Complete, closed bacterial genomes from complex samples

The latest estimates suggest that the human body contains as many microbes (30 trillion) as it does human cells[1]. It is perhaps therefore unsurprising that the composition of these microbes can profoundly influence human health and disease. At Stanford University, USA, Dr. Ami Bhatt's lab is examining how changes in the human gut microbiome affect the treatment outcomes of immunocompromised individuals.

The team at Stanford University identified that long sequencing reads that can span bacterial repeat elements could significantly enhance metagenome assembly and subsequent genomic and functional characterisation of the gut microbiome. However, according to their recent Nature Biotechnology paper (Moss, Maghini, and Bhatt *et al.*[2]), the application of long-read sequencing technologies are currently hindered by the lack of efficient methods for extracting high molecular weight (HMW) DNA from stool samples.

Bead beating is commonly used to lyse bacterial cells in microbiome samples, but standard protocols result in excessive DNA shearing, while more gentle protocols may fail to extract DNA from difficult to lyse organisms (e.g. Gram-positive bacteria). In order to overcome this challenge and generate enhanced insights into human gut microbiomes, Professor Bhatt's team developed a complete, end-to-end metagenomics workflow, incorporating HMW DNA extraction from stool, long-read nanopore sequencing, and streamlined analysis (**Figure 1**).

Their novel extraction method comprises enzymatic digestion of the microbial cell wall (using a mixture of different lytic enzymes), phenol-chloroform extraction, proteinase K and RNase A digestion, gravity column purification, and SPRI bead size selection of HMW DNA. In addition, they developed Lathe, an analysis workflow combining existing basecalling (Guppy), assembly (Flye or Canu), and polishing tools (Racon and Medaka, or Pilon), with refined approaches for misassembly detection and genome circularisation[2,3].

> ***'This approach produces microgram quantities of pure, HMW DNA suitable for long-read sequencing from as little as 300 mg of stool'*** [2]

Validating their approach on a microbial mock community of 12 mixed Gram-positive and Gram-negative bacteria — sequenced using the MinION device — resulted in seven complete, closed contigs from 30.3 Gb of data with an N50 of 5.9 kb. Even the most fragmented assembly, at 8 contigs, contained 83% of the genome in a single contig. All species were recovered and the total assembly size (for all genomes) was in agreement with the known total reference length.

Based on these successful results, the team applied their methodology to 10 human stool samples. It was demonstrated that, in comparison to bead beating, their technique provided higher species-level diversity and assembly contiguity. In total, 1,219 genome drafts were recovered, including 11 fully circularised, single-contig genomes. These complete genomes included the first circularised *Prevotella copri* genome from a human sample, and a putative member of the recently described *Cibiobacter* clade[4]. Interestingly, alignment of previously generated short

read *P. copri* genome assemblies against the closed genome obtained in this study revealed that the breaks in the short-read assemblies corresponded with repetitive elements; further highlighting the advantages of long sequencing reads for accurate genome assembly.

Furthermore, comparison with an alternative long-read sequencing technology revealed that nanopore sequencing produced more complete assemblies and at a lower cost per base.

Commenting on this work, Dr. Bhatt stated that their method *'represents an effective, straightforward solution for the complete and efficient* de novo *characterization of structurally complex bacterial genomes within the gut microbiome'*[5].

The first complete *P. copri* genome assembly was achieved by researchers at The Roslin Institute, UK, who used nanopore sequencing to analyse the microbiomes of the cow rumen[6,7]. In this single sample study, they obtained over 31 circular microbial contigs, including the 3.8 Mb *P. copri* genome and the first assembly for the species *Selenomonas ruminatium*. Twenty-six small circular contigs were also resolved, corresponding to novel and known plasmids (one of which contained at least three AMR genes) and novel bacteriophages.

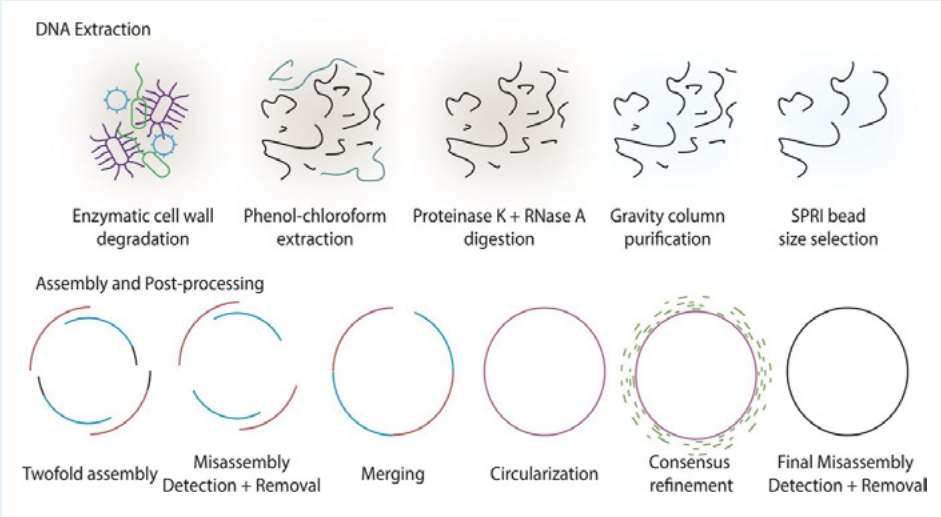

Figure 1
Optimised sample extraction and data analysis workflow for metagenomic analysis of gut microbiomes. Figure from Moss *et al.*[2] and available under Creative Commons license (creativecommons.org/licenses/by/4.0).

Find out more about metagenomic sequencing at www.nanoporetech.com/applications

References

1. Sender, R., Fuchs, S., and Milo, R. Revised estimates for the number of human and bacteria cells in the body. PLoS Biol. 14 (8), e1002533 (2016).
2. Moss, E.L., Maghini, D.G., and Bhatt, A.S. Complete, closed bacterial genomes from microbiomes using nanopore sequencing. Nat Biotechnol. [Online ahead of print] https://doi.org/10.1038/s41587-020-0422-6 (2020).
3. Maghini, D. Genomes from metagenomes: assembling bacterial genomes with nanopore sequencing. Presentation. Available at: https://nanoporetech.com/resource-centre/genomes-metagenomes-assembling-bacterial-genomes-nanopore-sequencing [Accessed: 20 February 2020]
4. Pasolli, E. et al. Extensive unexplored human microbiome diversity revealed by over 150,000 genomes from metagenomes spanning age, geography, and lifestyle. Cell 176, 649662. e20 (2019).
5. Bhatt, A.S. Closing highly repetitive bacterial genomes from metagenomic nanopore sequencing. Available at: https://nanoporetech.com/resource-centre/ami-bhatt-closing-highly-repetitive-bacterial-genomes-metagenomic-nanopore [Accessed: 20 February 2020]
6. Warr, A. Going full circle: assembly of high-quality, single-contig microbial genomes from the rumen microbiome using long-read sequencing. Presentation. Available at: https://nanoporetech.com/resource-centre/going-full-circle-assembly-high-quality-single-contig-microbial-genomes-rumen [Accessed: 26 February 2020]
7. Stewart, R.D. et al. Compendium of 4,941 rumen metagenome-assembled genomes for rumen microbiome biology and enzyme discovery. Nat. Biotechnol. 37(8):953-961 (2019).