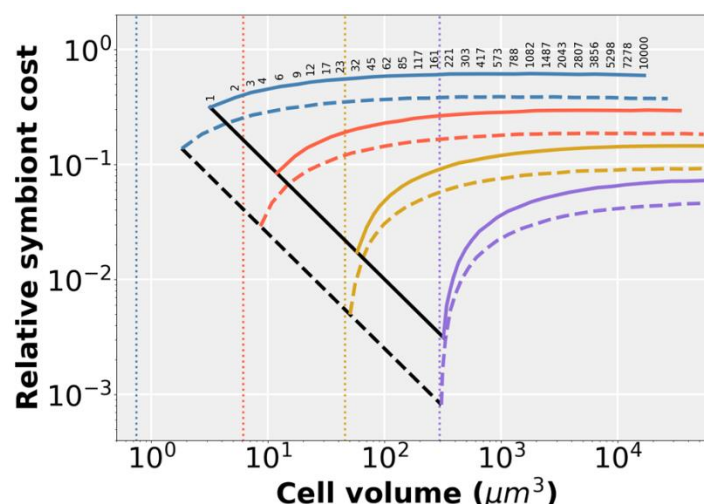
3

4   **Figure 6.** Costs and benefits of harboring ATP-exporting respiring symbionts. Relationship between maximum
5   cell volume and symbiont population cost for different symbiont population sizes at several cell division times.
6   The plot was generated from Eq. 8-11. Solid lines, $V_{sym} = 1$ μm³; dashed lines, $V_{sym} = 0.25$ μm³. Vertical
7   dotted lines show the host cell volumes in the absence of a symbiont population. The colors indicate the cell
8   division time: 1 h (blue), 2 h (orange), 4 h (gold), and 8 h (purple). For all the colored lines the cell division
9   time, t_d, is constant and the number of symbionts, N_{sym}, is varied. The numbers above the solid blue line
10   indicate the number of symbionts. For the black lines $N_{sym} = 1$ and t_d is varied from 1-8 h.

11   **Discussion**

12   The role of mitochondrial energetics in the origin and diversification of eukaryotes remains highly
13   contested (3, 10, 22, 26, 54). As an attempt to resolve this debate, we investigated the respiratory
14   deficit of mitochondrion-less cells and the maximum cell volume that can be supported by
15   respiration at the cell surface. We showed that the maximum volume that a cell can attain is
16   dependent on at least three major factors: cell body shape, cell division time, and maximum
17   respiratory membrane fraction. A combination of biologically plausible values for these factors may
18   allow mitochondrion-less cells to achieve volumes of up to 10^3-10^5 μm³ without a deficit in surface
19   area devoted to respiration (Fig. 3). Furthermore, we investigated the energetic consequences of
20   the contrasting genomic architectures of mitochondrion-less and mitochondrion-bearing cells. Our
21   results show that the asymmetrical genomic architecture of mitochondrion-bearing cells provide
22   slight energetic savings in DNA costs relative to mitochondrion-less cells across a wide range of
23   cell volumes and genome sizes (Fig. 5). The model further predicts that mitochondrion-less cells
24   can achieve a genome size of 3×10^7 bp, if they devote 10% of their ATP budget to DNA synthesis,
25   at an energetic disadvantage of 20% (or 1.2-fold) (Fig. 5E, 5F).

26   The upper cell volumes and genome sizes of mitochondrion-less cells can be predicted based on
27   energetic considerations, as done here. However, evolutionary success depends not only on the
28   energetic capacity of a cell to sustain its own features, but also on the selective or ecological
29   advantages conferred by such features. For example, a cell that has an energetic disadvantage by
30   investing a large proportion of energy in DNA (and thus less in ribosome biogenesis or growth) but
31   has a feature that confers a large ecological advantage (e.g., phagocytosis or antibiotic secretion)
32   may otherwise outcompete cells that invest less in DNA but lack such a feature. Similarly, the
33   reproductive disadvantage that may accompany longer cell division times in larger cells may be
34   overcome by ecological specialization to avoid competition. This is the sentiment behind some of
35   the criticisms of the energetic hypothesis for the origin of eukaryotic complexity previously raised

15

1   by others (24, 25). Moreover, even features that do not confer selective advantages large enough
2   to offset the selective disadvantages associated with their energetic costs can passively emerge
3   and be maintained in evolution. This becomes more likely in species with larger cells that have
4   longer cell division times and smaller effective population sizes. In such species, the relative
5   energetic cost of cellular features is lower, and the power of random genetic drift is stronger (10).

