The ring of life provides evidence for a genome fusion origin of eukaryotes

Maria C. Rivera[1,3,4] **& James A. Lake**[1,2,4]

[1]*Molecular Biology Institute, MCD Biology,* [2]*Human Genetics,* [3]*IGPP, and* [4]*Astrobiology Institute, University of California, Los Angeles 90095, USA*

Genomes hold within them the record of the evolution of life on Earth. But genome fusions and horizontal gene transfer seem to have obscured sufficiently the gene sequence record such that it is difficult to reconstruct the phylogenetic tree of life. Here we determine the general outline of the tree using complete genome data from representative prokaryotes and eukaryotes and a new genome analysis method that makes it possible to reconstruct ancient genome fusions and phylogenetic trees. Our analyses indicate that the eukaryotic genome resulted from a fusion of two diverse prokaryotic genomes, and therefore at the deepest levels linking prokaryotes and eukaryotes, the tree of life is actually a ring of life. One fusion partner branches from deep within an ancient photosynthetic clade, and the other is related to the archaeal prokaryotes. The eubacterial organism is either a proteobacterium, or a member of a larger photosynthetic clade that includes the Cyanobacteria and the Proteobacteria.

The origin of the eukaryotic cell is enigmatic and complex. Early studies of nuclear-encoded enzymes, transfer RNAs, ribosome structures and ribosomal RNA catalogues implied deep, but unresolved, connections between prokaryotes and eukaryotes[1–3]. Subsequent analyses of ribosomal sequences suggested that eukaryotes were either the sister group to the Archaea[4], or the sister group to the archaeal Eocyta[5,6] in the rooted tree of life[7,8]. But as additional eukaryotic genes were sequenced, their analyses frequently suggested just the opposite; that is, that many eukaryotic genes were more closely related to the Bacteria than to the Archaea[9–12]. A further complication arose from the finding that the function of each eukaryotic gene is strongly correlated with its ancestry. Informational genes (genes involved in transcription, translation and other related processes) are most closely related to archaeal genes, whereas operational genes (genes involved in cellular metabolic processes such as amino acid biosynthesis, cell envelope and lipid synthesis, and so on) are most closely related to eubacterial genes[13]. Recently, comprehensive analyses involving large numbers of genomes[14] have shown the strength of this correlation.

It has been difficult to reconcile these conflicting results with the origin of eukaryotes because of the complicating effects of genome fusions and horizontal (or lateral) gene transfer (HGT) on phylogenetic reconstructions. Genome fusions convert trees into rings, which cannot be analysed by conventional phylogenetic algorithms. Although multiple, independent events of HGT cannot make rings, the extensive horizontal transfer observed among prokaryotes[13,15–19] can obscure the identities of those prokaryotes that may have contributed genes to eukaryotes.

Recently a new algorithm, conditioned reconstruction, based on the two character states of gene presence and absence has been developed, which can reconstruct genome fusions[20]. Analyses of the presence or absence of genes have been used previously to reconstruct phylogenetic trees[21–26], but these methods cannot detect genome fusions. Conditioned reconstructions, when used in conjunction with Markov-based quartet methods, can rigorously analyse the fusion of two genomes and ameliorate biases due to HGT. Here we apply this new method to investigate the evolution of eukaryotes, and present evidence that the eukaryotic genome arose through the fusion of two diverse prokaryotic genomes.

The topology of the ring of life

Ten complete genomes from prokaryotic and eukaryotic organisms, comprising a representative sampling of the diversity of life, were analysed using the method of conditioned reconstruction to obtain a better understanding of eukaryotic origins. (See Supplementary Data for the analyses of 24 further prokaryotic genomes and additional control experiments.) The five most probable trees from a set of three Bacteria, three Archaea and two eukaryotes are shown in Fig. 1 (the names of the taxa are listed in the legend). The

