Text

Description automatically generated with medium confidence

**Project Report**

Quality control and adapter trimming was performed with bcl-convert [1]. Read count summaries are provided in the ‘*Read Counts.*tsv’ file. Paired end reads were processed using Centrifuge’s [2] version of the NCBI ‘*nt*’ database to obtain the relative abundance of community members. Relative abundances can be viewed in the ‘*Sample.*tsv files. Relative abundances were processed and converted into Kronagrams using KronaTools [3]. Kronagrams are available in the ‘*Sample.html*’ files. Paired end reads were then assembled using megahit [4]. Assemblies are available in the ‘*Sample.fasta*’ files. Single genome reconstruction and binning was performed using metabat2 [5]. Single genomes can be found in their respective bin folders in the ‘*bin-#.fasta*’ files. To check for contamination, centrifuge and KronaTools were used to assess the relative abundance of contigs within the binned, single genome. Relative abundances can be viewed in the ‘*bin-#.tsv*’ files. Kronagrams for binned, single genome can be found in their respective bin folders in the ‘*bin-#.html*’ files. The complete metagenome and binned, single genomes were benchmarked using QUAST [6]. The assembly statistics can be found in the *‘Assembly Metrics.tsv*’ file. If a single genome contained any trace of bacteria, archaea, or viruses it was annotated with prokka [7]. Annotated genomes can be found in their respective bin folders under the ‘*bin-#.gbk*. file name.

|  |  |  |
| --- | --- | --- |
| **Tool** | **Version** | **Parameters** |
| bcl-convert | 3.9.3 | Default Parameters |
| centrifuge | 1.0.3 | Default Parameters |
| KronaTools | 2.8 | Default Parameters |
| megahit | 1.2.9 | Default Parameters |
| metabat2 | 2.15 | Default Parameters |
| QUAST | 5.0.2 | Default Parameters |
| prokka | 1.14.5 | Default Parameters |

**References**

[1] bcl-convert: A proprietary Illumina software for the conversion of bcl files to basecalls. <https://support-docs.illumina.com/SW/BCL_Convert/Content/SW/FrontPages/BCL_Convert.htm>

[2] Kim D, Song L, Breitwieser FP, Salzberg SL. Centrifuge: rapid and sensitive classification of metagenomic sequences. Genome Res. 2016;26(12):1721-1729. doi:10.1101/gr.210641.116

[3] Ondov BD, Bergman NH, Phillippy AM. Interactive metagenomic visualization in a Web browser. BMC Bioinformatics. 2011 Sep 30;12:385. doi: 10.1186/1471-2105-12-385. PMID: 21961884; PMCID: PMC3190407.

[4] Li D, Liu CM, Luo R, Sadakane K, Lam TW. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics. 2015 May 15;31(10):1674-6. doi: 10.1093/bioinformatics/btv033. Epub 2015 Jan 20. PMID: 25609793.

[5] Kang DD, Li F, Kirton E, et al. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. PeerJ. 2019;7:e7359. Published 2019 Jul 26. doi:10.7717/peerj.7359

[6] Alexey Gurevich, Vladislav Saveliev, Nikolay Vyahhi and Glenn Tesler. QUAST: quality assessment tool for genome assemblies, Bioinformatics (2013) 29 (8): 1072-1075. doi: 10.1093/bioinformatics/btt086

[7] Torsten Seemann, Prokka: rapid prokaryotic genome annotation, Bioinformatics, Volume 30, Issue 14, 15 July 2014, Pages 2068–2069, https://doi.org/10.1093/bioinformatics/btu153