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

16   The maximum advantage of 3-fold for eukaryotes found here stems from the comparison of two
17   considerably different types of cells: mitochondrion-less and mitochondrion-bearing cells. Whereas
18   the former represents an average prokaryote, the latter arguably represents a derived (proto-
19   )eukaryote with a highly reduced mitochondrial genome and dynamic cytoskeleton. This is because
20   our model considers a mitochondrial genome that is less than 7% the size of the main genome (i.e.,
21   ≤70 Kbp, equivalent to that of jakobids and LECA; Last Eukaryote Common Ancestor), and
22   assumes a main nuclear genome whose copy number does not increase with a larger cell volume.
23   Such a reduced mitochondrial genome could only have evolved after the invention of a protein
24   import machinery that sped up gene transfer to the nucleus or main genome (i.e., by allowing import
25   of transferred gene products back into mitochondria). In addition, only the presence of active
26   intracellular transport (i.e., a dynamic cytoskeleton and motor proteins that bypassed diffusion
27   constraints) would have allowed the nuclear or main genome *not* to scale up with cell volumes
28   (unlike in prokaryotes (43, 48)). Thus, a great degree of evolutionary change (and time) separates
29   the two types of cells compared here. This suggests, then, that the energetic advantages between
30   immediate ancestor and descendant populations of proto-eukaryotes were necessarily much
31   smaller than the 3-fold energetic advantage for mitochondrion-bearing cells found here. Indeed,
32   using a model and a set of parameters that best reflects the genomic architecture of ancestral
33   eukaryotes at the onset of mitochondrial symbiosis (as viewed by mitochondrion-early scenarios)
34   shows that these early eukaryotes did not derive any energetic advantage relative to prokaryotes.
35   Eukaryotes only gained an energetic advantage after their mitochondrial genomes reduced
36   considerably and their main genome copy number was no longer required to scale with cell volume
37   (Fig. S5). Moreover, if the ancestral eukaryote that served as host for the pre-mitochondrial
38   symbionts possessed a dynamic cytoskeleton (as predicted by mitochondrion-late scenarios), the
39   energetic costs associated with DNA synthesis would have been slightly lower (Fig. S5).

40   Numerous cellular features other than mitochondria separate extant eukaryotes from prokaryotes.
41   Among these, the cytoskeleton and endomembranes provided major evolutionary advantages to
42   early proto-eukaryotes by synergistically enabling phagocytosis and intracellular digestion. These
43   adaptations massively increase the surface area by which nutrients can be taken up by cells (i.e.,
44   expanded nutritional membranes such as food and digestive vacuoles), and might compensate for
45   the absence of higher efficiency in energy harnessing (i.e., aerobic respiration in mitochondria).
46   Phagotrophs also have an ecological advantage by preying on their bacterial competitors.
47   Furthermore, a dynamic cytoskeleton with motors allowed for active intracellular transport, thereby
48   overcoming the diffusion constraints that burden prokaryotes. These advantages may have allowed
49   early proto-eukaryotes to achieve larger cell volumes entirely composed of active cytoplasm (as
50   opposed to most larger prokaryotes) in the absence of mitochondria. The recent discovery of a
51   phagocytic prokaryotes with a dynamic cytoskeleton and sizes of up to 10 µm in length (38) suggest
52   that the evolution of a larger cell volumes is possible in the absence of mitochondria. As shown
53   above, these larger cell volumes and their accompanying longer cell division times are also those

16

1 at which a host cell would benefit the most from harboring a population of small respiring symbionts
2 (Fig. 6).

