environmental microbiology reports

Environmental Microbiology Reports (2019) **11**(1), 9–12 doi:10.1111/1758-2229.12701

Towards a systematic understanding of differences between archaeal and bacterial diversity

Anja Spang[1,2*] and Pierre Offre[1]**

[1]*NIOZ, Royal Netherlands Institute for Sea Research, Department of Marine Microbiology and Biogeochemistry, and Utrecht University, NL-1790 AB, Den Burg, The Netherlands.*

[2]*Department of Cell and Molecular Biology, Science for Life Laboratory, Uppsala University, SE-75123, Uppsala, Sweden.*

Summary

In this crystal ball, we discuss emerging methodologies that can help reaching a synthesis on the biodiversity of Archaea and Bacteria and thereby inform a central enigma in microbiology, i.e. the fundamental split between these primary domains of life and the apparent lower diversity of the Archaea.

Humans have long marvelled at the variety of life forms that populate the Earth. This sense of an aesthetic of life together with the impending need to manage and preserve biological resources and ecosystem services is one of the driving forces that keeps pushing forward the exploration of life's biodiversity. Yet, in spite of continuing efforts to describe and categorize life forms, a complete inventory of all extant taxa and a comprehensive and systematic synthesis integrating the multiple aspects of biodiversity still lay far ahead. The invisible realm of microorganisms constitutes the largest and least explored reservoir of biodiversity and it is widely recognized that Bacteria represent an extraordinarily diverse group of organisms. In contrast, Archaea, which form a separate domain of life distinct from Bacteria and Eukarya, have only been described four decades ago and have been assumed to comprise much fewer taxa, many of which were initially thought to have only minor roles in global biogeochemical cycles. This point of view emerged from early discoveries, which described archaeal organisms thriving in environments characterized by extreme temperature, salt or pH regimes. The use of cultivation-independent approaches, most notably meta- and single cell genomics, has recently opened a window into the hidden world of microorganisms and has considerably expanded the number of known archaeal and bacterial taxa, revealing novel lineages of high taxonomic rank that are broadly distributed throughout the Biosphere (Adam *et al*., 2017; Spang *et al*., 2017; Castelle and Banfield, 2018). Yet, in spite of tremendous progress in sampling bacterial and archaeal genomes, archaeal taxa still appear to be outnumbered by bacterial counterparts, though estimates vary (Hug *et al*., 2016; Parks *et al*., 2018). This raises a number of fundamental questions, which need to be addressed carefully as the diversity of archaea and bacteria has deep connections to their ecology and mode of evolution and sampling microbial diversity is essential to uncover the deepest transitions in the evolution of life on Earth.

How accurate are the estimates of archaeal and bacterial taxonomic richness and how much diversity still defies discovery? Is the discrepancy of observed biodiversity between archaea and bacteria merely a result of evolutionary contingency or is there a deterministic explanation? For example; do archaea have a late origin from within bacteria, suggesting that bacteria had more time to evolve and diversify? Or are archaea characterized by different evolvability, e.g. do they experience different average rates of mutations, substitutions, recombination and horizontal gene transfer? Alternatively, did fundamental traits such as a different degree of adaptation to chronic energy stress (Valentine, 2007) of the primordial archaea and bacteria result in niche differentiation towards two distinct categories of environments? Finally, one may wonder, whether the biosphere realizes fewer niches in which archaeal organisms have a competitive advantage.

We want to take this opportunity to look into the crystal ball for discussing methodologies that can help reaching a synthesis on the biodiversity of archaea and bacteria and thereby inform these questions. Furthermore, we highlight some central insights that could emerge from such a synthesis.

