Vir-MinION User Manual

Vir-MinION requires few and intuitive arguments to run. The Vir-MinION, once invoked, returns its usage syntax:

```
Usage: VirMinION-Pipe.sh PreProcDataFolder PreProcOutFolder BaseCallConfFile
BarCodeKit NumOfThread Method [--read_based | --ass_based | --clust_based]
```

In details:

- PreProcOutFolder → Folder will contain basecalled files // For example: Superpools1-10_Yunnan_bat_anal_swabs
- BaseCallConfFile → Configuration file to use for basecalling // For Example: dna_r9.4.1_450bps_hac.cfg
- BarCodeKit → BarCode Kit used // For example: EXP-NBD104
- NumOfThread → Number of threads to use
- Method → Method to use. It can be --read_based | --ass_based | --clust_based

The two relevant information related to your experiment, BarCallConfFife and BarCodeKit, can be obtained from the kit you used for the experiment.

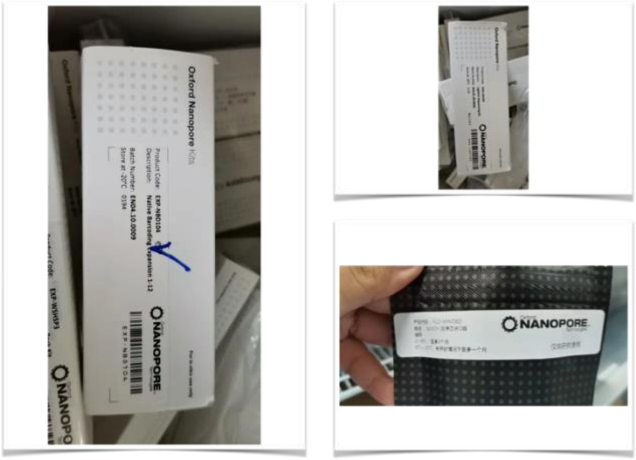
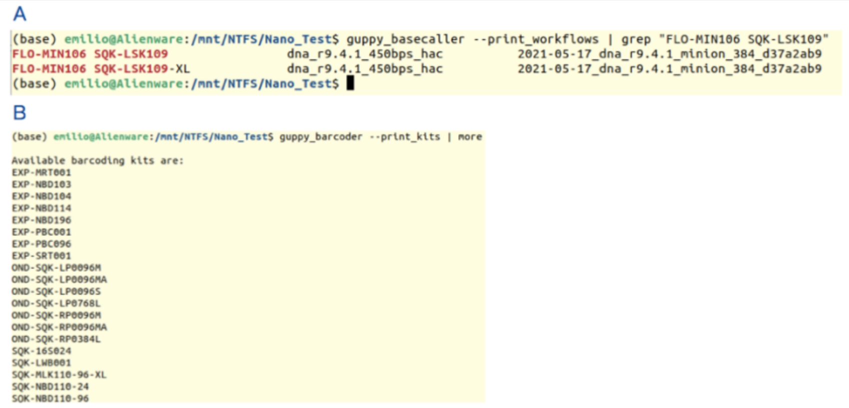


Figure 1 Lab Boxes from your experiment



A

```
(base) emilio@Alienware:/mnt/NTFS/Nano_Test$ guppy_basecaller --print_workflows | grep "FLO-MIN106 SQK-LSK109"
FLO-MIN106 SQK-LSK109                dna_r9.4.1_450bps_hac          2021-05-17_dna_r9.4.1_minion_384_d37a2ab9
FLO-MIN106 SQK-LSK109-XL             dna_r9.4.1_450bps_hac          2021-05-17_dna_r9.4.1_minion_384_d37a2ab9
(base) emilio@Alienware:/mnt/NTFS/Nano_Test$
```

