Review of ONT Sequencing Runs

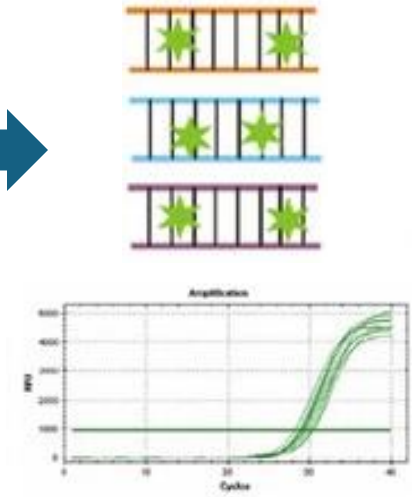
Dr. Louise Judd
Centre for Pathogen Genomics- Innovation Hub
Lead Laboratory Scientist and Genomics Trainer

How do we do genomic sequencing?

1. Clinical sample

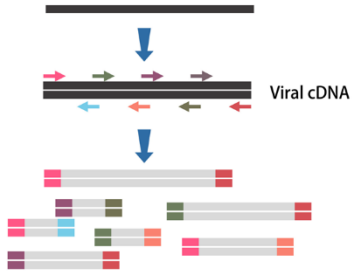
- High viral load
- Skin swabs
- Swab storage
- DNA extraction
- Negative extraction controls

2. Initial test to see if there's a bug



- Q-PCR
- High viral load
- Ct <29
- DNA storage

3. Make more copies of RNA or DNA



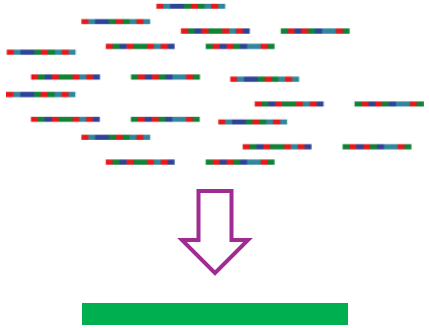
- Tiled amplicon PCR
- Artic 2500 bp scheme

4. Sequencing

- ONT platform
- Rapid libraries
- Sup or HAC basecalling
- 50K reads at N50 ~1200 bp

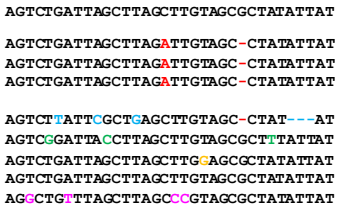
How do we do genomic sequencing (cont.)?

5. Put the pieces together



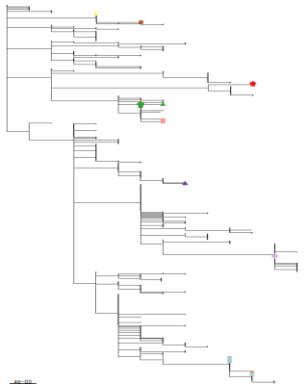
Generate consensus genome

6. Line them up



Align all the sequences to reference genome to compare

7. Map the similarities and differences



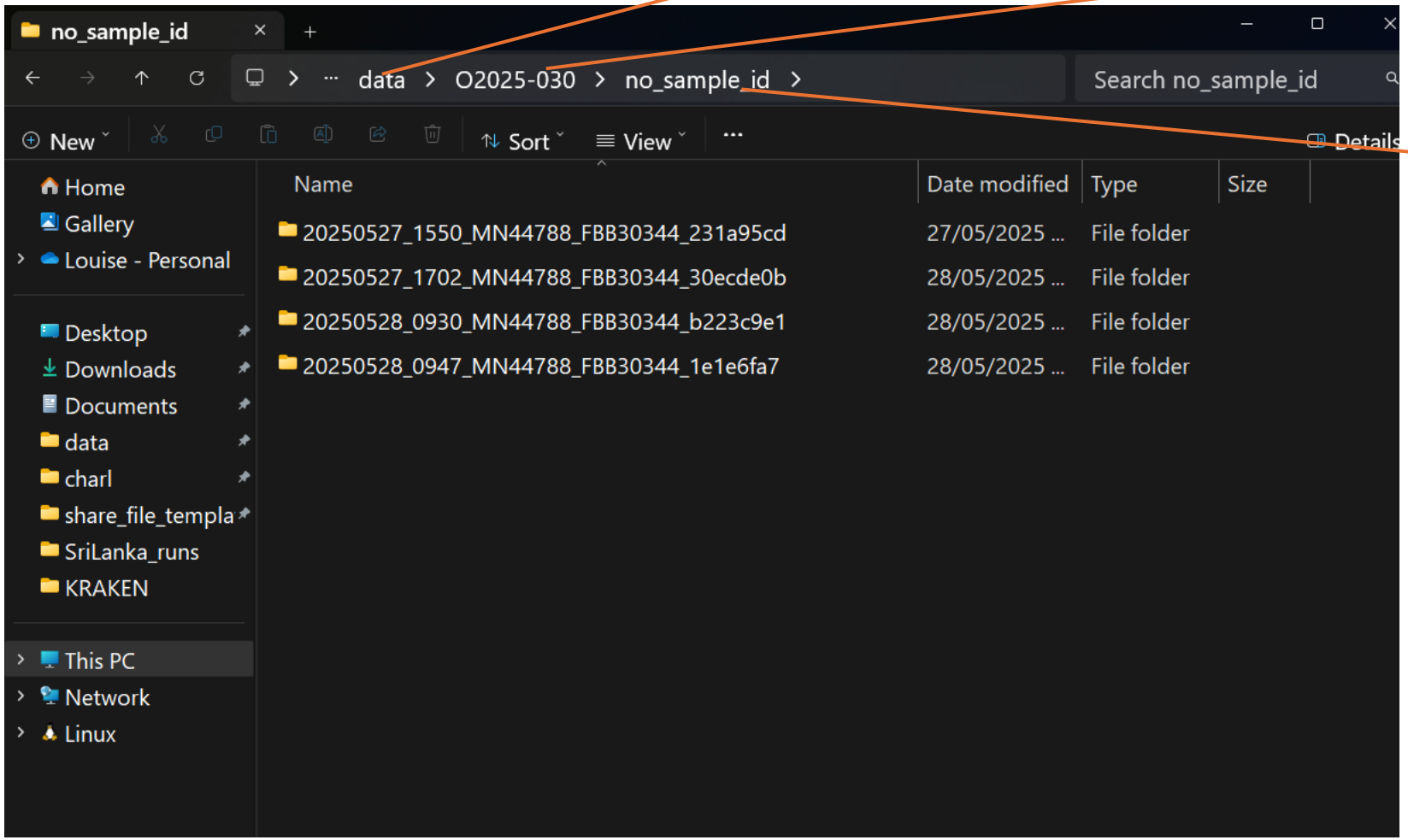
Visualize relationships between sequences

8. Put it all together

Mpox genomics recap

Source Details	Barcode	PCR DNA pool 1 ng/ul	PCR DNA pool 2 ng/ul	final library ng/ul
sample1	barcode41	7	16	
sample2	barcode42	41	46	
sample3	barcode43	5	11	
sample4	barcode44	23	26	
sample5	barcode45	11	11	
sample6	barcode46	30	45	
sample7	barcode47	5	5	
NTC	barcode48	6	12	38

Where do I find my sequencing data?



