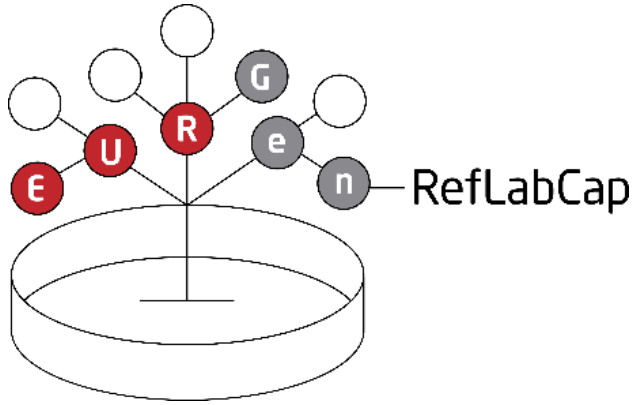
EURGen-RefLabCap

Best practice workshop – Nanopore sequencing

Fourth day (virtual)

Friday, 15 December 2023

9:00 - 12:30 CET



Funded by the European Union

Previously…

Four-day best practice workshop for Nanopore training

First day: Thursday 20 November (virtual) - Introduction to Nanopore sequencing

Second day: Wednesday 6 December at DTU - DNA extraction for Nanopore sequencing

Third day: Thursday 7 December at DTU - Library preparation and loading ONT flowcells

Fourth day: Friday 15 December (virtual) - Analysis and interpretation of ONT data

Funded by the European Union

Agenda for today

Fourth day (virtual) – Friday 15 December 2023, 9:00 - 12:30 CET

9:00 – 9:10: Introduction and agenda for the day (Ana Rita Rebelo, DTU)

9:10 – 9:30: What happened at DTU after the workshop (Ana Rita Rebelo, DTU)

9:30 – 9:50: Quality control of ONT sequence data (Niamh Lacy-Roberts, DTU)

9:50 – 10:00: Coffee break

10:00 – 10:20: Assembly of ONT sequence data and quality control (Niamh Lacy-Roberts, DTU)

10:20 – 10:50: Bioinformatics analyses of ONT data using an internal pipeline (Astrid Rasmussen, SSI)

10:50 – 11:20: Cluster analysis with ONT data (Astrid Rasmussen, SSI)

11:20 – 11:30: Coffee break

11:30 – 12:00: Updates to the harmonized EURGen-RefLabCap WGS protocols (Ana Rita Rebelo and Jette Sejer Kjeldgaard, DTU)

12:00 –12:30: Q&A, wrap-up (Ana Rita Rebelo, DTU)

Funded by the European Union

Ana Rita Rebelo
anrire@food.dtu.dk

Pending from previous days

Funded by the European Union

About the bubbles in the flowcell

https://help.nanoporetech.com/en/articles/6639683-how-to-proceed-if-my-flow-cell-has-a-bubble

Bubbles present in the inlet channel:

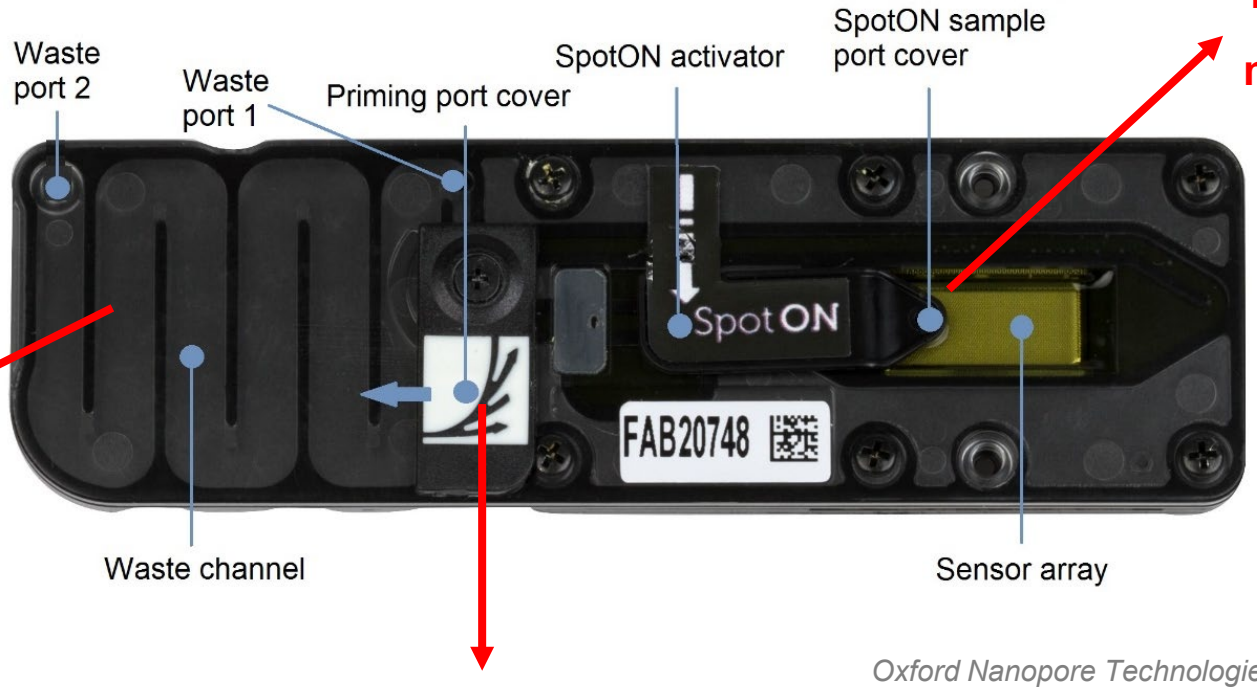
These bubbles will not be visible on an intact flow cell, but can be removed by re-drawing about 15 - 20 uL of buffer from the flow cell (via the priming port). There is enough buffer in the outlet channel so that even if you remove about 15 uL you will still have the flow cell covered at all times with liquid.

Monitor the sensor array while dialing up to ensure that the pores stay covered in buffer. If bubbles in the inlet channels are positioned in a corner and do not move by withdrawing liquid, it is best practice to leave them there and be careful while loading the flow cells so that the bubble does not get dislodged.

Oxford Nanopore Technologies – November 2023

Funded by the European Union

About the bubbles in the flowcell



The SpotON port cover must be closed !!!!

Bubbles in the waste channel do not matter.

Pipette back via priming port

Oxford Nanopore Technologies, Oxford, United Kingdom

Funded by the European Union

Ana Rita Rebelo
anrire@food.dtu.dk

What happened at DTU after the workshop

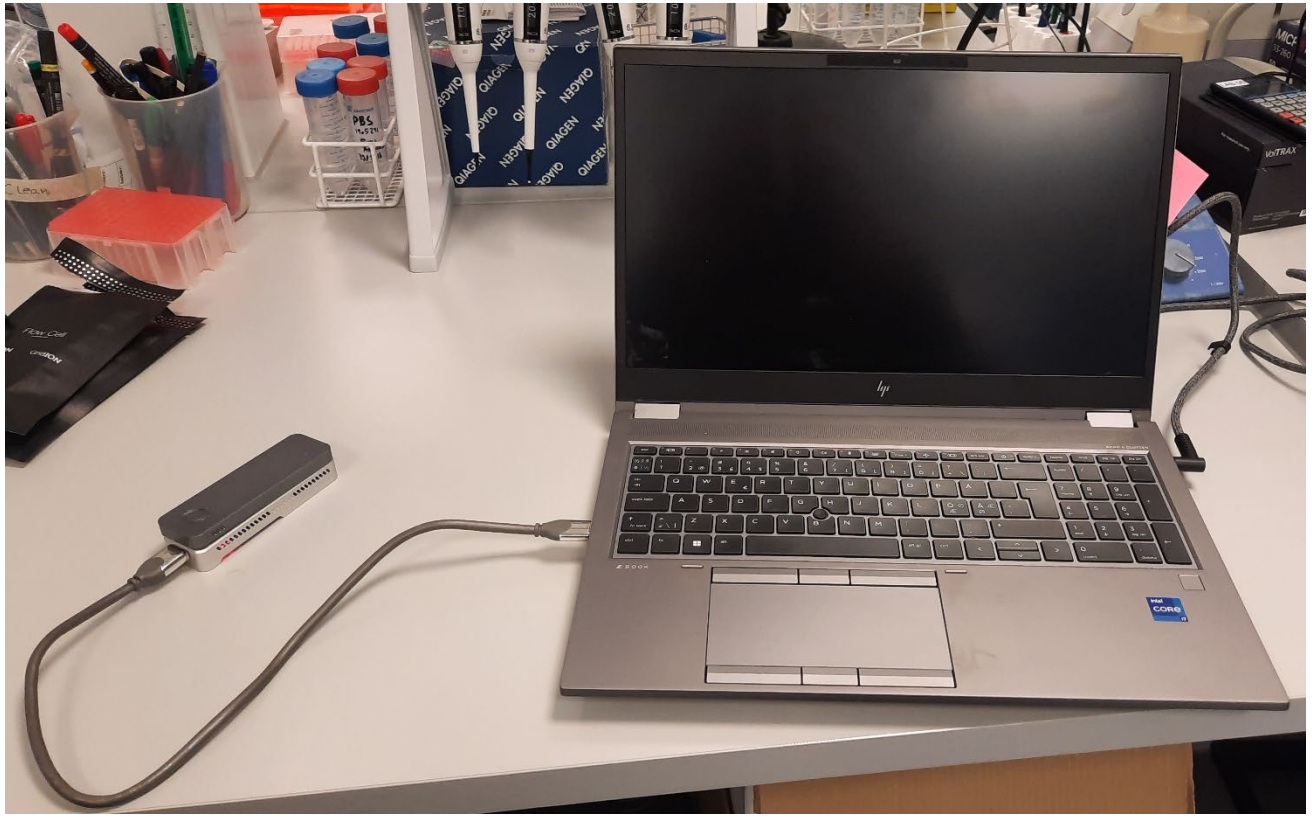
Funded by the European Union

During the course

We loaded one flowcell in the MinION in the classroom

Group 5 (FAW16535)

Samples 49-57



Funded by the European Union

After the course (same day)

We loaded four flowcells in the GridION

Group 1 (FAW16682)

Samples 1-12

Group 4 (FAX59268)

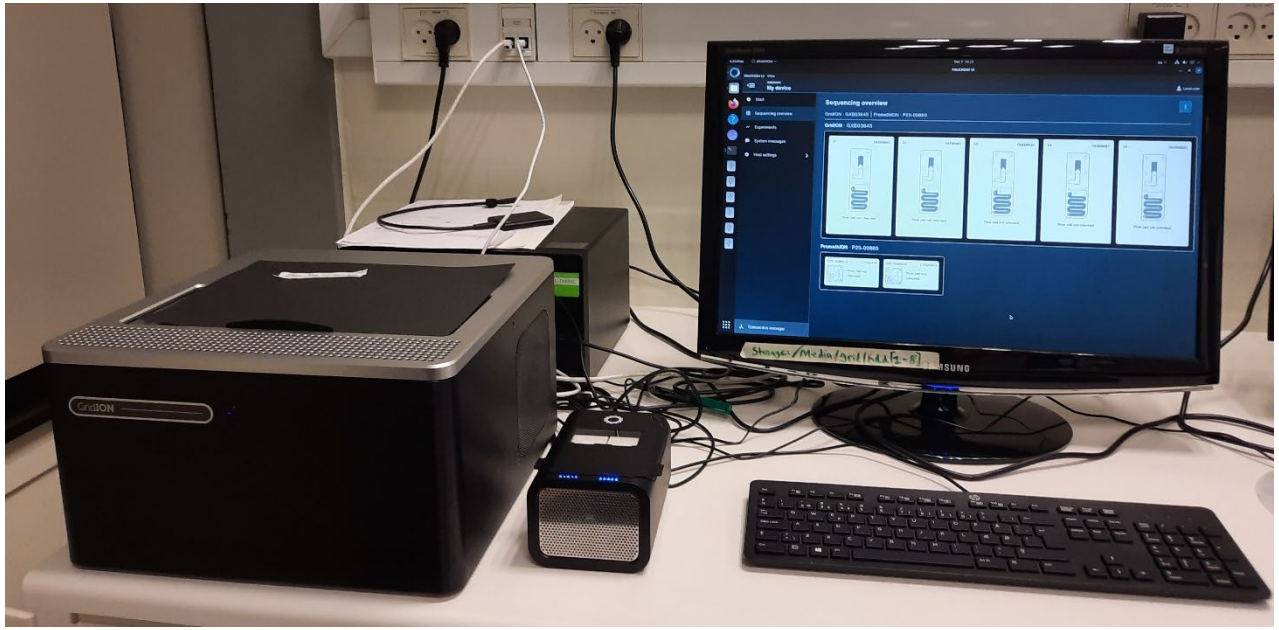
Samples 37-48

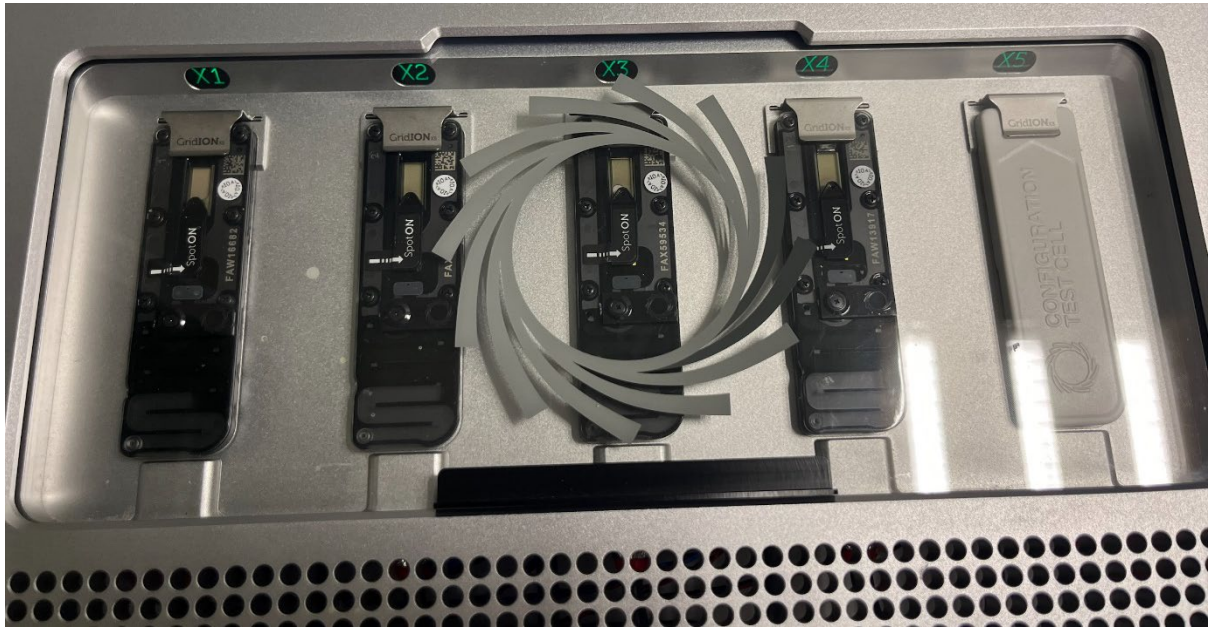
Group 2 (FAX59534)

Samples 13-24

Group 3 (FAW13917)

Samples 25-36

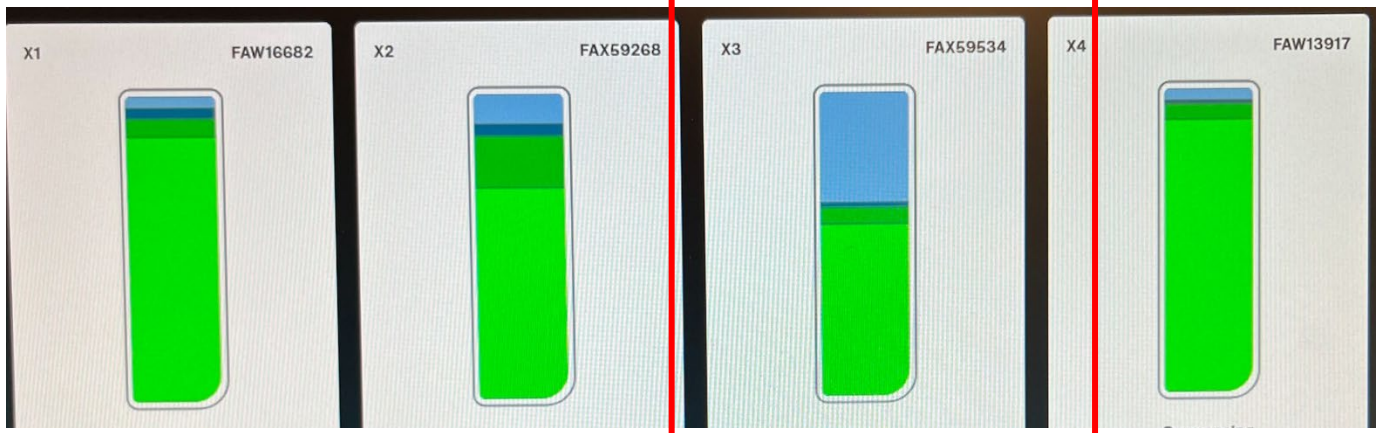




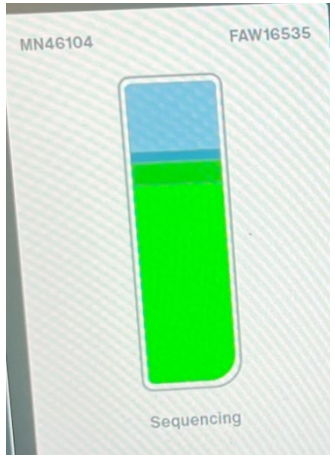
Funded by the European Union

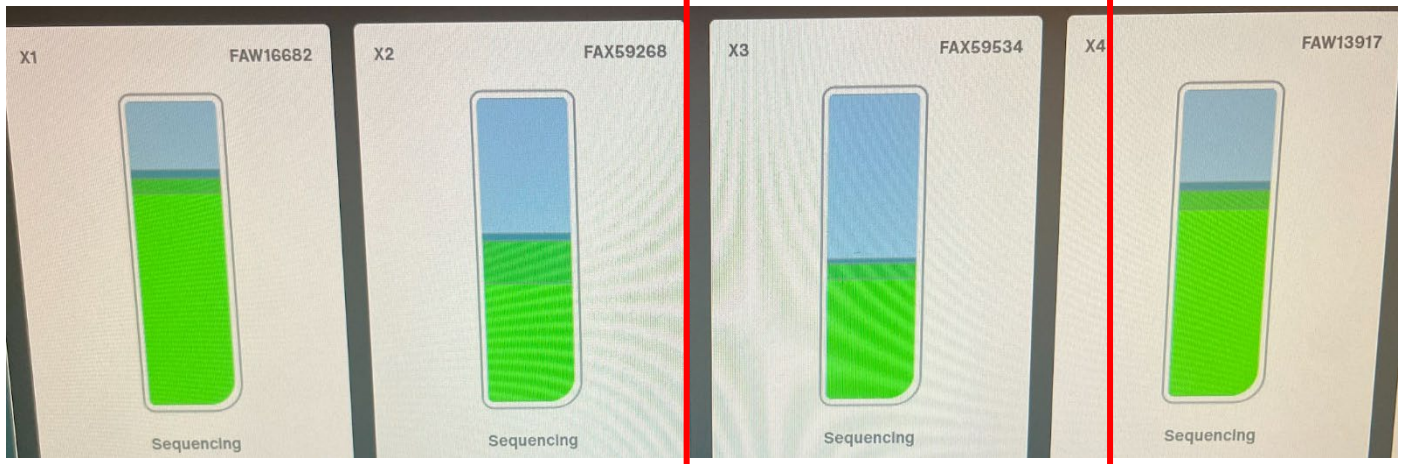
Sequencing progress

First day



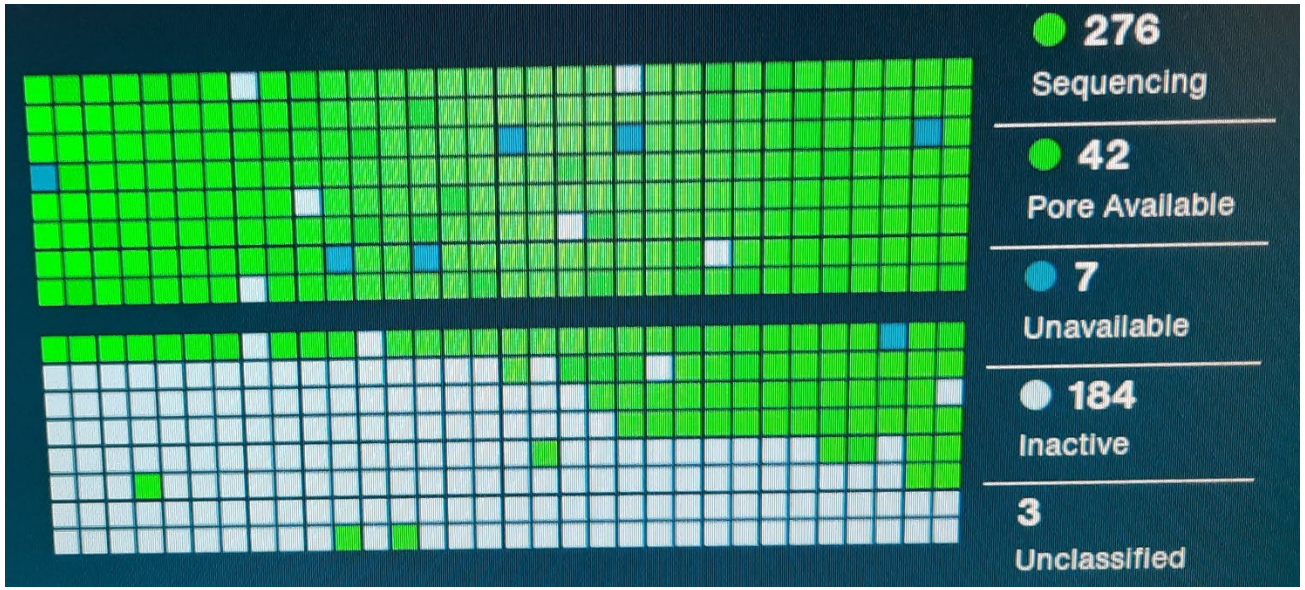
Second day





Funded by the European Union

Bubble!



Funded by the European Union

Brief overview of data quality

Run summary stats		Group 1 (FAW16682)	Group 2 (FAX59534)	Group 3 (FAW13917)	Group 4 (FAX59268)	Group 5 (FAW16535)
Data output	Estimated bases	27.44 Gb	17.01 Gb	19.37 Gb	12.6 Gb	23.9 Gb
	Reads generated	4.56 M	1.61 M	2.92 M	1.67 M	2.94 M
	Estimated N50	12.58 kb	23.37 kb	15.09 kb	17.99 kb	18.77 kb
	Total data produced (pass/ fail)	320.35 GB	199.86 GB	228.67 GB	155.91 GB	248.31 GB
Basecalling (min Q score: 10)	Reads called	100%	100%	100%	100%	91.53%
	Bases called Pass	23.31 Gb	14.26 Gb	16.9 Gb	10.35 Gb	18.57 Gb
	Bases called Fail	3.73 Gb	2.4 Gb	1.85 Gb	2.93 Gb	3.52 Gb

⇩ Still OK with bubble

⇩ 9 samples multiplexed

Funded by the European Union

Brief overview of data quality

See excel

Funded by the European Union

Niamh Lacy-Roberts
nlac@food.dtu.dk

Quality control of ONT sequence data

Funded by the European Union

QC for ONT data

- Before beginning any post sequencing analyses, QC assessments is a good way to understand if your data meets certain requirements or matches your expectations from the sequencing run.

