Oxford Nanopore Quality Control Work Flow v1.0

A Step-by-Step Guide

[DRAFT]

Michael Hall

Plant Breeding and Genetics Laboratory
FAO/IAEA Joint Division
Seibersdorf, Austria

Created: June, 2022
Last updated: 28 June 2022

Please note: *This is not an official IAEA publication but is made available as working material. The material has not undergone an official review by the IAEA. The views expressed do not necessarily reflect those of the International Atomic Energy Agency or its Member States and remain the responsibility of the contributors. The use of particular designations of countries or territories does not imply any judgement by the publisher, the IAEA, as to the legal status of such countries or territories, of their authorities and institutions or of the delimitation of their boundaries. The mention of names of specific companies or products (whether or not indicated as registered) does not imply any intention to infringe proprietary rights, nor should it be construed as an endorsement or recommendation on the part of the IAEA.*

Contents

1 Oxford Nanopore Minion Quality Control **2**
1.1 Standard Workflow 2
1.2 Data Tree Structure 2
1.3 Guppy Basecaller Log Standard Output 4
1.4 Conda/Mamba Environment Preview 4
1.5 Standard R Script Output 4
1.6 Figures/Plots 5

1 Oxford Nanopore Minion Quality Control

Author Michael Hall

Date 06/28/2022

Arabidopsis Thalania

1.1 Standard Workflow

MinionQC

Clone the Github Repository

git clone https://github.com/PBGLMichaelHall/MinionQC.git

Change Working Directory

cd/ MinionQC

Create a new enviroment

mamba env create –file env/MinionQCenv.yaml

Activate new Environment

conda activate MinionQC

Run Minion R Script

Rscript Minion_R_QC_Script/MinIONQC.R -i summary/sequencing_summary.txt -q 0

1.2 Data Tree Structure

```
(base) michael@mh-ubuntu:~/Desktop/Arabidopsis_Thalania_Run00123$ tree
.
├── fast5
│   ├── AJI658_535642224ed16dcaf6e8cab4fc0ebb5f60ddaf03_0.fast5
│   ├── AJI658_535642224ed16dcaf6e8cab4fc0ebb5f60ddaf03_10.fast5
│   ├── AJI658_535642224ed16dcaf6e8cab4fc0ebb5f60ddaf03_11.fast5
│   ├── AJI658_535642224ed16dcaf6e8cab4fc0ebb5f60ddaf03_12.fast5
│   ├── AJI658_535642224ed16dcaf6e8cab4fc0ebb5f60ddaf03_13.fast5
│   ├── AJI658_535642224ed16dcaf6e8cab4fc0ebb5f60ddaf03_14.fast5
│   ├── AJI658_535642224ed16dcaf6e8cab4fc0ebb5f60ddaf03_15.fast5
│   ├── AJI658_535642224ed16dcaf6e8cab4fc0ebb5f60ddaf03_16.fast5
│   ├── AJI658_535642224ed16dcaf6e8cab4fc0ebb5f60ddaf03_17.fast5
│   ├── AJI658_535642224ed16dcaf6e8cab4fc0ebb5f60ddaf03_18.fast5
│   ├── AJI658_535642224ed16dcaf6e8cab4fc0ebb5f60ddaf03_19.fast5
│   ├── AJI658_535642224ed16dcaf6e8cab4fc0ebb5f60ddaf03_1.fast5
│   ├── AJI658_535642224ed16dcaf6e8cab4fc0ebb5f60ddaf03_20.fast5
│   ├── AJI658_535642224ed16dcaf6e8cab4fc0ebb5f60ddaf03_21.fast5
│   ├── AJI658_535642224ed16dcaf6e8cab4fc0ebb5f60ddaf03_22.fast5
│   ├── AJI658_535642224ed16dcaf6e8cab4fc0ebb5f60ddaf03_23.fast5
│   ├── AJI658_535642224ed16dcaf6e8cab4fc0ebb5f60ddaf03_24.fast5
│   ├── AJI658_535642224ed16dcaf6e8cab4fc0ebb5f60ddaf03_25.fast5
│   ├── AJI658_535642224ed16dcaf6e8cab4fc0ebb5f60ddaf03_2.fast5
```

(continues on next page)

