High-quality Oxford Nanopore assemblies reveal structural completeness, plasmid diversity, and antimicrobial resistance genes in NCTC bacterial genomes

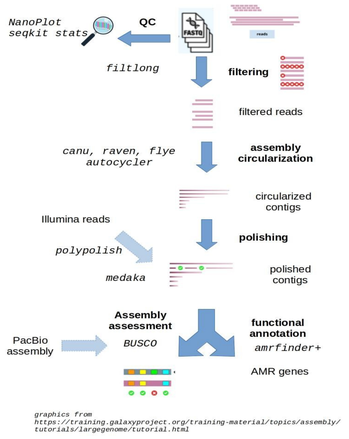
Marielle Vigouroux[1], Naomi Davis[2], Arinder Kohli[1], Dorothy Dewan[2], Alexander Dickinson[1], Rupa Rai[1], Jake Turnbull[1], Hannah McGregor[1], Sarah Alexander[1], Jon Hubb[2] and Jo Dicks[1]

1. Culture Collections, United Kingdon Health Security Agency, 61 Colindale Avenue, London
2. Central Sequencing Laboratory, United Health Security Agency, 61 Colindale Avenue, London

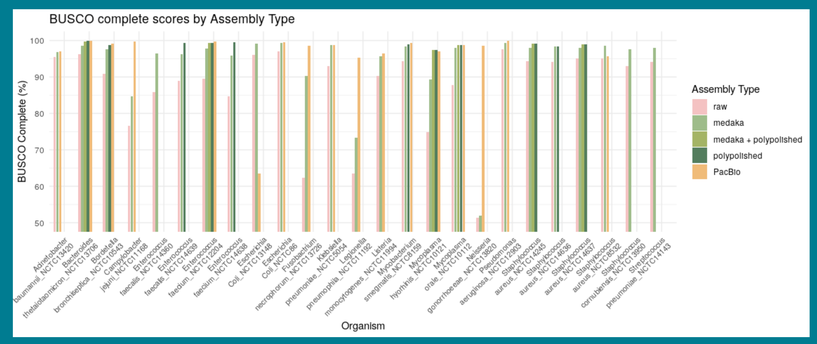
BACKGROUND:

The National Collection of Type Cultures (NCTC; https://www.culturecollections.org.uk/about-us/nctc/) comprises approximately 6,000 bacterial strains, many of historical, medical or veterinary significance. Culture Collections worldwide now recognise the immense value that can be added to such biological resources through whole genome sequencing. Previously, the NCTC3000 project generated long PacBio reads, genome assemblies and annotations for almost 3,000 strains (PRJEB6403). In this new pilot project, we investigated the pros and cons of using Oxford Nanopore Technologies (ONT) in generating further NCTC genome assemblies and annotations.

PIPELINE



Consensus-based ONT assembly pipeline can reliably produce complete bacterial genomes



UK Health Security Agency

Culture Collections

Data coming soon at ENA

Marielle Vigouroux
marielle.vigouroux@ukhsa.gov.uk

SEQUENCING:

We successfully sequenced 25 NCTC bacterial strains using ONT on a GridION X5 Mk1 (GXB03630) platform with R9.4.1 (FLO-MIN106) flow cells. Illumina short reads were also available for 11 out of the 25 strains.

Basecalling was performed using Guppy (versions 6.4.6 to 6.5.7) with the high-accuracy model (450 bps) for early runs, and Dorado (v7.3.11) for later runs.

