biology letters
Genome biology

Biol. Lett. (2007) **3**, 180–184
doi:10.1098/rsbl.2006.0582
Published online 19 December 2006

The origin of mitochondria in light of a fluid prokaryotic chromosome model

Christian Esser, William Martin and Tal Dagan*

Institut für Botanik III, Heinrich-Heine Universität Düsseldorf, Universitätsstr. 1, 40225 Düsseldorf, Germany
**Author for correspondence* (*tal.dagan@uni-duesseldorf.de*).

Biologists agree that the ancestor of mitochondria was an α-proteobacterium. But there is no consensus as to what constitutes an α-proteobacterial gene. Is it a gene found in all or several α-proteobacteria, or in only one? Here, we examine the proportion of α-proteobacterial genes in α-proteobacterial genomes by means of sequence comparisons. We find that each α-proteobacterium harbours a particular collection of genes and that, depending upon the lineage examined, between 97 and 33% are α-proteobacterial by the nearest-neighbour criterion. Our findings bear upon attempts to reconstruct the mitochondrial ancestor and upon inferences concerning the collection of genes that the mitochondrial ancestor possessed at the time that it became an endosymbiont.

Keywords: genomics; hydrogenosomes; lateral gene transfer; microbial evolution; endosymbiosis; *Magnetococcus*

1. INTRODUCTION

There is consensus among biologists that mitochondria descend from free-living prokaryotes and that the organelle arose only once during evolution (Gray *et al.* 1999; Dolezal *et al.* 2006). There is considerably less agreement concerning the biochemical capabilities and phylogenetic affinity of the mitochondrial ancestor. Various eubacterial groups have been proposed as the ancestor of mitochondria. Even before the time of molecular phylogenies, interest in this topic has focused upon the purple non-sulphur bacteria (John & Whatley 1975), later renamed as α-proteobacteria (Stackebrandt *et al.* 1988).

The conventional approach to identify the mitochondrial ancestor is founded in the comparison of mitochondrially encoded genes with those in the genomes of free-living prokaryotes. By this means, early analyses of 16S rRNA suggested *Agrobacterium tumefaciens* to be the closest relative of the mitochondrion (Yang *et al.* 1985). More recent studies have attributed the mitochondrial ancestor to the Rickettsiales order, which mostly contains parasitic species with highly reduced genomes (Lang *et al.* 1999; Emelyanov 2003). Other studies have pointed specifically to *Rickettsia prowazekii* as the ancestral genome (Andersson *et al.* 1998), or a common ancestor of *Rickettsia* and *Wolbachia* (Wu *et al.* 2004), while others still have implicated larger genomes, free-living representatives (Esser *et al.* 2004). A different approach to the issue was taken by Gabaldon & Huynen (2003), who inferred the kinds of biochemical pathways that the mitochondrion possessed, without addressing the nearest neighbour of the organelle among free-living groups. Yet, a different approach to the issue entails the study of nuclear-encoded proteins shared by mitochondria and hydrogenosomes—the ATP- and H_2-producing mitochondria of anaerobic eukaryotes (Müller 2003)—and inferences about the physiology of their free-living ancestor (van der Giezen & Tovar 2005; Embley & Martin 2006).

The nearest neighbour of mitochondria among free-living α-proteobacteria is still unknown (Lang *et al.* 1999; Esser *et al.* 2004). At the same time, gene content in bacterial genomes is variable over time owing to inheritance, mutation, gene loss and lateral gene transfer (LGT) events (Lawrence & Ochman 1998; Martin 1999; Doolittle 2004; Kunin *et al.* 2005; Lerat *et al.* 2005). Here, we examine the phylogenetic affinities of the 47 143 proteins encoded among 18 α-proteobacterial genomes by means of nearest-neighbour comparisons.

2. MATERIAL AND METHODS

(a) *Data*

Prokaryotic and mitochondrial genomes were downloaded from the NCBI website (http://www.ncbi.nlm.nih.gov/; versions of April 2005; table S1 in electronic supplementary material). All 288 prokaryotic genomes were formatted into a single Blast (Altschul *et al.* 1990) database. Eighteen α-proteobacterial and six mitochondrial genomes were used as queries, including only proteins longer than 50 amino acids.

(b) *Nearest-neighbour inference*

The nearest neighbour of each protein was inferred by a best Blast hit (BBH) approach and a phylogenetic tree approach using the neighbour-joining (Saitou & Nei 1987) and maximum-likelihood methods. Neither approach is infallible (Koski & Golding 2001; Penny *et al.* 2001); we used both approaches for comparison.

