4 Evolutionary History of Prokaryotes: Tree or No Tree?

J. O. McInerney and D. E. Pisani
Department of Biology, National University of Ireland Maynooth,
Maynooth, Co. Kildare, Ireland

M. J. O'Connell
Department of Biochemistry, University College Cork, Cork, Ireland

D. A. Fitzpatrick
Conway Institute, University College Dublin, Dublin, Ireland

C. J. Creevey
European Molecular Biology Laboratory, EMBL Heidelberg,
Heidelberg, Germany

CONTENTS

ABSTRACT

Prokaryotes are likely to be the most numerous and species rich organisms on the planet[1], occupying a more diverse set of ecological niches than eukaryotes. Knowledge of prokaryote diversity is severely limited by our inability to recreate the conditions in the laboratory that are needed to cultivate the majority. Discrepancies between direct microscopical counts and the numbers of colony-forming units can be as much as 100-fold, leading to speculation concerning how much we really know about prokaryotes. In contrast, genomic studies of prokaryotes are advanced. So, while on one hand we know that we have a poor overview of prokaryotic life on the planet, we have, paradoxically, succeeded in obtaining more completed genomic sequences of prokaryotes than of

eukaryotes. Therefore, even though taxon sampling has been restricted, we have now reached the stage where we can evaluate whether there is a meaningful prokaryotic phylogenetic tree or taxonomy. Questions remain as to whether the history of prokaryotic life has been overwritten by continuous and random interspecies gene transfer and occasional genome fusions, or whether these events have only been minor contributors, thereby enabling prokaryotic evolutionary history to be adequately described by a tree.

4.1 A BRIEF HISTORY OF PROKARYOTIC SYSTEMATICS

Haeckel formalised the concept of using a phylogenetic tree in order to depict the relationships between all the life forms on the planet (see below). The metaphor of the tree seemed to work quite well, and indeed, in Charles Darwin's magnum opus of 1859[2], the only diagram that was used was one depicting a phylogenetic tree. For botanists and zoologists the concept of a phylogenetic tree with large trunks giving rise to smaller branches and then to leaves had so many attractive properties that its position as a central metaphor is almost unshakeable. For microbiologists, however, phylogenetic trees of the prokaryotes have always been problematic; even the definition of prokaryotic and eukaryotic taxa was not satisfactorily resolved until the 1960s[3].

The ranges of morphological characters that have been the subject of analysis in animal and plant groups simply do not exist in the prokaryotes. Cell morphology in many prokaryotes can be described using adjectives as simple as 'rod shaped' or 'round'; nothing approaching the rich lexicon that can be used by botanical or zoological systematists to describe their study taxa. The description of the prokaryotes (called Monera, at the time) given by Haeckel is perhaps the most colourful. He described them as: "… not composed of any organs at all, but consist entirely of shapeless, simple homogeneous matter … nothing more than a shapeless, mobile, little lump of mucus or slime, consisting of albuminous combination of carbon"[4].

Stanier and van Niel finally settled on a definition of the prokaryotes that included three traits that they lacked: absence of true nuclei, absence of sexual reproduction and absence of plastids[5]. This lack of morphological diversity resulted in a situation where microbiologists settled for classification systems that were taxonomically based, rather than phylogenetically based. Naturally, this led to the downgrading of microbial phylogenetics, and with students of microbiology being presented with nothing more than lists of species names, prokaryotic systematics proceeded at a very much slower pace than was seen in plants and animals. The definitive authority on prokaryotic species, *Bergey's Manual of Determinative Bacteriology,* was published first in 1923 and made no attempt at presenting the prokaryotes in a hierarchical manner based on common ancestry, and indeed, the most recent version still does not[6]. While this approach is changing, with *Bergey's Manual of Systematic Bacteriology*[7] presenting the prokaryotes in a phylogenetic context, the absence of a phylogenetic paradigm in earlier editions of Bergey's manual was reflective of the prevailing attitude that the natural history of the prokaryotes was not knowable at that time and perhaps even that it was not important.

While the phylogenetic relationships between the prokaryotes did not receive much attention in the early part of the last century, it was becoming increasingly clear that metabolic diversity in the prokaryotes was extensive[8]. Prokaryotes could live at a wider range of temperatures than eukaryotes, could live on a very diverse range of diets and produced an almost endless range of secondary metabolites. The source of this metabolic diversity was obviously the result of differences in the genetic composition of the organisms. However, there was no reason at that stage to suggest that microorganisms varied enormously in their genomic composition; after all, they all needed to replicate, carry out transcription and translation and other housekeeping functions. Perhaps small numbers of genes were responsible for this huge amount of metabolic variation?

