Practical Exercise: Analysing RSV Genomic Data

Bioinformatics Tools for RSV Analysis

Jérôme Bourret, Institut Pasteur

10/12/2024

Intended Learning Objectives

Specific objectives of this session:

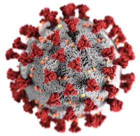
1. Know the key analyses of RSV sequencing data
2. Have a grasp on available tools and platforms
3. Perform a « on the go » analysis of an example dataset
4. Do a practical training on public data

NRC mission

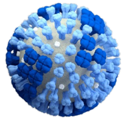
Reception of samples from North of France

Focus on:

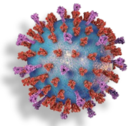
- SARS-CoV-2



- Influenza viruses

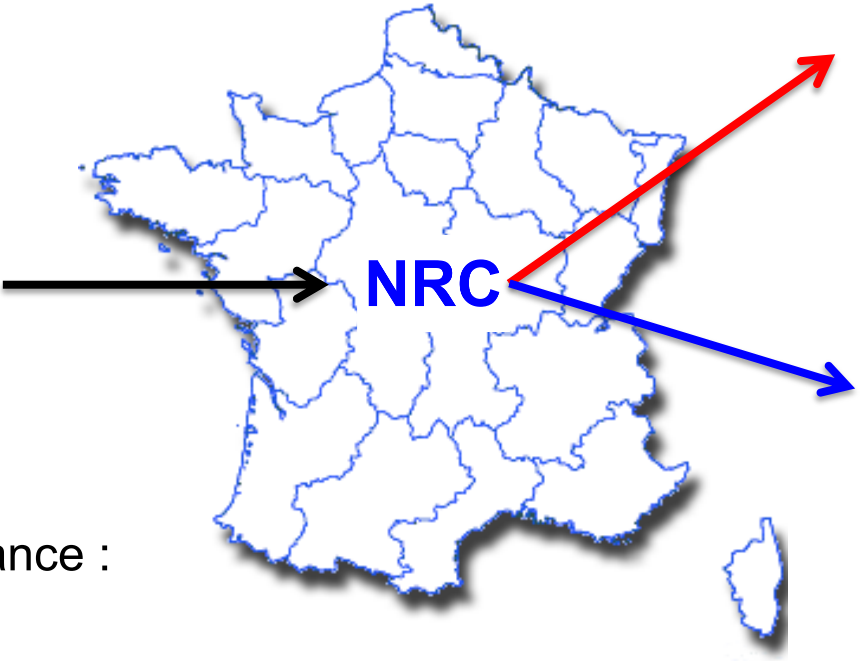


- **RSV**



Networks:

- ***RENAL***
- ***RELAB***
- ***Sentinelle***



NRC

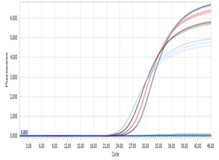
French health authorities

WHO

➔ Mission to inform health autorities on epidemiological surveillance :

- Antigenic variation
- Antiviral sensitivity

RT-PCR



Positive

- SARS-CoV-2
- Flu
- RSV

NGS Sequencing

Sequencing data Analysis & Epidemiology surveillance

Virus isolation

Phenotypic characterisation

How do we analyse RSV genomic data?

Goals

- Track the diversity and evolution of RSV samples
- Detect mutations of interest

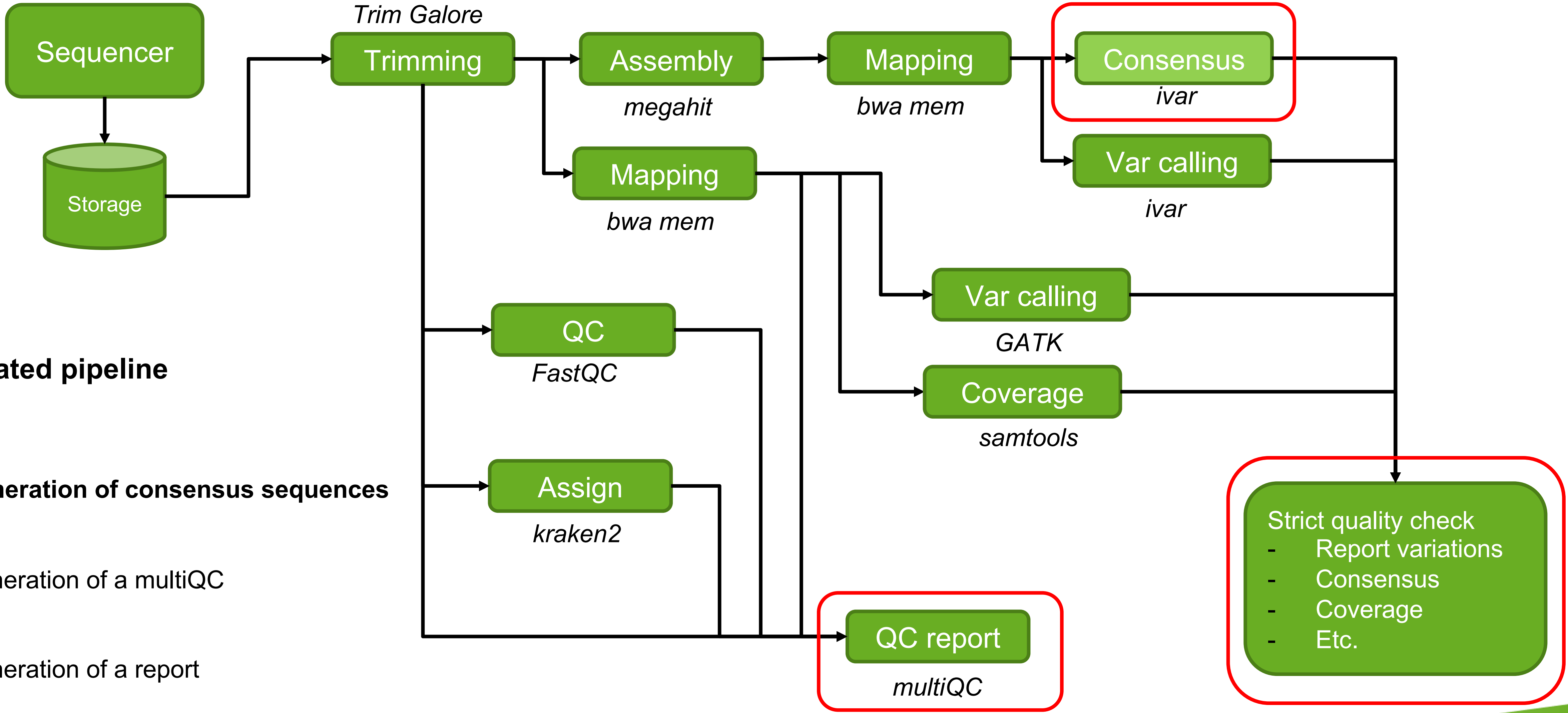
➔ First, we need to recreate the genomes from our samples using sequencing data!

Sequencing data analysis


S. Berreira


K. Da Silva

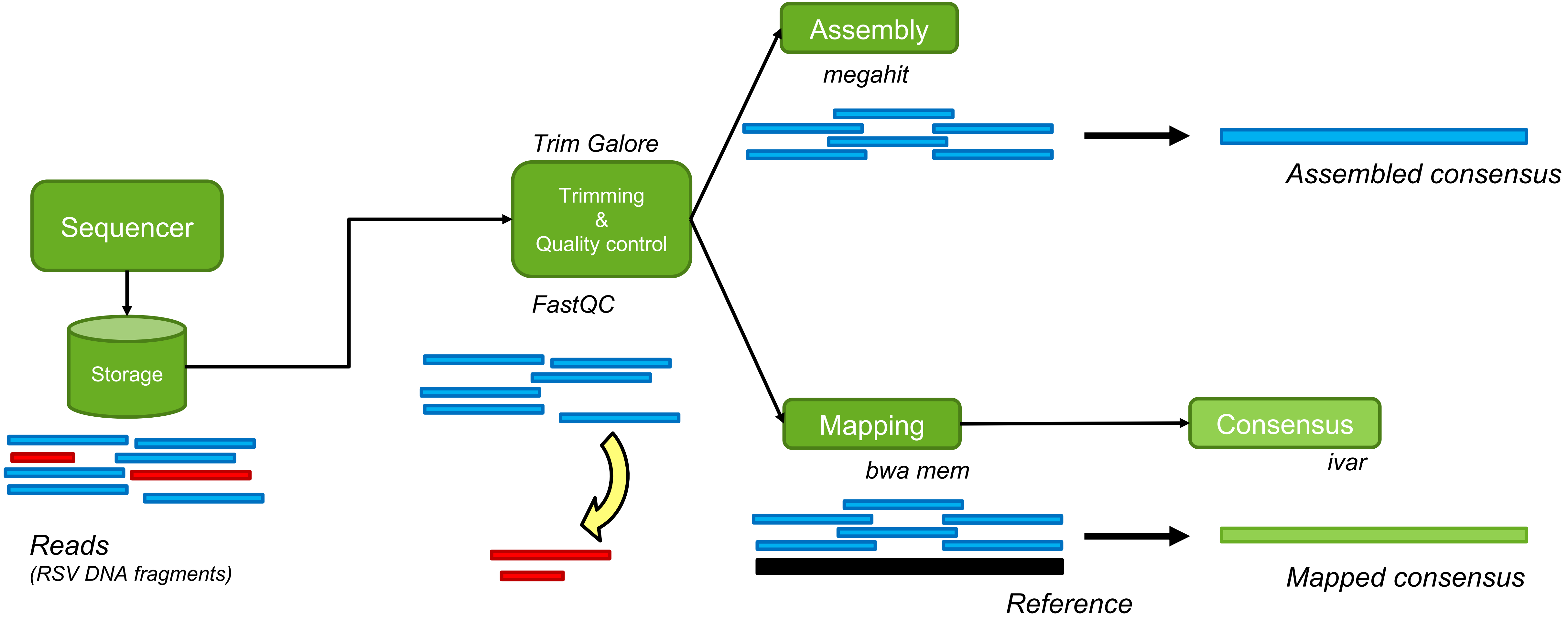

F. Lemoine



Dedicated pipeline

- **Generation of consensus sequences**
- Generation of a multiQC
- Generation of a report

Sequencing data analysis



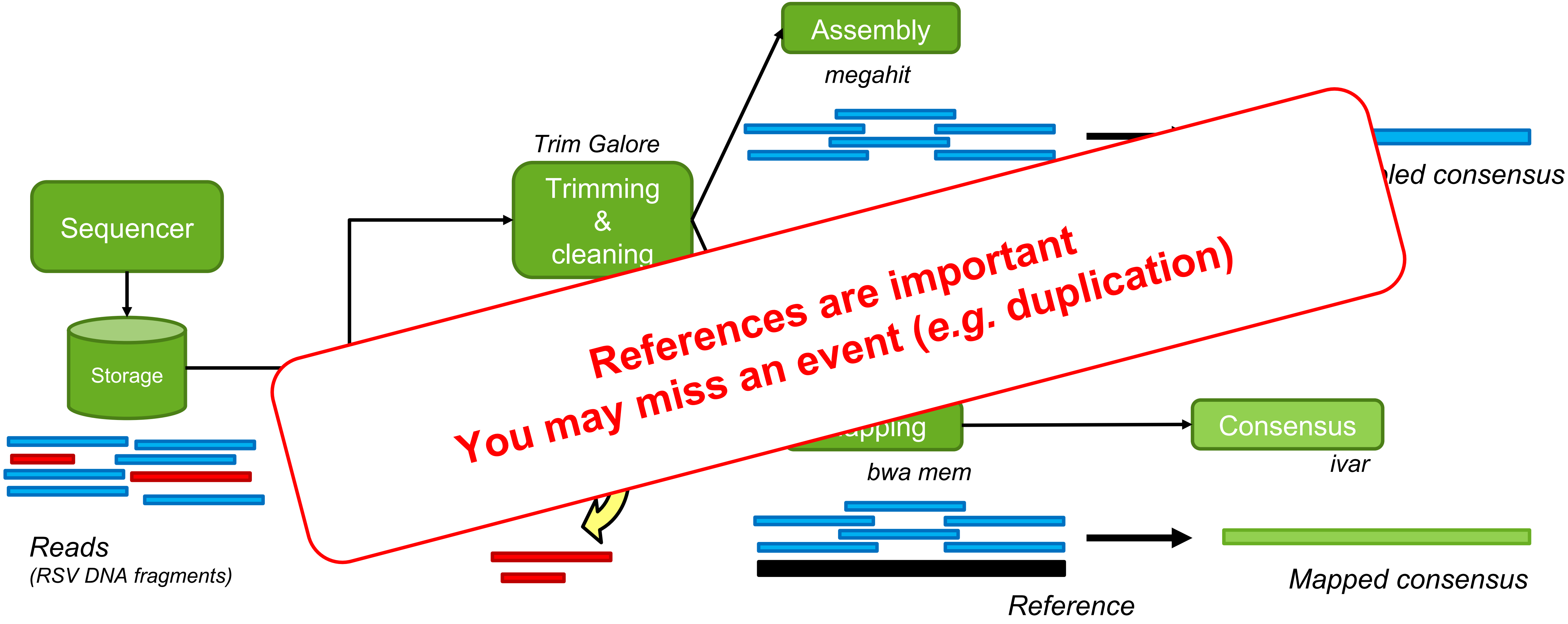
slido

Please download and install the
Slido app on all computers you use

When mapping reads, what is the best reference to use?