(continued from previous page)

```
│   ├── AJI658_535642224ed16dcaf6e8cab4fc0ebb5f60ddaf03_3.fast5
│   ├── AJI658_535642224ed16dcaf6e8cab4fc0ebb5f60ddaf03_4.fast5
│   ├── AJI658_535642224ed16dcaf6e8cab4fc0ebb5f60ddaf03_5.fast5
│   ├── AJI658_535642224ed16dcaf6e8cab4fc0ebb5f60ddaf03_6.fast5
│   ├── AJI658_535642224ed16dcaf6e8cab4fc0ebb5f60ddaf03_7.fast5
│   ├── AJI658_535642224ed16dcaf6e8cab4fc0ebb5f60ddaf03_8.fast5
│   └── AJI658_535642224ed16dcaf6e8cab4fc0ebb5f60ddaf03_9.fast5
└── guppy_out
├── fail
│   ├── fastq_runid_535642224ed16dcaf6e8cab4fc0ebb5f60ddaf03_0_0.fastq
│   ├── fastq_runid_535642224ed16dcaf6e8cab4fc0ebb5f60ddaf03_1_0.fastq
│   ├── fastq_runid_535642224ed16dcaf6e8cab4fc0ebb5f60ddaf03_2_0.fastq
│   ├── fastq_runid_535642224ed16dcaf6e8cab4fc0ebb5f60ddaf03_3_0.fastq
│   ├── fastq_runid_535642224ed16dcaf6e8cab4fc0ebb5f60ddaf03_4_0.fastq
│   ├── fastq_runid_535642224ed16dcaf6e8cab4fc0ebb5f60ddaf03_5_0.fastq
│   └── fastq_runid_535642224ed16dcaf6e8cab4fc0ebb5f60ddaf03_6_0.fastq
├── guppy_basecaller-core-dump-db
│   ├── attachments
│   ├── completed
│   ├── new
│   ├── pending
│   └── settings.dat
├── guppy_basecaller_log-2022-05-23_12-49-31.log
├── guppy_basecaller_log-2022-05-23_18-21-02.log
├── guppy_out
│   ├── guppy_basecaller-core-dump-db
│   │   ├── attachments
│   │   ├── completed
│   │   ├── new
│   │   ├── pending
│   │   └── settings.dat
│   ├── guppy_basecaller_log-2022-05-23_12-49-21.log
│   └── sequencing_summary.txt
├── pass
│   ├── fastq_runid_535642224ed16dcaf6e8cab4fc0ebb5f60ddaf03_0_0.fastq
│   ├── fastq_runid_535642224ed16dcaf6e8cab4fc0ebb5f60ddaf03_1_0.fastq
│   ├── fastq_runid_535642224ed16dcaf6e8cab4fc0ebb5f60ddaf03_2_0.fastq
│   ├── fastq_runid_535642224ed16dcaf6e8cab4fc0ebb5f60ddaf03_3_0.fastq
│   ├── fastq_runid_535642224ed16dcaf6e8cab4fc0ebb5f60ddaf03_4_0.fastq
│   ├── fastq_runid_535642224ed16dcaf6e8cab4fc0ebb5f60ddaf03_5_0.fastq
│   └── fastq_runid_535642224ed16dcaf6e8cab4fc0ebb5f60ddaf03_6_0.fastq
├── sequencing_telemetry.js
└── summary
    ├── MinIONQC.R
    └── sequencing_summary.txt

16 directories, 49 files
```

1.3 Guppy Basecaller Log Standard Output

```
2022-05-23 12:49:31.950838 [guppy/message] ONT Guppy basecalling software version 5.0.
↪17+99baa5b27
config file:        /opt/ont/guppy/data/dna_r9.4.1_450bps_hac.cfg
model file:         /opt/ont/guppy/data/template_r9.4.1_450bps_hac.jsn
input path:         fast5/
save path:          guppy_out/
chunk size:         2000
chunks per runner:  256
minimum qscore:     9
records per file:   4000
fastq compression:  ON
num basecallers:    1
cpu mode:           ON
threads per caller: 4
```

1.4 Conda/Mamba Environment Preview

```yaml
name: MinionQC

channels:
  - default
  - bioconda
  - conda-forge
  - r

dependencies:
  - r-data.table
  - r-futile.logger
  - r-ggplot2
  - r-optparse
  - r-plyr
  - r-readr
  - r-reshape2
  - r-scales
  - r-viridis
  - r-yaml
```

1.5 Standard R Script Output

```
mamba env create --file MinionQCenv.yaml

conda activate MinionQC

Rscript MinIONQC.R -i sequencing_summary.txt -q 0


INFO [2022-06-28 10:59:29] Loading input file: sequencing_summary.txt
INFO [2022-06-28 10:59:30] MinION flowcell detected
```

(continues on next page)

(continued from previous page)

```
INFO [2022-06-28 10:59:30] .: creating output directory:.
INFO [2022-06-28 10:59:30] .: summarising input file for flowcell
INFO [2022-06-28 10:59:30] .: plotting length histogram
INFO [2022-06-28 10:59:32] .: plotting mean Q score histogram
INFO [2022-06-28 10:59:33] .: plotting flowcell overview
INFO [2022-06-28 10:59:39] .: plotting flowcell yield over time
INFO [2022-06-28 10:59:40] .: plotting flowcell yield by read length
INFO [2022-06-28 10:59:41] .: plotting sequence length over time
INFO [2022-06-28 10:59:44] .: plotting Q score over time
INFO [2022-06-28 10:59:46] .: plotting reads per hour
INFO [2022-06-28 10:59:47] .: plotting read length vs. q score scatterplot
INFO [2022-06-28 10:59:49] .: plotting flowcell channels summary histograms
INFO [2022-06-28 10:59:50] .: plotting physical overview of output per channel
```

1.6 Figures/Plots

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
1
2
3
4
29
30
31
Read length
1e+05
1e+04
1e+03
1e+02
Hours into run
0
5
10
15
Q
20
10

All reads
Q>=0
10
20
30
channel row
4
8
12
16
4
8
12
16
channel column
GB/channel
1e-03
5e-04

Mean read length (bases)
750
500
250
0
0
5
10
15
Hours into run
Reads
All reads
Q>=0

All reads
Q>=0
Number of reads
600
400
200
0
1e+02
1e+03
1e+04
1e+05
Read length (bases)

30
20
10
Mean Q score of read
1e+02
1e+03
1e+04
1e+05
Read length (bases)
Events per base
(log scale)
1.48e+01
7.39e+00
5.44e+00
2.72e+00
2.00e+00

Mean Q score
9
6
3
0
0
5
10
15
Hours into run
Reads
All reads
Q>=0

All reads
Q>=0
Number of reads
Mean Q score of read
0
100
200
300
400
10
20
30

Number of reads per hour
2000
1500
1000
500
0
0
5
10
15
Hours into run
Reads
All reads
Q>=0

Total yield in gigabases
0.020
0.015
0.010
0.005
0.000
0
10000
20000
30000
40000
50000
Minimum read length (bases)
Reads
All reads
Q>=0

0.020
0.015
0.010
0.005
0.000
Total yield in gigabases
0
5
10
15
Hours into run
Reads
All reads
Q>=0