Genomics and the Irreducible Nature of Eukaryote Cells

C. G. Kurland,[1] L. J. Collins,[2] D. Penny[2]*

Large-scale comparative genomics in harness with proteomics has substantiated fundamental features of eukaryote cellular evolution. The evolutionary trajectory of modern eukaryotes is distinct from that of prokaryotes. Data from many sources give no direct evidence that eukaryotes evolved by genome fusion between archaea and bacteria. Comparative genomics shows that, under certain ecological settings, sequence loss and cellular simplification are common modes of evolution. Subcellular architecture of eukaryote cells is in part a physical-chemical consequence of molecular crowding; subcellular compartmentation with specialized proteomes is required for the efficient functioning of proteins.

Comparative genomics and proteomics have strengthened the view that modern eukaryote and prokaryote cells have long followed separate evolutionary trajectories. Because their cells appear simpler, prokaryotes have traditionally been considered ancestors of eukaryotes (1–4). Nevertheless, comparative genomics has confirmed a lesson from paleontology: Evolution does not proceed monotonically from the simpler to the more complex (5–9). Here, we review recent data from proteomics and genome sequences suggesting that eukaryotes are a unique primordial lineage.

Mitochondria, mitosomes, and hydrogenosomes are a related family of organelles that distinguish eukaryotes from all prokaryotes (10). Recent analyses also suggest that early eukaryotes had many introns (11, 12), and RNAs and proteins found in modern spliceosomes (13). Indeed, it seems that life-history parameters affect intron numbers (14, 15). In addition, "molecular crowding" is now recognized as an important physical-chemical factor contributing to the compartmentation of even the earliest eukaryote cells (16, 17).

Nuclei, nucleoli, Golgi apparatus, centrioles, and endoplasmic reticulum are examples of cellular signature structures (CSSs) that distinguish eukaryote cells from archaea and bacteria. Comparative genomics, aided by proteomics of CSSs such as the mitochondria (18, 19), nucleoli (20, 21), and spliceosomes (13, 22), reveals hundreds of proteins with no orthologs evident in the genomes of prokaryotes; these are the eukaryotic signature proteins (ESPs) (23, 24). The many ESPs within the subcellular structures of eukaryote cells provide landmarks to track the trajectory of eukaryote genomes from their origins. In contrast,

hypotheses that attribute eukaryote origins to genome fusion between archaea and bacteria (25–30) are surprisingly uninformative about the emergence of the cellular and genomic signatures of eukaryotes (CSSs and ESPs). The failure of genome fusion to directly explain any characteristic feature of the eukaryote cell is a critical starting point for studying eukaryote origins.
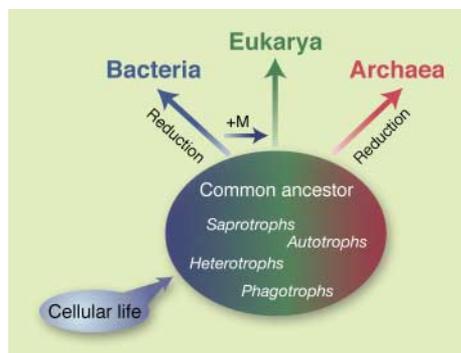

Fig. 1. The common ancestor of eukaryotes, bacteria, and archaea may have been a community of organisms containing the following: autotrophs that produced organic compounds from CO_2 either photosynthetically or by inorganic chemical reactions; heterotrophs that obtained organics by leakage from other organisms; saprotrophs that absorbed nutrients from decaying organisms; and phagotrophs that were sufficiently complex to envelop and digest prey. +M: endosymbiosis of mitochondrial ancestor.

It is agreed that, whether using gene content, protein-fold families, or RNA sequences (31–36), the unrooted tree of life divides into archaea, bacteria, and eukaryotes (Fig. 1). On such unrooted trees, the three domains diverge from a population that can be called the last universal common ancestor (LUCA). However, LUCA (37) means different things to different people, so we prefer to call it a common ancestor; in this case it is the hypothetical node at which the three domains coalesce in unrooted trees.

