GENOME WATCH

A bit of a mouthful

James Hadfield & Sophia David

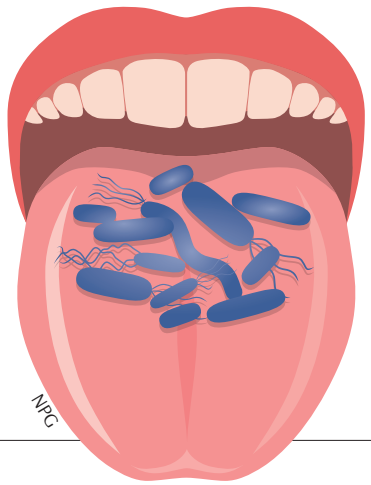
This month's Genome Watch explores recent advances in the identification of species-level and strain-level diversity in microbiome studies, and highlights how these have provided insights into the tropism and persistence of *Neisseria* spp. in the human oral cavity.

Metagenomics provides a powerful approach to survey complex microbial communities, such as those that are associated with humans, without the need for culture. Previous studies have identified microbial taxa linked to many disease states, including autoimmune disorders and obesity. However, until recently, most studies that probe the variation of microorganisms within metagenome samples have been limited to detecting high-level taxa. However, different species, and even strains of the same species, can have vastly different virulence and antimicrobial resistance profiles, as well as other important phenotypic differences. Recent advances in metagenomic analysis enable both species-level and strain-level detection and provide exciting new insights that were previously masked by the higher-level findings.

New methods for uncovering species-level and strain-level variation of bacteria within a community are taking advantage of the increasing number of reference genomes that are available. A common approach used by both Ahn *et al.*[1] and Sankar *et al.*[2] has been to map the sequence reads simultaneously to multiple reference genomes, or core genomes, of a target species. Statistical models are then used to infer the relative abundance of each species and/or strain based on patterns of polymorphisms. As the number of available reference genomes of a given species continues to increase, so will the resolution of these methods. For species that are not yet so well represented, other SNP-based methods such as those developed by Luo *et al.*[3] make use of just one reference genome for detecting strain-level variation within a target species.

One recent study applied various methods to explore the diversity of the *Neisseria* genus within the human oral cavity[4]. *Neisseria* species are closely related and undergo frequent recombination, which makes the identification of individual species and strains particularly challenging. Using a large reference collection of 241 draft and complete genomes that belong to the Neisseriaceae family, the authors used a custom mapping-based approach to analyse SNP patterns. They successfully identified different species from oral microbiome samples, revealing intricate associations between tissues and species.

Different oral sites within the same patients were enriched for different species, but, remarkably, the same sites across different patients showed marked similarities. For example, *Neisseria flavescens* and *Neisseria subflava* were generally found on the upper surface of the tongue, whereas *Neisseria sicca*, *Neisseria mucosa* and *Neisseria elongata* were found to populate the gingival (gum) plaque. *Neisseria meningitidis*, an important cause of sepsis and bacterial meningitis in young adults, was found only in a small number of throat samples. The authors confirmed their results by identifying short, 12-nucleotide-long features within the *Neisseria* genomes called DNA uptake sequences (DUSs), which are unique to distinct clades of *Neisseria* species. Analysis of the dominant multilocus sequence typing (MLST) alleles in each sample also supported the associations between specific species and tissues.



To probe even deeper than the species-level, the authors used a panel of strain-specific marker genes, revealing fascinating results. Although the same sites across different patients were colonized by the same species, individual patients harboured unique variants, or strains, of the species. Moreover, longitudinal sampling revealed that the presence of these strains was stable within individual patients. These important findings are only possible owing to the high accuracy of these new approaches.

Microbiome studies are advancing rapidly, and these latest tools bring us one step closer to fully characterizing and understanding the implications of the astounding diversity within microbial ecosystems. The increase in resolution beyond species-level detection will be particularly useful in identifying further associations with disease states, and understanding the spatial composition and function of microbial communities. These advances will also enable us to better understand disease risk, infer transmission events, characterize antibiotic resistance profiles and identify strains that are present in co-infections.

James Hadfield and Sophia David are at the Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK.
e-mail: microbes@sanger.ac.uk

doi:10.1038/nrmicro.2016.124
Published online 11 Aug 2016

1. Ahn, T. H., Chai, J. & Pan, C. Sigma:Strain-level inference of genomes from metagenomic analysis for biosurveillance. *Bioinformatics* **31**, 170–177 (2015).
2. Sankar, A. *et al.* Bayesian identification of bacterial strains from sequencing data. *Microb. Genom.* http://dx.doi.org/10.1099/mgen.0.000075 (2016).
3. Luo, C. W. *et al.* ConStrains identifies microbial strains in metagenomic datasets. *Nat. Biotechnol.* **33**, 1045–1052 (2015).
4. Donati, C. *et al.* Uncovering oral *Neisseria* tropism and persistence using metagenomic sequencing. *Nat. Microbiol.* **1**, 16070 (2016).

Competing interests statement
The authors declare no competing interests.