Review

The prokaryotic tree of life: past, present...and future?

James O. McInerney[1], James A. Cotton[2] and Davide Pisani[1]

[1] Department of Biology, National University of Ireland Maynooth, Maynooth, County Kildare, Ireland
[2] School of Biological and Chemical Sciences, Queen Mary, University of London, London E1 4NS, UK

No accepted phylogenetic scheme for prokaryotes emerged until the late 1970s. Prior to that, it was assumed that there was a phylogenetic tree uniting all prokaryotes, but no suitable data were available for its construction. For 20 years, through the 1980s and 1990s, rRNA phylogenies were the gold standard. However, beginning in the last decade, findings from genomic data have challenged this new consensus. Gene trees can conflict greatly, and strains of the same species can differ enormously in genome content. Horizontal gene transfer is now known to be a significant influence on genome evolution. The next decade is likely to resolve whether or not we retain the centuries-old metaphor of the tree for all of life.

Where does the tree come from?

The use of trees to model evolutionary relationships can be traced back to the work of Jean-Baptiste Lamark. However, it was Haeckel who was mainly responsible for popularising the idea [1]. Indeed, the tree metaphor was so effective that the search for a unique tree, representing the relationships among all cellular organisms, has continued to this day [2–9].

From the very beginning, plant and animal phylogenies could be based on embryological and morphological characters. However, there was no such luxury for the prokaryotes, as these organisms lack complex intracellular structures and have extremely simple external morphologies [1]. In the first half of the last century, the lack of phylogenetically informative prokaryotic characters resulted in a great debate centered on whether or not it would ever be possible to recover a sensible phylogenetic classification for this group. Many were convinced that a simple, 'phenetic' classification of the prokaryotes was more useful than a phylogenetic one and, from its inception in 1923 and for some time afterward, *Bergey's Manual of Determinative Bacteriology* disregarded any attempt to classify prokaryotes according to phylogenetic principles. The feeling among the editors of *Bergey's Manual* was summarised by Breed [10] in 1939 when he distinguished between 'realistic workers' and 'idealists,' stating that idealists had introduced "unjustified speculations regarding relationships between the various groups of Bacteria." Stanier and van Niel [11] rejected this viewpoint and opined in 1941 that when it came to classifying prokaryotes, "there is a good reason to prefer an admittedly imperfect natural system to a purely empirical one." Interestingly, in hindsight, Stanier and van Niel queried why the absence of sexual reproduction was not included by the editors of *Bergey's Manual* as a defining feature of prokaryotes. At that time it was generally accepted that prokaryotic reproduction was clonal and that all genetic material was vertically inherited.

Glossary

Archaebacteria: Single-celled organisms lacking a nucleus (prokaryotes), with ether-linked lipids in their membranes and lacking murein in their cell walls.
Compositional bias: Not all DNA and protein sequences in all organisms have the same nucleotide or amino acid composition. Deviation from an equal composition of all nucleotides or amino acids is typical and, unless these compositional differences are adequately accounted for in evolutionary analyses, the resulting phylogenetic tree can often appear to be more like a classification of sequences according to their composition rather than their evolutionary history.
Eubacteria: Single-celled organisms lacking a nucleus (prokaryotes), with ester-linked lipids in their membranes and murein in their cell walls.
Eukaryotes: All single-celled and multicellular organisms with a membrane-bound nucleus.
Long-branch attraction: In a situation where different lineages evolve at different rates, rapidly evolving sequences can converge on the same character state. Some phylogenetic methods, most notably maximum parsimony, can interpret these homoplastic characters as synapomorphies and place the long-branch lineages as sister taxa. Methods of avoiding long-branch attraction include the use of more appropriate evolutionary models and the breaking of long branches by the addition of appropriate taxa.
Orthology: Two genes that are members of the same gene family are said to be orthologs if they trace their most recent common ancestor to a speciation event.
Paralogy: Two genes that are members of the same gene family are said to be paralogs if they trace their most recent common ancestor to a gene duplication event.
Prokaryotes: Unicellular organisms without a membrane-bound nucleus (see Box 3).
Quartet-based tree construction: It is often desirable to break up an evolutionary analysis into a series of smaller analyses. The smallest nontrivial phylogenetic tree is composed of four sequences. The analysis of all or a large number of the possible quartets that can be made from a data set can provide useful information for evaluating congruent phylogenetic signals, for making phylogenetic trees and for making phylogenetic supertrees.
SSU rRNA: The RNA component of the small subunit of the ribosome.
Supertree: A supertree is a phylogenetic tree that is generated by amalgamating several phylogenetic trees into a single tree. If the leaf sets on the input trees are not identical, then the resulting tree is a supertree. If the leaf sets are identical on all the input trees, the result is a type of consensus tree.

Corresponding author: McInerney, J.O. (james.o.mcinerney@nuim.ie).

