Trends in Microbiology

CellPress

Letter

Retracing lineage history: time to emphasize genetic turnover

Andrew K. Watson,[1] Philippe Lopez,[1] and Eric Bapteste[1,*]

Debates between proponents of a tree of life and that of a web of life [1] encouraged a substantial rethink about which events genuinely compose the history of microbial lineages. Admittedly, the reconstruction of patterns of vertical descent alone cannot recapitulate the entire evolutionary history of lineages, which share some gene families due to introgressive processes [2–4]. Thus, both vertical and lateral inheritances of genetic material are recognized to contribute to the history of a lineage. Moreover, recent works by Koonin *et al*. [4–6] highlight how, during evolution, genomes are significantly (re)-invented through processes of gene gains and losses, and how our knowledge of microbial evolution is shifting. Consistently, we build upon these works to propose a further, deeper rethink on what makes up a lineage's genuine history. Specifically, as a consequence of the shifts characterized by [4], we argue that the genesis of phyla should explain more than the origins of shared genes: it should also explain the origins of lineage-specific gene families, at all levels of the biological classification, and include all historical genetic events of descent and modification that shape lineages.

Strikingly, a glaring proportion of gene families is simply not shared between most cellular lineages. This observation holds true at the microevolutionary scale, for strains and species with open pangenomes [4,7]. Consistently, the evolutionary history of these lineages is investigated using a diversity of complementary approaches [4]. However, the prevalence of nonshared gene families is even more true at broader evolutionary scales, that is, for lineages such as major microbial groups (Figure 1A). Their history is still typically analyzed using approaches that first aim at reconstructing ancient phylogenies, and therefore focus on shared genetic material, hence consider only a part of what, according to us, constitutes the bona fide history of these lineages. Consequently, further developments in phylogenomics should retrace ancient history more comprehensively.

Taking constant genomic (re)-invention into account will expand our views on the true history of microbial lineages and, at the same time, it may further unravel some limits of deep historical inferences in the face of gene turnover. Therefore, shifting towards a more inclusive definition of lineage history will impact studies of all evolving entities, from viruses to multicellular eukaryotes, and bring forward an emerging question relevant at all taxonomic scales: how and why is Life massively reinventing its genes?

To tackle this issue, novel practices are necessary to represent genomic (re)-invention, to track its possible rules, its timing of occurrence and its extent, to enhance historical accounts of lineage history in ways that explicitly show the genetic turnover by which nonshared genetic material becomes pervasive within lineages at all taxonomic levels. This ambitious research program has started in microevolutionary studies, but even there two future research avenues deserve specific mention. First, tools are lacking to analyze to what extent lineages are either reinventing their functions or rather exploring new functional spaces when they reinvent their genes (Figure 1B). Such tools are, however, critical to determine (i) whether gene turnover preserves the established functions within a lineage or (ii) whether gene turnover supports the integration of novel functions in a lineage. In the first case, a situation typical of the Ship of Theseus [8,9] problem occurs, where, although the constitutive material, for example, the planks forming the ship or the genes forming the genomes, are replaced by new ones, the form/function of the ship persists; in the latter case, gene turnover metaphorically produces new kinds of ships: lineage-specific genes are used to explore the evolutionary landscape in novel directions. One wants to know whether pangenomes and major microbial groups contain evidence of Ships of Theseus in the making.

Second, systematic analyses of which molecular interactions are fueling the genetic turnover in lineages are yet to be conducted, both at the microevolutionary and macroevolutionary levels. To fully comprehend lineage history, it becomes even more of a priority to explain the origins of genes without homologs outside a given lineage. This apparent lack of homologs encourages the deployment of models of gene origination that are more complicated than the divergence from a single last common ancestor. Typically, the extent of gene remodeling, involving unique combinations of genetic material already present within the lineage and of gene remodeling/gains and losses, resulting from collisions with genetic material external to the lineage [10,11], in particular from mobile elements [4–6], needs to be systematically quantified across lineages. Beyond standard homology detection approaches, this issue requires the development of scalable network-based methods to detect genes and genetic fragments with multiple phylogenetic origins, and their rules of associations in genomes. Critically, such methods could also be systematically coupled with population genomic approaches to analyze whether lineage-specific genes accumulate neutrally or under positive selection [4]. In the