There are links between comparative genomics and the ecology of organisms. These include the aerobic/anaerobic states of the environment and the adaptive fit of organelles such as mitochondria, hydrogenosomes, and mitosomes (10, 18, 19, 38–41). In addition to the advantages from oxidative metabolism and/or oxygen detoxification, other advantages must have accrued from having a cellular compartment with dense proteomes (15, 38, 42). Ecological specialization can account for the differences between prokaryote and eukaryote cell architectures and genome sizes. Small prokaryote cells with streamlined genomes may reflect adaptation to rapid growth and/or minimal resource use by autotrophs, heterotrophs, and saprotrophs. Divergent evolutionary paths may emerge with the adoption of a phagotrophic-feeding mode in an ancestor of eukaryotes. This uniquely eukaryote feeding mode requires a larger and more complex cell, consistent with earlier suggestions that a unicellular raptor (predator), which acquired a bacterial endosymbiont/mitochondria lineage, became the common ancestor of all modern eukaryotes (3, 4, 43). Indeed, predator/prey relationships may provide the ecological setting for the divergence of the distinctive cell types adopted by eukaryotes, bacteria, and archaea.

Proteomics of Cell Compartments

Comparative genomics and proteomics reveal phylogenetic relationships between proteins making up eukaryote subcellular features and those found in prokaryotes. We distinguish three main phylogenetic classes; the first are proteins that are unique to eukaryotes: the ESPs. The ESPs we place in three subclasses: proteins arising de novo in eukaryotes; proteins so divergent to homologs of other domains that their relationship is largely lost; or finally, descendants of proteins that are lost from other domains, surviving only as ESPs in eukaryotes.

The second class contains interdomain horizontal gene transfers; these are proteins occurring in two domains with the lineage of one domain rooted within their homologs in a second domain (44). The third class contains homologs found in at least two domains, but the proteins of one domain are not rooted within another domain(s); instead, the homologs appear to descend from the common ancestor (Fig. 1). Most eukaryote proteins shared by prokaryotes are distant, rather than close, relatives. Thus, proteins shared between domains appear to be descendants of the common ancestor; few seem to result from interdomain lateral gene transfer (31–35).

Although the genomes of mitochondria are clearly descendants of α-proteobacteria (45, 46), proteomics and comparative genomics identify relatively few proteins in yeast and human

[1]Department of Microbial Ecology, Lund University, Lund, Sweden. [2]Allan Wilson Center for Molecular Ecology and Evolution, Massey University, Palmerston North, New Zealand.

*To whom correspondence should be addressed. E-mail: d.penny@massey.ac.nz

mitochondria descended from the ancestral bacterium (*17*, *18*, *36*, *47*). Several hundred genes have been transferred from the ancestral bacterium to the nuclear genome, but most proteins from the original endosymbiont have been lost. For yeast, the largest protein class contains more than 200 eukaryote proteins (ESPs) targeted to the mitochondrion but encoded in the nucleus. In addition, the yeast nucleus encodes 150 mitochondrial proteins not uniquely identifiable with a single domain but apparently eukaryotic descendants from the common ancestor. Accordingly, the yeast and human mitochondria proteomes emerge largely as products of the eukaryotic nuclear genome (85%) and only to a lesser degree (15%) as direct descendants of endosymbionts (*17*, *18*, *36*, *45*). The strong representation of ESPs in their proteomes means that mitochondria and their descendants are usefully viewed as "honorary" CSSs.

There are substantial numbers of ESPs in the other CSSs. For the proteome of the reduced anaerobic parasite *Giardia lamblia* (*23*), searches of 2136 proteins found in each of *Saccharomyces cerevisiae*, *Drosophila melanogaster*, *Caenorhabditis elegans*, and *Arabidopsis thaliana* yielded 347 ESPs for *G. lamblia*. This was reduced to roughly 300 by rigorous screening, with ESPs distributed between nuclear and cytoplasmic compartments (Fig. 2) (*48*). The ubiquity of the ESPs and the absence of archaeal descendants are not easily explained by a prokaryote genome fusion model (*49*). The simplest interpretation is that the host for the endosymbiont/mitochondrial lineage was an ancestral eukaryote.

Similar results are obtained for another reduced eukaryote, the intracellular parasite *Encephalitozoon cuniculi*. A recent study (*24*) identified 401 ESPs, of which 295 had homologs among the ESPs of *G. lamblia* (*23*). Two major categories of ESPs in the *G. lamblia* and *E. cuniculi* genomes were distinguished: those associated with the CSSs (Fig. 2) and those involved in control functions such as guanosine triphosphate (GTP) binding proteins, kinases, and phosphatases (*7*). It was also observed (*23*) that many characteristic eukaryotic proteins with weak sequence homology to prokaryotic proteins but more convincing homologies of structural fold such as the actins, tubulins, kinesins,

ubiquitins, and some GTP binding proteins are among the most highly conserved eukaryotic proteins. These may be descendants of the common ancestor recruited early in the evolution of the eukaryotic nuclear genome.