3 Comparative genomic analyses have estimated that LECA had 4,431 gene domains (55), ~ 4,137–
4 5,938 gene families (56–58) or 7,447–21,840 genes (mean = 12,753) (11). This inferred number
5 of genes can be accommodated by a genome of ~20–50 Mbp in size that also devotes more than
6 a third of its size to regulatory and other non-coding DNA. Because nuclear DNA amounts scale
7 strongly with cell volume in eukaryotes as the power law $V = 1025.4 \cdot DNA^{0.97}$ (where V is in µm^3
8 and DNA in pg) (59–61), a haploid LECA with such a genome size may have had a cell volume of
9 ~23–57 µm^3 (i.e., the volume of a spherical cell of 3.5–4.8 µm in diameter). Indeed, these genome
10 and cell sizes are similar to those of small heterotrophic nanoflagellates such as jakobids and
11 malawimonads which also have the most ancestral-like mitochondrial genomes known (49, 62).
12 The gene number, genome size, and cell volume inferred for LECA fall within or is close to the
13 modern prokaryote-eukaryote overlap (i.e., 10^0–10^2 µm^3, 10^6–10^7 bp, and 4,000–13,000 genes),
14 and also encompass the cell volumes and genome sizes at which prokaryotes may not face a
15 shortage of surface area (Fig. 3) or a considerable energetic disadvantage due to increasing DNA
16 costs (Fig. 5). Thus, the prokaryote-eukaryote transition may have happened under these
17 conditions.

18 Even though our analyses suggest that mitochondrion-less cells may achieve relatively large
19 volumes and genome sizes under certain conditions, they also point at constraints that these
20 simpler cells inevitably face at even larger volumes or genome sizes. Because the amount of
21 respiratory membrane needed (i.e., number of ATP synthases) scales super-linearly with total
22 surface area (22) (Fig. 3C and S3), prokaryotes may experience a shortage of respiratory
23 membrane area at larger cell volumes (as long as their internal volumes remain active unlike in
24 giant bacteria). Eukaryotes, on the other hand, can maintain such a super-linear scaling and reach
25 much larger cell volumes by internalizing respiratory membranes in mitochondria. In other words,
26 mitochondria allow energy supply to continuously match energy demand at increasingly larger
27 volumes. Barring non-energetic constraints (e.g., DNA replication times), mitochondria may also
28 allow eukaryotes to have shorter cell division times and rounder (or less flattened) cell body shapes
29 than mitochondrion-less cells (e.g., prokaryotes) at comparable cell volumes. Furthermore, as
30 genome size increases, prokaryotes divest more and more of their ATP budgets to DNA synthesis,
31 due to their genomic symmetry. Therefore, the energetic advantage of eukaryotes over prokaryotes
32 increases with larger genome sizes. The maximum genome size that prokaryotes can theoretically
33 achieve is 3×10^8 bp if the entire ATP budget were devoted to DNA synthesis, or up to 3×10^7 bp at
34 10% of the ATP budget. In contrast, eukaryotes can drastically expand their genomes as their cell
35 volumes (and ATP budgets) grow larger, due to their genomic asymmetry. These theoretical
36 predictions are consistent with constraints on prokaryotes suggested by the cell volume and
37 genome size distributions (Fig. 2) and are at odds with the conclusions of Lynch and Marinov (10,
38 22).

39 **Conclusions**

40 It has been claimed that an energy gap underlies the large differences in size and complexity
41 between eukaryotic and prokaryotic cells. The proponents of this view further hold that the origin of
42 mitochondria was a pre-requisite for simple prokaryotic cells to bridge such a gap and evolve into
43 complex eukaryotic cells. Based on energetic considerations, we have shown here that prokaryotes
44 can theoretically achieve eukaryote-like cell volumes and genome sizes. These findings are
45 consistent with the modern prokaryote-eukaryote overlap in cell volumes and genome sizes.
46 Because LECA was probably a small heterotrophic flagellate similar to a modern jakobid or
47 malawimonad eukaryote, we suggest that the prokaryote-eukaryote transition did not necessarily
48 require an expansion of respiratory membranes or the savings in DNA costs that mitochondria can
49 provide. We also argue that the selective advantages conferred by mitochondria did not represent
50 a quantum leap in energy supply (or 'bioenergetic jump' (3)) at the origin of eukaryotes and were,
51 in principle, not different from those provided by other eukaryotic innovations, such as a dynamic
52 cytoskeleton or an endomembrane system. Mitochondria, however, became much more important