1. Default location for run is “data”

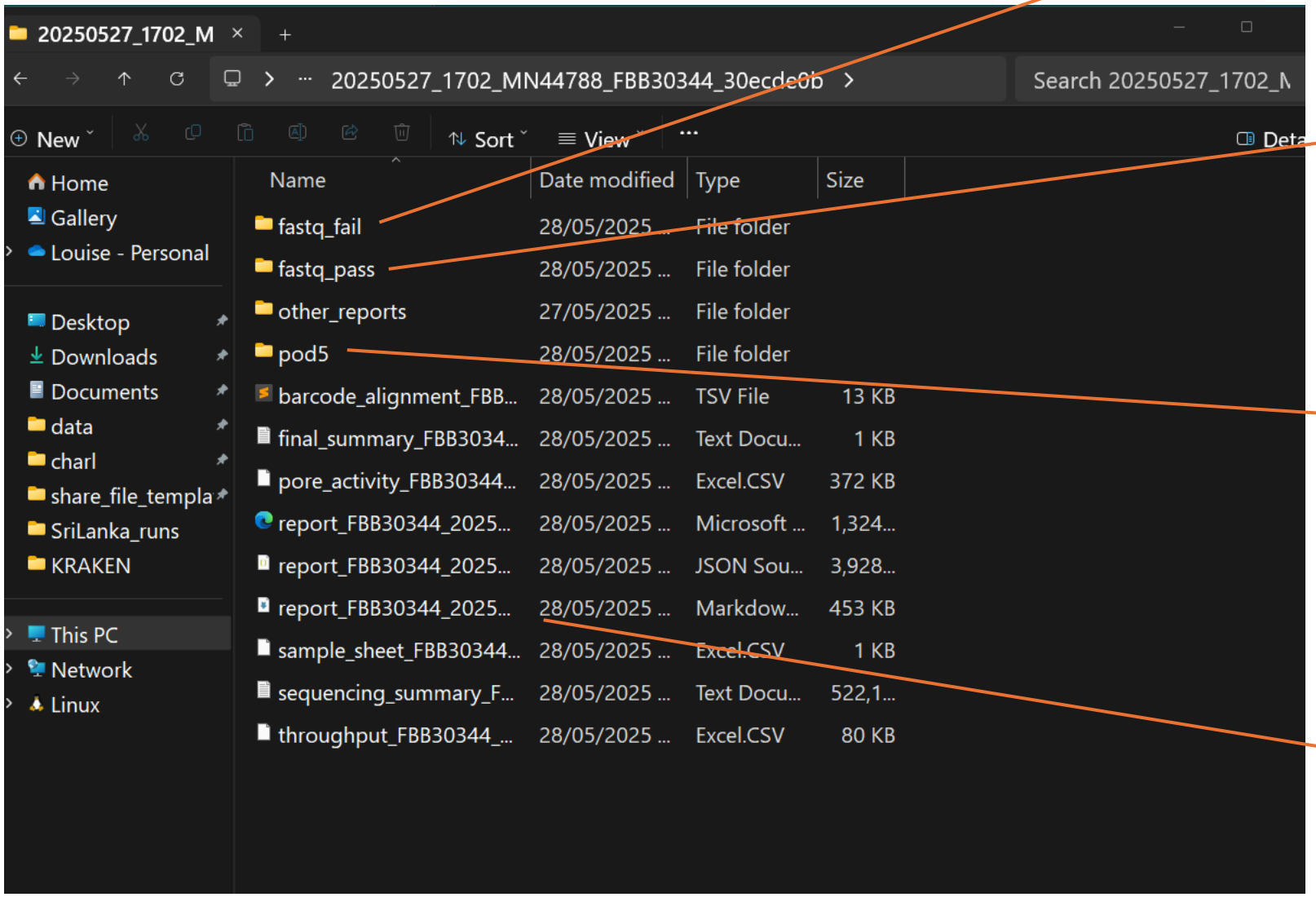
2. Run name will be as entered when the run was being started

3. Will by default say no_sample_id unless sample ids used at set up

New folder each time run is started.
Folder name

- Date stamp
- Time stamp
- Sequencing device ID (MN)
- Flow cell ID
- Random unique number

How is my sequencing data structured?



fastq_fail
- reads with Q score below cut off (<10)
- Discard

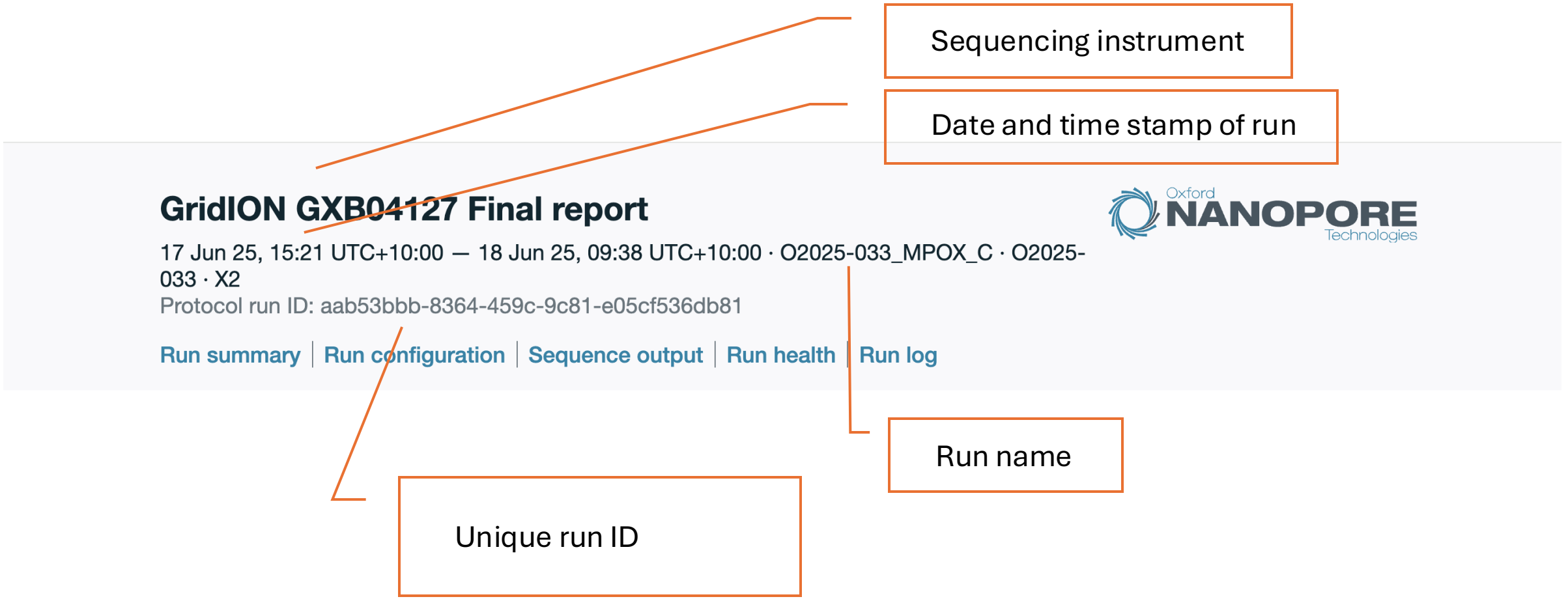
fastq_pass
- reads with Q score above cut off (<10)
- One directory per sample or barcode

pod5
- raw data files
- very large files
- Only needed if you wish to rebsecall
- Keep until basecalling successful

reports
- Move all files to other_reports
- Save whole directory

Sequencing run report

- html file that can be opened with any web browser
- Will only be made if the run finishes successfully



Sequencing run report- Run summary

Check basecalling is at 100%

Run summary

DATA OUTPUT

Estimated bases

3.25 Gb

Reads generated

3.39 M

Estimated N50

1.2 kb

N50: 1.0-1.5 kb

BASECALLING

Reads called

100%

Reads called (min Q score: 10)

