Evaluating a method for metagenomic sequencing of faecal specimens referred to the Gastrointestinal Bacteria Reference Unit

Israel Olonade, Ella Rodwell, Anais Painset, David Greig, Claire Jenkins, Craig Swift

UKHSA (Gastrointestinal Bacteria Reference Unit)

INTRODUCTION

Surveillance, outbreak detection and investigation of zoonotic, foodborne pathogens is undertaken at UKHSA by the Gastrointestinal Bacteria Reference Unit using Whole Genome Sequencing data generated via Illumina NextSeq1000 platforms from DNA extracted from single colony bacterial cultures.

We aimed to evaluate the application of metagenomic sequencing using Oxford Nanopore Technology to detect and characterize microbial pathogens from DNA extracts prepared from faecal material.

METHODS

We have optimised a manual in-house DNA extraction protocol incorporating chemical (DNA/RNA Shield™), enzymatic (Metapolzyme and Proteinase K) and mechanical (ZR BashingBead) lysis of bacterial cells present in faecal material.

DNA extracts were purified using MagBinding Beads prior to library preparation and sequencing on either a MinION or GridION platform. Bioinformatics analyses included basecalling using Guppy, read trimming using PoreChop and taxonomic profiling using Kraken (and Krakenuniq) (figure 1).

The in-house protocol was evaluated in a comparison matrix with commercial kits using different commercial reference standards.

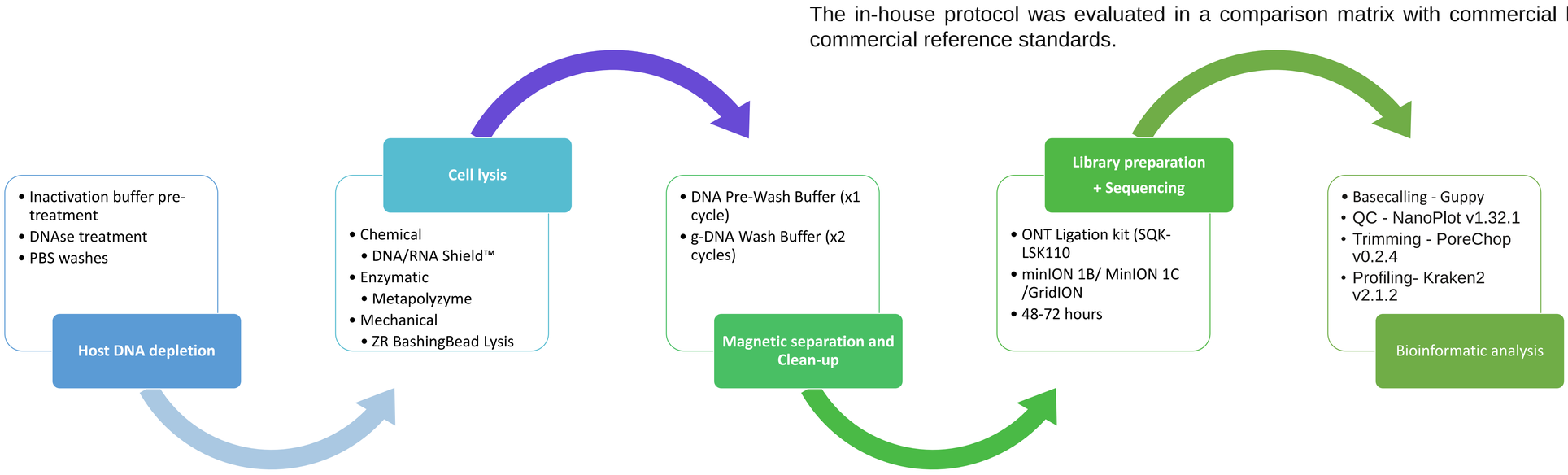

Figure. 1 Summary of the current metagenomics workflow in GBRU

RESULTS AND DISCUSSION

- Detected all expected bacteria species in correct proportions validating the library preparation, sequencing criteria and bioinformatic analysis protocols employed (figure 2).
- Yeast references were not present in the Kraken database used for microbial identification, therefore these species were represented as unclassified reads (figure 2).

