HYPOTHESIS

doi:10.1038/nature09486

The energetics of genome complexity

Nick Lane[1] & William Martin[2]

All complex life is composed of eukaryotic (nucleated) cells. The eukaryotic cell arose from prokaryotes just once in four billion years, and otherwise prokaryotes show no tendency to evolve greater complexity. Why not? Prokaryotic genome size is constrained by bioenergetics. The endosymbiosis that gave rise to mitochondria restructured the distribution of DNA in relation to bioenergetic membranes, permitting a remarkable 200,000-fold expansion in the number of genes expressed. This vast leap in genomic capacity was strictly dependent on mitochondrial power, and prerequisite to eukaryote complexity: the key innovation en route to multicellular life.

Despite boundless biochemical ingenuity, prokaryotes have not evolved morphological complexity beyond the rudimentary level seen in cyanobacteria or planctomycetes in four billion years of evolution. In contrast, complex multicellular organisms have evolved independently in at least six different eukaryotic groups[1]. Eukaryotic cells are generally larger and more highly structured than prokaryotic cells, with much bigger genomes and proteomes, but the critical difference enabling that complexity has remained elusive. Virtually every 'eukaryotic' trait is also found in prokaryotes, including nucleus-like structures[2], recombination[3], linear chromosomes[4], internal membranes[5], multiple replicons[6], giant size[7], extreme polyploidy[8], dynamic cytoskeleton[9], predation[10], parasitism[11], introns and exons[12], intercellular signalling[13] (quorum sensing), endocytosis-like processes[14] and even endosymbionts[15,16]. Bacteria made a start up virtually every avenue of eukaryotic complexity, but then stopped short. Why?

Population genetic approaches addressing the evolutionary divide between prokaryotes and eukaryotes focus on population size: if the first eukaryotes were few in numbers, they could tolerate many new mutations and a larger genome size through weakened purifying selection[17]. But why do prokaryotes with small population sizes not tend to become eukaryotic? If the constraint was circular chromosomes[18], why didn't bacteria with straight chromosomes and multiple replicons become complex? If phagocytosis offered the decisive advantage[19,20], why didn't eukaryotes evolve repeatedly from prokaryotes for the same reasons?

The answer hinges upon the uniqueness of eukaryote origins. All eukaryotes share a common ancestor, which arose from prokaryotes just once in four billion years. Genomic chimaerism points to the origin of eukaryotes in an endosymbiosis between prokaryotes[21–24]. All eukaryotes either possess mitochondria, or once did and later lost them[25,26], placing the origin of mitochondria and the eukaryotic cell as plausibly the same event[27]. Was the acquisition of mitochondria the critical step towards eukaryote genome complexity? If so, what salient advantage did they confer?

It is not aerobic respiration—many mitochondria are anaerobic[28], and many free-living prokaryotes aerobic[29]. Whereas mitochondria enabled aerobic respiration in large eukaryotes, and oxygen is all but essential for multicellular life, oxygen itself cannot explain why there are no aerobic multicellular prokaryotes more complex than cyanobacteria. Mitochondria did not protect their host cell against an 'oxygen catastrophe'[30]. There is no evidence for such a catastrophe in the geological record, or in microbial phylogeny; anaerobes are not a branch of microbial diversity, either eukaryotic or prokaryotic. Higher oxygen levels did not scour the oceans of anaerobes, but produced sulphidic oceans, which persisted for more than a billion years[31]. Oxygen is not reactive in the absence of single-electron donors (hence its accumulation in the air); but single-electron donors are ubiquitous in mitochondria, making them anything other than protective. Mitochondria do not even increase respiratory rate: gram for gram, many prokaryotes respire faster than eukaryotes[32,33]. Mitochondria do compartmentalize respiration within the cell[34]; but prokaryotes can compartmentalize themselves too, and some respire over locally invaginated membranes[5]. With faster respiration and internal compartments, bacteria would even seem to have an energetic advantage over mitochondrion-bearing cells. Why did they not realize that advantage, why did only mitochondrion-bearing cells evolve true complexity?

The answer, we posit, resides ultimately in mitochondrial genes. By enabling oxidative phosphorylation across a wide area of internal membranes, mitochondrial genes enabled a roughly 200,000-fold rise in genome size compared with bacteria. Whereas the energetic cost of possessing genes is trivial, the cost of expressing them as protein is not and consumes most of the cell's energy budget. Mitochondria increased the number of proteins that a cell can evolve, inherit and express by four to six orders of magnitude, but this requires mitochondrial DNA. How so? A few calculations are in order.

Energy per gene expressed

The massive difference in mean genome size between prokaryotes and eukaryotes is most revealingly quantified in terms of energy available per gene. By 'energy per gene', we mean the cost of expressing the gene. The cost of DNA replication itself accounts for just 2% of the energy budget of microbial cells during growth[29]. In contrast, protein synthesis accounts for a remarkable ~75% of a cell's total energy budget[29]. If the bacterial genome is increased tenfold in size, the cost of replicating the genome itself would still only account for about 20% of the cell's existing energy budget (although 100 times more DNA would treble the cell's energy budget, and 1,000 times more DNA would raise the energy budget 20-fold, so copying the DNA of a eukaryote-sized genome would be a serious cost for bacteria). But the most immediate and pressing constraint of increasing genome size even tenfold is that ten times as many proteins would need to be expressed[35]. If *Escherichia coli* had 44,000 proteins instead of 4,400, it would need to allocate a portion of its 75% protein energy dedication to the synthesis of these new proteins. *E. coli* normally devotes on average 0.017% of its total energy budget to each protein. If it could halve this expenditure, only 9×10^{-4}% of the energy budget could be dedicated to each of the 40,000 new proteins, a mere 1/20th that for each pre-existing protein, hardly a viable proposition. And were the

[1]Department of Genetics, Evolution and Environment, University College London, Gower Street, London W1E 6BT, UK. [2]Institut für Botanik III Heinrich-Heine-Universität, Düsseldorf, Universitätsstrasse 1, 40225 Düsseldorf, Germany.

energy budget for existing proteins halved, which genes should compensate via reduced expression? A cell that halved its ribosome number, carbon metabolism or respiratory chain would hardly prosper. Plainly, to raise gene number tenfold, *E. coli* must also increase its energy budget by close to tenfold; and therein lies the problem.

