PacBio Assembly and Annotation (Eukaryote)

Assemblies were generated via Flye[1] with the --pacbio-hifi read input flag, a stringent read error cutoff, and additional options to initiate the assembly by first using reads longer than an estimated N50 based on a genome size of 12Mbp. Assemblies were annotated using Funannotate[2]. The output assembly and annotation files are *"SampleName.fasta", "SampleName.gff",* and *"SampleName.gbk"*. Finally, QUAST[3] was used to generate assembly metrics which are reported in the *"Assembly metrics.tsv"* file.

Kinetics data from the PacBio HiFi reads were used to determine the methylation states of the assembled genomes. Kinetics data was collected from the HiFi reads using ccs-kinetics-bystrandify, an algorithm in the pbbam[4] suite. Exported kinetics data were mapped to the generated assembly using pbmm2 from the SMRT Tools suite in SMRT Link 11.0[5]. The ipdSummary algorithm putatively identified methylation states from the kinetics data as either undetermined, m4C or m6A. Methylation calls are reported in the file *"all_base_modifications.bigwig"* which can be viewed in a genome browser and *"all_base_modifications.gff3"* which can be loaded as an annotation file. The all modifications gff3 file was subset into files for m4C and m6A gff3 files, respectively.

Tool	Version	Parameters
flye	2.9.2	--asm-coverage 50 --genome-size 12000000 --pacbio-hifi --read-error 0.02
funannotate	1.8.15	interproscan and eggnog mapper enabled
quast	5.2.0	default parameters
pbbam	2.1.0	ccs-kinetics-bystrandif (default parameters)
smrtlink	12.0.0.177059	pbmm2 (default parameters) ipdSummary (--identify m6A,m4C)

[1] Mikhail Kolmogorov, Jeffrey Yuan, Yu Lin and Pavel Pevzner. Assembly of Long Error-Prone Reads Using Repeat Graphs. Nature Biotechnology (2019)
[2] Palmer J (2017) Funannotate: Fungal genome annotation scripts. https://github.com/nextgenusfs/funannotate
[3] Alexey Gurevich, Vladislav Saveliev, Nikolay Vyahhi and Glenn Tesler, QUAST: quality assessment tool for genome assemblies, Bioinformatics (2013) 29 (8): 1072-1075.
[4] PacificBiosciences/pbbam: Pacbio BAM C++ library. GitHub. https://github.com/PacificBiosciences/pbbam
[5] SMRT® Tools (v11.0). https://www.pacb.com/support/software-downloads/

91 43rd Street
Suite 250
Pittsburgh, PA 15201
878-227-4915