4.2 THE RIBOSOMAL RNA REVOLUTION

In a seminal paper in 1965, Zuckerkandl and Pauling compared the degree of divergence between α-globin proteins of various animals and the separation times of these animals as judged by the fossil record[9]. The result was a generally linear increase in protein divergence with time. The implications were that cellular macromolecules could be used to make inferences concerning historical events, and if this was so, then these molecules could potentially be used to infer phylogenetic relationships, and ultimately, the tree of life might be inferred using these data.

By the early 1970s, manipulation of the macromolecules of the cell became more tractable, and this led Woese and coworkers to the development of classification systems based on ribosomal RNA oligonucleotide cataloguing[10]. Within a few years, enough information was available for the first really big change in our views concerning prokaryotic evolutionary relationships. This change in perception centred on the discovery that prokaryotes could be divided into two groups, with neither group being particularly closely related to each other and certainly no more closely related to each other than either was to eukaryotes[11]. Suddenly, a complete revolution took place. The pace of change in molecular biology facilitated some of this revolution. Rapid DNA sequencing technologies would develop over the following fifteen years[12,13], resulting in the sequencing of tens of thousands of ribosomal RNA molecules from prokaryotes and eukaryotes. In part, the renewed interest was driven by one of the most comprehensive and incisive manuscripts to have ever been written on the subject of bacterial evolution[14], and in part, it seemed that microbiologists were making up for lost time. By the late 1980s, ribosomal RNA phylogenetic trees became the gold standard for inferring evolutionary relationships across all levels.

Quite ironically, Darwin had cautioned that "The importance, for classification, of trifling characters, mainly depends on their being correlated with several other characters of more or less importance. The value indeed of an aggregate of characters is very evident in natural history"[2]. Even though Darwin knew nothing of genes and the cellular macromolecules, he was clear that classification systems should be based on a broad spectrum of traits whose functions were also diverse. Woese[14], to his credit, accepted that it was possible to see differences between the phylogenies that were inferred using small subunit ribosomal RNA molecules and the phylogenies that were being inferred using Cytochrome C genes. Quite likely, these differences were due to interspecies gene transfer, a form of prokaryotic sex, first described by Lederberg and Tatum in 1946[15]. Woese also noted that the phylogenetic trees were otherwise almost identical and concluded that it was "safe to assume" that there was a unique prokaryotic evolutionary history and that some of the cellular macromolecules would have recorded this history[14].

Molecular biology continued to advance, and with the arrival of automated sequencing methods[16] the first genome sequence of a prokaryote, that of *Haemophilus influenzae,* became available[17]. The genome was followed soon afterwards by the genome sequence of an archaeon, *Methanosarcina janaschii*[18], and the genome sequence of the smallest known autonomously replicating organism, *Mycoplasma genitalium*[19]. The genome sequence of *Escherichia coli* K12 was a relatively late arrival[20], given that it was the first organism for which a genome sequencing effort had started. However, when three 'strains' of this species were sequenced completely, the full extent of the nature of gene transfer in prokaryotes was seen[20–22]. These three genomes have no more than 39% of their genes in common and vary in sequence length by almost one million base pairs. Clearly, if any pair of plants or animals differed in genome content by more than 20%, they would not be considered to be the same species; however, in prokaryotes, the standard taxonomic tools had grouped these organisms together as a species. The underlying cause of this genome content difference appears to be the independent acquisition of large numbers of genes in the process known by varying terms including lateral gene transfer (LGT), horizontal gene transfer, or simply gene transfer. This presents us with a problem for inferring phylogenetic relationships. If this pattern is replicated throughout the prokaryotic world, then perhaps the inference of phylogenies based on genomic data may not be possible.

4.3 CONFLICTING TREES

An examination of many phylogenetic trees derived from single-copy genes reveals that there is a considerable degree of similarity across these trees. Consider the situation in Figure 4.1. These two trees are constructed from a ribonuclease gene (left) and a DNA polymerase gene family (right). There is a large degree of similarity between these two trees. The differences are to be seen in the absence of an ortholog for the ribonuclease gene in *Xylella fastidiosa* and *Vibrio cholera,* an alternative resolution of the branching order within the *Escherichia coli* strains and a resolution of the branching order within the *Neisseria meningitidis* strains. Overall, these two trees generally suggest highly congruent, but not identical, evolutionary histories. These minor differences are probably attributable to errors in phylogeny reconstruction and to lineage specific gene loss.

