

# Användande av Oxford Nanopore-data för bakteriell core SNP-analys

---

Oscar Aspelin<sup>1</sup>, Fredrik Dyrkell<sup>1</sup>

<sup>1</sup>1928 Diagnostics AB - Göteborg (Sverige)

## Introduktion

Sekvensering med Oxford Nanopore har visat stor potential för snabb helgenomsekvensering av bakterier genom att drastiskt korta ner tiden från provtagning till resultat, dock med nackdelen att felfrekvensen i datan är relativt hög.

Om sekvensering med Oxford Nanopore skulle uppvisa tillräcklig hög kvalitet och upplösning, skulle den potentiellt kunna ersätta konventionella sekvenseringsmetoder såsom Illumina och potentiellt revolutionera bakteriell utbrotsanalys. I detta projekt så utvecklades en core SNP-pipeline med stöd för Nanopore-data i 1928-plattformen och utvärderades med hjälp av ett dubbelsekvenserat dataset.

## Metod

Datasetet som användes för utvärdering (Hallgren et al., ENA projekt PRJEB38543) bestod av 12 stycken dubbelsekvenserade (Illumina + Nanopore) *Escherichia coli*.

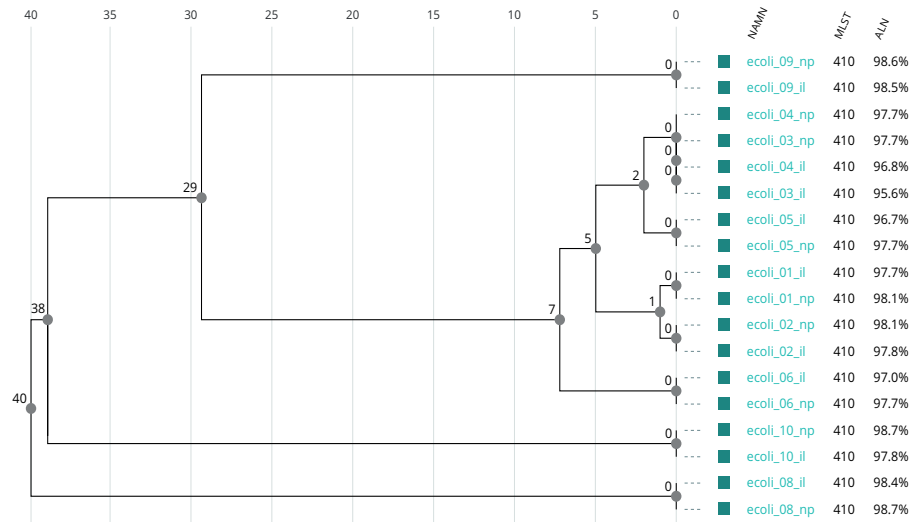
Illumina-datan laddades upp och analyserades i 1928-plattformens core SNP-pipeline för shortread-data (Werner et al.). Nanopore-datan laddades upp till 1928-plattformen och en kvalitetskontroll av datan (minimum 40x sekvenseringsdjup) genomfördes. Sekvenserna mappades mot referensen NZ\_LR595691.1 (RefSeq) med minimap2 (v2.20-r1061) och SNPs identifierades med Clair3 (v0.1-r10).

Endast SNPs med tillräckligt hög kvalitet (AF  $\geq$  0.7, QUAL  $\geq$  5) behölls, medan positioner med SNPs av låg kvalitet maskerades. Positioner med DCM-metyleringsmotiv maskerades också eftersom detta sedan tidigare visats påverka resultatet negativt.

Fylogenetiska träd genererades slutligen med hjälp av UPGMA.

## Resultat

De dubbelsekvenserade proverna klustrades tillsammans och även separat per plattform. För samtliga prover beräknades parvisa SNP-avstånd mellan de två sekvenseringsplattformarna, vilket resulterade i ett genomsnittligt avstånd på 4 SNPs (median 0 SNPs, spann 0-17 SNPs). En del av SNP-trädet illustreras i figur 1.



**Figur 1.** Core SNP-träd för en del av det dubbelsekvenserade datasetet där NZ\_LR595691.1 användes som referens. Genomsnittliga SNP-avstånd beräknades med hjälp av UPGMA.

## Konklusion

Nanopore-data kan användas i bakteriell core SNP-analys för att generera resultat som är jämförbara med Illumina. Med ytterligare förbättringar inom sekvenseringskemi och basecalling förutspås sekvensering med Oxford Nanopore bli en stor tillgång inom rutinmässig epidemiologisk analys.

## Referenser

Hallgren et al., MINTyper: A method for generating phylogenetic distance matrices with long read sequencing data, bioRxiv 2020 (preprint).  
Werner et al., Whole genome sequencing of *Clostridioides difficile* PCR ribotype 046 suggests transmission between pigs and humans, PLoS ONE 2021; 15(12): 1-13.

## Kontakt

oscar.aspelin@1928diagnostics.com