B

```
(base) emilio@Alienware:/mnt/NTFS/Nano_Test$ guppy_barcoder --print_kits | more

Available barcoding kits are:
EXP-MRT001
EXP-NBD103
EXP-NBD104
EXP-NBD114
EXP-NBD196
EXP-PBC001
EXP-PBC096
EXP-SRT001
OND-SQK-LP0096M
OND-SQK-LP0096MA
OND-SQK-LP0096S
OND-SQK-LP0768L
OND-SQK-RP0096M
OND-SQK-RP0096MA
OND-SQK-RP0384L
SQK-16S024
SQK-LWB001
SQK-MLK110-96-XL
SQK-NBD110-24
SQK-NBD110-96
```

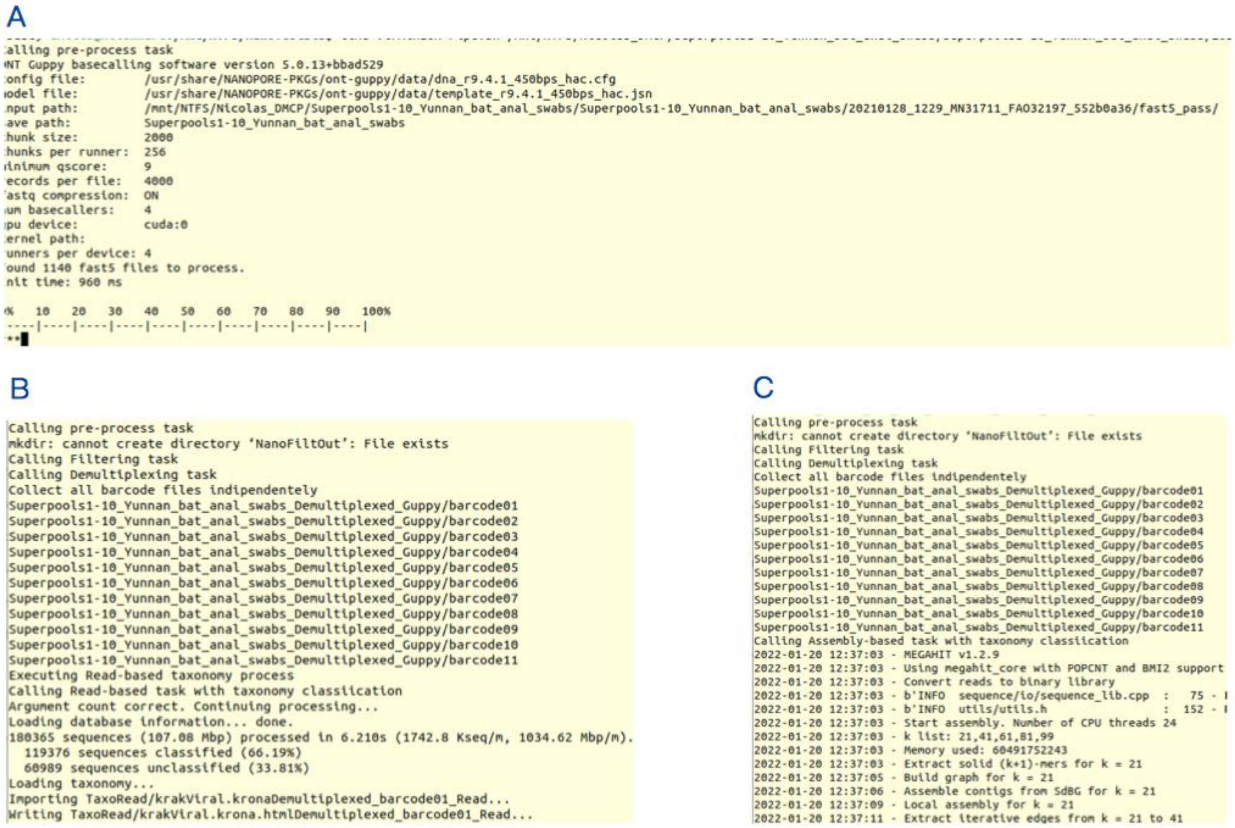
Figure 2 Commands to get Kit and basecall information

The Figure 2A reports the command to execute in order to get the list of available BaseCallerCode on your Ubuntu machine, while Figure 2B shows how to obtain the list of available kits. Once this information is available, you are ready to start the Vir-MinION pipeline.