By 1998, the first large-scale comparative genome analyses were being carried out, and one of the first findings was that 755 of the identified 4,288 open reading frames in the *E. coli* genome (547.8 Kb) were introduced by LGT in at least 234 lateral transfer events since its divergence from *Salmonella* approximately 100 million years before present[23]. If this was true and these LGT events were stable, then it was relatively easy to conclude that LGT was indeed a major feature, perhaps the most important feature, of prokaryotic evolution. The implication also was that ribosomal RNA phylogenetic trees were no more than gene trees and did not reflect organismal phylogeny.

In 1999, Doolittle, writing in *Science,* made the statement that “If ‘chimerism’ or ‘lateral gene transfer’ cannot be dismissed as trivial in extent or limited to special categories of genes, then no hierarchical universal classification can be taken as natural”[24]. The reason for making a statement such as this, which was a radical departure from the questions that were being asked at the time (such as “What is the shape of the universal tree and how should we try to infer this shape?”), had to do with what genomic data was beginning to tell us. Increasingly, ortholog-derived trees were being produced that were not in agreement with the ribosomal RNA tree and were not in agreement with one another. In another paper at the time, Doolittle suggested that it was more appropriate to visualise the evolutionary history of life on the planet as a web[25]. This would reflect the central role of LGT in life’s evolution and would be more accurate. This caused controversy and was seen in some quarters as an effort to hark back to the dark days when it was accepted that a prokaryotic phylogeny was unknowable.

Defending the phylogenetic tree concept (tree thinking), Kurland and coworkers[26] refuted the suggestion that LGT was the “essence of phylogeny”. They pointed out the difficulties of incorporating a new gene into a genome, particularly when there may be an incumbent gene that is performing a similar or identical function. Their conclusion was that stable integration of a new gene into a genome is at such a low rate that it has little or no influence on the idea of a core phylogenetic tree uniting all organisms.

Woese had already put forward the ‘genetic annealing model’ of organismal evolution[27]. In this model, Woese suggested that, prior to organismal diversification, the planet was populated by ‘progenotes’, and gene transfer between these progenotes was high. Subsequently, gene transfer became more difficult, and currently there are high barriers to LGT. Interestingly, Woese stated that “By now, it is obvious that what we have come to call the universal phylogenetic tree is no conventional organismal tree. Its primary branchings reflect the common history of central components of the ribosome, components of the translation apparatus, and a few other genes. But that is all. In its deep branches, the tree is merely a gene tree”[27]. Subsequently, Woese extended his hypothesis, stating explicitly that the very importance of LGT is in part evidenced by the universality of the genetic code; if it was not universal, then LGT would not be possible[28]. However, Woese, in sticking with his doctrine of espousing the view that vertical inheritance is the most important mode of organismal evolution, defined the ‘Darwinian Threshold’ as the critical point that is reached when vertical inheritance becomes more important than horizontal transfer. According to Ge and coworkers, the evolutionary history of life is somewhat like a great tree with occasional cobwebs joining branches[29]. They estimate the extent of LGT to be 2% per genome.