- While live basecalling, MinKNOW provides real-time feedback, such as read quality and length. This information is presented in an interactive run report and can be exported during and after the sequencing run.

- Additional QC metrics can be generated post-sequencing using one of several tools.

- Command line tools exist such as PycoQC which can also be used to generate QC metrics from the `sequencing_summary.txt` file, or NanoPlot which generates summary statistics from the fastq files.

- You can run EPI2ME workflows on the cloud or locally on your computer.

Read quality, read length and N50

- bp = base pair
- kb (= kbp) = kilo–base-pair = 1,000 bp
- Mb (= Mbp) = mega–base-pair = 1,000,000 bp
- Gb (= Gbp) = giga–base-pair = 1,000,000,000 bp

- Nanopore technology routinely generates sequencing reads that are tens of kilobases in length

- The longest DNA fragment sequenced to date using nanopore technology is 4.2 Mb, which was achieved using the Ultra-Long DNA Sequencing Kit.

- N50 — the length at which half of the nucleotides in the fastq/assembly belong in reads/contigs of this length or longer.

- We set the Q score to > 10 - the basecalling quality of each base in the sequence.

Read quality – Q scores

- Q scores are defined as a property that is logarithmically related to the base calling error probabilities $(P)^2$.
- $Q = -10 \log_{10} P$

Phred Quality Score	Probability of Incorrect Base Call	Base Call Accuracy
10	1 in 10	90%
20	1 in 100	99%
30	1 in 1000	99.9%
40	1 in 10,000	99.99%
50	1 in 100,000	99.999%

Let’s take a look at the output files and data reports… ☺

NanoStat

- https://github.com/wdecoster/nanostat
- Report from barcode1 group1:

- General summary:
- Mean read length: 8,475.2
- Mean read quality: 14.7
- Median read length: 4,474.0
- Median read quality: 16.7
- Number of reads: 160,000.0
- Read length N50: 17,338.0
- STDEV read length: 10,960.8
- Total bases: 1,356,026,721.0

NanoStat

- https://github.com/wdecoster/nanostat
- Report from barcode1 group1:

- Number, percentage and megabases of reads above quality cutoffs
- >Q5: 159998 (100.0%) 1356.0Mb
- >Q7: 159725 (99.8%) 1356.0Mb
- >Q10: 155931 (97.5%) 1353.7Mb
- >Q12: 145376 (90.9%) 1305.2Mb
- >Q15: 111514 (69.7%) 1121.5Mb

NanoStat

- https://github.com/wdecoster/nanostat
- Report from barcode1 group1:

- Top 5 highest mean basecall quality scores and their read lengths
- 1: 24.5 (48535)
- 2: 24.4 (24898)
- 3: 24.3 (11702)
- 4: 24.2 (15821)
- 5: 23.8 (8094)
- Top 5 longest reads and their mean basecall quality score
- 1: 141782 (12.7)
- 2: 130858 (20.4)
- 3: 125465 (18.2)
- 4: 120832 (17.8)
- 5: 119562 (19.6)

Niamh Lacy-Roberts
nlac@food.dtu.dk

Assembly of ONT sequence data and quality control

Funded by the European Union

Genome Assembly

- High-quality genome assemblies are crucial for their use as reliable reference sequences and downstream analyses.

- The short reads produced by traditional sequencing technologies lead to highly fragmented, incomplete assemblies.
  - Short reads cannot span important genomic regions such as repeats and structural variants, resulting in them being assembled incorrectly.

- In contrast, nanopore technology can deliver long and ultra-long sequencing reads that can span complex genomic regions, enabling the generation of highly contiguous genome assemblies.
  - ONT shows lack of bias in GC-rich regions
  - Can span repeat-rich sequences and structural variants
  - Possible to get complete assembly (one contig)

Command line tools

- Filtlong to QC reads before assembly.
  - Remove reads shorter than a certain length exclude the worst % of reads.

- Raven
  - Fast yet reliable assembly, less computationally expensive

- Flye
  - Slower but more accurate, more computationally expsive
  - Around 32 GB of RAM and 1 hour for most read sets

- Unicycler
  - Hybrid assembly, combining short and long reads
  - Your short-read set is deep but your long-read set is shallow
-
- Medaka
  - ONT assembly polishing tool

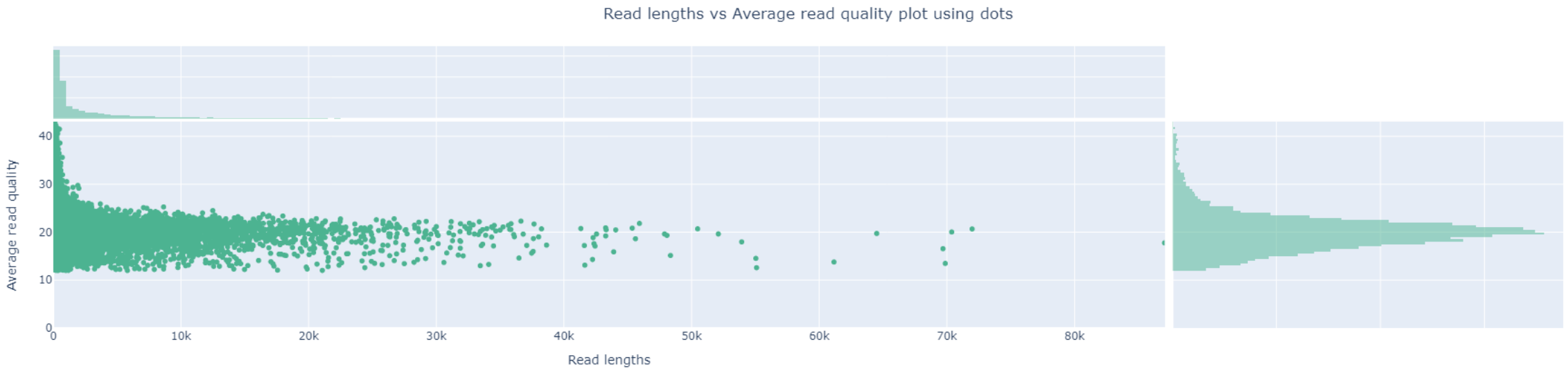
Epi2me labs

Astrid Rasmussen, SSI

Bioinformatics analyses of ONT data using an internal pipeline

Funded by the European Union

STATENS SERUM INSTITUT



Bioinformatic analyses of ONT data using an internal pipeline

STATENS SERUM INSTITUT

Astrid Rasmussen

MSc in Bioinformatics and Systems Biology

Bioinformatician at SSI since January 2022

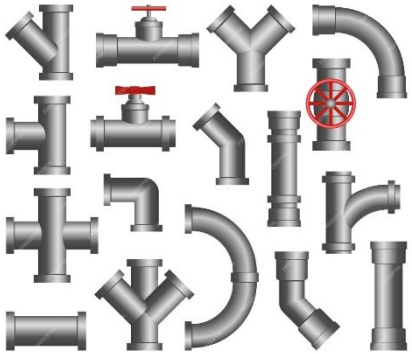
Main focus area is analysis of (Nanopore) sequencing data

Outline

STATENS SERUM INSTITUT

- Bioinformatic pipelines
- SSI pipeline for analysing ONT data from bacterial isolates
- SSI pipeline output
- Cluster analysis with ONT data

Bioinformatic pipelines

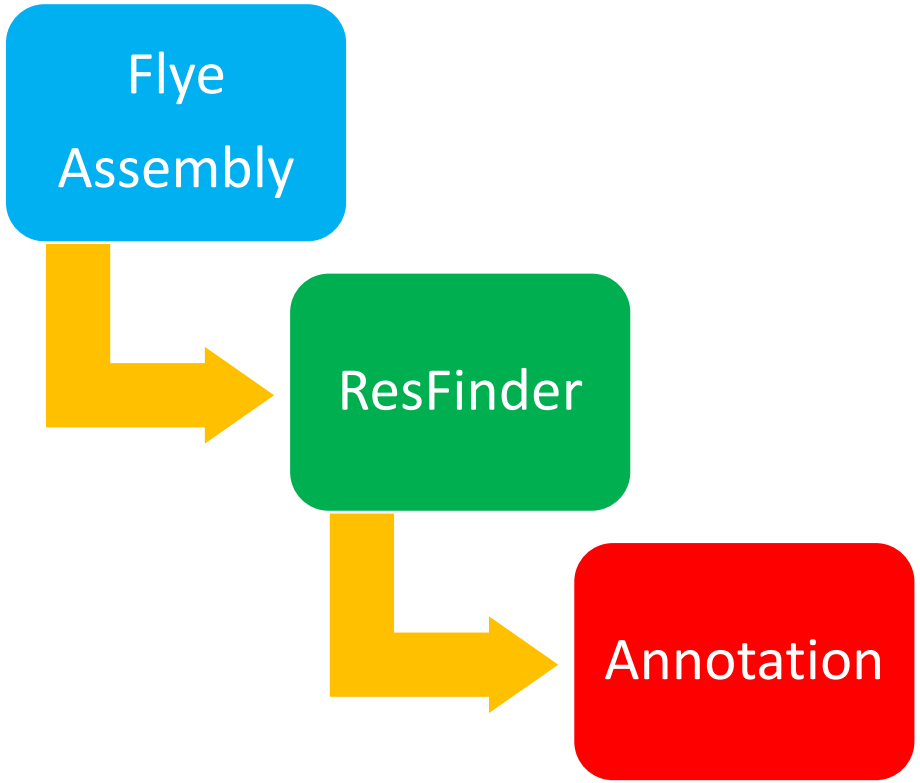


STATENS SERUM INSTITUT

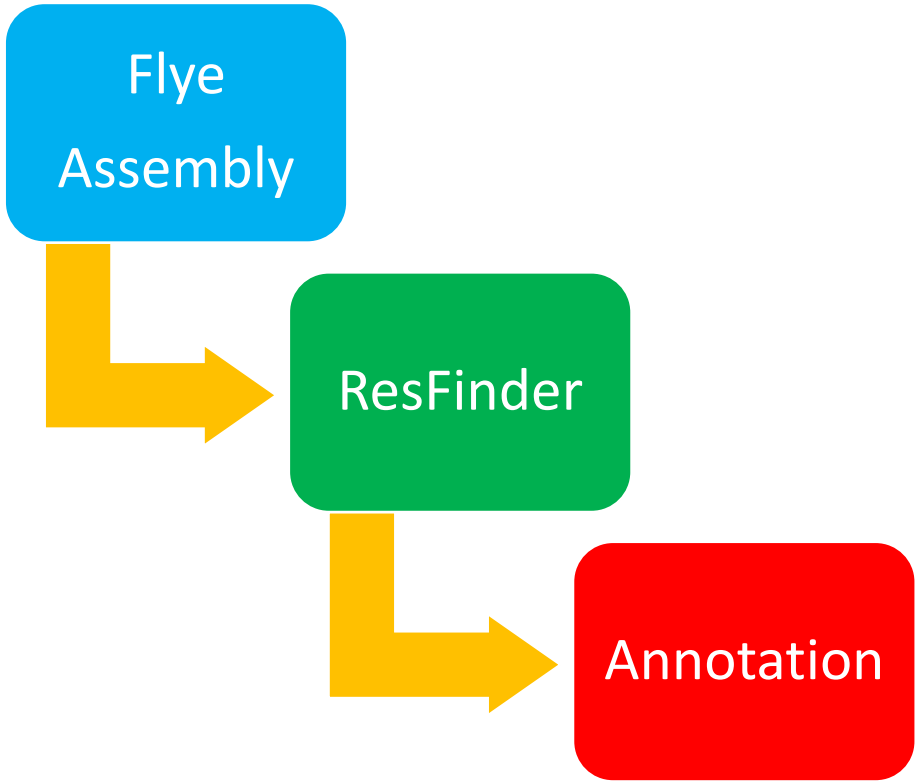
- A number of analyses connected into a series
- The output from one analysis is used as input for the next analysis
- A way to automatically perform multiple bioinformatic analyses in a row
- Requires no manual interference
- Minimizes human errors
- Promotes reproducibility (version control, logging)

Bioinformatic pipelines come with different levels of complexity

STATENS SERUM INSTITUT



Bioinformatic pipelines come with different levels of complexity

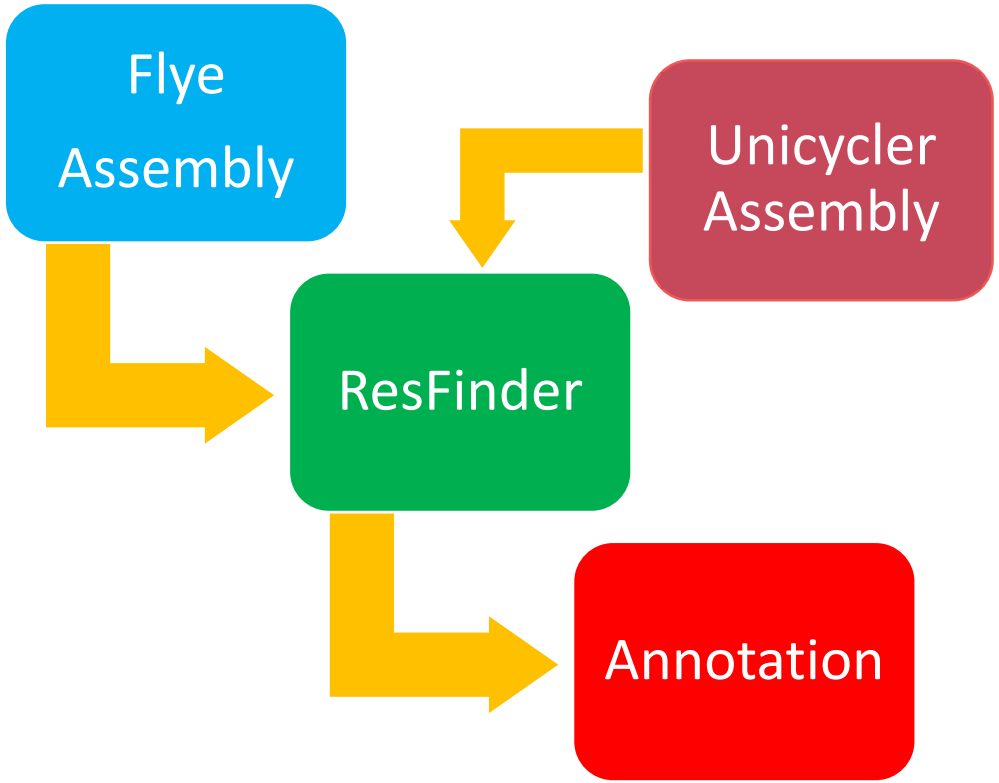


Command?

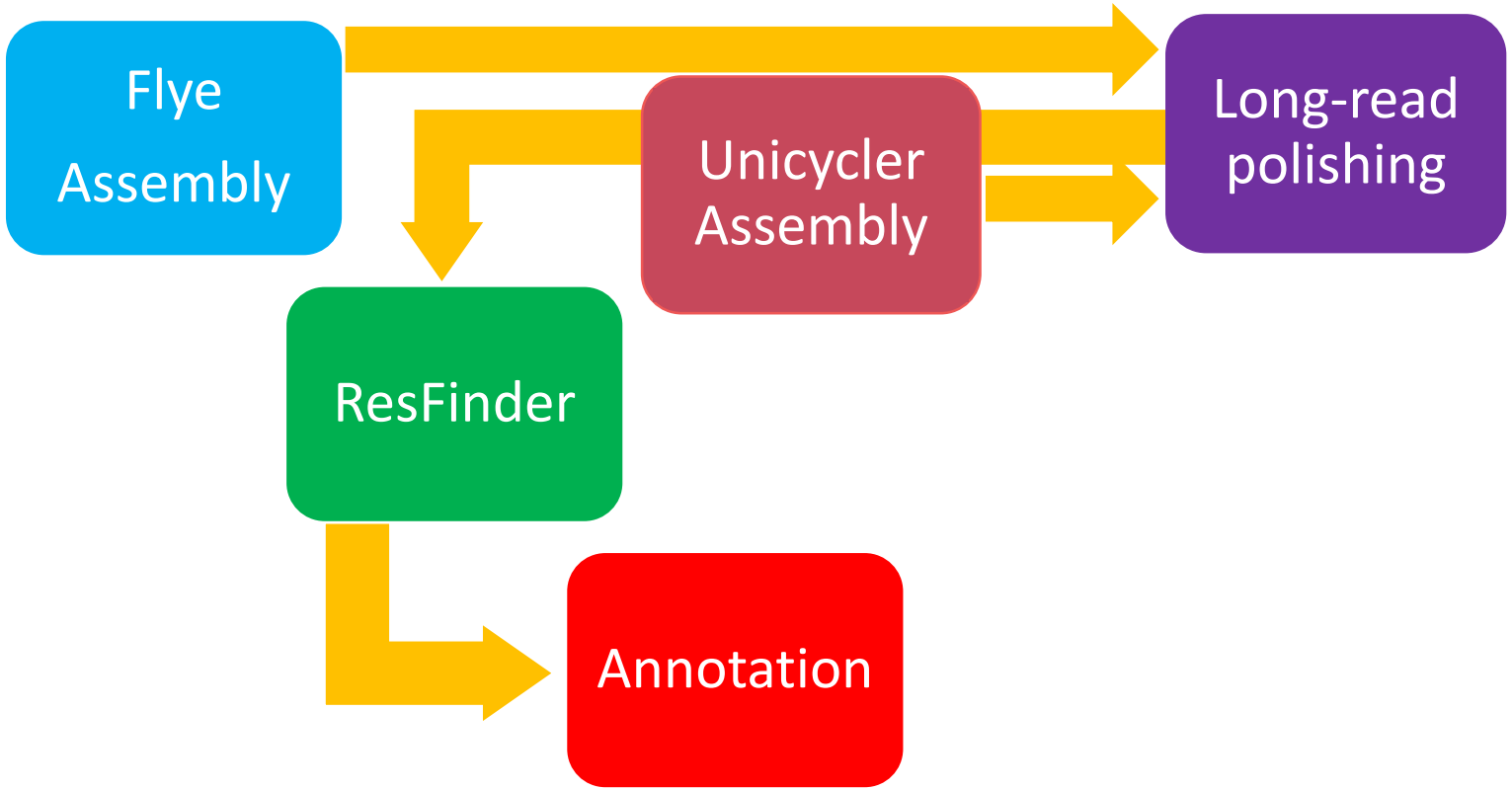
Version?

Database?