The following command line can be taken as an example of usage, where the read_based strategy will be applied for the taxonomic classification of the reads.

```
VirMinION-Pipe.sh /mnt/NTFS/Nicolas_DMCP/Superpools1-10_Yunnan_bat_anal_swabs/Superpools1-
10_Yunnan_bat_anal_swabs/20210128_1229_MN31711_FAO32197_552b0a36/fast5_pass/ Superpools1-
10_Yunnan_bat_anal_swabs dna_r9.4.1_450bps_hac.cfg EXP-NBD104 24 --read_based
```

Figure 3A-B-C represent the screen output during the basecalling task, the read_based and ass_based strategies execution, respectively.



```
alling pre-process task
NT Guppy basecalling software version 5.0.13+bbad529
onfig file:        /usr/share/NANOPORE-PKGs/ont-guppy/data/dna_r9.4.1_450bps_hac.cfg
odel file:         /usr/share/NANOPORE-PKGs/ont-guppy/data/template_r9.4.1_450bps_hac.jsn
nput path:         /mnt/NTFS/Nicolas_DMCP/Superpools1-10_Yunnan_bat_anal_swabs/Superpools1-10_Yunnan_bat_anal_swabs/20210128_1229_MN31711_FA032197_552b0a36/fast5_pass/
ave path:          Superpools1-10_Yunnan_bat_anal_swabs
hunk size:         2000
hunks per runner:  256
inimum qscore:     9
ecords per file:   4000
astq compression:  ON
um basecallers:    4
pu device:         cuda:0
ernel path:
unners per device: 4
ound 1140 fast5 files to process.
nit time: 960 ms

%  10   20   30   40   50   60   70   80   90   100%
----|----|----|----|----|----|----|----|----|----|
**
```

```
Calling pre-process task
mkdir: cannot create directory 'NanoFiltOut': File exists
Calling Filtering task
Calling Demultiplexing task
Collect all barcode files indipendentely
Superpools1-10_Yunnan_bat_anal_swabs_Demultiplexed_Guppy/barcode01
Superpools1-10_Yunnan_bat_anal_swabs_Demultiplexed_Guppy/barcode02
Superpools1-10_Yunnan_bat_anal_swabs_Demultiplexed_Guppy/barcode03
Superpools1-10_Yunnan_bat_anal_swabs_Demultiplexed_Guppy/barcode04
Superpools1-10_Yunnan_bat_anal_swabs_Demultiplexed_Guppy/barcode05
Superpools1-10_Yunnan_bat_anal_swabs_Demultiplexed_Guppy/barcode06
Superpools1-10_Yunnan_bat_anal_swabs_Demultiplexed_Guppy/barcode07
Superpools1-10_Yunnan_bat_anal_swabs_Demultiplexed_Guppy/barcode08
Superpools1-10_Yunnan_bat_anal_swabs_Demultiplexed_Guppy/barcode09
Superpools1-10_Yunnan_bat_anal_swabs_Demultiplexed_Guppy/barcode10
Superpools1-10_Yunnan_bat_anal_swabs_Demultiplexed_Guppy/barcode11
Executing Read-based taxonomy process
Calling Read-based task with taxonomy classiication
Argument count correct. Continuing processing...
Loading database information... done.
180365 sequences (107.08 Mbp) processed in 6.210s (1742.8 Kseq/m, 1034.62 Mbp/m).
  119376 sequences classified (66.19%)
  60989 sequences unclassified (33.81%)
Loading taxonomy...
Importing TaxoRead/krakViral.kronaDemultiplexed_barcode01_Read...
Writing TaxoRead/krakViral.krona.htmlDemultiplexed_barcode01_Read...
```

```
Calling pre-process task
mkdir: cannot create directory 'NanoFiltOut': File exists
Calling Filtering task
Calling Demultiplexing task
Collect all barcode files indipendentely
Superpools1-10_Yunnan_bat_anal_swabs_Demultiplexed_Guppy/barcode01
Superpools1-10_Yunnan_bat_anal_swabs_Demultiplexed_Guppy/barcode02
Superpools1-10_Yunnan_bat_anal_swabs_Demultiplexed_Guppy/barcode03
Superpools1-10_Yunnan_bat_anal_swabs_Demultiplexed_Guppy/barcode04
Superpools1-10_Yunnan_bat_anal_swabs_Demultiplexed_Guppy/barcode05
Superpools1-10_Yunnan_bat_anal_swabs_Demultiplexed_Guppy/barcode06
Superpools1-10_Yunnan_bat_anal_swabs_Demultiplexed_Guppy/barcode07
Superpools1-10_Yunnan_bat_anal_swabs_Demultiplexed_Guppy/barcode08
Superpools1-10_Yunnan_bat_anal_swabs_Demultiplexed_Guppy/barcode09
Superpools1-10_Yunnan_bat_anal_swabs_Demultiplexed_Guppy/barcode10
Superpools1-10_Yunnan_bat_anal_swabs_Demultiplexed_Guppy/barcode11
Calling Assembly-based task with taxonomy classiication
2022-01-20 12:37:03 - MEGAHIT v1.2.9
2022-01-20 12:37:03 - Using megahit_core with POPCNT and BMI2 support
2022-01-20 12:37:03 - Convert reads to binary library
2022-01-20 12:37:03 - b'INFO  sequence/io/sequence_lib.cpp  :   75 - I
2022-01-20 12:37:03 - b'INFO  utils/utils.h                 :  152 - I
2022-01-20 12:37:03 - Start assembly. Number of CPU threads 24
2022-01-20 12:37:03 - k list: 21,41,61,81,99
2022-01-20 12:37:03 - Memory used: 60491752243
2022-01-20 12:37:03 - Extract solid (k+1)-mers for k = 21
2022-01-20 12:37:05 - Build graph for k = 21
2022-01-20 12:37:06 - Assemble contigs from SdBG for k = 21
2022-01-20 12:37:09 - Local assembly for k = 21
2022-01-20 12:37:11 - Extract iterative edges from k = 21 to 41
```

