Nanopore Sequencing Technology and Tools: Computational Analysis of the Current State, Bottlenecks and Future Directions

Damla Senol[1], Jeremie Kim[1], Saugata Ghose[1], Can Alkan[2] and Onur Mutlu[3,1]

[1] Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA, USA
[2] Department of Computer Engineering, Bilkent University, Bilkent, Ankara, Turkey
[3] Department of Computer Science, Systems Group, ETH Zürich, Zürich, Switzerland

Nanopore sequencing [1-4], a promising single-molecule DNA sequencing technology, exhibits many attractive qualities and, in time, could potentially surpass current sequencing technologies. Nanopore sequencing promises higher throughput, lower cost, and increased read length, and it does not require a prior amplification step. Nanopore sequencers rely solely on the electrochemical structure of the different nucleotides for identification and measure the change in the ionic current as long strands of DNA (ssDNA) pass through the nano-scale protein pores.

Biological nanopores for DNA sequencing was first proposed in the 1990s, but it was only just recently made commercially available in May 2014 by Oxford Nanopore Technologies (ONT). The first commercial nanopore sequencing device, MinION, is an inexpensive, pocket-sized, portable, high-throughput sequencing apparatus that produces real-time data. These properties enable new potential applications of genome sequencing, such as rapid surveillance of Ebola, Zika or other epidemics, near-patient testing, and other applications that require real-time data analysis. In addition, this technology is capable of generating very long reads (~50,000bp) with minimal sample preparation. Despite all these advantageous characteristics, it has one major drawback: high error rates. In order to provide higher accuracy and higher speed, in May 2016, ONT released a new version of MinION with a new nanopore chemistry called R9, which replaced the previous version R7. Although R9 chemistry improves the data accuracy, the tools used for nanopore sequence analysis are of critical importance as they should overcome the high error rates of the technology.

Our goal in this work is to comprehensively analyze tools for nanopore sequence analysis, with a focus on understanding the advantages, disadvantages, and bottlenecks of the various tools. To this end, we rigorously examine multiple steps in the nanopore genome analysis pipeline. The first step, *basecalling*, translates the raw signal output of MinION into nucleotides to generate DNA sequences. Currently, *Metrichor* [5] is the cloud-based basecaller of ONT; and also *Nanocall* [6] and *Nanonet* [7] are the publicly available nanopore basecallers. The second step performs genome assembly with assemblers for noisy long reads. Using only the basecalled DNA reads, assemblers generate longer contiguous fragments called *draft assemblies*. Currently, *Canu* [8] and *Miniasm* [9] are the commonly used long-read assemblers. The third step, *read mapping*, maps the basecalled DNA reads to the draft assembly. Currently, *BWA-MEM* [10] is the commonly used long-read mappers. After this step, an improved consensus sequence is generated from the draft assembly, with the help of aligned reads from the previous step, with *Nanopolish* [11] or *Racon* [12], and a complete whole genome is obtained.

We analyze the aforementioned nanopore sequencing tools in terms of their speed and accuracy, with the goals of determining their bottlenecks and finding improvements to these tools. We also discuss potential future works in nanopore basecallers and assemblers, to take better advantage of nanopore sequencing and to overcome its current disadvantage of high error rates.

References:

[1] Branton, Daniel, et al. "The potential and challenges of nanopore sequencing." Nature biotechnology 26.10 (2008): 1146-1153.

[2] Quick, Joshua, Aaron R. Quinlan, and Nicholas J. Loman. "A reference bacterial genome dataset generated on the MinION™ portable single-molecule nanopore sequencer." Gigascience 3.1 (2014): 1.

[3] Laver, T., et al. "Assessing the performance of the Oxford Nanopore Technologies MinION." Biomolecular detection and quantification 3 (2015): 1-8.

[4] Ip, Camilla LC, et al. "MinION Analysis and Reference Consortium: Phase 1 data release and analysis." F1000Research 4 (2015).

[5] Metrichor. [https://nanoporetech.com/products/metrichor]

[6] David, Matei, et al. "Nanocall: An Open Source Basecaller for Oxford Nanopore Sequencing Data." bioRxiv (2016): 046086.

[7] Nanonet. [https://github.com/nanoporetech/nanonet]

[8] Berlin, Konstantin, et al. "Assembling large genomes with single-molecule sequencing and locality-sensitive hashing." Nature biotechnology 33.6 (2015): 623-630.

[9] Li, Heng. "Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences." Bioinformatics (2016): btw152.

[10] Li, Heng. "Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM." arXiv preprint arXiv:1303.3997 (2013).

[11] Loman, Nicholas J., Joshua Quick, and Jared T. Simpson. "A complete bacterial genome assembled de novo using only nanopore sequencing data." Nature methods (2015).

[12] Vaser, Robert, et al. "Fast and accurate de novo genome assembly from long uncorrected reads." bioRxiv (2016): 068122.