

Assignment 1: Abstract and Introduction summary

- 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. Currently, genome sequencing and analysis have become a crucial component in biology .
- The problem is the exponential increase of reported genomes on GenBank (e.g., from 30,000 sequenced prokaryotic genomes in 2014 to 183,000 in 2018) , " a 6-fold increase in only 4 years " .
- Thus, You must choose a more effective tool and method of obtaining to make the generated signals more similar to the real ones, we added a low-pass filter to post-process the pore model signals.
- To solve this problem we update all three modules from DS1.0. to DS1.5
- As for the sequence generator, we updated the sample read length distribution to reflect the newest real reads' features.
- added one more pore model, the context-independent pore model, which is much faster than the previous context-dependent one.
- added a low-pass filter to post-process the pore model signals.
- added the support for the newest official basecaller, Guppy, which can support both GPU and CPU.
- multiple optimizations, related to multiprocessing control, memory and storage management, have been implemented to make DS1.5 a much more amenable and lighter simulator than DS1.0.
- contents of paper will be Workflow , implementation , Performances , Conclusions and discussion of this update.
- Although the first version of DeepSimulator (DS1.0) has been recognized and used by a number of users " <https://github.com/lykaust15/DeepSimulator> " .
- there is still a large room for improvement. For example, though the final simulated reads have almost the same error distribution as the real reads, for some sequences, the divergence between the simulated raw signals and the real signals can be large, which can be inconvenient for the users who care about the signal outputs. In addition, the Nanopore technology has evolved greatly since DS1.0 was released. It is thus necessary to update DS significantly to accommodate those changes, such as the extended reads' length. Here, we present a substantially updated version of DS.