Evaluation of rapid Nanopore-based 16s rRNA gene sequencing as diagnostic tool for clinical service

R. Baldan[1], P. Cliff[2], R. Batra[3], A. Patel[3], S. Lewis[2], A. Cerda[2], M. Mashhadani[4], D. Jeyaratnam[5], J. Edgeworth[6], M. Chand[7]

[1]King's College London - GSTT, London, UK; [2]Viapath - GSTT, London, UK; [3]St Thomas' Hospital, London, UK; [4]Viapath – KCH, London, UK; [5]King's College Hospital, London, UK; [6]Viapath - GSTT, King's College London, UK; [7]Public Health England, London, UK

Background

Traditionally, the diagnosis of bacterial infections relies on culture based techniques considered the gold standard of pathogen identification in clinical microbiology.

However, standard microbiology has limited sensitivity, specificity and slow turnaround time leading to overuse of broad spectrum antibiotics and emergence of antimicrobial resistance. Poor diagnostics also lead to poor patient outcomes, prolonged hospital stay and increased costs. Oxford Nanopore Technologies (ONT) long read sequencing holds promise of a microbiology revolution generating data analysable in real-time.

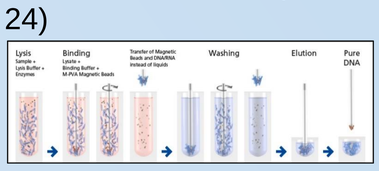
We started evaluating an ONT based 16S rRNA gene assay, to investigate whether low accuracy long reads and high speed analysis can combine to a clinically useful diagnostics.

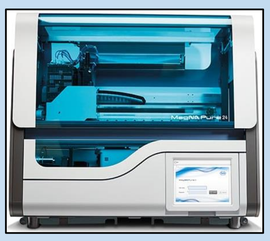
Methods

A bank of bacterial isolates representing pathogenic species received by the clinical laboratory over 1 year was assembled.
Sample preparation comprised bead beating, MagNAPure extraction (Roche) and 16S barcoding library kit (ONT) multiplexing 12 samples per flowcell. Sequencing was performed for 16 hours on the MinION and GridION X5 platforms using R9.4.1 flowcells with live basecalling.
Fastq files were analysed with the ONT Epi2Me 16S pipeline which assigns reads to taxa using BLAST results and the NCBI Bacterial 16S database (Figure 1).

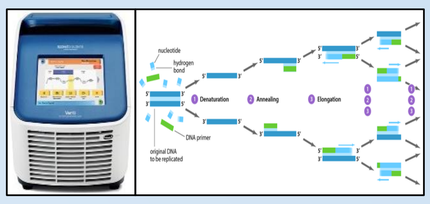
A. Bacterial cells lysis (FastPrep 24)

B. DNA extraction & purification (MagNA Pure 24)

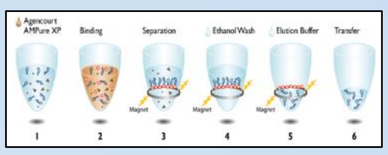




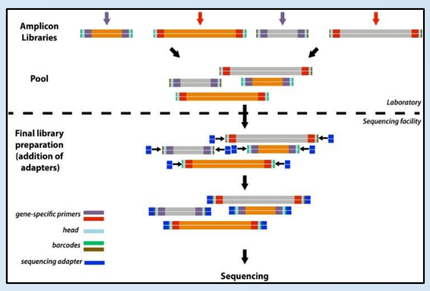
C. PCR



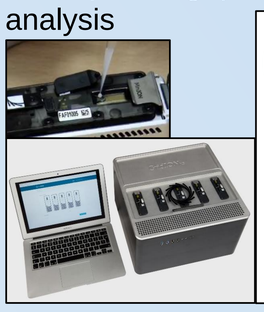
D. Amplicons clean-up (Biomek 4000)



E. Library preparation



F. Sequencing (MinION, GridION X5) and data analysis



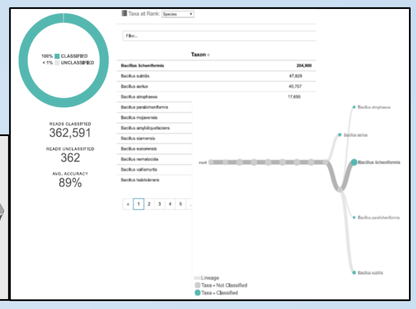


Figure 1. ONT 16S rRNA sequencing workflow

Results & Discussion

- 155 bacterial isolates representing 139 different species confirmed by MALDI-TOF were sequenced
- 119 species were correctly identified at the species level (i.e. most reads were assigned to the correct species)
- For the other 20 species, a species in the same genus claimed the majority of reads, with the true species being matched to <3%-35.5% of reads
- The average proportion of classified reads assigned to the correct species was 62.7%, specifically 66.8% for non-Enterobacteria and 38.1% for Enterobacteria
- As proof of concept, 4 clinical samples (3 BALs, positive by culture for (1) *K. pneumoniae*, (2) *S. pneumoniae* and (3) *S. pneumoniae* and *S. enterica* s. typhimurium, and 1 bone positive for *P. aeruginosa*) were also analysed and sequencing results matched culture microbiology.

