Conference Abstract

Decona: From demultiplexing to consensus for Nanopore amplicon data

Saskia Oosterbroek[‡], Karlijn Doorenspleet[‡], Reindert Nijland[‡], Lara Jansen[‡]

‡ Wageningen University, Wageningen, Netherlands

Abstract

Sequencing of long amplicons is one of the major benefits of Nanopore technologies, as it allows for reads much longer than Illumina. One of the major challenges for the analysis of these long Nanopore reads is the relatively high error rate. Sequencing errors are generally corrected by consensus generation and polishing. This is still a challenge for mixed samples such as metabarcoding environmental DNA, bulk DNA, mixed amplicon PCR's and contaminated samples because sequence data would have to be clustered before consensus generation.

To this end, we developed Decona (https://github.com/Saskia-Oosterbroek/decona), a command line tool that creates consensus sequences from mixed (metabarcoding) samples using a single command. Decona uses the CD-hit algorithm to cluster reads after demultiplexing (qcat) and filtering (NanoFilt). The sequences in each cluster are subsequently aligned (Minimap2), consensus sequences are generated (Racon) and finally polished (Medaka). Variant calling of the clusters (Medaka) is optional. With the integration of the BLAST+ application Decona does not only generate consensus sequences but also produces BLAST output if desired. The program can be used on a laptop computer making it suitable for use under field conditions.

Amplicon data ranging from 300-7500 nucleotides was successfully processed by Decona, creating consensus sequences reaching over 99,9% read identity. This included fish datasets (environmental DNA from filtered water) from a curated aquarium, vertebrate

datasets that were contaminated with human sequences and separating sponge sequences from their countless microbial symbionts.

Decona considerably simplifies and speeds up post sequencing processes, providing consensus sequences and BLAST output through a single command. Classifying consensus sequences instead of raw sequences improves classification accuracy and drastically decreases the amount of sequences that need to be classified. Overall it is a user friendly option for researchers with limited knowledge of script based data processing.

Keywords

consensus, amplicon, barcoding, bioinformatics, nanopore, long reads

Presenting author

Saskia Oosterbroek

Presented at

1st DNAQUA International Conference (March 9-11, 2021)