2.81 M	**687.53 k**
Pass	Fail

Bases called (min Q score: 10)

1.86 Gb	**313.57 Mb**
Pass	Fail

Proportion reads failing Q Score <10

RUN DURATION

Run time

18 hrs 12 mins / 24 hrs 0 mins

Elapsed time Run limit

Run status

STOPPED · *By user*

Sequencing run report- Run configuration

Run configuration

RUN SETUP

Flow cell type	FLO-MIN114
Flow cell type alias	FLO-MIN114
Flow cell ID	FBA40009
Kit type	SQK-RBK114-96

RUN SETTINGS

Run limit	24 hrs
Pore scan freq.	1.5 hrs
Reserved pores	On
Basecalling	Super-accurate basecalling v4.3.0, 400 bps
Modified basecalling	Off
Trim barcodes	On
Mid-read barcode filtering	Off
Min Q score	10

DATA OUTPUT SETTINGS

FAST5 output	Off
FASTQ data output	One file per hour
POD5 data output	One file per hour, or 500000000 bases per batch
BAM file output	Off
Bulk file output	Off
Data location	/data/./O2025-033_MPOX_C/O2025-033/20250617_1521_X2_FBA40009_aab53bbb

SOFTWARE VERSIONS

MinKNOW	24.11.8
Bream	8.2.5
Configuration	6.2.12
Dorado	7.6.7
MinKNOW Core	6.2.6

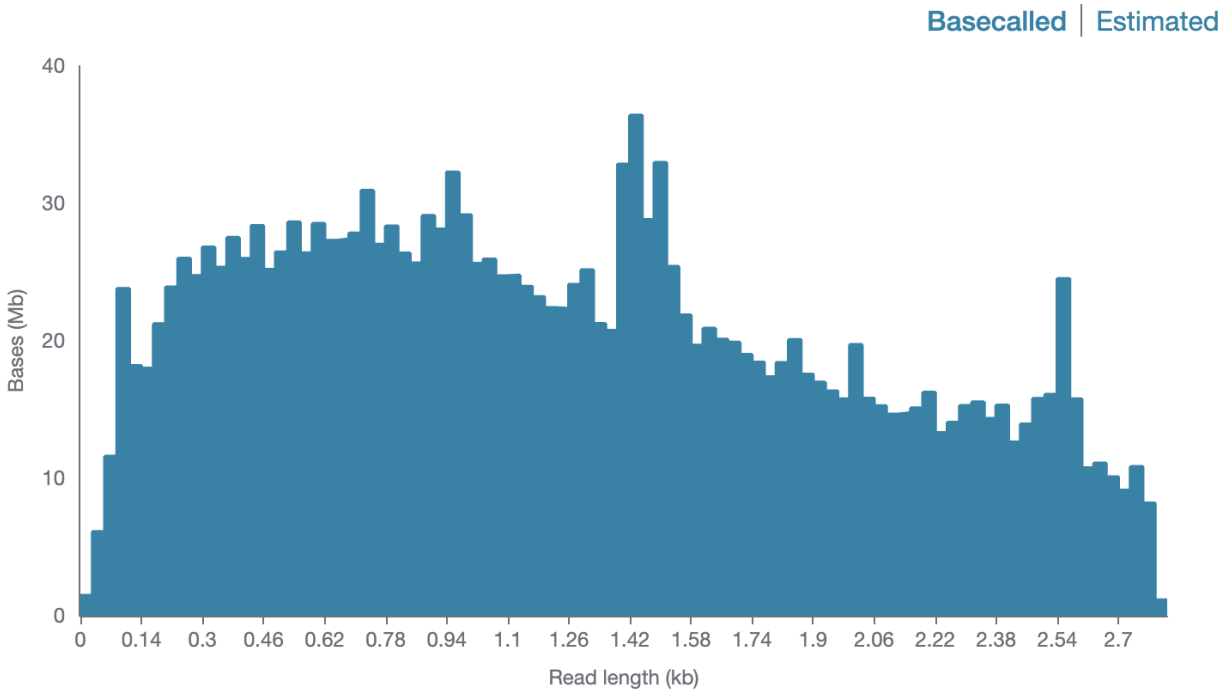
Sequencing run report- Read length distribution

Sequence output

READ LENGTHS · OUTLIERS REMOVED

The read length graph shows the total number of bases vs the read length. The longest 1% of strands are classified as outliers, and excluded to allow focus on the main body of data.

N50*	% Basecalled
1.2 kb	100 %



*N50 calculated from basecalled read length histogram.

OUTLIERS

The longest 1% of strands are classified as outliers, and aggregated into groups to show their relative amounts.

Read length (kb)	Bases (Mb)
0 - 131.072	17.82
131.072 - 262.144	0.44
262.144 - 389.12	0.39

Sequencing run report- Base output

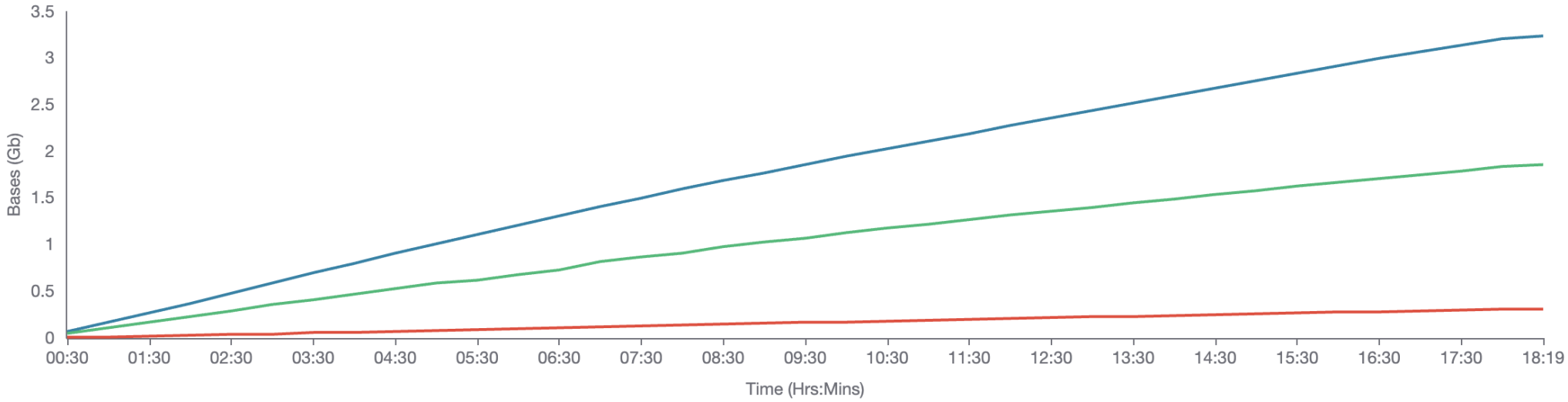
CUMULATIVE OUTPUT

The cumulative output shows the total amount of bases or reads sequenced over time by your device.

Bases

Legend

Estimated	Passed	Failed
Predicted total number of bases, prior to basecalling	Bases equal to or above the quality score threshold.	Bases below the quality score threshold.