Bioinformatic pipelines come with different levels of complexity

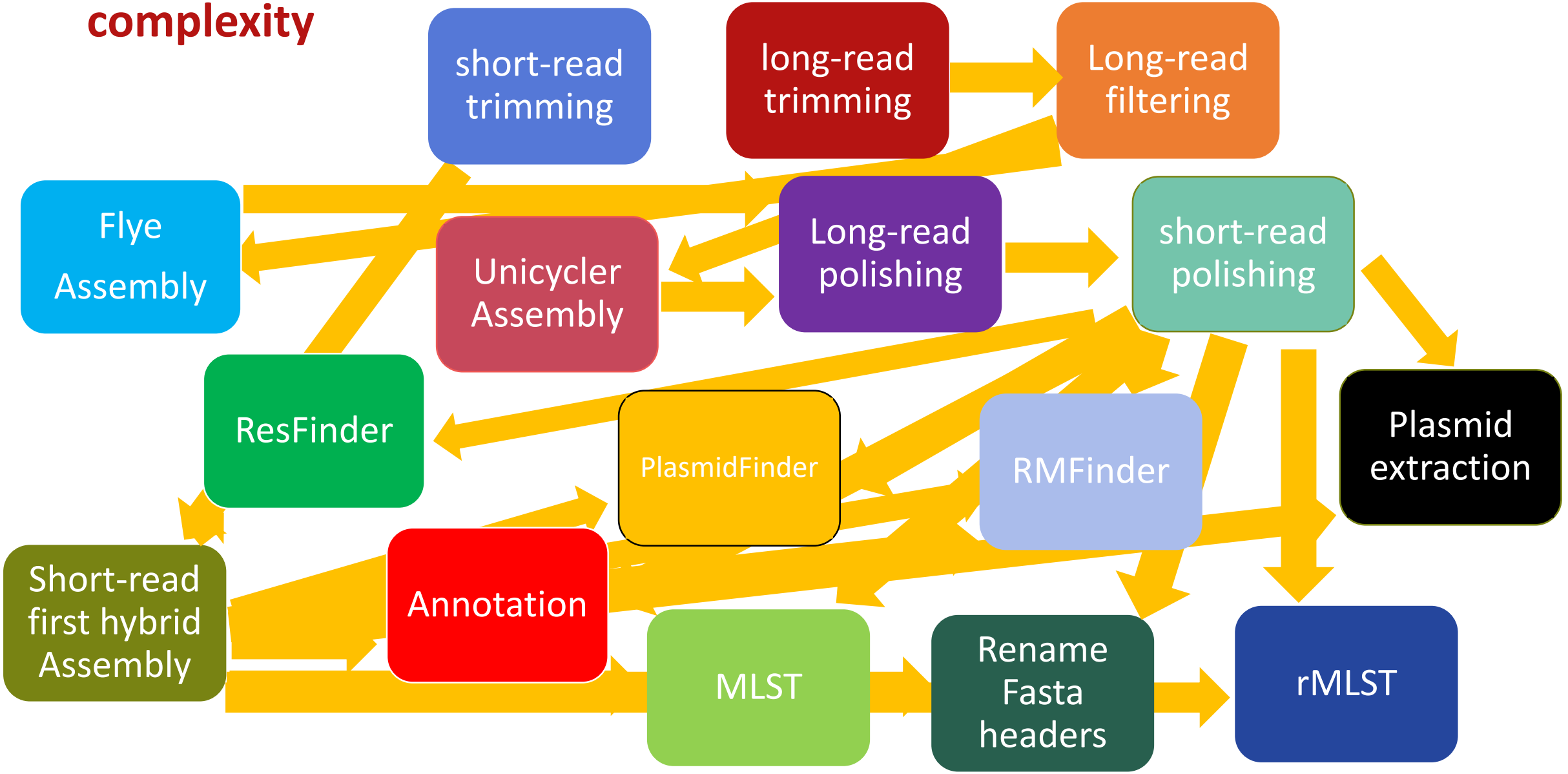


Bioinformatic pipelines come with different levels of complexity

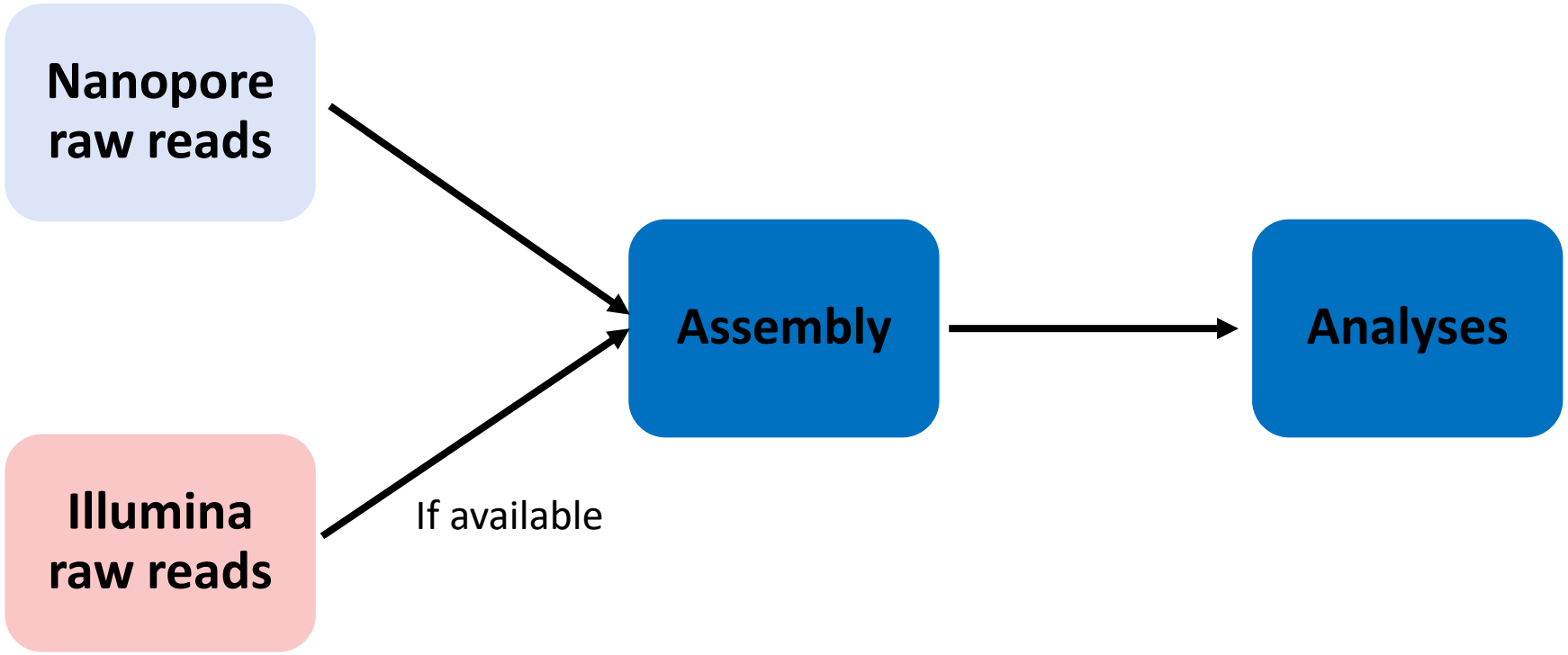


Bioinformatic pipelines come with different levels of complexity

STATENS SERUM INSTITUT

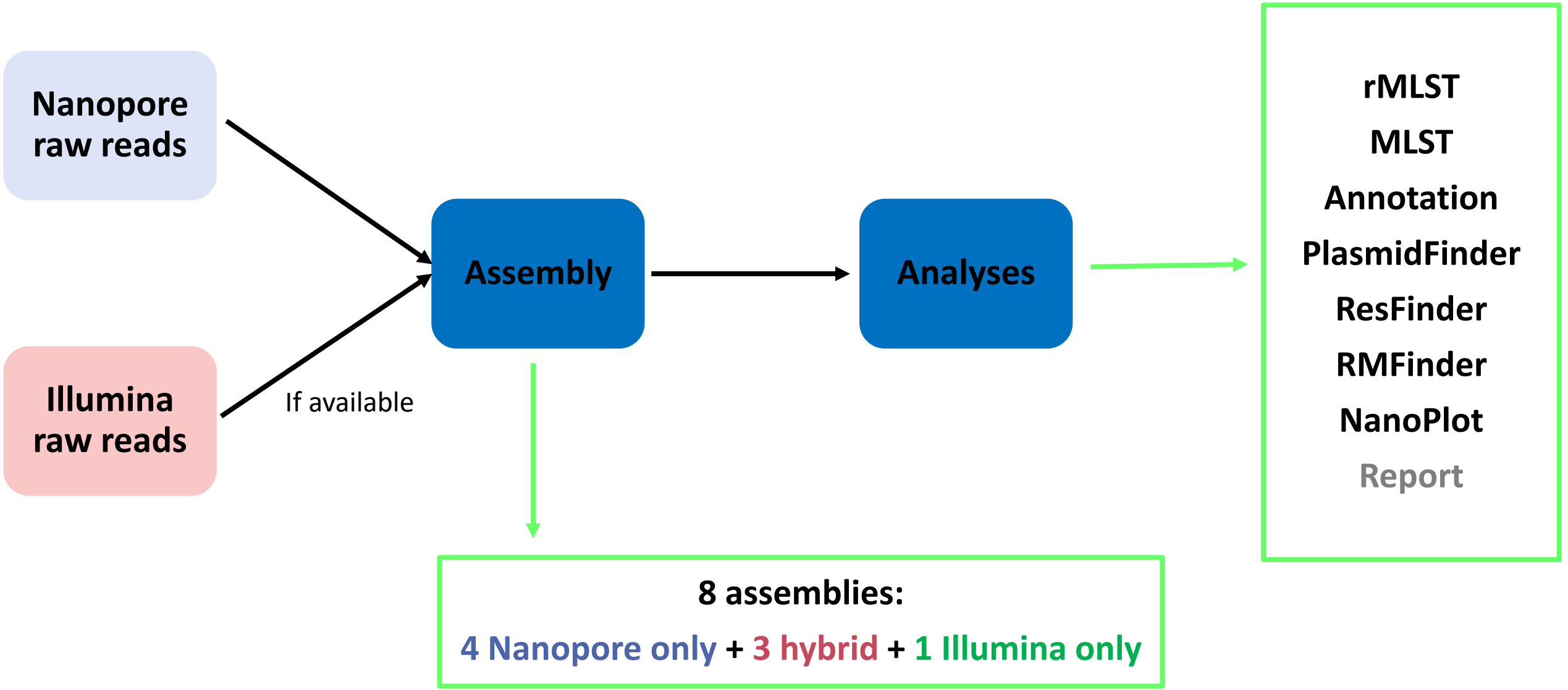


SSI internal pipeline



SSI internal pipeline

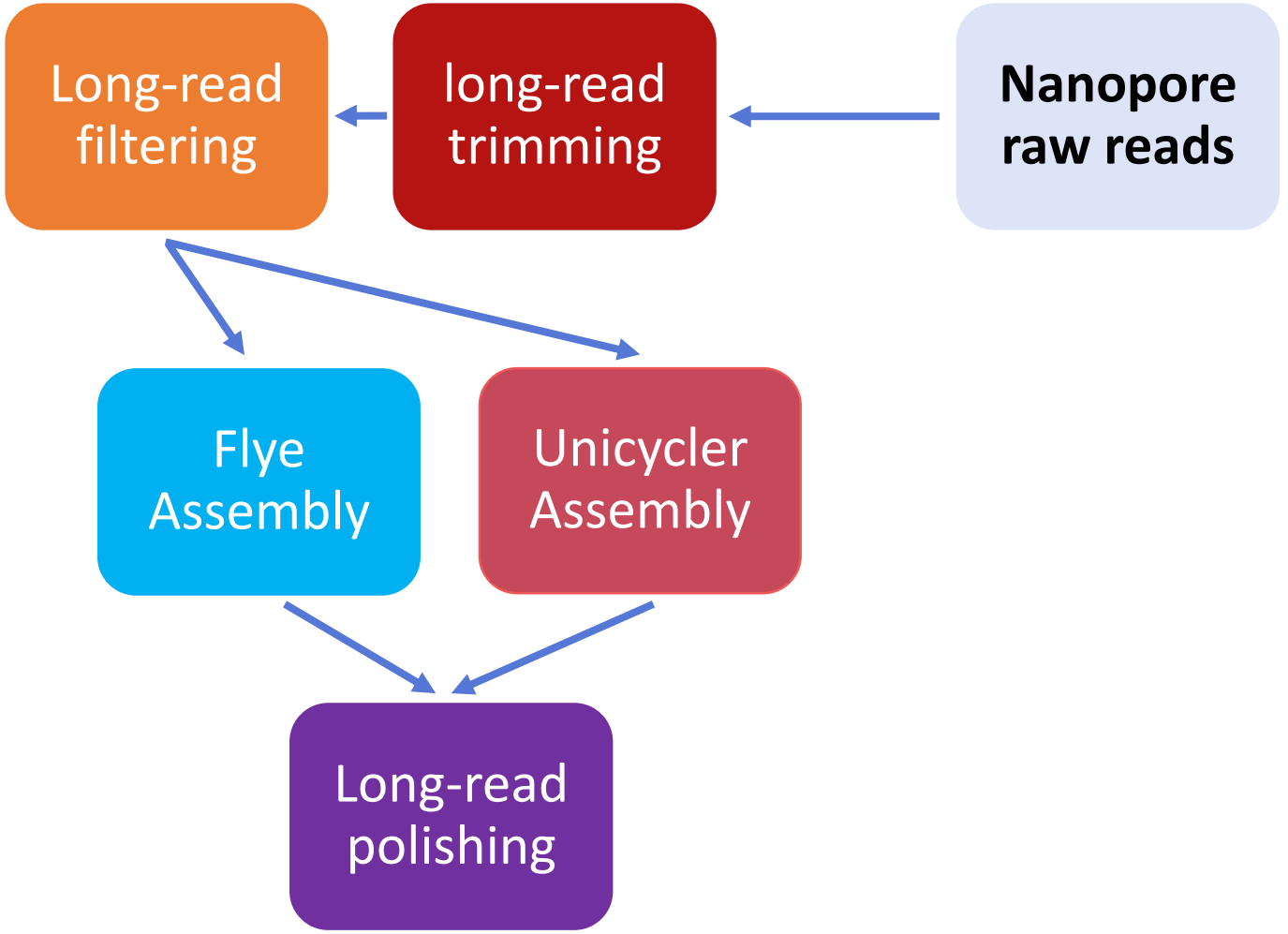
STATENS SERUM INSTITUT



SSI internal pipeline

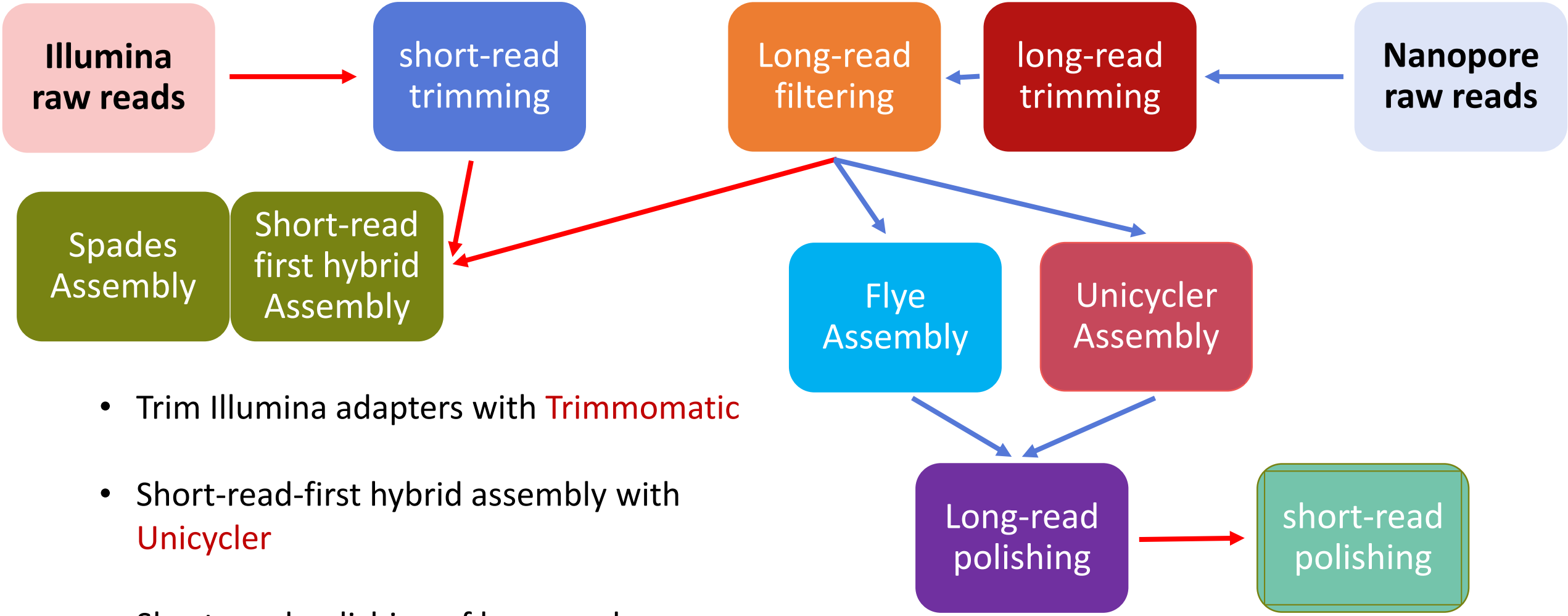
STATENS SERUM INSTITUT

- Trim barcodes with qcat
- Filter based on Q-score with Chopper
- Filter with FiltLong to max 500Mbp based on Q-score and read length
- Assemble with Flye and Unicycler
- Polish assemblies with Medaka



SSI internal pipeline

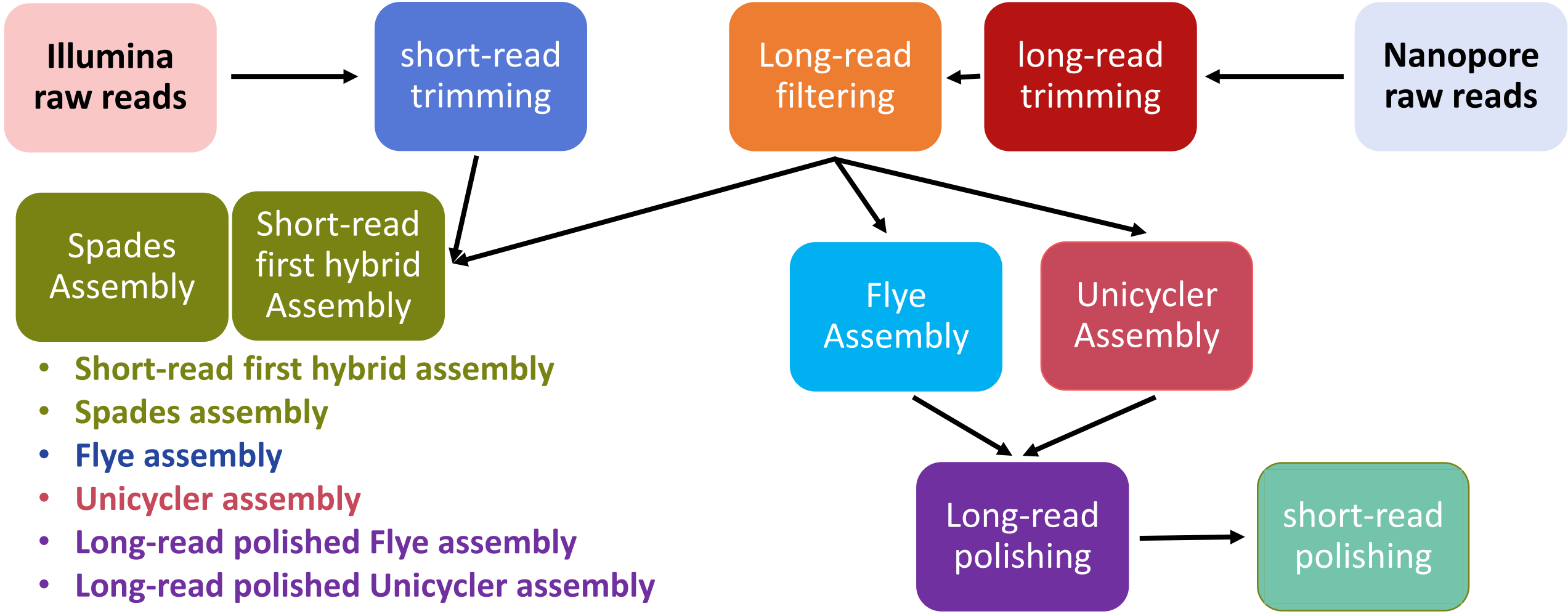
STATENS SERUM INSTITUT



- Trim Illumina adapters with Trimmomatic
- Short-read-first hybrid assembly with Unicycler
- Short-read polishing of long-read assemblies with Polypolish

SSI internal pipeline

STATENS SERUM INSTITUT



- **Short-read first hybrid assembly**
- **Spades assembly**
- **Flye assembly**
- **Unicycler assembly**
- **Long-read polished Flye assembly**
- **Long-read polished Unicycler assembly**
- **Short- and long-read polished Flye assembly**
- **Short- and long-read polished Unicycler assembly**

- Short-read first hybrid assembly
- Spades assembly
- Flye assembly
- Unicycler assembly
- Long-read polished Flye assembly
- Long-read polished Unicycler assembly
- Short- and long-read polished Flye assembly
- Short- and long-read polished Unicycler assembly