0169-5347/$ – see front matter doi:10.1016/j.tree.2008.01.008 Available online 25 March 2008

Five years later, Lederberg and Tatum [12] presented evidence for "the existence of a sexual stage" in *Escherichia coli*, showing for the first time that genetic recombination existed in prokaryotes. A decade later, Stanier and colleagues wrote about van Niel that he had completely revised his position and "has expressed the opinion that it is a waste of time to attempt a natural classification for bacteria" [13]. This skepticism pervaded the microbiology community for almost another 20 years, during which bacteriology had to proceed without an evolutionary framework.

The rise of the small-subunit rRNA-based tree of life

The advent of molecular phylogenetics heralded another sea-change in the perceived usefulness of a prokaryotic phylogeny. In the 1960s, Zuckerkandl and Pauling [14] defined the new research area of molecular evolution. Within a decade, Woese and colleagues [2,3], using indirect methods of oligonucleotide cataloguing from small-subunit rRNA (SSU rRNA), identified one particularly important split within the prokaryotes: that separating the Archaebacteria from the Eubacteria. The identification of this split [2,3] meant that there might be a third major category of living things on the planet, as divergent from Eubacteria and Eukaryota as each is from the other. Confirmation of the monophyly of these three potential groups seemed to be found in 1989, when the SSU rRNA tree was rooted using anciently duplicated genes [15,16].

These early studies showed that there was substantial phylogenetic structure within the prokaryotes, and prompted microbiologists to look at their diversity from a new evolutionary point of view. From then on, phylogenetic studies of prokaryotic evolution took off. By 1980, SSU rRNA had been characterised from around 170 prokaryotic species, giving a broad outline of the major eubacterial and archaebacterial groups [17]. By 1987, when Carl Woese published his authoritative 50-page review "Bacterial Evolution" summarising his earlier work [18], around 500 Eubacteria had already been characterised by oligonucleotide cataloguing, and complete SSU rRNA sequences were coming onstream. This led to the view of the SSU rRNA as "the ultimate molecular chronometer" [18]. Optimism concerning the existence of a single unifying tree of life was at its peak; the SSU rRNA tree was crowned as the "universal tree of life," and was used to devise a new (three-domain-based) phylogenetic classification of life [19]. The legacy of this work is that today there are more than 400 000 SSU rRNA sequences in the public sequence repositories, the vast majority of which have been sequenced for phylogenetic purposes.

Genomics and the rise of the network of life hypothesis

During the 1990s, gene sequences started to accumulate at an ever-increasing rate, and by 1995 the first complete eubacterial genome was publicly available [20]. As more genetic markers became available, the universal tree of life was put to the test ever more frequently. Interestingly, phylogenies inferred from alternative markers were often found to be incongruent with the topology of the SSU rRNA tree [21]. Some of these incongruent phylogenies were based on genes whose biological function was as essential as that of the SSU rRNA (e.g. the RNA polymerase gene [22]). Accordingly, it became less clear why the SSU rRNA should have been preferred over other markers to derive the universal tree of life [23]. The impression that the SSU rRNA tree might be inadequate was further exacerbated by the realization that trees derived from markers that initially seemed to support the SSU rRNA tree (e.g. ATP synthase and the elongation factors) were unreliable either because they were affected by horizontal gene transfer (HGT [24]) or because of phylogenetic artifacts (e.g. long-branch attraction [25,26]). Finally, the presumed superiority of the SSU rRNA gene as a phylogenetic marker was revealed to be untenable when it was discovered that the placement of many groups (e.g. the Microsporidia [22] – a group of fungi) in the SSU rRNA tree was the result of systematic errors (e.g. long-branch attraction).

These findings caused a shift in thinking. Among the first to state that at least for some genes and organisms there was a network rather than a tree of life were Hilario and Gogarten [24] and Martin and colleagues [27]. This viewpoint became increasingly popular as the first comparative genomic studies showed that the genomes of organisms that were closely related in the universal tree of life had significantly different gene content. In particular, Lawrence and Ochman claimed that the genome of *E. coli* acquired 18% of its genes via HGT after its divergence from its closest relative (*Salmonella enterica*) [28]. Because *E. coli* and *S. enterica* diverged ~100 million years ago, this might have meant that the vertical phylogenetic signal in the *E. coli* genome could be completely erased every 500 million years, assuming the rate of HGT per gene was homogeneous through time. Even vertebrates have a fossil history that is longer [29], whereas cellular life is likely to have existed on Earth for more than 3.8 billion years [30]. However, rates of HGT are gene specific, and vertically inherited genes might persist in bacterial genomes for more (or less) than 500 million years. On the other hand, even if a gene can persist in a genome for more than 500 million years, it seems unlikely that it might persist in a genome for billions of years. In fact, it has recently been shown that every gene family can be transferred into *E. coli* [31], and that (at the least in the laboratory) barriers to HGT are very low. We do not know whether this result will hold for prokaryotes in general. Nonetheless, it implies that at most prokaryotes can swap their genes with *E. coli* very easily, and that a gene that could be considered a universal chronometer that has kept track of the clonal history of cellular life since its origin is thus unlikely to exist [32]. Indeed, if the HGT rates observed in *E. coli* were found to apply to all prokaryotes, their evolutionary history would be essentially horizontal. That is, prokaryotes would effectively share a common gene pool and, from an evolutionary point of view, they would behave in a way that is not unlike that of the populations of a single species.

