An educational guide for nanopore sequencing in the classroom

Abstract And Introduction Summary

- Abstract :

1-The last decade has witnessed a remarkable increase in our ability to measure genetic information. Advancements of sequencing technologies are challenging the existing methods of data storage and analysis.

2-Future generations of biologists must be more computationally aware and capable. This means they should be trained to give them the computational skills to keep pace with technological developments.

- Introduction :

1-what defines a biologist? In short, a biologist is a person who studies life and living organisms.

2-Biology covers diverse topics such as molecular biology, structural biology, ecology, evolution, genetics, microbiology, immunology, and biotechnology. Importantly, most (if not all) of these topics have undergone incredible progress due to rapid discoveries and technological advances [1,2]. As such, a modern biologist has the inevitable tasks of adapting to rapid change and mastering new knowledge and technology.

3-One of the most important revolutions in the field of biology was caused by the development of next-generation sequencing (NGS) technologies. Using massively parallel processing of samples, NGS dramatically reduces sequencing time and costs, enabling the sequencing of entire genomes.

4-Currently, genome sequencing and analysis have become a crucial component in biology, as evidenced by recent scientific breakthroughs [3,4] and by the exponential increase of reported genomes on GenBank (e.g., from 30,000 sequenced prokaryotic genomes in 2014 [5] to 183,000 in 2018.

5-not only do biologists need to adapt and learn how to use these emerging technologies, they also need to learn how to mine the ever-growing mountain of genomic information they generate, which requires bioinformatics skills. Now, the question is how do we train this generation of biologists so that they have the required computational skills?

-Related works:

Twenty-four groups of 4 students (96 total) prepared their own DNA libraries of various
single-isolate bacterial, bacteriophage, and metagenomic samples in the classroom. Number of groups and their size were determined to allow for sufficient supervision within the available lab space. If possible, smaller groups are preferable to increase the hands-on time of each student
We would like to emphasize the benefits of having multiple groups working on different
related samples (e.g., each barcode represents a similar but different microbial isolate). This allows groups to initiate discussions about differences in their own findings.
1-Karas M, Hillenkamp F. Laser desorption ionization of proteins with molecular masses exceeding 10000 daltons.

Anal Chem. 1988; 60: 2299–301.
https://doi.org/10.1021/ac00171a028 PMID:3239801

2-Norton ME. Noninvasive prenatal testing to analyze the fetal genome. Proc Natl Acad Sci U S A. 2016;
113: 14173–14175. https://doi.org/10.1073/pnas.1617112113
PMID: 27911833

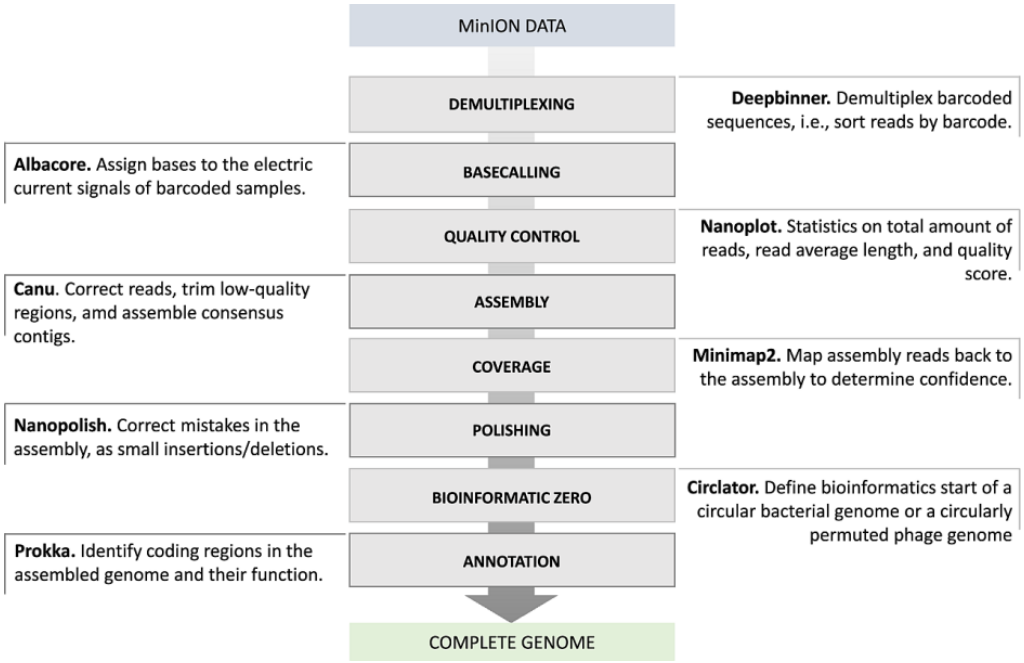
3- Kim D, Song L, Breitwieser FP, Salzberg SL. Centrifuge: rapid and sensitive classification of metagenomic sequences. Genome Res. 2016; 26: 1721–1729.
https://doi.org/10.1101/gr.210641.116 PMID: 27852649

-Methodology:

1-The challenge set for students in our course was to identify and discover novel phages from environmental samples and to reconstruct complete genomes from single-isolate and metagenomics samples.

2-The DNA libraries were prepared using the rapid barcoding kit (SQK-RBK004), which has fewer steps than other available kits and thus allows the procedure to be completed within the 3-hour timeframe of the class. For longer sessions, the ligation sequencing kit (SQK-LSK109) could be used, increasing the robustness and throughput of the experiment. Both kits allow for barcoding of multiple genomic DNA samples. Samples were prepared individually by each group and then barcoded and pooled together at different proportions depending on the success of each group. When sequencing runs failed, the student was supplied with previously generated backup data.

3-After running DNA samples in MinION, students performed quality control of their data and then assembled the genomes. As we focused on teaching technical concepts of bioinformatics, we provided a computational guide.



4-Once data processing was completed, students pursued a variety of research questions, such as investigating the genomic composition of their bacterial sample as well as the population composition of their metagenomics sample.

5-students found that their assembly had little overlap with the reference, prompting discussions about the novelty of the genetic content in their phage.

6-This process stimulated discussion about a number of course-related topics: (1) limitations of k-mer-based tools (e.g., k-mers are not always unique to individual species), (2) biases when comparing against a reference data set (e.g., you can only classify what you have previously observed), (3) understanding bacteriophage biology (e.g., phages can integrate their DNA in a bacterial host; therefore, sequences that are labeled as "bacteria" may actually correspond to integrated phage DNA), and (4) understanding whether long-read sequencing is advantageous to the scientific question addressed.

Results:

1-We experienced increased interest and engagement in our course from both the instructors and the students. Students were much more interested in the course content because they could assume scientific responsibility and ownership.

2-the experiments were chosen such that they contribute to ongoing research in the lab. As a result, we generated several follow-up project ideas, one of which resulted in a master's thesis on heterogeneity of bacteriophage genomes detected by nanopore sequencing, as well as a tripling of the number of undergraduate lab-rotations in the area of bioinformatics.

3-Naturally, many of the assignments, including interpretation and comparison of a genome assembly from single bacterial isolates to that of viral samples, were open-ended and initially challenged the students. However, the experience gave them a more realistic impression of academic research and foundational skills to help them in their future career as modern biologists.

4-Future editions of such an integrated course could consider even developing the student ownership further by explaining the "problem" and asking students to design the DNA sequencing experiments given the boundaries of the reagents available. With adequate supervision and coaching to include proper controls and experiments, this could lead to even greater collaboration and ownership by the students.

Name : Omnya Abd El Fattah Ali

3rd year Bioinformatics Department