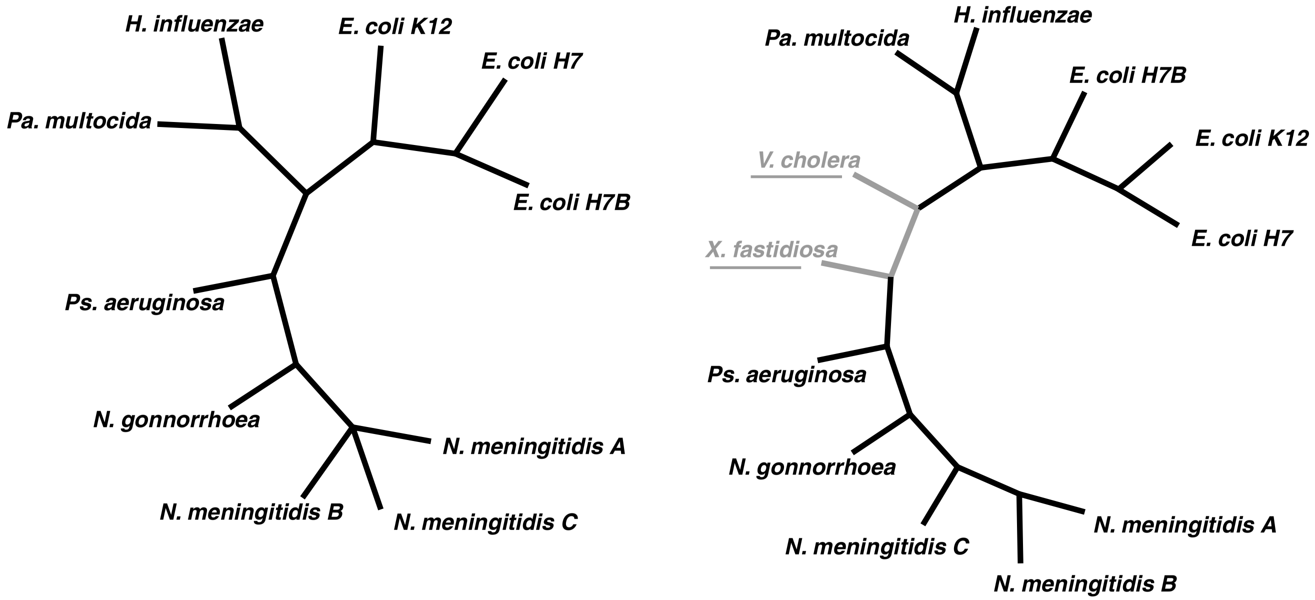

FIGURE 4.1 Phylogenetic trees of *Escherichia, Haemophilus, Neisseria, Pseudomonas* (Ps), *Pasteurella* (Pa), *Vibrio* and *Xylella.* On the left is a phylogenetic tree derived using orthologs from the ribonuclease family. On the right is a phylogenetic tree derived using orthologs of DNA polymerase III. The completed genomes of these species were searched for orthologs, and all available orthologs were used. For the DNA polymerase III family, there was no ortholog present in the genomes of *V. cholera* and *X. fastidiosa.*

Therefore, the early part of this century has resulted in the formation of two camps, one that emphasises evolution by vertical inheritance and focuses on the identification of 'core' genomic components that tend to be inherited together, using this information to define prokaryotic relationships (tree thinkers), and the other that emphasises LGT and attempts to accommodate it (net thinkers or web thinkers).

4.4 METHODOLOGICAL DEVELOPMENTS

Initially, the post genome era debate was being fought on philosophical grounds with ad hoc invocation of analyses of small amounts of data; the finding of the extraordinary level of plasticity in the *E. coli* genome being the most highlighted case to argue in favour of the pan transfer advocates[20–23].

At the moment, the ground on which the debate is being fought is becoming technical, and arguments are being made on the basis of sophisticated methods of analysis and large amounts of data. In general, the sequences of large numbers of presumed orthologous genes are being collected, trees constructed, and phylogenetic hypotheses based on these trees postulated. The LGT debate centres on the analysis of these trees, and usually, though not always, the analysis method of choice involves some kind of consensus or supertree approach. We will briefly review supertree methods, before describing the kinds of findings that these methods are producing.

The use of supertree methods in phylogenetics can trace its origin back to Gordon's classic paper[30], although the earliest supertree algorithm actually predates Gordon's work[31]. Supertree approaches seek to amalgamate the information contained in a set of phylogenetic trees (that is, dendrograms or cladograms), the only requirement being that they overlap in a specific way. Whilst there is no requirement for any given tree to contain the entire set of leaves, there is a requirement that the combined trees can be linked to one another through common subsets of their leaves. Supertrees cannot be constructed from sets of input trees with disjoint leaf sets. The output from a supertree analysis is a supertree, that is, a tree summarising, according to a defined set of rules, the information contained in the input trees. Different supertree methods are based on different sets of rules. From this point of view, supertrees are no different from standard consensus methods, and supertrees can be considered generalisations of consensus tree methods[32]. However, in contrast to consensus methods, supertrees combine partially overlapping trees. This can provide an inference based on the information contained in the input trees and may result in clades being present in the supertree that do not appear in the input set. In any case, the relationships that are present in the supertree but not in the input trees must be implied by some of the trees in the input set and should never be contradicted by all input trees[33,34].

Many supertree methods exist, and a classification of these methods is now difficult, in part because there are such a diverse range of methods and in part because, in some cases, a method could be said to belong to more than one kind of approach. Broadly speaking, two categories can be distinguished, that of the strict/semistrict supertree methods, and that of the liberal supertree methods. Strict and semistrict supertree methods do not allow conflict among the input trees to be resolved, while liberal supertree methods allow for conflict resolution[35]. Strict and semistrict methods are generally not used in practical studies because they tend to return artificially highly unresolved supertrees. The most frequently used supertree methods are the liberal ones, and amongst these the most common involve the generation of matrices that are representations of the input treesy[36]. Alternative matrix representation-based methods are characterised by the way the trees are recoded (for example, as sets of splits or quartets) and by the optimality criterion used to analyse these matrices, such as parsimony[36] compatibility[37], or the minimum number of flips (state changes) necessary to eliminate all the incompatibilities from the matrix representation of a set of trees (Min Flip supertrees)[38]. In any case, it is important to note that all supertree methods can be defined in terms of the tree-to-tree distance they use as an optimality criterion; for example tree length in the case of Matrix Representation using Parsimony (MRP), the Robinson-Foulds distance[39] in the case