ⓘ Start presenting to display the poll results on this slide.

Sequencing data analysis



Sequencer

Storage

Reads

(RSV DNA fragments)

Trim Galore

Trimming & cleaning

Assembly

megahit

led consensus

apping

bwa mem

Consensus

ivar

Reference

Mapped consensus

References are important

You may miss an event (e.g. duplication)

How do we analyse RSV genomic data?

Goals

- Track the diversity and evolution of RSV samples
- Detect mutations of interest

➔ First, we need to recreate the genomes from our samples using sequencing data!

➔ We need to validate our genomes!

- Confirm the subtype
- Determine the clades
- Get infos on your data such as:
    - Coverage of the genome
    - Frameshifts
    - Other quality criteria

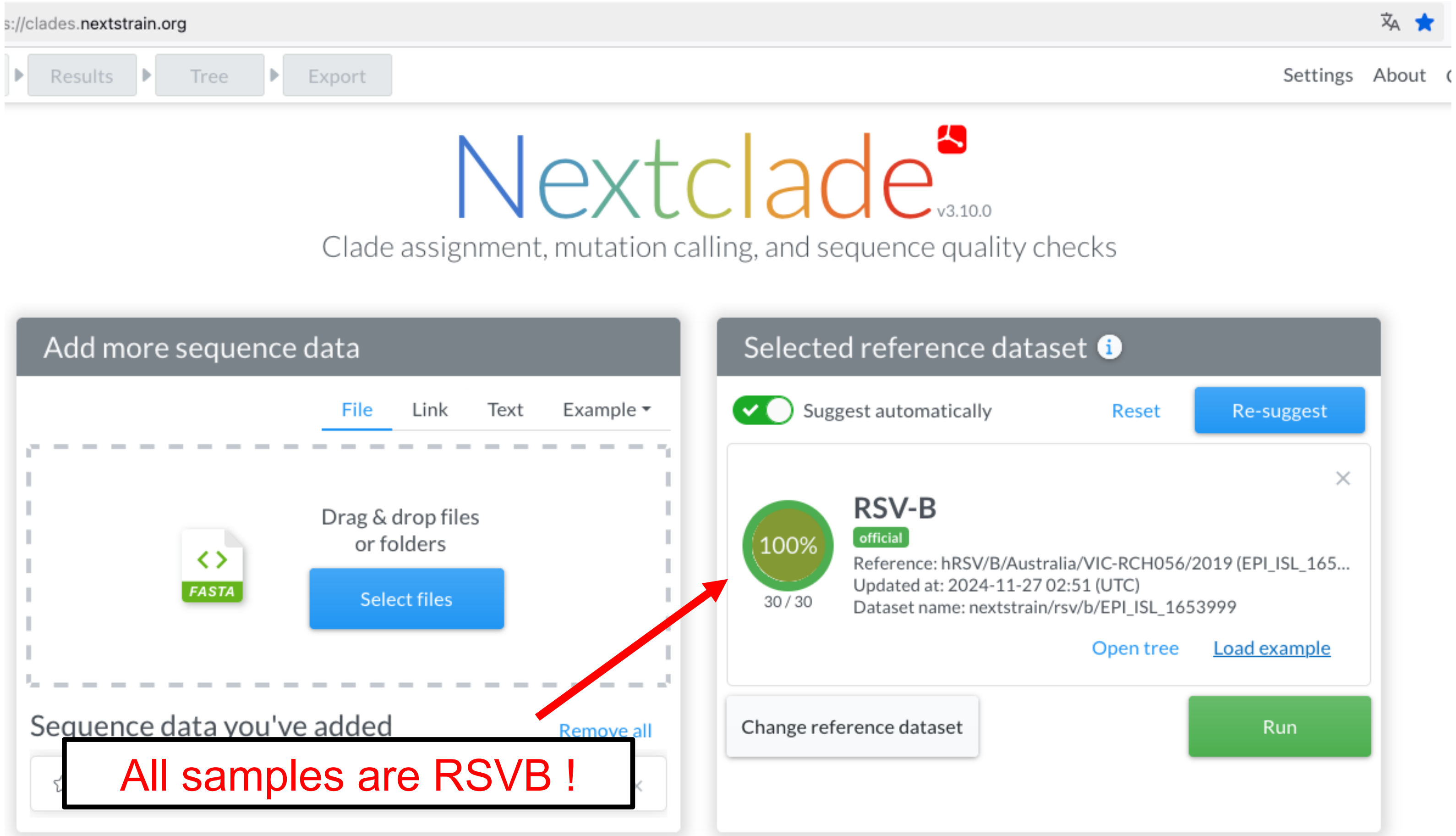
A dedicated tool can perform all these tasks! Nextclade

Nextclade

Takes fasta file as input

Uses a reference to work.
(the reference is automatically selected, but can be chosen manually)

- Determine quickly if RSVA or RSVB
- Align sequences to the reference
- Determine the mutational profile along genome / genes
- Give a quality score to the sequence
- Place them onto a phylogeny



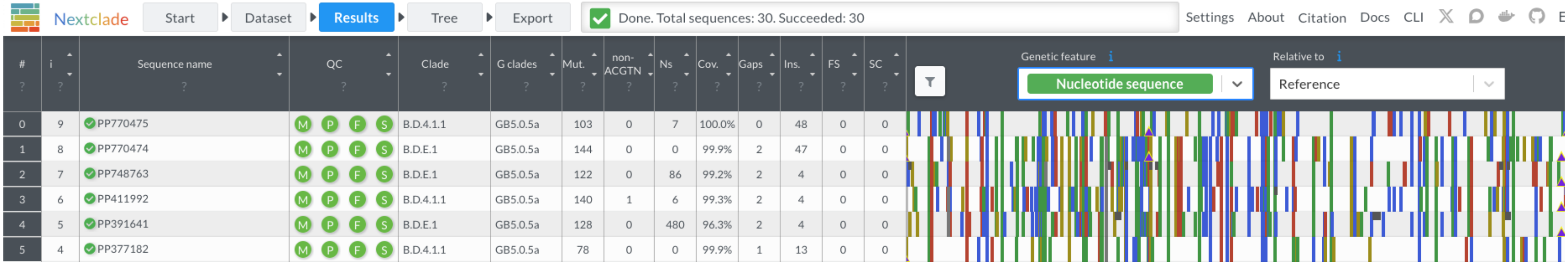
Clade assignment, mutation calling, and sequence quality checks

Reference: hRSV/B/Australia/VIC-RCH056/2019 (EPI_ISL_165...
Updated at: 2024-11-27 02:51 (UTC)
Dataset name: nextstrain/rsv/b/EPI_ISL_1653999

All samples are RSVB !

https://clades.nextstrain.org

Nextclade



Done. Total sequences: 30. Succeeded: 30

#	i	Sequence name	QC	Clade	G clades	Mut.	non-ACGTN	Ns	Cov.	Gaps	Ins.	FS	SC
0	9	PP770475	M P F S	B.D.4.1.1	GB5.0.5a	103	0	7	100.0%	0	48	0	0
1	8	PP770474	M P F S	B.D.E.1	GB5.0.5a	144	0	0	99.9%	2	47	0	0
2	7	PP748763	M P F S	B.D.E.1	GB5.0.5a	122	0	86	99.2%	2	4	0	0
3	6	PP411992	M P F S	B.D.4.1.1	GB5.0.5a	140	1	6	99.3%	2	4	0	0
4	5	PP391641	M P F S	B.D.E.1	GB5.0.5a	128	0	480	96.3%	2	4	0	0
5	4	PP377182	M P F S	B.D.4.1.1	GB5.0.5a	78	0	0	99.9%	1	13	0	0

Quality scores

Clades
(based on genome or G ORF)

Overview of sample

Mutations
(v.s. reference)

All of these metrics / informations can be downloaded!

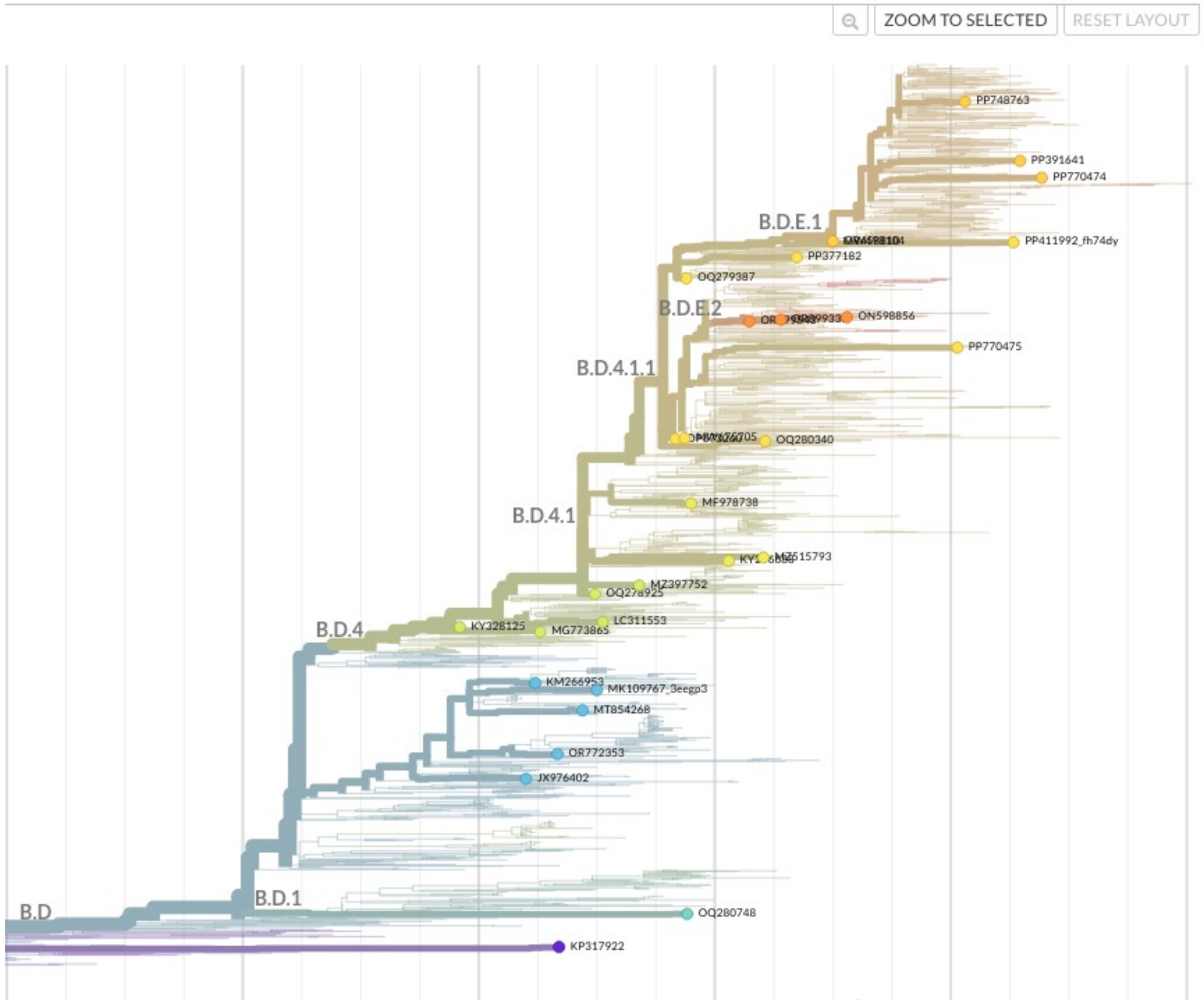
https://clades.nextstrain.org

Nextclade

Nextclade places samples onto a phylogeny with sequences that represents RSV-A/B diversity