For BBH analysis, each protein in each query genome was blasted against the 288 genome database. The nearest neighbour was defined as the BBH above an *E*-value of 10^{-20} that is neither the query protein nor stems from the same genus as the query protein. The taxonomic group of the nearest neighbours was specified as the phylum according to the NCBI taxonomy (http://www.ncbi.nlm.nih.gov/Taxonomy/), or as the class in the case of proteobacteria.

In the NJ approach, the BBHs from each genus were selected. These were aligned with ClustalW (Thompson *et al.* 1994), protein distances were calculated with Protdist (Felsenstein 2005) using the JTT matrix, and used to reconstruct an NJ tree with Neighbour (Felsenstein 2005) using 100 bootstrap replicates. Maximum-likelihood trees were reconstructed using FastML (Pupko *et al.* 2000). The nearest neighbour was defined as the operational taxonomic unit with smallest sum of branch lengths to the query protein that appears in more than or equal to 90% of the replicates.

3. RESULTS

If every gene contained within an α-proteobacterial genome were of an α-proteobacterial origin (i.e. most closely related to homologues in other α-proteobacteria), then every nearest neighbour of every gene in each α-proteobacterial genome would be found in another α-proteobacterium. Our results (figure 1) indicate that α-proteobacterial genomes are mosaic to varying degrees.

The highest proportion of α-proteobacterial BBH nearest neighbours (92%) was found in *Sinorhizobium meliloti*, while the lowest proportion (64%) was found in *Magnetospirillum magnetotacticum*. An even lower proportion of α-proteobacterial BBH nearest neighbours

Electronic supplementary material is available at http://dx.doi.org/10.1098/rsbl.2006.0582 or via http://www.journals.royalsoc.ac.uk.