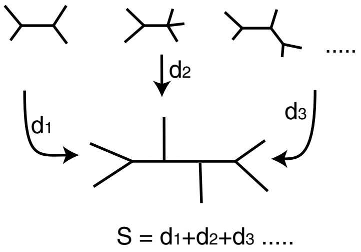

FIGURE 4.2 Outline of the procedure for evaluating a supertree using **DFIT** or **SFIT** measures as implemented in CLANN. For each input tree, its similarity to an appropriately pruned supertree is measured. The overall score for the supertree is either the sum or average distance computed for all input trees. The difference between the **DFIT** and the **SFIT** measures is to be found in the way in which the distance is computed. S = supertree score, d = distance between the input tree and the appropriately pruned supertree.

of Split Fit, or a flip distance in the case of Min Flip[40]. An alternative is to use a path length distance-based approach to infer the optimal supertree[41]. Approaches using path length distances include the Distance Fit (DFIT) method[42], and the Average Consensus[43], the latter having the potential advantage that it can use branch length information if available.

All of these methods (with the exclusion of Min Flip) are implemented in the program CLANN[44], which also implements a fast Neighbour Joining Average Consensus (NJAC) procedure, and Quartet Fit (QFIT). For the DFIT approach, a supertree can be proposed for the dataset; this supertree can be randomly generated, or an initial rapid supertree construction method such as NJAC can be used to provide a starting tree. The proposed supertree is compared with any input tree, even when the input tree only contains a subset of the total complement of leaves. This can be achieved by pruning the supertree appropriately. Once the pruned supertree and the input tree have the same leaf set, a simple comparison can be made to evaluate their similarity (see Figure 4.2). The DFIT approach involves the calculation of a path length distance from every taxon to the others. The distance is simply the number of nodes that separates the taxa on the tree. If the pruned supertree and the input tree are identical, then the distance matrix that is derived from the pruned supertree and the distance matrix derived from the input tree will also be identical. If the two are different, then the distance matrices will be different, and with increasing dissimilarity in tree shape, there will be increasing dissimilarity in the distances derived from the trees. The supertree that is chosen is therefore the one that is most similar to the input trees.

Other methods like QFIT and Split Fit (SFIT), although originally thought as matrix representation based methods[34,41], can be similarly derived. QFIT involves breaking up the pruned supertree and the input trees into the quartets they entail. Naturally, the two collections of quartets will be identical in terms of leaf content. Again, if both the pruned supertree and the input tree have identical topologies, their quartets will be identical. However, increasing dissimilarity in tree shape will result in fewer quartets with identical topologies. Therefore for QFIT, the score of any given supertree will be proportional to the number of quartets that it contains that have identical topologies to those found in the input trees. SFIT involves breaking up the pruned supertree and the input trees into the splits they entail. SFIT can then be seen as comparing an appropriately pruned supertree with each input tree. The measure of similarity in this case will be the Robinson-Foulds distance[39], and the best supertree will be the one minimising the distance between it and the input trees.

The first large-scale supertree that was constructed for prokaryotes was constructed by Daubin and coworkers[45]. The dataset included a total of 33 prokaryotes and four eukaryotes. They indicated that they could produce a robust supertree when they used ortholog trees with a broader taxon sampling, that is when they avoided using gene trees with small numbers of leaves, and they also indicated that this genome phylogenetic tree was very much in agreement with the ribosomal RNA trees. Subsequently, this work was followed up with an analysis of differences between these ortholog trees, using a multivariate analysis method to identify a core of gene trees with similar topologies and then using these gene trees in order to construct a MRP supertree. For many of the groups on this supertree there is strong support (support being assessed using the bootstrap method); however, the spine of the tree appeared to only have low to medium levels of support.