The new uncertainty at the end of the last century was different from that of half a century before. Prior to the molecular era, the absence of data caused uncertainty. With the arrival of complete genomes, uncertainty was not anymore caused by lack of data. Instead, now the very existence of the universal tree of life was on trial.

The sound of ideologies clashing

Toward the end of the 1990s, as more and more prokaryotic genomes were being sequenced, it became customary to report the proportion of their genes that were of foreign origin. Methods used to identify horizontally transferred genes were based on the identification of compositional biases (either codon usage biases or nucleotide composition biases; e.g. Ref. [28]) and BLAST score analysis (reviewed in Ref. [33]). These methods (see Box 1) have their limitations, and their results were relatively inaccurate, and often incongruent [21]. However, they consistently showed that HGT was anything but rare, and in some cases foreign genes seemed to have been transferred between organisms that were very distantly related (according to the SSU rRNA tree). For example, it was reported that 24% of the genome of the bacterium *Thermatoga maritima* [34] was acquired from archaebacterial donors, whereas 30% of the genes in the genome of the Archaebacteria *Methanosarcina mazei* and *M. acetivorans* were of eubacterial origin [35,36]. Obviously, many researchers remained skeptical of HGT [37–39], and in particular Kurland went as far as stating that "global LGT" (i.e. the view that HGT was a major evolutionary force; LGT = HGT) was "an ideology that is begging for deconstruction" [38]. Nevertheless, the view that HGT was a "rampant" phenomenon gained credibility, and was eventually formalised by Doolittle, who concluded "the history of life cannot properly be represented as a tree" [23]. According to Doolittle, HGT was such a powerful force that the evolutionary history of the prokaryotes was better represented using a network in which edges represent HGTs (see Figure 1).

The last word was still far from said. Studies attempting to estimate global rates of HGT could not agree (see above), and methods to identify these events (see Box 1) were questioned [40–43]. This cast doubt on the real frequency of HGT, and on the network of life hypothesis [38,39]. Yet, when the genomic sequences of three ecologically distinct *E. coli* strains were compared, it was shown that they

Box 1. Identification of HGT

A variety of approaches have been developed to identify genes that have been laterally transferred. These include BLAST-based approaches and methods that compare base composition or codon usage biases of different genes in a set of genomes. Discovering HGT is best carried out by identifying significant disagreement between phylogenetic trees inferred from different genes. This approach assumes that there exists an underlying species phylogeny with which alternative gene trees can be compared, and with which they should be congruent in the absence of HGT, paralogy and systematic biases (like those caused by long-branch attraction and amino acid composition bias). Any phylogenetic tree is a statistical inference from the sequence data and, like any inference, is subject to error. Careful analysis is thus needed to identify significant incongruence which cannot be explained as being a result of estimation error or paralogy, and so is potentially indicative of HGT. The tree-based approach to identifying HGT has the disadvantage that one has to assume a reference tree with which the individual gene trees are compared. Ideally, this reference tree should be the organismal phylogeny of the considered taxa. However, as we pointed out (see main text), this tree might not exist. This is not necessarily a problem, as any nonrandom tree, for example, a genomic tree (see Box 2), which might not represent a species tree (see text), might be used as a reference to infer HGTs. In fact, because in the absence of HGT every tree derived from a set of orthologs should have the same topology, even a single gene tree, such as the SSU rRNA [68], can be used as a reference against which the congruence of other gene trees might be compared to infer potential HGT events. However, if the source of incongruence between a gene tree and a reference tree is phylogenetic inaccuracy rather than HGT, tree-based approaches will overestimate HGT.