Short-read-first hybrid Assembly

Spades Assembly

Unicycler Assembly

Flye Assembly

Long-read polished Unicycler

Long-read polished Flye

Short- and long-read polished Unicycler

Short- and long-read polished Flye

Annotation

MLST

rMLST

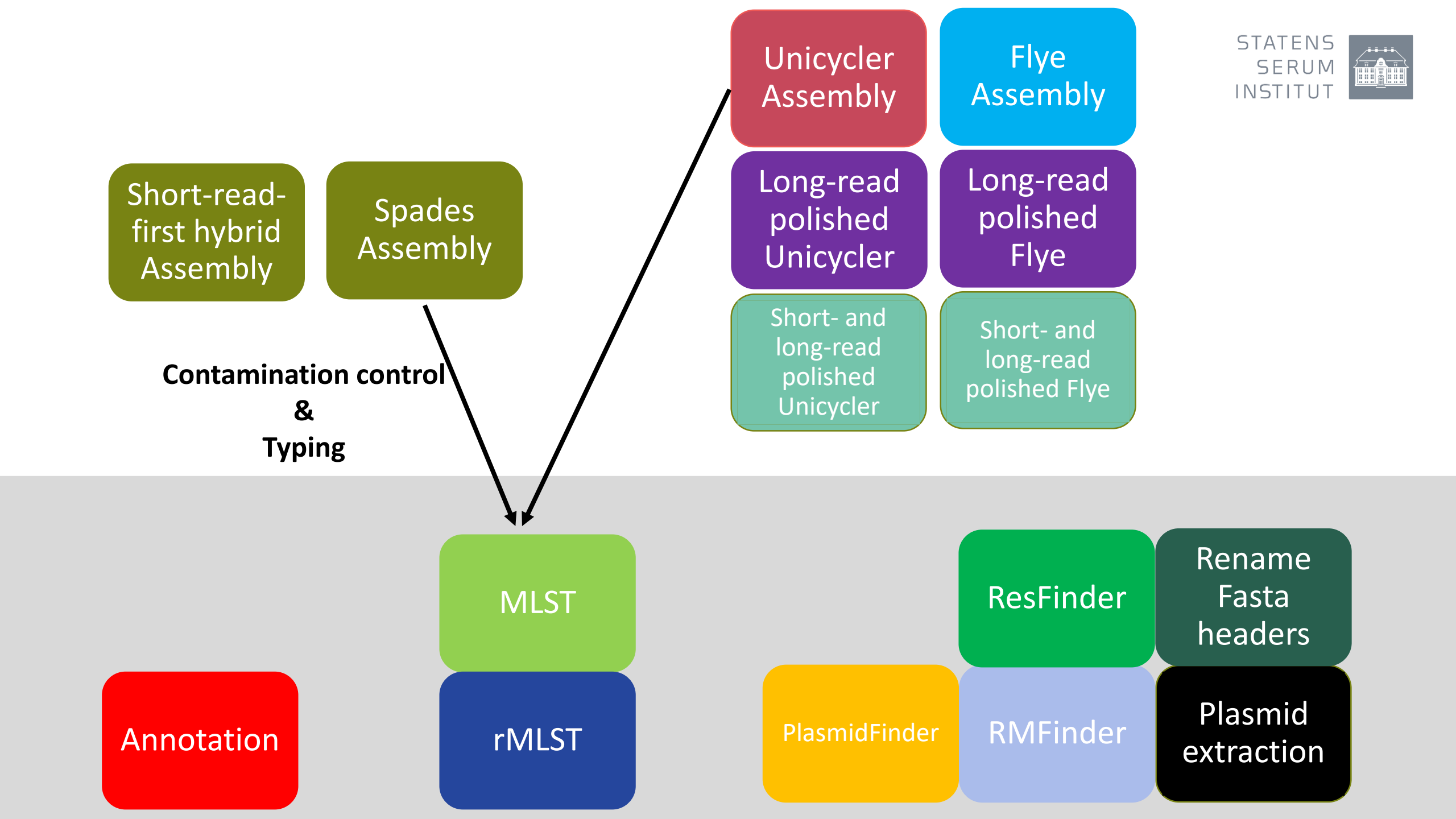
ResFinder

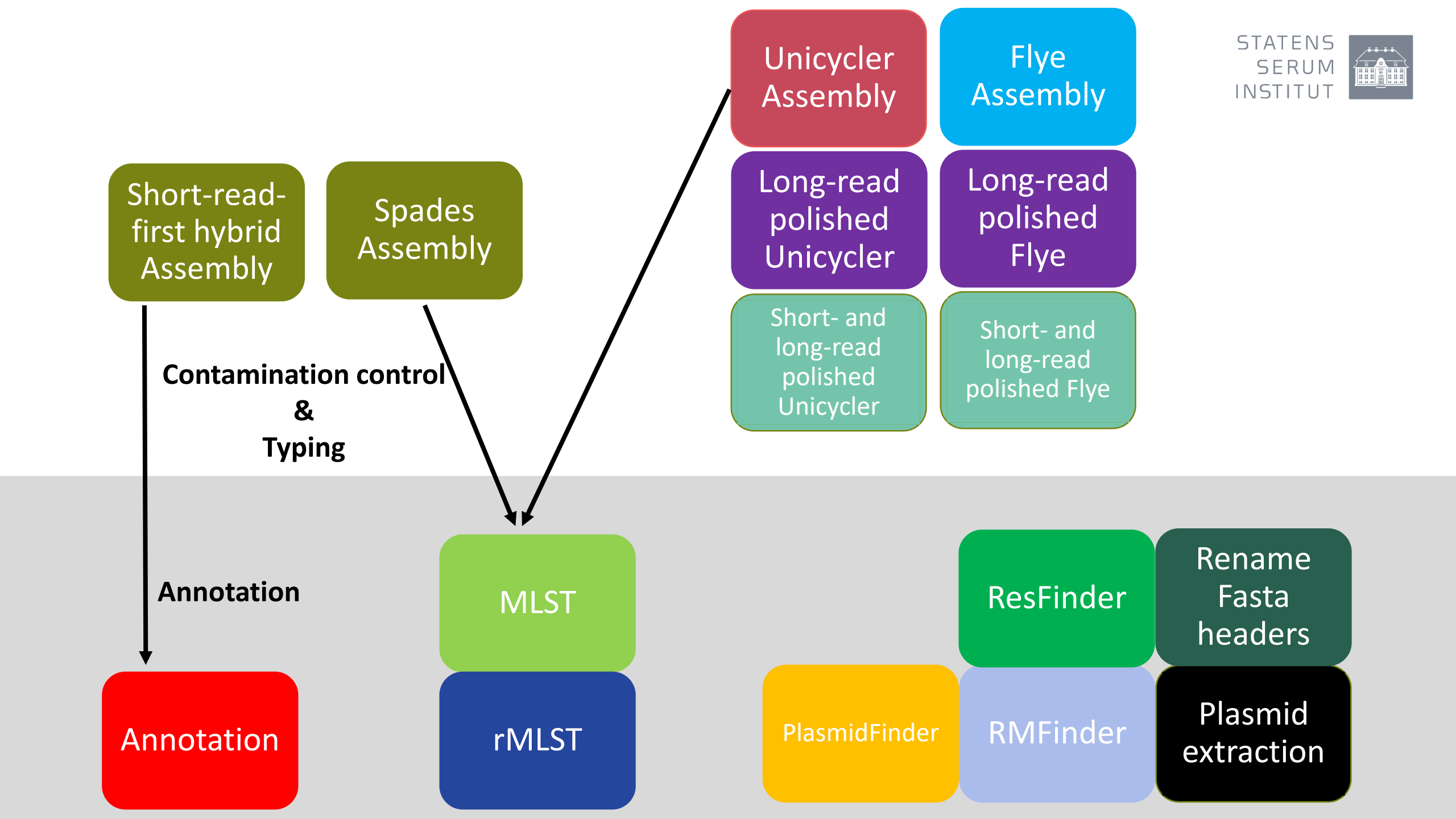
Rename Fasta headers

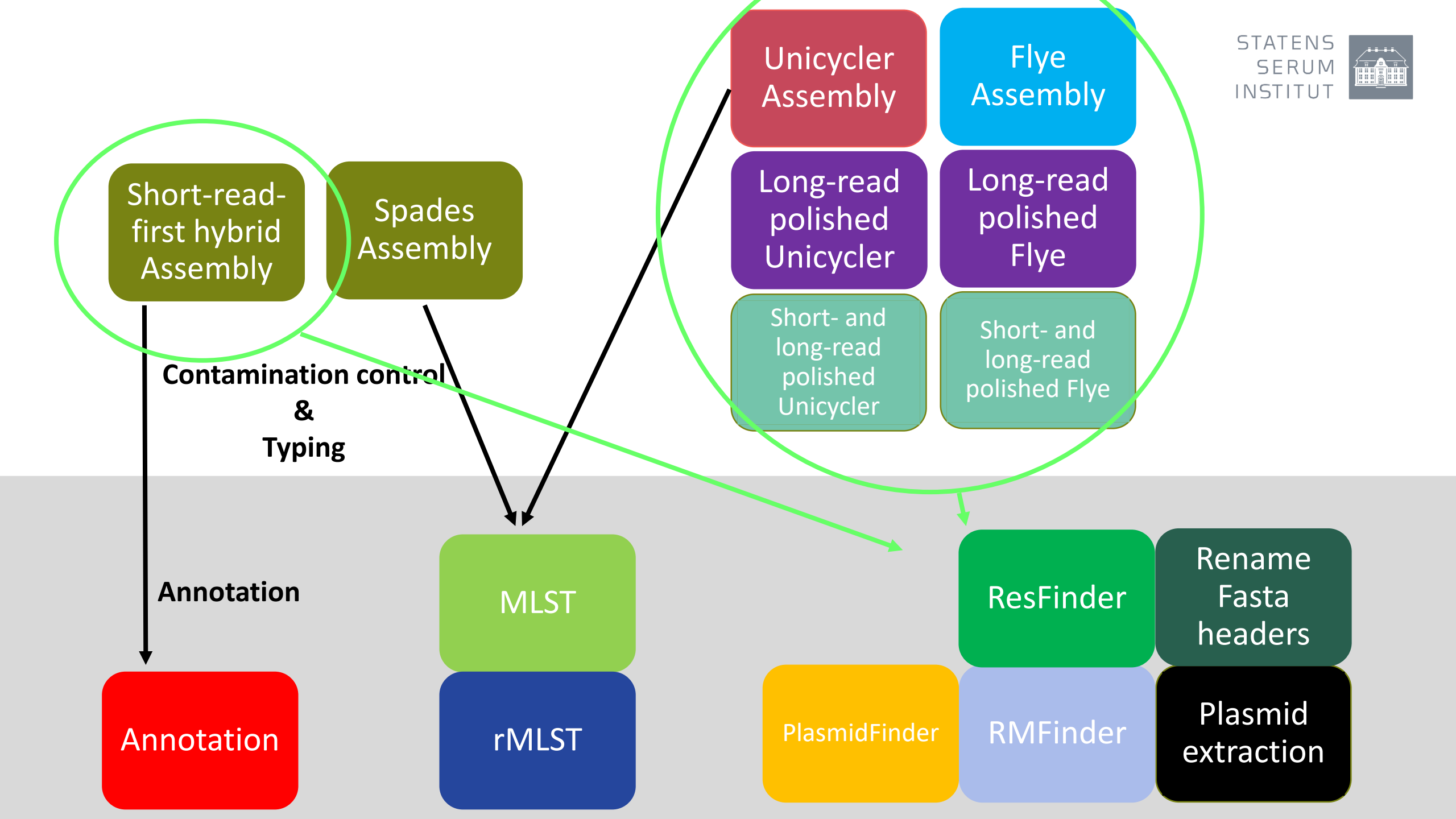
PlasmidFinder

RMFinder

Plasmid extraction


STATENS SERUM INSTITUT
Unicycler Assembly
Flye Assembly
Long-read polished Unicycler
Long-read polished Flye
Short- and long-read polished Unicycler
Short- and long-read polished Flye
Short-read-first hybrid Assembly
Spades Assembly
Contamination control & Typing
MLST
rMLST
Annotation
ResFinder
Rename Fasta headers
PlasmidFinder
RMFinder
Plasmid extraction


STATENS SERUM INSTITUT
Unicycler Assembly
Flye Assembly
Long-read polished Unicycler
Long-read polished Flye
Short- and long-read polished Unicycler
Short- and long-read polished Flye
Short-read-first hybrid Assembly
Spades Assembly
Contamination control & Typing
Annotation
Annotation
MLST
rMLST
ResFinder
Rename Fasta headers
PlasmidFinder
RMFinder
Plasmid extraction


STATENS SERUM INSTITUT
Unicycler Assembly
Flye Assembly
Long-read polished Unicycler
Long-read polished Flye
Short- and long-read polished Unicycler
Short- and long-read polished Flye
Short-read-first hybrid Assembly
Spades Assembly
Contamination control & Typing
Annotation
Annotation
MLST
rMLST
ResFinder
Rename Fasta headers
PlasmidFinder
RMFinder
Plasmid extraction

Rename Fasta headers

Plasmid extraction

Assembly.fasta

```
>1
>2
>3
>4
>5
>6
```

[Isolate_name]_Flye.fasta

```
>1_length=4918942_depth=1.00_circular
>2_length=147642_depth=1.67_circular
>3_length=106405_depth=1.68_circular
>4_length=76271_depth=1.43_circular
>5_length=67767_depth=1.51_circular
>6_length=36245_depth=2.00_circular
```

Assembly.fasta

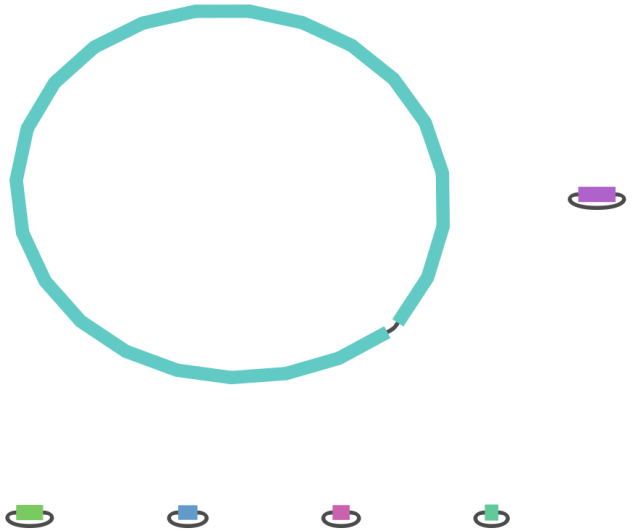
>1
>2
>3
>4
>5
>6

[Isolate_name]_Flye.fasta

>1_length=4918942_depth=1.00_circular
>2_length=147642_depth=1.67_circular
>3_length=106405_depth=1.68_circular
>4_length=76271_depth=1.43_circular
>5_length=67767_depth=1.51_circular
>6_length=36245_depth=2.00_circular

Rename Fasta headers

Plasmid extraction



WGS043_41_AMA005046_contig1_4918942bp_circular_UNPO.fasta
WGS043_41_AMA005046_contig2_147642bp_circular_UNPO.fasta
WGS043_41_AMA005046_contig3_106405bp_circular_UNPO.fasta
WGS043_41_AMA005046_contig4_76271bp_circular_UNPO.fasta
WGS043_41_AMA005046_contig5_67767bp_circular_UNPO.fasta
WGS043_41_AMA005046_contig6_36245bp_circular_UNPO.fasta

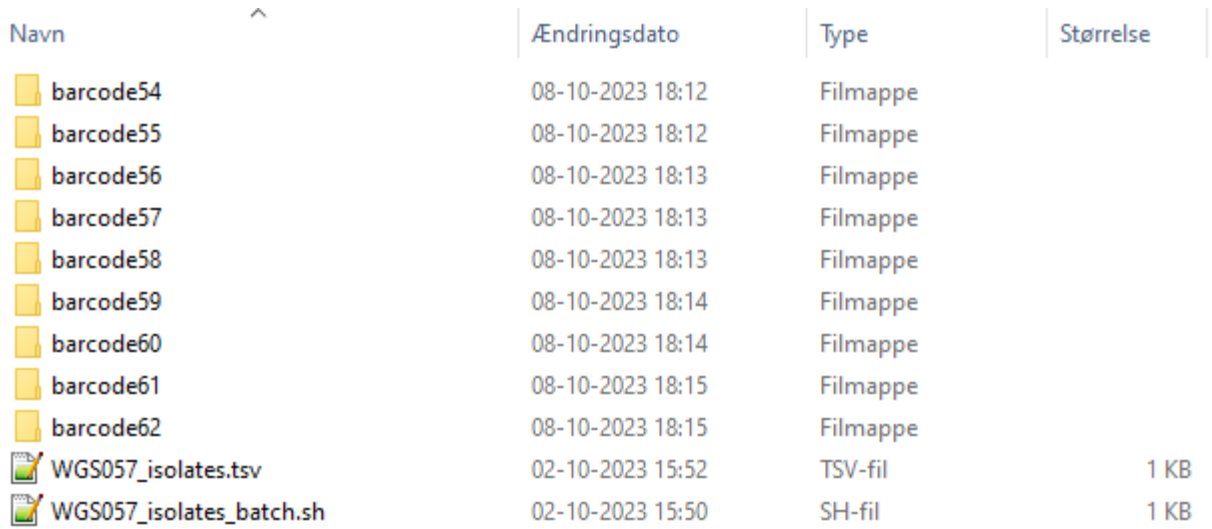
Old version of SSI pipeline

STATENS SERUM INSTITUT

MBHallgren/FullForcePlasmidAssembler: Full Force Plasmid Assembler Tool (github.com)

SSI pipeline output

STATENS SERUM INSTITUT



Navn	Ændringsdato	Type	Størrelse
barcode54	08-10-2023 18:12	Filmappe	
barcode55	08-10-2023 18:12	Filmappe	
barcode56	08-10-2023 18:13	Filmappe	
barcode57	08-10-2023 18:13	Filmappe	
barcode58	08-10-2023 18:13	Filmappe	
barcode59	08-10-2023 18:14	Filmappe	
barcode60	08-10-2023 18:14	Filmappe	
barcode61	08-10-2023 18:15	Filmappe	
barcode62	08-10-2023 18:15	Filmappe	
WGS057_isolates.tsv	02-10-2023 15:52	TSV-fil	1 KB
WGS057_isolates_batch.sh	02-10-2023 15:50	SH-fil	1 KB

STATENS
SERUM
INSTITUT

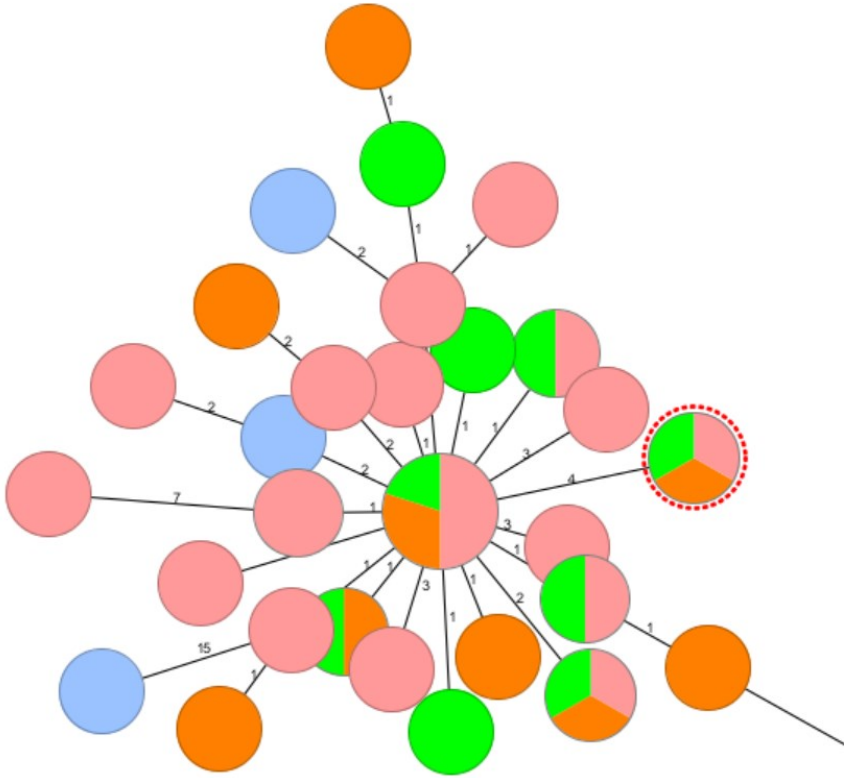
Questions?

Astrid Rasmussen, SSI

Cluster analysis with ONT data

Funded by the European Union

STATENS SERUM INSTITUT



Cluster analysis with Nanopore data

Henrik Hasman & Astrid Rasmussen

Outline

STATENS SERUM INSTITUT

- Nanopore vs Illumina data
- cgMLST with Nanopore and Illumina data
  - When it works
  - When it does not work
- MINTyper

Nanopore – the new(ish) kid on the block

STATENS SERUM INSTITUT

6-15 days

6-48 hours

Relatively..

- low price per isolate
- well-proven tech[illegible]gy
- high precisi[illegible]w error rate)
- Slow (d[illegible]ng on the setup)

..but r[illegible]s in real-time

Tools for outbreak detection validated

Relatively..

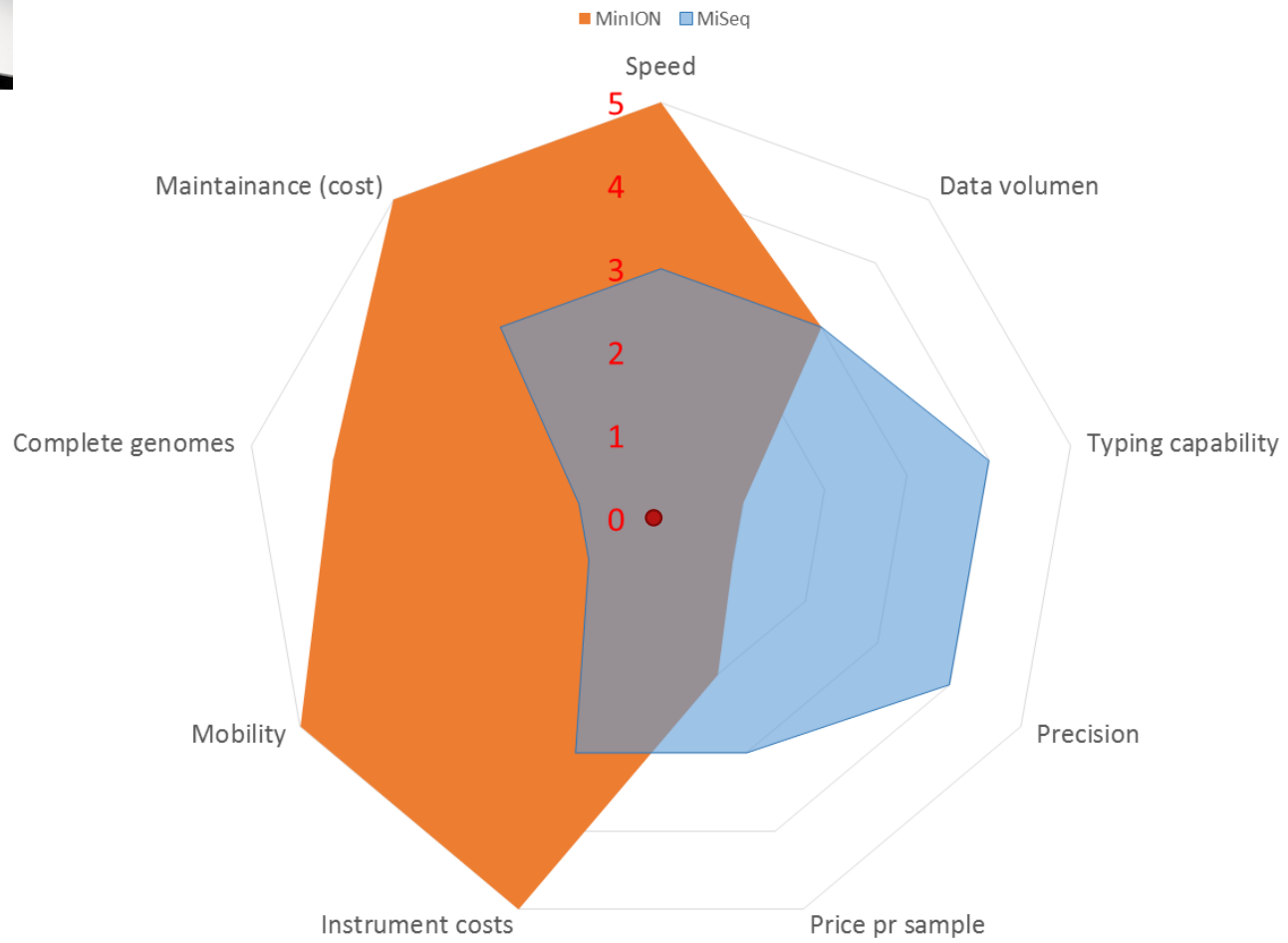
- Low-to medium [illegible]per isolate
- experimenta[illegible]hology
- low pre[illegible](high error rate)
- fast

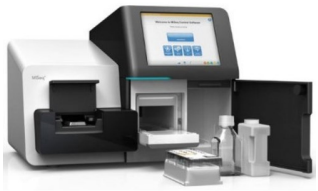
..a[illegible]ds available in real-time

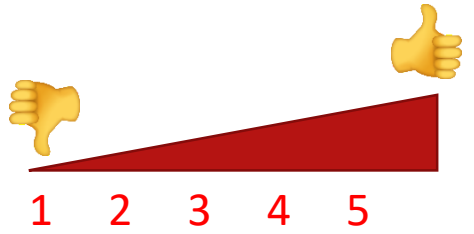
Tools for outbreak detection emerging

Sequencing Instruments (pros and cons)