1. Confirms the subtype / clade assignation
2. Is a good way to determine samples with a lot of mutations
3. A good first measure of diversity

It’s a phylogenetic placment, not a phylogeny!



https://clades.nextstrain.org

Check also Nextstrain, that tracks phylogeny in almost real-time !

How do we analyse RSV genomic data?

Goals

- Track the diversity and evolution of RSV samples
- Detect mutations of interest

➔ First, we need to recreate the genomes from our samples using sequencing data!

➔ We need to validate our genomes!

➔ We need to analyze the profiles of our genomes and compare them to other RSVs!

But where do we find other samples?

Where to contextualize my data?

Apart from our own samples, it is important to gather sequences to:

- Contextualize samples in global evolutionary tendencies
  (we need to have this kind of data to do phylogenies!)
- Detect patterns of transmission
  (can you retrace history of transmission?)
- Check for new mutations / rare mutations in your samples
  (determine frequencies of mutations)

Thousands of RSV samples can be found in several databases:

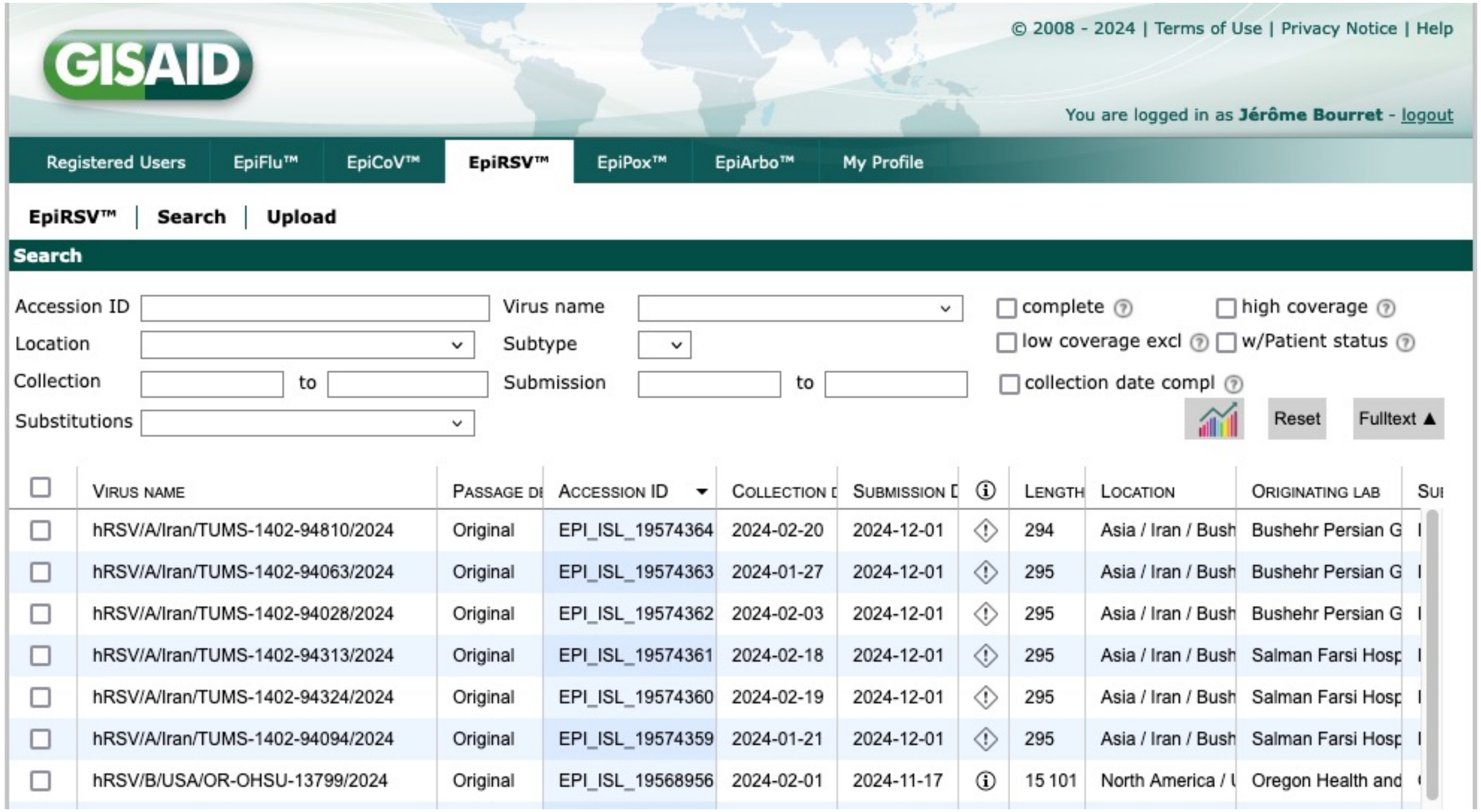
- Public ones (NCBI ; BV-BRC)
- Private ones (GISAID ; account and access must be granted)

GISAID database

« Search » tab allow to gather data

You can filter by:

- RSV subtype
- Date
- Location
- size of sequences
  (to get complete genomes)
- …



☐	Virus name	Passage de	Accession ID	Collection d	Submission d	ⓘ	Length	Location	Originating lab	Su
☐	hRSV/A/Iran/TUMS-1402-94810/2024	Original	EPI_ISL_19574364	2024-02-20	2024-12-01	◇	294	Asia / Iran / Bush	Bushehr Persian G	I
☐	hRSV/A/Iran/TUMS-1402-94063/2024	Original	EPI_ISL_19574363	2024-01-27	2024-12-01	◇	295	Asia / Iran / Bush	Bushehr Persian G	I
☐	hRSV/A/Iran/TUMS-1402-94028/2024	Original	EPI_ISL_19574362	2024-02-03	2024-12-01	◇	295	Asia / Iran / Bush	Bushehr Persian G	I
☐	hRSV/A/Iran/TUMS-1402-94313/2024	Original	EPI_ISL_19574361	2024-02-18	2024-12-01	◇	295	Asia / Iran / Bush	Salman Farsi Hosp	I
☐	hRSV/A/Iran/TUMS-1402-94324/2024	Original	EPI_ISL_19574360	2024-02-19	2024-12-01	◇	295	Asia / Iran / Bush	Salman Farsi Hosp	I
☐	hRSV/A/Iran/TUMS-1402-94094/2024	Original	EPI_ISL_19574359	2024-01-21	2024-12-01	◇	295	Asia / Iran / Bush	Salman Farsi Hosp	I
☐	hRSV/B/USA/OR-OHSU-13799/2024	Original	EPI_ISL_19568956	2024-02-01	2024-11-17	ⓘ	15 101	North America / U	Oregon Health and	(

https://gisaid.org/

To date, one of the most complete database on RSV sequences

slido

Please download and install the
Slido app on all computers you use

Do you have access to the GISAID database?

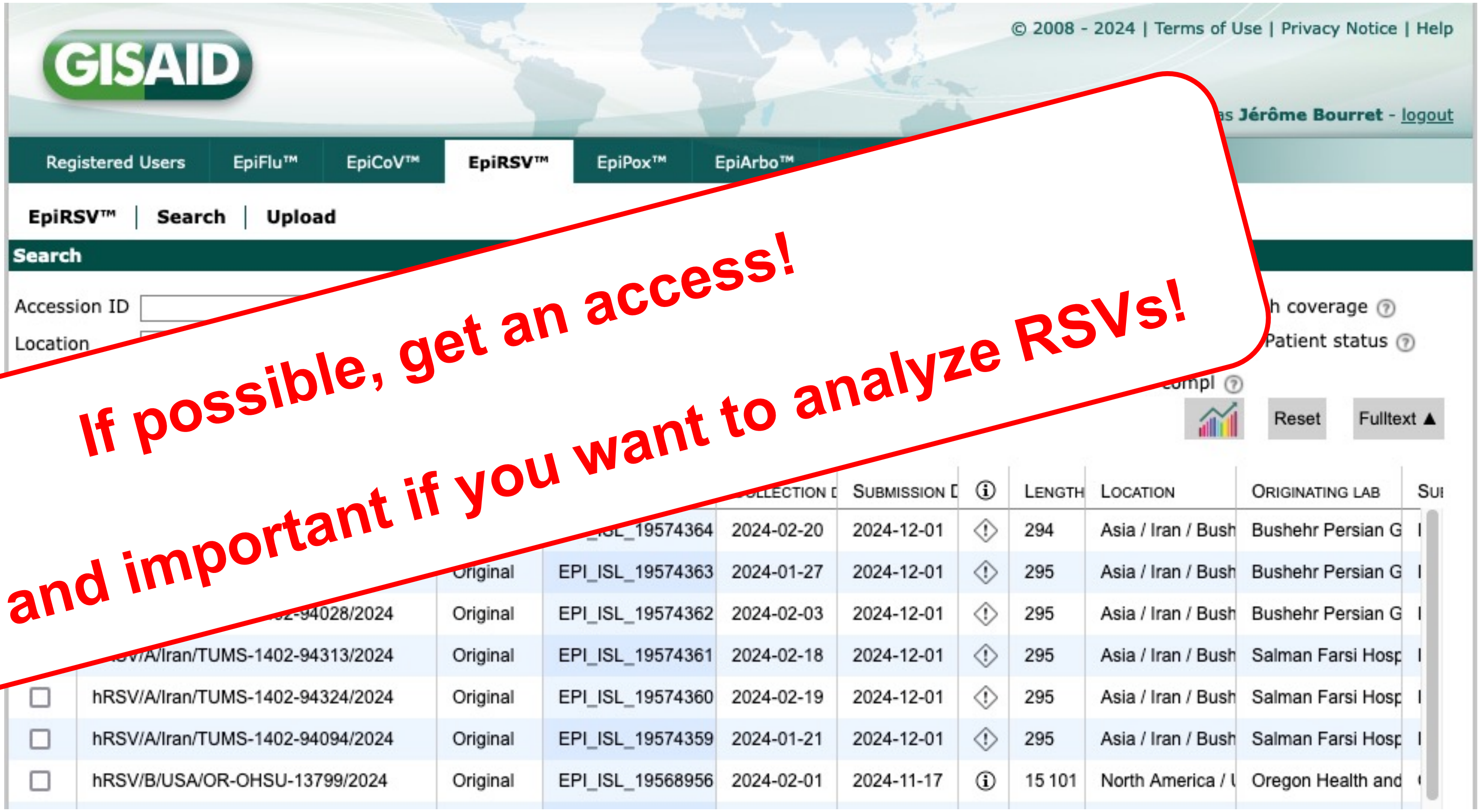
ⓘ Start presenting to display the poll results on this slide.