Received 25 October 2006
Accepted 30 November 2006

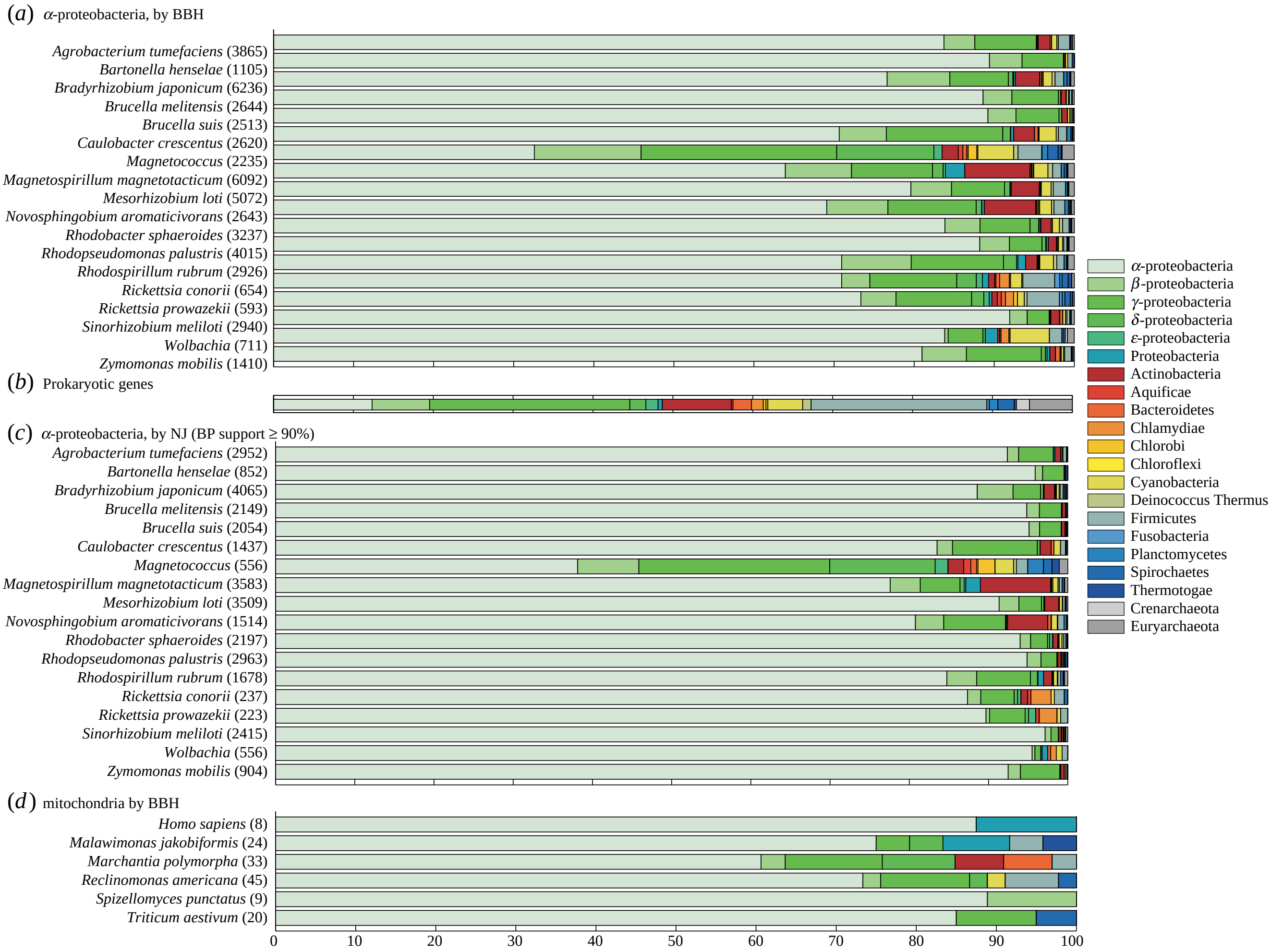

Figure 1. Distribution across taxonomic groups for (*a*) α-proteobacterial nearest neighbours by BBH, (*b*) prokaryotic genes, (*c*) α-proteobacterial nearest neighbours by NJ and (*d*) mitochondrial nearest neighbours by BBH. The number of query proteins used for the analysis is shown to the left of the species name (*y*-axis).

(33%) was detected in *Magnetococcus* sp., which is currently classified as an unclassified proteobacteria, but shows clear resemblance to α-proteobacteria (see figure 2). On average, 77±13% of the proteins in α-proteobacterial genomes, sampled here, had their nearest neighbour in another α-proteobacterium. The remainder of the BBH nearest neighbours of α-proteobacterial genes was found in other proteobacterial classes (β, 5±3%; δ, 1±3%; ε, 0.2±0.2%; and γ, 9±5%) or outside the proteobacterial phylum (7±4%). The most frequent non-proteobacterial nearest neighbours are actinobacterial, cyanobacterial and firmicute genes (figure 1*a*; table S2 in electronic supplementary material). In all α-proteobacterial genomes, these frequencies deviate significantly ($p<0.05$, using χ^2-test with Bonferroni correction) from the taxonomic distribution of the prokaryotic genes in our data (figure 1*b*), hence our results are not random.

The distribution of NJ nearest neighbours of α-proteobacterial genes is similar to the results of the BBH method (figure 1*c*). The majority of the genes (87±13%) have an α-proteobacterial nearest neighbour, while other frequent nearest neighbours are γ-proteobacterial and actinobacterial genes (table S3 of electronic supplementary material).

The BBH nearest-neighbour analysis of the mitochondrial genes results in a phyletic distribution that is similar to that of the α-proteobacterial genes (figure 1*d*). The majority of the nearest neighbours are α-proteobacterial (82±7%), and additional frequent taxa are γ-proteobacteria and firmicutes (table S2 of electronic supplementary material).

The vast majority (93±4%) of the BBH nearest neighbours among proteins encoded within the 18 α-proteobacterial genomes sampled here reside within genomes of the proteobacterial phylum (figure 1). The α-proteobacterial genomes sampled encode many open reading frames, sequences that have no known homologues (17 953 in all α-proteobacteria), and whose history cannot presently be addressed by sequence comparisons. In the present study, the BBH approach detected, on average, no nearest neighbour for 27±10% of the proteins in each α-proteobacterial genome (figure S1 in electronic supplementary material). Our results were found to be independent of the tree size or the sampled species (figures S2–S4 in electronic supplementary material). Using the ML reconstruction, the proportion of α-proteobacterial nearest neighbours was lower by approximately 10% (figure S5 in electronic supplementary material).