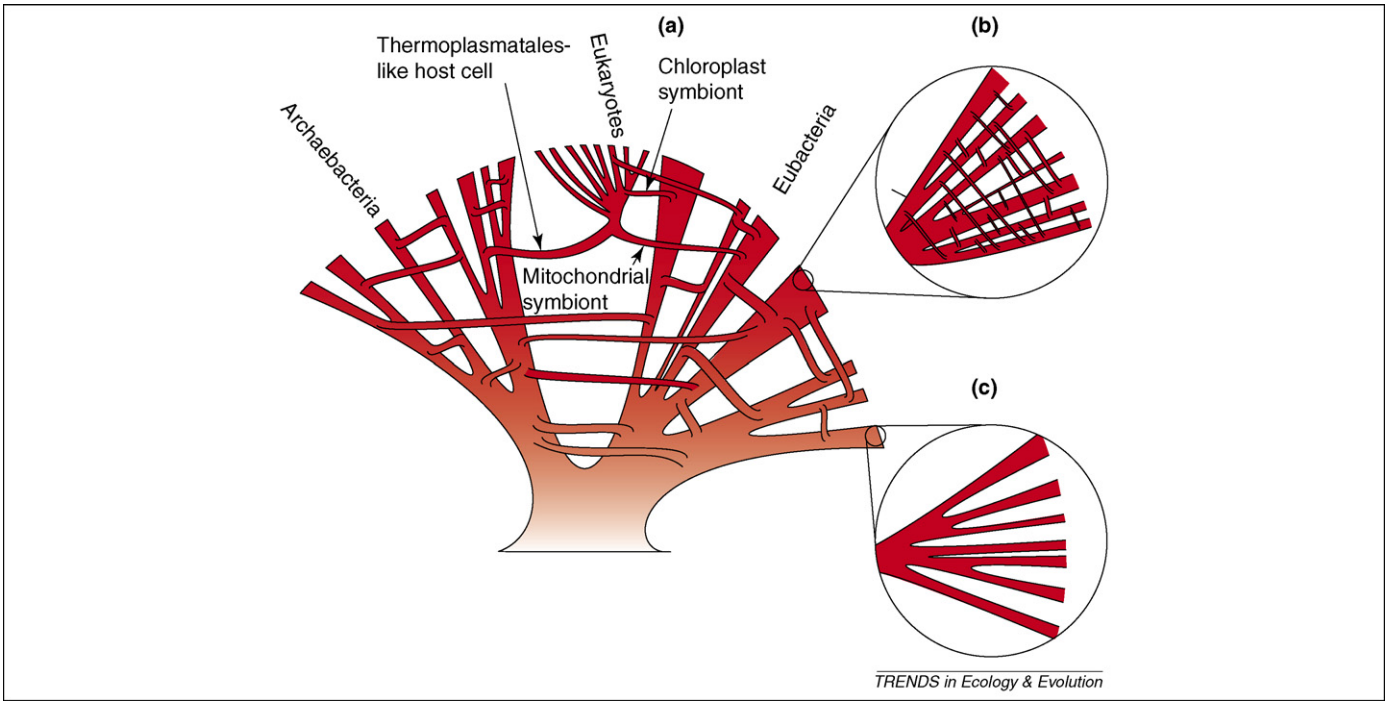

Figure 1. A network of life. **(a)** Eukaryotes are known to be chimeric, with chloroplast and mitochondrial genes having a different origin from nuclear genes, and originating from different groups of Eubacteria. There is more debate over whether lateral gene transfer between Eubacteria, Archaebacteria and Eukaryota and between major groups of prokaryotes is common or rare. **(b)** Lateral gene transfer is known to be so common between members of some groups of bacteria that they are effectively panmictic, whereas **(c)** it can be completely absent from other groups.

shared only 40% of their genes [44]. This patchy distribution of genes could be explained by assuming that the genome of the common ancestor of the three *E. coli* strains included all the genes found in the genomes of its descendants. The partially overlapping distribution of genes would then be the result of lineage-specific gene deletions. However, this explanation would entail that the last common ancestor of these strains had an unrealistically large genome. A more reasonable explanation would be that most of the genes that are not shared among the three strains (~60% of the total) have been independently acquired by HGT. Similarly, the analysis of the genomes of five Cyanobacteria showed that also the chromosomes of these organisms were littered with horizontally transferred genes [45].

Evidence accumulating during the 1990s and the earliest years of the new millennium seemed to suggest that there was no tree of life, but counterarguments were also being proposed. Phylogenetic model misspecification and misidentification of orthologs are confounding influences that might cause the overestimation of HGT (see Box 1). Additionally, even if HGT were rampant, it seemed that not all genes were equally likely to be transferred [46,47]. This observation was formalised as the complexity hypothesis by Lake and coworkers [46], who suggested that operational genes (those involved in the day-to-day processes of cell maintenance) are more likely to be transferred than informational genes (those involved in DNA replication, transcription and translation), and that a core set of nontransferable or rarely transferable genes might exist.

In August 2003, Daubin and his collaborators [48], using a variety of interspecies and intraspecies data sets and a quartet-based method, attempted to test whether trees inferred from protein-coding genes significantly disagreed with the SSU rRNA tree. Using this approach, they concluded that "orthologs available for phylogenetic reconstruction are compatible with a single tree." The same month, Kurland and his coworkers [39] published a trenchant criticism of the notion that interspecies gene transfer had obliterated the signal of vertical descent. This criticism focused on the methods that were being used to infer HGTs, which did not seem particularly robust (see above and Box 1). Additionally, they also suggested that although HGT might result in a gene finding its way into a genome, this did not necessarily mean that it would stay for an appreciable length of time; its presence might only be transient.

