

Branching out

Robert L. Charlebois,
Robert G. Beiko and Mark A. Ragan

Seven years after the publication of the first microbial genome sequence — that of *Haemophilus influenzae* — the roster of microbial genomes has topped 100. Despite early fears that whole-genome sequencing might be economically justified only for human pathogens, this list represents a gratifyingly broad range of microbial phenotypes — soil bacteria and photosynthesizers, thermophiles and halophiles, animal and plant pathogens, and more. At least 12 prokaryotic phyla are represented, as are a few eukaryotes — enough to allow a meaningful examination of the Tree of (microbial) Life.

In the early days of molecular phylogenetics (the mid-1960s to the early 1990s), it was thought that sequencing was the path to enlightenment — more sequences of more genes could only improve the depth and resolution of our knowledge of life's history. But instead, our 100-genome world is riven by seemingly irreconcilable conflicts; ambiguities and discrepancies are the norm, rather than the exception. Some of modern biology's fundamental tenets — notably the darwinian–mendelian model of parent-to-offspring ('vertical') gene flow — have once again, at least for microbes, been thrown into doubt. Lateral (horizontal) gene flow — in which genes are transmitted across, rather than along, branches in family trees — is no longer an explanation of last resort, but a competitive model for the origin of microbial biodiversity.

Although seldom correct, highly polarized views often serve to delineate a problem. Vertical inheritance with tree-like speciation fails to explain why so many gene families are distributed as they are among microbial genomes — that is, in highly diverse, sparse patterns that often fail to support accepted taxonomy. Yet genome evolution based largely on lateral (recombination-dependent) events seems

impossible for prokaryotes that live solitary lives inside eukaryotic cells, for example, or more generally reproduce for generation after generation by simple binary fission.

As recently as the 1970s, authoritative textbooks stated that microbiology might never be put on a phylogenetic footing. Carl Woese and colleagues showed, however, how the history of organismal life could be reconstructed from sequences of small-subunit ribosomal RNA, part of the protein-synthesizing machinery. The first molecular phylogenetic trees were sparse and not always trustworthy. But the 1980s brought automated technology for sequencing DNA, the polymerase chain reaction (PCR), and much-improved methods for inferring phylogenies. Ribosomal RNA (rRNA) genes are ubiquitous, with highly conserved termini that make amplification by PCR easy. Tens of thousands of rRNA sequences became available. Not surprisingly, the rRNA tree quickly became the 'gold standard' for determining microbial relationships.

Nonetheless, rRNAs provide only a narrow window on the microbial genome: for every gene that encodes an rRNA, there may be 1,000 that encode a protein. Protein-coding genes are less universal, more difficult to amplify by PCR, and often shorter and less information-rich than rRNA genes. What's more, trees inferred from individual protein-coding genes (or from their proteins) often disagree irreconcilably with the rRNA tree. Why is there such discrepancy, and what does it tell us about microbial genomes? These questions spark profound disagreement in the microbial-phylogenetics community.

Some theorists — let's call them the verticalists — remind us of the (real or supposed) inadequacies of single-gene phylogenetics. For verticalists, protein-based trees disagree because their true phylogenetic signal is too often obscured by noise and bias. Only by overcoming these obstacles — through using better models, perhaps, or cleaner data — can we understand how microbial genomes have diversified and evolved.

But others — the lateralists — point to the sophistication and power of existing methods, and argue that trees disagree because genes really do have different histories. Microbial genomes are, to a lateralist, more or less ephemeral entities that are maintained, if only fleetingly, by the vagaries of selection and chance. The apparent woesian hierarchy of taxa is only an epiphenomenon of differential barriers — whether environmental, geographical or more intrinsically biological — to lateral gene flow.

We and others have been exploring 'whole-genome trees' as a means of overcoming the noise and bias of single-protein

Microbial phylogenomics

Has genomics overturned the family tree of microbial life? Thanks in part to often polarized debate, elements of a new synthesis are emerging.

analyses, to extract the bulk phylogenetic signals that are inherent in genomes. The input data for genome trees can be the proportions of genes or proteins that genomes hold in common, or (as we prefer) the mean pairwise similarities between shared proteins. Despite some early indications to the contrary, whole-genome trees have now largely converged on the rRNA-sequence tree.

For us — as, presumably, for the verticalists — this convergence means that lateral gene transfer has not undermined descent with modification as the default explanation for microbial biodiversity, nor (as recently suggested by Ford Doolittle) has it thrown microbial classification into disarray. Lateral transfer is not both quantitatively important and directional. One of the few widely accepted instances of lateral gene transfer — the origin of chloroplasts from relatives of cyanobacteria — is clearly visible in our whole-genome trees, and even more so in 'sub-genome trees' based on functional subsets of genomes.

The most enthusiastic lateralists reply, however, that convergence between whole-genome and rRNA trees merely demonstrates that rRNA genes — unlike most individual protein-coding genes, but like the genome as a whole — are but pastiches that are produced by lateral gene transfer.

Fascinating as these conflicts are, the important point is not whether a given tree is right or wrong. Rather, we should use these trees as frameworks upon which to construct and test hypotheses about the rate and mode of microbial evolution, and to improve our analytical methods. Without conflicts, we might all be far more complacent about evolutionary theory. In microbial phylogenomics, the scientific process is alive and well! ■

Robert L. Charlebois and Robert G. Beiko are in the Department of Biology, University of Ottawa, Ottawa, Ontario K1N 6N5, Canada. Mark A. Ragan is at the Institute for Molecular Bioscience, University of Queensland, Brisbane, Queensland 4072, Australia.

FURTHER READING

- Clarke, G. D. P. *et al.* *J. Bacteriol.* **184**, 2072–2080 (2002).
Doolittle, W. F. *Science* **284**, 2124–2128 (1999).
Gogarten, J. P., Doolittle, W. F. & Lawrence, J. G. *Mol. Biol. Evol.* **19**, 2226–2238 (2002).
Woese, C. R. *Proc. Natl Acad. Sci. USA* **99**, 8742–8747 (2002).
Wolf, Y. I. *et al.* *BMC Evol. Biol.* **1**, 8 (2001).



How close are we to a true impression of life's tree?