Nucleolar proteomes (*20*, *21*) are examples of essential eukaryote compartments not wrapped in double membranes and where there is no suspicion of an endosymbiotic origin. From 271 proteins in the human nucleolar proteome, 206 protein folds were identified and classified phylogenetically (*20*, *21*). Of these, 109 are eukaryotic signature folds, and the remaining ones appear to be descendants of the common ancestor, occurring in two or three domains.

The spliceosome is a unique molecular machine that removes introns from eukaryote mRNAs (*22*). Even though we do not know the ancestral processing signals for the earliest eukaryotes (*50*), roughly half of the 78 spliceosomal proteins likely to be present in the ancestral spliceosome are ESPs, (*13*) whereas the other half containing the Sm/LSm proteins (*51*) have homologs in bacteria and archaea (*13*). These distributions of both ESPs as well as of putative descendants of the common ancestor suggest that many components of modern spliceosomes were present in the common ancestor (*52*).

The subdivision into subcellular compartments (CSSs) with characteristic proteomes restricts proteins to volumes considerably smaller than the whole cell. Concentrations of macromolecules in cells are very high, typically be-

tween 20 and 30% of weight or volume (*53*). Such densities are described as "molecular crowding" because the space between macromolecules is much less than their diameters; consequently, diffusion of proteins in cells is retarded (*54*). Molecular crowding favors macromolecular associations, large complexes, and networks of proteins that support biological functions (*16*, *17*, *53*).

High densities enhance the association kinetics of small molecules with proteins because the excluded volumes of the proteins reduce the effective volume through which small molecules diffuse (*55*). The sum of these effects is that the high macromolecular densities within CSSs enhance the kinetic efficiencies of proteins. The same principles apply to the smaller prokaryotic cells, but the effects are accentuated in larger cells. Subdividing high densities of proteins into more or less distinct compartments containing functionally interactive macromolecules is expected to be an early feature of the eukaryote lineage. The distinctive proteome of nucleoli demonstrates that compartmentation does not require an enclosing membrane. Furthermore, cell fusion is not required to account for, nor does it explain (*49*), the large number of eukaryote cell compartments.
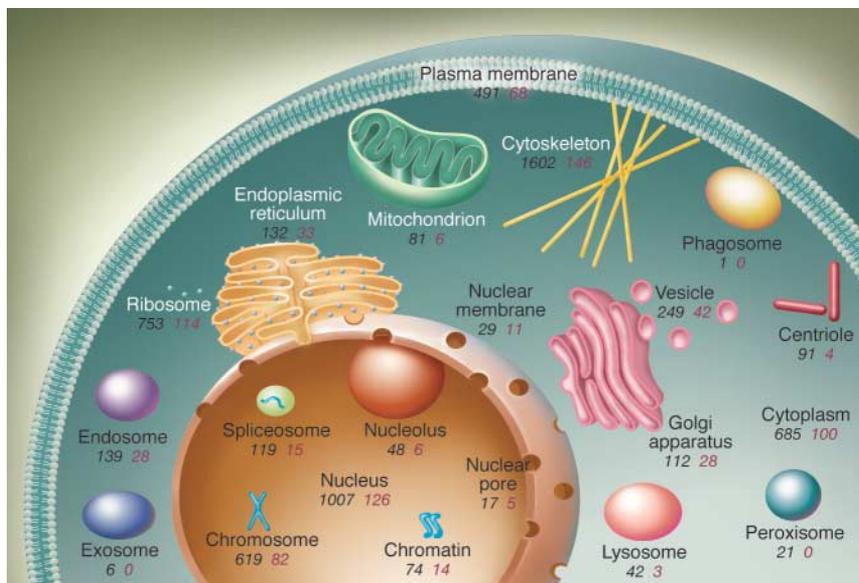
Selection Gives and Selection Takes

Genomes evolve continuously through the interplay of unceasing mutation, unremitting competition, and ever-changing environments. Both sequence loss and sequence gain can result. In general, expanded genome size, along with augmented gene expression, increases the costs of cell propagation so the evolution of larger genomes and larger cells requires gains in fitness that compensate (*15*, *56*, *57*). Conversely, genome reduction is expected to lower the costs of propagation. There is an ever-present potential to improve the efficiency of cell propagation by reductive evolution.