1 for increasingly larger and faster-dividing eukaryotic cells, and may have thus allowed eukaryotes
2 to successfully diversify and occupy novel adaptive zones throughout their evolutionary history.
3
4 **Acknowledgments**
5

10
11 **References**

12 1. T. Cavalier-Smith, The Neomuran Revolution and Phagotrophic Origin of Eukaryotes and
13   Cilia in the Light of Intracellular Coevolution and a Revised Tree of Life. *Cold Spring Harb*
14   *Perspect Biol* **6**, a016006 (2014).
15 2. R. Y. Stanier, M. Douderoff, E. Adelberg, *The microbial world*, 2nd Ed. (Prentice-Hall,
16   1963).
17 3. N. Lane, W. Martin, The energetics of genome complexity. *Nature* **467**, 929–934 (2010).
18 4. W. Martin, M. Müller, The hydrogen hypothesis for the first eukaryote. *Nature* **392**, 37–41
19   (1998).
20 5. T. Cavalier-Smith, Predation and eukaryote cell origins: A coevolutionary perspective. *The*
21   *International Journal of Biochemistry & Cell Biology* **41**, 307–322 (2009).
22 6. P. López-García, D. Moreira, The Syntrophy hypothesis for the origin of eukaryotes
23   revisited. *Nature Microbiology* **5**, 655–667 (2020).
24 7. D. A. Baum, B. Baum, An inside-out origin for the eukaryotic cell. *BMC Biology* **12**, 76
25   (2014).
26 8. L. Sagan, On the origin of mitosing cells. *J Theor Biol* **14**, 255–274 (1967).
27 9. R. Y. Stanier, Some aspects of the biology of cells and their possible evolutionary
28   significance in *Symp Soc Gen Microbiol*, (1970), pp. 1–38.
29 10. M. Lynch, G. K. Marinov, The bioenergetic costs of a gene. *PNAS* **112**, 15690–15695
30   (2015).
31 11. J. Vosseberg, *et al.*, Timing the origin of eukaryotic cellular complexity with ancient
32   duplications. *Nature Ecology & Evolution* **5**, 92–100 (2021).
33 12. A. A. Pittis, T. Gabaldón, Late acquisition of mitochondria by a host with chimaeric
34   prokaryotic ancestry. *Nature* **531**, 101–104 (2016).
35 13. I. Zachar, E. Szathmáry, Breath-giving cooperation: critical review of origin of mitochondria
36   hypotheses. *Biology Direct* **12**, 19 (2017).
37 14. T. Vellai, K. Takács, G. Vida, A New Aspect to the Origin and Evolution of Eukaryotes. *J*
38   *Mol Evol* **46**, 499–507 (1998).
39 15. T. Vellai, G. Vida, The origin of eukaryotes: the difference between prokaryotic and
40   eukaryotic cells. *Proc Biol Sci* **266**, 1571–1577 (1999).
41 16. N. Lane, *Power, Sex, Suicide: Mitochondria and the Meaning of Life* (Oxford University
42   Press, USA, 2006).
43 17. N. Lane, Energetics and genetics across the prokaryote-eukaryote divide. *Biology Direct* **6**,
44   35 (2011).
45 18. N. Lane, Bioenergetic Constraints on the Evolution of Complex Life. *Cold Spring Harb*
46   *Perspect Biol* **6**, a015982 (2014).
47 19. N. Lane, How energy flow shapes cell evolution. *Current Biology* **30**, R471–R476 (2020).
48 20. K. Chiyomaru, K. Takemoto, Revisiting the hypothesis of an energetic barrier to genome
49   complexity between eukaryotes and prokaryotes. *Royal Society Open Science* **7**, 191859
50   (2020).
51 21. A. Booth, W. F. Doolittle, Eukaryogenesis, how special really? *PNAS* **112**, 10278–10285
52   (2015).
53 22. M. Lynch, G. K. Marinov, Membranes, energetics, and evolution across the prokaryote-
54   eukaryote divide. *eLife* **6**, e20437 (2017).