The total number of archaeal and bacterial species (defined as groups of 16S amplicons sharing 97% sequence identity) has recently been predicted to be in the order of 1 trillion, less than 10^7 of which have been

Received 1 October, 2018; accepted 2 October, 2018. For correspondence. *E-mail anja.spang@nioz.nl; Tel. 0031-222-369526. **E-mail pierre.offre@nioz.nl; Tel. 0031-222-369489.

catalogued thus far (Locey and Lennon, 2016). While this estimate provides a powerful message in terms of commonness and rarity of 16S phylotypes, it is unclear how well it reflects the taxonomic richness of archaea and bacteria and to which extent it represents their genetic diversity. Certainly, the use of PCR-based 16S rRNA gene surveys often results in a biased picture of the microbial diversity in a sample, as for example revealed through primer-independent approaches. Thus, even though the use of 16S amplicon-based surveys is still commonly applied to assess microbial diversity, successes of PCR-independent approaches advocate for their widespread adoption. However, several challenges still stand in the way of a complete shift in methodology, particularly if metagenome sequencing takes over 16S amplicon surveys. For example, integrating the wealth of available genomic data will depend on the establishment of an improved computational infrastructure, that enables a centralized data storage and coherent data analysis (Koonin, 2018). Further, it remains to be proven that recovering the genome sequence of rare microorganisms can be routinely achieved. Yet, we are confident that the coming years will see further developments and standardization of metagenomics and single cell genomics helped by the introduction of novel library preparation protocols, sequencing technologies and bioinformatics tools and resources. We expect continuing discoveries of novel archaeal and bacterial taxa but also wonder whether broader taxonomic groups such as the CPR (Brown *et al*., 2015), DPANN (Rinke *et al*., 2013; Castelle *et al*., 2015) and Asgard archaea (Spang *et al*., 2015; Zaremba-Niedzwiedzka *et al*., 2017) still await discovery.

Beyond the sampling of bacterial and archaeal genome sequences, evaluating and comparing the taxonomic richness of the Archaea and Bacteria depends ultimately on the use of normalized taxonomic ranks and a resolved phylogeny that adequately reconstructs the evolutionary history of Bacteria and Archaea. For example, most taxonomies in common use are largely unstandardized: e.g. different genera may represent non-equivalent categories of organisms, which should be assigned to different ranks. Recently, however, a novel standardized taxonomy was proposed for Bacteria, which took into account the varying degrees of evolutionary divergence of distinct taxa for the assignment of ranks (Parks *et al*., 2018). This taxonomy established 111 phyla including 9911 species based on 94,759 genomes, 58% of which had a change to their existing taxonomy (Parks *et al*., 2018). For comparison, a similar approach established 11 phyla and 509 species for the archaea based on 2075 genomes (http://gtdb.ecogenomic.org/). Although this approach certainly represents a significant step towards an adequate taxonomy, it relies on the assumption that the topology, branch length and rooting of the phylogeny used for deriving the taxonomy are correct. Furthermore, it remains to be established whether the taxonomic ranks, which were assigned based on separately inferred phylogenies for bacteria and archaea, are comparable. Inferring the phylogeny of all archaea and bacteria with high accuracy remains a major challenge. For example, while models of protein evolution taking into account the compositional heterogeneities of amino acid sequences across branches and sites have already been developed (Blanquart and Lartillot, 2008; Heaps *et al*., 2014), current implementations of these models are computationally intensive and can only be applied to relatively small sets of archaeal and bacterial taxa. We foresee that the need to organize the current and forthcoming wealth of genome sequences and the perspective of a coherent and powerfully predictive taxonomy together with the wish of elucidating major evolutionary transitions will stimulate novel developments in phylogenetics and phylogenomics. This may include the computationally efficient implementation of models of evolution that better capture the different constraints on gene and protein sequences (Koonin and Wolf, 2010) and the development of novel methodologies that integrate the horizontal and vertical components of genome evolution and therefore provide a comprehensive representation of the relatedness of bacterial and archaeal genomes simultaneously (Dagan, 2011; Chan *et al*., 2013). We also hope that this will facilitate agreements upon a conceptual framework for the standardized ranking of taxa using genome sequence data (Konstantinidis *et al*., 2017).