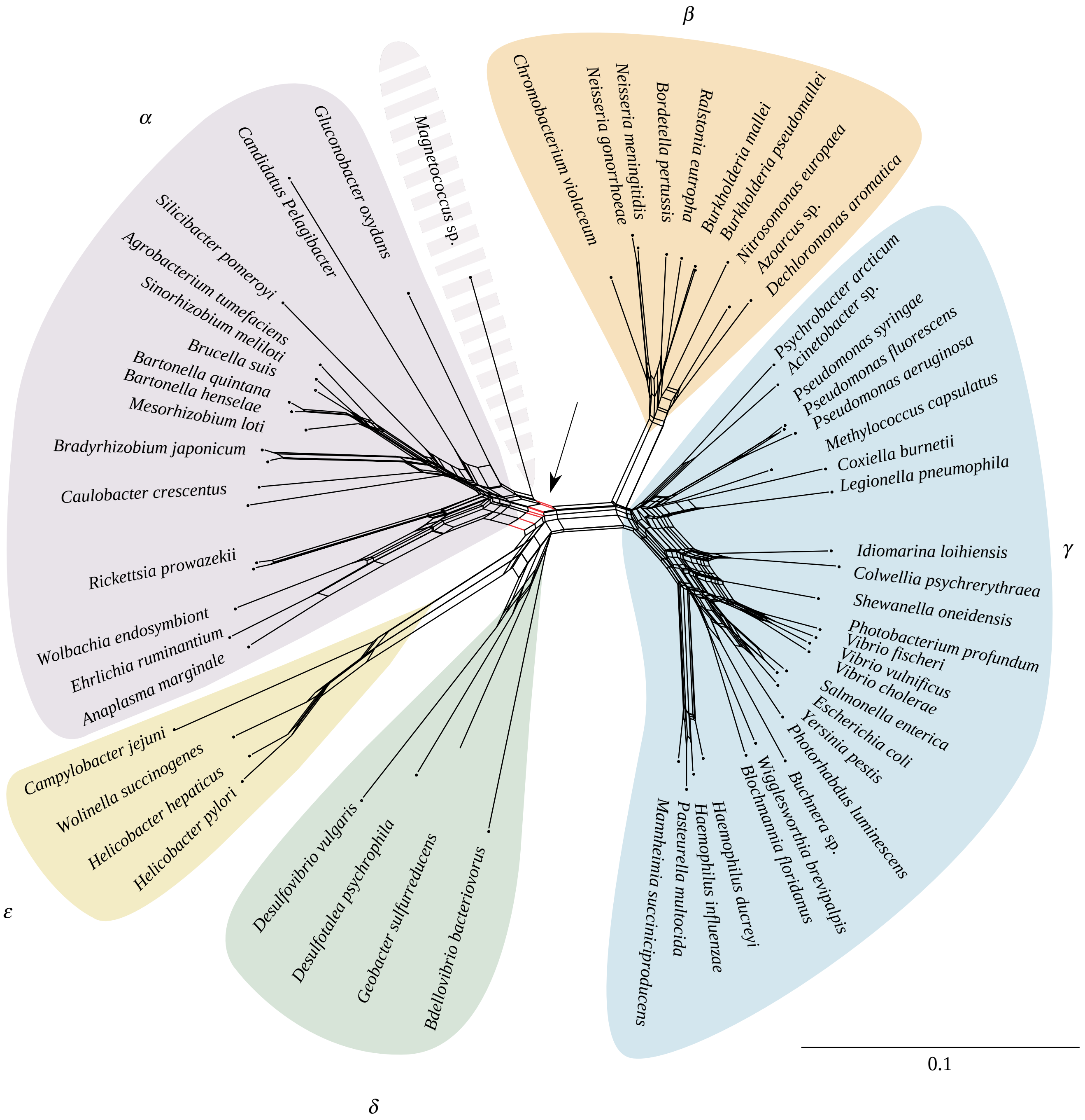

Figure 2. Neighbour-Net (Huson & Bryant 2006) of proteobacterial 16S rRNA. The bootstrap support for the split of *Magnetococcus* with α-proteobacteria (highlighted in red and with arrow) is 73% using neighbour-joining, 61% using Neighbour-Net and 45% using maximum likelihood (see electronic supplementary material).