Figure 2 Screenshots of Vir-MinION execution

At the end of the Vir-MinION pipeline execution, the folder structure depicted in Figure 4 will be available. In detail, the html files obtained from the krona-blast.sh execution and containing the taxonomic classification will be placed in the following folders:

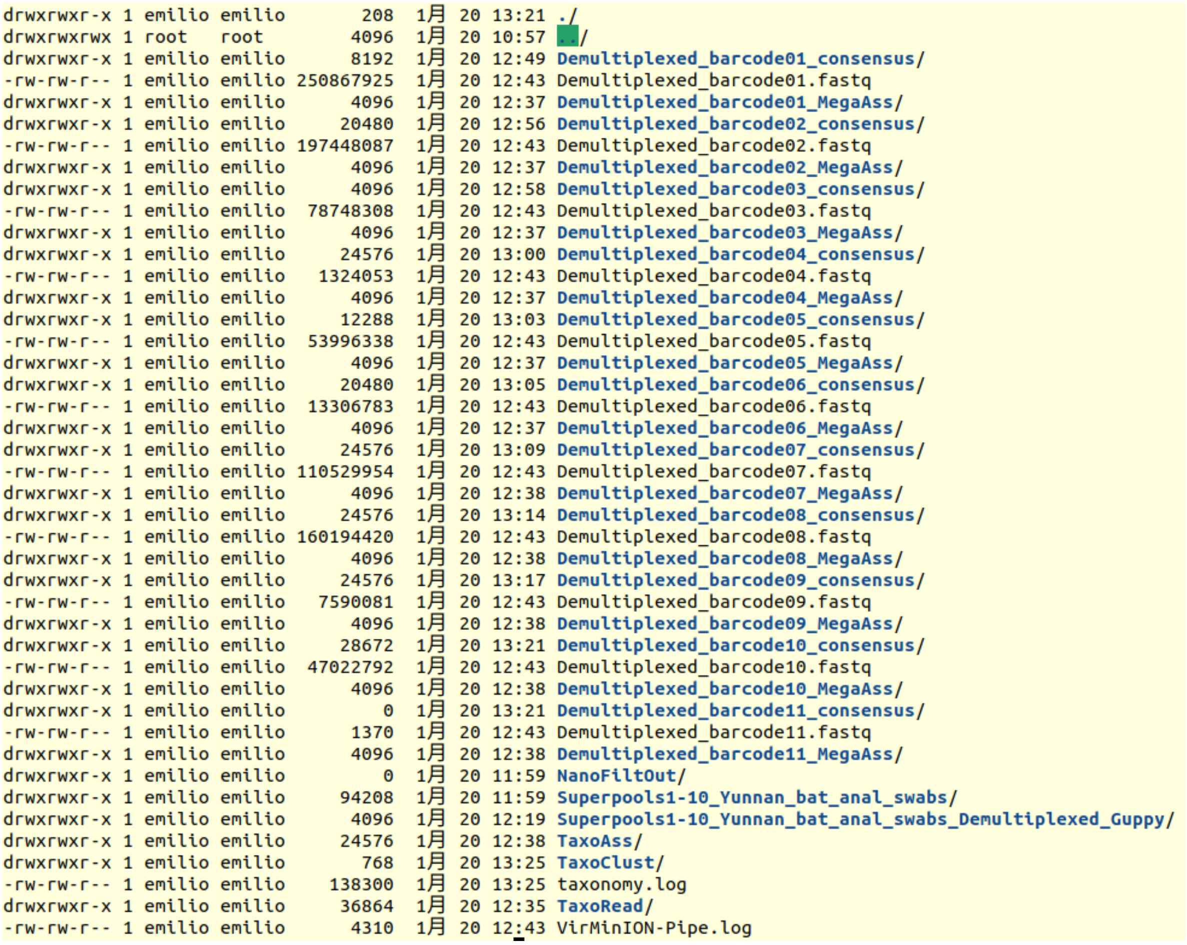
```
-rw-rw-r-- 1 virminion virminion 237232 4 月  21 12:56 KronaBlast_flye/krona_out.html
-rw-rw-r-- 1 virminion virminion 247583 4 月  21 12:57 KronaBlast_mega/krona_out.html
-rw-rw-r-- 1 virminion virminion 461699 4 月  21 13:11 KronaBlast_reads/krona_out.html
-rw-rw-r-- 1 virminion virminion 371653 4 月  22 16:07 KronaBlast_spades/krona_out.html
```

The folders containing the taxonomy classification based on kraken and kaiju, are the following:

```
drwxrwxr-x  2 virminion virminion     4096 4 月  21 11:47 flye_taxonomy/
drwxrwxr-x  2 virminion virminion     4096 4 月  21 11:46 megahit_taxonomy/
drwxrwxr-x  3 virminion virminion     4096 4 月  21 10:43 reads_taxonomy/
drwxrwxr-x  3 virminion virminion     4096 4 月  22 15:58 spades_taxonomy/
```