This energetic barrier cannot be circumvented by synthesizing regulatory proteins at low copy number, as ribosomal numbers attest. An average bacterium such as *E. coli* has up to 13,000 ribosomes[36], whereas a human liver cell has 13 million on the rough endoplasmic reticulum alone[37]—1,000 to 10,000-fold more. This large difference entails energetic costs that are orders of magnitude higher in eukaryotic cells. Eukaryotes are enormously expanded in cell volume, their cytosol is packed with a massive variety of highly expressed structural proteins (Fig. 1). If a bacterial cell could add thousands of new regulatory proteins produced at very few copies per cell, the additional energetic cost might be affordable and an advance in complexity imaginable. But a regulatory protein arsenal of eukaryotic dimensions is unknown in prokaryotes, for without the additional structural and behavioural complexity of eukaryotic cells, there is no need for additional regulation.

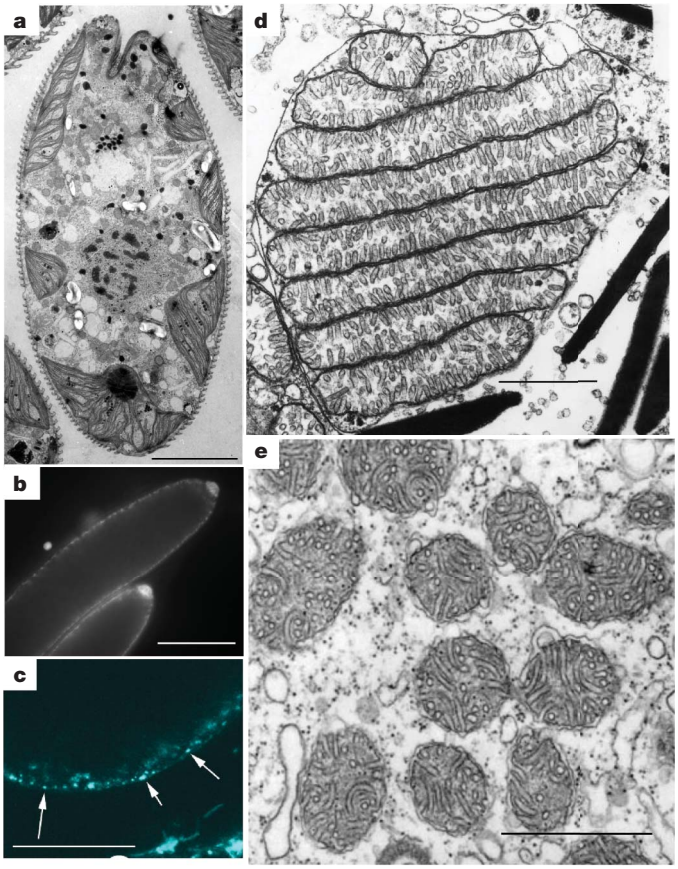

Figure 1 | Cell complexity, cell simplicity and energy supply for both. **a**, Transmission electron micrograph of a eukaryote, a complex cell, the protist *Euglena gracilis* (scale bar, 5 μm). **b**, **c**, Fluorescence micrographs of DAPI-stained giant prokaryotes *Epulopiscium fishelsoni* (**b**) and *Thiomargarita namibiensis* (**c**) (scale bars in **b** and **c**, 50 μm). Although the prokaryotes are 5–20 times larger than the eukaryote (see scale bars), they lack true complexity. Their nucleoids (active chromosomes: fluorescent white dots in **b**, white arrows in **c**) are tightly co-localized with the plasma membrane, the site of chemiosmotic ATP synthesis in prokaryotes[29,51]. The dark area above the nucleoids in **c** is a large vacuole. **d**, **e**, Transmission electron micrographs of mitochondria, site of chemiosmotic ATP synthesis in eukaryotes[29,51]. All mitochondria retain core genomes of their own, which are necessary for the control of membrane potential across a circumscribed area of membrane, enabling a 10^4–10^5-fold increase in the total area of internalized bioenergetic membrane. **d**, A single folded mitochondrion in the dinoflagellate *Oxyrrhis marina* (osmium-fixed). **e**, Multiple mitochondria in the ciliate *Paramecium bursaria* (collidine buffer-fixed) (scale bars in **d** and **e**, 1 μm). Photos: **a**, **d**, M. Farmer; **b**, E. Angert; **c**, H. Schulz-Vogt; **e**, R. Allen.

The calculations below chart the bioenergetic discrepancy between bacteria and eukaryotes, per gene expressed, assuming their copy number remains roughly constant. For clarity we assume haploid genomes here, but take ploidy into consideration in Table 1.

Actively growing proteobacteria have a mean metabolic rate of $0.19 \pm 0.5\ W\ g^{-1}$ ($1\ W = 1\ J\ s^{-1}$) and mass of 2.6×10^{-12} g (based on a mean of 55 samples[32]). Actively growing protozoa have a mean metabolic rate of $0.06 \pm 0.1\ W\ g^{-1}$ and mass of $40{,}100 \times 10^{-12}$ g (based on 12 samples[32,33]). Consider metabolic rates per cell. The average rate for proteobacteria is 0.49 pW per cell; for protozoa 2,286 pW. Because the metabolic rate per gram is not particularly different (a factor of three) but cell size is very different (a factor of 15,000) an average protozoan has nearly 5,000 times more metabolic power (W) than a single bacterium.

