High-accuracy, long-read sequencing of microbial isolates and standards enables AMR profiling, methylation detection, and taxonomic annotation

Lynn Ly, Richard Fetherston, Samriti Midha, Phillip James, Patrick West, John Beaulaurier, Scott Hickey, George Dodd, Matthew Heatly, Sean McKenzie, Philipp Rescheneder, Sissel Juul
Oxford Nanopore Technologies Inc, New York, NY
Contact: @nanoporetech.com

Abstract

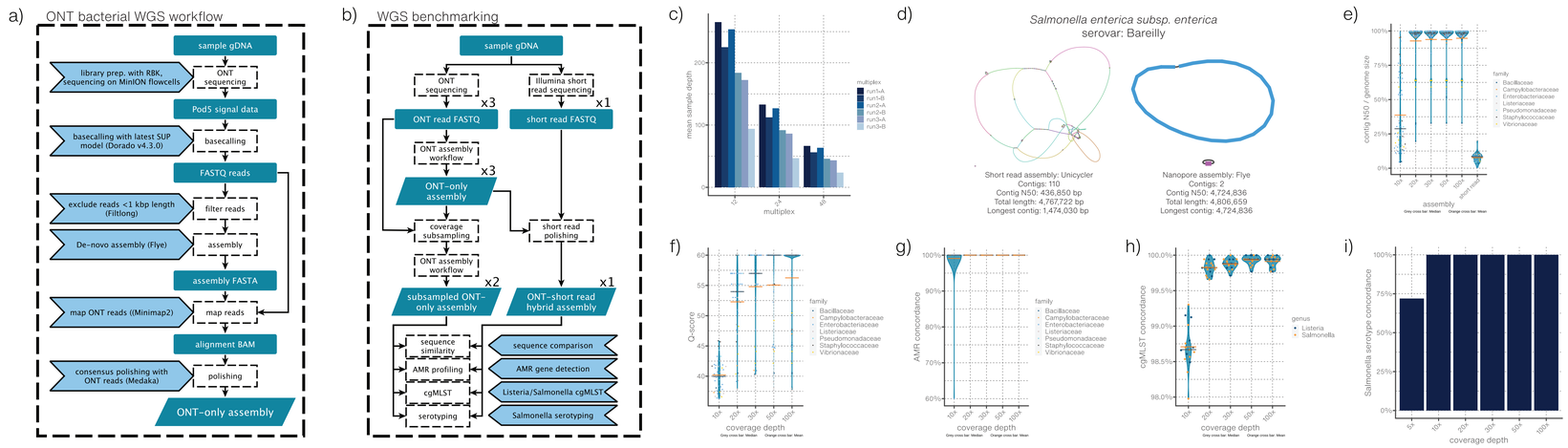
Genomic and metagenomic sequencing is revolutionizing the field of microbiology. Previously, tradeoffs between read length and accuracy in short read sequencing and early long read sequencing have limited the completeness and accuracy of genomic and metagenomic analyses; however, recent advancements in nanopore sequencing have significantly improved long read accuracy for microbial genomic applications[1,2].

Here, we use nanopore sequencing with the latest chemistry and basecalling models to characterize the genomes of 32 bacterial isolates sourced from the FDA CFSAN proficiency panel in triplicate, multiplexed on MinION Flow Cells. Nanopore-only assemblies achieved high completeness and accuracy (mean/median >Q50) when compared to publicly available reference genomes. We found that 20x coverage was sufficient for performing AMR profiling and *Salmonella* serotyping with 100% accuracy.

Because DNA methylation has been implicated in virulence in a variety of bacterial pathogens, we performed epigenetic characterization of a subset of the FDA strains. Detection of 4mC, 5mC, and 6mA modified bases in four near-identical *Salmonella enterica* isolates revealed complex bacterial epigenomes and uncovered a methylation-dependent mechanism affecting serotype.