4.5 AN EMERGING CONSENSUS?

Recently Creevey et al.[42] carried out an analysis of completed genome information in a supertree context. The question that was being addressed was whether or not it was possible to identify a robust phylogenetic tree among the deepest branches of the prokaryotic domains. This was to be compared and contrasted with an analysis of a similarly sized dataset that spanned a relatively well-characterised but less ancient group of prokaryotes, the γ-proteobacteria. The analysis involved using the single-gene families from completed genomes, inferring the phylogenetic relationships between these gene sequences and only retaining the inferences that, according to the most widely used methods of analysis, were the most robust. These phylogenetic hypotheses were combined using the MSSA (DFIT) supertree approach implemented in CLANN[44], and the input trees were compared with the supertree to evaluate the goodness of fit of the data to the tree. The results were interesting. While the gene trees derived from the γ-proteobacteria were in good agreement with each other and were in good agreement with the supertree, the trees that spanned the deepest branches of the prokaryotes strongly conflicted with each other. This was not a simple case of lacking signal; it was a case that the gene trees were strongly supported but conflicting. In fact, a comparison of the congruence across these trees and congruence across randomised sets of trees, using the YAPTP test[42], showed that the trees derived from the data were no more in agreement with one another than trees that were completely random. The conclusion from this work was that the prokaryotic phylogeny inferred is strongly supported in parts and not so in other parts. The γ-proteobacteria were also examined using an entirely different approach[46], but the conclusions were the same: congruence across different gene trees is excellent. Lerat et al.[46] recorded that for 205 gene trees examined, there was concordance across 203. In another study using the same methods, we examined the relationships within the α-proteobacteria and found, once again, that there was good agreement between the individual gene trees[47]. However, these analyses have only concentrated on gene families where there are no apparent paralogs. Obviously, if duplicated genes were also taken into consideration, there would be much more data to examine. In addition, taxonomic sampling is sparse and it remains to be seen if the conclusions still hold when sampling is improved.

In a recent report by Beiko and coworkers[48], ‘highways’ of gene sharing between prokaryotic groups were identified. Their analysis centered on using edit distances to transform ortholog derived trees into a topology that is consistent with a supertree. The finding was that vertical inheritance of genes was in the majority, but the patterns of LGT could not be ignored, and that LGT mostly took place between closely related organisms (presumably using homologous recombination as a means of integrating new genetic material) or between distantly related organisms that occupied the same environment (presumably using illegitimate recombination as the means of integration). The frequency with which each category of genes was transferred was not uniform, with genes involved in ‘metabolism’, and ‘cellular processes’ being significantly more frequently involved in a LGT event.

This leads to the question of whether there is or there is not a meaningful prokaryotic phylogenetic tree. If there is, then the paradigm of a tree still stands; if there is not, then the paradigm falls and we need to revert to the descriptive taxonomy of yesteryear, and any evolutionary indications would refer to some subset of the organism’s genes, but not the organism. There is clearly an emerging lack of consensus. We can easily find instances where congruence is excellent, and we can find instances where congruence is impossibly poor.

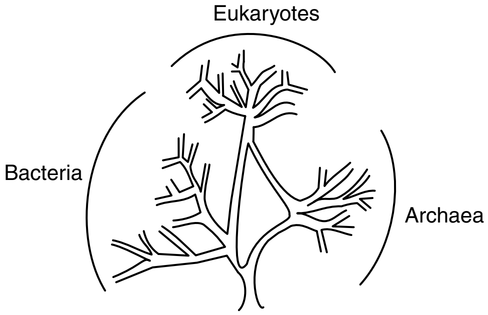

FIGURE 4.3 A stylised outline of how the evolutionary history of cellular life could be represented using the ring of life theory.

4.6 THE PROKARYOTIC INFLUENCE ON THE EUKARYOTE

The ribosomal RNA tree of all cellular life is a metaphor that is very widely recognised. The three main divisions of life, Bacteria, Archaea and Eucarya, are widely recognised, and even though there are disagreements about the importance of this classification system, there is general agreement that these three life forms are very different from one another. The ribosomal RNA tree also has an important inference, that the first microorganisms on the planet were prokaryotes, but eukaryotes evolved from this prokaryote world. The ribosomal RNA tree suggests that there was some kind of discrete event that led to the development of the nucleus early in eukaryote evolution. The ribosomal RNA tree also led to the conclusion that mitochondria were α-proteobacteria-like and had evolved via some kind of symbiosis[14].

Challenges to this dogma have been in circulation for some time, but recently, the first evidence has been produced for a discrete event that suddenly resulted in the development of the eukaryotic cell. Using a new method of genome analysis, conditioned reconstruction, Lake and coworkers have suggested that the eukaryotic cell was created as a result of a fusion of the genome of a bacterium and an archaeum[49–51]. They then suggest that there is no tree of life; if anything, there is a ring of life (see Figure 4.3 for an illustration of what is inferred). If this analysis proves correct, then the consequences for prokaryotic systematics are profound. This would mean that neither the bacteria nor the archaea are monophyletic and both would have the eukaryotic lineage as one of their descendents. This could also mean that the development of our ideas concerning prokaryotic evolution may be incorrect. If true, it also begs the question concerning whether or not there are other ‘rings’ of life.