During the execution, Vir-MinION will produce a collection of logs to document out every task:

```
-rw-rw-r-- 1 virminion virminion  27487 4 月  22 12:39 ass_flye/flye.log
-rw-rw-r-- 1 virminion virminion 119605 4 月  22 13:47 ass_spades/spades.log
-rw-rw-r-- 1 virminion virminion    794 4 月  22 13:47 ass_spades/warnings.log
-rw-rw-r-- 1 virminion virminion    535 4 月  22 12:56 KronaBlast_flye/krona-blast.log
-rw-rw-r-- 1 virminion virminion    540 4 月  22 12:57 KronaBlast_mega/krona-blast.log
-rw-rw-r-- 1 virminion virminion    554 4 月  22 13:11 KronaBlast_reads/krona-blast.log
-rw-rw-r-- 1 virminion virminion    542 4 月  22 13:07 KronaBlast_spades/krona-blast.log
-rw-rw-r-- 1 virminion virminion   4930 4 月  22 13:06 taxonomy.log
-rw-rw-r-- 1 virminion virminion   4930 4 月  22 12:06 VirMinION-Pipe.log
```



```
drwxrwxr-x 1 emilio emilio       208  1月  20 13:21 ./
drwxrwxrwx 1 root   root        4096  1月  20 10:57 ../
drwxrwxr-x 1 emilio emilio      8192  1月  20 12:49 Demultiplexed_barcode01_consensus/
-rw-rw-r-- 1 emilio emilio 250867925  1月  20 12:43 Demultiplexed_barcode01.fastq
drwxrwxr-x 1 emilio emilio      4096  1月  20 12:37 Demultiplexed_barcode01_MegaAss/
drwxrwxr-x 1 emilio emilio     20480  1月  20 12:56 Demultiplexed_barcode02_consensus/
-rw-rw-r-- 1 emilio emilio 197448087  1月  20 12:43 Demultiplexed_barcode02.fastq
drwxrwxr-x 1 emilio emilio      4096  1月  20 12:37 Demultiplexed_barcode02_MegaAss/
drwxrwxr-x 1 emilio emilio      4096  1月  20 12:58 Demultiplexed_barcode03_consensus/
-rw-rw-r-- 1 emilio emilio  78748308  1月  20 12:43 Demultiplexed_barcode03.fastq
drwxrwxr-x 1 emilio emilio      4096  1月  20 12:37 Demultiplexed_barcode03_MegaAss/
drwxrwxr-x 1 emilio emilio     24576  1月  20 13:00 Demultiplexed_barcode04_consensus/
-rw-rw-r-- 1 emilio emilio   1324053  1月  20 12:43 Demultiplexed_barcode04.fastq
drwxrwxr-x 1 emilio emilio      4096  1月  20 12:37 Demultiplexed_barcode04_MegaAss/
drwxrwxr-x 1 emilio emilio     12288  1月  20 13:03 Demultiplexed_barcode05_consensus/
-rw-rw-r-- 1 emilio emilio  53996338  1月  20 12:43 Demultiplexed_barcode05.fastq