Yet the metabolic power per megabase (Mb) of DNA is similar in bacteria and protozoa, to within an order of magnitude. Assuming 6 Mb of DNA, an average bacterium has about $0.08\ pW\ Mb^{-1}$. The 'average protozoan' has a power of $0.76\ pW\ Mb^{-1}$. If the bacterial genome is small (in the same size cell), for example 1.5 Mb, this difference shrinks, with the bacterium having $0.33\ pW\ Mb^{-1}$. Conversely, if the protozoan genome is large (30,000 Mb in the case of *Amoeba proteus*, which weighs about $1{,}000{,}000 \times 10^{-12}$ g; ref. 32) the power per cell is 57 nW, giving $0.19\ pW\ Mb^{-1}$, again similar to an 'average' protozoan. These mean and range values are similar to the specific derivations calculated from actual cell size, ploidy and metabolic rate in Table 1.

In other words, despite the fact that bacteria have a faster metabolic rate per gram than protozoa, their small size disguises the fact that the power dedicated to each Mb of DNA has remained roughly constant, to within one or two orders of magnitude, whereas eukaryotic genome sizes have expanded relative to prokaryotes by at least 10,000-fold. Protozoa and multicellular eukaryotes face no bioenergetic penalty for having enormously expanded genome sizes. On the contrary, despite having much larger genomes, eukaryotes tend to have more power per Mb of DNA than prokaryotes.

This conclusion is true even when taking into consideration the genomic weight of mitochondrial DNA (mtDNA) in eukaryotes. Over evolutionary time, mtDNA has been whittled away to between 6 kilobase (kb) and 77 kb in protozoa[38]. Taking an average mitochondrial genome of 30 kb in 200,000 mitochondria (as in large amoebae[39]), the total mtDNA content per cell is 6,000 Mb, twice the size of the 'average' haploid nuclear genome; or 9,000 Mb in total. The genomic power corresponds to $0.25\ pW\ Mb^{-1}$: more than most bacteria, yet unlike bacteria sustaining a nuclear genome of 3,000 Mb.

Consider what happens if the genome size of a bacterium is scaled up to that of an average protozoan, 3,000 Mb, without scaling up metabolic rate—a reasonable assumption if cell size is kept constant, as prokaryotes respire over their plasma membrane. The power falls to $0.16\ fW\ Mb^{-1}$, some 4,600 times less than the protozoan. The situation is similar for gene number. An average bacterial genome contains about 5,000 genes, compared with about 20,000 in an average protozoan (ranging up to 40,000 in *Paramecium*[40]). At a metabolic rate of 0.49 nW per cell, a bacterium with 5,000 genes would have 0.1 fW per gene. At a power of 2,286 pW per cell and 20,000 genes, an average protist has 115 fW per gene, giving each eukaryotic gene 1,200 times more power (W) than its counterpart in even the most energetic bacteria. If bacterial gene number is scaled up to the size of the average protist (20,000 genes) the metabolic power per gene falls to 0.03 pW per gene: 4,600 times less than the eukaryote. For a haploid, in the presence of the same terminal electron acceptor (oxygen or otherwise), a eukaryotic nuclear gene governs nearly 5,000 times more energy flux than a prokaryotic gene. Again, these derivations are consistent with the values in Table 1.

Large size benefits eukaryotes but not prokaryotes

Eukaryotic cells are usually far larger than bacteria in physical size as well as genome size, and this too affects metabolic power per gene. Consider

Table 1 | Energetics of bacteria and eukaryotes by cell and genome size

	Prokaryotes					Eukaryotes				
Parameter	Mean	S	M	L	XL	Mean	S	M	L	XL
Weight of cell ($\times 10^{-12}$ g)	2.6	0.2	1.2	4	1×10^6	40,100	250	7,000	33,000	1×10^6
Power (W g^{-1})	0.19	0.07	0.3	0.11	0.0005	0.06	0.09	0.03	0.05	0.01
Power per cell (pW)	0.49	0.014	0.36	0.44	500	2,286	21.5	224	1,782	10,000
Ploidy level	4	1	6	4	10,000	2	2	2	100	3
Haploid genome size (Mb)	6	1.9	4.6	9	7.5	3,000	300	3,000	100	11,000
Power per haploid Mb (pW)	0.02	0.01	0.01	0.01	0.01	0.38	0.04	0.04	0.18	0.3
No. of haploid genes $\times 10^3$	5	2	4.4	6	6	20	12	20	25	15
Power per gene (fW)	0.03	0.01	0.01	0.02	0.01	57.15	0.90	5.6	0.71	222.2
Power per genome (fW)	0.12	0.01	0.06	0.11	0.05	1,143	10.75	112	17.8	3,333

For prokaryotes, the mean is from 55 values given in ref. 32; specific examples are derived from ref. 32, Supplementary data. For eukaryotes, the mean is from 12 values re-calculated independently from ref. 33; specific examples from data given in Table 1, ref. 33. We have converted from nl O_2 per cell per hour to watts using the same conversion factor as Makarieva *et al.*[31] (complete aerobic oxidation of endogenous substrates yields 20 J per ml O_2). Metabolic rate for *Thiomargarita namibiensis* is from ref. 73. The standard deviations in metabolic rate per gram (given in main text) are not transformed further here, but the variance of around twice the mean falls significantly short of the differences calculated. There is an appreciable range of uncertainty in measurement for both cell mass and metabolic rates for microbes: values differing by one or two orders of magnitude might not be meaningfully different. Nonetheless, differences of four to six orders of magnitude, as calculated, certainly are. Power per gene depends partly on ploidy. Very high ploidy, as in *Thiomargarita* (Schulz-Vogt, personal communication) and to a lesser extent *Bresslaua insidiatrix*[74], lowers energy per gene. Genome sizes are from the Joint Genome Institute (http://img.jgi.doe.gov/cgi-bin/pub/main.cgi). For prokaryotes: S, small (*Streptococcus pyogenes*); M, medium-sized (*Escherichia coli*); L, large (*Azospirillum lipoferum*); XL, very large (*Thiomargarita namibiensis*). For eukaryotes: S, small (*Ochromonas sp.*); M, medium-sized (*Euglena gracilis*); L, large (*B. insidiatrix*); XL, very large (*Amoeba proteus*). Power per genome is power per haploid gene times haploid gene number.

an average bacterium that is scaled up to the cell volume of an average protozoan. Because ATP synthesis scales with plasma membrane surface area but protein synthesis scales with cell volume, larger prokaryotic cells are energetically less efficient[41]. For simplicity, consider cells as spheres, the bacterium with a radius of 1 μm (surface area over volume ratio (SA/V) = 3 μm^{-1}), the protozoan with radius 50 μm (SA/V = 0.06 μm^{-1}), a fall in SA/V of 50-fold. Were the bacterium scaled up to the same volume, shape and genome size as an average protozoan, it would now have available 0.003 fW Mb^{-1}, a factor of 250,000 less than the protozoan. In terms of energy per gene, the scaled-up bacterium could muster a mere 0.0005 fW per gene, a 230,000-fold reduction.