Environmental shifts may neutralize sequences, leaving no selective pressure to maintain them against the persistent flux of deleterious mutations. Such neutralized sequences eventually and inevitably disappear because of "mutational meltdown" (*14*, *15*, *56*, *57*). Genome reduction can be achieved through differential loss of coding and noncoding sequences (com-

Fig. 2. Distribution of ESPs in the proteome of *G. lamblia*. ESPs (*23*) were matched to the human International Protein Index data set (*48*) and then assigned to individual CSSs based on their gene ontology annotations. A protein may be present in more than one CSS (e.g., a protein involved in transport from the nucleus to the cytoplasm will be assigned to both CSSs). Black numbers are the number of proteins assigned to each CSS from the total *G. lamblia* proteome (AACB00000000) (3077 ORFs matched and linked to gene ontology); red numbers are the ESPs assigned to each CSS (320 proteins matched and linked to gene ontology).

paction) (*57*). *Theileria* has evolved through gene loss as well as compaction of its intergenic spaces, whereas *Paramecium* has eliminated only a small length of genes but markedly reduced the number of its introns (*57*). The complex genomes of some vertebrates (pufferfish, *Takifugu*) are so highly compacted that their genome lengths are reduced to one-eighth that of other vertebrates (*58*). Extreme cellular simplification is observed among anaerobic protists, including simplification of CSSs such as mitochondria and the Golgi apparatus (*59–64*). *S. cerevisiae*, which underwent a whole-genome duplication, subsequently purged ~85% of the duplicated sequences (*65, 66*). The evolution of genome content is clearly not monotonic (Fig. 3) (*67, 68*). Genome sizes on the branches of a phylogenetic tree of fungi show irregular genome enlargement (including duplication) and reduction. Examples of ecological circumstances driving genome reduction are seen in many intracellular endosymbionts and parasites, which gain few genes but lose many genes responsible for metabolic flexibility (*6–8, 69*).

The mitochondrion is even more extreme in its reductive evolution; its ancestral bacterial genome has been reduced to a vestigial microgenome supported by a predominantly eukaryote proteome (*18, 19*). Genomes of modern mitochondria encode between 3 and 67 proteins (*44*), whereas the smallest known free-living α-proteobacterium (*Bartonella quintana*) encodes ~1100 proteins (*70*). Taking *Bartonella* as a minimal genome for the free-living ancestor of mitochondria, nearly all of the bacterial coding sequences have been lost from the organelle, though not necessarily from the eukaryote cell. The mitochondrial genome of the protist *Reclinomonas americana* is the largest known but has still lost more than 95% of its original coding capacity.

This abbreviated account of genome reduction illustrates the Darwinian view of evolution as a reversible process in the sense that "eyes can be acquired and eyes can be lost." Genome evolution is a two-way street. This bidirectional sense of reversibility is important as an alternative to the view of evolution as a rigidly monotonic progression from simple to more complex states, a view with roots in the 18th-century theory of orthogenesis (*71*). Unfortunately, such a model has been tacitly favored by molecular biologists who appeared to view evolution as an irreversible march from simple prokaryotes to complex eukaryotes, from unicellular to multicellular. The many well-documented instances of genome reduction provide a necessary corrective measure to the often-unstated assumption that eukaryotes must have originated from prokaryotes.

The Hunt for the Phagotrophic Unicellular Raptor

Proteomics, together with comparative genomics, allows glimpses of the cell structure of eukaryote ancestors. They are likely to have had introns as well as the complex machinery for removing them, and much of that RNA processing machinery still exists in their descendants (*13, 22, 51*). Because of molecular crowding, it is expected that interacting proteins would tend to accumulate in functional domains, making rudimentary CSSs early features of the large-celled eukaryotes. We cannot say whether there was a substantial period of time after the emergence of cells

Fig. 3. Genome sizes (in megabases) can increase and decrease in lineages because of events such as genome duplication and reductive evolution, as illustrated in this fungal phylogeny [adapted from (*67, 68*)]. Genome sizes were obtained from the National Center for Biotechnology Information (NCBI) Genome biology (www.ncbi.nih.gov/Genomes/) database. GD, genome duplication; RE, reductive evolution.

when there were no unicellular raptors or predators—a Garden of Eden. However, the identification among prokaryotes of orthologs with structural affinities to actins, tubulins, kinesins, and ubiquitins (*72, 73*) is consistent with some early organisms having evolved a phagotrophic life-style. This echoes a recurrent theme (*3, 4, 43*) in which it was supposed that the earliest eukaryotes could feed as unicellular "raptors."