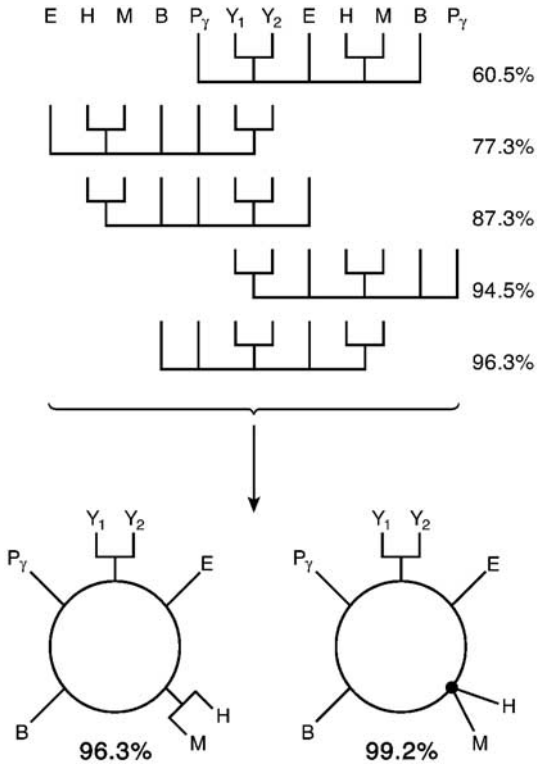

Figure 1 Conditioned reconstructions provide evidence for the ring of life. The genomes are from two yeasts (Y_1, *Schizosaccharomyces pombe* and Y_2, *Saccharomyces cerevisiae*), a γ-proteobacterium (P_γ, *Xylella fastidiosa*), a bacillus (B, *Staphylococcus aureus* MW2), a halobacterium (H, *Halobacterium* sp. NRC-1), a methanococcus (M, *Methanosarcina mazei* Goe1), an eocyte (E, *Sulfolobus tokodaii*) and an archaeoglobium (not shown; the conditioning genome *Archaeoglobus fulgidus* DSM4304). The five most probable unrooted trees are shown with leaves pointing upward to emphasize that each is part of a repeating pattern. Cumulative probabilities are shown at the right of each tree. Fully and partially resolved rings are at the lower left and right, respectively.

cumulative probabilities of these five trees are shown at the right of each tree. The upper tree is supported by 60.5% of the conditioned reconstruction bootstraps, the second tree by 16.8% of the conditioned reconstruction bootstraps, for a cumulative probability of 77.3%, and so on. It initially appears that the resolution of the tree is poor because the most probable tree is supported by a low bootstrap value, and the other trees are supported by even lower values. However, when the five most probable unrooted trees are aligned by shifting each to the left or the right, until their leaves match, they form a repeating pattern indicating that the five trees are simply permutations of an underlying cyclic pattern. This implies that they are derived from the single cycle graph (ring) shown at the bottom left of Fig. 1 (ref. 20). When that ring is cut at any one of the five central arcs and then unfolded, the resulting unrooted tree will correspond to one of the five most probable trees. In other words, the data are not tree-like they are ring-like.

A rigorous combinatorial analysis of the genomic fusion of two organisms valid for all possible initial pre-fusion trees has shown that the conditioned reconstruction algorithm recovers all permutations of the cycle graph, and only those permutations[20]. Hence, these results may be interpreted, in a manner analogous to the interpretation of restriction digests of a circular plasmid or the mapping of a circular chromosome, as implying a ring of life. The ring shown at the lower left of Fig. 1 is completely consistent with all five of the fully resolved trees shown at the top of the figure, hence we refer to it as a fully resolved ring. That ring explains 96.3% of the bootstrap replicates, and the partially resolved ring at the lower right of Fig. 1 explains 99.2% of the bootstrap replicates. The ring is supported for all possible conditioning genomes (see Supplementary Data). This provides strong evidence for the completely resolved ring on the left, and even stronger evidence for the less-resolved ring on the right. From these results we infer that the eukaryotic nuclear genome was formed from the fusion of the genomes of a relative of a proteobacterium (P_γ) and a relative of an archaeal eocyte (E).