Put another way, a eukaryotic gene commands some 200,000 times more energy than a prokaryotic gene, or at a similar energy per gene, the eukaryote could in principle support a genome 200,000 times larger. The implications for complexity can hardly be overstated. Whereas prokaryotes frequently make a start towards eukaryotic complexity, they rarely exhibit more than one complex eukaryotic trait at a time. This is because each trait has energy costs in terms of evolving and expressing novel protein families, and unless these costs can be met generously, complexity is counter-selected for energetic reasons. The prokaryote-to-eukaryote transition involved the origin of a multiplicity of new complex traits underpinned by some 3,000 new protein families[42,43]. That evolutionary leap required energy to burn, orders of magnitude more energy than any prokaryote can offer. Mitochondria bestowed upon their host 10^5–10^6 times more power per gene.

For four billion years bacteria have remained in a local minimum in the complexity fitness landscape, a deep canyon bounded on all sides by steep energetic constraints. The possession of mitochondria enabled eukaryotes to tunnel through this mountainous energetic barrier. Mitochondria allowed their host to evolve, explore and express 200,000-fold more genes with no energetic penalty. This is because mitochondria obliterated the heavy selection pressure to remove superfluous DNA (and potential proteins), which is among the most pervasive selective forces in prokaryote genome evolution[44–46].

Eukaryotes harbour approximately 12 genes per Mb, compared with about 1,000 in bacteria. If an average bacterium had a eukaryotic gene density, at 6 Mb of DNA it would encode fewer than 100 genes. With only 0.08 pW Mb^{-1}, it lacks the energy to support much regulatory or non-coding DNA. Bacteria must therefore maintain high gene density, around 500–1,000 genes per Mb, and do so by eliminating intergenic and intragenic material, including regulatory elements and microRNAs, by organizing genes into operons, and by restricting the median length of proteins[47]—all of which reduce the energetic costs. The high gene density and small protein size of bacteria can be explained in bioenergetic terms. In comparison, at a gene density of 12 genes per Mb and a metabolic power of 0.76 pW Mb^{-1}, an average protozoan could in principle sustain nearly 350,000 genes, allowing it to evolve, express and explore novel genes and gene families, increase the size of proteins[47], and invest freely in regulatory microRNAs[48].

These calculations give a thrillingly large empirical benefit to having mitochondria. But if the compartmentalization of energy coupling within cells gives such a massive energetic advantage, why don't prokaryotes just compartmentalize themselves? Many prokaryotes, including cyanobacteria and many nitrifying bacteria, do have extensively invaginated internal membranes[5]. What stopped them from becoming more complex, like eukaryotes?

Mitochondrial genes, key to nuclear genome expansion

Mitochondria that generate ATP by oxidative phosphorylation always retain a core genome that encodes proteins of the respiratory electron-transport chain[38]. This small genome holds the key to successful internalization of bioenergetic membranes. Mitochondria that lose their genome (hydrogenosomes and mitosomes) lose the ability to synthesize ATP by chemiosmotic coupling[25,26].

Mitochondria must respond quickly to changes in membrane potential and the penalty for any failure to do so is serious. The electron and proton transfers of chemiosmotic energy coupling generate a transmembrane potential of 150–200 mV over the membrane (~5 nm across), giving a field strength of about 30 million volt per metre, equal to that discharged by a bolt of lightning. This high membrane potential sets the inner membrane of bioenergetic organelles (mitochondria and chloroplasts) apart from all other eukaryotic membrane systems. Failure to maintain the mitochondrial membrane potential is penalized by a collapse in energy charge, blocking active transport across the cell membrane, and a rise in free-radical leak, which in eukaryotes and many prokaryotes leads directly to programmed cell death[49].

By encoding proteins of the respiratory chain, mtDNA allows individual mitochondria to respond, by gene expression, to changes in membrane potential. According to the CORR hypothesis (co-location for redox regulation[50,51]), this is the selective pressure that maintains DNA in bioenergetic organelles, in turn predicting that the presence of mtDNA should correlate with respiratory capacity and ATP availability. Respiration rates do correlate with the amount of mtDNA in the cell[52–54], and mutations that deplete mtDNA usually cause mitochondrial diseases[55]. Oxidative phosphorylation is under tight control by the amount of mtDNA in the cell, and the full complement of mtDNA is necessary to maintain a normal energy production level[56,57]. In mammals, the rate-limiting step in the assembly of new respiratory complexes (which determines respiratory rate) is the rate of transcription of the ND5 subunit of NADH dehydrogenase, encoded by mtDNA[58,59]. The expression of mtDNA-encoded complex I genes is sensitive to changes in oxygen tension, with a specific downregulation of ND4 and ND5 transcripts within 30 min of moderate hypoxia[60]. Thus, the presence of mtDNA is essential for chemiosmotic ATP synthesis in mitochondria regardless of whether the reason is a requirement for

redox regulation, as predicted by the CORR hypothesis (and recently demonstrated in the analogous chloroplast system[61]) or not.