Kurland and his coworkers forcefully claimed that a prokaryotic tree existed even in the presence of HGT, and provided some sensible arguments to support their claim [39]. However, even if a universal tree of life exists, it does not mean that this tree is also recoverable. We have used supertree approaches (see Box 2) to investigate congruence across large sets of ortholog trees. These studies concluded that for relatively recent groups of prokaryotes such as the γ- and the α-Proteobacteria, congruence is high. Conversely, at the deepest levels of prokaryotic history, congruence fades significantly [7,9]. Even if a prokaryotic tree exists [37–39,49,50], its most basal nodes might not be recoverable because of phylogenetic signal erosion [51], HGT, ortholog misidentification or a combination of all three.

Box 2. Inferring phylogenies from multigene data sets

Three main approaches can be used to infer phylogenies from large collections of genes

The gene concatenation approach

The sequences of different genes are concatenated into a single data set, which is then analysed using one of the available phylogenetic methods (e.g. maximum likelihood, parsimony, Bayesian analysis and so on; see e.g. Refs [8,69]). The obvious advantage of this approach is that it can reduce (potentially even eliminate [69]) stochastic error. It also allows the combination of weak phylogenetic subsignals in different genes [51], which might result in the discovery of new clades that do not obtain significant support from single gene analyses. This approach ignores HGT, as it assumes that all loci have the same history. Gene concatenation is the method generally used when recovering core gene-based phylogenies [8].

The supertree approach

In the supertree approach, individual gene trees are inferred using standard phylogenetic methods. These trees are then combined using one of the available supertree methods to derive a consensus phylogeny; see Ref. [9] for an example. The advantage of the supertree approach is that the gene sequences are not combined before the phylogenetic inference. This can avoid the combination of genes with incompatible phylogenetic histories (i.e. genes vexed by HGT). A second advantage of the supertree approach is that it can be used to derive phylogenies based on extremely large numbers of genes (on the order of thousands [9]).

Gene content-based methods

These approaches use the presence or absence of specific genomic features (e.g. protein families), rather than sequence alignments, as discrete characters to be used in phylogenetic analyses; see Ref. [6] for an example. Matrices construed using such characters are analysed using standard phylogenetic methods.

Future prospects for the tree of life

In 1998, Woese proposed his genetic annealing model for the earliest stages of prokaryotic life [52]. In this model, HGT was initially the dominant mode of evolution within single-celled communities. Later on, HGT became less frequent and vertical inheritance became dominant (as in the species seen today). However, more recently, Goldenfeld and Woese [53] stated that HGT seems to have been so pervasive that it must be one of the most significant parts of any discussion about species or phylogenies. The minimal assumption for the existence of a prokaryotic tree is the existence of a core set of nontransferable (i.e. vertically inherited) genes, but evidence suggests that such a core set of genes cannot exist. Informational genes, including rRNA, can and have been horizontally transferred [21,54–58]. Some of these genes have been extensively swapped between domains. Probably the most extreme case is that of the aminoacyl tRNA synthetases [55], as none of the genes in this family agree with the SSU rRNA tree. Additionally, Miller and colleagues found a hybrid SSU rRNA gene in a chlorophyll d-producing cyanobacterium [59]. Finally, in *E. coli* [31], barriers to HGT are very low, and no single gene family is completely untransferable (by laboratory approaches) into this species [32].

As currently defined, the core represents less than 1% of the average prokaryotic genome [60] and its existence is

difficult to defend [32]: as more genomes are being sequenced, more genes are found to have been horizontally transferred and are thus removed from the core [57]. Indeed, 44 putative core genes have been identified for Eubacteria [61] and 45 putative core genes have been identified for the Archaebacteria [62], but only 31 core genes could be identified when Archaebacteria, Eubacteria and Eukaryota were considered [8].

Ernst Mayr defined species as "groups of interbreeding natural populations that are reproductively isolated from other such groups" [63]. Mayr was a zoologist and his biological species concept is one that essentially fits animals. Microbiologists might surely modify Mayr's definition to accommodate some level of HGT. However, when HGT seems to be able to move virtually every gene in the genome [32], the meaning of a tree derived from a tiny minority of putatively vertically inherited core genes become fuzzy [58]. This caused much disagreement on what is the real meaning of the core gene-based trees [49,50,60,64–66] (see Box 2). Indeed, in agreement with others [58,60,66], we think that the congruence of these core genes is unlikely to have any special significance. A core gene tree could only recapitulate clonal history if the core genes truly were vertically inherited. Unfortunately, the only approach to identifying vertically inherited genes is the congruence of gene trees with a species phylogeny. These putatively vertically inherited genes are mostly coding for proteins that are part of, or tightly associated with, the ribosomal machinery (e.g. ribosomal proteins [8]). Phylogenetic congruence among these genes might therefore reflect coevolution, not clonal history. In any case (see also Ref. [66]), the phylogeny derived from these genes cannot represent a species tree, as prokaryotic lineages are not reproductively isolated in any meaningful sense and thus are not species (*sensu* Mayr [63]).