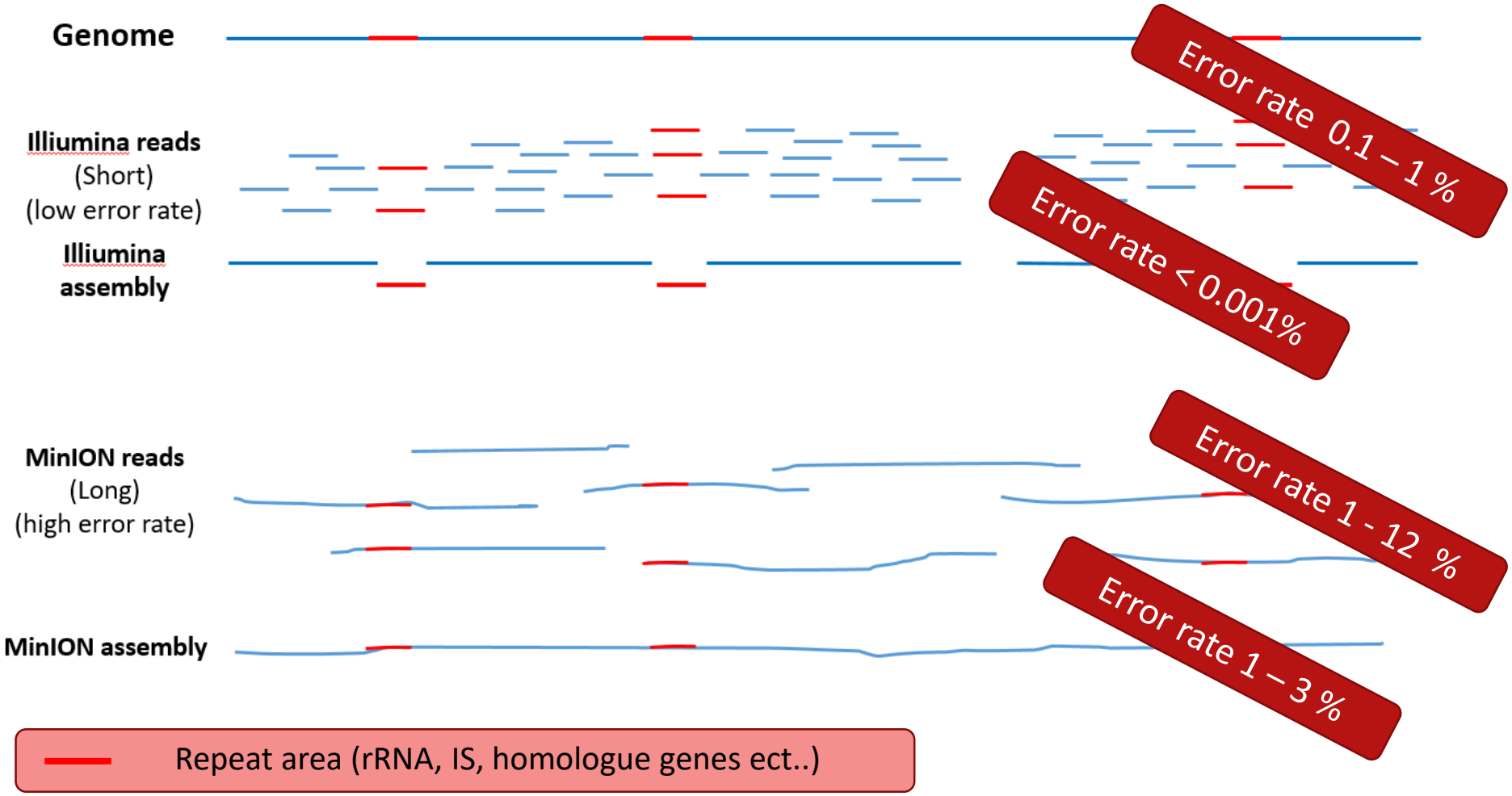
MinION vs MiSeq





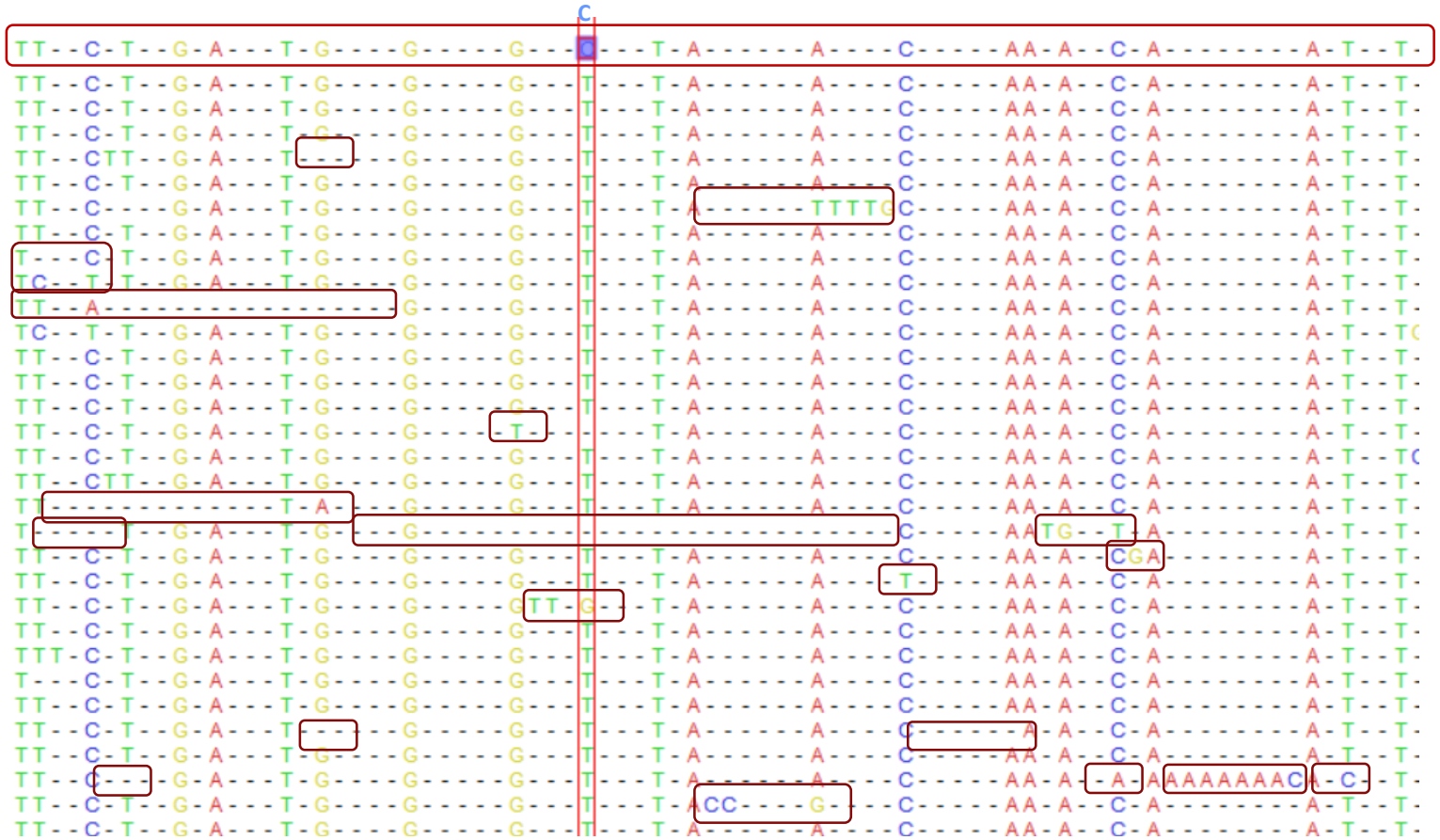


Illumina vs Nanopore data



Illumina vs Nanopore data

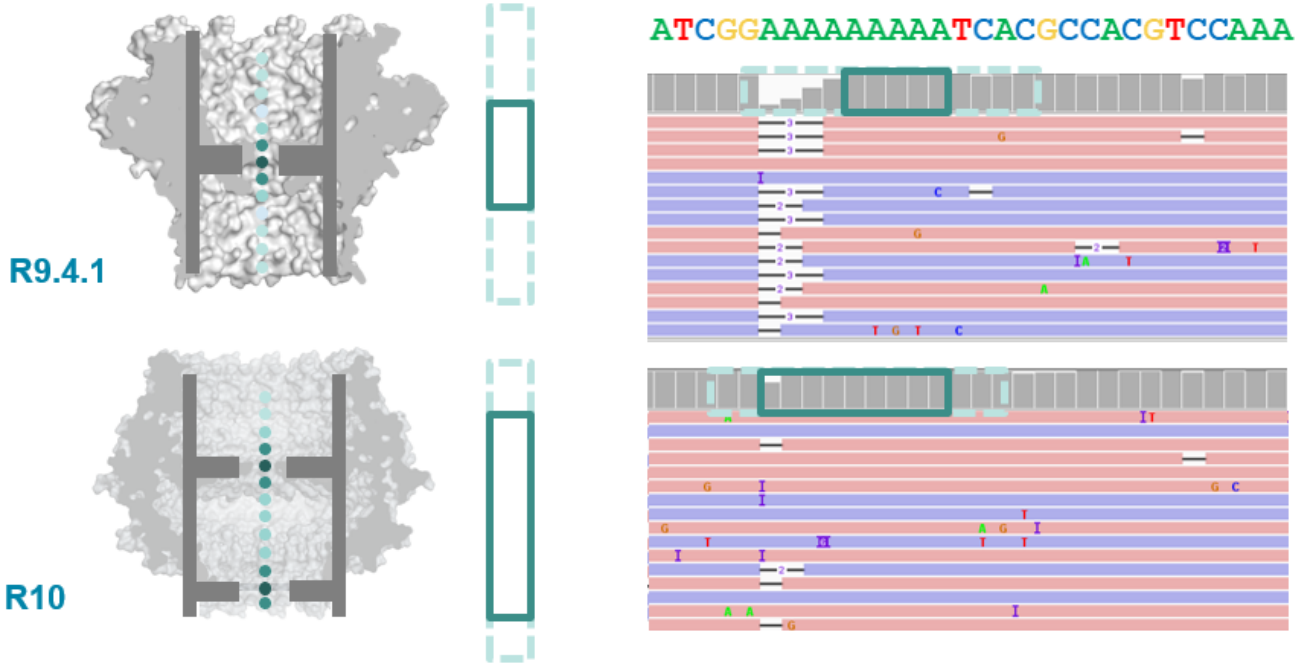
MinION raw data



Nanopore common errors

STATENS SERUM INSTITUT

- DNA modification (methylation, glycosylation, …)
- Homopolymeric regions



E. Hormaechei ST79 - cgMLST

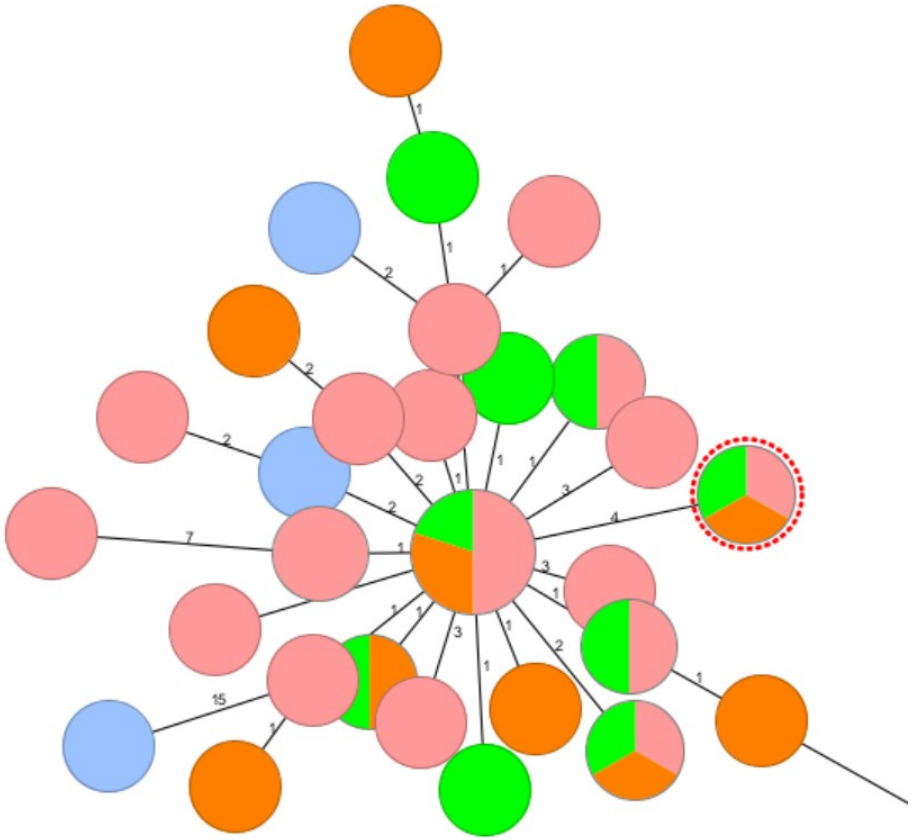
STATENS SERUM INSTITUT

Illumina draft genomes

ONT r9.4.1 UniCycler + Medaka

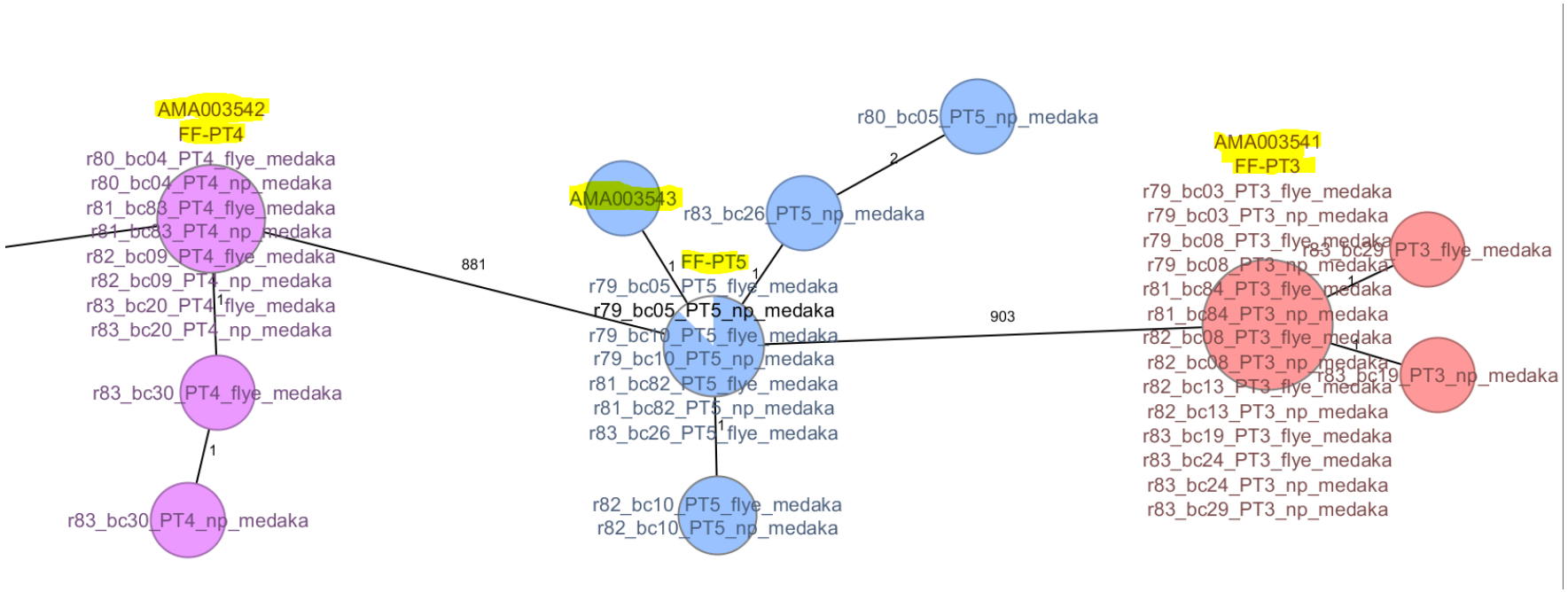
ONT r10.4.1 UniCycler + Medaka

ONT r10.4.1 Flye + Medaka



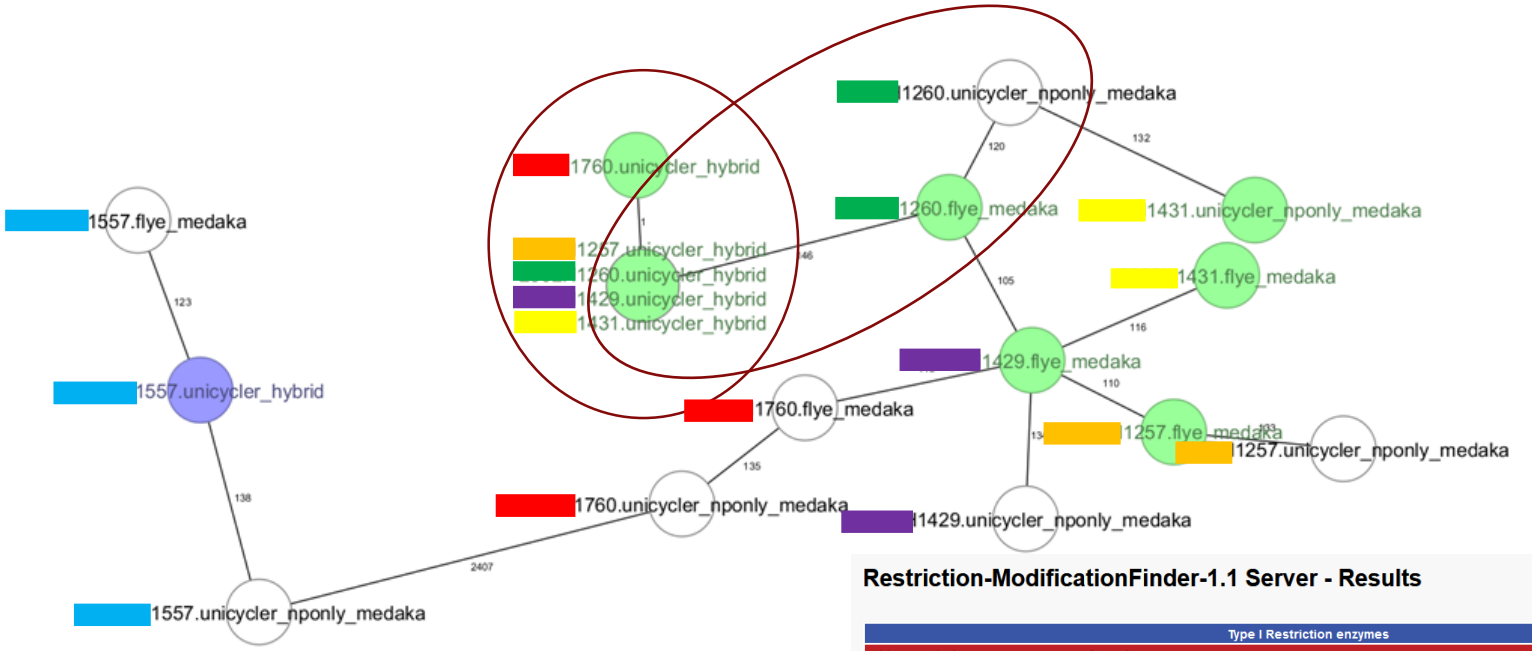
SeqSphere cgMLST – Illumina vs ONT r10.3

STATENS SERUM INSTITUT



Salmonella Infantis - *SinI*

STATENS SERUM INSTITUT



Restriction-ModificationFinder-1.1 Server - Results

Type I Restriction enzymes

No restriction enzyme genes found.

Type II Restriction enzymes

Gene	%Identity	HSP/Query length	Contig	Position in contig	Type	Function	Recognition Seq	Accession number
M.SenAboDcm	99.37	1431 / 1431	1	1849386..1850816	Type II	methyltransferase	CCWGG	CP007534,
M.SinI	100.00	1386 / 1386	1	3987543..3988928	Type II	methyltransferase	GGWCC	J03391,
SinI	100.00	693 / 693	1	3988960..3989652	Type II	restriction enzyme	GGWCC	J03391,
M.Sen641III	100.00	885 / 885	1	490189..491073	Type II	methyltransferase	ATGCAT	CP007249,

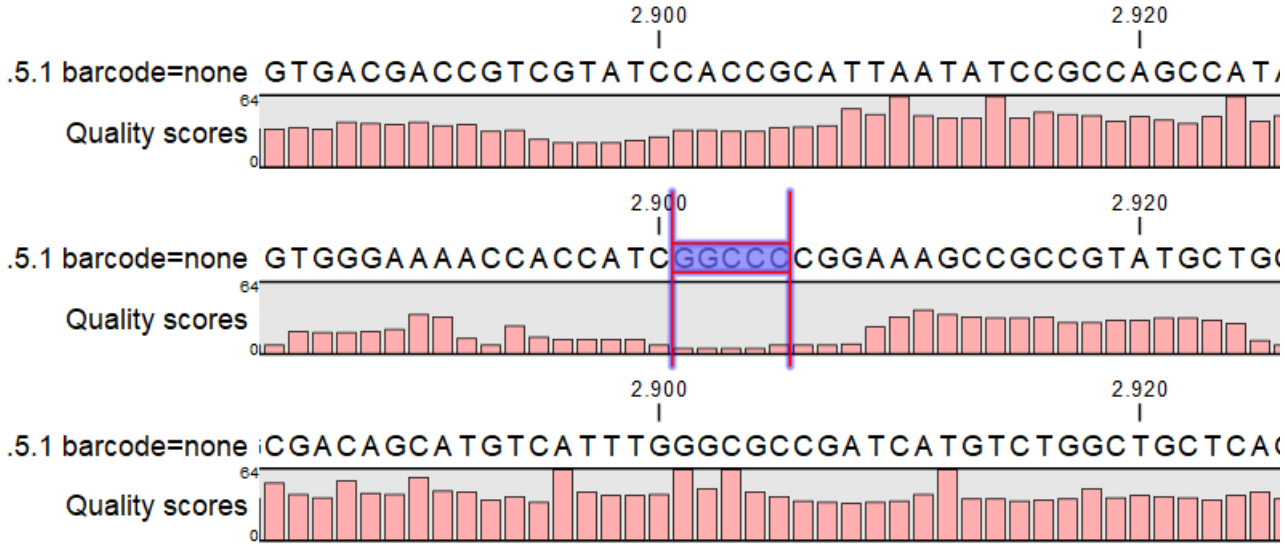
Type III Restriction enzymes

Gene	%Identity	HSP/Query length	Contig	Position in contig	Type	Function	Recognition Seq	Accession number
SenAZII	96.27	2976 / 2976	1	3468747..3471722	Type III	restriction enzyme	CAGAG	CP000880,
M.SenSPBII	99.64	1959 / 1959	1	3471729..3473687	Type III	methyltransferase	CAGAG	CP000886,

Type IV Restriction enzymes

No restriction enzyme genes found.

Salmonella Infantis - *Sin*I



G → A

6021

AGCTGTGGGAAACCACCATGGACCCGGAAAGCCGCCGTATG

AGCTGTGGGAAACCACCATGAACCCGGAAAGCCGCCGTATG

50288

TTATTGGCTCTAACATGGTGGACCTTATCACCTACACCAACC

TTATTGGCTCTAACATGGTGAACCTTATCACCTACACCAACC

MINTyper

STATENS SERUM INSTITUT

JOURNAL ARTICLE

MINTyper: an outbreak-detection method for accurate and rapid SNP typing of clonal clusters with noisy long reads

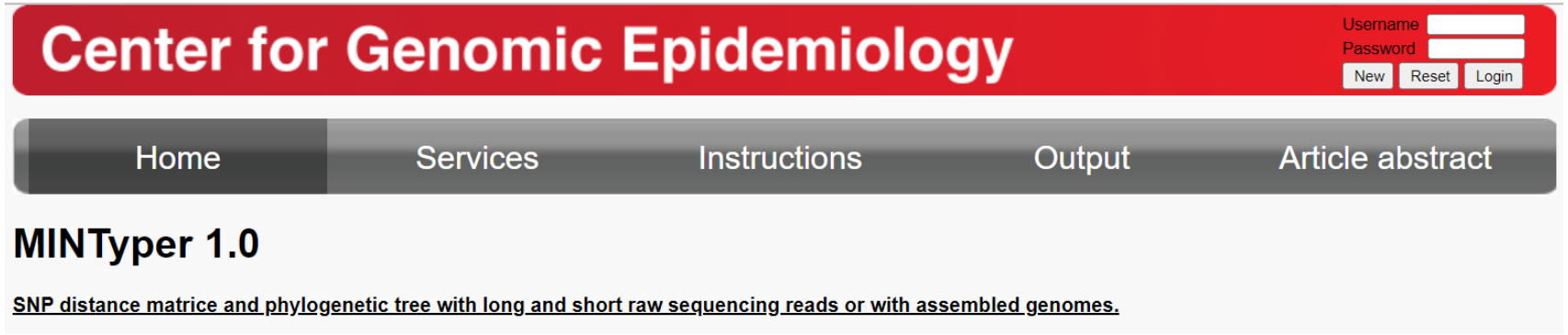
Malte B Hallgren, Søren Overballe-Petersen, Ole Lund, Henrik Hasman, Philip T L C Clausen

Biology Methods and Protocols, Volume 6, Issue 1, 2021, bpab008,
https://doi.org/10.1093/biomethods/bpab008
Published: 21 April 2021 **Article history** ▾

Available online on CGE and as a command line tool

The MINTyper tool at CGE

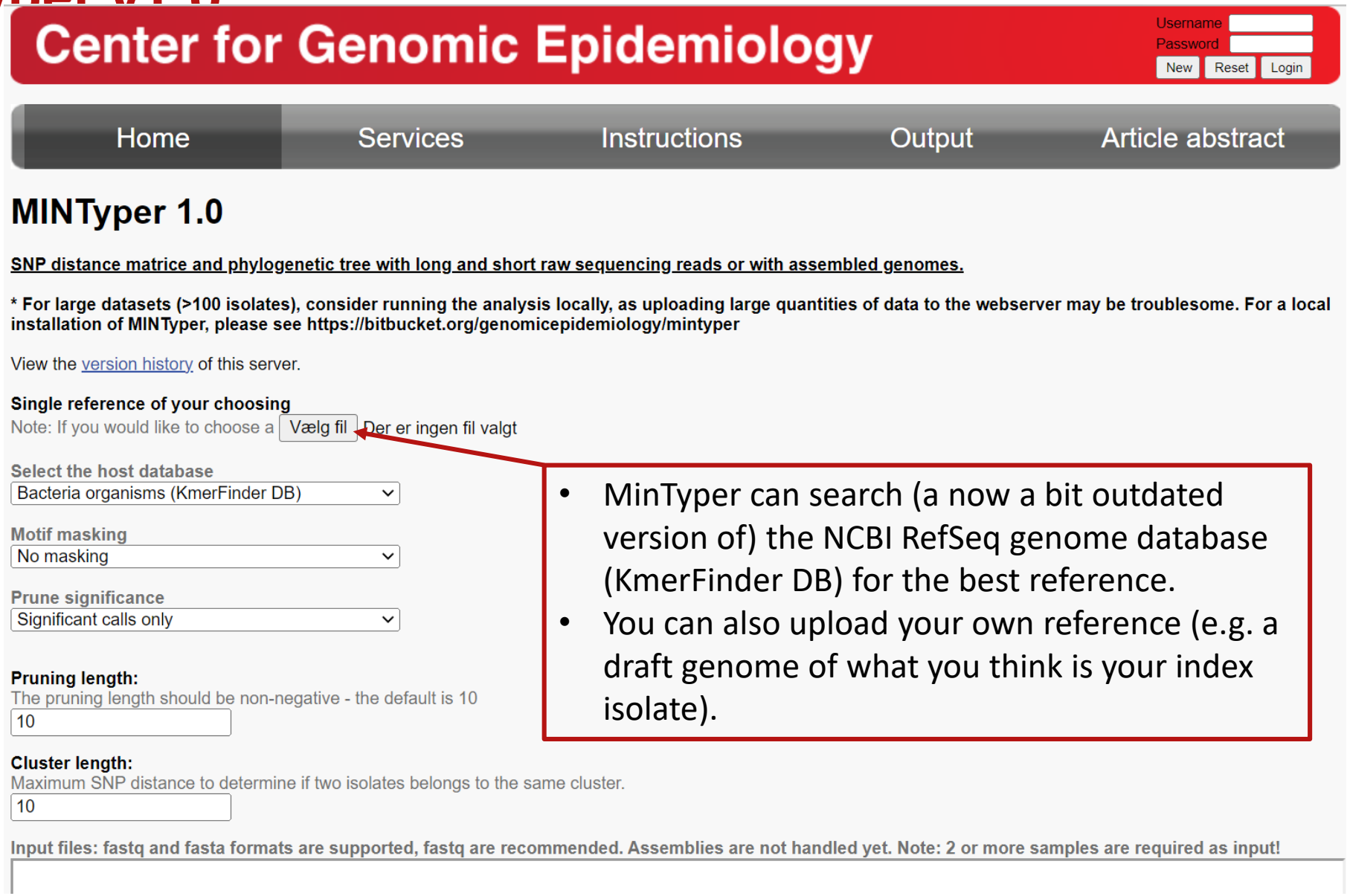
STATENS SERUM INSTITUT



- Will only accept raw data (Illumina and ONT)
- Will fail if not all input data (strains) cover at least 50% of the reference
- Allow for the user to give her own reference genome (fasta format)
- Allow to filter out Dcm methylation signals, which may give issues with the fast basecaller (at least in old versions of Guppy).
- Exists as a command-line tool (genomicepidemiology / mintyper — Bitbucket).