This requirement for physical association of genes with bioenergetic membranes to maintain ATP synthesis constrains both the genomes and the complexity of prokaryotes. If some genes for oxidative phosphorylation must be physically associated with a certain unit area of bioenergetic membranes, then beyond that threshold prokaryotes could not maintain membrane potential homeostasis unless additional genomes are co-localized with the membranes[41]. Examples of giant prokaryotes confirm that this is the case. *Epulopiscium fishelsoni* grows up to 600 µm in length and exhibits extreme polyploidy, with as many as 600,000 copies of the full genome per cell distributed at regular intervals along the plasma membrane (Fig. 1; ref. 8). Likewise *Thiomargarita namibiensis* has 6,000-17,000 nucleoids (Heidi Schulz-Vogt, personal communication), again with a regular peripheral distribution around the plasma membrane[62]. Bacteria can thus sequester DNA at their bioenergetic membranes to attain giant size, so what stops them from attaining true eukaryotic complexity?

Only endosymbiosis restructures genomes

The main difference between endosymbiosis and polyploidy relates to the size and distribution of genomes over evolutionary time. In endosymbiosis, surplus organelle genes are lost or transferred to the host's chromosomes, streamlining endosymbiont replication via cytoplasmic inheritance[11,17,63]. The outcome is a massive reduction in genome size, both in prokaryotic endosymbionts[11] and organelles[64], with a reciprocal relocation of genes in low copy number to nuclear chromosomes in the latter. By contrast, in giant polyploid prokaryotes, all genomes are essentially the same. Without cytoplasmic inheritance, no genomic specialization ensues.

In principle, prokaryotes could control respiration using specialized, membrane-associated plasmids that emulate organelle genomes in gene content and function. In practice, such plasmids are not found. Bacteria usually have small, high-copy-number plasmids that segregate randomly at cell division, or very few giant plasmids that co-segregate with chromosomes on filaments from midpoint[65]. For plasmids in a prokaryote to support electron flux as organelle genomes do, high-copy-number giant plasmids encoding components of the electron-transport chain would need to associate with the plasma membrane, and evolve counter to the tendency to segregate with size rather than function[46]. That no mtDNA-like plasmids are known indicates that high energetic barriers preclude their evolution: unlike organelles, which pay back energetically from the start, substantial energetic costs must be paid up front (high copy number of the correct plasmids, and the machinery to associate them with the membrane at regular intervals) before any energetic advantage can accrue.

The penalty for not having mitochondria or dedicated mtDNA-like giant plasmids is that *Epulopiscium* must replicate its 3.8 Mb genome hundreds of thousands of times every generation. This giant bacterium with 200,000 3.8 Mb genomes harbours 760,000 Mb DNA; a similarly sized eukaryote with 200,000 copies of an average mitochondrial genome must sustain only 6,000 Mb of DNA (and for small mitochondrial genomes potentially as little as 1,200 Mb). If the metabolic rate of *Epulopiscium* were around 0.01 W g^{-1} (similar to *Amoeba proteus*) and its mass 4,000,000 × 10^{-12} g, its metabolic rate would be 40 nW per cell, similar to eukaryotes. However, because *Epulopiscium* has thousands of complete genomes, this translates into only 0.075 pW Mb^{-1}, similar to other bacteria. At a mean gene density of 12 genes per Mb, *Epulopiscium* could sustain fewer than 50 genes, and hence should have high gene density, typical of bacteria, despite its energetic tolerance for a massive amount of DNA. Bioenergetic considerations grant *Epulopiscium* lots of DNA per cell, but organized as complete compact prokaryotic genomes.

Thus, being large and having masses of DNA is not enough to attain complexity: cells need to control energy coupling across a wide area of membranes using small, high copy, bioenergetically specialized genomes like mtDNA (Fig. 2). Segregating the genes relinquished by the endosymbiont (mtDNA) into low copy number in the host's chromosomes, specialization of the endosymbiont into an ATP-generating organelle[50,51] and increasing organelle copy number provides sufficient energy per gene to support the evolution, maintenance and expression of some 10^5 more host genes, affording the cell the chance—but not the necessity—of becoming complex.

This critical redistribution of DNA in relation to bioenergetic membranes seems to be the fundamental hurdle en route to complexity. Shifting from fermentation or anaerobic respiration to aerobic respiration increases ATP availability by at best an order of magnitude, a difference already manifest in many aerobic bacteria. By contrast, mitochondria enabled an increase in host genome size by four to six orders of magnitude, regardless of the electron acceptor. *Epulopiscium* is as big as a eukaryote and has as much DNA, but its genomic symmetry stipulates that it remains a prokaryote. The same applies to *Thiomargarita* (Fig. 2). Neither giant prokaryote has evolved genuinely eukaryotic traits because the metabolic power to support the suite of additional genes necessary to attain true complexity is lacking.

Rather than maintaining 200,000 copies of the energetically unnecessary majority of the genome (that is lost from mtDNA), a genome supported by mitochondria has that much room to expand in size, to evolve new, larger and expressable genes, gene families and regulatory elements, opening the door to a realm of protein evolution that is inaccessible to cells that lack mitochondria and hence harness energy across their plasma membrane only. That is why mitochondrial DNA is the key to complexity.

Cell complexity requires symbiotic energetics

The cornerstone of eukaryotic complexity is a vastly expanded repertoire of novel protein folds, protein interactions and regulatory cascades. The eukaryote common ancestor increased its genetic repertoire by some 3,000 novel gene families[42,43]. The invention of new protein folds in the eukaryotes was the most intense phase of gene invention since the origin of life[66]. Eukaryotes invented five times as many protein folds as eubacteria, and ten