4. DISCUSSION

On the basis of sequence similarity to α-proteobacterial homologues, it has been estimated that 630 eukaryotic genes trace to α-proteobacteria (Gabaldon & Huynen 2003). But there are thousands of eukaryotic nuclear genes that are clearly eubacterial, but not specifically α-proteobacterial, in terms of their patterns of sequence similarity (Esser *et al.* 2004; Rivera & Lake 2004; Embley & Martin 2006). Finding a eukaryotic gene that branches with a group other than α-proteobacteria is often taken as evidence for an origin from that group (for example, Baughn & Malamy 2002), the methodological problems of deep phylogenetic trees notwithstanding (Susko *et al.* 2006). But if we let go of the static prokaryotic chromosome model and assume a fluid chromosome model for prokaryotes, then the expected phylogeny for a gene acquired from the mitochondrion would be common ancestry for all eukaryotes, but not necessarily tracing to α-proteobacteria, because the ancestor of mitochondria possessed an as yet unknown collection of genes. A previous investigation of genome evolution in α-proteobacteria considered the genome size and functional classes (Boussau *et al.* 2004), but not sequence similarities. Hence, we wished to know how many of the α-proteobacterial genes pass the test of being α-proteobacterial by the nearest-neighbour criterion.

The answer, based upon the current sample, ranges from approximately 97% for *Sinorhizobium* to approximately 33% for *Magnetococcus* sp. The mitochondrial genomes studied (figure 1*d*) did not differ in terms of the nearest-neighbour composition from α-proteobacterial genomes.

Prokaryotic gene content is shaped not only by inheritance, but also by gene loss and LGT (Doolittle 2004; Kunin *et al.* 2005; Lerat *et al.* 2005). But this realization is only slowly being assimilated into thinking on the mitochondrial origin and eukaryotic gene origins (Esser *et al.* 2004). Our findings indicate that modern α-proteobacterial genomes represent transient collections of genes that stem from diverse sources. By inference, the ancestor of mitochondria had a mosaic genome as well; hence, a criterion that is often used to infer whether a eukaryotic nuclear gene of eubacterial origin stems from the mitochondrion or not—namely branching with an α-proteobacterial gene (Kurland & Andersson 2000)—is probably too strict, because it tacitly assumes a static model of bacterial chromosome evolution in which LGT and gene loss do not exist, either now or in the past. Incorporating a fluid bacterial chromosome model into endosymbiotic theory generates the prediction that nuclear genes acquired by eukaryotes from the ancestor of mitochondria should tend to reflect a single common eubacterial ancestry—provided that molecular phylogeny can accurately recover events that occurred more than 1.5 billion years ago (Embley & Martin 2006)—but that they should not necessarily belong to the known set of contemporary α-proteobacterial genes, regardless of how one were to define it.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. 1990 Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410. (doi:10.1006/jmbi.1990.9999)

Andersson, S. G. E. *et al.* 1998 The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. *Nature* **396**, 133–140. (doi:10.1038/24094)

Baughn, A. D. & Malamy, M. H. 2002 A mitochondrial-like aconitase in the bacterium *Bacteroides fragilis*: implications for the evolution of the mitochondrial Krebs cycle. *Proc. Natl Acad. Sci. USA* **99**, 4662–4667. (doi:10.1073/pnas.052710199)

Boussau, B., Karlberg, E. O., Frank, A. C., Legault, B. A. & Andersson, S. G. 2004 Computational inference of scenarios for alpha-proteobacterial genome evolution. *Proc. Natl Acad. Sci. USA* **101**, 9722–9727. (doi:10.1073/pnas.0400975101)

Dolezal, P., Likic, V., Tachezy, J. & Lithgow, T. 2006 Evolution of the molecular machines for protein import into mitochondria. *Science* **313**, 314–318. (doi:10.1126/science.1127895)

Doolittle, W. F. 2004 If the tree of life fell, would it make a sound? In *Microbial phylogeny and evolution: concepts and controversies* (ed. J. Sapp), pp. 119–133. New York, NY: Oxford University Press.

Embley, T. M. & Martin, W. 2006 Eukaryotic evolution, changes and challenges. *Nature* **440**, 623–630. (doi:10.1038/nature04546)

Emelyanov, V. V. 2003 Common evolutionary origin of mitochondrial and rickettsial respiratory chains. *Arch. Biochem. Biophys.* **420**, 130–141. (doi:10.1016/j.abb.2003.09.031)

Esser, C. *et al.* 2004 A genome phylogeny for mitochondria among alpha-proteobacteria and a predominantly eubacterial ancestry of yeast nuclear genes. *Mol. Biol. Evol.* **21**, 1643–1660. (doi:10.1093/molbev/msh160)

Felsenstein, J. 2005 *PHYLIP (phylogeny inference package)*. Seattle, DC: Department of Genome Sciences, University of Washington.

