

Task 1: Abstract and Introduction summary

- Sequencing machine is divided into short-read sequencing and Long-read sequencing.
- Short-read sequencers such as Illumina's NovaSeq, HiSeq, NextSeq, and MiSeq instruments BGI's MGISEQ and BGISEQ models produce reads of up to 600 bases.
- Long-read sequencing, or third-generation sequencing such as: Pacific Biosciences (PacBio), single molecule real-time (SMRT), sequencing and Oxford Nanopore Technologies (ONT), nanopore sequencing.
- Long-read sequencing technologies routinely generate reads in excess of 10 kb.
- Long-read sequencing offers several advantages over short-read sequencing lie improving de novo assembly.
- Short-read sequencing is cost-effective, accurate, and supported by a wide range of analysis tools.
- As well as a discussion of long-read analysis that require improvements. they introduce the complementary open-source catalogue of long-read analysis tools: long-read-tools.org. The long-read-tools.org database allows users to search and filter tools based on various parameters.
- They further focus on the principles of error correction, base modification detection, and long-read transcriptomics analysis and highlight the challenges.

Task 2: Related work summary

- Chemistry v3 in 2018 introduced the faster polymerase for the Sequel sequencer increased the read lengths to an average 30-kb polymerase read length, SMRT sequencing range from 250 bp to 50 kbp, Nanopore sequencing provides the longest read lengths.

- Nanopore and SMRT long-read sequencing technologies rely on very distinct principles. Nanopore sequencers (MinION, GridION, and PromethION) measure the ionic current fluctuations when single stranded nucleic acids pass through biological nanopores, Different SMRT sequencers (RSII, Sequel, and Sequel II) detect fluorescence events that correspond to the addition of one specific nucleotide by a polymerase tethered to the bottom of a tiny well.

- There are 354 long read analysis tools, developed for nanopore read analyses and 170 tools developed to analyze SMRT data, categorized them into 31 groups based on their functionality. They do not attempt to provide a comprehensive review of tool performance for all long-read applications. Instead, they present the principles and potential pitfalls of long-read data analysis with a focus on some of the main types of downstream analyses.