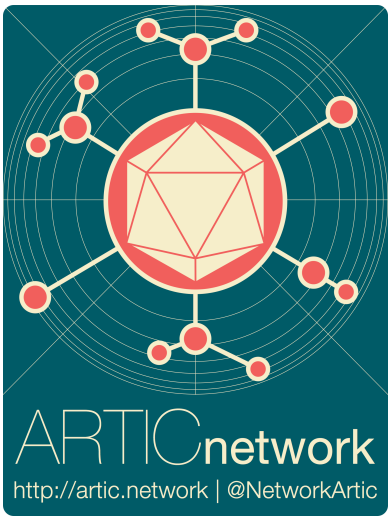

Ebola virus bioinformatics protocol

Nanopore | bioinformatics

Document: ARTIC-EBOV-bioinformaticsSOP-v1.0.0
Creation Date: 2018-05-26
Author: Nick Loman

Overview: A complete bioinformatics protocol to take the output from the sequencing protocol (/ebov/ebov-seq-sop.html) to consensus genome sequences. Includes basecalling, de-multiplexing, mapping, polishing and consensus generation.

This document is part of the Ebola virus Nanopore sequencing protocol package:

http://artic.network/ebov/

Related documents:

Ebola virus Nanopore sequencing protocol:

http://artic.network/ebov/ebov-seq-sop.html (/ebov/ebov-seq-sop.html)

Setting up the laptop computing environment using Conda:

http://artic.network/ebov/ebov-it-setup.html

Phylogenetic analysis and visualization:

http://artic.network/ebov/ebov-phylogenetics-sop.html

Funded by the Wellcome Trust
Collaborators Award 206298/Z/17/Z --- ARTIC network (artic.network)

Preparation

Set up the computing environment as described here in this document: ebov-it-setup (ebov-it-setup.html). This should be done and tested prior to sequencing, particularly if this will be done in an environment without internet access or where this is slow or unreliable. Once this is done, the bioinformatics can be performed largely off-line.

Nanopore Bioinformatics

Activate the ARTIC environment:

```
source activate artic-ebov
```

Basecalling with Albacore (MinION on laptop)

Run the Albacore basecaller on the new MinION run folder:

```
read_fast5_basecaller.py -c r94_450bps_linear.cfg -i /path/to/reads -s run_nam
e -o fastq -t 4 -r --barcoding
```

You need to substitute `/path/to/reads` to the directory where the FAST5 files from your run are. Common locations are:

- Mac: `/Library/MinKNOW/data/reads/run_name`
- Linux: `/var/lib/MinKNOW/data/reads`
- Windows `c:/data/reads`

Consensus sequence generation

Gather up the FASTQ output from Albacore:

```
artic gather --min-length 400 --max-length 700 --prefix run_name output_direct
ory
```

We use a length filter here of between 400 and 700 to remove obviously chimeric reads.

Basecalling using MinIT or GridION

If running on MinIT or GridION and you have used Guppy to basecall through Dogfish, instead you can do:

```
artic gather --guppy --min-length 400 --max-length 700 --prefix run_name /dat
a/basecalled/path/to/reads
```

You will now have a file called: `run_name_all.fastq` and a file called `run_name_sequencing_summary.txt`, as well as individual files for each barcode (if previously demultiplexed).

Demultiplex with Porechop with stringent settings

This stage is obligatory, even if you have already demultiplexed with Albacore, due to significant barcoding misassignments that can confound results:

```
artic demultiplex --threads 4 --prefix run_name_final run_name_all.fastq
```

Now you will have new files called:

```
run_name_final_BC01.fastq
run_name_final_BC02.fastq
run_name_final_BC03.fastq
```

Create the nanopolish index (once per sequencing run, not per sample)

```
nanopolish index -s run_name_sequencing_summary.txt -d /path/to/reads run_name
_all.fastq
```

Again, alter `/path/to/reads` to point to the original location of the FAST5 files.

Run the MinION pipeline

For each barcode you wish to process:

```
artic minion --normalise 200 --threads 4 --scheme-directory primer-schemes --r
ead-file run_name_final_NB01.fastq --nanopolish-read-file run_name_all.fastq Z
aireEbola/V2 samplename
```

Replace `samplename` as appropriate:

Output files

- `samplename.primertrimmed.bam` - BAM file for visualisation after primer-binding site trimming
- `samplename.vcf` - detected variants in VCF format
- `samplename.variants.tab` - detected variants
- `samplename.consensus.fasta` - consensus sequence