4.7 CONCLUSIONS, FUTURE DIRECTIONS AND OPEN QUESTIONS

The consensus at the moment is that the prokaryote phylogeny is more tree like than random. There are clear instances of groups of prokaryotes where the agreement across their ortholog phylogenies is high[42,47]. Speciation in prokaryotes is not well understood; however, it is likely that inheritance patterns are generally divergent, and in that respect, the evolutionary history of the prokaryote cells are tree like. What is at question is whether there are groups of genes that make the inference of this history deviate from a tree like pattern. Various metaphors have been used to describe the evolutionary history of prokaryotes such as tree, web, ring or cobweb. However, it is clear that one single description is insufficient to describe the entire history of the group. Future work will centre on more precise descriptions of prokaryote genes, genome and cellular evolution.

ACKNOWLEDGEMENTS

This work was supported by a Science Foundation Ireland Research Frontiers Programme grant to James McInerney, and a Marie Curie Intra European Fellowship to Davide Pisani (contract number MEIF-CT-2005-01002). The authors would like to thank the two referees for their helpful advice.

REFERENCES

1. Whitman, W.B., Coleman, D.C., and Wiebe, W.J., Prokaryotes: the unseen majority, *Proc. Natl. Acad. Sci. USA,* 95, 6578, 1998.
2. Darwin, C., *On the Origin of Species by Means of Natural Selection,* John Murray, London, 1859.
3. Sapp, J., The prokaryote-eukaryote dichotomy: meanings and mythology, *Microbiol. Mol. Biol. Rev.,* 69, 292, 2005.
4. Haeckel, E., *The History of Creation,* Trench and Co., London, 1883.
5. Stanier, R.Y. and van Niel, C.B., The concept of a bacterium, *Arch. Mikrobiol.,* 42, 17, 1961.
6. Holt, J.G., *Bergey's Manual of Determinative Bacteriology,* Williams and Wilkins, Baltimore, 1994.
7. Garrit, G.M., Ed., *Bergey's Manual of Systematic Bacteriology*, Springer, New York, 2001.
8. Breed, R.S., Murray, E.G.D., and Hitchens, A.P., *Bergey's Manual of Determinative Bacteriology.* The Williams and Wilkins Company, 1948.
9. Zuckerkandl, E. and Pauling, L., Molecules as documents of evolutionary history, *J. Theor. Biol.,* 8, 357, 1965.
10. Sogin, S.J., Sogin, M.L., and Woese, C.R., Phylogenetic measurement in procaryotes by primary structural characterization, *J. Mol. Evol.,* 1, 173, 1971.
11. Woese, C.R. and Fox, G.E., Phylogenetic structure of the prokaryotic domain: the primary kingdoms, *Proc. Natl. Acad. Sci. USA,* 74, 5088, 1977.
12. Sanger, F., Nicklen, S., and Coulson, A.R., DNA sequencing with chain-terminating inhibitors, *Proc. Natl. Acad. Sci. USA,* 74, 5463, 1977.
13. Smith, L.M. et al., Fluorescence detection in automated DNA sequence analysis, *Nature,* 321, 674, 1986.
14. Woese, C.R., Bacterial evolution, *Microbiol. Rev.,* 51, 221, 1987.
15. Lederberg, J. and Tatum, E., Gene recombination in *Escherichia coli*, *Nature,* 158, 558, 1946.
16. Wilson, R.K. et al., Development of an automated procedure for fluorescent DNA sequencing, *Genomics,* 6, 626, 1990.
17. Fleischmann, R.D. et al., Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd, *Science,* 269, 496, 1995.
18. Bult, C.J. et al., Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii, Science,* 273, 1058, 1996.
19. Fraser, C.M. et al, The minimal gene complement of *Mycoplasma genitalium, Science,* 270, 397, 1995.
20. Blattner, F.R. et al., The complete genome sequence of *Escherichia coli* K-12, *Science,* 277, 1453, 1997.
21. Hayashi, T. et al., Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12, *DNA Res.,* 8, 11, 2001.
22. Welch, R.A. et al., Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli, Proc. Natl. Acad. Sci. USA,* 99, 17020, 2002.
23. Lawrence, J.G. and Ochman, H., Molecular archaeology of the Escherichia coli genome, *Proc. Natl. Acad. Sci. USA,* 95, 9413, 1998.
24. Doolittle, W.F., Phylogenetic classification and the universal tree, *Science,* 284, 2124, 1999.
25. Doolittle, W.F., Lateral genomics, *Trends Cell. Biol.,* 9, M5, 1999.
26. Kurland, C.G., Canback, B., and Berg, O.G., Horizontal gene transfer: a critical view, *Proc. Natl. Acad. Sci. USA,* 100, 9658, 2003.
27. Woese, C.R, The universal ancestor, *Proc. Natl. Acad. Sci. USA,* 95, 6854, 1998.
28. Woese, C.R., On the evolution of cells, *Proc. Natl. Acad. Sci. USA,* 99, 8742, 2002.
29. Ge, F., Wang, L.S., and Kim, J., The cobweb of life revealed by genome-scale estimates of horizontal gene transfer, *PLoS Biol.,* 3, e316, 2005.
30. Gordon, A.D., Consensus supertrees: the synthesis of rooted trees containing overlapping sets of labeled leaves, *J. Classif.,* 3, 31, 1986.

