Assignment 1

➢ abstract :

Nanopore sequencing is one of the most exciting new technologies that undergo dynamic development. With its development, a growing number of analytical tools are becoming available for researchers. To help them better navigate this ever changing field, we discuss a range of software available to analyze sequences obtained using nanopore technology.

➢ Introduction :

Beginning of twenty-first century witnessed dynamic development of sequencing technology. First, so-called "Next Generation Sequencing" brought increased sequencing yield and decline of sequencing cost. However, it was with a certain price, namely the length of the reads, which is much shorter than in traditional Sanger sequencing. As a result, we are flooded with a number of genomes (as of May 30, 2019, there are over 200,000 prokaryotic and almost 9000 eukaryotic genomes deposited at the NCBI database). However, most of these genomes are in the so-called draft form. It means that chromosomes are presented in rather small pieces for which order and orientation on a chromosome is unknown.

Moreover, gene annotation in these genomes is quite poor or does not exist at all.

Understandably, the emergence of the third generation of sequencing technology that enables a single molecule long reads was met with a great excitement within the community. There are currently two competing products on the market, namely single molecule, real-time sequencing by PacBio [1] and nanopore sequencing by Oxford Nanopore Technologies (ONT). The latter is especially exciting technology thanks to its portability and very low initial cost of hardware [3], which brings the sequencing "democratization" one step closer. Initially, ONT did not provide any analytical tools for the sequences obtained by nanopore sequencing. Moreover, the sequences delivered by the native base caller were in a peculiar format (FAST5) that none of the existing at that time software could handle. As a matter of fact, the very first tool developed outside of ONT was a toolkit called poretools.Nevertheless, long reads with relatively low-sequencing accuracy require different computational approaches than short highly accurate reads. Consequently, many algorithms and analytical tools have been developed to aid nanopore sequences, some of them are specific to this particular method and others are more generic that can be used to any long reads.

Here, we describe a few dozens of tools developed in recent years that are suitable for the nanopore sequences' analyzes. Although the list is by no means comprehensive, we try to cover the whole range of software that reflects diversity of the nanopore sequencing applications. For the reading convenience, we group these programs by tasks they can perform. some more generic programs are listed several times as they can handle different tasks.