Using Nanopore sequencing data for bacterial core SNP analysis

Oscar Aspelin[1], Fredrik Dyrkell[1]

[1]1928 Diagnostics AB – Gothenburg (Sweden)

Background

Nanopore sequencing has shown great potential for rapid sequencing of bacterial genomes, substantially reducing the sample-to-result turnaround time at the expense of higher error rates. If Nanopore sequencing would provide enough resolution, it could potentially replace conventional shortread sequencing and revolutionize the field of bacterial outbreak analysis. In this project, a core SNP-pipeline for bacterial Nanopore sequencing data was developed in the 1928 platform and evaluated using a double sequenced dataset.

Materials

The dataset, described previously (ENA project PRJEB38543, Hallgren M et al., bioRxiv 2020) in which 12 E. coli isolates were sequenced with both Illumina and Nanopore, was used for evaluation. The Illumina data was uploaded and analysed in the 1928 shortread SNP-pipeline (Werner A et al., Plos One 2020).

Nanopore samples were uploaded to the 1928 platform and checked for quality and a minimum sequence depth of 40x. The reads were mapped against the GenBank sequence CP024801.1 using minimap2 (v2.20-r1061). Variants were called using Clair3 (v0.1-r4). High quality variants (AF >= 0.7, QUAL >= 5) were kept whilst low quality variant positions were masked. DCM-methylation sites were also masked, since these have been shown to introduce bias. Phylogenetic trees were constructed using UPGMA.

Results

The double sequenced samples were clustered together, as well as separately by sequencing platform. Pairwise SNP distances between sequencing platforms were calculated, yielding a mean difference of 6 SNPs (range 0-23). A subset of the double sequenced cluster is shown in figure 1.

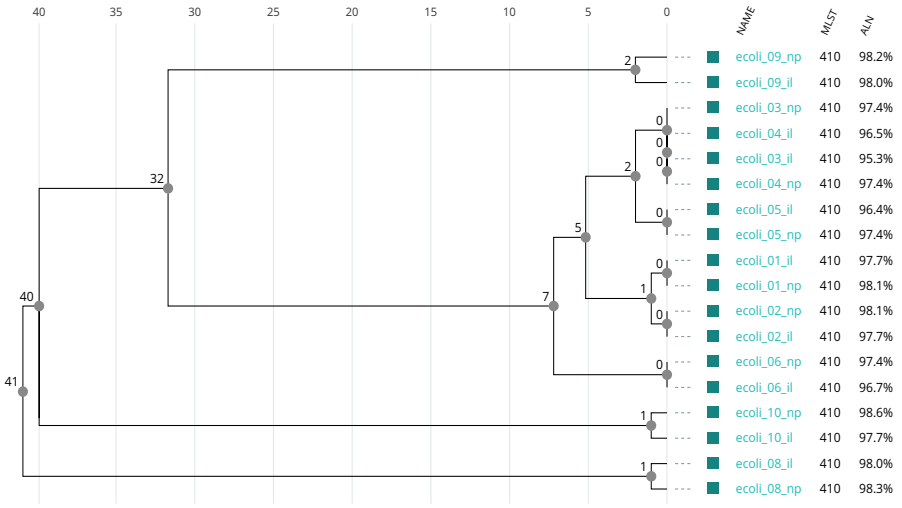

Figure 1. Core SNP tree for a subset of the double sequenced dataset, using CP024801.1 as reference. Average SNP distances were calculated using the UPGMA algorithm.

Conclusions

Nanopore data can be used in bacterial core SNP analysis to generate results that are comparable to Illumina. With further increases in accuracy for chemistry and basecalling, Nanopore sequencing is projected to generate great value in routine epidemiological analysis.

Conflict of interest statement

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

Contact

oscar.aspelin@1928diagnostics.com