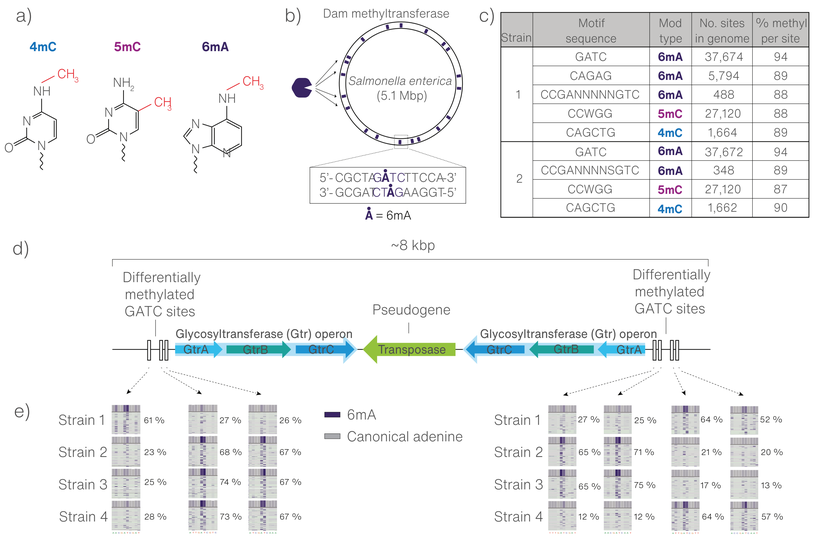
We also investigated the performance of nanopore sequencing for metagenomic assembly and composition profiling using the ZymoBIOMICS Fecal Reference sample. We generated almost 400 medium or high-quality metagenome assembled genomes from this sample, spaning a wide range of taxa, and show that the *Sylph* tool is able to classify species within the sample with high precision and recall.

1. Assembly of FDA CFSAN bacterial isolates



To evaluate the suitability of nanopore sequencing as a standalone platform for bacterial isolate WGS, 32 CFSAN bacterial isolate strains were sequenced in triplicate on MinION Flow Cells (RBK, 72 hrs sequencing time, 16 samples per flow cell) and compared to extraction-matched hybrid assemblies generated with both nanopore and short read data. Nanopore-only assemblies were produced by basecalling with *Dorado*[3] (v4.3.0 SUP model), assembling with *Flye*[4], and polishing with *Medaka*[5], as implemented in the Epi2Me™ *wf-bacterial-genomes* workflow[6] (Fig. 1a). *In silico* resampling showed samples could achieve a coverage of ~90-240x with 12 samples per flowcell, 40-120x with 24 samples, or 20-70x with 48 samples (Fig. 1c). 25 isolate samples achieving >= 100X coverage were used for WGS benchmarking (Fig. 1b). For benchmarking, reads were subsampled nanopore in duplicate to 10-100x depth (n=44), assembled, and then compared against matched nanopore-short read hybrid assemblies. We show that nanopore assemblies are highly contiguous with as little as 20x coverage (Fig. 1d, e) – especially compared to short read sequencing (Fig 1d). Nanopore-only assemblies with 100x coverage achieved a mean/median per-base accuracy of Q57/Q60 (Q60 assemblies in 33/44 replicate), or Q55/Q60 with just 50x (Fig. 1f). In several common sequencing applications for foodborne pathogen typing, nanopore-only genomes showed 100% concordance with hybrid assemblies with as little as 20X coverage (Fig 1g), achieved a mean cgMLST profiling concordance of 99.9% with >=50x coverage (Fig. 1h), and needed only 10x coverage for 100% accurate *Salmonella* serotyping (Fig. 1i).

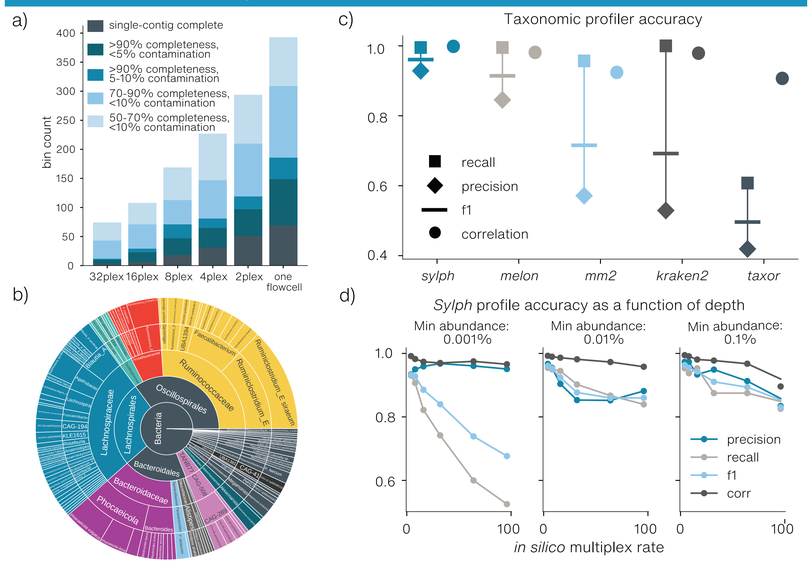
2. Microbial strain epigenetics



Strain	Motif sequence	Mod type	No. sites in genome	% methyl per site
1	GATC	6mA	37,674	94
	CAGAG	6mA	5,794	89
	CCGANNNNNGTC	6mA	488	88
	CCWGG	5mC	27,120	88
	CAGCTG	4mC	1,664	89
2	GATC	6mA	37,672	94
	CCGANNNNNSGTC	6mA	348	89
	CCWGG	5mC	27,120	87
	CAGCTG	4mC	1,662	90

