

EVOLUTION

Illuminating the first bacteria

A new analysis aims to uncover the root of the bacterial tree of life

By Laura A. Katz

The ability to sequence genes and more recently whole genomes has transformed our understanding of the tree of life by elucidating the tremendous diversity of microorganisms, and by placing plants, animals and fungi as branches nested among microbial lineages (1–3). The resulting evolutionary tree divides life into three ‘domains’: the exclusively microbial bacteria and archaea, and eukaryotes, organisms whose cells contain nuclei (including ciliates, amoebae, and animals). Yet the ordering of the earliest branching events on the tree and the nature of now-extinct ancestors remains unclear. On page XXX of this issue, Coleman *et al.* (4) provide a new estimate of the root of the bacterial tree of life, that is, the ancestor from which all bacterial species are derived. Knowledge of the root of the bacterial tree is important because it defines the evolutionary starting point for the tremendous diversity of bacteria and offers glimpses into the nature of the first bacterial cells.

Although genome sequencing has given biologists a wealth of data for estimating phylogenies (i.e., evolutionary trees), it has also revealed the importance of non-vertical inheritance as genes, and sometimes whole genomes, have been transferred across species boundaries (2, 3, 5). This lateral gene transfer (LGT; also called horizontal gene transfer) has substantially altered bacterial evolution, for example through the spread of antibiotic resistance and the acquisition of metabolic pathways (2). In some lineages, such as the bacterial clade Thermatogales, a large number of LGTs have generated a chimeric lineage: an estimated 20% of the genes in the *Thermotoga* genus have been acquired from archaea rather than by vertical transmission through binary division (5). Eukaryotes represent another chimera having arisen through the merger of an archaeon and a bacterium, the latter eventually becoming a mitochondrion (6, 7). Events such as these make estimation of the root of a tree challenging because transferred genes reflect different histories than vertically-inherited genes, and thus can confound inferences.

Coleman *et al.* use amalgamated likelihood estimation (ALE), a method developed by their group called, to estimate the position of the root of the bacterial tree of life while

attempting to account for gene duplications and losses as well as LGTs. To determine the root of the bacterial tree, Coleman *et al.* first construct an unrooted phylogeny based on 62 single-copy genes sampled from 265 bacterial genomes, and then use ALE to evaluate various positions of the root on this tree through analyses of the evolutionary history of 11,000 gene families sampled from across the same lineages.

Coleman *et al.* estimate that two-thirds of the transmissions among gene families analyzed are vertical, and that the root of the bacterial tree of life likely lies between “Gracilicutes”, predominantly Gram-negative bacteria including Proteobacteria, Acidobacteria and Spirochaeta, and “Terrabacteria”, including Gram positive bacteria, Cyanobacteria, Firmicutes, and “CPR”; taxonomic names in quotes indicate uncertainty (see the figure). Intriguingly, the placement of two clades – “DST” (Deinococcota, Synergistota and Thermotogota) and Fusobacteriota (an anaerobic clade of Gram-negative bacteria) – that are the recipients of a disproportionate number of LGTs, is unresolved. This is exemplary of the uncertainty introduced by LGT in reconstructing a bacterial tree of life.

Knowing the root of the bacterial tree of life allows both polarization of the evolution of bacterial traits and reconstruction of the last bacterial common ancestor (LBCA), at least for those aspects that have evolved through vertical transmission. Coleman *et al.* infer that LBCA was a free-living rod-shaped cell surrounded by a double-membrane and with a flagellum. The genome of this ancestor of all bacteria encoded key components of informational processing (e.g., DNA transcription, replication), metabolic pathways including CO₂ fixation and the Krebs cycle, as well as elements of the CRISPR-CAS9 adaptive immune system. Overall, these inferences agree with findings of others on the physiological and morphological features of LBCA (8, 9).

Did they get it right? It is difficult to know because reconstructing such ancient events [the earliest fossil bacteria are estimated to have existed around 3–3.4 billion years ago (10)] presents a challenge even in the absence of LGT. Results are dependent on the parameters used in the mathematical models, and although Coleman *et al.* extensively

evaluate these parameters through simulations, assumptions about the relative rates of gene duplication and LGT have been questioned (11). To further complicate inferences, Coleman *et al.* note the importance of taxon sampling because changing representative species within major bacterial clades affects the estimate of the position of the root. This suggests that future analyses including different, and perhaps newly discovered, lineages may alter both estimates of the root of the bacterial tree of life and inferences about LBCA.