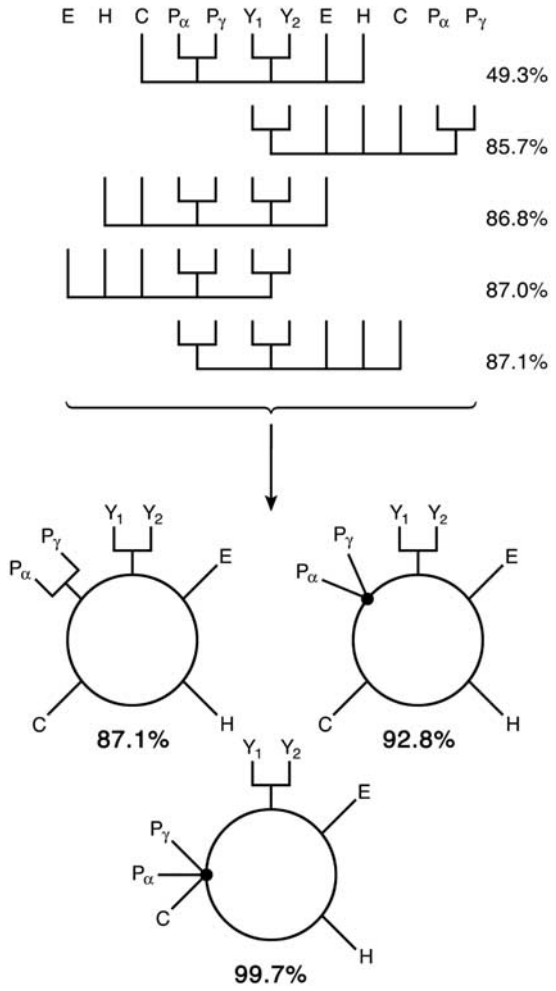

Figure 2 Eubacterial relationships within the ring of life. The genomes are from two yeasts (Y_1, *S. pombe* and Y_2, *S. cerevisiae*), a γ-proteobacterium (P_γ, *X. fastidiosa* 9a5c), an α-proteobacterium (P_α, *Brucella melitensis* 16M), a cyanobacterium (C, *Synechocystis* sp. PCC6803), a halobacterium (H, *Halobacterium* sp. NRC-1), an eocyte (E, *S. tokodaii*) and a bacillus (not shown; the conditioning genome *S. aureus* MW2). The five unrooted trees consistent with the ring are shown with leaves pointing upward to emphasize that each is part of a repeating pattern. Cumulative probabilities are at the right of each tree. Fully and partially resolved rings are at the lower left, right and centre, respectively.

Eubacterial roots of the ring of life

By analysing additional taxa, one can discover whether the eubacterial fusion partner branches even deeper than the γ-proteobacterium *Xylella*. The data set for this second analysis, described in the legend for Fig. 2, specifically included the α-proteobacterium *Brucella* and the cyanobacterium *Synechocystis*. The results are shown in Fig. 2. The cumulative probability of the five trees, shown at the top of the figure, corresponds to 87.1% of the total bootstraps, indicating significant statistical support for the ring shown at the lower left of Fig. 2. The percentage increases to 92.8% when the α-proteobacterium and the γ-proteobacterium are allowed to form an unresolved bifurcation, and rises to 99.7% when the α-proteobacterium, γ-proteobacterium and cyanobacterium are allowed to form an unresolved trifurcation. The ring is also supported for all possible conditioning genomes (see Supplementary Data). The combined analyses in Figs 1 and 2 provide strong evidence for the ring in Fig. 3, and indicate that the immediate relative of the eubacterial fusion partner was probably a primitive proteobacterium, or possibly a primitive photosynthetic eubacterium.

The second analysis, illustrated in Fig. 2, did not detect an additional ring connecting the α-Proteobacteria, the phylum from which mitochondria arose[27], to the eukaryotes. There is increasing evidence that eukaryotes have received a deluge of DNA from organelles[28]. In one study, 630 nuclear-encoded genes in eukaryotes