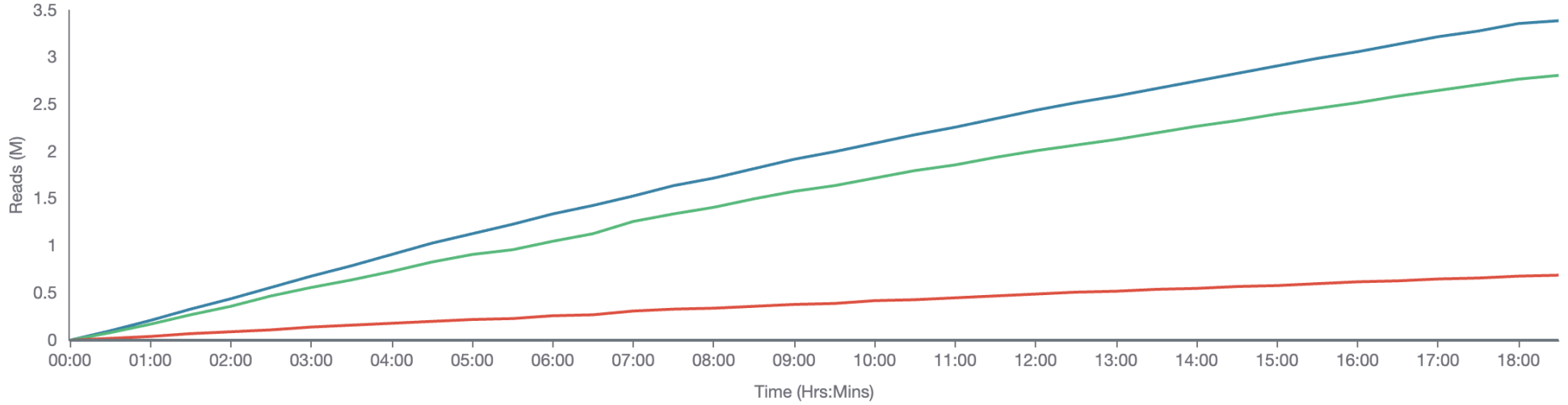


Sequencing run report- Read output

Reads

Legend

- Total — Total number of reads, including passed, failed and skipped.
- Passed — Reads equal to or above the quality score threshold.
- Failed — Reads below the quality score threshold.
- Skipped — Reads that will not be basecalled. Post run basecalling is possible.



Sequencing run report- Barcode distribution table

BARCODES

Detected barcodes Bases graph Reads graph

Detected barcodes

The total number of bases and reads for each barcode detected are displayed in table below. Reads/bases must have a quality score above 9 to pass.

Unclassified data

133.22 k (4.7%) Passed Reads

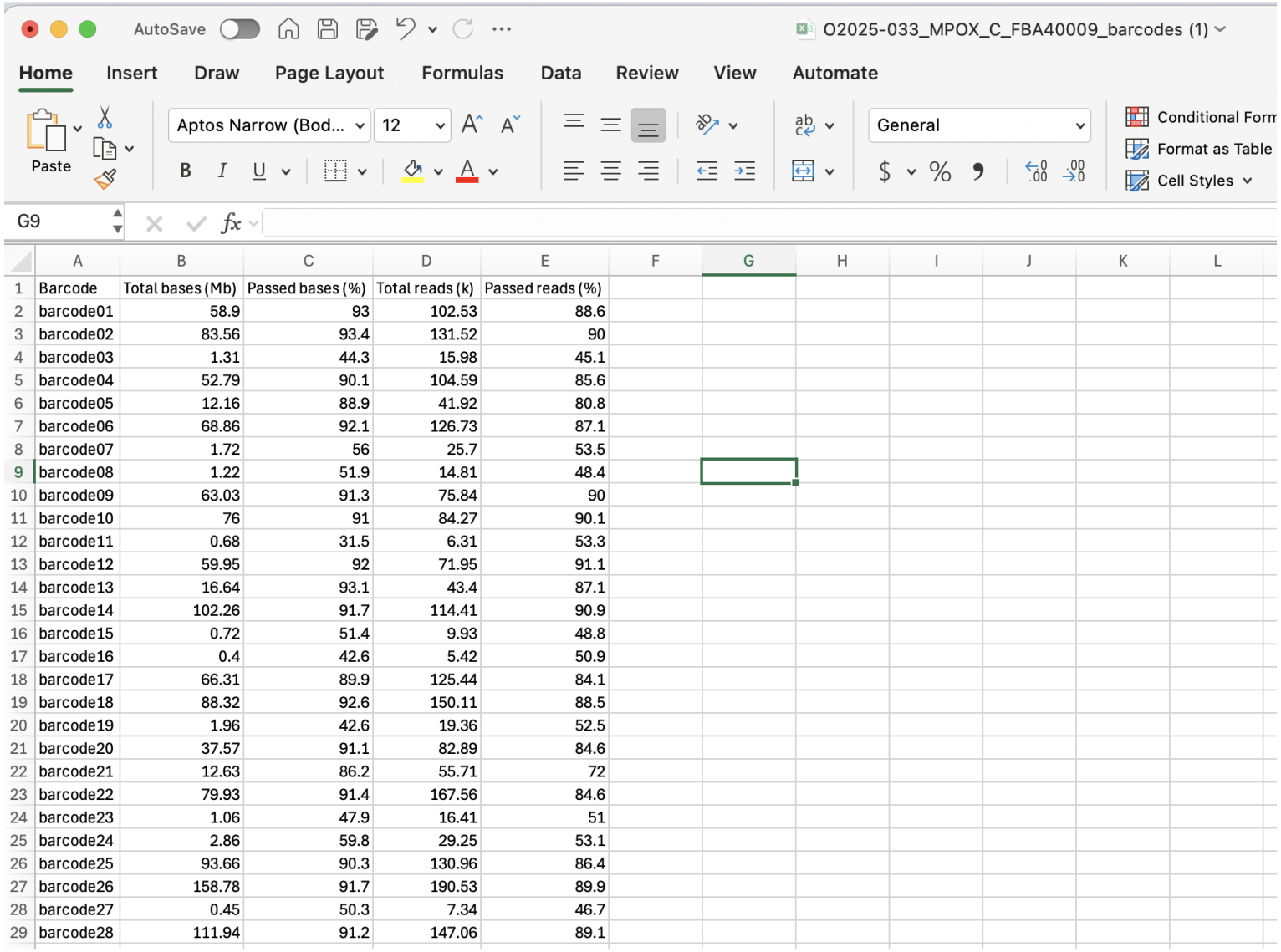
103.62 Mb (5.6%) Passed Bases

Search barcodes

Export CSV

Barcode	Total bases (Mb)	Passed bases (%)	Total reads (k)	Passed reads (%)
barcode01	58.9	93	102.53	88.6
barcode02	83.56	93.4	131.52	90
barcode03	1.31	44.3	15.98	45.1
barcode04	52.79	90.1	104.59	85.6
barcode05	12.16	88.9	41.92	80.8
barcode06	68.86	92.1	126.73	87.1
barcode07	1.72	56	25.7	53.5
barcode08	1.22	51.9	14.81	48.4
barcode09	63.03	91.3	75.84	90
barcode10	76	91	84.27	90.1