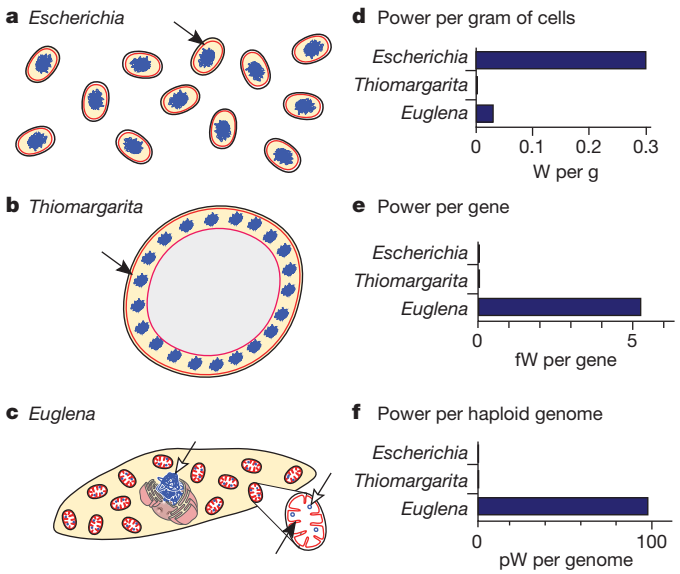

Figure 2 | The cellular power struggle. **a**–**c**, Schematic representations of a medium sized prokaryote (*Escherichia*), a very large prokaryote (*Thiomargarita*), and a medium-sized eukaryote (*Euglena*). Bioenergetic membranes across which chemiosmotic potential is generated and harnessed are drawn in red and indicated with a black arrow; DNA is indicated in blue. In **c**, the mitochondrion is enlarged in the inset, mitochondrial DNA and nuclear DNA are indicated with open arrows. **d**–**f**, Power production of the cells shown in relation to fresh weight (**d**), per haploid gene (**e**) and per haploid genome (power per haploid gene times haploid gene number) (**f**). Note that the presence or absence of a nuclear membrane in eukaryotes, although arguably a consequence of mitochondrial origin[70], has no impact on energetics, but that the energy per gene provided by mitochondria underpins the origin of the genomic complexity required to evolve such eukaryote-specific traits (see text).

times as many as archaea[42]. Even median protein length is 30% greater in eukaryotes than in prokaryotes[47].

Our considerations reveal why the exploration of protein sequence space en route to eukaryotic complexity required mitochondria. Without mitochondria, prokaryotes—even giant polyploids—cannot pay the energetic price of complexity; the lack of true intermediates in the prokaryote-to-eukaryote transition has a bioenergetic cause. The conversion from endosymbiont to mitochondrion provided a freely expandable surface area of internal bioenergetic membranes, serviced by thousands of tiny specialized genomes that permitted their host to evolve, explore and express massive numbers of new proteins in combinations and at levels energetically unattainable for its prokaryotic contemporaries. If evolution works like a tinkerer, evolution with mitochondria works like a corps of engineers.

Although conventional wisdom holds that the origin of complexity was key to the acquisition of mitochondria[19,20], the energetics of genome expression polarize this major evolutionary transition to the contrary: mitochondria are prerequisite to complexity. It follows that the host for mitochondria was a prokaryote[24,27], and complex, multigenic, eukaryote-specific traits such as the cell cycle, sex[67], phagocytosis[68], endomembrane trafficking[69], the nucleus[70] and multicellularity arose after the mitochondrial endosymbiosis. Similarly, prokaryotes cannot have evolved from eukaryotes[71,72] because the energy per gene required to bring forth the complex eukaryotic starting point for prokaryotic evolution under such views requires a prokaryotic endosymbiont to begin with.

Any bacterium with a haploid genome of average eukaryotic size (thousands of Mb) supported by chemiosmosis at the plasma membrane would falsify this hypothesis. Any primitively amitochondriate eukaryote would too. A giant bacterium with a high respiratory rate but without extreme polyploidy would falsify it, as would giant prokaryotes with thousands of membrane-associated, bioenergetically dedicated plasmids.

The transition to complex life on Earth was a unique event that hinged on a bioenergetic jump afforded by spatially combinatorial relations between two cells and two genomes (endosymbiosis), rather than natural selection acting on mutations accumulated gradually among physically isolated prokaryotic individuals. Given the energetic nature of these arguments, the same is likely to be true of any complex life elsewhere.