23. E. Szathmáry, Toward major evolutionary transitions theory 2.0. *PNAS* **112**, 10104–10111 (2015).

24. T. Cavalier-Smith, E. E.-Y. Chao, Multidomain ribosomal protein trees and the planctobacterial origin of neomura (eukaryotes, archaebacteria). *Protoplasma* **257**, 621–753 (2020).

25. V. Hampl, I. Čepička, M. Eliáš, Was the Mitochondrion Necessary to Start Eukaryogenesis? *Trends Microbiol* **27**, 96–104 (2019).

26. M. Lynch, G. K. Marinov, Reply to Lane and Martin: Mitochondria do not boost the bioenergetic capacity of eukaryotic cells. *PNAS* **113**, E667–E668 (2016).

27. M. W. Gray, *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 Biology* **18**, 22 (2020).

28. J. S. Huxley, *Problems Of Relative Growth* (Methuen And Company Limited., 1935) (March 27, 2021).

29. D. W. Thompson, *On Growth and Form* (Cambridge University Press, 1992).

30. O. Snell, Die Abhängigkeit des Hirngewichtes von dem Körpergewicht und den geistigen Fähigkeiten. *Archiv f. Psychiatrie* **23**, 436–446 (1892).

31. S. A. Muñoz-Gómez, J. G. Wideman, A. J. Roger, C. H. Slamovits, The Origin of Mitochondrial Cristae from Alphaproteobacteria. *Mol. Biol. Evol.* **34**, 943–956 (2017).

32. G. Mahmoudabadi, R. Phillips, M. Lynch, R. Milo, Defining the Energetic Costs of Cellular Structures. *bioRxiv*, 666040 (2019).

33. C. Etzold, G. Deckers-Hebestreit, K. Altendorf, Turnover number of Escherichia coli F0F1 ATP synthase for ATP synthesis in membrane vesicles. *Eur J Biochem* **243**, 336–343 (1997).

34. K. Valgepea, K. Adamberg, A. Seiman, R. Vilu, Escherichia coli achieves faster growth by increasing catalytic and translation rates of proteins. *Mol Biosyst* **9**, 2344–2358 (2013).

35. M. Szenk, K. A. Dill, A. M. R. de Graff, Why Do Fast-Growing Bacteria Enter Overflow Metabolism? Testing the Membrane Real Estate Hypothesis. *Cell Systems* **5**, 95–104 (2017).

36. M. Lindén, P. Sens, R. Phillips, Entropic Tension in Crowded Membranes. *PLOS Computational Biology* **8**, e1002431 (2012).

37. J. Larsen, D. J. Patterson, Some flagellates (Protista) from tropical marine sediments. *Journal of Natural History* **24**, 801–937 (1990).

38. T. Shiratori, S. Suzuki, Y. Kakizawa, K. Ishida, Phagocytosis-like cell engulfment by a planctomycete bacterium. *Nature Communications* **10**, 5529 (2019).

39. H. N. Schulz, B. B. Jorgensen, Big bacteria. *Annu Rev Microbiol* **55**, 105–137 (2001).

40. H. N. Schulz, *et al.*, Dense Populations of a Giant Sulfur Bacterium in Namibian Shelf Sediments. *Science* **284**, 493–495 (1999).

41. K. D. Clements, S. Bullivant, An unusual symbiont from the gut of surgeonfishes may be the largest known prokaryote. *J Bacteriol* **173**, 5359–5362 (1991).