Barcode distribution csv file- open with Excel



	A	B	C	D	E	F	G	H	I	J	K	L
1	Barcode	Total bases (Mb)	Passed bases (%)	Total reads (k)	Passed reads (%)							
2	barcode01	58.9	93	102.53	88.6							
3	barcode02	83.56	93.4	131.52	90							
4	barcode03	1.31	44.3	15.98	45.1							
5	barcode04	52.79	90.1	104.59	85.6							
6	barcode05	12.16	88.9	41.92	80.8							
7	barcode06	68.86	92.1	126.73	87.1							
8	barcode07	1.72	56	25.7	53.5							
9	barcode08	1.22	51.9	14.81	48.4							
10	barcode09	63.03	91.3	75.84	90							
11	barcode10	76	91	84.27	90.1							
12	barcode11	0.68	31.5	6.31	53.3							
13	barcode12	59.95	92	71.95	91.1							
14	barcode13	16.64	93.1	43.4	87.1							
15	barcode14	102.26	91.7	114.41	90.9							
16	barcode15	0.72	51.4	9.93	48.8							
17	barcode16	0.4	42.6	5.42	50.9							
18	barcode17	66.31	89.9	125.44	84.1							
19	barcode18	88.32	92.6	150.11	88.5							
20	barcode19	1.96	42.6	19.36	52.5							
21	barcode20	37.57	91.1	82.89	84.6							
22	barcode21	12.63	86.2	55.71	72							
23	barcode22	79.93	91.4	167.56	84.6							
24	barcode23	1.06	47.9	16.41	51							
25	barcode24	2.86	59.8	29.25	53.1							
26	barcode25	93.66	90.3	130.96	86.4							
27	barcode26	158.78	91.7	190.53	89.9							
28	barcode27	0.45	50.3	7.34	46.7							
29	barcode28	111.94	91.2	147.06	89.1							

Sequencing run report- Barcode distribution graph

BARCODES

Detected barcodes **Bases graph** Reads graph

Bases graph

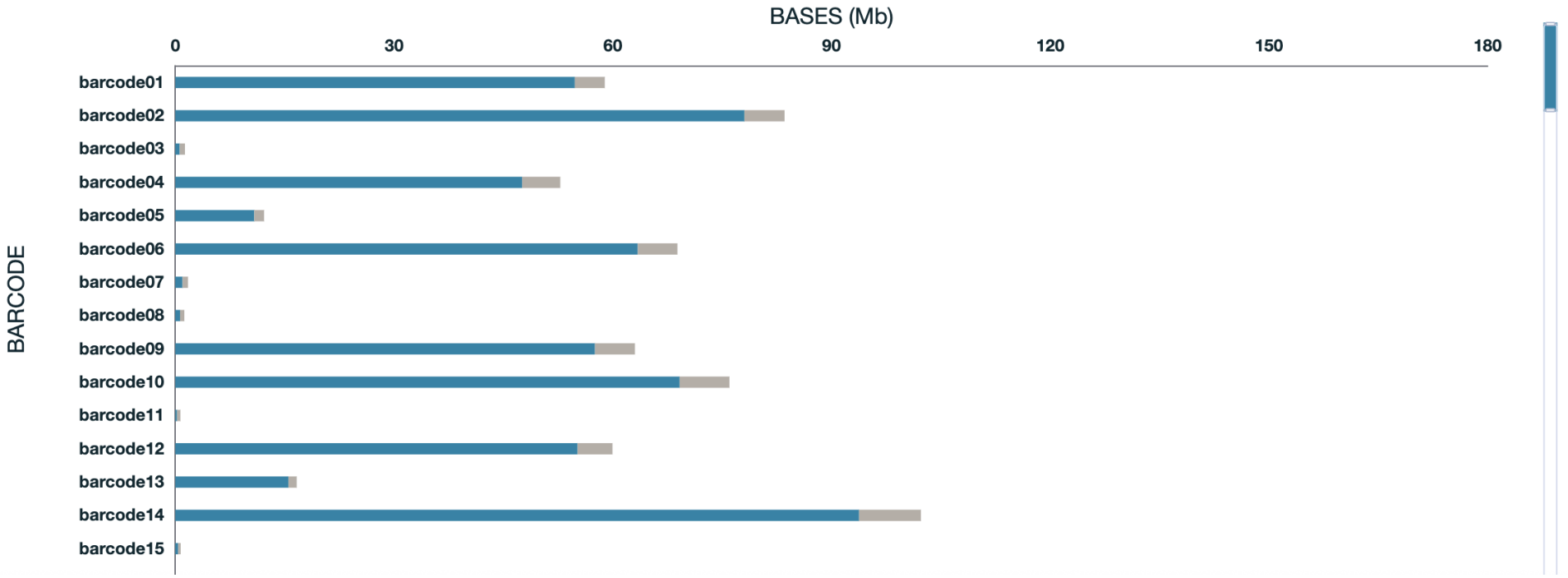
The total number of bases for each barcode detected and displayed in the graph below. Bases must have a quality score above 9 to pass.

Unclassified data

103.62 Mb	5.6%
Passed Bases	Percentage

Barcodes ascending

Expand

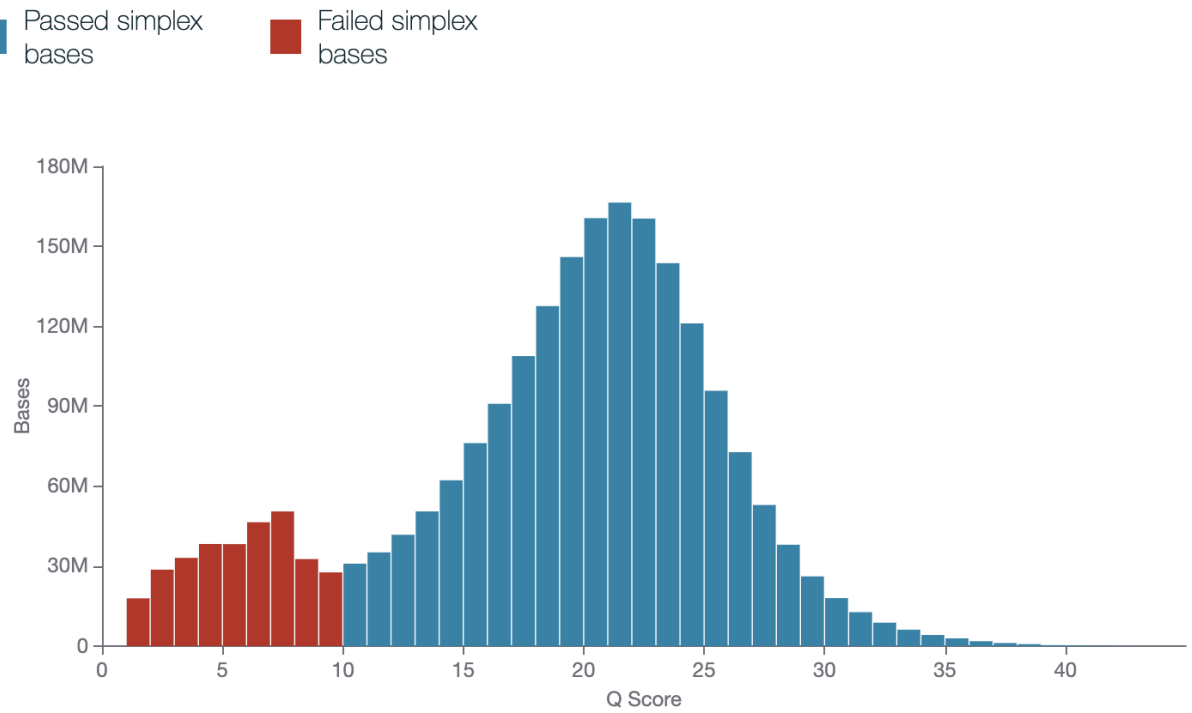


Sequencing run report- Q Score distribution graph

QUALITY SCORE

The quality score is calculated as basecalling is performed on your device. Reads that fall below the minimum value of 9 will be classified as failed reads. You can alter the accepted minimum quality score in MinKNOW.