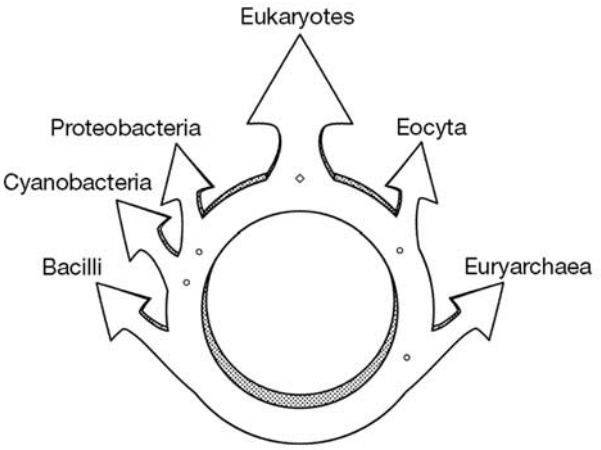

Figure 3 A schematic diagram of the ring of life. The eukaryotes plus the two eukaryotic root organisms (the operational and informational ancestors) comprise the eukaryotic realm (see Supplementary Discussion). Ancestors defining major groups in the prokaryotic realm are indicated by small circles on the ring. The Archaea[49], shown on the bottom right, includes the Euryarchaea, the Eocyta and the informational eukaryotic ancestor. The Karyota[5], shown on the upper right of the ring, includes the Eocyta and the informational eukaryotic ancestor. The upper left circle includes the Proteobacteria[49] and the operational eukaryotic ancestor. The most basal node on the left represents the photosynthetic prokaryotes and the operational eukaryotic ancestor.

were identified as having mitochondrial antecedents[29]. A substantial and continuous influx of mitochondrial DNA to the eukaryotic nucleus has been documented[28–33], and the α-proteobacterial relatives of mitochondria might have undergone HGT with other prokaryotic groups[16–19]. Hence the present data do not reveal whether the eubacterial fusion partner was distinct from the ancestor of mitochondria or identical with it. Further lineage sampling with conditioned reconstruction among α-Proteobacteria and other prokaryotes might help to resolve this issue.

Identifying the fusion organism

We have, so far, interpreted these results as strongly supporting the conclusion that two prokaryotes fused their genomes, thereby closing the ring of life and creating the first eukaryote. However the formal alternative exists that it was a prokaryote rather than a eukaryote that resulted from the genome fusion and closed the ring of life. Hence we explicitly tested the identity of the fusion organism as follows.

In tree analyses, each leaf contacts only one node of the tree, so that eliminating one taxon from the analysis will delete that leaf, but not otherwise change the tree. Similarly, in a conditioned reconstruction analysis, eliminating a non-fusion organism from the analysis will delete that leaf from the tree, or ring, without affecting the ring. However, eliminating a fusion organism, which necessarily contacts two nodes of a ring, will delete the leaf, open the ring and convert it into a tree[20]. When the taxa that comprise the ring in Fig. 2 were systematically removed, the ring opened only when both yeast genomes were simultaneously removed, indicating that the eukaryotic genome had inherited genes from its prokaryotic fusion partners. When the eukaryotes were removed, the fully resolved tree (Fig. 2, lower left) minus eukaryotes was supported by 88.6% of the bootstraps. The poorly resolved tree (Fig. 2, bottom) minus eukaryotes was supported by 99.96% of the bootstraps, thereby demonstrating that the yeast lineage is the fusion product of prokaryotes, as illustrated in Fig. 3. Furthermore, the conditioned reconstruction analysis of an additional 24 prokaryotic genomes in the absence of eukaryotic genomes reconstructed only trees, and not rings (Supplementary Fig. S1). This adds further support to the conclusion that eukaryotes are indeed the products of genome fusions.

Discussion

Various theories have been proposed for the origin of the nuclear genes of eukaryotes. These include autogenous theories, chimaeric theories and genome fusion theories. The results derived in this paper argue against autogenous theories; that is, tree of life theories in which eukaryotes are proposed to have evolved clonally from a single, possibly very ancient, prokaryote. In this paper chimaeric theories refer to the acquisition of genes by eukaryotes from multiple sources through unspecified mechanisms. The data presented here argue against them, except of course for chimaeric theories that specifically propose genome fusions.