GISAID database

« Search » tab allow to gather data

You can filter by:

- RSV subtype
- Date
- Location
- size of sequences
  (to get complete genom
- …



				ollection d	Submission d	ⓘ	Length	Location	Originating lab	Su
			ISL_19574364	2024-02-20	2024-12-01	◈	294	Asia / Iran / Bush	Bushehr Persian G	I
		Original	EPI_ISL_19574363	2024-01-27	2024-12-01	◈	295	Asia / Iran / Bush	Bushehr Persian G	I
	2-94028/2024	Original	EPI_ISL_19574362	2024-02-03	2024-12-01	◈	295	Asia / Iran / Bush	Bushehr Persian G	I
	V/A/Iran/TUMS-1402-94313/2024	Original	EPI_ISL_19574361	2024-02-18	2024-12-01	◈	295	Asia / Iran / Bush	Salman Farsi Hosp	I
☐	hRSV/A/Iran/TUMS-1402-94324/2024	Original	EPI_ISL_19574360	2024-02-19	2024-12-01	◈	295	Asia / Iran / Bush	Salman Farsi Hosp	I
☐	hRSV/A/Iran/TUMS-1402-94094/2024	Original	EPI_ISL_19574359	2024-01-21	2024-12-01	◈	295	Asia / Iran / Bush	Salman Farsi Hosp	I
☐	hRSV/B/USA/OR-OHSU-13799/2024	Original	EPI_ISL_19568956	2024-02-01	2024-11-17	ⓘ	15 101	North America /	Oregon Health and	

If possible, get an access!

It's easy and important if you want to analyze RSVs!

https://gisaid.org/

To date, one of the most complete database on RSV sequences

How do we analyse RSV genomic data?

Goals

- Track the diversity and evolution of RSV samples
- Detect mutations of interest

➔ First, we need to recreate the genomes from our samples using sequencing data!

➔ We need to validate our genomes!

➔ We need to analyze the profiles of our genomes and compare them to other RSVs!

But where do we find other sequences? ➔ GISAID, NCBI, BV-BRC, …

We need to uniformize them!

Alignment

Alignment uniformize the sequences. This allows:

- Comparison position-wise of sequences
  (they all begin and end at the same position)
- Inference of phylogenies

Perform a multiple alignment (e.g. **MAFFT**)

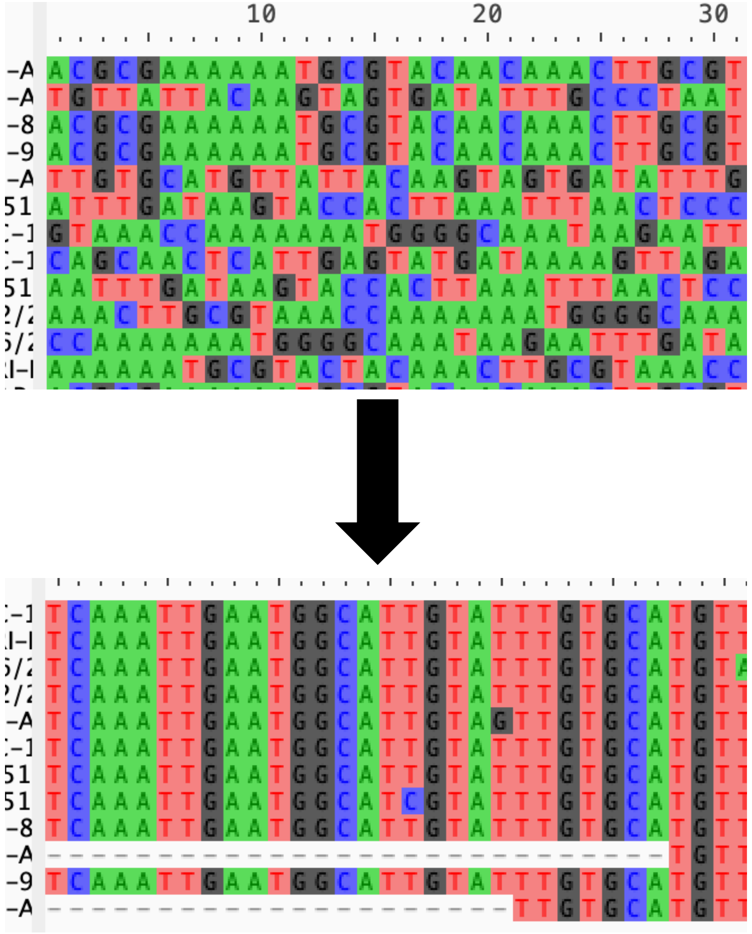
https://mafft.cbrc.jp/

- numbering depends of the indels
- No trimming on sequences

Align samples against reference (e.g. **Nextalign**)

https://clades.nextstrain.org

- Uniformity of positions (➔ reference ones)
- Insertions are suppressed!



Implemented in nextclade!

How do we analyse RSV genomic data?

Goals

- Track the diversity and evolution of RSV samples
- Detect mutations of interest

➔ First, we need to recreate the genomes from our samples using sequencing data!

➔ We need to validate our genomes!

➔ We need to analyze the profiles of our genomes and compare them to other RSVs!

But where do we find other sequences? ➔ GISAID, NCBI, BV-BRC, …

We need to uniformize them! ➔ alignment

We need to retrace the evolutionary history of RSVs!

Inference of phylogeny

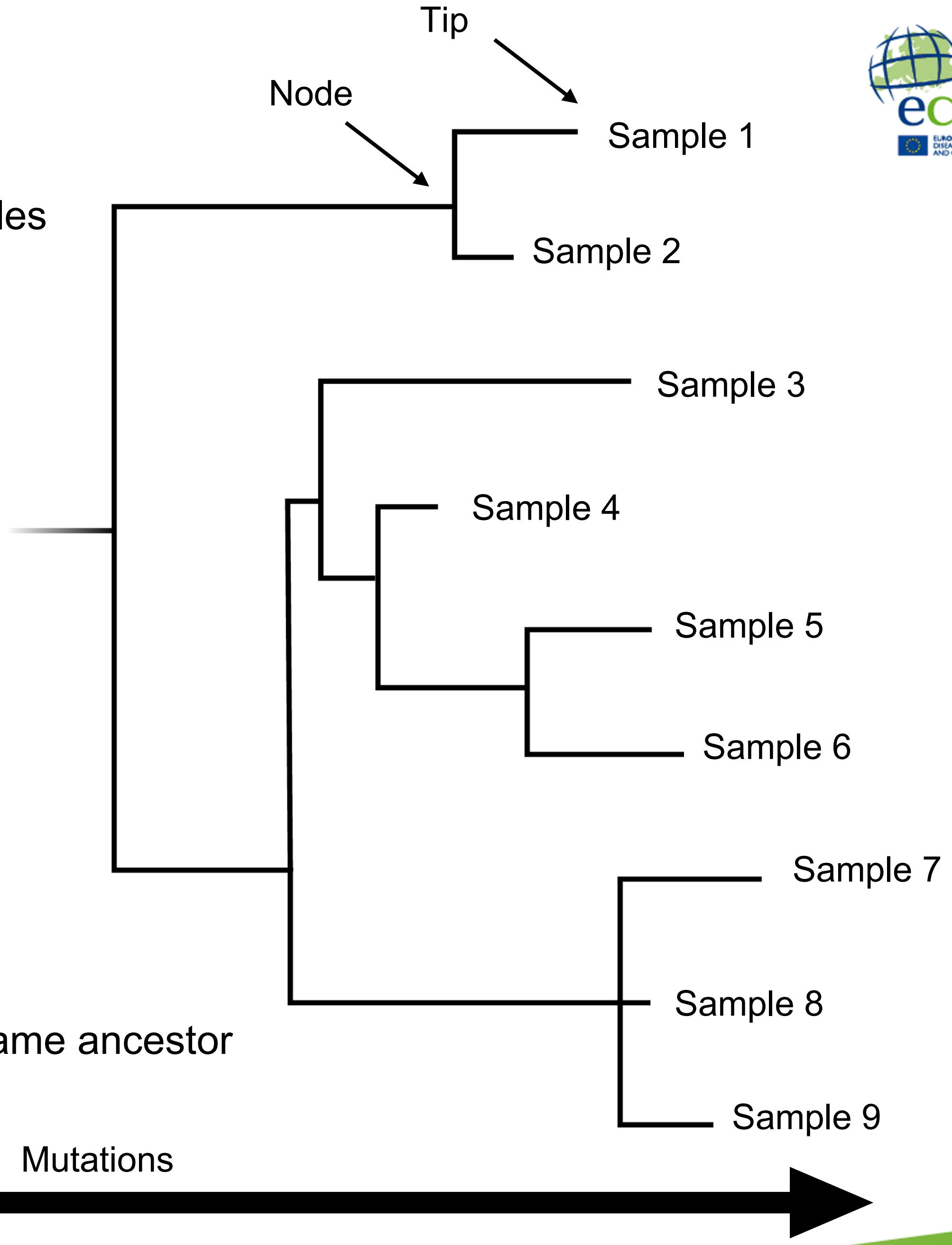
Phylogeny draws the evolutionary history (relationship) between samples (*and thus of an epidemic episode)*

Tip / external node: samples (what is collected)

Internal node : common ancestor / transmission event

Brench length: number of mutations (but can be time for example)

Monophyletic group: Variant, clade, lineage, Anything that share the same ancestor



Inference of phylogeny

Numerous tools exists to infer a phylogeny!

- IQ-TREE
- RAxML
- PhyML

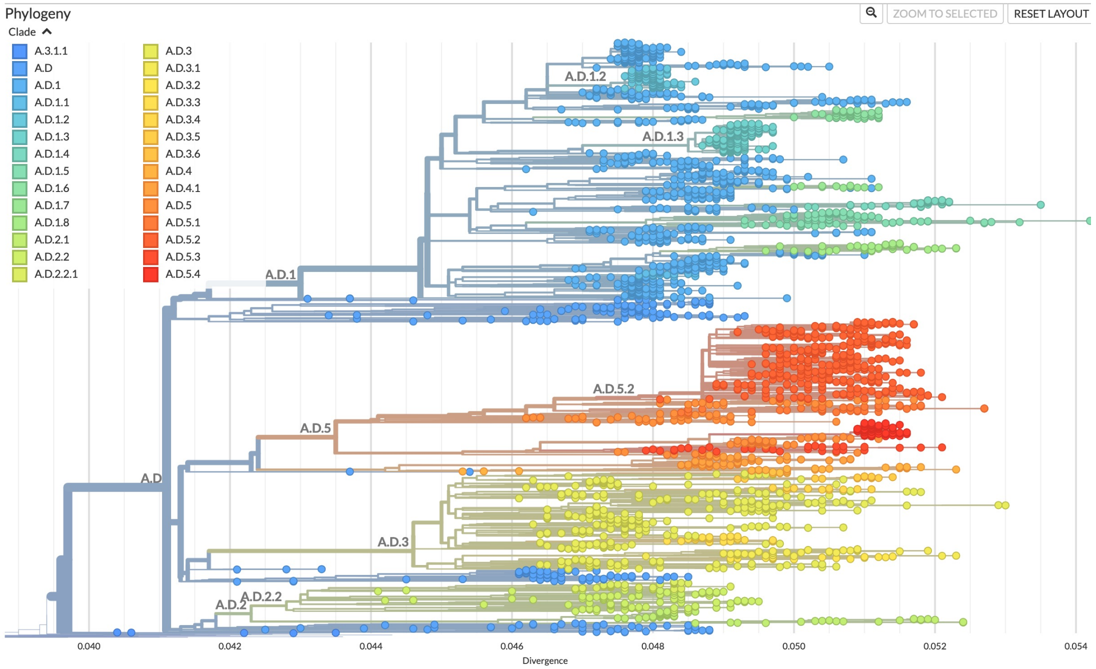
Some platforms allow an easy inference:

- Nextstrain-cli
- NGPhylogeny
- RAxML-web

You can visualize your own phylogenies / prepared ones with:

- iTOL
- Nextclade / nextstrain
- FigTree

Nextstrain



ITOL INTERACTIVE TREE OF LIFE

How do we analyse RSV genomic data?