We expect that the earliest organisms were primarily auxotrophs, heterotrophs, and saprotrophs—an excellent community to support raptors. Phagotrophy is a hallmark of eukaryotic cells and is unknown among modern prokaryotes, and so it is natural to reconsider this feeding mode as a defining feature of ancestral eukaryotes. Cavalier-Smith (*43*) suggested that the ancestors of eukaryotes were phagotrophic, anaerobic free-living protists, called archeozoa. He also identified present-day anaerobic parasites such as *Entamoeba*, *Giardia*, and Microsporidia as archeozoa. However, these organisms are descendants of aerobic, mitochondriate eukaryotes (*10*). Genome reduction and cellular simplification are hallmarks of parasites and symbionts (*6–8, 46, 69*). Indeed, most of the eukaryotic anaerobes studied so far are parasites or symbionts of multicellular creatures.

For the reasons outlined above, we favor the idea (*3, 4*) that the host that acquired the mitochondrial endosymbiont was a unicellular eukaryote predator, a raptor. The emergence of unicellular raptors would have had a major ecological impact on the evolution of the gentler descendants of the common ancestor. These may have responded with several adaptive strategies: They might outproduce the raptors by rapid growth or hide from raptors by adapting to extreme environments. Thus, the hypothetical eukaryote raptors may have driven the evolution of their autotrophic, heterotrophic, and saprotrophic cousins in a reductive mode that put a premium on the relatively fast-growing, streamlined cell types we call prokaryotes (*74*).

Concluding Remarks

Genomics and proteomics have greatly increased our awareness of the uniqueness of eukaryote cells. This, together with increased understanding of molecular crowding, as well as the dynamic, often reductive nature of genome evolution, offers a new view of the origin of eukaryote cells. The eukaryotic CSSs define a unique cell type that cannot be deconstructed into features inherited directly from archaea and bacteria. Only a small fraction (~15%) of α-proteobacterial proteins are identified in the yeast and human mitochondrial proteomes; none seem to be direct descendants of archaea, and roughly half seem to be exclusively eukaryotic (*18, 19, 38, 47*). The identification of the α-proteobacterial descendants in this proteome validates the phylogenetic distinction between direct descent from genes transferred to the host from the bacterial endosymbiont, as opposed to descent from a hypothetical common ancestor.

ESPs are important markers of the novel evolutionary trajectory of modern eukaryotes. In contrast, most proteins occur in more than one domain (*31–36*), and most of these could derive from the common ancestor. We take the relative abundance of signature proteins among eukaryotes to indicate that their genomes typically have a greater coding capacity than those of prokaryotes. It remains to be seen which ESPs have been lost from prokaryotes and which have been acquired by eukaryotes during their evolution.
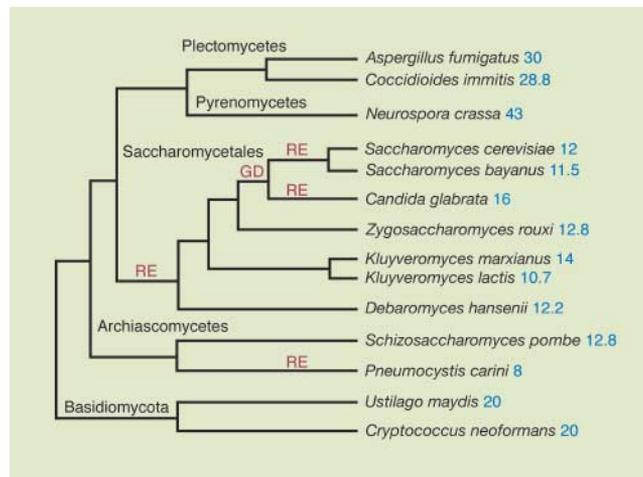
The hypothetical fusion of an archaeon and a bacterium explains nothing about the special features of the modern eukaryote cell (*49*), nor the many signature proteins. Nothing in global phylogenies based on ribosomal RNA, pooled proteins, and protein-fold families indicates that genome fusion generated the eukaryote lineage. Perhaps interest in fusion models arose because BLAST searches suggest that different eukaryotic coding sequences are sometimes more closely related to archaeal homologs and other times more closely related to bacterial homologs (*49*). These weak domain-specific affinities do need to be understood and alternative explanations found. However, in our view (*49*), they do not indicate that the eukaryote genome arose as a mosaic pieced together from archaeal and bacterial genomes.