Another confounding factor is the strength of the assumption that bacteria are monophyletic – that bacteria are descended from a common ancestor that existed more recently than their shared common ancestor with archaea. This belief requires either the independent evolution of bacteria and archaea at the origin of life (12) or a complex pattern of diversification and extinction events to give rise to reciprocally monophyletic bacteria and archaea (i.e., the evolution of two distinct lineages of microorganisms). Coleman *et al.* state that they are agnostic as to the monophyly of bacteria, pointing out the limitations of previous studies that rely on an archaeal outgroup for rooting the bacterial tree of life. Instead, their model excludes archaea and the nested eukaryotes (6, 7) in their estimates of the bacterial root. Alternative models suggest that archaea descended from LBCA (13, 14), in which case archaea would have to be included to determine the root of the bacterial tree of life.

Although some will contest the proposed position of the root of the bacterial tree of life, others challenge the concept of a tree as a model for the diversification of life on Earth (14, 15). Though Coleman *et al.* may have accurately captured the vertical portion of the bacterial tree of life, the tree that they root is missing the history of biological innovations and ecological adaptations that are derived from LGTs. To portray these lateral events, some authors have suggested that the tree of life be replaced with a circle, a web, or even a network depicting the complex history of inheritance (2, 3, 5). It has been suggested that “to save the trees, one might define organisms as more than the sums of their genes and imagine organismal lineages to have a sort of emergent reality” (2). In other words, there is room for alternative methods and

Department of Biological Sciences, Smith College, Northampton, MA, USA. Email: lkatz@smith.edu

innovation to encompass both vertical and lateral inheritance, focusing on the evolutionary history of phenotypic features such as metabolism, morphology, and life history.

REFERENCES AND NOTES

1. L. A. Hug *et al.*, *Nat Microbiol* **1**, 16048 (2016).
2. W. F. Doolittle, *Science* **284**, 2124 (1999).
3. T. Dagan, W. Martin, *Genome Biol.* **7**, 118 (2006).
4. G. A. Coleman *et al.*, *Science* **372**, XXX (2021).
5. J.P. Gogarten, J. P. Townsend, *Nat. Rev. Microbiol.* **3**, 679 (2005).
6. T.A. Williams, P. G. Foster, C. J. Cox, T. M. Embley, *Nature* **504**, 231 (2013).
7. A. Collens, L. A. Katz, *J. Heredity* **112**, 140 (2021).
8. W. F. Martin, *Front Microbiol* **11**, 817 (2020).
9. M. Miyata *et al.*, *Genes Cells* **25**, 6 (2020).
10. E. J. Javaux, *Nature* **572**, 451 (2019).
11. T. J. Treangen, E. P. C. Rocha, *PLOS Genet.* **7**, e1001284 (2011).
12. F. L. Sousa *et al.*, *Philo. Trans. R. Soc. B Biol. Sci.* **368**, 1 (2013).
13. T. Cavalier-Smith, *Int. J. Syst. Evol. Microbiol.* **52**, 7 (2002).
14. J. A. Lake, R. G. Skophammer, C. W. Herbold, J. A. Servin, *Philos. Trans. R. Soc. B Biol. Sci.* **364**, 2177 (2009).
15. S. Gribaldo, C. Brochier, *Res. Microbiol.* **160**, 513 (2009).

ACKNOWLEDGMENTS

I am grateful to A. Cote-L'Heureux and several other colleagues for thought-provoking conversations on this topic. L.A.K. is supported by grants from the National Science Foundation (OCE-1924570, DEB-1651908, DEB-1541511) and National Institute of Health (R15HG010409).

10.1126/science. abh2814

FIGURE AND LEGEND BELOW ON NEXT PAGE

Fig. Illuminating the root of the bacterial tree of life

Coleman et al (4) infer that the root of the bacterial tree of life likely lies between the “Gracilicutes” and the “Terrabacteria” (see text). Unresolved are two clades – “DST” (Deinococcota, Synergistota and Thermotogota), and Fusobacteriota – that are estimated to have a substantial number of LGT events across the gene families analyzed; only ~30-40% of gene transfers in these lineages are estimated to be vertical. Here, names in quotes represent provisional taxonomic terms. Left inset shows effect of lateral gene transfer across a three domain tree of life from (redrawn from (2)) and right insert represents an alternative in which both archaea and eukaryotes emerge from among bacterial lineages.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59