Goals

- Track the diversity and evolution of RSV samples
- Detect mutations of interest

➔ First, we need to recreate the genomes from our samples using sequencing data!

➔ We need to validate our genomes!

➔ We need to analyze the profiles of our genomes and compare them to other RSVs!

But where do we find other sequences? ➔ GISAID, NCBI, BV-BRC, …

We need to uniformize them! ➔ alignment

We need to retrace the evolutionary history of RSVs! ➔ phylogeny

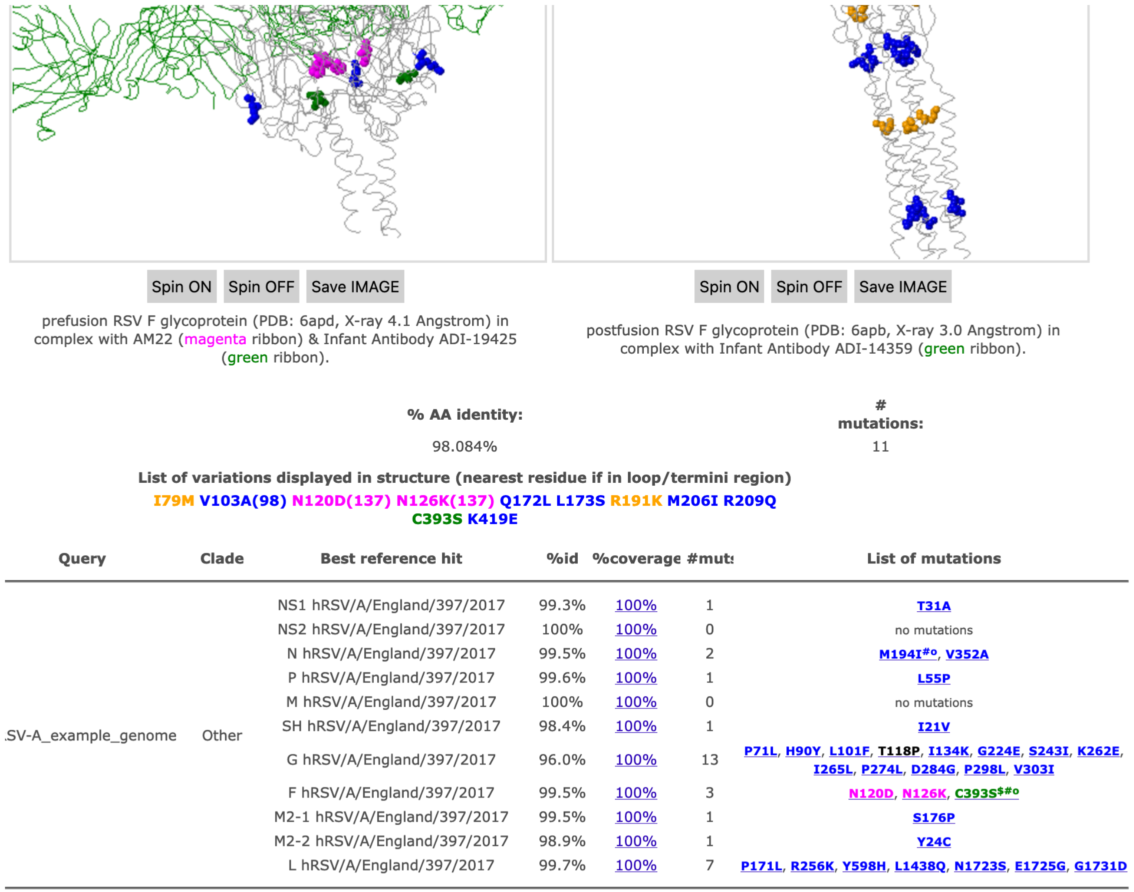
We need to check for specific mutations!

RSVSurver

A webserver that takes as input sequences and:

- Detect the RSV subtype
- List mutations against a reference, for each gene!
- List mutations that are important (especially on gene F)

➔ A good way to detect mutations of interest!



prefusion RSV F glycoprotein (PDB: 6apd, X-ray 4.1 Angstrom) in complex with AM22 (magenta ribbon) & Infant Antibody ADI-19425 (green ribbon).

postfusion RSV F glycoprotein (PDB: 6apb, X-ray 3.0 Angstrom) in complex with Infant Antibody ADI-14359 (green ribbon).

% AA identity:	# mutations:
98.084%	11

List of variations displayed in structure (nearest residue if in loop/termini region)

I79M V103A(98) N120D(137) N126K(137) Q172L L173S R191K M206I R209Q C393S K419E

Query	Clade	Best reference hit	%id	%coverage	#muts	List of mutations
SV-A_example_genome	Other	NS1 hRSV/A/England/397/2017	99.3%	100%	1	T31A
		NS2 hRSV/A/England/397/2017	100%	100%	0	no mutations
		N hRSV/A/England/397/2017	99.5%	100%	2	M194I$^{\#o}$, V352A
		P hRSV/A/England/397/2017	99.6%	100%	1	L55P
		M hRSV/A/England/397/2017	100%	100%	0	no mutations
		SH hRSV/A/England/397/2017	98.4%	100%	1	I21V
		G hRSV/A/England/397/2017	96.0%	100%	13	P71L, H90Y, L101F, T118P, I134K, G224E, S243I, K262E, I265L, P274L, D284G, P298L, V303I
		F hRSV/A/England/397/2017	99.5%	100%	3	N120D, N126K, C393S$^{\$\#o}$
		M2-1 hRSV/A/England/397/2017	99.5%	100%	1	S176P
		M2-2 hRSV/A/England/397/2017	98.9%	100%	1	Y24C
		L hRSV/A/England/397/2017	99.7%	100%	7	P171L, R256K, Y598H, L1438Q, N1723S, E1725G, G1731D

https://rsvsurver.bii.a-star.edu.sg/

How do we analyse RSV genomic data?

Goals

- Track the diversity and evolution of RSV samples
- Detect mutations of interest

➔ First, we need to recreate the genomes from our samples using sequencing data!

➔ We need to validate our genomes!

➔ We need to analyze the profiles of our genomes and compare them to other RSVs!

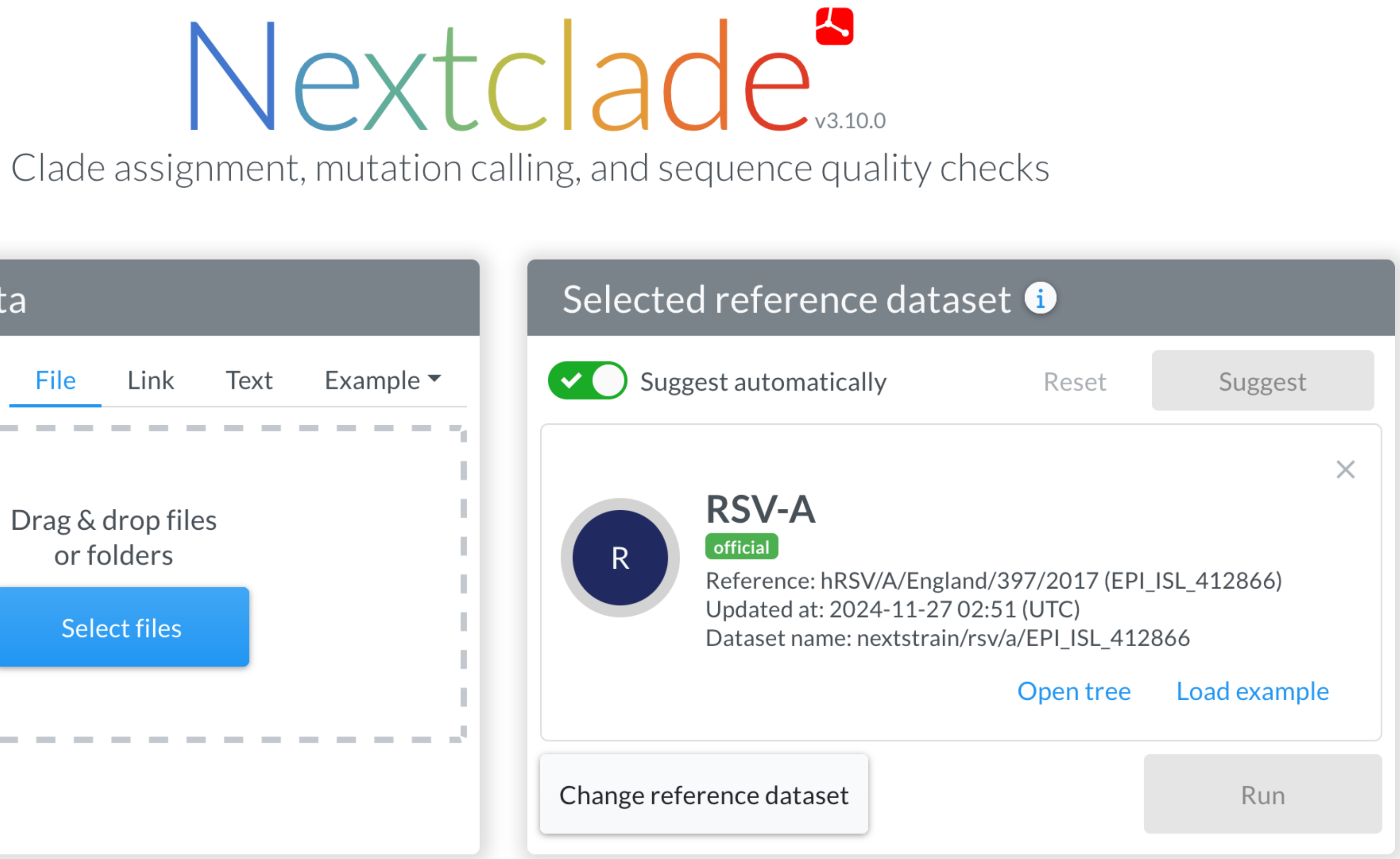
But where do we find other genomes? ➔ GISAID, NCBI, BV-BRC, …

We need to uniformize them! ➔ alignment

We need to retrace the evolutionary history of RSVs! ➔ phylogeny

We need to check for specific mutations! ➔ RSVSurver

Tool demonstration: Nextclade



Nextclade v3.10.0

Clade assignment, mutation calling, and sequence quality checks

Provide sequence data

Drag & drop files or folders

Selected reference dataset

Suggest automatically

RSV-A

official

Reference: hRSV/A/England/397/2017 (EPI_ISL_412866)

Updated at: 2024-11-27 02:51 (UTC)

Dataset name: nextstrain/rsv/a/EPI_ISL_412866

https://clades.nextstrain.org/

Practical Exercise: Analyzing RSV Genomic Data

Practical Exercise: Analyzing RSV Genomic Data

10/12/2024

Now is your turn!

You'll conduct a RSV genome analysis on:

- Two contextual datasets (RSVA and RSVB)
- Two query datasets (RSVA and RSVB)

You will analyse these datasets using solutions described previously, such as:

- Nextclade
- NGPhylogeny
- iTOL
- RSVSurver

Now is your turn!