Genome fusion theories, in which eukaryotic nuclear genes are obtained through the fusion of two diverse genomes, are strongly supported by the analyses presented here. By default, an endosymbiosis between two prokaryotes is probably the mechanism responsible for the genome fusion described here, although the fusion signal may have been augmented by gene contributions from eukaryotic organelles. Symbiotic relationships are fairly common among organisms living together, and in rare cases this leads to endosymbiosis, the intracellular capture of former symbionts[34]. We and others have previously proposed endosymbiotic theories for the origin of eukaryotes[26,35–39]. Given a genome fusion, and in the absence of other mechanisms that could produce fusions, we conclude that an endosymbiosis was the probable cause.

The ring of life is consistent with, and confirms and extends, a number of previously reported results. It implies that prokaryotes pre-date eukaryotes, as two pre-existing prokaryotes contributed their genomes to create the first eukaryotic genome. This probably places the root of the ring below the eubacterial–eukaryotic last common ancestor and the eocytic–eukaryotic last common ancestor, as shown in Fig. 3. This partial rooting of the ring of life is consistent with the eukaryotic rooting implied by the EF-1α insert that is present in all known eukaryotic and eocytic EF-1α sequences and lacking in all paralogous EF-G sequences[12,40]. Given the limited sampling here, the prokaryotic connections in the ring are broadly consistent with the results from insert distributions and concatenated gene trees[12,41–43].

The ring of life explains a number of confusing observations that motivated this study. Because the eukaryotic genome resulted from a fusion it is expected that in some gene trees eukaryotes will be related to Bacteria, whereas in other gene trees eukaryotes will be related to Archaea, in accord with the results of others[9–12]. Our observations and those of others[13,14] suggesting that the informational genes of eukaryotes are primarily derived from Archaea and the operational genes are primarily derived from Bacteria are also consistent with the ring of life. Those observations suggest that the operational genes have come from the eubacterial fusion partner and the informational genes from the archaeal fusion partner. The ring of life does not explain why this happened, but it does provide a broad phylogenetic framework for testing theories for the origin and evolution of the eukaryotic genome. □

Methods

Conditioned reconstructions[20] use Markov-based methods to determine phylogenetic trees and graphs based on the two character states 'gene presence' and 'gene absence' defined by gene orthologue sets. Gene orthologue sets used in these reconstructions were calculated in a three-step process to ensure that each gene orthologue set was (1) a globally optimal, strongly connected digraph, with orthologues explicitly distinguished from (2) paralogues and (3) recent duplications.

Globally optimal orthologue sets were constructed using gapped BLAST analyses (BLASTP, v.2) using the BLOSUM62 matrix and default parameters[44]. Because the top BLAST score is often not the most closely related orthologue[41], we used the best three scores to construct preliminary orthologue sets (for details see ref. 20). Each gene orthologue set was defined by using a specific gene in the conditioning genome for reference. For example, if *Escherichia coli* was used as the conditioning genome, and the EF-Tu gene was used for reference, each non-*E. coli* genome would be searched for the top three BLAST hits for *E. coli* EF-Tu. From this collection, 2,187 unique sets of orthologues (3^{n-1} where n equals eight genomes total) were constructed. The sum of all possible pairwise BLAST scores between orthologues was calculated for each set. The set with the highest sum was chosen as a tentative orthologue set. In the second step of the procedure, gene paralogues were detected through computations based on the properties of digraphs. We have observed that paralogues are frequently unilaterally or even weakly connected to other nodes in a BLAST-based digraph. Hence each tentative orthologue set was systematically analysed to determine all non-empty, strongly connected subsets of the tentative orthologue set. These subsets of putative orthologous genes were ordered according to the sum of their reciprocal BLAST scores, from largest to smallest. The final step identified recently duplicated genes based on the observation that they are separated from each other by less time than they are from other members of an orthologue set. Hence BLAST scores between recently duplicated genes are, on average, larger than scores between orthologous genes. The top three tentative recent duplications for each member of the orthologue sets were examined and putative recent duplications were identified. Simultaneously, through a sequential, iterative process, gene orthologue sets were selected from largest to smallest BLAST scores making sure that every gene was contained in one, and only one, orthologue set.