drwxrwxr-x 1 emilio emilio      4096  1月  20 12:37 Demultiplexed_barcode05_MegaAss/
drwxrwxr-x 1 emilio emilio     20480  1月  20 13:05 Demultiplexed_barcode06_consensus/
-rw-rw-r-- 1 emilio emilio  13306783  1月  20 12:43 Demultiplexed_barcode06.fastq
drwxrwxr-x 1 emilio emilio      4096  1月  20 12:37 Demultiplexed_barcode06_MegaAss/
drwxrwxr-x 1 emilio emilio     24576  1月  20 13:09 Demultiplexed_barcode07_consensus/
-rw-rw-r-- 1 emilio emilio 110529954  1月  20 12:43 Demultiplexed_barcode07.fastq
drwxrwxr-x 1 emilio emilio      4096  1月  20 12:38 Demultiplexed_barcode07_MegaAss/
drwxrwxr-x 1 emilio emilio     24576  1月  20 13:14 Demultiplexed_barcode08_consensus/
-rw-rw-r-- 1 emilio emilio 160194420  1月  20 12:43 Demultiplexed_barcode08.fastq
drwxrwxr-x 1 emilio emilio      4096  1月  20 12:38 Demultiplexed_barcode08_MegaAss/
drwxrwxr-x 1 emilio emilio     24576  1月  20 13:17 Demultiplexed_barcode09_consensus/
-rw-rw-r-- 1 emilio emilio   7590081  1月  20 12:43 Demultiplexed_barcode09.fastq
drwxrwxr-x 1 emilio emilio      4096  1月  20 12:38 Demultiplexed_barcode09_MegaAss/
drwxrwxr-x 1 emilio emilio     28672  1月  20 13:21 Demultiplexed_barcode10_consensus/
-rw-rw-r-- 1 emilio emilio  47022792  1月  20 12:43 Demultiplexed_barcode10.fastq
drwxrwxr-x 1 emilio emilio      4096  1月  20 12:38 Demultiplexed_barcode10_MegaAss/
drwxrwxr-x 1 emilio emilio         0  1月  20 13:21 Demultiplexed_barcode11_consensus/
-rw-rw-r-- 1 emilio emilio      1370  1月  20 12:43 Demultiplexed_barcode11.fastq
drwxrwxr-x 1 emilio emilio      4096  1月  20 12:38 Demultiplexed_barcode11_MegaAss/
drwxrwxr-x 1 emilio emilio         0  1月  20 11:59 NanoFiltOut/
drwxrwxr-x 1 emilio emilio     94208  1月  20 11:59 Superpools1-10_Yunnan_bat_anal_swabs/
drwxrwxr-x 1 emilio emilio      4096  1月  20 12:19 Superpools1-10_Yunnan_bat_anal_swabs_Demultiplexed_Guppy/
drwxrwxr-x 1 emilio emilio     24576  1月  20 12:38 TaxoAss/
drwxrwxr-x 1 emilio emilio       768  1月  20 13:25 TaxoClust/
-rw-rw-r-- 1 emilio emilio    138300  1月  20 13:25 taxonomy.log
drwxrwxr-x 1 emilio emilio     36864  1月  20 12:35 TaxoRead/
-rw-rw-r-- 1 emilio emilio      4310  1月  20 12:43 VirMinION-Pipe.log
```