Bacterial genomes have modified bases that play important biological roles (Fig. 2a). Bacterial methyltransferases deposit methylation in a motif-specific manner; e.g., the Dam methyltransferase in *Salmonella enterica* places 6mA at all instances of the motif GATC in the genome (Fig. 2b). Epigenetic profiling of multiple strains of *S. enterica* revealed 4mC, 5mC, and 6mA modified motifs (Fig. 2c). In *S. enterica,* the glycosyltransferase (Gtr) operon modifies the O-antigen in the lipopolysaccharides in the outer membrane of the cell wall. Transcription of the Gtr operon is controlled by the methylation status of several GATC sites upstream of the transcription start site (TSS)[7] (Fig. 2d). The *S. enterica* isolates sequences, while differing by <200 SNVs genome wide, showed heterogeneous 6mA methylation signatures at these important GATC sites (Fig. 2e). This suggests that the O-antigens are likely modified differently between these nearly identical strains, affecting serotype.

3. Metagenomic assembly and classification



MetaMDBG[8] assembly and *Semibin2*[9] binning of the ZymoBIOMICS Fecal Reference sample[10] sequenced on a single PromethION Flow Cell recovered hundreds of medium-or-greater quality metagenomic assembled genomes (MAGs), including 148 high-quality MAGs (HQ MAGS), 69 of which were assembled to a single contig (Fig. 1a). *In silico* multiplexing of samples still recovers dozens of HQ MAGS (Fig. 1a). Genomes span a range of taxa (Fig. 1b). A subset of reads mapping to 204 MAGs with unambiguous species assignment was analyzed with multiple taxonomic profiling tools, most of which showed high recall and abundance correlation with the truth. However, only *Sylph*[11] and *Melon*[12] showed high precision in this analysis (Fig. 1c). For all but the lowest abundance species, Sylph showed high accuracy even when multiplexing many samples per PromethION Flow Cell (Fig. 1d)

Conclusions

Our study demonstrates that nanopore sequencing yields reliable, high quality results for genomic investigations of both microbial isolate and metagenomic samples.

Isolate sequencing is the backbone of microbial genomics. We show that for many applications, 10-20X isolate coverage with nanopore reads - obtainable while multiplexing 48 samples on a single MinION Flow Cell - is sufficient for high concordance with the highest quality hybrid assemblies, AMR and Salmonella serotyping. Essentially perfect genomes (Q60) are often obtainable by sequencing to 50X coverage.

The ability of nanopore sequencing to simultaneously identify both canonical and modified bases in a single read allows for simultaneous genetic and epigenetic profiling. We show that strain-specific modification motifs can be uncovered with nanopore sequencing, and furthermore show an example where nanopore sequencing elucidates the epigenetic control over a serotype-influencing locus in *Salmonella enterica*.

We present evidence that nanopore sequencing is highly effective for analyzing complex metagenomic samples, showcasing its potential for *de novo* assembly and taxonomic profiling. Our results demonstrate that, depending on the sequencing depth, a single sample can yield dozens to hundreds of assembled genomes. Notably, we achieve remarkable accuracy in classifying organisms, even at a mere 0.01% abundance, when multiplexing 96 samples on a PromethION Flow Cell.

References

1. Zhao et al. *Frontiers in Microbiology* 14:1179966 (2003)
2. Kim et al. *Journal of Translational Medicine volume* 22: 111 (2024)
3. https://github.com/nanoporetech/dorado/releases
4. Kolmogorov et al. *Nature Biotechnology* 37: 540–546 (2019)
5. https://github.com/nanoporetech/medaka
6. https://github.com/epi2me-labs/wf-bacterial-genomes
7. Broadbent et al. *Molecular Microbiology* 77(2):337-53 (2010)
8. Benoit et al. *Nature Biotechnology* doi:10.1038/s41587-023-01983-6 (2024)
9. Pan et al. *Bioinformatics* 39: i21–i29 (2023)
10. https://zymoresearch.eu/pages/zymobiomics-fecal-reference-with-trumatrix-technology
11. Shaw and Yu. *bioRxiv* 2023.11.20.567879 (2023)
12. Xi et al. *bioRxiv* 2023.12.17.572079 (2023)

Oxford Nanopore Technologies, the Wheel icon, MinION are registered trademark of Oxford Nanopore Technologies plc in various countries. All other brands and names contained are the property of their respective owners. © 2024 Oxford Nanopore Technologies plc. All rights reserved.
Oxford Nanopore Technologies products are not intended for use for health assessment or to diagnose, treat, mitigate, cure, or prevent any disease or condition.