Open the zip archive:

https://dl.pasteur.fr/fop/KxpfUnHt/practical_training.zip

Follow the guidelines and answer the questions!

Nextclade

RSVsurver

iTOL INTERACTIVE TREE OF LIFE

Now is your turn!

PART 1: QUESTIONS AND ANSWERS

slido

Please download and install the
Slido app on all computers you use

Were you able to download the zip archive?

ⓘ Start presenting to display the poll results on this slide.

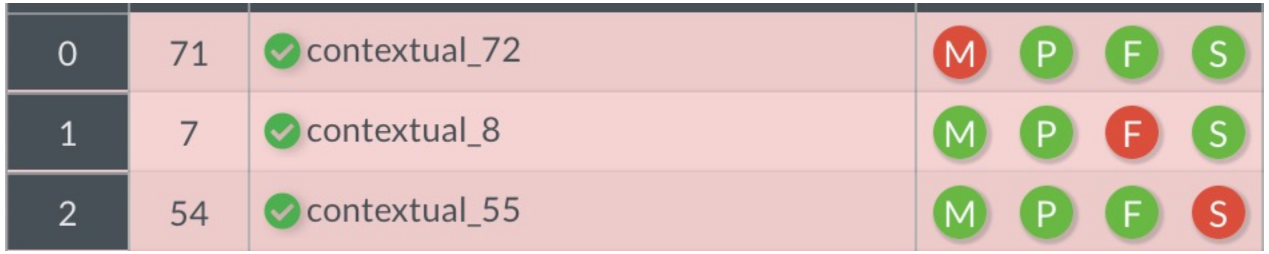
slido

Please download and install the
Slido app on all computers you use

In your opinion, which RSVA sequences should be removed from the contextual dataset?

ⓘ Start presenting to display the poll results on this slide.

Now is your turn!



0	71	✓ contextual_72	M	P	F	S
1	7	✓ contextual_8	M	P	F	S
2	54	✓ contextual_55	M	P	F	S

➔ Lots of mixtures!

➔ Frameshifts!

➔ Lots of private mutations / Stop codons!

3	✕ contextual_4	When calculating seed matc

➔ Not a RSVA!

The contextual_1 sample lacks coverage, especially on the F gene. It is acceptable to remove it also!

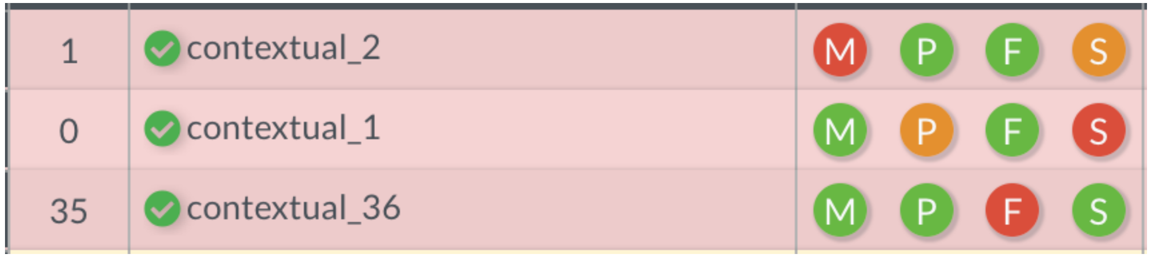
slido

Please download and install the
Slido app on all computers you use

In your opinion, which RSVB sequences should be removed from the contextual dataset?

ⓘ Start presenting to display the poll results on this slide.

Now is your turn!



1	✓ contextual_2	M	P	F	S
0	✓ contextual_1	M	P	F	S
35	✓ contextual_36	M	P	F	S

➔ Lots of mixtures / Stop codons!

➔ Lots Private mutations / Stop codons!

➔ Frameshifts!

✕ contextual_41	When calculating seed matc

➔ Not a RSVB!

slido

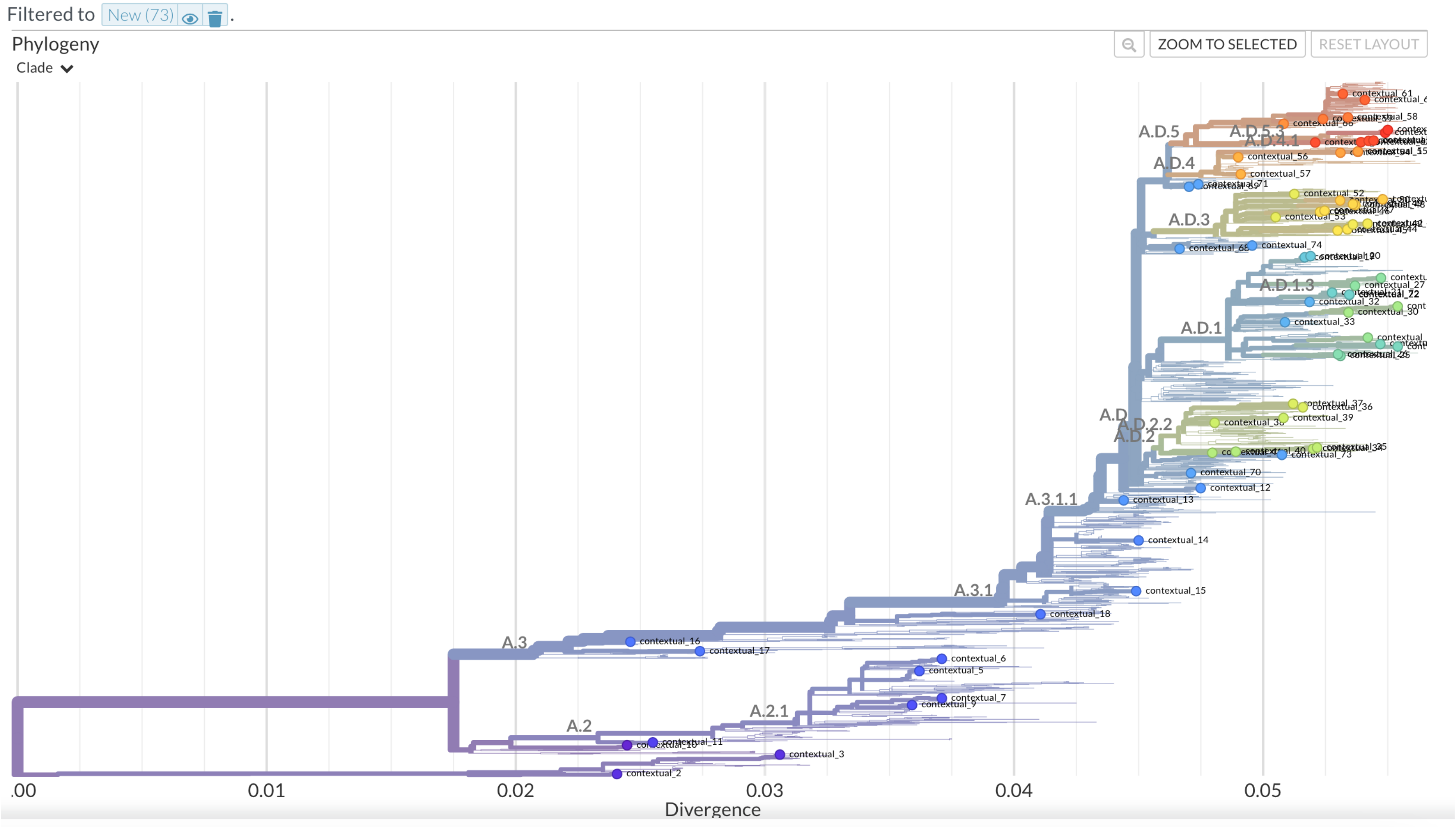
Please download and install the
Slido app on all computers you use

Do you think that both these datasets represent well the diversity of RSVA and RSVB ?

ⓘ Start presenting to display the poll results on this slide.

Now is your turn!

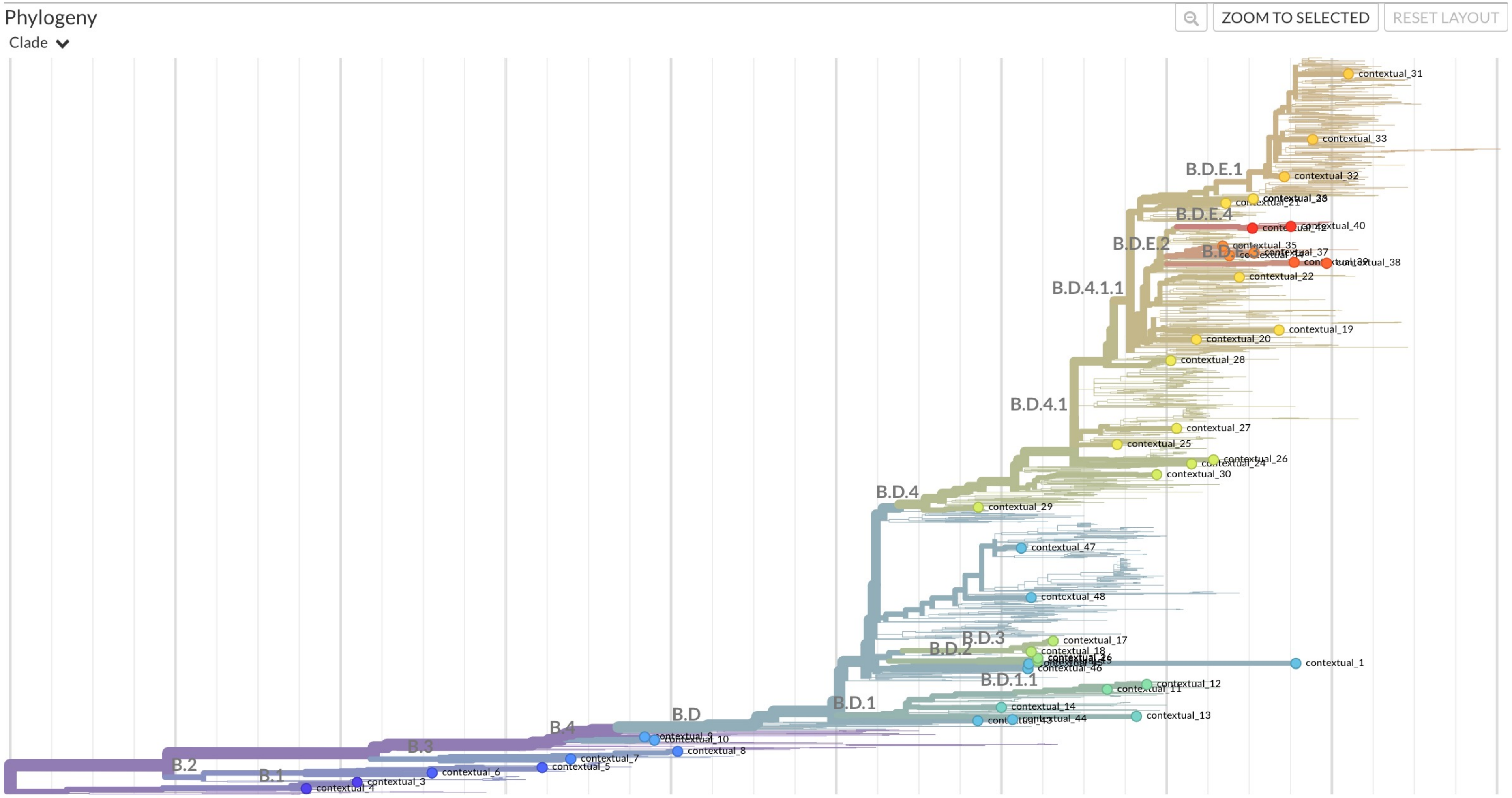
RSVA



When placing contextual samples onto the existing phylogeny ➔ good representatitivity!!!

Now is your turn!