Figure 3 Vir-MinION output data structure

The following table reports relevant parameters used during the script's execution and its relative program. The user can refer to it to customize the pipeline, according to the data/experiment.

Pipeline	Script name	Program	Parameters
Vir-MinION	VirMinION-Pipe.sh	NGSpeciesID	NGSpeciesID --ont --fastq $i –outfolder $folder --rc_identity_threshold 0.8 --medaka --t 12
Vir-MinION	VirMinION-Pipe.sh	megahit	megahit -t 24 --read $i --k-list 21,41,61,81,99 --no-mercy --min-count 2 --out-dir $mega_outFolder
Vir-MinION	VirMinION-Pipe.sh	flye	--nano-raw $i --threads $JThread --out-dir $flye_outFoder --debug --meta
Vir-MinION	VirMinION-Pipe.sh	guppy_basecaller	-i $PreProcDataFolder -r -s $PreProcOutFolder -c $BaseCallConfFile -x "cuda:0" --compress_fastq --require_barcodes_both_ends --trim_barcodes
Vir-MinION	VirMinION-Pipe.sh	NanoFilt	--maxlength 1500 -l 500 -q 10 --minGC 0.4 --maxGC 0.6
Vir-MinION	VirMinION-Pipe.sh	guppy_barcoder	-i $NanoFiltFolder --save_path $DemuOutFold -x "cuda:0" --barcode_kits $BarCodeKit
Vir-MinION	taxonomy.sh	kraken2	--db $KrakenViralDB $FilteredReads --output $krakenViralOut --classified-out $krakenViralClassified --unclassified-out $krakenViralUnClassified --report $krakenViralReport
Vir-MinION	taxonomy.sh	kaiju	-t $kaijuNodes -f $kaijuDB -i $FilteredReads -o $kaijuOut
Vir-MinION	taxonomy.sh	kaiju2krona	-t $kaijuNodes -n $kaijuNames -i $kaijuOut -o $kronaOut -u -v
Vir-MinION	krona-blast.sh	blastn	-db $ref_viruses_rep_genomes -query $merged_seq -evalue 1e-3 -word_size 11 -outfmt "6 std staxid staxids" -num_threads $Jtrim