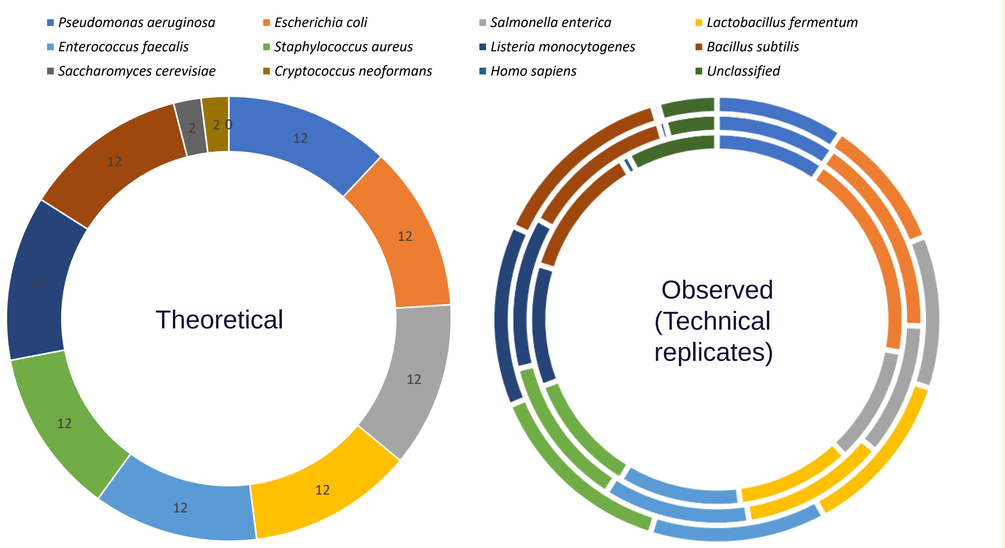

Figure. 2: Comparison of the theoretical composition of ZymoBIOMICS™ Microbial Community Standard (left) with the observed composition in triplicate (right)

- Preliminary run metrics for High molecular weight DNA showed good reproducibility with N50 greater than 25kb (figure 3), validating our ability to sequence long reads when present in a sample.

Run ID	Date	Guppy V + model	Number of reads	N50 (bp)	Average read length (bp)	Yield (Mbp)
0143XBD	12.12.2022	v5.0.12 HAC	832,680	25,036	4,595.10	3,826.22
0155XBD	21.02.2023	v6.4.6 HAC	1,584,081	25,049	4,451.20	7,051.11
0157XBD	06.03.2023	v6.4.6 HAC	2,774,884	25,433	4,522.50	12,549.44

Figure. 3: Demonstration of reproducibility of library preparation protocol and bioinformatics analysis using ZymoBIOMICS™ High molecular weight DNA standard

- Similar microbial species were detected across all three extraction methods demonstrating the suitability of all three protocols for total microbiome studies (figure 4).
- Although similar percentages of spiked in *E. coli* was detected when the reference material was extracted with commercial kits (14.77% & 14.14% respectively) compared to <3% for the in-house protocol, very high percentage of unclassified reads makes it difficult to draw firm conclusions (figure 4).
- More compute power is required to process the large sequence database required for the complete analysis of these high molecular weight standards.

