

## GENOME WATCH

# Microbial diversity knows no borders

Lindsay J. Pike, Elisa Viciani and Nitin Kumar



This month's genome watch discusses the vast global microbial diversity revealed by the Earth Microbiome Project and the value of its characterization.

It is well documented that only a small number of all microbial species thought to exist have been described. With estimates suggesting that more than 1 trillion ( $10^{12}$ ) species of bacteria, archaea and microscopic fungi exist<sup>1</sup> and our increasing understanding of their importance, large-scale studies of microbial diversity are required.

The Earth Microbiome Project (EMP) is a landmark study that is making such an attempt. Since its launch in 2010, it has overcome many challenges through an impressive effort in crowdsourcing and development of standardized protocols<sup>2</sup>. Now, the first meta-analysis results have been published: Thompson *et al.*<sup>3</sup> report on bacterial and archaeal 16S rRNA diversity in 27,751 samples from 97 independent studies. Producing 2.2 billion sequence reads, this study is 100-fold larger in both sampling and sequencing than previous meta-analyses of bacterial and archaeal diversity. Typically, 16S rRNA studies cluster sequence reads into operational taxonomic units (OTUs) by mapping against 16S references. A 97–98% similarity cut-off is used as a proxy for species-level classification, and clusters of reads are merged to form a 'consensus' 16S sequence per OTU for onward analyses. Only two-thirds of EMP reads could be mapped to existing 16S references, preventing meaningful OTU analysis. Therefore, a novel reference-free approach was developed: Deblur<sup>4</sup> removes reads that contain suspected errors and groups remaining reads by 100% identity. Using these exact sequences, or 'sub-OTUs', the EMP captured approximately 50% of currently described bacteria and archaea.

These accounted for fewer than 15% of all the sub-OTUs present, emphasizing the gaps in our understanding of microbial diversity. The authors also described broad ecological patterns in community structure. For example, samples showed high levels of 'nestedness' at the genus level or above. This is a measure of structure in which low-diversity samples have a subset of taxa also found in high-diversity samples, rather than harbouring a unique set of taxa. However, at lower taxonomic levels nestedness was reduced and unique species or strains not found in more diverse samples were identified. This was interpreted as individual species or strains showing specificity to particular environments or habitats, with genera and higher taxonomic groups reflecting shared microbial physiological features. However, the current standardized protocols may not recover all bacterial and archaeal DNA in a sample, which could particularly affect microorganisms with very different biology from what we currently understand. Additional coordinated procedures and analytical methods could help to discover novel taxa, especially those specific to extreme habitats.

The next phase of the EMP (informally 'EMP500') intends to perform shotgun metagenomic sequencing and analysis of 500 samples with high levels of unclassified data, from diverse environments. This follows a path laid out by the Human Microbiome Project (HMP), in which HMP1 analysed 16S data sampled from various sites on the human body<sup>5</sup> followed by shotgun metagenome analysis in HMP2 (REF. 6). As HMP2 has demonstrated, moving from 16S to shotgun metagenomics opens up new realms

— genome reconstruction, strain-level

analysis and pathway modelling — adding higher resolution to diversity and function. This may help to confirm the current findings and uncover the adaptation 'secrets' of microorganisms without their physical observation — a particular issue for the EMP. Although many challenges remain, Thompson *et al.* have proved that problems are there to be tackled: just 5 years ago the EMP was deemed unfeasible.

EMP500 is crucial to really understand microbial community diversity, structure and function from within particular environments and on a global scale. This is fundamental to conservation of these communities and their many ecological roles, as well as providing opportunities for translational studies. For example, a bacterium might be found that produces a completely novel antibiotic, or a community capable of reducing carbon dioxide in its environment. The possibilities are endless, and exciting.

Lindsay J. Pike, Elisa Viciani and Nitin Kumar are at the Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK.  
e-mail: [microbes@sanger.ac.uk](mailto:microbes@sanger.ac.uk)

doi:10.1038/nrmicro.2017.166

Published online 2 Jan 2018

1. Locey, K. J. & Lennon, J. T. Scaling laws predict global microbial diversity. *Proc. Natl Acad. Sci. USA* **113**, 5970–5975 (2016).
2. Gilbert, J. A., Jansson, J. K. & Knight, R. The Earth Microbiome project: successes and aspirations. *BMC Biol.* **12**, 69 (2014).
3. Thompson, L. R. *et al.* A communal catalogue reveals Earth's multiscale microbial diversity. *Nature* <https://doi.org/10.1038/nature24621> (2017).
4. Amir, A. *et al.* Deblur rapidly resolves single-nucleotide community sequence patterns. *mSystems* <https://doi.org/10.1128/mSystems.00191-16> (2017).
5. The Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature* **486**, 207–214 (2012).
6. Lloyd-Price, J. *et al.* Strains, functions and dynamics in the expanded Human Microbiome Project. *Nature* **550**, 61–66 (2017).

### Competing interests statement

The authors declare no competing interests.