Eight genomes were analysed to compute the global orthologue alignments. Organisms included in the analyses were chosen according to two criteria: they should span the diversity of life (see Supplementary Data for an analysis of 24 additional prokaryotic genomes) and their genomes should contain similar numbers of genes[20]. The slowly evolving yeast sequences facilitated comparisons with prokaryotic sequences (*Schizosaccharomyces pombe*, 5,000 genes, and *Saccharomyces cerevisiae*, 6,329). Both yeast genomes contained all nuclear genes, including those annotated as mitochondrial. The prokaryotes included four Bacteria (a γ-proteobacterium, *Xylella fastidiosa* 9a5c, 2,766 genes; an α-proteobacterium, *Brucella melitensis* 16M, 3,198; a cyanobacterium, *Synechocystis* sp. PCC6803, 3,169; and a bacillus, *Staphylococcus aureus* MW2, 2,632) and four Archaea (a halobacterium, *Halobacterium* sp. NRC-1, 2,605; a methanococcus, *Methanosarcina mazei* Goe1, 3,371; an archaeoglobium, *Archaeoglobus fulgidus* DSM4304, 2,408; and an eocyte, *Sulfolobus tokodaii*, 2,826). *Archaeoglobus* and *Staphylococcus* were used as the conditioning genomes for Figs 1 and 2, respectively. The analyses shown in Fig. 1 are based on 2,408 orthologous gene sets. Of these, 433 sets contained all seven orthologues, whereas 239, 285, 292, 329, 293, 204 and 333 contained six, five, four, three, two, one and no orthologues, respectively. The analyses shown in Fig. 2 are based on 2,632 orthologous gene sets. Of these, 432 sets contained all seven orthologues, whereas 191, 208,

208, 230, 218, 266 and 879 contained six, five, four, three, two, one and no orthologues, respectively.

The big genome attraction artefact was partially compensated by using Paralinear/LogDet distances[45,46] for tree reconstruction, and the artefact was further reduced by using genomes of similar size; that is, containing similar numbers of genes. Bootstrappers Gambit[47] was used for the analyses as it, and possibly most quartet methods, can naturally accommodate graphs. Corrections for site-to-site variation used Pattern Filtering[48] and were applied to the gene order found in the conditioning genome, as this can considerably reduce errors introduced by violations of the assumption that sites are independent and identically distributed. The big genome attraction artefact did not appear to be significant for this set as the largest prokaryotic genomes were not directly connected to the eukaryotes in Figs 1 and 2.

Received 29 January; accepted 15 July 2004; doi:10.1038/nature02848.