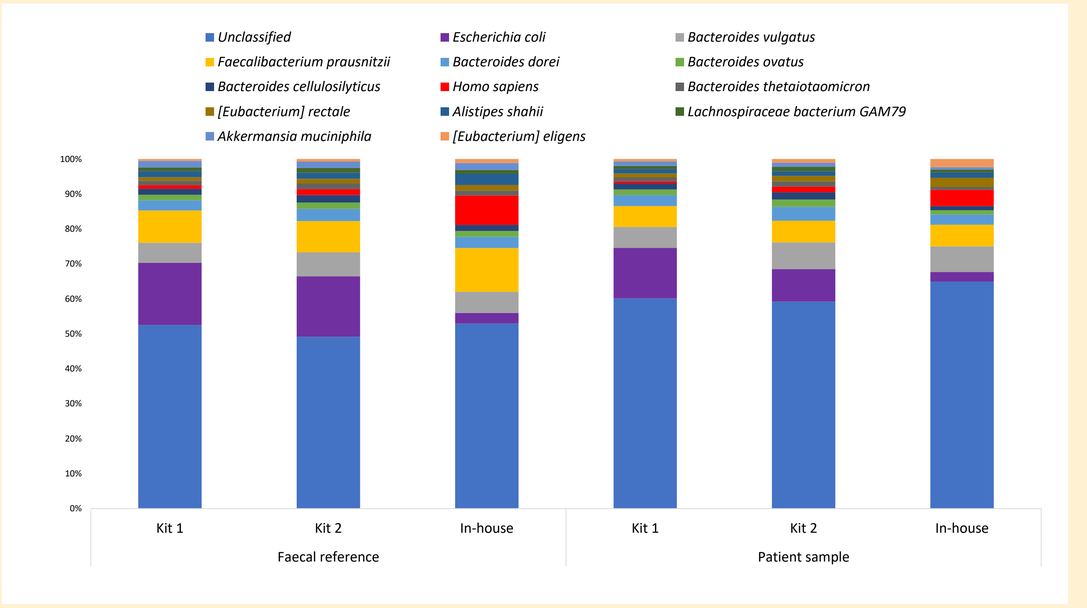

Figure. 4: Comparison of the in-house protocol with two commercial kits using ZymoBIOMICS™ TruMatrix™ Technology Fecal reference material and a patient sample both spiked with 10^6cfu/ml of *E. coli*

- These preliminary results suggest that Kits 1 & 2 are potential candidates for routine metagenomic investigations because of the ease of use and possibility of high throughput DNA extraction.
- The in-house method might be the most applicable to 'unique' questions like resistome exploration if we can establish the lack of bias. Conversely, the presence of bias can be an advantage if targeted sequencing is the method of choice for the downstream processing of that 'unique' sample.

CONCLUSIONS

- Metagenomic sequencing of DNA extracted directly from clinical faecal specimens can enable us to detect and characterise microbial pathogens without bias inherent in the culture and isolation of single colonies.
- Long read sequencing which bypasses PCR amplification can be applied to faecal specimens for studies of the gut microbiome with an understanding of current limitations.
- Biases observed during genetic material extraction can be further explored as a form of 'targeted sequencing'.

FUTURE WORK

- Re-analyse unclassified reads with a more diverse kraken database.
- Challenge all DNA extraction protocols with more patient samples.
- Complete validation of bioinformatic pipeline for antimicrobial resistance determinants.

REFERENCES

- Bertrand, D. *et al.*.(2018). Nanopore sequencing enables high-resolution analysis of resistance determinants and mobile elements in the human gut microbiome. preprint. *Bioinformatics*. Doi: 10.1101/456905.
- Wood, D.E., Lu, J. & Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. Genome Biol 20, 257. Doi: 10.1186/s13059-019-1891-0.

ACKNOWLEDGEMENTS

The National Institute for Health Research Health Protection Research Unit (NIHR HPRU) in Gastrointestinal Infections, a partnership between the UK Health Security Agency, the University of Liverpool and the University of Warwick.

The National Institute for Health Research Health Protection Research Unity (NIHR HPRU) in Gastrointestinal infections for funding conference attendance.

Katherine Lock, Marwa Al-Sabah, Dawn Hedges, Amaal Ibrahim, and Shukri Mohamed for their contribution to sample preparation and general laboratory duties during this study.