It is an attractively simple idea that a primitive eukaryote took up the endosymbiont/mitochondrion by phagocytosis (*3, 4, 43*). A unicellular raptor with a larger, more complex cell structure than that of present-day prokaryotes is envisioned as the host of the ancestral endosymbiont. This scenario, which is not contradicted by new data derived from comparative genomics and proteomics, is a suitable starting point for future work. Acquisition of genome sequences from free-living eukaryotes among basal lineages is a high priority.

References and Notes

1. A. H. Knoll, *Science* **256**, 622 (1992).
2. S. L. Baldauf, *Science* **300**, 1703 (2003).
3. R. Y. Stanier, C. B. van Niel, *Arch. Mikrobiol.* **42**, 17 (1962).
4. C. de Duve, *Ann. N. Y. Acad. Sci.* **386**, 1 (1982).
5. G. G. Simpson, *Tempo and Mode in Evolution* (Columbia Univ. Press, New York, 1944).
6. S. G. Andersson, C. G. Kurland, *Trends Microbiol.* **6**, 263 (1998).
7. L. Klasson, S. G. E. Andersson, *Trends Microbiol.* **12**, 37 (2004).
8. M. V. Olsen, *Am. J. Hum. Genet.* **64**, 18 (1999).
9. P. Forterre, H. Philippe, *Bioessays* **21**, 871 (1999).
10. T. M. Embley, W. Martin, *Nature* **440**, 623 (2006).
11. S. W. Roy, W. Gilbert, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 5773 (2005).
12. F. Raible *et al.*, *Science* **310**, 1325 (2005).
13. L. J. Collins, D. Penny, *Mol. Biol. Evol.* **22**, 1053 (2005).
14. M. Lynch, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 6118 (2002).
15. D. C. Jeffares, T. Mourier, D. Penny, *Trends Genet.* **22**, 16 (2006).
16. G. J. Pielak, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 5901 (2005).
17. R. J. Ellis, *Curr. Opin. Struct. Biol.* **11**, 114 (2001).
18. O. Karlberg, B. Canback, C. G. Kurland, S. G. Andersson, *Yeast* **17**, 170 (2000).
19. C. G. Kurland, S. G. Andersson, *Microbiol. Mol. Biol. Rev.* **64**, 786 (2000).
20. J. S. Andersen *et al.*, *Nature* **433**, 77 (2005).
21. E. Staub, P. Fiziev, A. Rosenthal, B. Hinzmann, *Bioessays* **26**, 567 (2004).
22. M. S. Jurica, M. J. Moore, *Mol. Cell* **12**, 5 (2003).
23. H. Hartman, A. Fedorov, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 1420 (2002).
24. A. Fedorov, H. Hartman, *J. Mol. Evol.* **59**, 695 (2004).
25. J. P. Gogarten, L. Olendzenski, E. Hilario, C. Simon, K. E. Holsinger, *Science* **274**, 1750 (1996).
26. D. F. Feng, G. Cho, R. F. Doolittle, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 13028 (1997).
27. S. Ribeiro, G. B. Golding, *Mol. Biol. Evol.* **15**, 779 (1998).
28. P. Lopez-Garcia, D. Moreira, *Trends Biochem. Sci.* **24**, 88 (1999).
29. T. Cavalier-Smith, *Int. J. Syst. Evol. Microbiol.* **52**, 297 (2002).
30. M. C. Rivera, J. A. Lake, *Nature* **431**, 152 (2004).
31. C. R. Woese, O. Kandler, M. L. Wheelis, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 4576 (1990).
32. J. O. Korbel, B. Snel, M. A. Huynen, P. Bork, *Trends Genet.* **18**, 158 (2002).
33. B. Snel, P. Bork, M. A. Huynen, *Genome Res.* **12**, 17 (2002).
34. R. F. Doolittle, *Annu. Rev. Biochem.* **64**, 287 (1995).
