logo

Output description for SARS-Cov2 Nanopore Artic pipeline

Pipeline overview

The pipeline was created by Artic and follows these steps:

- Guppy v3.4.5+fb1fbfb - High accuracy base calling.
- Gather Merge base called fastq files.
- Porechop v0.3.2rc0 - Sample demultiplexing
- Nanopolish v0.11.3 - Indexing gathered files.
- Main Artic v1.0.0 - Variant calling, mapping and genome consensus

Reference genome: NC_045512.2

Depending on the analysis, we will have some ANALYSIS_IDs. This ANALYSIS_IDs are going to be composed of the date of the analysis, and some identification of the type of analysis.

Preprocessing

Guppy

Guppy from Oxford Nanopore Technologies [1] is a base caller similar to Albacore but can use GPUs for improved basecalling speed. While the two basecallers have coexisted for about a year, ONT has discontinued development on Albacore in favour of the more performant Guppy. We performed the high accuracy base calling with Guppy.

Output directory: `00-guppy`

- `SARS-Cov-2/`
  - Output directory where all the fastq files of tha basecalling are going to be stored.

Note: These files are not sent to the scientiest because they are just RAW basecalled reads.

Gather

Gather is a module from the artic pipeline used to merge all the basecalled fastq files in a unique fastq file and generated summary stats.

Output directory: `01-gather`

- `{run_id}_all.fastq`
  - Fastq file with all the gathered reads that passed the sequencing.
- `{run_id}_all.fastq.demultiplexreport.txt`
  - Txt file with the demultiplexing report if we'd demultiplexed with guppy (We didn't).
- `{run_id}_all.fastq`
  - Fastq file with all the gathered reads.
- `{run_id}_sequencing_summary`
  - Sequencing summary file.

Porechop

Artic uses a customized version of Porechop a tool for finding and removing adapters from Oxford Nanopore reads. Adapters on the ends of reads are trimmed off, and when a read has an adapter in its middle, it is treated as chimeric and chopped into separate reads. Porechop performs thorough alignments to effectively find adapters, even at low sequence identity. Porechop also supports demultiplexing of Nanopore reads that were barcoded with the Native Barcoding Kit, PCR Barcoding Kit or Rapid Barcoding Kit.

Output directory: `02-demultiplex`

- `{run_id}-{barcode}.fastq`
  - Fastq file containing all the reads that were assigned to that barcode.

Nanopolish

Artic uses Nanopolish [2], which is a software package for signal-level analysis of Oxford Nanopore sequencing data. Nanopolish can calculate an improved consensus sequence for a draft genome assembly, detect base modifications, call SNPs and indels with respect to a reference genome and more (see Nanopolish modules, below). Here they used it to index the gathered fastq file.

Output directory: `01-gather`

- `{run_id}_all.fastq.index*`

- Index files created by nanopolish for the gathered fastq.

Variant calling, mapping and genome consensus

Main Artic

Artic pipeline is complete bioinformatics protocol to take the output from the sequencing protocol to consensus genome sequences. Includes basecalling, de-multiplexing, mapping, polishing and consensus generation.

It uses the following programs:

- BWA[3]
- Samtools[4]
- Nanopolish [2]

Output directory: `04-artic-minion`

- `{sample_id}.alignreport.er`
  - Error report from the alignment and amplicon trimming.
- `{sample_id}.alignreport.txt`
  - Report from the alignment and amplicon trimming.
- `{sample_id}.consensus.fasta`
  - Consensus genome.
- `{sample_id}.minion.log`
  - Log file with all the commands runned by Artic.
- `{sample_id}.primertrimmed.sorted.bam`
  - Bam file with the primers trimmed.
- `{sample_id}.primertrimmed.sorted.bam.bai`
  - Index of the Bam file with the primers trimmed.
- `{sample_id}.primertrimmed.vcf`
  - VCF file of the primers trimmed reads.
- `{sample_id}.variants.tab`
  - Table with the variants in all the VCF files.
- `{sample_id}.vcf`
  - VCF file of the reads trimmed in the start of the read.

Final Results

We hace collected the most significant files for you.

Output directory: `RESULTS`

- draft_genomes: this folder contains the draft genomes.
- ordered_contigs: this folder contains the ordered contigs for each sample.
- circos_images: circos images for the reconstructed genomes.
- reads_stats: statistics for mapped reads against host and virus, with coments.

References

1. Wick R.R., Judd L.M., Holt K.E. Performance of neural network basecalling tools for Oxford Nanopore sequencing. Genome Biol. 2019 Jun 24;20(1):129.
2. Loman N.J., Quick J., Simpson J.T. A complete bacterial genome assembled de novo using only nanopore sequencing data. Nat Methods. 2015 Aug;12(8):733-5.
3. Li H., Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics. 2010 Mar 1;26(5):589-95.
4. Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., Durbin R.; 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009 Aug 15;25(16):2078-9.