

Illumina and Nanopore Combo Assembly and Annotation (Prokaryotic)

Porechop¹ was used to trim residual adapter sequence from the Oxford Nanopore Technology (ONT) reads that may have been missed during basecalling and demultiplexing. *De novo* genome assemblies were generated from the Oxford Nanopore Technologies (ONT) read data with Flye² under the nano-hq (ONT high-quality reads) model. Additional Flye options initiate the assembly by first using reads longer than an estimated N50 based on a genome size of 6Mbp. Subsequent polishing used the Illumina read data with Pilon³ under default parameters. To reduce erroneous assembly artifacts caused by low quality nanopore reads, long read contigs with an average short read coverage of 15x or less were removed from the assembly.

Assembled contigs were evaluated for circularization via circulator⁴ using the ONT long reads. Assembly annotation was then performed with Bakta⁵ using the Bakta v5 database. Finally, assembly statistics were recorded with QUAST⁶ which can be found in the “*Assembly Metrics.tsv*” file. The assembly and annotation files are “*SampleName.fasta*”, “*SampleName.gff*”, and “*SampleName.gbk*” respectively. A summary of the assembled contigs can be found in each sample’s respective ‘*Contig Summary.tsv*’ file.

Tool	Version	Parameters
porechop	0.2.4	default parameters
flye	2.9.2	--asm-coverage 50 --genome-size 6000000 --nano-hq
bowtie2	2.5.1	--very-sensitive
pilon	1.24	default parameters
circulator	1.5.5	all; 6-hour timeout
bakta	1.8.1	default parameters; db version 5.0
quast	5.2.0	default parameters

¹ An open source software for the QC and adapter trimming of ONT technologies.

<https://github.com/rrwick/Porechop>

² Mikhail Kolmogorov, Jeffrey Yuan, Yu Lin and Pavel Pevzner. Assembly of Long Error-Prone Reads Using Repeat Graphs. *Nature Biotechnology* (2019)

³ Walker, Bruce J et al. “Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement.” *PloS one* vol. 9,11 e112963. 19 Nov. 2014, doi:10.1371/journal.pone.0112963

⁴ Hunt, M., Silva, N.D., Otto, T.D. et al. Circlator: automated circularization of genome assemblies using long sequencing reads. *Genome Biol* 16, 294 (2015).

⁵ Schwengers O., Jelonek L., Dieckmann M. A., Beyvers S., Blom J., Goesmann A. (2021). Bakta: rapid and standardized annotation of bacterial genomes via alignment-free sequence identification. *Microbial Genomics*, 7(11).

⁶ Alexey Gurevich, Vladislav Saveliev, Nikolay Vyahhi and Glenn Tesler, QUAST: quality assessment tool for genome assemblies, *Bioinformatics* (2013) 29 (8): 1072-1075.