Mintyper v1.0

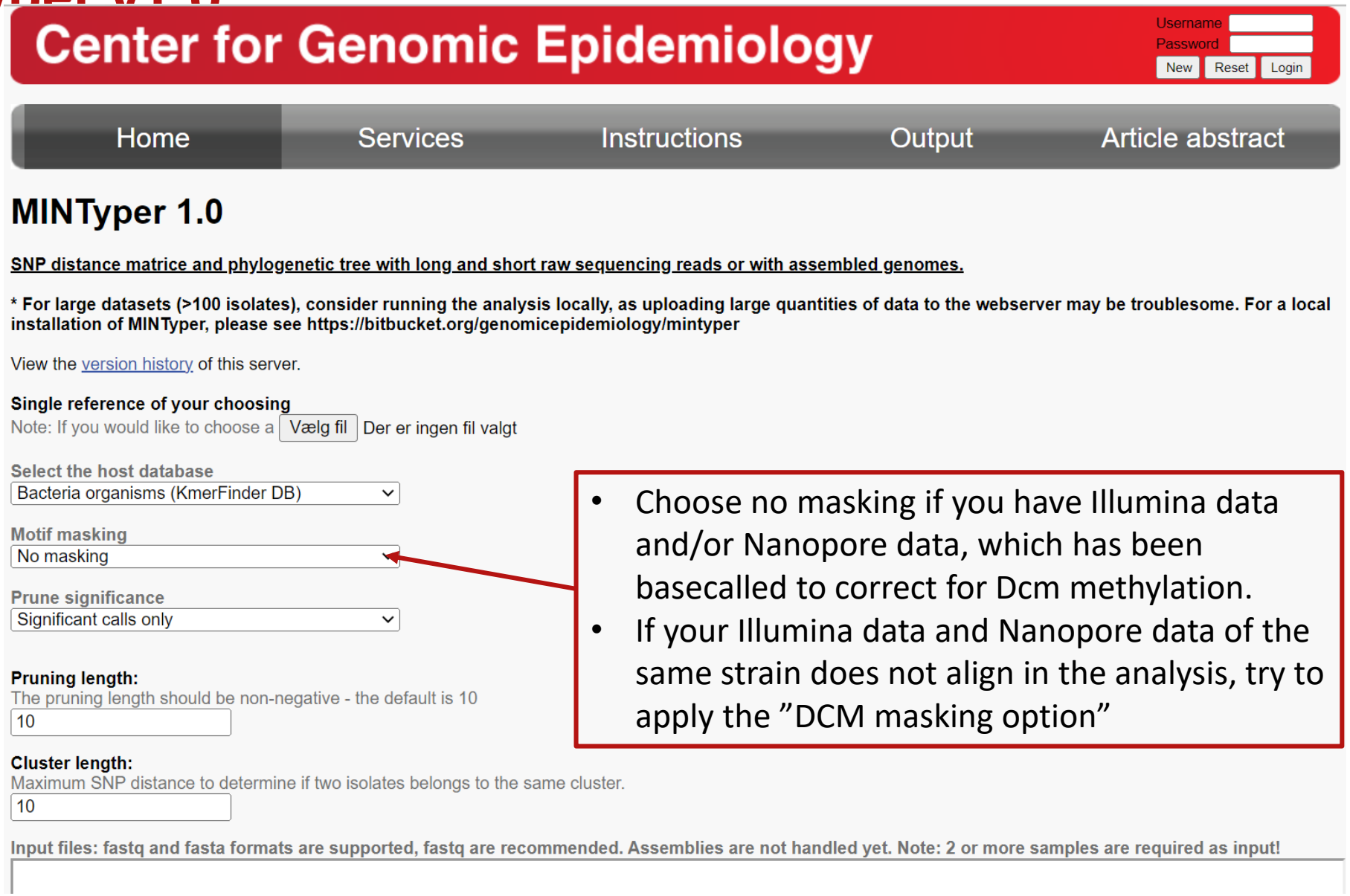
STATENS SERUM INSTITUT



- MinTyper can search (a now a bit outdated version of) the NCBI RefSeq genome database (KmerFinder DB) for the best reference.
- You can also upload your own reference (e.g. a draft genome of what you think is your index isolate).

Mintyper v1.0

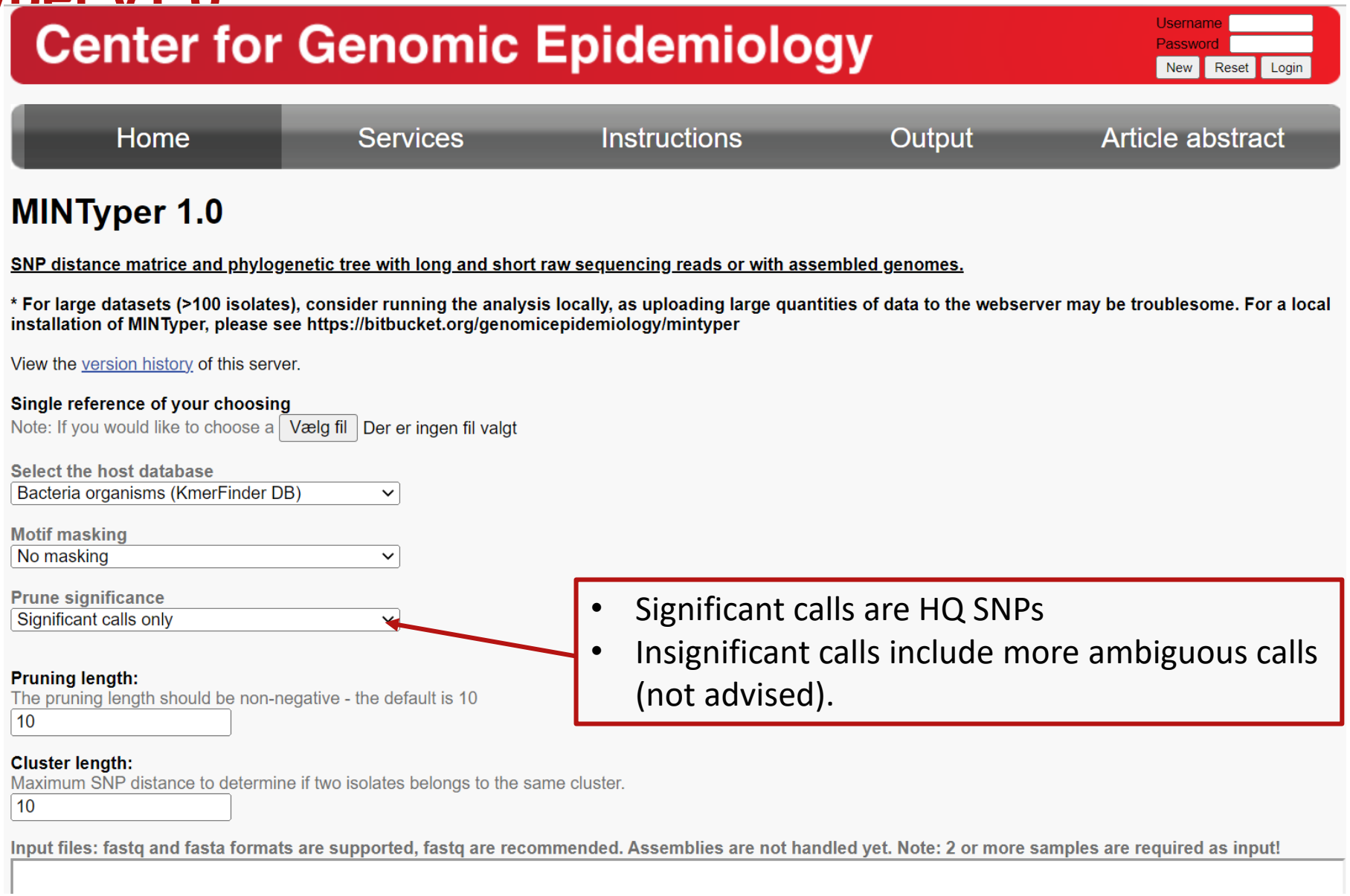
STATENS SERUM INSTITUT



- Choose no masking if you have Illumina data and/or Nanopore data, which has been basecalled to correct for Dcm methylation.
- If your Illumina data and Nanopore data of the same strain does not align in the analysis, try to apply the ”DCM masking option”

Mintyper v1.0

STATENS SERUM INSTITUT



- Significant calls are HQ SNPs
- Insignificant calls include more ambiguous calls (not advised).

Mintyper v1.0

STATENS SERUM INSTITUT

Center for Genomic Epidemiology

Username
Password
New Reset Login

Home | Services | Instructions | Output | Article abstract

MINTyper 1.0

SNP distance matrice and phylogenetic tree with long and short raw sequencing reads or with assembled genomes.

*** For large datasets (>100 isolates), consider running the analysis locally, as uploading large quantities of data to the webserver may be troublesome. For a local installation of MINTyper, please see https://bitbucket.org/genomicepidemiology/mintyper**

View the version history of this server.

Single reference of your choosing
Note: If you would like to choose a Vælg fil Der er ingen fil valgt

Select the host database
Bacteria organisms (KmerFinder DB)

Motif masking
No masking

Prune significance
Significant calls only

Pruning length:
The pruning length should be non-negative - the default is 10
10

Cluster length:
Maximum SNP distance to determine if two isolates belongs to the same cluster.
10

Input files: fastq and fasta formats are supported, fastq are recommended. Assemblies are not handled yet. Note: 2 or more samples are required as input!

- Select pruning distance.
- Use default or perhaps 100 bp.

Mintyper v1.0

STATENS SERUM INSTITUT

Center for Genomic Epidemiology

Username
Password
New Reset Login

Home | Services | Instructions | Output | Article abstract

MINTyper 1.0

SNP distance matrice and phylogenetic tree with long and short raw sequencing reads or with assembled genomes.

*** For large datasets (>100 isolates), consider running the analysis locally, as uploading large quantities of data to the webserver may be troublesome. For a local installation of MINTyper, please see https://bitbucket.org/genomicepidemiology/mintyper**

View the version history of this server.

Single reference of your choosing
Note: If you would like to choose a Vælg fil Der er ingen fil valgt

Select the host database
Bacteria organisms (KmerFinder DB)

Motif masking
No masking

Prune significance
Significant calls only

Pruning length:
The pruning length should be non-negative - the default is 10
10

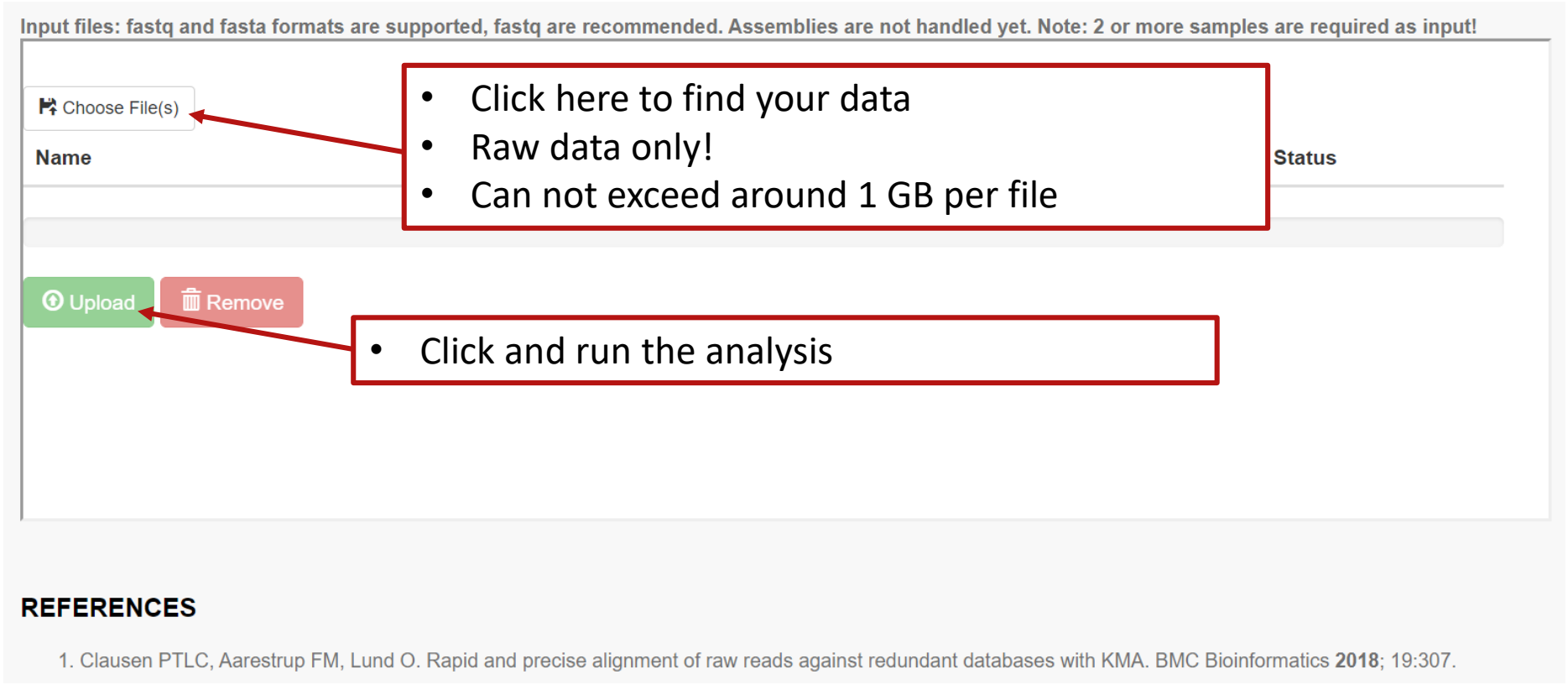
Cluster length:
Maximum SNP distance to determine if two isolates belongs to the same cluster.
10

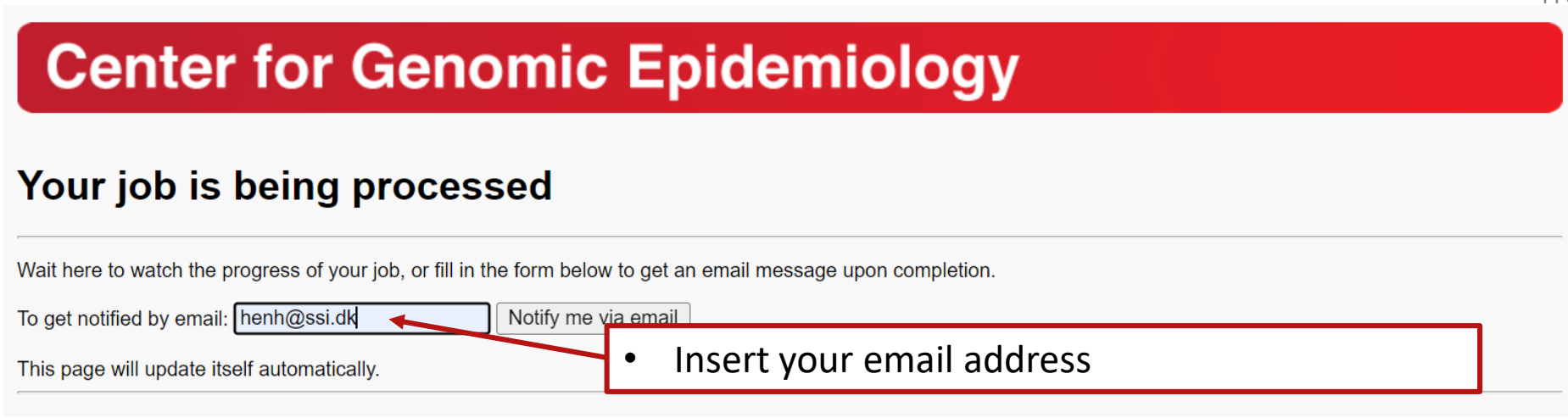
Input files: fastq and fasta formats are supported, fastq are recommended. Assemblies are not handled yet. Note: 2 or more samples are required as input!

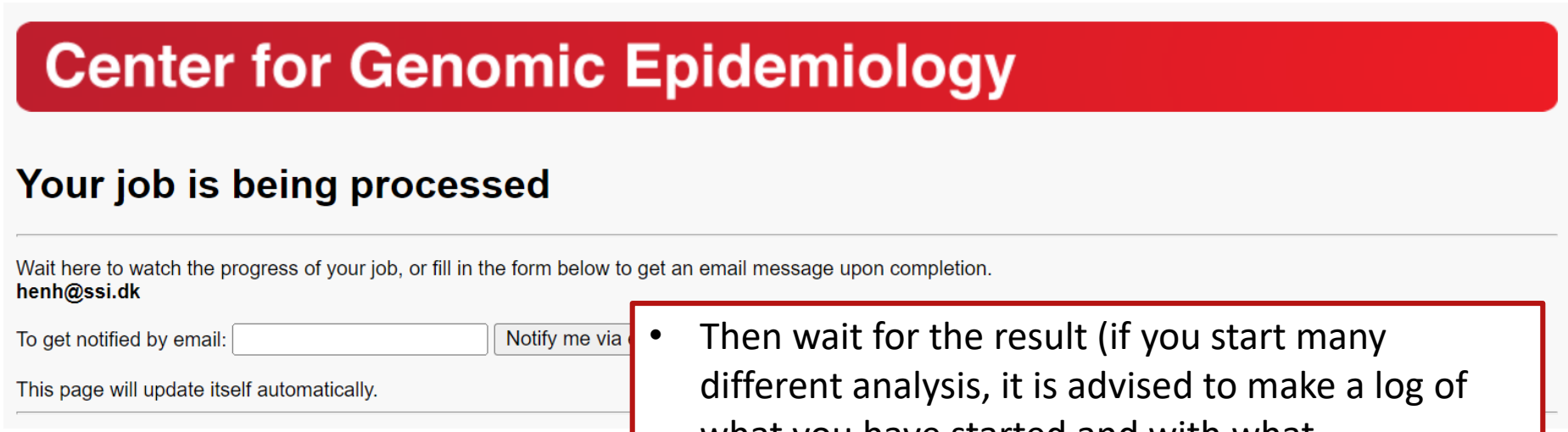
- Define a SNP distance for clusters
- Often between 10 and 20 (but depends on the length and nature of the outbreak).

Uploading data

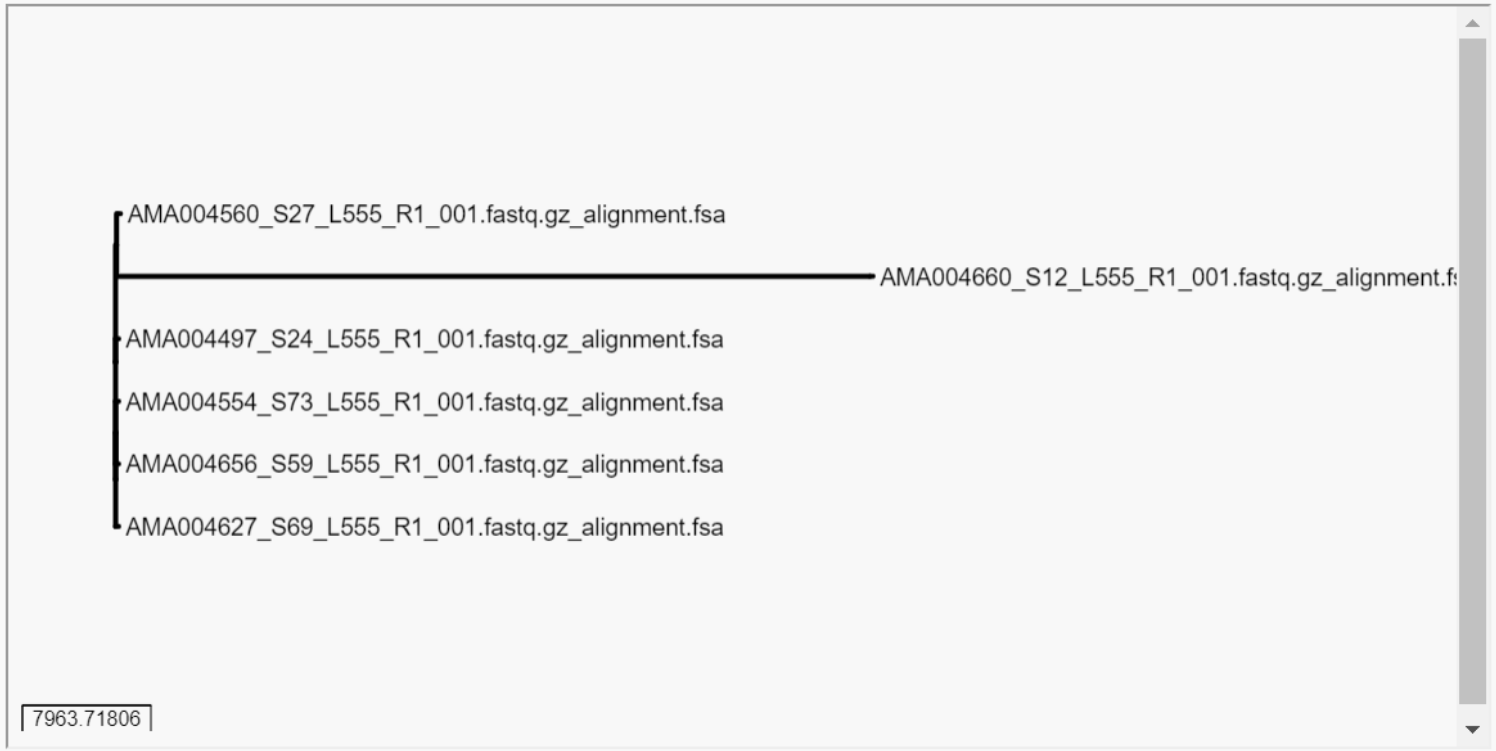
STATENS SERUM INSTITUT




Center for Genomic Epidemiology
Your job is being processed
Wait here to watch the progress of your job, or fill in the form below to get an email message upon completion.
To get notified by email: henh@ssi.dk
Notify me via email
This page will update itself automatically.
• Insert your email address


Center for Genomic Epidemiology
Your job is being processed
Wait here to watch the progress of your job, or fill in the form below to get an email message upon completion.
henh@ssi.dk
To get notified by email:
Notify me via
This page will update itself automatically.
• Then wait for the result (if you start many different analysis, it is advised to make a log of what you have started and with what settings…and perhaps also the hypothesis).