35. G. Caetano-Anolles, D. Caetano-Anolles, *Genome Res.* **13**, 1563 (2003).
36. S. Yang, R. F. Doolittle, P. E. Bourne, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 373 (2005).
37. D. Penny, A. Poole, *Curr. Opin. Genet. Dev.* **9**, 672 (1999).
38. T. Gabaldon, M. A. Huynen, *Science* **301**, 609 (2003).
39. O. G. Berg, C. G. Kurland, *Mol. Biol. Evol.* **17**, 951 (2000).
40. T. M. Embley *et al.*, *IUBMB Life* **55**, 387 (2003).
41. B. A. Williams, R. P. Hirt, J. M. Lucocq, T. M. Embley, *Nature* **418**, 865 (2002).
42. A. Regoes *et al.*, *J. Biol. Chem.* **280**, 30557 (2005).
43. T. Cavalier-Smith, *Nature* **326**, 332 (1987).
44. B. Canback, S. G. Andersson, C. G. Kurland, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 6097 (2002).
45. S. G. Andersson *et al.*, *Nature* **396**, 133 (1998).
46. M. W. Gray, G. Burger, B. F. Lang, *Science* **283**, 1476 (1999).
47. T. Gabaldon, M. A. Huynen, *Bioinformatics* **21** (suppl. 2), ii144 (2005).
48. P. J. Kersey *et al.*, *Proteomics* **4**, 1985 (2004).
49. See Supporting Online Material.
50. L. J. Collins, D. Penny, *Mol. Biol. Evol.* **23**, 901 (2006).
51. J. D. Beggs, *Biochem. Soc. Trans.* **33**, 433 (2005).
52. A. Poole, D. Jeffares, D. Penny, *Bioessays* **21**, 880 (1999).
53. R. J. Ellis, *Trends Biochem. Sci.* **26**, 597 (2001).
54. K. Luby-Phelps, *Int. Rev. Cytol.* **192**, 189 (2000).
55. M. Ehrenberg, C. G. Kurland, *Q. Rev. Biophys.* **17**, 45 (1984).
56. O. G. Berg, C. G. Kurland, *Mol. Biol. Evol.* **19**, 2265 (2002).
57. P. J. Keeling, C. H. Slamovits, *Curr. Opin. Genet. Dev.* **15**, 601 (2005).
58. S. Aparicio *et al.*, *Science* **297**, 1301 (2002).
59. J. B. Dacks *et al.*, *Proc. Biol. Sci.* **270** (suppl. 2), S168 (2003).
60. C. G. Clark, A. J. Roger, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 6518 (1995).
61. A. Germot, H. Philippe, H. Le Guyader, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 14614 (1996).
62. J. Tovar *et al.*, *Nature* **426**, 172 (2003).
63. A. Akhmanova *et al.*, *Mol. Microbiol.* **30**, 1017 (1998).
64. B. Boxma *et al.*, *Nature* **434**, 74 (2005).
65. K. H. Wolfe, D. C. Shields, *Nature* **387**, 708 (1997).
66. B. Dujon *et al.*, *Nature* **430**, 35 (2004).
67. J. Cai, I. N. Roberts, M. D. Collins, *Int. J. Syst. Bacteriol.* **46**, 542 (1996).
68. A. C. Padovan, G. F. Sanson, A. Brunstein, M. R. Briones, *J. Mol. Evol.* **60**, 726 (2005).
69. N. A. Moran, *Cell* **108**, 583 (2002).
70. C. M. Alsmark *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 9716 (2004).
71. P. J. Bowler, *The Non-Darwinian Revolution: Reinterpreting a Historical Myth* (Johns Hopkins Univ. Press, Baltimore, 1988).
72. R. R. Copley, J. Schultz, C. P. Ponting, P. Bork, *Curr. Opin. Struct. Biol.* **9**, 408 (1999).
73. C. P. Ponting, L. Aravind, J. Schultz, P. Bork, E. V. Koonin, *J. Mol. Biol.* **289**, 729 (1999).
74. J. E. Darnell Jr., *Science* **202**, 1257 (1978).
75. We thank G. Jameson and M. Oliveberg for very useful discussions and help with the literature. Supported by The New Zealand Marsden Fund (L.J.C.) and The Royal Physiographic Society, Lund (C.G.K.).