poreTally: run and publish *de novo* Nanopore assembler benchmarks - supplementary information

Carlos de Lannoy*1, Judith Risse2, and Dick de Ridder1

1Bioinformatics Group, Wageningen University, 6700AP, Wageningen, The Netherlands
2Netherlands Institute of Ecology (NIOO-KNAW), 6708PB, Wageningen, The Netherlands
*Corresponding author, e-mail: carlos.delannoy@wur.nl

A The collective benchmark

poreTally was conceived to facilitate standardization of *de novo* nanopore assembly pipeline benchmarks and fast publication on the user’s personal Github/Gitlab page. Although this method of publication allows for fast and independent dissemination of benchmark results, it may result in a fragmented view of overall assembly pipeline performance.

![Figure A.1: Schematic representation of the poreTally collective benchmark submission process. If the user consents to submitting benchmark results, an empty repository will be forked to their account (1) after which the benchmark results are pushed to the fork (2) and a pull request is submitted (3). Periodically, we will accept made pull requests, summarize the results and publish the aggregated report in another repository (4).](image-url)
Therefore we also included the option to submit results to a collective benchmark effort. Benchmarks submitted to this effort will be summarized and published online, thus giving a more unified indication of de novo assembly pipeline performance.

If a user chooses to support the collective benchmark, a dedicated Github repository will automatically be cloned to their account, after which results are pushed to that repository and a pull request is submitted (Figure A.1). The same data that would appear on a regular poreTally repository publication is submitted to the collective benchmark, thus no actual read or assembly sequences are shared. The submission process through Github allows for scalable and transparent data collection. Through the Github account used to submit a benchmark, its source can be traced back to the user, thus allowing us to duly credit submitters for their contribution and providing credibility to the collective benchmark.

Periodically we will accept pull requests and present a summary of submitted benchmarks in a separate repository: github.com/cvdelannoy/poreTallyCommunity. Given sufficient and sufficiently diverse submissions, we may eventually be able to better characterize factors that influence assembler performance for often-used pipelines, such as sequencing hardware, read set quality and species. As nanopore sequencers – and the MinION in particular – uniquely allow any individual researcher to complete the genome sequencing process for any species from DNA sample to assembly, we envision that the collective benchmark will serve both the growing group of novel users and veterans alike when selecting the appropriate assembly pipeline for their data set and biological source.