An update on minSNPs: working with Nanopore sequence data

Nanopore sequencing technology has improved and is being used more widely. In contrast to short-read sequencing technology, Nanopore sequencing devices are more portable and generate data in real-time. The technology and the large volume of publicly available genomic data allow for the development of optimised tools for genomic surveillance of microbial pathogens. We previously presented a tool – minSNPs, that can identify genomic markers for species with extensive known genomic diversity and identified optimized markers for *Staphylococcus aureus*. We extended minSNPs to make use of the previously identified surveillance markers with Nanopore sequencer data.

Abstract

Nanopore sequencing technology has improved and is increasingly used in public health genomic surveillance. This is because it is portable and it has the ability to generate long-read sequence data in real-time. This, coupled with the large volume of publicly available sequencing data, allow for the development of optimised tools for genomic surveillance of microbial pathogens. We previously presented minSNPs, an R package to mine resolution-optimised SNP sets from genome-wide orthologous Single Nucleotide Polymorphism (SNP) matrix. Using minSNPs and a genomic pipeline, we mined 200 highly optimised 5-SNPs sets to genotype the clonal complexes of *Staphylococcus aureus* with publicly available geographically diverse short-read sequencing data from 4 different initiatives. We have extended this work by developing an SNP-based genotyping method that uses Nanopore sequence data. Besides diagnosing the clonal complex of the samples, we have also created a Kmer-based method to quickly test the presence/absence of antibiotic-resistant and virulence genes. The method has been tested by simulated long-read sequence data and by re-sequencing isolates of known clonal complexes. The preliminary result suggests that the method works well. minSNPs is an efficient and flexible tool for mining resolution-optimised sets of SNP markers. It has been successfully used to analyse a large SNP matrix derived from multiple *S. aureus* BioProjects, showing the potential to identify globally applicable microbial surveillance markers for biological entities for which there is extensive known genomic diversity. Extending minSNPs to make use of Nanopore sequencer provided a simple and quick way to make use of the surveillance markers.