1. Rokas, A. The origins of multicellularity and the early history of the genetic toolkit for animal development. *Annu. Rev. Genet.* **42,** 235–251 (2008).
2. Lindsay, M. R. *et al.* Cell compartmentalisation in planctomycetes: novel types of structural organisation for the bacterial cell. *Arch. Microbiol.* **175,** 413–429 (2001).
3. Smith, J. M., Smith, N. H., O'Rourke, M. & Spratt, B. G. How clonal are bacteria? *Proc. Natl Acad. Sci. USA* **90,** 4384–4388 (1993).
4. Bentley, S. D. *et al.* Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* **417,** 141–147 (2002).
5. Pinevich, A. V. Intracytoplasmic membrane structures in bacteria. *Endocyt. Cell Res.* **12,** 9–40 (1997).
6. Robinson, N. P. & Bell, S. D. Extrachromosomal element capture and the evolution of multiple replication origins in archaeal chromosomes. *Proc. Natl Acad. Sci. USA* **104,** 5806–5811 (2007).
7. Schulz, H. N. & Jorgensen, B. B. Big bacteria. *Annu. Rev. Microbiol.* **55,** 105–137 (2001).
8. 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).
9. Vats, P., Yu, J. & Rothfield, L. The dynamic nature of the bacterial cytoskeleton. *Cell. Mol. Life Sci.* **66,** 3353–3362 (2009).
10. Davidov, Y. & Jurkevitch, E. Predation between prokaryotes and the origin of eukaryotes. *Bioessays* **31,** 748–757 (2009).
11. Moran, N. A. Symbiosis as an adaptive process and source of phenotypic complexity. *Proc. Natl Acad. Sci. USA* **104,** 8627–8633 (2007).
12. Simon, D. M. & Zimmerly, S. A diversity of uncharacterized retroelements in bacteria. *Nucleic Acids Res.* **36,** 7219–7229 (2008).
13. Waters, C. M. & Bassler, B. L. Quorum sensing: cell-to-cell communication in bacteria. *Annu. Rev. Cell Dev. Biol.* **21,** 319–346 (2005).
14. Lonhienne, T. G. A. *et al.* Endocytosis-like protein uptake in the bacterium *Gemmata obscuriglobus*. *Proc. Natl Acad. Sci. USA* **107,** 12883–12888 (2010).
15. von Dohlen, C. D., Kohler, S., Alsop, S. T. & McManus, W. R. Mealybug β-proteobacterial symbionts contain γ-proteobacterial symbionts. *Nature* **412,** 433–436 (2001).
A rare example of a prokaryote residing as an endosymbiont within a prokaryotic host, demonstrating that phagocytosis is not prerequisite to endosymbiosis.
16. Wujek, D. E. Intracellular bacteria in the blue-green-alga *Pleurocapsa minor*. *Trans. Am. Microsc. Soc.* **98,** 143–145 (1979).
17. Lynch, M. & Conery, J. S. The origins of genome complexity. *Science* **302,** 1401–1404 (2003).
18. Smith, J. M. & Szathmary, E. *The Major Transitions in Evolution* (Oxford Univ.Press, 1995).
19. Cavalier-Smith, T. Predation and eukaryote cell origins: a coevolutionary perspective. *Int. J. Biochem. Cell Biol.* **41,** 307–322 (2009).
20. de Duve, C. The origin of eukaryotes: a reappraisal. *Nature Rev. Genet.* **8,** 395–403 (2007).
21. Rivera, M. C. & Lake, J. A. The ring of life provides evidence for a genome fusion origin of eukaryotes. *Nature* **431,** 152–155 (2004).
22. Koonin, E. V. Darwinian evolution in the light of genomics. *Nucleic Acids Res.* **37,** 1011–1034 (2009).
23. Pisani, D., Cotton, J. A. & McInerney, J. O. Supertrees disentangle the chimeric origin of eukaryotic genomes. *Mol. Biol. Evol.* **24,** 1752–1760 (2007).
24. Cox, C. J., Foster, P. G., Hirt, R. P., Harris, S. R. & Embley, T. M. The archaebacterial origin of eukaryotes. *Proc. Natl Acad. Sci. USA* **105,** 20356–20361 (2008).
An important contribution, using a state of the art phylogenetic repertoire, to show that the host that acquired the mitochondrion was an archaebacterium (a prokaryote).
25. Tovar, J. *et al.* Mitochondrial remnant organelles of *Giardia* function in iron-sulphur protein maturation. *Nature* **426,** 172–176 (2003).
26. van der Giezen, M. Hydrogenosomes and mitosomes: conservation and evolution of functions. *J. Eukaryot. Microbiol.* **56,** 221–231 (2009).
27. Martin, W. & Müller, M. The hydrogen hypothesis for the first eukaryote. *Nature* **392,** 37–41 (1998).
28. Tielens, A. G. M. *et al.* Mitochondria as we don't know them. *Trends Biochem. Sci.* **27,** 564–572 (2002).
29. Harold, F. M. *The Vital Force: A Study of Bioenergetics* (Freeman, 1986).
30. Walker, J. C., Margulis, L. & Rambler, M. Reassessment of roles of oxygen and ultraviolet light in Precambrian evolution. *Nature* **264,** 620–624 (1976).
31. Johnston, D. T., Wolfe-Simon, F., Pearson, A. & Knoll, A. H. Anoxygenic photosynthesis modulated Proterozoic oxygen and sustained Earth's middle age. *Proc. Natl Acad. Sci. USA* **106,** 16925–16929 (2009).
32. Makarieva, A. M., Gorshkov, V. G. & Li, B. L. Energetics of the smallest: do bacteria breathe at the same rate as whales? *Proc. R. Soc. Lond. B* **272,** 2219–2224 (2005).
33. Fenchel, T. & Finlay, B. J. Respiration rates in heterotrophic, free-living protozoa. *Microb. Ecol.* **9,** 99–122 (1983).
34. Vellai, T. & Vida, G. The origin of eukaryotes: the difference between prokaryotic and eukaryotic cells. *Proc. R. Soc. Lond. B* **266,** 1571–1577 (1999).
35. Wagner, A. Energy constraints on the evolution of gene expression. *Mol. Biol. Evol.* **22,** 1365–1374 (2005).
36. Nilsson, M., Bülow, L. & Wahlund, K. Use of flow field-flow fractionation for the rapid quantitation of ribosome and ribosomal subunits in *Escherichia coli* at different protein production conditions. *Biotechnol. Bioeng.* **54,** 461–467 (1997).
37. Weibel, E. R. *et al.* Correlated morphometric and biochemical studies of the liver cell. *J. Cell Biol.* **42,** 68–91 (1969).
38. Gray, M. W., Lang, B. F. & Burger, G. Mitochondria of protists. *Annu. Rev. Genet.* **38,** 477–524 (2004).
39. Daniels, E. W. & Breyer, E. P. Starvation effects on the ultrastructure of amoeba mitochondria. *Z. Zellforsch.* **91,** 159–169 (1968).
40. Aury, J.-M. *et al.* Global trends of whole genome duplications revealed by the ciliate *Paramecium tetraurelia*. *Nature* **444,** 171–178 (2006).
41. Lane, N. *Power, Sex, Suicide: Mitochondria and the Meaning of Life* (Oxford Univ. Press, 2005).
42. Koonin, E. V. *et al.* A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. *Genome Biol.* **5,** R7 (2004).
A seminal contribution that underscores the uniqueness of eukaryotic genomes with respect to their enriched protein content relative to prokaryotic forebears.
43. Fritz-Laylin, L. K. *et al.* The genome of *Naegleria gruberi* illuminates early eukaryotic versatility. *Cell* **140,** 631–642 (2010).
44. Kunin, V. & Ouzounis, C. A. The balance of driving forces during genome evolution in prokaryotes. *Genome Res.* **13,** 1589–1594 (2003).
45. Kuo, C. H. & Ochman, H. The extinction dynamics of bacterial pseudogenes. *PLoS Genet.* **6,** e1001050 (2010).
46. Vellai, T., Takacs, K. & Vida, G. A new aspect to the origin and evolution of eukaryotes. *J. Mol. Evol.* **46,** 499–507 (1998).
47. Brocchieri, L. & Karlin, S. Protein length in eukaryotic and prokaryotic proteomes. *Nucleic Acids Res.* **33,** 3390–3400 (2005).
48. Peterson, K. J., Dietrich, M. R. & McPeek, M. A. MicroRNAs and metazoan macroevolution: insights into canalization, complexity, and the Cambrian explosion. *Bioessays* **31,** 736–747 (2009).
49. Bidle, K. D. & Falkowski, P. G. Cell death in planktonic, photosynthetic microorganisms. *Nature Rev. Microbiol.* **2,** 643–655 (2004).
50. Allen, J. F. Control of gene expression by redox potential and the requirement for chloroplast and mitochondrial genomes. *J. Theor. Biol.* **165,** 609–631 (1993).
51. Allen, J. F. The function of genomes in bioenergetic organelles. *Philos. Trans. R. Soc. Lond. B* **358,** 19–38 (2003).
Presents compelling bioenergetic reasons, necessary and sufficient, to account for the retention of genes involved in membrane-associated electron transport in mitochondria (and chloroplasts).
52. Williams, R. S. Mitochondrial gene expression in mammalian striated muscle: evidence that variation in gene dosage is the major regulatory event. *J. Biol. Chem.* **261,** 12390–12394 (1986).
53. Williams, R. S. *et al.* Regulation of nuclear and mitochondrial gene expression by contractile activity in skeletal muscle. *J. Biol. Chem.* **261,** 376–380 (1986).