1. Lake, J. A., Sabatini, D. D. & Nonomura, Y. in *Ribosomes* (eds Nomura, M., Tissieres, A. & Lengyel, P.) 543–557 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1974).
2. Woese, C. R. Archaebacteria. *Sci. Am.* **244,** 98–105 (1981).
3. Dayhoff, M. O. *Atlas of Protein Sequence and Structure* (National Biomedical Research Foundation, Silver Spring, Maryland, 1972).
4. Pace, N. R., Olsen, G. J. & Woese, C. R. Ribosomal RNA phylogeny and the primary lines of evolutionary descent. *Cell* **45,** 325–326 (1986).
5. Lake, J. A. Origin of the eukaryotic nucleus determined by rate-invariant analysis of ribosomal RNA sequences. *Nature* **331,** 184–186 (1988).
6. Galtier, N., Tourasse, N. & Gouy, M. A nonhyperthermophilic common ancestor to extant life forms. *Science* **283,** 220–221 (1999).
7. Gogarten, J. P. *et al.* Evolution of the vacuolar H+-Atpase—implications for the origin of eukaryotes. *Proc. Natl Acad. Sci. USA* **86,** 6661–6665 (1989).
8. Iwabe, N., Kuma, K., Hasegawa, M., Osawa, S. & Miyata, T. Evolutionary relationship of archaebacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes. *Proc. Natl Acad. Sci. USA* **86,** 9355–9359 (1989).
9. Martin, W., Mustafa, A. Z., Henze, K. & Schnarrenberger, C. Higher-plant chloroplast and cytosolic fructose-1,6-bisphosphatase isoenzymes: Origins via duplication rather than prokaryote-eukaryote divergence. *Plant Mol. Biol.* **32,** 485–491 (1996).
10. Brown, J. R. & Doolittle, W. F. Archaea and the prokaryote-to-eukaryote transition. *Microbiol. Mol. Biol. Rev.* **61,** 456–502 (1997).
11. Feng, D. F., Cho, G. & Doolittle, R. F. Determining divergence times with a protein clock: Update and reevaluation. *Proc. Natl Acad. Sci. USA* **94,** 13028–13033 (1997).
12. Gupta, R. S. Protein phylogenies and signature sequences: A reappraisal of evolutionary relationships among archaebacteria, eubacteria, and eukaryotes. *Microbiol. Mol. Biol. Rev.* **62,** 1435–1491 (1998).
13. Rivera, M. C., Jain, R., Moore, J. E. & Lake, J. A. Genomic evidence for two functionally distinct gene classes. *Proc. Natl Acad. Sci. USA* **95,** 6239–6244 (1998).
14. Esser, C. *et al.* A genome phylogeny for mitochondria among α-Proteobacteria and a predominantly eubacterial ancestry of yeast nuclear genes. *Mol. Biol. Evol.* doi:10.1093/molbev/msh160 (2004).
15. Karlin, S., Mrazek, J. & Campbell, A. M. Compositional biases of bacterial genomes and evolutionary implications. *J. Bacteriol.* **179,** 3899–3913 (1997).
16. Gogarten, J. P., Hilario, E. & Olendzenski, L. The tree of life. *ASM News* **63,** 404–405 (1997).
17. Doolittle, W. F. Phylogenetic classification and the universal tree. *Science* **284,** 2124–2128 (1999).
18. Campbell, A. M. Lateral gene transfer in prokaryotes. *Theor. Popul. Biol.* **57,** 71–77 (2000).
19. Ochman, H. & Jones, I. B. Evolutionary dynamics of full genome content in *Escherichia coli*. *EMBO J.* **19,** 6637–6643 (2000).
20. Lake, J. A. & Rivera, M. C. Deriving the genomic tree of life in the presence of horizontal gene transfer: Conditioned Reconstruction. *Mol. Biol. Evol.* **21,** 681–690 (2004).
21. Dickerson, R. E. in *Diffraction and Related Studies* (ed. Srinivasan, R.) 227–249 (Pergamon, Oxford/New York, 1980).
22. Snel, B., Bork, P. & Huynen, M. A. Genome phylogeny based on gene content. *Nature Genet.* **21,** 108–110 (1999).
23. Fitz-Gibbon, S. T. & House, C. H. Whole genome-based phylogenetic analysis of free-living microorganisms. *Nucleic Acids Res.* **27,** 4218–4222 (1999).
24. Tekaia, F., Lazcano, A. & Dujon, B. The genomic tree as revealed from whole proteome comparisons. *Genome Res.* **9,** 550–557 (1999).
25. Montague, M. G. & Hutchison, C. A. Gene content phylogeny of herpesviruses. *Proc. Natl Acad. Sci. USA* **97,** 5334–5339 (2000).
26. Lake, J. A., Henderson, E., Clark, M. W. & Matheson, A. T. Mapping evolution with ribosome structure: Intralineage constancy and interlineage variation. *Proc. Natl Acad. Sci. USA* **79,** 5948–5952 (1982).
27. Gray, M. W., Burger, G. & Lang, B. F. Mitochondrial evolution. *Science* **283,** 1476–1481 (1999).