RSVB



When placing contextual samples onto the existing phylogeny ➔ good representatitivity!!!

slido

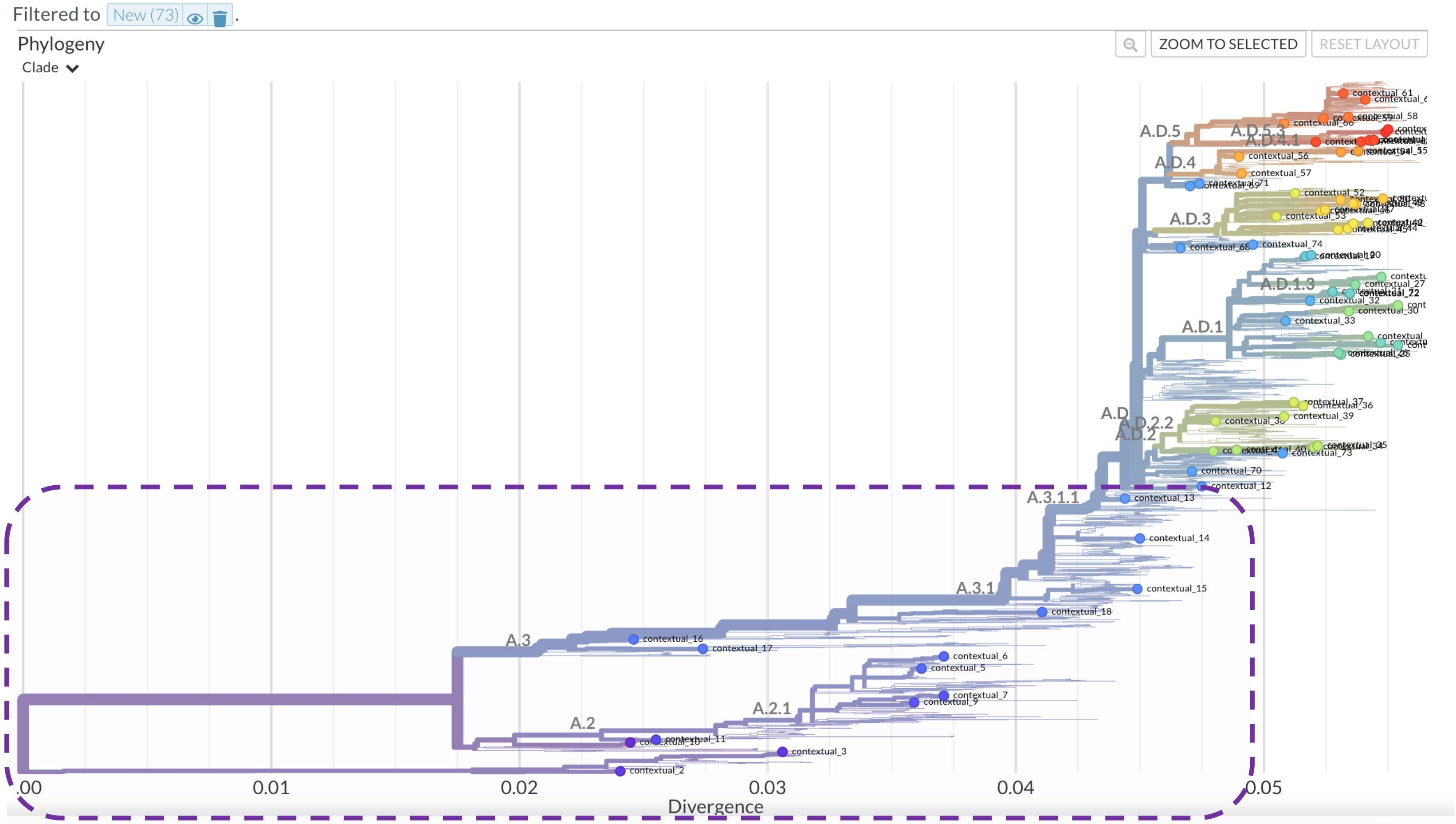
Please download and install the
Slido app on all computers you use

Do you have any pre-duplication individuals in the datasets? If so, to which clades do they belong?

ⓘ Start presenting to display the poll results on this slide.

Now is your turn!

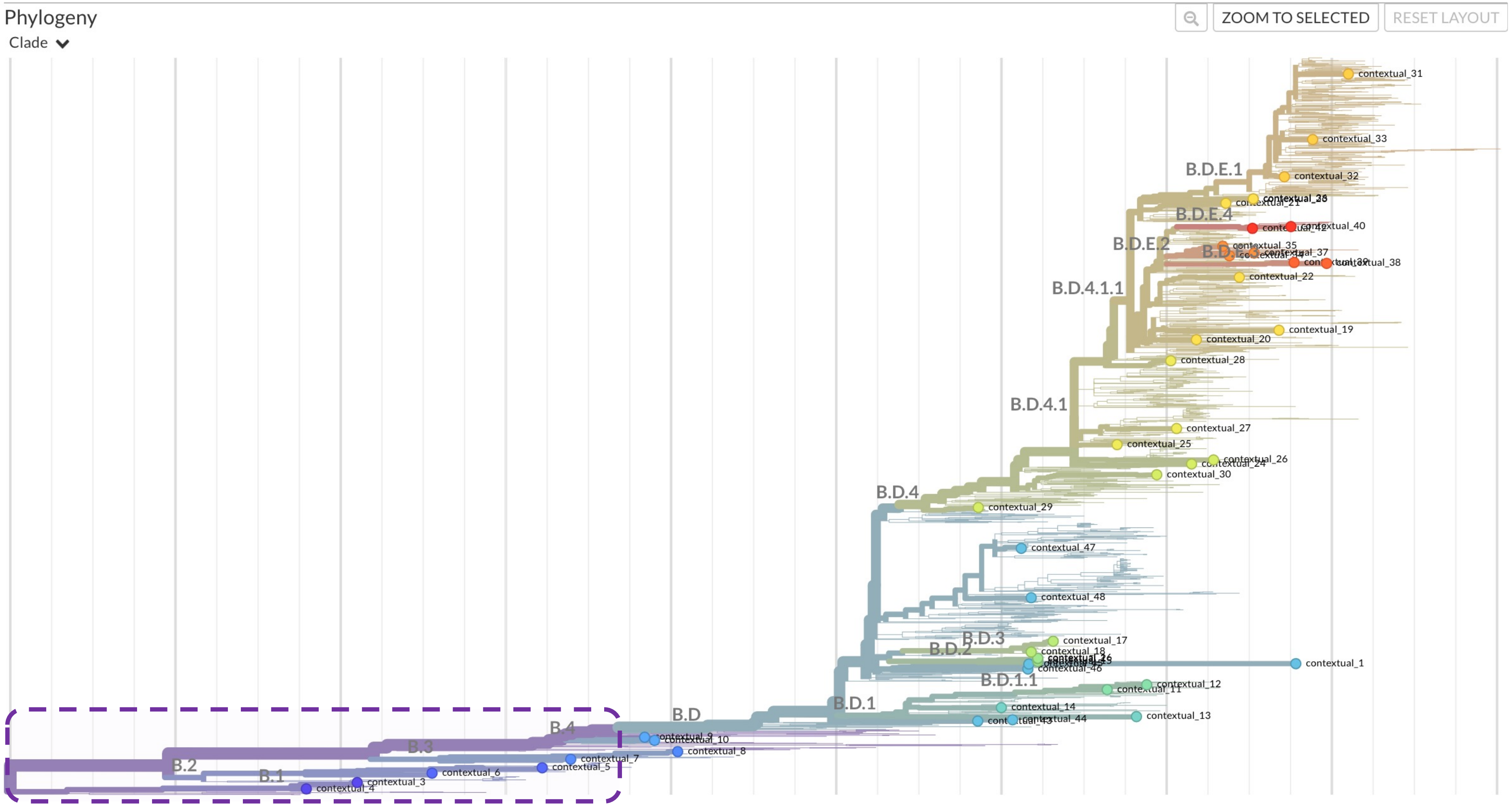
RSVA



All samples that belong to clades A.1, A.2 and A.3

Now is your turn!

RSVB



All samples that belong to clades B.1, B.2, B.3 and B.4

Now is your turn!

PART 2: QUESTIONS AND ANSWERS

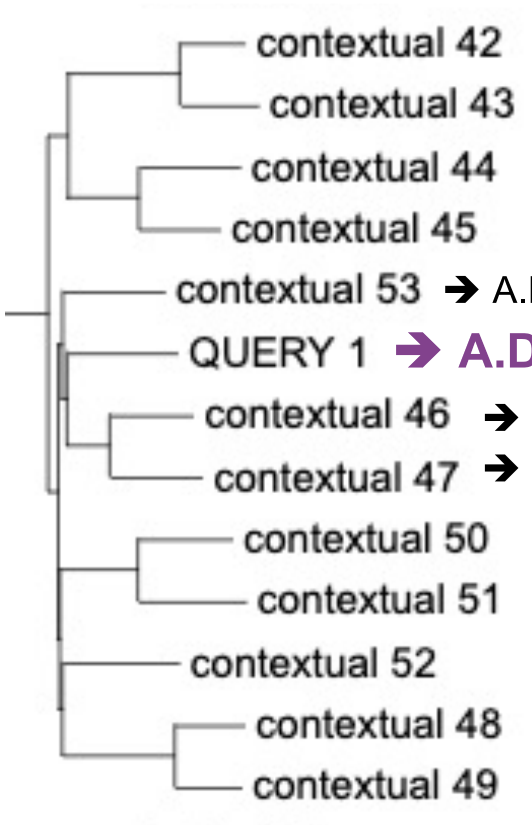
slido

Please download and install the
Slido app on all computers you use

To which clades belong the RSVA queries?

ⓘ Start presenting to display the poll results on this slide.

Now is your turn!



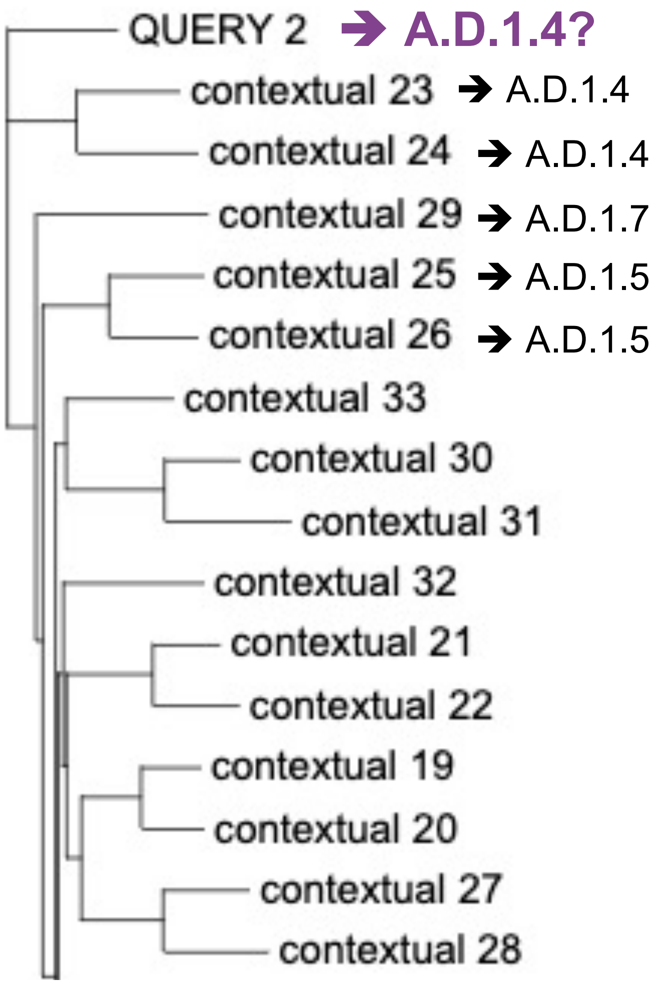
contextual 53 ➔ A.D.3

QUERY 1 ➔ **A.D.3? A.D.3.3?**

contextual 46 ➔ A.D.3.3

contextual 47 ➔ A.D.3.3

Now is your turn!