54. Shay, J. W., Pierce, D. J. & Werbin, H. Mitochondrial DNA copy number is proportional to total cell DNA under a variety of growth conditions. *J. Biol. Chem.* **265,** 14802–14807 (1990).
55. Schapira, A. H. Mitochondrial disease. *Lancet* **368,** 70–82 (2006).
56. Rocher, C. *et al.* Influence of mitochondrial DNA level on cellular energy metabolism: implications for mitochondrial diseases. *J. Bioenerg. Biomembr.* **40,** 59–67 (2008).
A systematic study demonstrating the linear dependence of metabolic rate on mtDNA copy number.
57. Moreno-Loshuertos, R. *et al.* Differences in reactive oxygen species production explain the phenotypes associated with common mouse mitochondrial DNA variants. *Nature Genet.* **38,** 1261–1268 (2006).
An important paper showing that free-radical signals modulate mtDNA copy number and the rate of ATP synthesis.
58. Bai, Y., Shakeley, R. M. & Attardi, G. Tight control of respiration by NADH dehydrogenase ND5 subunit gene expression in mouse mitochondria. *Mol. Cell. Biol.* **20,** 805–815 (2000).
A seminal contribution, showing that the rate of transcription of a mtDNA-encoded respiratory subunit controls the overall rate of respiration.
59. Chomyn, A. Mitochondrial genetic control of assembly and function of complex I in mammalian cells. *J. Bioenerg. Biomembr.* **33,** 251–257 (2001).
60. Piruat, J. I. & López-Barneo, J. Oxygen tension regulates mitochondrial DNA-encoded complex I gene expression. *J. Biol. Chem.* **280,** 42676–42684 (2005).
61. Shimizu, M. *et al.* Sigma factor phosphorylation in the photosynthetic control of photosystem stoichiometry. *Proc. Natl Acad. Sci. USA* **107,** 10760–10764 (2010).
62. Schulz, H. N. The genus *Thiomargarita*. *Prokaryotes* **6,** 1156–1163 (2006).
63. Timmis, J. N. *et al.* Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nature Rev. Genet.* **5,** 123–135 (2004).
64. Lane, C. E. & Archibald, J. M. The eukaryotic tree of life: endosymbiosis takes its TOL. *Trends Ecol. Evol.* **23,** 268–275 (2008).
65. Ebersbach, G. & Gerdes, K. Plasmid segregation mechanisms. *Annu. Rev. Genet.* **39,** 453–479 (2005).
66. Yang, S., Doolittle, R. F. & Bourne, P. E. Phylogeny determined by protein domain content. *Proc. Natl Acad. Sci. USA* **102,** 373–378 (2005).
67. Lane, N. *Life Ascending: The Ten Great Inventions of Evolution* (Norton, 2009).
68. Yutin, N., Wolf, M. Y., Wolf, Y. I. & Koonin, E. V. The origins of phagocytosis and eukaryogenesis. *Biol. Direct* **4,** 9 (2009).
69. Brighouse, A., Dacks, J. B. & Field, M. C. Rab protein evolution and the history of the eukaryotic endomembrane system. *Cell. Mol. Life Sci.* **67,** 3449–3465 (2010).
70. Martin, W. & Koonin, E. V. Introns and the origin of nucleus–cytosol compartmentalization. *Nature* **440,** 41–45 (2006).
71. Forterre, P. & Gribaldo, S. Bacteria with a eukaryotic touch: a glimpse of ancient evolution? *Proc. Natl Acad. Sci. USA* **107,** 12739–12740 (2010).
72. Kurland, C. G., Collins, L. J. & Penny, D. Genomics and the irreducible nature of eukaryote cells. *Science* **312,** 1011–1014 (2006).
73. Schulz, H. N. & de Beer, D. Uptake rates of oxygen and sulphide measured with individual *Thiomargarita namibiensis* cells by using microelectrodes. *Appl. Environ. Microbiol.* **68,** 5746–5749 (2002).
74. Parfrey, L. W., Lahr, D. J. G. & Katz, L. A. The dynamic nature of eukaryotic genomes. *Mol. Biol. Evol.* **25,** 787–794 (2008).

Acknowledgements We are indebted to A. Hidalgo, D. Braben, F. Harold, J. Ellis, H. Schulz-Vogt, J. Allen, G. Shields and L. Sweetlove for many discussions and comments on the manuscript, M. Farmer, H. Schulz-Vogt and R. Allen for microscopic images. N.L. is very grateful to the UCL Provost's Venture Research Fellowship, W.M. to the German Research Foundation and the European Research Council (Networkorigins) for funding.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence should be addressed to N.L. (nick.lane@ucl.ac.uk).