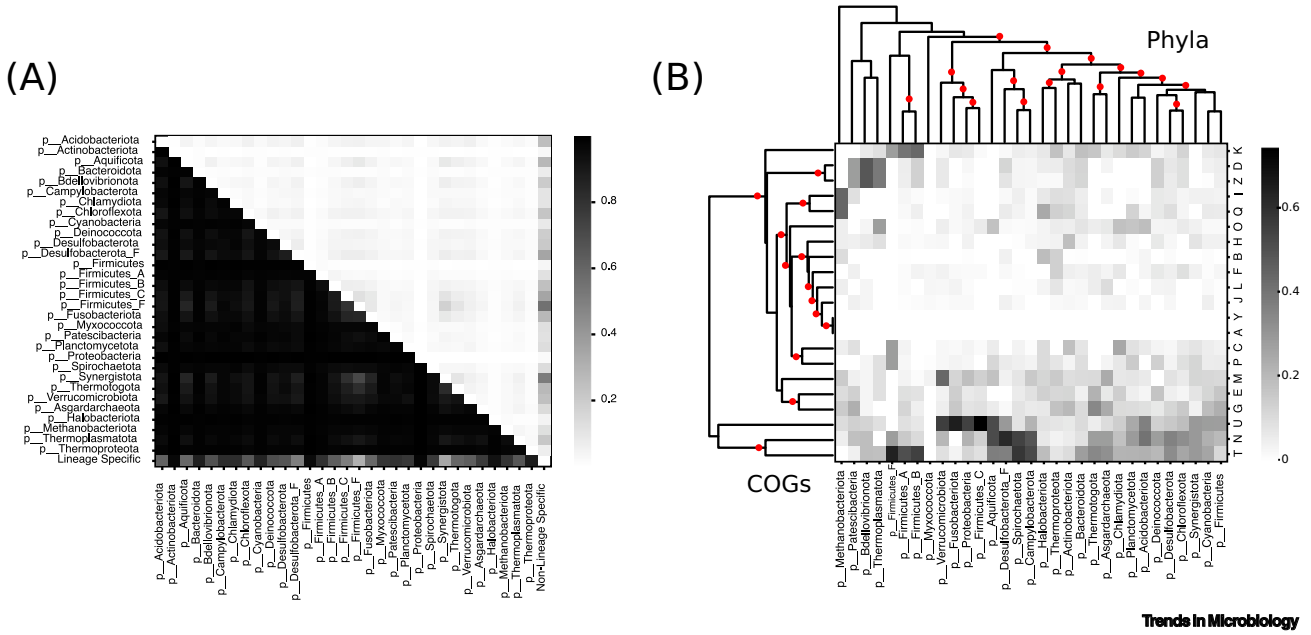

Figure 1. Distribution and functional profiles of lineage-specific genes across major microbial phyla. (A) This heatmap summarizes statistics relative to gene family distributions across major phyla of the Bacteria and Archaea, defined in the Genome Taxonomy Database (GTDB) (https://rdcu.be/b3OI7), computed from a diamond blastp all-against-all search, followed by clustering into gene families (connected components at the 30% identity and 80% mutual coverage thresholds) for 6078 complete genomes (5601 Bacteria and 477 Archaea), available at the NCBI at the time of the analysis. Major groups are organized by Domains along rows and columns. Cells are color-coded by % of gene families associated with the pangenome of each group. The bottom right triangle, devoted to 'non-gene-sharing metrics', reports the % of nonshared gene families between each pair of lineages. The final row indicates the % of lineage-specific gene families for each group. The upper left triangle, devoted to 'gene-sharing metrics', reports the % of shared gene families between each pair of lineages. The final column indicates the % of shared gene families for each group. (B) This clustered heatmap summarizes the functions of lineage-specific gene families for which eggnog emapper-1.0.3 (DB version 4.5.1) functional predictions were available [12], thus showing whether and which groups use their exclusive genes to reinvent similar functions. The order of Cluster of Orthologous Groups (COG) categories (rows) and GTDB phyla (columns) is dictated by hierarchical clustering (Euclidean distance, average method). Red dots on dendrogram branches indicate an approximately unbiased test support of >0.9, as calculated by pvclust.

latter case, a Red Queen-type of process, promoting the emergence of lineage-specific gene contents and the active erosion of the inherited genetic material would account for the genetic divergence between lineages, beyond the accumulation of substitutions within shared gene families.

Ultimately, rethinking the concept of lineage history to put more emphasis on processes leading to genetic turnover, which result in (abundant) lineage-specific genes, could lead both microevolutionary and macroevolutionary studies in underexplored directions. Unraveling prevalent signals for a Red Queen process and for a flotilla of Ships of Theseus in the Web of Life would shift our views on microbial evolution further away from its initial (phylogenetic) simplicity!

[1]Institut de Systématique, Evolution, Biodiversité (ISYEB), Sorbonne Université, CNRS, Museum National d'Histoire Naturelle, EPHE, Université des Antilles, 7, quai Saint Bernard, 75005 Paris, France

*Correspondence:
eric.bapteste@mnhn.fr (E. Bapteste).

https://doi.org/10.1016/j.tim.2021.08.001

References

1. Blais, C. and Archibald, J.M. (2021) The past, present and future of the tree of life. *Curr. Biol.* 31, R314–R321
2. Dagan, T. *et al.* (2008) Modular networks and cumulative impact of lateral transfer in prokaryote genome evolution. *Proc. Natl. Acad. Sci. U. S. A.* 105, 10039–10044
3. 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
4. Koonin, E.V. *et al.* (2021) Evolution of microbial genomics: conceptual shifts over a quarter century. *Trends Microbiol.* 29, 582–592
5. Iranzo, J. *et al.* (2019) Gene gain and loss push prokaryotes beyond the homologous recombination barrier and accelerate genome sequence divergence. *Nat. Commun.* 10, 5376
6. Wolf, Y.I. *et al.* (2016) Two fundamentally different classes of microbial genes. *Nat. Microbiol.* 2, 16208
7. Medini, D. *et al.* (2005) The microbial pan-genome. *Curr. Opin. Genet. Dev.* 15, 589–594
8. Doolittle, W.F. (2004) W. Ford Doolittle. *Curr. Biol.* 14, R176–R177
9. Drinnenberg, I.A. *et al.* (2016) Evolutionary turnover of kinetochore proteins: a Ship of Theseus? *Trends Cell Biol.* 26, 498–510
10. Méheust, R. *et al.* (2018) Hundreds of novel composite genes and chimeric genes with bacterial origins contributed to haloarchaeal evolution. *Genome Biol.* 19, 75
11. Méheust, R. *et al.* (2018) Formation of chimeric genes with essential functions at the origin of eukaryotes. *BMC Biol.* 16, 30
12. Huerta-Cepas, J. *et al.* (2017) Fast genome-wide functional annotation through orthology assignment by eggNOG-Mapper. *Mol. Biol. Evol.* 34, 2115–2122