These approaches will help to represent and organize the diversity of archaeal and bacterial taxa based on the genetic variation within a small subset of genes shared by these organisms (Fig. 1). However, defining biologically and ecologically cohesive units will ultimately require the complementary use of bottom-up approaches to the definition of minimum units of diversity (Achtman and Wagner, 2008). For instance, various studies suggest that a considerable amount of biological and ecological variation almost invariably exists within any taxon (Cordero and Polz, 2014; Larkin and Martiny, 2017). Population genomics has emerged more than a decade ago as a powerful approach to identify genetically cohesive units within natural populations but also to investigate diversification and speciation events as well as study the genetic basis of adaptation (Luikart *et al*., 2003; Shapiro and Polz, 2014). The general approach of population genomics to the ordering of microbial genetic diversity consists in comparing the genomes of closely related organisms (i.e. having identical or near identical 16S rRNA genes) derived from one or several populations followed by the clustering of the genomes based on single nucleotide polymorphisms within shared genes (Shapiro and Polz, 2014). This forms the basis for further

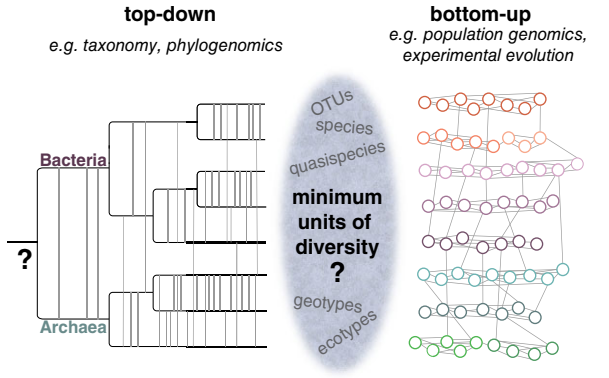

Fig. 1. Top-down and bottom-up approaches towards a synthesis on the diversity of Archaea and Bacteria. Highly simplified illustration of some of the top-down and bottom-up approaches, that will help organize archaeal and bacterial biodiversity and to determine, which factors have contributed to apparent differences in archaeal and bacterial diversity. Both types of approaches will inform the definition of the minimum units of diversity. The scheme on the left hand side represents a highly reticulated tree-like network, which represents the vertical (black lines of the tree) and horizontal component (grey lines) of archaeal and bacterial genome evolution. The scheme on the right-hand side illustrates populations of genomes corresponding to minimum units of diversity and the varying degree of exchange of their genetic information within and between these populations (grey lines).

investigations including for example the study of environmental selection, recombination patterns and ecological diversification. Experimental evolution (Hindre *et al*., 2012; Barrick and Lenski, 2013; Rainey *et al*., 2017) is another approach that has proven to be extremely valuable to test, from the bottom-up, the mechanisms driving diversification and niche differentiation in clonal populations of bacterial model systems. But in spite of the application of population genomics as well as experimental evolution to a limited number of microbial taxa, the genetic diversity of microorganisms remains overwhelmingly uncharacterized. In particular, only few studies thus far focus on archaeal population dynamics (Zhang *et al*., 2013; Papke *et al*., 2015) or provide empirical data on the mechanisms driving the evolution and diversification of archaeal model systems (Hillesland *et al*., 2014). We expect that the coming years will see an expansion of knowledge on the biology, genetics and evolution of natural microbial populations enabled by ever decreasing sequencing costs. Additionally, the intensification of high-throughput and highly parallelizable experimental evolution experiments (Cottinet *et al*., 2016) of an increased number of model systems as well as the application of experimental evolution to more complex communities will be of considerable value to determine fundamental principles of microbial genome evolution and speciation. And, while a detailed description of these methods is beyond the scope of this crystal ball, we believe that a holistic understanding of microbial diversity will also require the integration of approaches aiming at a systematic and high-throughput phenotyping of both archaea and bacteria as envisioned within the phenomics discipline (Houle *et al*., 2010). Finally, we believe that substantial progress will rely on the collective effort of a large number of networked research teams and the development of interoperable standards for collecting, analyzing and reporting data as for instance applied for the Earth microbiome project (Thompson *et al*., 2017).