31. Aho, A.V. et al., Inferring a tree from lowest common ancestors with an application to the optimisation of relational expressions, *SIAM J. Comput.,* 10, 405, 1981.
32. Semple, C. and Steel, M., A supertree method for rooted trees., *Discrete Appl. Math.,* 105, 2000.
33. Pisani, D.E. and Wilkinson, M., Matrix representation with parsimony, taxonomic congruence and total evidence, *Syst. Biol.,* 51, 151, 2002.
34. Wilkinson, M. et al., Measuring support and finding unsupported relationships in supertrees, *Syst. Biol.,* 54, 823, 2005.
35. Wilkinson, M. et al., Some desiderata for meta-analytical supertrees, in *Phylogenetic Supertrees: Combining Information to Reveal the Tree of Life.,* Bininda-Emonda, O.R.P., Ed., Kluwer Academic, Dordrecht, 2004, 227.
36. Ragan, M.A., Phylogenetic inference based on matrix representation of trees, *Mol. Phylogenet. Evol.,* 1, 53, 1992.
37. Ross, H.A. and Rodrigo, A.G., An assessment of matrix representation with compatibility in supertree construction, in *Phylogenetic Supertrees: Combining Information to Reveal the Tree of Life,* Bininda-Emonda, O.R.P., Ed., Kluwer Academic, Dordrecht, 2004, 35.
38. Burleigh, J.G et al., MRF supertrees, in *Phylogenetic Supertrees: Combining Information to Reveal the Tree of Life,* Bininda-Emonds, O.R.P., Ed., Kluwer Academic, 2004, 65.
39. Robinson, D. and Foulds, L., Comparison of phylogenetic trees, *Math. Biosci.,* 53, 131, 1981.
40. Chen, D. et al., *Flipping: A Supertree Construction Method,* American Mathematical Society, Providence, Rhode Island, 2003, 135.
41. Steel, M. and Penny, D., Distributions of tree comparison metrics—some new results, *Syst. Biol.,* 42, 126, 1993.
42. Creevey, C.J. et al., Does a tree-like phylogeny only exist at the tips in the prokaryotes? *Proc. R. Soc. Lond. B. Biol. Sci.,* 271, 2551, 2004.
43. Lapointe, F.-J. and Cucumel, G., The average consensus procedure: combination of weighted trees containing identical or overlapping sets of taxa, *Syst. Biol.,* 46, 306, 1997.
44. Creevey, C.J. and McInerney, J.O., Clann: investigating phylogenetic information through supertree analyses, *Bioinformatics,* 21, 390, 2005.
45. Daubin, V., Gouy, M., and Perriere, G., Bacterial molecular phylogeny using supertree approach, *Genome Inform. Ser. Workshop Genome Inform.,* 12, 155, 2001.
46. Lerat, E., Daubin, V., and Moran, N.A., From gene trees to organismal phylogeny in prokaryotes: the case of the gamma-Proteobacteria, *PLoS Biol.,* 1, e19, 2003.
47. Fitzpatrick, D.A., Creevey, C.J., and McInerney, J.O., Genome phylogenies indicate a meaningful {alpha}-proteobacterial phylogeny and support a grouping of the mitochondria with the Rickettsiales, *Mol. Biol. Evol.,* 23, 74, 2006.
48. Beiko, R.G., Harlow, T.J., and Ragan, M.A., Highways of gene sharing in prokaryotes, *Proc. Natl. Acad. Sci. USA,* 102, 14332, 2005.
49. Rivera, M.C. and Lake, J.A., The ring of life provides evidence for a genome fusion origin of eukaryotes, *Nature,* 431, 152, 2004.
50. Lake, J.A. and Rivera, M.C., Deriving the genomic tree of life in the presence of horizontal gene transfer: conditioned reconstruction, *Mol. Biol. Evol.,* 21, 681, 2004.
51. McInerney, J.O. and Wilkinson, M., New methods ring changes for the tree of life, *Trends Ecol. Evol.,* 20, 105, 2005.