Q Score histogram



Filters

Min Q score	10

Q Score over time

Legend

Mode

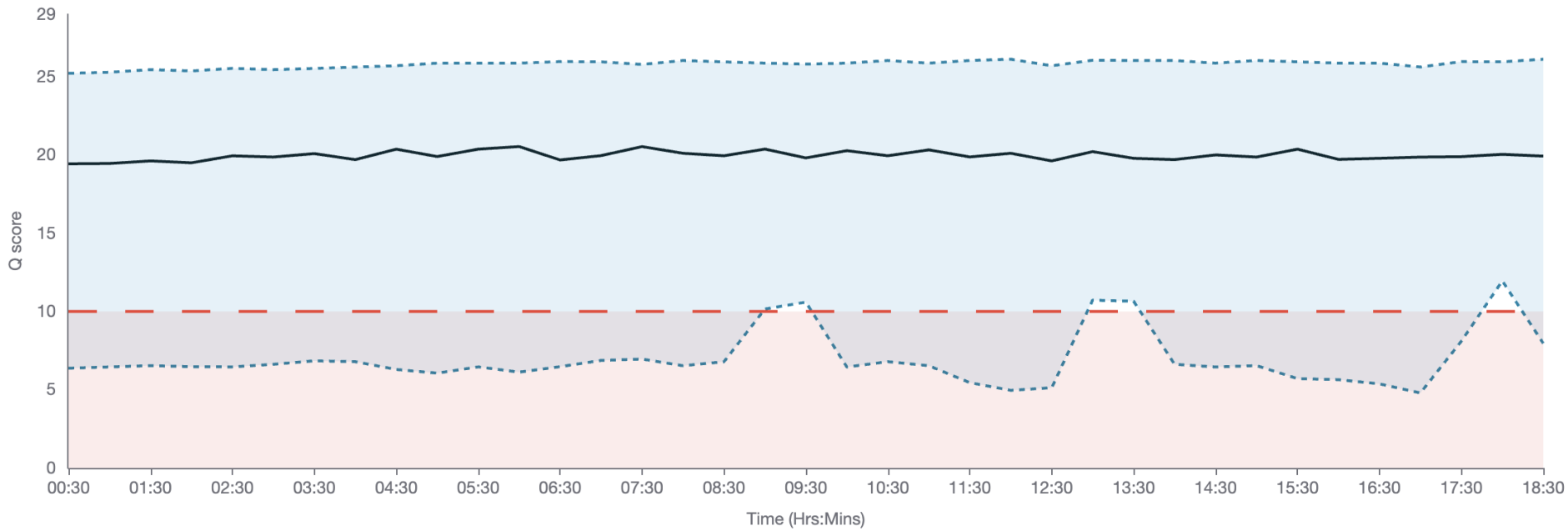
The most frequent quality score of reads in the run.

Spread

The spread of quality scores, found by calculating full width half maximum.

Min. quality score

Minimum quality score to be accepted as a passed read.



Sequencing run report- Pore Activity

PORE ACTIVITY

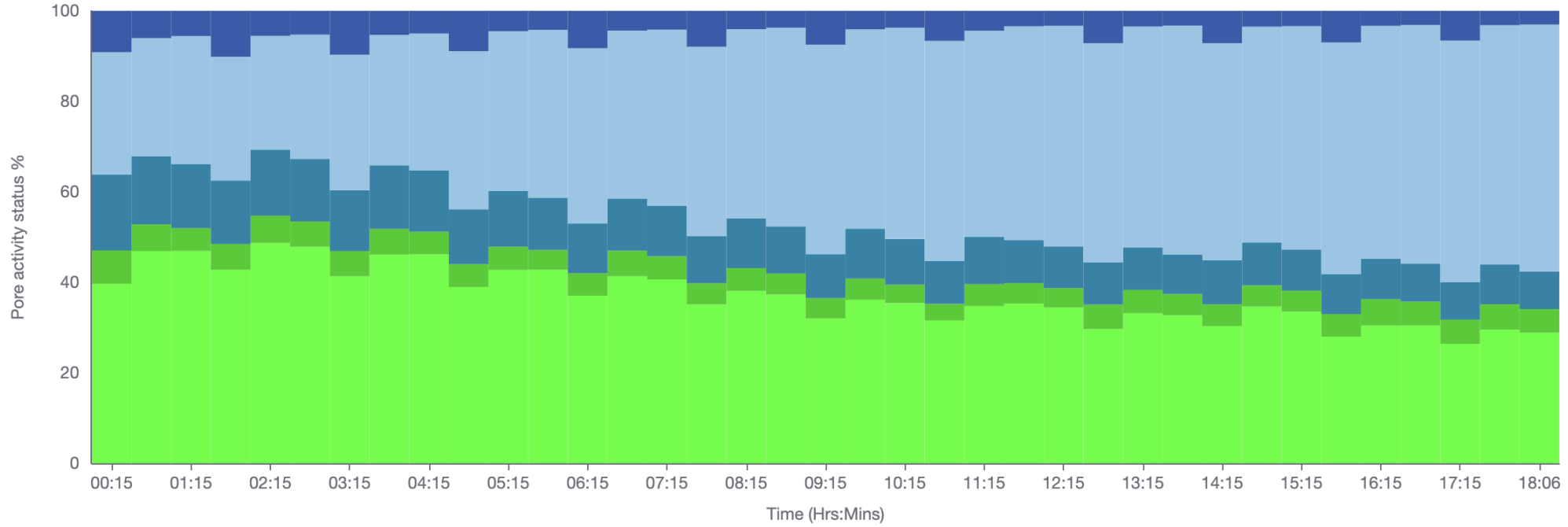
The Pore activity graph shows the performance of your sample as it is being sequenced during a run.

Show grouped

Legend

- Sequencing — Pore currently sequencing
- Pore available — Pore available for sequencing
- Unavailable — Pore currently unavailable for sequencing
- Inactive — Pore no longer suitable for further sequencing
- Unclassified — Pore status unknown