The subsequent integration of knowledge from these complementary approaches will eventually bring us a step closer to shed light on a central enigma in microbiology, i.e. the fundamental split between the primary domains of life – Bacteria and Archaea – and the apparently lower diversity of the latter. We feel privileged to live in this 'era of big data' and methodological breakthroughs and are hopeful, that the coming years will witness substantial progress towards a better understanding of the diversity, nature and evolutionary trajectories of archaea and bacteria.

Originality – significance statement

In this crystal ball article, we point out a number of questions that connect the diversity of archaea and bacteria to their ecology, evolution and emphasize processes that may be the cause of their varying degree of diversity. We discuss emerging methodological approaches that can help establish a systematic understanding of archaeal and bacterial diversity.

Acknowledgments

This work was supported by a grant of the Swedish Research Council (VR starting grant 2016-03559 to A.S.), the NWO-I foundation of the Netherlands Organization for Scientific Research (WISE fellowship to A.S.).

References

Achtman, M., and Wagner, M. (2008) Microbial diversity and the genetic nature of microbial species. *Nat Rev Microbiol* **6**: 431–440.

Adam, P.S., Borrel, G., Brochier-Armanet, C., and Gribaldo, S. (2017) The growing tree of archaea: new perspectives on their diversity, evolution and ecology. *ISME J* **11**: 2407–2425.

Barrick, J.E., and Lenski, R.E. (2013) Genome dynamics during experimental evolution. *Nat Rev Genet* **14**: 827–839.

Blanquart, S., and Lartillot, N. (2008) A site- and time-heterogeneous model of amino acid replacement. *Mol Biol Evol* **25**: 842–858.

Brown, C.T., Hug, L.A., Thomas, B.C., Sharon, I., Castelle, C.J., Singh, A., *et al*. (2015) Unusual biology across a group comprising more than 15% of domain bacteria. *Nature* **523**: 208–2011.

Castelle, C.J., and Banfield, J.F. (2018) Major new microbial groups expand diversity and alter our understanding of the tree of life. *Cell* **172**: 1181–1197.

Castelle, C.J., Wrighton, K.C., Thomas, B.C., Hug, L.A., Brown, C.T., Wilkins, M.J., *et al*. (2015) Genomic expansion of domain archaea highlights roles for organisms from new phyla in anaerobic carbon cycling. *Curr Biol* **25**: 690–701.

Chan, J.M., Carlsson, G., and Rabadan, R. (2013) Topology of viral evolution. *Proc Natl Acad Sci U S A* **110**: 18566–18571.

Cordero, O.X., and Polz, M.F. (2014) Explaining microbial genomic diversity in light of evolutionary ecology. *Nat Rev Microbiol* **12**: 263–273.

Cottinet, D., Condamine, F., Bremond, N., Griffiths, A.D., Rainey, P.B., de Visser, J.A., *et al*. (2016) Lineage tracking for probing heritable phenotypes at single-cell resolution. *PLoS One* **11**: e0152395.

Dagan, T. (2011) Phylogenomic networks. *Trends Microbiol* **19**: 483–491.

Heaps, S.E., Nye, T.M., Boys, R.J., Williams, T.A., and Embley, T.M. (2014) Bayesian modelling of compositional heterogeneity in molecular phylogenetics. *Stat Appl Genet Mol Biol* **13**: 589–609.

Hillesland, K.L., Lim, S., Flowers, J.J., Turkarslan, S., Pinel, N., Zane, G.M., *et al*. (2014) Erosion of functional independence early in the evolution of a microbial mutualism. *Proc Natl Acad Sci U S A* **111**: 14822–14827.