42. M. Schlame, Protein crowding in the inner mitochondrial membrane. *Biochim Biophys Acta Bioenerg* **1862**, 148305 (2021).

43. D. Ionescu, M. Bizic, "Giant Bacteria" in *ELS*, (American Cancer Society, 2019), pp. 1–10.

44. R. Jajoo, *et al.*, Accurate concentration control of mitochondria and nucleoids. *Science* **351**, 169–172 (2016).

45. C. Kukat, *et al.*, Super-resolution microscopy reveals that mammalian mitochondrial nucleoids have a uniform size and frequently contain a single copy of mtDNA. *PNAS* **108**, 13534–13539 (2011).

46. H. S. Ilamathi, *et al.*, Mitochondrial fission is required for proper nucleoid distribution within mitochondrial networks. *bioRxiv*, 2021.03.17.435804 (2021).

47. R. Ohbayashi, *et al.*, Coordination of Polyploid Chromosome Replication with Cell Size and Growth in a Cyanobacterium. *mBio* **10**, e00510-19 (2019).

48. J. E. Mendell, K. D. Clements, J. H. Choat, E. R. Angert, Extreme polyploidy in a large bacterium. *PNAS* **105**, 6730–6734 (2008).

49. A. J. Roger, S. A. Muñoz-Gómez, R. Kamikawa, The Origin and Diversification of Mitochondria. *Current Biology* **27**, R1177–R1192 (2017).

50. J. Janouškovec, *et al.*, A New Lineage of Eukaryotes Illuminates Early Mitochondrial Genome Reduction. *Current Biology* **27**, 3717-3724.e5 (2017).

51. T. Fenchel, B. J. Finlay, Respiration rates in heterotrophic, free-living protozoa. *Microb Ecol* **9**, 99–122 (1983).

52. K. Komaki, H. Ishikawa, 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).

53. P. Mergaert, *et al.*, Eukaryotic control on bacterial cell cycle and differentiation in the Rhizobium-legume symbiosis. *Proc Natl Acad Sci U S A* **103**, 5230–5235 (2006).

54. N. Lane, W. F. Martin, Mitochondria, complexity, and evolutionary deficit spending. *PNAS* **113**, E666–E666 (2016).

55. C. M. Zmasek, A. Godzik, Strong functional patterns in the evolution of eukaryotic genomes revealed by the reconstruction of ancestral protein domain repertoires. *Genome Biology* **12**, R4 (2011).

56. K. S. Makarova, Y. I. Wolf, S. L. Mekhedov, B. G. Mirkin, E. V. Koonin, Ancestral paralogs and pseudoparalogs and their role in the emergence of the eukaryotic cell. *Nucleic Acids Res* **33**, 4626–4638 (2005).

57. L. K. Fritz-Laylin, *et al.*, The genome of Naegleria gruberi illuminates early eukaryotic versatility. *Cell* **140**, 631–642 (2010).

58. D. Newman, F. J. Whelan, M. Moore, M. Rusilowicz, J. O. McInerney, Reconstructing and Analysing The Genome of The Last Eukaryote Common Ancestor to Better Understand the Transition from FECA to LECA. *bioRxiv*, 538264 (2019).

59. B. J. Shuter, J. E. Thomas, W. D. Taylor, A. M. Zimmerman, Phenotypic Correlates of Genomic DNA Content in Unicellular Eukaryotes and Other Cells. *The American Naturalist* **122**, 26–44 (1983).

60. T. Cavalier-Smith, M. J. Beaton, The skeletal function of non-genic nuclear DNA: new evidence from ancient cell chimaeras. *Genetica* **106**, 3–13 (1999).

61. T. Cavalier-Smith, Economy, Speed and Size Matter: Evolutionary Forces Driving Nuclear Genome Miniaturization and Expansion. *Ann Bot* **95**, 147–175 (2005).

62. G. Burger, M. W. Gray, L. Forget, B. F. Lang, Strikingly bacteria-like and gene-rich mitochondrial genomes throughout jakobid protists. *Genome Biol Evol* **5**, 418–438 (2013).