Gabaldon, T. & Huynen, M. A. 2003 Reconstruction of the proto-mitochondrial metabolism. *Science* **301**, 609. (doi:10.1126/science.1085463)

Gray, M. W., Burger, G. & Lang, B. F. 1999 Mitochondrial evolution. *Science* **283**, 1476–1481. (doi:10.1126/science.283.5407.1476)

Huson, D. H. & Bryant, D. 2006 Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* **23**, 254–267. (doi:10.1093/molbev/msj030)

John, P. & Whatley, F. R. 1975 *Paracoccus denitrificans* and the evolutionary origin of the mitochondrion. *Nature* **254**, 495–498. (doi:10.1038/254495a0)

Koski, L. B. & Golding, G. B. 2001 The closest BLAST hit is often not the nearest neighbor. *J. Mol. Evol.* **52**, 540–542.

Kunin, V., Goldovsky, L., Darzentas, N. & Ouzounis, C. A. 2005 The net of life: reconstructing the microbial phylogenetic network. *Genome Res.* **15**, 954–959. (doi:10.1101/gr.3666505)

Kurland, C. G. & Andersson, S. G. 2000 Origin and evolution of the mitochondrial proteome. *Microbiol. Mol. Biol. Rev.* **64**, 786–820. (doi:10.1128/MMBR.64.4.786-820.2000)

Lang, B. F., Gray, M. W. & Burger, G. 1999 Mitochondrial genome evolution and the origin of eukaryotes. *Annu. Rev. Genet.* **33**, 351–397. (doi:10.1146/annurev.genet.33.1.351)

Lawrence, J. G. & Ochman, H. 1998 Molecular archaeology of the *Escherichia coli* genome. *Proc. Natl Acad. Sci. USA* **95**, 9413–9417. (doi:10.1073/pnas.95.16.9413)

Lerat, E., Daubin, V., Ochman, H. & Moran, N. A. 2005 Evolutionary origins of genomic repertoires in bacteria. *PLoS Biol.* **3**, e130. (doi:10.1371/journal.pbio.0030130)

Martin, W. 1999 Mosaic bacterial chromosomes: a challenge on route to a tree of genomes. *Bioessays* **21**, 99–104. (doi:10.1002/(SICI)1521-1878(199902)21:2<99::AID-BIES3>3.0.CO;2-B)

Müller, M. 2003 Energy metabolism. Part 1: anaerobic protozoa. In *Molecular medical parasitology* (ed. J. Marr), pp. 125–139. London, UK: Academic Press.

Penny, D., McComish, B. J., Charleston, M. A. & Hendy, M. D. 2001 Mathematical elegance with biochemical realism: the covarion model of molecular evolution. *J. Mol. Evol.* **53**, 711–723. (doi:10.1007/s002390010258)

Pupko, T., Pe'er, I., Shamir, R. & Graur, D. 2000 A fast algorithm for joint reconstruction of ancestral amino acid sequences. *Mol. Biol. Evol.* **17**, 890–896.

Rivera, M. C. & Lake, J. A. 2004 The ring of life provides evidence for a genome fusion origin of eukaryotes. *Nature* **431**, 152–155. (doi:10.1038/nature02848)

Saitou, N. & Nei, M. 1987 The Neighbor-Joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406–425.

Stackebrandt, E., Murray, R. G. E. & Trüper, H. G. 1988 *Proteobacteria classis* nov. a name for the phylogenetic taxon that includes the "purple bacteria and their relatives". *Int. J. Syst. Bacteria* **38**, 321–325.

Susko, E., Leigh, J., Doolittle, W. F. & Bapteste, E. 2006 Visualizing and assessing phylogenetic congruence of core gene sets: a case study of the γ-Proteobacteria. *Mol. Biol. Evol.* **23**, 1019–1030. (doi:10.1093/molbev/msj113)

Thompson, J. D., Higgins, D. G. & Gibson, T. J. 1994 CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680.

van der Giezen, M. & Tovar, J. 2005 Degenerate mitochondria. *EMBO Rep.* **6**, 525–530. (doi:10.1038/sj.embor.7400440)

Wu, M. *et al.* 2004 Phylogenomics of the reproductive parasite *Wolbachia pipientis* wMel: a streamlined genome overrun by mobile genetic elements. *PLoS Biol.* **2**, E69. (doi:10.1371/journal.pbio.0020069)

Yang, D., Oyaizu, Y., Oyaizu, H., Olsen, G. J. & Woese, C. R. 1985 Mitochondrial origins. *Proc. Natl Acad. Sci. USA* **82**, 4443–4447. (doi:10.1073/pnas.82.13.4443)