NCTC ID	Organism	Family	Contigs	Plasmids	N50 (bp) raw / filtered reads	Assembly Size (bp)	Average read quality raw data / after filtering	BUSCO (C/F/D/M)	Coverage raw / filtered reads	Illumina data	Pacbio assem
86	Escherichia Coli	Enterobacteriaceae	1	0	15,502/18,319	5,100,493	12.3/12.9	99.4 / 0.7 / 0.5 / 0.1	1,756x / 1,295x	No	Yes
13148	Escherichia Coli	Enterobacteriaceae	9	8	18,746	5,180,748	12.6	99.2 / 0.6 / 0.0 / 0.2	390x	No	Yes
5054	Klebsiella pneumoniae	Enterobacteriaceae	5	4	21,358/29,822 / 24,586	5,242,133 / 5,242,120	12/12.7 / 13.1	98.7 / 0.7 / 0.5 / 0.6	2,405/1,617x 69x	No	Yes
11168	Campylobacter jejuni	Campylobacteraceae	1	0	5,895/9,813	1,641,423	12/12.5	94.7 / 8.8 / 0.2 / 6.5	2,750x/1,187	No	Yes
11994	Listeria monocytogenes	Listeriaceae	1	0	15,254 / 12,630	2,979,902	12.6/12.7	98.5 / 0.8 / 0/ 0.7	2,381x / 1,829x	No	Yes
12204	Enterococcus faecium	Enterococcaceae	8	7	21,599/22,442	3,325,898	12.8/14.2	99.3 / 0.5 / 1.0 / 0.2	2,107x/1,821x	Yes	Yes
14638	Enterococcus faecium	Enterococcaceae	9	8	25,599	3,056,430	12.3	99.6 / 0.1 / 0.6 / 0.3	100x	Yes	No
14360	Enterococcus faecalis	Enterococcaceae	3	2	24,387	2,959,033	12.6	96.6 / 2.2 / 0.7 / 1.4	489x	No	No
14639	Enterococcus faecalis	Enterococcaceae	3	2	23,817	3,057,694	12.4	99.4 / 0.0 / 1.0 / 0.6	184x	Yes	No
13820	Neisseria gonorrhoeae	Neisseriaceae	2	1	18,382/19,218	2,176,796	12.3/13.4	46.7 / 13.4 / 0.0 / 39.9	4,853x	No	Yes
8159	Mycobacterium smegmatis	Mycobacteriaceae	1	0	11,827	6,983,200	13.0	98.9 / 0.4 / 0.7 / 0.3	30x	Yes	Yes
12903	Pseudomonas aeruginosa	Pseudomonadaceae	1	0	8,918	6,839,827	12.9	99.0 / 0.7 / 0.7 / 0.3	65x	No	Yes
13706	Bacteroides thetaiotaomicron	Bacteroidaceae	3	1	15,872 / 16,962 / 9,962	6,166?,802	12.2 / 12.9 / 13.9	99.9 / 0.1 / 0.0 / 0.0	2,069x 96x	Yes	Yes
10543	Bordetella bronchiseptica	Alcaligenaceae	2	1 Klebsiella plasmid	10,303	5,204,802	12.4	98.7 / 0.6 / 0.4 / 0.7	473x	Yes	Yes
13420	Acinetobacter baumannii	Moraxellaceae	3	2	19,014	4,000,591	12.7	96.8 / 1.2 / 0.3 / 2.0	495x	No	Yes
8532	Staphylococcus aureus	Staphylococcaceae	2	1	11,967	2,753,071	13.3	98.6 / 0.3 / 0.2 / 1.1	1,482x	No	Yes
14636	Staphylococcus aureus	Staphylococcaceae	1	0	22,594	2,864,322	12.7	98.3 / 0.3 / 0.2 / 1.4	475x	Yes	No
14637	Staphylococcus aureus	Staphylococcaceae	2	1	25,260	2,984,463	12.5	98.9 / 0.2 / 0.4 / 0.9	118x	Yes	No
14245	Staphylococcus aureus	Staphylococcaceae	3	2	26,424	2,958,199	12.3	99.1 / 0.3 / 0.2 / 0.8	169x	Yes	No
13950	Staphylococcus cohnii/delenis	Staphylococcaceae	1	0	8,210	2,667,978	12.3	97.4 / 0.7 / 0.0 / 1.7	1,729x	No	No
11192	Legionella pneumophila	Legionellaceae	1	0	9,070	3,406,409	12.8	73.4 / 6.0 / 0.5 / 20.7	552x	No	Yes
13726	Fusobacterium necrophorum	Fusobacteriaceae	1	0	3,764	2,674,243	12.8	90.2 / 4.5 / 0.4 / 5.3	243x	No	No
10112	Mycoplasma orale	Mycoplasmataceae	1	0	19,194	732,496	12.9	98.7 / 0.9 / 0.0 / 0.4	2,759x	Yes	Yes
10121	Mycoplasma hyorhinis	Mycoplasmataceae	1	0	17,801	844,167	12.4	97.4 / 0.4 / 0.4 / 2.1	2,905x	Yes	Yes
14143	Streptococcus pneumoniae	Streptococcaceae	1	0	15,850	2,110,936	12.2	98.0 / 0.4 / 0.3 / 1.4	646x	No	No
4518	Streptomyces sp.	Streptomycetaceae	-	-	23,225	-	12.0	-	27.5x	No	No

Summary of Raw Read, Assembly, and Polishing metrics for each NCTC strain

RESULTS

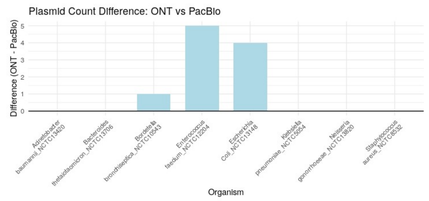
-All 25 assemblies produced circularised chromosomes and plasmids

-*BUSCO* analysis (see central figure) showed high completeness scores (>95%) in 21/25 of the assemblies, with polishing steps leading to noticeable gains in gene recovery

-19/25 *Autocycler* assemblies were resolved automatically, while 6/25 required manual intervention

-We identified and assembled plasmids in 13/25 strains, including small plasmids as short as 1,819 bp, and with greater plasmid recovery in ONT vs. PacBio reads (see figure below)

-An *AMRFinder+* analysis of seven Gram-positive strains confirmed the presence of expected antimicrobial resistance genes.



Detection of plasmids ONT vs PacBio