Adapted from <https://artic.network/ncov-2019/ncov2019-bioinformatics-sop.html>

V1: 17th Nov 2021

V2: 19th Nov 2021

V3: 11th Feb 2022 (changes in yellow)

**Make a new directory for analysis**

Give your analysis directory a meaningful name, e.g.. analysis/run\_name

mkdir analysis

cd analysis

mkdir run\_name

cd run\_name

**Activate the ARTIC environment:**

All steps should be performed in the artic-rabv conda environment:

source activate artic-rabv

**Basecalling with Guppy**

If you did basecalling with MinKNOW, you can skip this step and go to Demultiplexing.

Run the Guppy basecaller on the new MinION run folder:

For fast mode basecalling:

guppy\_basecaller -c dna\_r9.4.1\_450bps\_fast.cfg -i /path/to/reads -s run\_name -x auto -r

You need to substitute /path/to/reads to the folder where the FAST5 files from your

run are.

Common locations are:

Mac: /Library/MinKNOW/data/run\_name

Linux: /var/lib/MinKNOW/data/run\_name

Windows c:/data/reads

This will create a folder called run\_name with the base-called reads in it.

**Demultiplexing**

For the current version of the ARTIC protocol it is essential to demultiplex using strict

Parameters to ensure barcodes are present at each end of the fragment.

guppy\_barcoder --require\_barcodes\_both\_ends -i run\_name -s output\_directory --barcode\_kits "EXP-NBD104"

The following steps must be done for every barcode

**Read filtering**

This step is performed for each barcode in the run.

We first collect all the FASTQ files (typically stored in files each containing 4000 reads)

into a single file.

To collect and filter the reads for barcode03, we would run:

artic guppyplex --skip-quality-check --min-length 350 --directory output\_directory/barcode03 --prefix run\_name

You will now have a files called: run\_name\_barcode03.fastq

**Run the MinION pipeline: medaka**

For each barcode you wish to process (e.g. run this command 12 times for 12 barcodes),

replacing the file name and sample name as appropriate:

E.g. for barcode03

artic minion --medaka --medaka-model r941\_min\_fast\_g303 --normalise 200 --threads 4 --scheme-directory ~/Github/artic-rabv\_ea/primer\_schemes --read-file run\_name\_barcode02.fastq rabv\_ea/V1 samplename

Replace samplename as appropriate.

Double check the filepath to the artic-rabv folder. If this does not match the location on *your* laptop, modify as necessary.

**Output files**

samplename.rg.primertrimmed.bam - BAM file for visualisation after primer-binding site

trimming

samplename.trimmed.bam - BAM file with the primers left on (used in variant calling)

samplename.merged.vcf - all detected variants in VCF format

samplename.pass.vcf - detected variants in VCF format passing quality filter

samplename.fail.vcf - detected variants in VCF format failing quality filter

samplename.primers.vcf - detected variants falling in primer-binding regions

samplename.consensus.fasta - consensus sequence

To put all the consensus sequences in one file called my\_consensus\_genomes.fasta , run

cat \*.consensus.fasta > my\_consensus\_genomes.fasta

**To visualise genomes in Tablet**

Open a new Terminal window:

tablet

Go to “Open Assembly”

Load the BAM (binary alignment file) as the first file.

Load the reference file (in artic/artic-rabv/primer\_schemes/rabv\_ea/V1/rabv\_ea\_reference.fasta) as the second file.

Select Variants mode in Color Schemes for ease of viewing variants.