MINTyper output



Percentage of reference covered by all isolates: 84.71 (4149824 / 4899014)
Below is the single isolate stats on covered and trusted positions with respect to the reference.

Isolate	Valid positions	Pct. of reference
AMA004497_S24_L555_R1_001.fastq.gz	4435406	90.54
AMA004554_S73_L555_R1_001.fastq.gz	4427220	90.37
AMA004560_S27_L555_R1_001.fastq.gz	4465781	91.16
AMA004627_S69_L555_R1_001.fastq.gz	4412663	90.07
AMA004656_S59_L555_R1_001.fastq.gz	4442114	90.67
AMA004660_S12_L555_R1_001.fastq.gz	4327141	88.33

Log | Distance matrix | Phylogentic tree | Vcf files of mutations | Reference Sequence | Cluster.dbscan

MINTyper output

STATENS SERUM INSTITUT

Percentage of reference covered by all isolates: 84.71 (4149824 / 4899014)

Below is the single isolate stats on covered and trusted positions with respect to the reference.

Isolate	Valid positions	Pct. of reference
AMA004497_S24_L555_R1_001.fastq.gz	4435406	90.54
AMA004554_S73_L555_R1_001.fastq.gz	4427220	90.37
AMA004560_S27_L555_R1_001.fastq.gz	4465781	91.16
AMA004627_S69_L555_R1_001.fastq.gz	4412663	90.07
AMA004656_S59_L555_R1_001.fastq.gz	4442114	90.67
AMA004660_S12_L555_R1_001.fastq.gz	4327141	88.33

Log | Distance matrix | Phylogentic tree | Vcf files of mutations | Reference Sequence | Cluster.dbscan

```
# Running mintyper 1.1.0 with following input conditions:
Namespace(bc=0.7, cge=True, cluster_length=10, exe_path='/home/data1/services/MINTyper/MINTyper-1.0/scripts/bin/MINTyper/', fast_tree=False, i_a
/MINTyper/MINTyper-1.0/IO/1_25_9_2022_239_804_64033/uploads//AMA004627_S69_L555_R2_001.fastq.gz', '/home/data1/services/MINTyper/MINTyper-1.0/IO
# Finding best template
# Best template found was NZ_CP024672.1 Citrobacter freundii strain HM38 chromosome, complete genome
# Template number was: 1901
# Mapping reads to template
# Paired-end illumina input not given but determined by the eval_pe function
/home/data1/services/MINTyper/MINTyper-1.0/scripts/bin/MINTyper/kma/kma -ipe /home/data1/services/MINTyper/MINTyper-1.0/IO/1_25_9_2022_239_804_6
/home/data1/services/MINTyper/MINTyper-1.0/scripts/bin/MINTyper/kma/kma -ipe /home/data1/services/MINTyper/MINTyper-1.0/IO/1_25_9_2022_239_804_6
/home/data1/services/MINTyper/MINTyper-1.0/scripts/bin/MINTyper/kma/kma -ipe /home/data1/services/MINTyper/MINTyper-1.0/IO/1_25_9_2022_239_804_6
/home/data1/services/MINTyper/MINTyper-1.0/scripts/bin/MINTyper/kma/kma -ipe /home/data1/services/MINTyper/MINTyper-1.0/IO/1_25_9_2022_239_804_6
/home/data1/services/MINTyper/MINTyper-1.0/scripts/bin/MINTyper/kma/kma -ipe /home/data1/services/MINTyper/MINTyper-1.0/IO/1_25_9_2022_239_804_6
/home/data1/services/MINTyper/MINTyper-1.0/scripts/bin/MINTyper/kma/kma -ipe /home/data1/services/MINTyper/MINTyper-1.0/IO/1_25_9_2022_239_804_6
# Alignment completed succesfully
# 4149824 / 4899014 bases included in distance matrix.

mintyper total runtime: 383.13289737701416 seconds
```

MINTyper output

STATENS SERUM INSTITUT

Percentage of reference covered by all isolates: 84.71 (4149824 / 4899014)

Below is the single isolate stats on covered and trusted positions with respect to the reference.

Isolate	Valid positions	Pct. of reference
AMA004497_S24_L555_R1_001.fastq.gz	4435406	90.54
AMA004554_S73_L555_R1_001.fastq.gz	4427220	90.37
AMA004560_S27_L555_R1_001.fastq.gz	4465781	91.16
AMA004627_S69_L555_R1_001.fastq.gz	4412663	90.07
AMA004656_S59_L555_R1_001.fastq.gz	4442114	90.67
AMA004660_S12_L555_R1_001.fastq.gz	4327141	88.33

ST18

ST91

Log | Distance matrix | Phylogentic tree | Vcf files of mutations | Reference Sequence | Cluster.dbscan

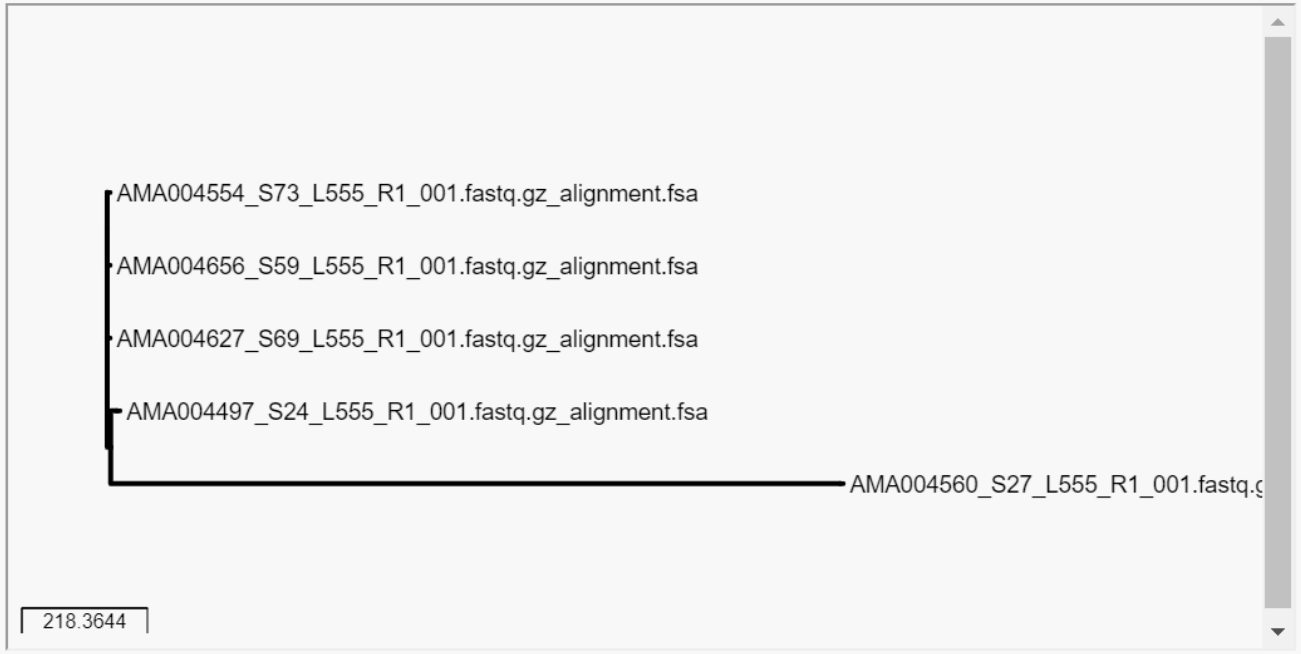
		1	2	3	4	5	6
1	AMA004497_S24_L555_R1_001.fastq.gz_alignment.fsa	0					
2	AMA004554_S73_L555_R1_001.fastq.gz_alignment.fsa	15	0				
3	AMA004560_S27_L555_R1_001.fastq.gz_alignment.fsa	133	130	0			
4	AMA004627_S69_L555_R1_001.fastq.gz_alignment.fsa	15	0	130	0		
5	AMA004656_S59_L555_R1_001.fastq.gz_alignment.fsa	15	0	130	0	0	
6	AMA004660_S12_L555_R1_001.fastq.gz_alignment.fsa	46761	46758	46758	46758	46758	0

Rerun without AMA004660

STATENS SERUM INSTITUT

Center for Genomic Epidemiology

Home | Services | Instructions | Output



Percentage of reference covered by all isolates: 89.18 (4368832 / 4899014)
Below is the single isolate stats on covered and trusted positions with respect to the reference.

Isolate	Valid positions	Pct. of reference
AMA004497_S24_L555_R1_001.fastq.gz	4435406	90.54
AMA004554_S73_L555_R1_001.fastq.gz	4427220	90.37
AMA004560_S27_L555_R1_001.fastq.gz	4465781	91.16
AMA004627_S69_L555_R1_001.fastq.gz	4412663	90.07
AMA004656_S59_L555_R1_001.fastq.gz	4442114	90.67

Log | Distance matrix | Phylogentic tree | Vcf files of mutations | Reference Sequence | Cluster.dbscan

Rerun without AMA004660

STATENS SERUM INSTITUT

Percentage of reference covered by all isolates: 89.18 (4368832 / 4899014)
Below is the single isolate stats on covered and trusted positions with respect to the reference.

Isolate	Valid positions	Pct. of reference
AMA004497_S24_L555_R1_001.fastq.gz	4435406	90.54
AMA004554_S73_L555_R1_001.fastq.gz	4427220	90.37
AMA004560_S27_L555_R1_001.fastq.gz	4465781	91.16
AMA004627_S69_L555_R1_001.fastq.gz	4412663	90.07
AMA004656_S59_L555_R1_001.fastq.gz	4442114	90.67

Log | Distance matrix | Phylogentic tree | Vcf files of mutations | Reference Sequence | Cluster.dbscan

```
# Running mintyper 1.1.0 with following input conditions:
Namespace(bc=0.7, cge=True, cluster_length=10, exe_path='/home/data1/services/MINTyper/MINTyper-1.0/scripts/bin/MINTyper/', fast
ervices/MINTyper/MINTyper-1.0/IO/1_25_9_2022_230_605_513390/uploads//AMA004627_S69_L555_R2_001.fastq.gz', '/home/data1/services/
# Finding best template
# Best template found was NZ_CP024672.1 Citrobacter freundii strain HM38 chromosome, complete genome
# Template number was: 1901
# Mapping reads to template
# Paired-end illumina input not given but determined by the eval_pe function
/home/data1/services/MINTyper/MINTyper-1.0/scripts/bin/MINTyper/kma/kma -ipe /home/data1/services/MINTyper/MINTyper-1.0/IO/1_25_
/home/data1/services/MINTyper/MINTyper-1.0/scripts/bin/MINTyper/kma/kma -ipe /home/data1/services/MINTyper/MINTyper-1.0/IO/1_25_
/home/data1/services/MINTyper/MINTyper-1.0/scripts/bin/MINTyper/kma/kma -ipe /home/data1/services/MINTyper/MINTyper-1.0/IO/1_25_
/home/data1/services/MINTyper/MINTyper-1.0/scripts/bin/MINTyper/kma/kma -ipe /home/data1/services/MINTyper/MINTyper-1.0/IO/1_25_
/home/data1/services/MINTyper/MINTyper-1.0/scripts/bin/MINTyper/kma/kma -ipe /home/data1/services/MINTyper/MINTyper-1.0/IO/1_25_
# Alignment completed succesfully
# 4368832 / 4899014 bases included in distance matrix.

mintyper total runtime: 370.7805440425873 seconds
```

Rerun without AMA004660

STATENS SERUM INSTITUT

Percentage of reference covered by all isolates: 89.18 (4368832 / 4899014)

Below is the single isolate stats on covered and trusted positions with respect to the reference.

Isolate	Valid positions	Pct. of reference
AMA004497_S24_L555_R1_001.fastq.gz	4435406	90.54
AMA004554_S73_L555_R1_001.fastq.gz	4427220	90.37
AMA004560_S27_L555_R1_001.fastq.gz	4465781	91.16
AMA004627_S69_L555_R1_001.fastq.gz	4412663	90.07
AMA004656_S59_L555_R1_001.fastq.gz	4442114	90.67

Log | Distance matrix | Phylogentic tree | Vcf files of mutations | Reference Sequence | Cluster.dbscan

		1	2	3	4	5
1	AMA004497_S24_L555_R1_001.fastq.gz_alignment.fsa	0				
2	AMA004554_S73_L555_R1_001.fastq.gz_alignment.fsa	17	0			
3	AMA004560_S27_L555_R1_001.fastq.gz_alignment.fsa	1280	1275	0		
4	AMA004627_S69_L555_R1_001.fastq.gz_alignment.fsa	17	0	1275	0	
5	AMA004656_S59_L555_R1_001.fastq.gz_alignment.fsa	17	0	1275	0	0

Rerun without AMA004660

STATENS SERUM INSTITUT

Percentage of reference covered by all isolates: 89.18 (4368832 / 4899014)

Below is the single isolate stats on covered and trusted positions with respect to the reference.

Isolate	Valid positions	Pct. of reference
AMA004497_S24_L555_R1_001.fastq.gz	4435406	90.54
AMA004554_S73_L555_R1_001.fastq.gz	4427220	90.37
AMA004560_S27_L555_R1_001.fastq.gz	4465781	91.16
AMA004627_S69_L555_R1_001.fastq.gz	4412663	90.07
AMA004656_S59_L555_R1_001.fastq.gz	4442114	90.67

Log | Distance matrix | Phylogentic tree | Vcf files of mutations | Reference Sequence | Cluster.dbscan

		1	2	3	4	5
1	AMA004497_S24_L555_R1_001.fastq.gz_alignment.fsa	0				
2	AMA004554_S73_L555_R1_001.fastq.gz_alignment.fsa	17	0			
3	AMA004560_S27_L555_R1_001.fastq.gz_alignment.fsa	1280	1275	0		
4	AMA004627_S69_L555_R1_001.fastq.gz_alignment.fsa	17	0	1275	0	
5	AMA004656_S59_L555_R1_001.fastq.gz_alignment.fsa	17	0	1275	0	0

MINTyper output - visualizations

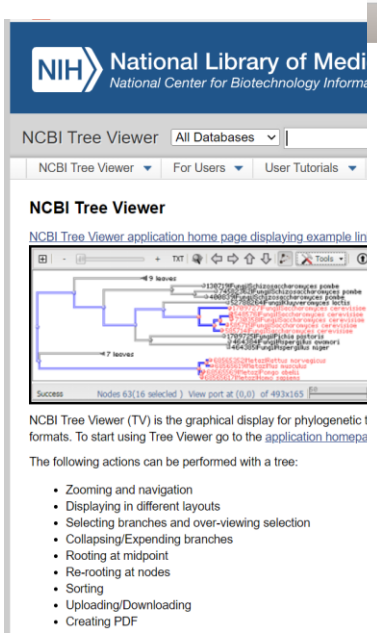
STATENS SERUM INSTITUT

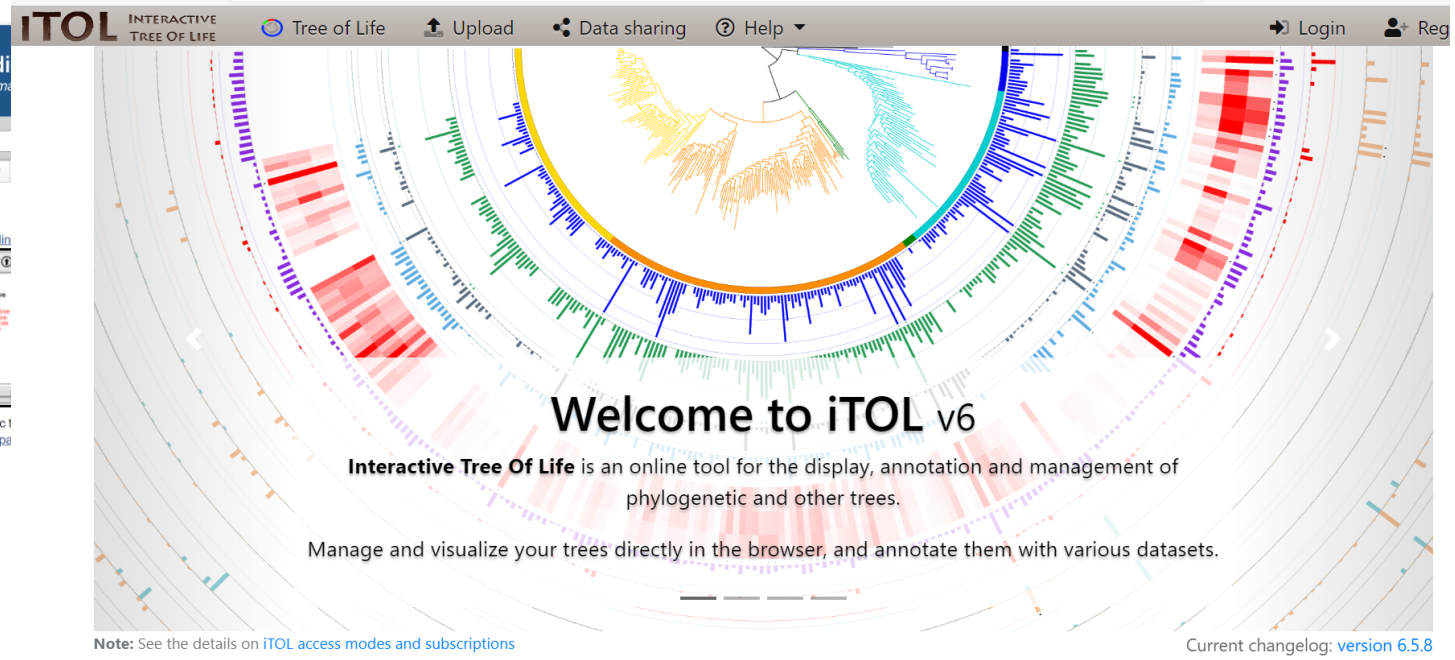
Percentage of reference covered by all isolates: 89.18 (4368832 / 4899014)

Below is the single isolate stats on covered and trusted positions with respect to the reference.

Isolate	Valid positions	Pct. of reference
AMA004497_S24_L555_R1_001.fastq.gz	4435406	90.54
AMA004554_S73_L555_R1_001.fastq.gz	4427220	90.37
AMA004560_S27_L555_R1_001.fastq.gz	4465781	91.16
AMA004627_S69_L555_R1_001.fastq.gz	4412663	90.07
AMA004656_S59_L555_R1_001.fastq.gz	4442114	90.67

Log | Distance matrix | Phylogentic tree | Vcf files of mutations | Reference Sequence | Cluster.dbscan





MINTyper output – VCF data

STATENS SERUM INSTITUT

Percentage of reference covered by all isolates: 89.18 (4368832 / 4899014)

Below is the single isolate stats on covered and trusted positions with respect to the reference.