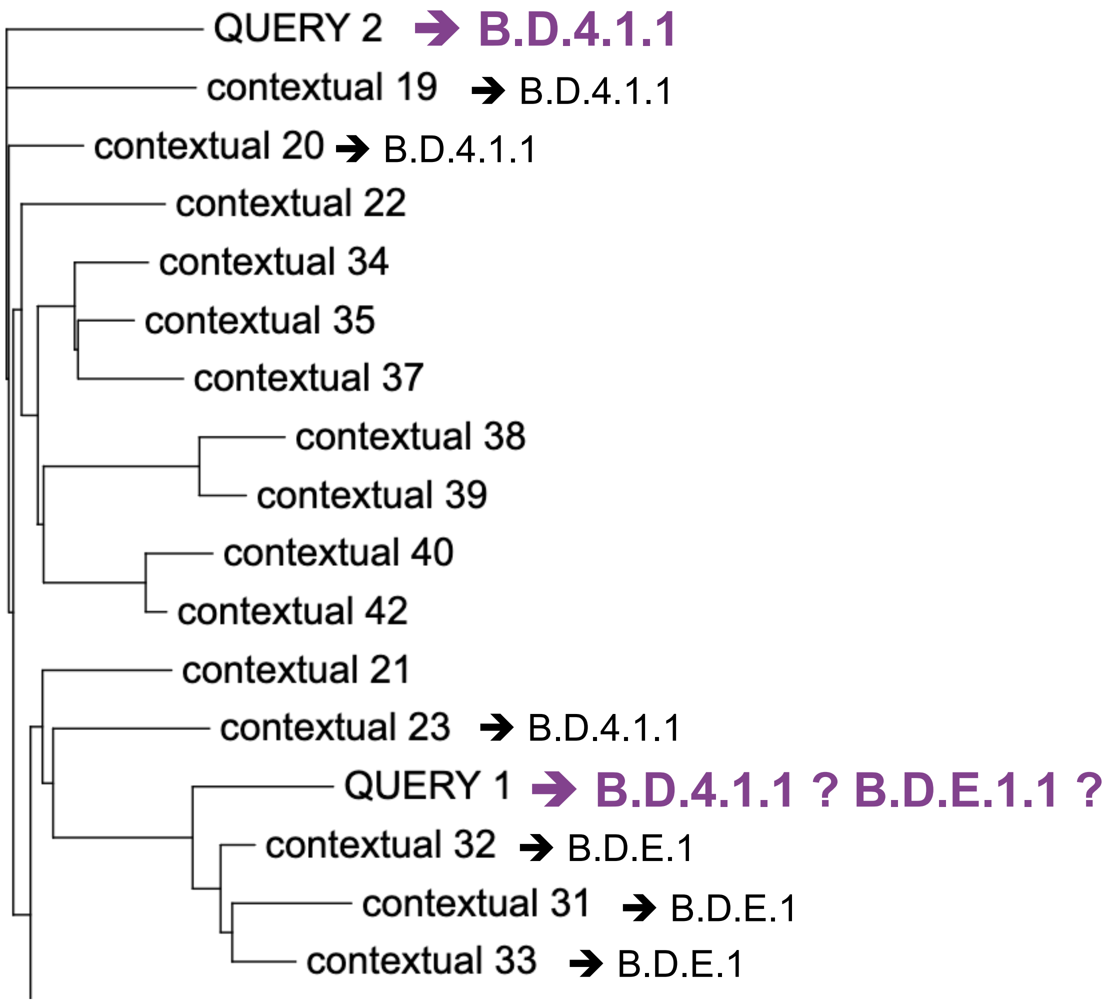
slido

Please download and install the
Slido app on all computers you use

To which clades belong the RSVB queries?

ⓘ Start presenting to display the poll results on this slide.

Now is your turn!



QUERY 2 ➔ **B.D.4.1.1**
contextual 19 ➔ B.D.4.1.1
contextual 20 ➔ B.D.4.1.1
contextual 22
contextual 34
contextual 35
contextual 37
contextual 38
contextual 39
contextual 40
contextual 42
contextual 21
contextual 23 ➔ B.D.4.1.1
QUERY 1 ➔ **B.D.4.1.1 ? B.D.E.1.1 ?**
contextual 32 ➔ B.D.E.1
contextual 31 ➔ B.D.E.1
contextual 33 ➔ B.D.E.1

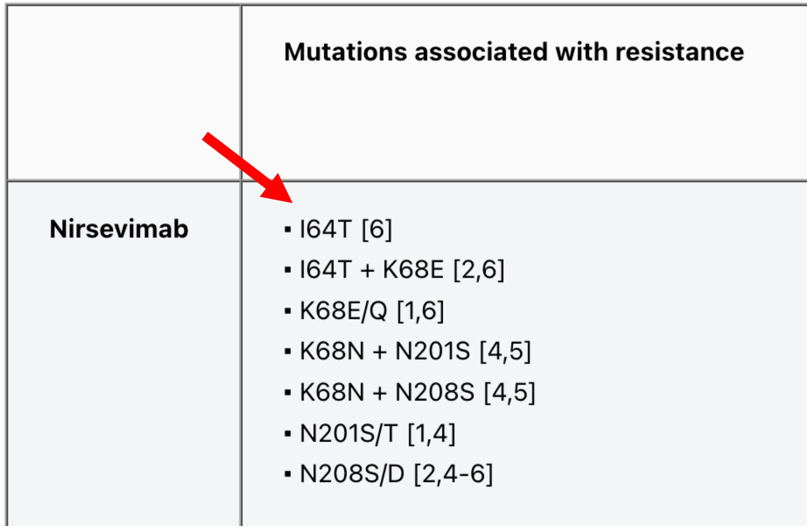
slido

Please download and install the
Slido app on all computers you use

Did you see any Drug-Resistance Mutations ? If so, which ones in which queries?

ⓘ Start presenting to display the poll results on this slide.

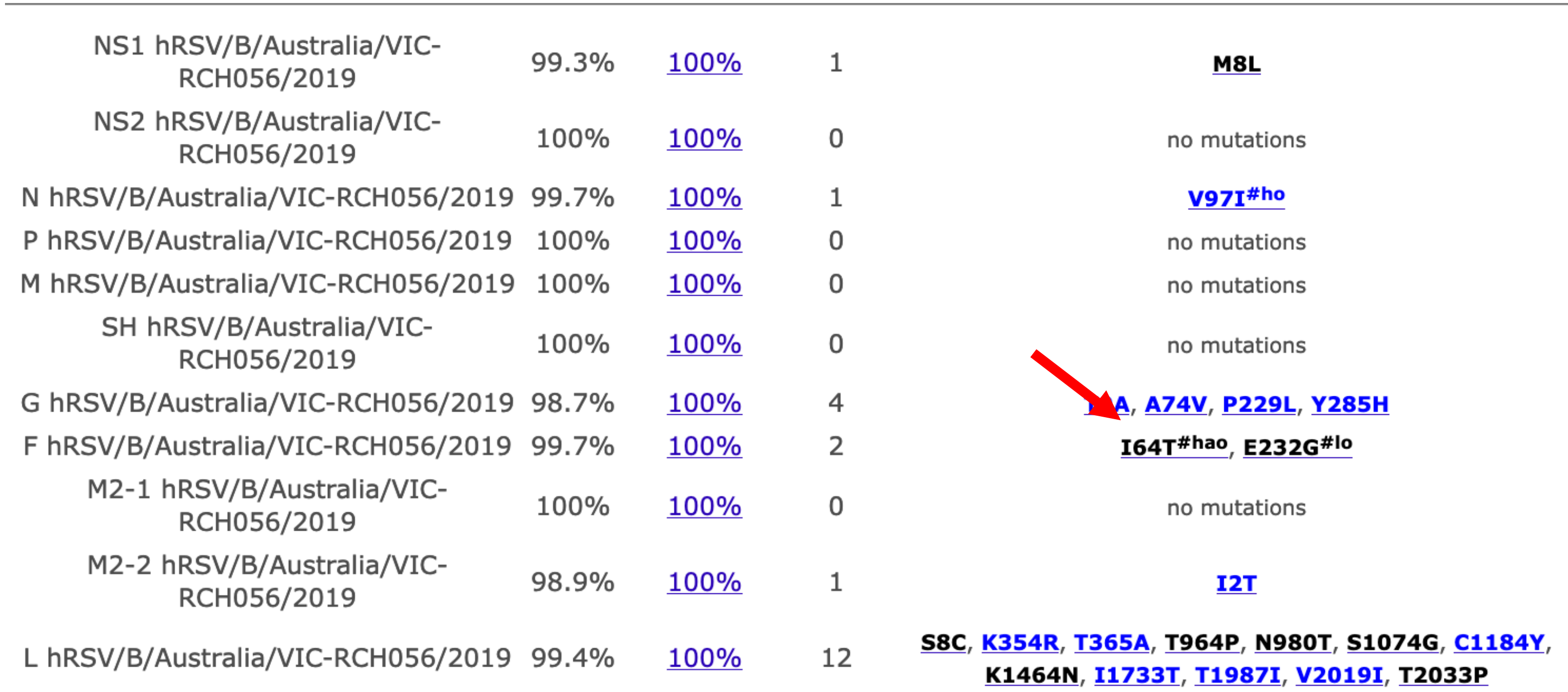
Now is your turn!



	Mutations associated with resistance
Nirsevimab	▪ I64T [6] ▪ I64T + K68E [2,6] ▪ K68E/Q [1,6] ▪ K68N + N201S [4,5] ▪ K68N + N208S [4,5] ▪ N201S/T [1,4] ▪ N208S/D [2,4-6]

Virusfrenchresistance table

RSVSurver results



NS1 hRSV/B/Australia/VIC-RCH056/2019	99.3%	100%	1	**M8L**
NS2 hRSV/B/Australia/VIC-RCH056/2019	100%	100%	0	no mutations
N hRSV/B/Australia/VIC-RCH056/2019	99.7%	100%	1	**V97I$^{\#ho}$**
P hRSV/B/Australia/VIC-RCH056/2019	100%	100%	0	no mutations
M hRSV/B/Australia/VIC-RCH056/2019	100%	100%	0	no mutations
SH hRSV/B/Australia/VIC-RCH056/2019	100%	100%	0	no mutations
G hRSV/B/Australia/VIC-RCH056/2019	98.7%	100%	4	**[illegible]A**, **A74V**, **P229L**, **Y285H**
F hRSV/B/Australia/VIC-RCH056/2019	99.7%	100%	2	**I64T$^{\#hao}$**, **E232G$^{\#lo}$**
M2-1 hRSV/B/Australia/VIC-RCH056/2019	100%	100%	0	no mutations
M2-2 hRSV/B/Australia/VIC-RCH056/2019	98.9%	100%	1	**I2T**
L hRSV/B/Australia/VIC-RCH056/2019	99.4%	100%	12	**S8C**, **K354R**, **T365A**, **T964P**, **N980T**, **S1074G**, **C1184Y**, **K1464N**, **I1733T**, **T1987I**, **V2019I**, **T2033P**

RSVB: QUERY_2 has a Drug Resistance Mutation!

National Center of Reference for Respiratory Viruses, Institut Pasteur, Paris, France

NRC: Marie-Anne Rameix-Welti

Bioinfo@NRC: Frédéric Lemoine, Kévin Da-Silva, Samar Berreira Ibraim

Acknowledgements

The creation of this training material was commissioned by ECDC to Institut Pasteur with the direct involvement of Jérôme Bourret