28. Timmis, J. N., Ayliffe, M. A., Huang, C. Y. & Martin, W. Endosymbiotic gene transfer: Organelle genomes forge eukaryotic chromosomes. *Nature Rev. Genet.* **5,** 123–135 (2004).
29. Gabaldon, T. & Huynen, M. A. Reconstruction of the proto mitochondrial metabolism. *Science* **301,** 609 (2003).
30. Adams, K. L., Daley, D. O., Qiu, Y. L., Whelan, J. & Palmer, J. D. Repeated, recent and diverse transfers of a mitochondrial gene to the nucleus in flowering plants. *Nature* **408,** 354–357 (2000).
31. Gray, M. W. Evolution of organellar genomes. *Curr. Opin. Genet. Dev.* **9,** 678–687 (1999).
32. Collura, R. V. & Stewart, C. B. Insertions and duplications of mtDNA in the nuclear genomes of old-world monkeys and hominoids. *Nature* **378,** 485–489 (1995).
33. Zischler, H., Geisert, H., vonHaeseler, A. & Paabo, A. A nuclear fossil of the mitochondrial D-loop and the origin of modern humans. *Nature* **378,** 489–492 (1995).
34. Margulis, L. *Origin of the Eukaryotic Cells* (Yale Univ. Press, New Haven, 1970).
35. Gupta, R. S., Aitken, K., Falah, M. & Singh, B. Cloning of *Giardia lamblia* heat-shock protein Hsp70 homologs—Implications regarding origin of eukaryotic cells and of endoplasmic-reticulum. *Proc. Natl Acad. Sci. USA* **91,** 2895–2899 (1994).
36. Martin, W. & Muller, M. The hydrogen hypothesis for the first eukaryote. *Nature* **392,** 37–41 (1998).
37. Lake, J. A. & Rivera, M. C. Was the nucleus the 1st endosymbiont. *Proc. Natl Acad. Sci. USA* **91,** 2880–2881 (1994).
38. Moreira, D. & Lopez-Garcia, P. Symbiosis between methanogenic archaea and δ-proteobacteria as the origin of eukaryotes: The syntrophic hypothesis. *J. Mol. Evol.* **47,** 517–530 (1998).
39. Horiike, T., Hamada, K., Kanaya, S. & Shinozawa, T. Origin of eukaryotic cell nuclei by symbiosis of Archaea in Bacteria is revealed by homology-hit analysis. *Nature Cell Biol.* **3,** 210–214 (2001).
40. Rivera, M. C. & Lake, J. A. Evidence that eukaryotes and eocyte prokaryotes are immediate relatives. *Science* **257,** 74–76 (1992).
41. Daubin, V., Gouy, M. & Perriere, B. A phylogenomic approach to bacterial phylogeny: Evidence of a core of genes sharing a common history. *Genome Res.* **12,** 1080–1090 (2002).
42. Brochier, C., Forterre, P. & Gribaldo, S. Archaeal phylogeny based on proteins of the transcription and translation machineries: tackling the *Methanopyrus kandleri* paradox. *Genome Biol.* **5,** R17 (2004).
43. Wolf, Y. I., Rogozin, I. B., Grishin, N. V., Tatusov, R. L. & Koonin, E. V. Genome trees constructed using five different approaches suggest new major bacterial clades. *BMC Evol. Biol.* **1,** 8 (2001).
44. Altschul, S. F. *et al.* Gapped BLAST and PSI_BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25,** 3389–3402 (1997).
45. Lake, J. A. Reconstructing evolutionary trees from DNA and protein sequences—Paralinear distances. *Proc. Natl Acad. Sci. USA* **91,** 1455–1459 (1994).
46. Lockhart, P. J., Steel, M. A., Hendy, M. D. & Penny, D. Recovering evolutionary trees under a more realistic model of sequence evolution. *Mol. Biol. Evol.* **11,** 605–612 (1994).
47. Lake, J. A. Calculating the probability of multitaxon evolutionary trees—Bootstrappers Gambit. *Proc. Natl Acad. Sci. USA* **92,** 9662–9666 (1995).
48. Lake, J. A. Optimally recovering rate variation information from genomes and sequences: Pattern filtering. *Mol. Biol. Evol.* **15,** 1224–1231 (1998).
49. Woese, C. R., Kandler, O. & Wheelis, M. L. Towards a natural system of organisms: Proposal for the domains *Archaea, Bacteria* and *Eucarya*. *Proc. Natl Acad. Sci. USA* **87,** 4576–4579 (1990).

Supplementary Information accompanies the paper on **www.nature.com/nature**.

Acknowledgements We thank A. B. Simonson for helpful discussions and editing, M. Kowalczyk for illustrations, J. Garey for suggesting the original title 'One ring to rule them all', and J. A. Servin and R. G. Skophammer for suggestions. This work was supported by grants from the National Science Foundation, NASA Astrobiology Program, The Department of Energy, and the National Institutes of Health to J.A.L.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.A.L. (lake@mbi.ucla.edu).