Isolate	Valid positions	Pct. of reference
AMA004497_S24_L555_R1_001.fastq.gz	4435406	90.54
AMA004554_S73_L555_R1_001.fastq.gz	4427220	90.37
AMA004560_S27_L555_R1_001.fastq.gz	4465781	91.16
AMA004627_S69_L555_R1_001.fastq.gz	4412663	90.07
AMA004656_S59_L555_R1_001.fastq.gz	4442114	90.67

Log | Distance matrix | Phylogentic tree | Vcf files of mutations | Reference Sequence | Cluster.dbscan

AMA004497_S24_L555_R1_001.fastq.gz_alignment.vcf - Notesblok

Filer Rediger Formater Vis Hjælp

```
##fileformat=VCFv4.2
##kmaVersion=1.4.2
##FILTER=<ID=LowQual,Description="Low quality">
##INFO=<ID=DP,Number=1,Type=Integer,Description="Total Depth">
##INFO=<ID=AD,Number=1,Type=Integer,Description="Allele Depth">
##INFO=<ID=AF,Number=1,Type=Float,Description="Allele Fraction">
##INFO=<ID=RAF,Number=1,Type=Float,Description="Revised Allele Fraction">
##INFO=<ID=DEL,Number=1,Type=Float,Description="Fraction of Reads Containing Spanning Deletions">
##INFO=<ID=AD6,Number=6,Type=Integer,Description="Count of all alternative alleles: A,C,G,T,N,-">
##FORMAT=<ID=Q,Number=1,Type=Float,Description="McNemar quantile">
##FORMAT=<ID=P,Number=1,Type=Float,Description="McNemar p-value">
##FORMAT=<ID=FT,Number=1,Type=String,Description="Filter">
#CHROM  POS     ID      REF     ALT     QUAL    FILTER  INFO    FORMAT  bacteria.ATG
NZ_CP024672.1 Citrobacter freundii strain HM38 chromosome, complete genome     338     .       A       a       277     .       DP=76;AD=65;AF=0.86;RAF=0.86
NZ_CP024672.1 Citrobacter freundii strain HM38 chromosome, complete genome     471     .       A       G       367     .       DP=61;AD=61;AF=1.00;RAF=1.00
NZ_CP024672.1 Citrobacter freundii strain HM38 chromosome, complete genome     489     .       C       T       325     .       DP=54;AD=54;AF=1.00;RAF=1.00
NZ_CP024672.1 Citrobacter freundii strain HM38 chromosome, complete genome     492     .       G       T       314     .       DP=56;AD=55;AF=0.98;RAF=0.98
NZ_CP024672.1 Citrobacter freundii strain HM38 chromosome, complete genome     508     .       T       C       264     .       DP=44;AD=44;AF=1.00;RAF=1.00
NZ_CP024672.1 Citrobacter freundii strain HM38 chromosome, complete genome     672     .       C       T       273     .       DP=49;AD=48;AF=0.98;RAF=0.98
NZ_CP024672.1 Citrobacter freundii strain HM38 chromosome, complete genome     756     .       A       a       200     .       DP=50;AD=44;AF=0.88;RAF=0.88
NZ_CP024672.1 Citrobacter freundii strain HM38 chromosome, complete genome     760     .       A       a       194     .       DP=49;AD=43;AF=0.88;RAF=0.88
NZ_CP024672.1 Citrobacter freundii strain HM38 chromosome, complete genome     894     .       T       C       270     .       DP=45;AD=45;AF=1.00;RAF=1.00
NZ_CP024672.1 Citrobacter freundii strain HM38 chromosome, complete genome     1251    .       C       T       338     .       DP=60;AD=59;AF=0.98;RAF=0.98
NZ_CP024672.1 Citrobacter freundii strain HM38 chromosome, complete genome     1548    .       T       G       559     .       DP=97;AD=96;AF=0.99;RAF=0.99
NZ_CP024672.1 Citrobacter freundii strain HM38 chromosome, complete genome     1549    .       T       t       361     .       DP=94;AD=82;AF=0.87;RAF=0.87
NZ_CP024672.1 Citrobacter freundii strain HM38 chromosome, complete genome     1568    .       C       c       355     .       DP=88;AD=78;AF=0.89;RAF=0.89
NZ_CP024672.1 Citrobacter freundii strain HM38 chromosome, complete genome     1569    .       A       G       529     .       DP=88;AD=88;AF=1.00;RAF=1.00
NZ_CP024672.1 Citrobacter freundii strain HM38 chromosome, complete genome     1594    .       A       a       336     .       DP=87;AD=76;AF=0.87;RAF=0.87
NZ_CP024672.1 Citrobacter freundii strain HM38 chromosome, complete genome     1597    .       A       a       324     .       DP=87;AD=75;AF=0.86;RAF=0.86
NZ_CP024672.1 Citrobacter freundii strain HM38 chromosome, complete genome     1604    .       T       t       361     .       DP=89;AD=79;AF=0.89;RAF=0.89
NZ_CP024672.1 Citrobacter freundii strain HM38 chromosome, complete genome     1612    .       A       a       304     .       DP=81;AD=70;AF=0.86;RAF=0.86
NZ_CP024672.1 Citrobacter freundii strain HM38 chromosome, complete genome     1743    .       G       T       385     .       DP=64;AD=64;AF=1.00;RAF=1.00
NZ_CP024672.1 Citrobacter freundii strain HM38 chromosome, complete genome     1753    .       T       G       379     .       DP=63;AD=63;AF=1.00;RAF=1.00
NZ_CP024672.1 Citrobacter freundii strain HM38 chromosome, complete genome     1764    .       C       T       385     .       DP=64;AD=64;AF=1.00;RAF=1.00
NZ_CP024672.1 Citrobacter freundii strain HM38 chromosome, complete genome     1773    .       C       T       391     .       DP=65;AD=65;AF=1.00;RAF=1.00
NZ_CP024672.1 Citrobacter freundii strain HM38 chromosome, complete genome     1777    .       T       C       379     .       DP=63;AD=63;AF=1.00;RAF=1.00
NZ_CP024672.1 Citrobacter freundii strain HM38 chromosome, complete genome     1816    .       G       T       392     .       DP=69;AD=68;AF=0.99;RAF=0.99
NZ_CP024672.1 Citrobacter freundii strain HM38 chromosome, complete genome     2047    .       A       C       270     .       DP=45;AD=45;AF=1.00;RAF=1.00
NZ_CP024672.1 Citrobacter freundii strain HM38 chromosome, complete genome     2100    .       A       G       344     .       DP=61;AD=60;AF=0.98;RAF=0.98
```

MINTyper output – Reference

STATENS SERUM INSTITUT

Percentage of reference covered by all isolates: 89.18 (4368832 / 4899014)

Below is the single isolate stats on covered and trusted positions with respect to the reference.

Isolate	Valid positions	Pct. of reference
AMA004497_S24_L555_R1_001.fastq.gz	4435406	90.54
AMA004554_S73_L555_R1_001.fastq.gz	4427220	90.37
AMA004560_S27_L555_R1_001.fastq.gz	4465781	91.16
AMA004627_S69_L555_R1_001.fastq.gz	4412663	90.07
AMA004656_S59_L555_R1_001.fastq.gz	4442114	90.67

Log | Distance matrix | Phylogentic tree | Vcf files of mutations | Reference Sequence | Cluster.dbscan

template_sequence (2) - Notesblok

Filer Rediger Formater Vis Hjælp

```
>NZ_CP024672.1 Citrobacter freundii strain HM38 chromosome, complete genome
TCGGAATGTGATCAATTTAAAAATTTATTGACTTAGGTCAGCAGATACTTTAACCCTAAAAGAATACAAGACAGACAGATAATAATTACAGAGCACACAACATCATGAAACGCATCAGCATCACTATTACCACAACCATCACCATTACCACAGC
ACTTCGGCGCCAAAGTGCTGCACCCGCGCACCATTACCCCTATTGCCCAGTTCCAGATCCCTTGCCTGATTAAAAATACCGGCAATCCACAAGCGCCTGGCACGTTGATTGGCGCCAGCAGTGATGACGACGATTTGCCGGTAAAAGGTATTTC
GCGTAAATTCCTCTACGACACCAACGTGGGCGCAGGCTTGCCGGTAATTGAAAACCTGCAAAACCTGCTCAGCGCAGGTGATGAATTGCAGCGTTTCTCCGGTATTCTTTCCGGCTCGCTGTCGTTTATTTTCGGCAAGCTGGATGAAGGCATC
TACGCGCCTGTTGGCGATAATGGGCGAGCTGGAAGGGCGTATCTCCGGCAGTATTCATTACGATAATGTTGCCCCGTGCTTTTTGGGCGGCATGCAGTTAATGATCGAAGAAAACGGCATTATCAGCCAGCAGGTGCCAGGCTTTGATGAGTGC
GTGCGCGTGGTGATCCTCTACCCGAACGGCAAGATCAGCCCATTACAGGAAAAACTGTTCTGTACGCTGGGCGGTAACATTGAAACTGTAGCGATTGATGGCGATTTTGATGCCTGCCAGGCGCTGGTGAAACAGGCGTTTGATGACGAAGAGC
AAGCGGTTCTCAATGATGTAGCGGCTCATCAGGCCGCGCGCTTTCTTGGCATAAAAGCTGATGACTTTGAACTTGCCGTTTTTCTCGTCTAGGAAGACTGGTTTAATCAGCTCGGCATTGAGCTTTTTCGGCTTCACGGATTTAAAATACTCAT
GATGCGCATATTGTCCAGTTTCAGAAAGACGAGATTGTTCTCCGCATAGATGTAGTTGGCGACGATCGAGCTAAAGGCAAACAGAATCACAATAAAAGCCACGAATCCGGCTccCCATTCTCCCGTCAGCGTTACCATTGCTTTTTGAAGTAAC
CCAAACTGGCGAATGTAGCGGAACTGCACGAACCCTGTACGCCAGGTGAACCAACATCCTGCGCCAAGCAGGAGGTAGATCATCACCGACCCCCAGAGTATTTCATTAATAAATGAGAAAAAATCAGGCATTAACATCCCTCTTGTTGATGCCC
CGGAAATCTACGAGTACTACAAACAGCACGGTTACGAAACCGTGGTGATGGGTGCGAGCTTCCGTAACATCGGCGAAATCATTGAGCTGGCAGGCTGCGATCGTCTGACGATTGCTCCGGCACTGCTGAAAGAGCTGGCGGAAAGCGAAGGCGC
AACAGCTGAAAGTAGCCCGATCGTGACATAAGTGCCAATCACAGTGGAGTATTGAATTGTTTACGATATAGTAATTTTTGCTTTACATATATCGACAGATAGATTCAATAACTCCACTGACTGATGGTTTGCTATGTCACAACACACTCGACCC
GGATTTATTGTCGACCGGGTCGGCGCCAGCAAAACCTTTATTGTCGGCAGCCTGCTGCTGGCCTGTTCGAGCTGGTTTTTCTATCACCTGACCGGCAGCCATCCTCAGCACTTGTTCCTGTTATACGGGCTGGTGGGATTATGCGTTGGCGTGC
CAAAATCATCATTGATTGAATGGTGAAATAGTTTCCCCAAATAACGATCACTGTCTTCGGGGCGCGGCATAATAATCAGGGGAGAGGGCACTGTCTATGATCTAACGAAGGGAAAACGAATTATTTTCCCTGTGATGGGCATCACGCTTGTGCC
TTGATGGAGGAAGGCACGTTCAGACTCTGGCTGGACATGCGGGCAGCTTTGAAATAAACCGATGCACCACTGAGCTGTAAATCACCATGATCGGCCGTAAGTTGAATGCGTTTCACCACGCGGCAAACGGGAAGTTTCAGCGTCAGATCGTTGC
AAAACGGCTGAAGATTACCTGGGTGAGCCGGTAACTGAAGCAGTTATCACCGTACCTGCATACTTTAACGATGCTCAGCGTCAGGCAACTAAAGACGCCGGTCGTATCGCAGGTCTGGAAGTAAAACGTATCATCAACGAACCAACCGCAGCTC
CGCGCGGTATGCCGCAGATCGAAGTTACTTTTGACATCGATGCCGACGGTATCCTGCACGTTTCCGCGAAAGACAAAAACAGCGGTAAAGAGCAGAAGATCACCATCAAGGCTTCTTCTGGTCTGAACGAAGAAGAAATTCAGAAAATGGTTCC
GTATTCCGACTCTGGAAGAGTGTGACGTCTGCCACGGTAGCGGCGCGAAGGCGGGCACTCAGCCGCAGACCTGTCCAACCTGTCACGGTTCTGGTCAGGTACAGATGCGTCAGGGCTTTTTTGCCGTACAGCAGACCTGTCCACACTGTCAGGC
ACACTACCAGTAATAACGATTCACCTTCCTTAATGTGAGATTTACATCACTATCCTGTCATTCTATTTCATTATTATTTGTCACCGCGGTAACACAATATTTATTGATGGATTAATTATAATTATGGGGAATACATTCAATGCCTCGGATAAAA
TGATTATTATGGCTGATGATTTGGGCTATGGCGATCTCGCGACTTATGGCCATCAAATCGTCAAAACGCCTAATATCGACAAGCTGGCACAGGAAGGGGTGAAATTTACCGACTACTATGCGCCCGCGCCGCTGTGCTCTCCTTCTCGTGCGGC
CATGATGAATTTTAAACTGCCAACAGATCGTACCTACGATGGGCAGTCTCTGGTTCCATTACTTGAACAGAAAACGTTAGCACGTCAGAAACCACTCATCTTTGGCATTGATATGCCGTTCCAGGATGATCCTACTGACGAATGGGCGATCCGT
GAAAAGCATTGGCAGTCGCTATATGCAGTTTATTCCGCTCATCGAAAGAAGCGCTGAAGAGGGCGGTATTGGTCTGGCACATCCTGAAGGCGGACAAACAAGAATGATGCCGTGGTCGGTTGATTCACTGCAATTTGGCCATTTTCTCAACGCT
```

MINTyper output – Cluster analysis

STATENS SERUM INSTITUT

Percentage of reference covered by all isolates: 89.18 (4368832 / 4899014)

Below is the single isolate stats on covered and trusted positions with respect to the reference.

Isolate	Valid positions	Pct. of reference
AMA004497_S24_L555_R1_001.fastq.gz	4435406	90.54
AMA004554_S73_L555_R1_001.fastq.gz	4427220	90.37
AMA004560_S27_L555_R1_001.fastq.gz	4465781	91.16
AMA004627_S69_L555_R1_001.fastq.gz	4412663	90.07
AMA004656_S59_L555_R1_001.fastq.gz	4442114	90.67

Log | Distance matrix | Phylogentic tree | Vcf files of mutations | Reference Sequence | Cluster.dbscan

other isolates Cluster

```
# 5     3       10.000000       1
"AMA004497_S24_L555_R1_001.fastq.gz_alignment.fsa"      0       0
"AMA004554_S73_L555_R1_001.fastq.gz_alignment.fsa"      2       1
"AMA004560_S27_L555_R1_001.fastq.gz_alignment.fsa"      0       2
"AMA004627_S69_L555_R1_001.fastq.gz_alignment.fsa"      2       1
"AMA004656_S59_L555_R1_001.fastq.gz_alignment.fsa"      2       1
```

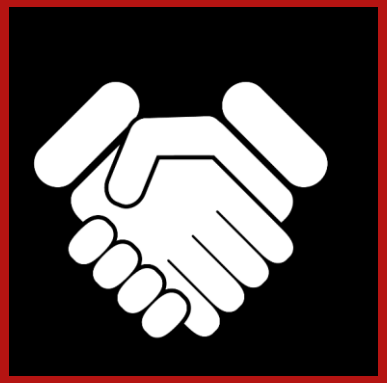
```
        5
AMA004497_S24_L555_R1_001.fastq.gz_alignment.fsa
AMA004554_S73_L555_R1_001.fastq.gz_alignment.fsa        17
AMA004560_S27_L555_R1_001.fastq.gz_alignment.fsa        1280    1275
AMA004627_S69_L555_R1_001.fastq.gz_alignment.fsa        17      0       1275
AMA004656_S59_L555_R1_001.fastq.gz_alignment.fsa        17      0       1275    0
```

STATENS
SERUM
INSTITUT

Questions?

STATENS SERUM INSTITUT

Thank you for listening!



Ana Rita Rebelo and Jette Sejer Kjeldgaard
anrire@food.dtu.dk ; jetk@food.dtu.dk

Updates to the harmonized EURGen-RefLabCap WGS protocols

Funded by the European Union

WGS protocols – EURGen-RefLabCap project

Agreed common WGS-based genome analysis methods and standard protocols for national CCRE surveillance and integrated outbreak investigations

Version n°: 1.0
Date: 01-09-2022

EURGen-RefLabCap agreed common WGS-based genome analysis methods and standard protocols for national surveillance and integrated outbreak investigations of carbapenem- and/or colistin-resistant *Pseudomonas aeruginosa* and *Acinetobacter baumannii*

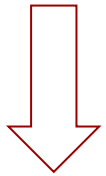
Version n°: 1.0
Date: 14-04-2023

https://www.eurgen-reflabcap.eu/resources/protocols-and-guidelines

Funded by the European Union

WGS protocols – EURGen-RefLabCap project

> This guidance document provides a framework to perform WGS directed towards short-read paired-end massive parallel synthesis sequencing, specifically using Illumina platforms (Illumina, Inc., San Diego, CA, USA) such as MiSeq and NextSeq. In addition, it



Some of the steps are not directly applicable for analyses of ONT data

Funded by the European Union

Adjusting the protocols for analyses of ONT data

- From a primary culture, select one single isolated colony to prepare a subculture.
- Inspect the subculture carefully to ensure purity. If the culture is not pure, prepare a new subculture.
- Extract bacterial DNA using in-house protocols or commercial kits.
- Measure UV 260/280 and 260/230 absorbance ratio values of the DNA samples to confirm that they are in the interval 1.8–2.0 and 2.0-2.2, respectively.
- Measure the concentration of the undiluted DNA samples.
- Confirm the DNA concentration of the diluted samples.
- The DNA dilution and confirmation of the DNA concentration should be repeated until the desired concentration is achieved.

Funded by the European Union

Adjusting the protocols for analyses of ONT data

Extract bacterial DNA using in-house protocols or commercial kits:

For gram-negative bacterial gDNA, ONT recommends using the QIAGEN Genomic-tip 500/G columns.

In addition to the kits recommended by ONT, there are many other kits designed for long read sequencing and ONT platforms. For example, the Monarch® HMW DNA Extraction Kit for Tissue provides a rapid and reliable process for extracting HMW genomic DNA from bacteria. There are separate protocols for processing Gram-negative and Gram-positive bacteria are provided and differ slightly in the initial lysis step. DNA size ranges from 50 – ≥ 500 kb for the standard protocols and can be adjusted to produce longer DNA into the Mb range for bacteria. Another example is the Zymo Quick-DNA HMW MagBead Kit generates pure DNA up to 150 kb.

Be aware that extraction methods that vortex or mix samples can break the DNA and create shorter reads, therefore it is recommended to flick or invert the tubes gently and avoid rigorous mixing. Additionally, when using kits with magnetic beads, make sure there are no beads in the final eluate. Moreover, extraction methods based on salt and ethanol precipitation can result in poor plasmid extraction, which can be problematic for determination of antimicrobial resistance (AMR) genes, as these often reside on plasmids.

Take care to note any additional reagents, materials, or equipment you may need to order that is not included in the kit but required for the DNA extraction procedure.

Funded by the European Union

Adjusting the protocols for analyses of ONT data

And some additional considerations regarding the DNA concentration:

Example: a concentration of 50 ng/µl per sample is recommended for the Rapid Sequencing Barcoding Library Prep kit.

The samples should be adjusted to a final volume of 10 µl (diluted with nuclease free water if the concentration is too high).

It is recommended to double or even quadruple the concentration of DNA for library preparation. Thus, in the example above, it is recommended to use 100ng/ µl or even 200ng/ µl if possible. This is because typically the more DNA in your starting material, the more successful your sequencing run will be with more data generated.

Funded by the European Union

Adjusting the protocols for analyses of ONT data

- Perform library preparation.
- Pool libraries and load the sequencer following the manufacturer’s instructions.
- Extract the raw reads and store them locally.
- Perform QC of the sequence data.

Funded by the European Union

Adjusting the protocols for analyses of ONT data

Perform library preparation:

ONT is one of the most widely used long read sequencing platforms, and protocols with preparation guidelines for specific library kits and guidelines for sequencing on the specific machinery are frequently updated and available on the ONT website. Examples are the Rapid sequencing gDNA barcoding Kit or the Rapid sequencing DNA PCR Barcoding Kit.

Funded by the European Union

Adjusting the protocols for analyses of ONT data

And some additional considerations regarding preparing the sequencing:

You need a specific software: MinKNOW

MinKNOW is ONTs software which controls sequencing devices (e.g., MinION). The software performs several core tasks, including sequencing, real-time analysis, basecalling, and data generation. The software writes out the sequencing data into .POD5, .FAST5 and FASTQ files.

To install MinKNOW for MinION sequencing, first ensure your computer or laptop meets the requirements. Then, download MinKNOW from the ONT community Downloads page. If you are using a GridION, MinION Mk1C or a PromethION 24/48, MinKNOW can be directly installed on the sequencing device.

Funded by the European Union

Adjusting the protocols for analyses of ONT data

And more additional considerations regarding preparing the sequencing:

You need to perform flowcell check before starting the sequencing.

This should be performed when you are ready to use the flow cell, and within 3 months of purchase.

Before the library is loaded onto the flow cell and sequencing can begin, it is important to perform a flow cell check to ensure there is an adequate number of active pores that are covered by the ONT warranty. For a MinION/GridION Flow Cell, the minimum number of active pores covered by the warranty is 800. It is not recommended to perform a sequencing run on a flow cell with below minimum active pores.

It should be noted Oxford Nanopore Technologies will replace any flow cell that falls below the warranty number of active pores, provided the result is reported within two days of performing the Flow Cell Check and the storage recommendations have been followed (stored unopened at room temperature for one month, or at 2–8°C for 12 weeks).

Funded by the European Union

Adjusting the protocols for analyses of ONT data

Pool libraries and load the sequencer following the manufacturer's instructions:

The libraries (DNA samples) are already pooled (multiplexed) during the library preparation protocols.

After the flow cell check, the library can be loaded onto the flow cell. The flow cell is then inserted into the MinION which should be connected to your laptop. The sequencing experiment can then be set up in MinKNOW and the appropriate parameters can be chosen for your experiment.

Log into MinKnow using log using your community credentials (recommended) or continue as a guest. Select the sequencing device connected to the computer (e.g MinION) and select start sequencing. Type in the experiment name, sample ID and choose flow cell type and then the kit used.

During kit selection, if you have used a barcoding kit or a barcoding expansion pack for your library preparation, MinKNOW will split your reads by barcode without having to use command line tools. Demultiplexing places reads into barcode-specific folders.

Funded by the European Union

Adjusting the protocols for analyses of ONT data

Pool libraries and load the sequencer following the manufacturer's instructions (cont.):

The run options tab provides variables for run time and minimum read length. To enable Short Fragment Mode (SFM), select the preferred minimum read length, from as short as 20 bp. This directs the software to write sequencing files from the minimum size read length selected.

The basecaller algorithm in MinKnow, Guppy, provides multiple models for basecalling. It is recommended to choose high-accuracy or super-accurate basecalling. However, it should be noted the higher accuracy basecalling algorithms require more computational power.

Choose the local destination for where the reads will be stored and start the sequencing experiment.

Funded by the European Union

Adjusting the protocols for analyses of ONT data

Extract the raw reads and store them locally:

ONT data is stored in POD5 and FASTQ files. The raw data will be stored locally in the destination chosen when setting up the sequencing run on MinKNOW.

POD5 is an Oxford Nanopore-developed file format which stores ONT data in an accessible way and replaces the legacy .fast5 format. This output also reads and writes data faster, uses less compute and has smaller raw data file size than .fast5. fast5 is a customised file format based upon the .hdf5 file type, which is designed to contain all information needed for analysing nanopore sequencing data and tracking it back to its source.

FASTQ is a universal text-based sequence storage format, containing both the sequence of DNA and its quality scores. By default, the MinION saves up to 4000 DNA sequences in one FASTQ file. File size can vary from < 1Mb to tens of Mb depending on the number and length of sequences.

Funded by the European Union

Adjusting the protocols for analyses of ONT data

Perform QC of the sequence data:

During live basecalling, MinKnow provides real-time QC assessments including read quality and length. These metrics are presented as the run report and can be exported post sequencing. The run report includes the run summary and configuration, sequence output such as read lengths sequenced with and without their outliers, along with the cumulative output and quality score, run health and run log. Troubleshooting suggestions are also available throughout the report, with links to further information available on the ONT Community site.

Additional QC metrics can be generated using ONT's EPI2ME labs platform*.

* ONT has two data analysis platforms called EPI2ME and EPI2ME labs. EPI2ME is cloud-based and performs all workflows in real time with no bioinformatics expertise needed, whereas EPI2ME labs is local or distributed (e.g runs on your laptop of cluster) and only selected workflows are real-time.

Funded by the European Union

Adjusting the protocols for analyses of ONT data


Assemble the reads (FASTQ files) into contigs (FASTA files).


Perform QC of the assembly.



Analyse your data.

Funded by the European Union

Adjusting the protocols for analyses of ONT data

Assemble the reads (FASTQ files) into contigs (FASTA files):

EPI2ME labs has a bacterial genome assembly workflow which can be employed to assemble, variant call and annotate bacterial genomes. A FASTA file is generated in addition to a QC HTML report.

Genome assembly can also be performed using command-line tools such as Flye or Unicycler. Assemblies can then be polished with tools such as Medaka.

Most assembly programs can be installed locally, and many institutions performing WGS routinely have this step incorporated into their analysis pipeline.

Funded by the European Union

Adjusting the protocols for analyses of ONT data

Analyse your data:

Use assembled genomes as input and follow the same procedures described in the project protocols.

Use EPI2ME or EPI2ME labs to analyse your ONT data.

Use other strategies discussed previously in the workshop.

Funded by the European Union

EURGen-RefLabCap@food.dtu.dk

Thank you on behalf of the EURGen-RefLabCap team

Funded by the European Union