Hindre, T., Knibbe, C., Beslon, G., and Schneider, D. (2012) New insights into bacterial adaptation through in vivo and in silico experimental evolution. *Nat Rev Microbiol* **10**: 352–365.

Houle, D., Govindaraju, D.R., and Omholt, S. (2010) Phenomics: the next challenge. *Nat Rev Genet* **11**: 855–866.

Hug, L.A., Baker, B.J., Anantharaman, K., Brown, C.T., Probst, A.J., Castelle, C.J., *et al*. (2016) A new view of the tree of life. *Nat Microbiol* **1**: 16048.

Konstantinidis, K.T., Rossello-Mora, R., and Amann, R. (2017) Uncultivated microbes in need of their own taxonomy. *ISME J* **11**: 2399–2406.

Koonin, E.V. (2018) Environmental microbiology and metagenomics: the brave New World is here, what's next? *Environ Microbiol*. https://doi.org/10.1111/1462-2920.14403.

Koonin, E.V., and Wolf, Y.I. (2010) Constraints and plasticity in genome and molecular-phenome evolution. *Nat Rev Genet* **11**: 487–498.

Larkin, A.A., and Martiny, A.C. (2017) Microdiversity shapes the traits, niche space, and biogeography of microbial taxa. *Environ Microbiol Rep* **9**: 55–70.

Locey, K.J., and Lennon, J.T. (2016) Scaling laws predict global microbial diversity. *Proc Natl Acad Sci U S A* **113**: 5970–5975.

Luikart, G., England, P.R., Tallmon, D., Jordan, S., and Taberlet, P. (2003) The power and promise of population genomics: from genotyping to genome typing. *Nat Rev Genet* **4**: 981–994.

Papke, R.T., Corral, P., Ram-Mohan, N., de la Haba, R.R., Sanchez-Porro, C., Makkay, A., and Ventosa, A. (2015) Horizontal gene transfer, dispersal and haloarchaeal speciation. *Life (Basel)* **5**: 1405–1426.

Parks, D.H., Chuvochina, M., Waite, D.W., Rinke, C., Skarshewski, A., Chaumeil, P.A., and Hugenholtz, P. (2018) A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat Biotechnol* **36**: 996–1004.

Rainey, P.B., Remigi, P., Farr, A.D., and Lind, P.A. (2017) Darwin was right: where now for experimental evolution? *Curr Opin Genet Dev* **47**: 102–109.

Rinke, C., Schwientek, P., Sczyrba, A., Ivanova, N.N., Anderson, I.J., Cheng, J.F., *et al*. (2013) Insights into the phylogeny and coding potential of microbial dark matter. *Nature* **499**: 431–437.

Shapiro, B.J., and Polz, M.F. (2014) Ordering microbial diversity into ecologically and genetically cohesive units. *Trends Microbiol* **22**: 235–247.

Spang, A., Caceres, E.F., and Ettema, T.J.G. (2017) Genomic exploration of the diversity, ecology, and evolution of the archaeal domain of life. *Science* **357**: eaaf3883.

Spang, A., Saw, J.H., Jorgensen, S.L., Zaremba-Niedzwiedzka, K., Martijn, J., Lind, A.E., *et al*. (2015) Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature* **521**: 173–179.

Thompson, L.R., Sanders, J.G., McDonald, D., Amir, A., Ladau, J., Locey, K.J., *et al*. (2017) A communal catalogue reveals Earth's multiscale microbial diversity. *Nature* **551**: 457–463.

Valentine, D.L. (2007) Adaptations to energy stress dictate the ecology and evolution of the Archaea. *Nat Rev Microbiol* **5**: 316–323.

Zaremba-Niedzwiedzka, K., Caceres, E.F., Saw, J.H., Backstrom, D., Juzokaite, L., Vancaester, E., *et al*. (2017) Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* **541**: 353–358.

Zhang, C., Krause, D.J., and Whitaker, R.J. (2013) Sulfolobus islandicus: a model system for evolutionary genomics. *Biochem Soc Trans* **41**: 458–462.