Trees inferred from complete genomes [6,9], are also unlikely to represent species phylogenies. However, we conjecture that these trees might be more useful than those based on core genes because they represent some sort of central tendency and can be used as general frameworks in the study of prokaryotic evolution (e.g. Ref. [67]), irrespective of the existence of a prokaryotic tree of life. Indeed, the same genomic tree can either be considered a representation of the prokaryotic tree of life, if such a tree can be shown to exist, or a phenetic dendrogram representing the proximity (under a given distance measure) of the prokaryotic lineages (if the prokaryotic tree of life does not exist). By contrast, a core gene-based tree will loose its predictive power in the absence of a tree of life.

Eukaryotes are outside the scope of this review. However, it is important to point out that the current genomic evidence suggests that eukaryotes emerged from the 'fusion' of an archaebacterium and a eubacterium [6,9]. If these results are confirmed then there is no tree of life, but a ring-like network of life [32]. In the next 10 years, as an ever-increasing number of genomes will become available, we might see either the rebirth of the tree of life hypothesis or its ultimate downfall. It is still too early to say in which direction the evidence will swing. Nonetheless, it is not overly presumptive to state that 70 years after Breed (1939) made the suggestion, it is possible that we might have to revert to a realistic approach to prokaryotic classifications, eventually divorcing ourselves from the idea that a natural classification of the prokaryotes can be defined.

Box 3. Should we be using the term Prokaryota?

Some authors assign evolutionary significance to the group Prokaryota, that is, they assume prokaryotes to be monophyletic (e.g. Refs [25,49]). However, in the traditionally rooted [15,16] SSU-based universal tree of life, the prokaryotes do not form a monophyletic group: Archaebacteria are recovered as the sister group of the Eukaryota, instead of grouping with Eubacteria. Because the traditionally rooted SSU rRNA tree does not support the monophyly of the Prokaryota, Pace [50] recently started to campaign for the dismissal of the word ''prokaryotes.'' To corroborate his view, this author claimed that no positive definition has ever been given for this group, that is, they have always been clustered because they lack some feature (e.g. a membrane-bound nucleus), rather than because they share a defining phenotypic character. As a reaction to Pace [50], a positive definition of the term ''prokaryotes'' was provided by Martin and Koonin [70], who pointed out that the prokaryotes can be defined as the organisms that co-transcriptionally translate their main chromosomes. Because of the current uncertainties about the tree of life, in this article we have not assumed the monophyly of any prokaryotic group. However, we think that abandoning the word prokaryotes is unnecessary, as it unambiguously identifies an organizational grade: that of the single-celled organism without a membrane-bound nucleus, all of whom perform co-translational transcription of their chromosomes.

Acknowledgements

The authors would like to thank Robert Beiko and four anonymous reviewers for their comments and suggestions. This work was partially supported by a Science Foundation Ireland Research Frontiers Programme grant to J.O.McI.

References

1 Haeckel, E. (1866) *Generelle Morphologie der Organismen: Allgemeine Grundzüge der Organischen Formen-Wissenschaft, Mechanisch Begründet durch die von Charles Darwin Reformirte Descendenz–Theorie,* Georg Reimer
2 Fox, G.E. *et al.* (1977) Comparative cataloging of 16S ribosomal ribonucleic acid: molecular approach to procaryotic systematics. *Int. J. Syst. Bacteriol.* 27, 44–57
3 Woese, C.R. and Fox, G.E. (1977) Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc. Natl. Acad. Sci. U. S. A.* 74, 5088–5090
4 Gupta, R.S. (1998) Protein phylogenies and signature sequences: a reappraisal of evolutionary relationships among archaebacteria, eubacteria, and eukaryotes. *Microbiol. Mol. Biol. Rev.* 62, 1435–1491
5 Daubin, V. *et al.* (2001) Bacterial molecular phylogeny using supertree approach. *Genome Inform.* 12, 155–164
6 Rivera, M.C. and Lake, J.A. (2004) The ring of life provides evidence for a genome fusion origin of eukaryotes. *Nature* 431, 152–155
7 Creevey, C.J. *et al.* (2004) Does a tree-like phylogeny only exist at the tips in the prokaryotes? Proc. *Biol. Sci.* 271, 2551–2558
8 Ciccarelli, F.D. *et al.* (2006) Toward automatic reconstruction of a highly resolved tree of life. *Science* 311, 1283–1287
9 Pisani, D. *et al.* (2007) Supertrees disentangle the chimerical origin of eukaryotic genomes. *Mol. Biol. Evol.* 24, 1752–1760
10 Breed, R. (ed.) (1939) *Bergey's Manual of Determinative Bacteriology*, Williams and Wilkins
11 Stanier, R.Y. and van Niel, C.B. (1941) The main outlines of bacterial classification. *J. Bacteriol.* 42, 437–466
12 Lederberg, J. and Tatum, E.L. (1946) Gene recombination in Escherichia coli. *Nature* 158, 558
13 Stanier, R.Y. *et al.* (1957) *The Microbial World,* Prentice-Hall
14 Zuckerkandl, E. and Pauling, L. (1965) Molecules as documents of evolutionary history. *J. Theor. Biol.* 8, 357–366

