Using Nanopore sequencing data for bacterial core genome MLST analysis

Oscar Aspelin[1], Fredrik Dyrkell[1]

[1]1928 Diagnostics AB – Gothenburg (Sweden)

Background

The rapid development within Oxford Nanopore Technologies (Nanopore) sequencing has enabled new kinds of genomic analyses to be performed with unprecedented speed. We have previously described how Nanopore data can be used in SNP analysis for bacterial outbreak analysis. Here, we investigate the feasibility of using Nanopore sequencing data in bacterial core genome MLST (cgMLST) analysis.

In this project, the fully automated and cloud based 1928 platform (www.1928diagnostics.com) was used for bioinformatic cgMLST analysis (n=1770 genes) of Illumina sequenced *Enterococcus faecalis* and used as a baseline for comparison of corresponding Nanopore sequenced samples.

Methods

Nineteen double sequenced *Enterococcus faecalis* (Illumina and Nanopore, PRJNA774104) were included. Illumina samples were analyzed in the 1928 platform (analysis includes quality control, cgMLST, AMR prediction and virulence factors) and cgMLST profiles were extracted. Novel alleles per sample ranged from 18 to 1438 (mean 693). Novels are defined as alleles not previously present in the allele database.

Each Nanopore sample was de-novo assembled with Flye and alleles were called using BLAST against the cgMLST database. Two different databases were used – with and without first populating with Illumina novels. The cgMLST profiles were compared between the sequencing platforms.

To evaluate novel prediction with Nanopore data, a FASTA file, containing the best BLAST gene hits from the non-populated database, was created for each sample. SNPs were called with Clair3, applied using BCFtools consensus and the final sequences were compared to the Illumina result.

Results

The results are shown in Table 1. BLAST works well when no novels are expected. Identifying true novel sequences from only Nanopore data proved to be difficult (total accuracy 20%, span 11%–56%). A subset of the cgMLST cluster (Illumina and Nanopore) is shown in Figure 1.

Table 1. Nanopore cgMLST results.

	BLAST (non-populated)	**BLAST** (populated)	**BLAST + Clair3** (non-populated)
Accuracy (novels)	0.0%	99.7%	19.8%
Accuracy (total)	59.9%	99.1%	67.9%

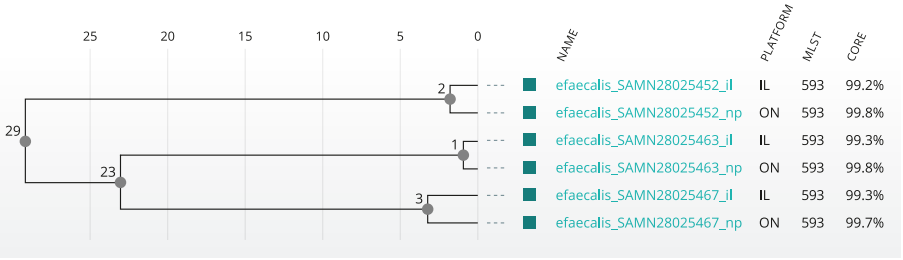

Figure 1. A subset of the cgMLST cluster.

Conclusions

Core genome MLST analysis using Nanopore data is feasible for samples where novels are expected to be few. In a clinical setting, Nanopore data could be used to investigate bacterial outbreaks in those cases where closely related Illumina samples have previously been analyzed. However, more reliable methods are still needed to identify true novel sequences from only Nanopore data.

Conflict of interest statement

O.A. and F.D. are employees of 1928 Diagnostics AB.

Contact

oscar.aspelin@1928diagnostics.com