SPECIES	READS CLASSIFIED	READS MATCHING GENUS	READS MATCHING SPECIES	OTHER SPECIES DETECTED ABOVE 3% (n. reads)
A. ursingii	169,020	**168,135**	**50,647 (29.9%)**	**A. septicus (104360); A. calcoaceticus (6141)**
C. braakii	143,912	**45,021**	**3,532 (2.4%)**	**C. murliniae (24572), C. freundii (15555)**
D. acidovorans	235,891	**156,339**	**2,673 (1.1%)**	**D. lacustris (79455); D. tsuruhatensis (74081)**
E. miricola	83,979	**55,623**	**6,090 (7.2%)**	**E. meningoseptica (30152), E. anophelis (19381), Chryseobacterium molle (7660), C. formosense (4471)**
E. asburiae	313,368	**20,287**	**827 (0.2%)**	**E. cancerogenus (7819); E. bugandensis (5145); E. xiangfangensis (2325); E. ludwigii (1715)**
E. kobei	214,190	**13,409**	**103 (0.04%)**	**E. tabaci (10987)**
E. ludwigii	12,459	**1,214**	**265 (2.1%)**	**E. cloacae (914)**
E. casseliflavus	114,581	**100,474**	**25,128 (21.9%)**	**E. gallinarum (35861), E. pseudoavium (13957), E. saccharolyticus (5246), E. malodoratus (3881)**
E. raffinosus	121,691	**102,555**	**19,494 (16%)**	**E. avium (21284); E. malodoratus (20199); E. pseudoavium (13785); E. gilvus (12756)**
E. coli	626,910	**27,919**	**2,061 (0.3%)**	**E. fergusonii (24,212) Shigella sonnei (2,729)**
H. alvei	286,836	**3,885**	**1621 (1.35%)**	**H. paralvei (2264); Serratia quinivorans (2044)**
K. variicola	355,606	**162,460**	**3,963 (1.1%)**	**K. pneumoniae (156484)**
L. monocytogenes	477,349	**425,336**	**15,334 (3.2%)**	**L. welshimeri (277696); L. innocua (97002); L. seeligeri (20777)**
P. gergoviae	359,853	**1,333**	**0**	**K. oxytoca (2272)**
P. stuartii	297,190	**177,105**	**9,267 (3.1%)**	**P. vermicola (130219), P. rettgeri (22434), P. burhodogranariae (10545)**
P. koreensis	192,357	**190,108**	**26,724 (13.8%)**	**P. graminis (55145), P. moraviensis (19665), P. vancouverensis (15880), P. baetica (14440), P. libanensis (10112), P. helmanticensis (8790), P. agarici (8587)**
P. putida	308,270	**305,362**	**1,436 (0.4%)**	**P. plecoglossicida (133637); P. graminis (45024); P. taiwanensis (41847); P. monteilii (22617); P. flavescens (19469)**
P. taiwanensis	132,489	**131,501**	**16,756 (12.6%)**	**P. plecoglossicida (57315), P. graminis (26156), P. flavescens (8683), P. monteilii (7585), P. pseudoalcaligenes (4456)**
S. caprae	79,617	**73,341**	**28,318 (35.5%)**	**S. capitis (33272); S. saccharolyticus (5071)**
S. oralis	96,354	**95,573**	**8,709 (9%)**	**S. mitis (75832); S. pneumoniae (3584)**

Concerning misidentified species:

- 9/20 are Enterobacteria for which only a small proportion of reads is assigned to genus and species by Epi2Me
- *A. ursingii* was misidentified as *A. septicus* that has not been proven to be different from the former (Nemec, JCM 2008)
- *D. acidovorans*: according to Ranc *et al.* (EID 2018) *D. tsuruhatensis* can be misidentified by MALDI as *D. acidovorans* in 50% of the cases
- MALDI is not reliable for the discrimination between *S. oralis* and *S. mitis*

Our early results show that despite reduced accuracy, nanopore sequencing with raw read analysis is sufficiently discriminatory to speciate the majority of species.

Pipeline refinement is required for Enterobacteriaceae and subsequent confirmatory consensus based identification may be a helpful adjunct. Nanopore based 16s shows potential for rapid clinical service development.