15 Gogarten, J.P. *et al.* (1989) Evolution of the vacuolar H^+-ATPase: implications for the origin of eukaryotes. *Proc. Natl. Acad. Sci. U. S. A.* 86, 6661–6665
16 Iwabe, N. *et al.* (1989) Evolutionary relationship of archaebacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes. *Proc. Natl. Acad. Sci. U. S. A.* 86, 9355–9359
17 Fox, G.E. *et al.* (1980) The phylogeny of prokaryotes. *Science* 209, 457–463
18 Woese, C.R. (1987) Bacterial evolution. *Microbiol. Rev.* 51, 221–271
19 Woese, C.R. *et al.* (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Natl. Acad. Sci. U. S. A.* 87, 4576–4579
20 Fleischmann, R.D. *et al.* (1995) Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. *Science* 269, 496–512
21 Doolittle, W.F. *et al.* (2003) How big is the iceberg of which organellar genes in nuclear genomes are but the tip? *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 358, 39–57
22 Hirt, R.P. *et al.* (1999) Microsporidia are related to fungi: evidence from the largest subunit of RNA polymerase II and other proteins. *Proc. Natl. Acad. Sci. U. S. A.* 96, 580–585
23 Doolittle, W.F. (1999) Phylogenetic classification and the universal tree. *Science* 284, 2124–2129
24 Hilario, E. and Gogarten, J.P. (1993) Horizontal transfer of ATPase genes – the tree of life becomes a net of life. *Biosystems* 31, 111–119
25 Philippe, H. and Forterre, P. (1999) The rooting of the universal tree of life is not reliable. *J. Mol. Evol.* 49, 509–523
26 Brinkmann, H. and Philippe, H. (1999) Archaea sister group of bacteria? Indications from tree reconstruction artifacts in ancient phylogenies. *Mol. Biol. Evol.* 16, 817–825
27 Martin, W. *et al.* (1993) Evidence for a chimeric nature of nuclear genomes: eubacterial origin of eukaryotic glyceraldehyde-3-phosphate dehydrogenase genes. *Proc. Natl. Acad. Sci. U. S. A.* 90, 8692–8696
28 Lawrence, J.G. and Ochman, H. (1998) Molecular archaeology of the Escherichia coli genome. *Proc. Natl. Acad. Sci. U. S. A.* 95, 9413–9417
29 Shu, D-G. *et al.* (1999) Lower Cambrian vertebrates from south China. *Nature* 402, 42–46
30 Knoll, A.H. (2003) *Life on a Young Planet: The First Three Billion Years of Evolution on Earth,* Princeton University Press
31 Sorek, R. *et al.* (2007) Genome-wide experimental determination of barriers to horizontal gene transfer. *Science* 318, 1449–1452
32 McInerney, J.O. and Pisani, D. (2007) Genetics. Paradigm for life. *Science* 318, 1390–1391
33 Koonin, E.V. *et al.* (2001) Horizontal gene transfer in prokaryotes: quantification and classification. *Annu. Rev. Microbiol.* 55, 709–742
34 Nelson, K.E. *et al.* (1999) Evidence for lateral gene transfer between archaea and bacteria from genome sequence of Thermotoga maritima. *Nature* 399, 323–329
35 Deppenmeier, U. *et al.* (2002) The genome of Methanosarcina mazei: evidence for lateral gene transfer between bacteria and archaea. *J. Mol. Microbiol. Biotechnol.* 4, 453–461
36 Galagan, J.E. *et al.* (2002) The genome of M. acetivorans reveals extensive metabolic and physiological diversity. *Genome Res.* 12, 532–542
37 Pace, N.R. (1997) A molecular view of microbial diversity and the biosphere. *Science* 276, 734–740
38 Kurland, C.G. (2000) Something for everyone. Horizontal gene transfer in evolution. *EMBO Rep.* 1, 92–95
39 Kurland, C.G. *et al.* (2003) Horizontal gene transfer: a critical view. *Proc. Natl. Acad. Sci. U. S. A.* 100, 9658–9662
40 Eisen, J.A. (2000) Horizontal gene transfer among microbial genomes: new insights from complete genome analysis. *Curr. Opin. Genet. Dev.* 10, 606–611
41 Ragan, M.A. (2001) Detection of lateral gene transfer among microbial genomes. *Curr. Opin. Genet. Dev.* 11, 620–626