Aim: Pore availability to stay constant for first 24 hours of run



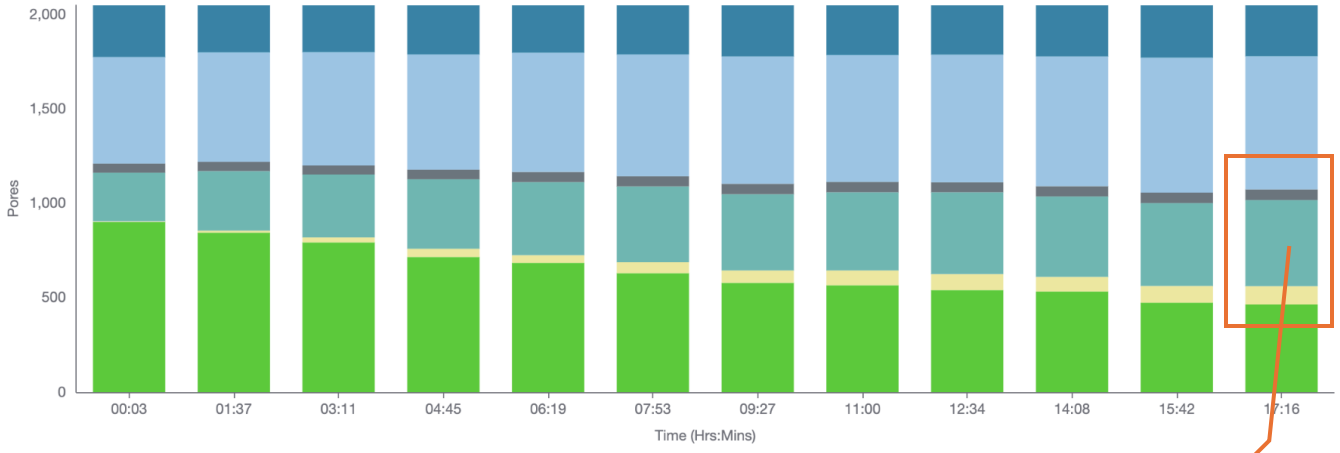
Sequencing run report- Pore Scan

PORE SCAN

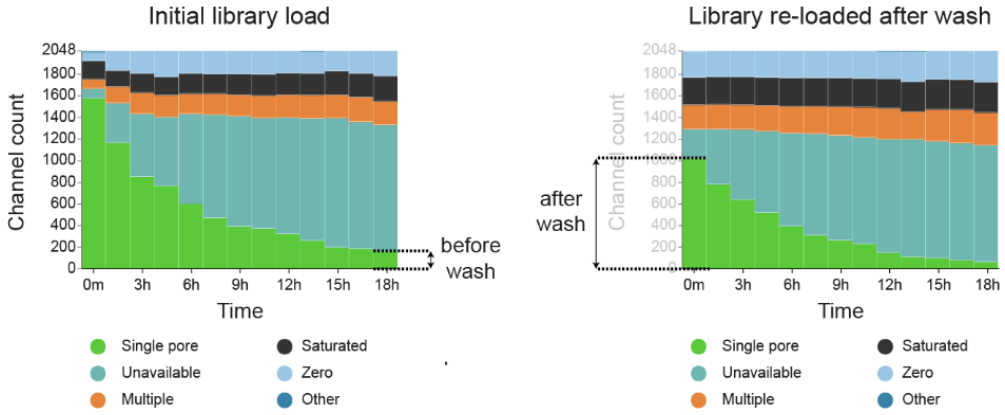
A Pore scan is performed at configurable time intervals to determine the current status of pores within channels on a Flow Cell. For this run a Pore scan is performed every 1.5 hrs.

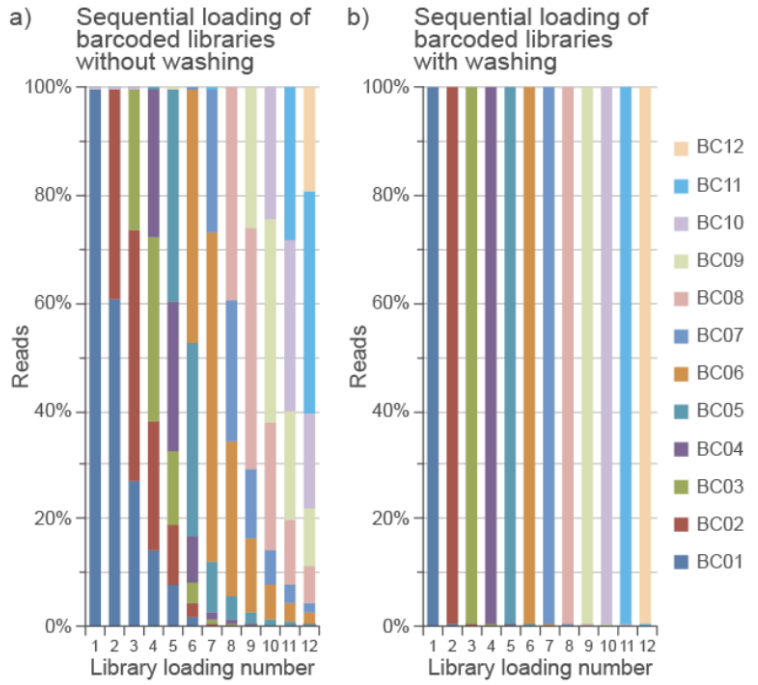
Legend

Pore available	Reserved pore	Unavailable	Saturated	Zero	Inactive
Pore in channel available for sequencing	Pore in reserve, will return to available when required	Pore inhibited from sequencing	Possible contamination in the sample	No current is passing through this pore, possibly due to bubbles on the membrane	Pore no longer suitable for further sequencing

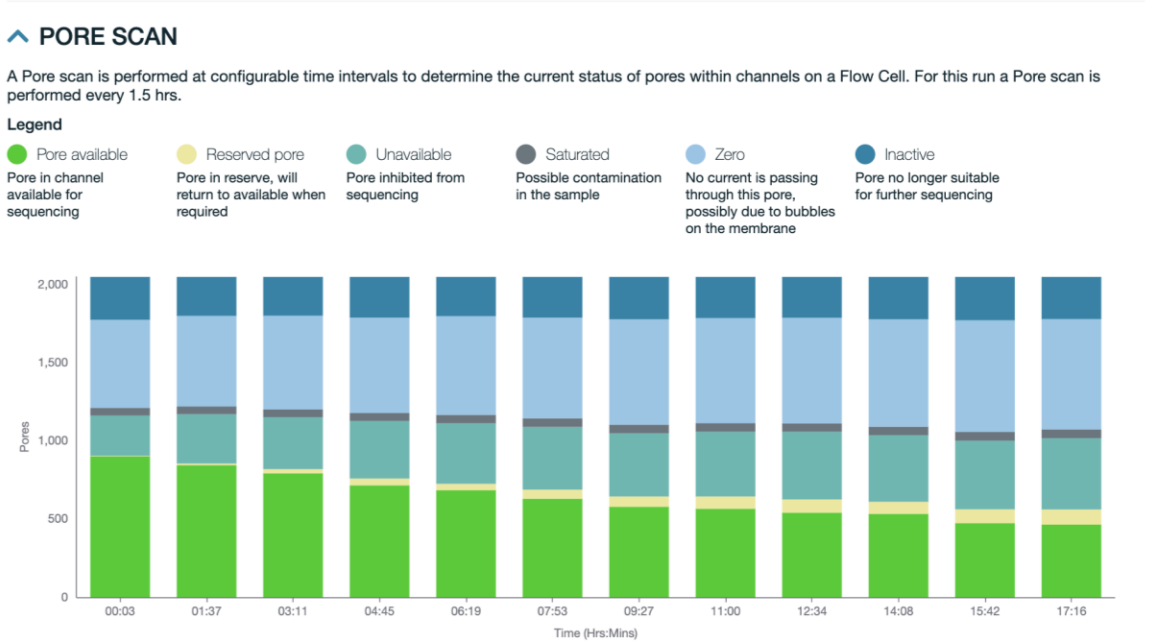


Unavailable pores: Can become available again after nuclease wash





Washing and reusing flow cells



Flow Cell Wash Kit (EXP-WSH004 or EXP-WSH004-XL)

Oxford NANOPORE Technologies

Version: WFC_9120_v1_revQ_08Dec2020
Last update: 4/25/2024

Kit batch number ..... **Flow cell number** ..... **DNA Samples** .....

Checklist: Flushing a MinION/GridION Flow Cell

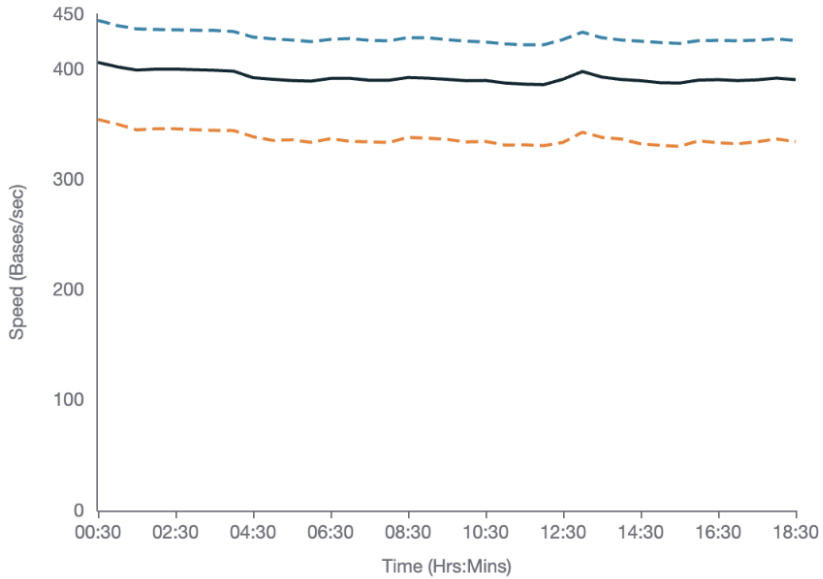
Materials	Consumables	Equipment
☐ Flow Cell Wash Kit (EXP-WSH004) or Flow Cell Wash Kit XL (EXP-WSH004-XL)		☐ P1000 pipette and tips ☐ P20 pipette and tips ☐ Ice bucket with ice

TRANSLOCATION SPEED

The translocation speed is the rate at which DNA/RNA travels through pores as it is being sequenced.

Legend

Median
75% quartile
25% quartile
Acceped range

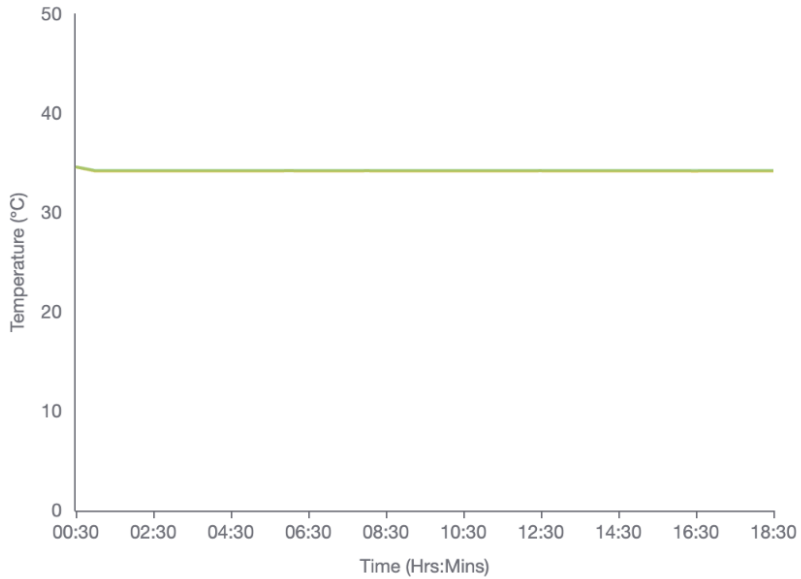


TEMPERATURE

The temperature of the Flow Cell over the run time.

Legend

Measured
Target



Mpox genomics recap

Source Details	Barcode	PCR DNA pool 1 ng/ul	PCR DNA pool 2 ng/ul	final library ng/ul
sample1	barcode41	7	16	
sample2	barcode42	41	46	
sample3	barcode43	5	11	
sample4	barcode44	23	26	
sample5	barcode45	11	11	
sample6	barcode46	30	45	
sample7	barcode47	5	5	
NTC	barcode48	6	12	38

Seqkit Stats to generate read stats

https://bioinf.shenwei.me/seqkit/

```
simple statistics of FASTA/Q files

Columns:

  1.  file      input file, "-" for STDIN
  2.  format    FASTA or FASTQ
  3.  type      DNA, RNA, Protein or Unlimit
  4.  num_seqs  number of sequences
  5.  sum_len   number of bases or residues       , with gaps or spaces counted
  6.  min_len   minimal sequence length           , with gaps or spaces counted
  7.  avg_len   average sequence length           , with gaps or spaces counted
  8.  max_len   miximal sequence length           , with gaps or spaces counted
  9.  Q1        first quartile of sequence length , with gaps or spaces counted
  10. Q2        median of sequence length         , with gaps or spaces counted
  11. Q3        third quartile of sequence length , with gaps or spaces counted
  12. sum_gap   number of gaps
  13. N50       N50. https://en.wikipedia.org/wiki/N50,_L50,_and_related_statistics#N50
  14. N50_num   N50_num or L50. https://en.wikipedia.org/wiki/N50,_L50,_and_related_statistics#L50
  15. Q20(%)    percentage of bases with the quality score greater than 20
  16. Q30(%)    percentage of bases with the quality score greater than 30
  17. AvgQual   average quality
  18. GC(%)     percentage of GC content
  19. sum_n     number of ambitious letters (N, n, X, x)
```

Source Details	PCR DNA pool 1 ng/ul	PCR DNA pool 2 ng/ul	final library ng/ul	QPCR Ct values	num_seqs	sum_len	N50	Q20(%)	Q30(%)	depth
sample1	7	16		25.6	40,871	23,601,986	1047	83	74.11	118.01
sample2	41	46		21.9	42,570	30,341,778	1233	83.04	74.39	151.71
sample3	5	11		Not detected	6,365	2,029,810	830	76.62	62.16	10.15
sample4	23	26		21	69,823	47,729,766	1245	84.77	76.02	238.65
sample5	11	11		29.1	47,565	15,985,503	529	83	70.01	79.93
sample6	30	45		19	71,881	52,654,060	1365	81.48	72.95	263.27
sample7	5	5		Not detected	3,918	518,858	425	55.61	39.69	2.59
NTC	6	12	38	N/A	8,848	3,008,851	808	74.93	62.26	15.04

Source Details	PCR DNA pool 1 ng/ul	PCR DNA pool 2 ng/ul	final library ng/ul	QPCR Ct values	Kraken Hit 1	% reads	Kraken Hit 2	% reads	Kraken Hit 3	% reads
sample1	7	16		25.6	**Monkeypox virus**	41.8	Photobacterium leiognathi	0.6	Vaccinia virus	0.1
sample2	41	46		21.9	**Monkeypox virus**	64.4	Homo sapiens	1.6	Photobacterium leiognathi	0.6
sample3	5	11		Not detected	Photobacterium leiognathi	1.4	Homo sapiens	0.1	Streptomyces sp. NBC_01420	0.0
sample4	23	26		21	**Monkeypox virus**	79.7	Homo sapiens	2.8	Vaccinia virus	0.3
sample5	11	11		29.1	Homo sapiens	46.4	Monkeypox virus	6.7	Streptomyces sp. NBC_01353	3.1
sample6	30	45		19	**Monkeypox virus**	80.2	Homo sapiens	4.1	Photobacterium leiognathi	0.5
sample7	5	5		Not detected	Photobacterium leiognathi	0.4	Monkeypox virus	0.0	Escherichia coli	0.0
NTC	6	12	38	N/A	Photobacterium leiognathi	0.6	Homo sapiens	0.1	Streptomyces sp. NBC_01420	0.0

Mpox primers to take home

- ARTIC INRB pool1 and pool2
- 100 uM stock of pooled primers
- 30 ul volume
- PLEASE store them at -20°C freezer as soon as possible
  - In the freezer in your hotel room
  - In your freezer at home
  - In your freezer at the lab

		Concentration ng/ul					
	Mpox QPCR Ct	pool1 from training	pool2 from training	pool1 -20oC	pool2 -20oC	pool1 RT O/N then -20oC	pool2 RT O/N then -20oC
sample1	25.6	7	16	12	25	4	18
sample2	21.9	41	46	44	50	23	44
sample3	Not detected	5	11	3	4	2	3
sample4	21	23	26	36	38	29	35
sample5	29.1	11	11	9	9	7	8
sample6	19	30	45	43	49	42	43
sample7	Not detected	5	5	3	4	4	2
NTC	N/A	6	12	3	4	4	2