42 Ragan, M.A. (2001) On surrogate methods for detecting lateral gene transfer. *FEMS Microbiol. Lett.* 201, 187–191
43 Koski, L.B. *et al.* (2001) Codon bias and base composition are poor indicators of horizontally transferred genes. *Mol. Biol. Evol.* 18, 404–412
44 Welch, R.A. *et al.* (2002) Extensive mosaic structure revealed by the complete genome sequence of uropathogenic Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.* 99, 17020–17024
45 Raymond, J. *et al.* (2002) Whole-genome analysis of photosynthetic prokaryotes. *Science* 298, 1616–1620
46 Jain, R. *et al.* (1999) Horizontal gene transfer among genomes: the complexity hypothesis. *Proc. Natl. Acad. Sci. U. S. A.* 96, 3801–3806
47 Rivera, M.C. *et al.* (1998) Genomic evidence for two functionally distinct gene classes. *Proc. Natl. Acad. Sci. U. S. A.* 95, 6239–6244
48 Daubin, V. *et al.* (2003) Phylogenetics and the cohesion of bacterial genomes. *Science* 301, 829–832
49 Kurland, C.G. *et al.* (2006) Genomics and the irreducible nature of eukaryote cells. *Science* 312, 1011–1014
50 Pace, N.R. (2006) Time for a change. *Nature* 441, 289
51 Pisani, D. and Wilkinson, M. (2002) Matrix representation with parsimony, taxonomic congruence, and total evidence. *Syst. Biol.* 51, 151–155
52 Woese, C. (1998) The universal ancestor. *Proc. Natl. Acad. Sci. U. S. A.* 95, 6854–6859
53 Goldenfeld, N. and Woese, C. (2007) Biology's next revolution. *Nature* 445, 369
54 Ke, D. *et al.* (2000) Evidence for horizontal gene transfer in evolution of elongation factor Tu in enterococci. *J. Bacteriol.* 182, 6913–6920
55 Woese, C.R. *et al.* (2000) Aminoacyl-tRNA synthetases, the genetic code, and the evolutionary process. *Microbiol. Mol. Biol. Rev.* 64, 202–236
56 Parker, M.A. (2001) Case of localized recombination in 23S rRNA genes from divergent bradyrhizobium lineages associated with neotropical legumes. *Appl. Environ. Microbiol.* 67, 2076–2082
57 Charlebois, R.L. and Doolittle, W.F. (2004) Computing prokaryotic gene ubiquity: rescuing the core from extinction. *Genome Res.* 14, 2469–2477
58 Gogarten, J.P. *et al.* (2002) Prokaryotic evolution in light of gene transfer. *Mol. Biol. Evol.* 19, 2226–2238
59 Miller, S.R. *et al.* (2005) Discovery of a free-living chlorophyll d-producing cyanobacterium with a hybrid proteobacterial/cyanobacterial small-subunit rRNA gene. *Proc. Natl. Acad. Sci. U. S. A.* 102, 850–855
60 Dagan, T. and Martin, W. (2006) The tree of one percent. *Genome Biol.* 7, 118
61 Brochier, C. *et al.* (2002) Eubacterial phylogeny based on translational apparatus proteins. *Trends Genet.* 18, 1–5
62 Matte-Tailliez, O. *et al.* (2002) Archaeal phylogeny based on ribosomal proteins. *Mol. Biol. Evol.* 19, 631–639
63 Mayr, E. (1942) *Systematics and the Origin of Species,* Columbia University Press
64 Lake, J.A. (2007) Disappearing act. *Nature* 446, 983
65 Martin, W. *et al.* (2007) The evolution of eukaryotes. *Science* 316, 542–543
66 Doolittle, W.F. and Bapteste, E. (2007) Pattern pluralism and the Tree of Life hypothesis. *Proc. Natl. Acad. Sci. U. S. A.* 104, 2043–2049
67 Beiko, R.G. *et al.* (2005) Highways of gene sharing in prokaryotes. *Proc. Natl. Acad. Sci. U. S. A.* 102, 14332–14337
68 Dagan, T. and Martin, W. (2007) Ancestral genome sizes specify the minimum rate of lateral gene transfer during prokaryote evolution. *Proc. Natl. Acad. Sci. U. S. A.* 104, 870–875
69 Rokas, A. *et al.* (2003) Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature* 425, 798–804
70 Martin, W. and Koonin, E.V. (2006) A positive definition of prokaryotes. *Nature* 442, 868