

# RSV Consensus Genome Analysis

10<sup>th</sup> of December, 2024

Practical Exercise of

## ***RSV Insights: Genomic Evolution, Treatments, and Public Health Strategies***

This exercise will ensure that participants gain hands-on experience of real-world RSV genome analysis workflows using widely available online tools. For both **RSVA** and **RSVB**, participants will focus on **quality control, clade assignment, diversity/evolutionary history analysis** and **resistance mutation identification**. Participants will use online versions of **Nextclade**, **RSVSurver** and **NGPhylogeny.fr**.

Participants will be given two sets of datasets. The first ones are comprising of “contextual” sequences that gives a good representation of RSVA and RSVB diversity:

- **RSVA\_contextual.fasta**
- **RSVB\_contextual.fasta**

The second ones are comprising of “query” sequences that must be processed with sequences from the first dataset:

- **RSVA\_queries.fasta**
- **RSVB\_queries.fasta**

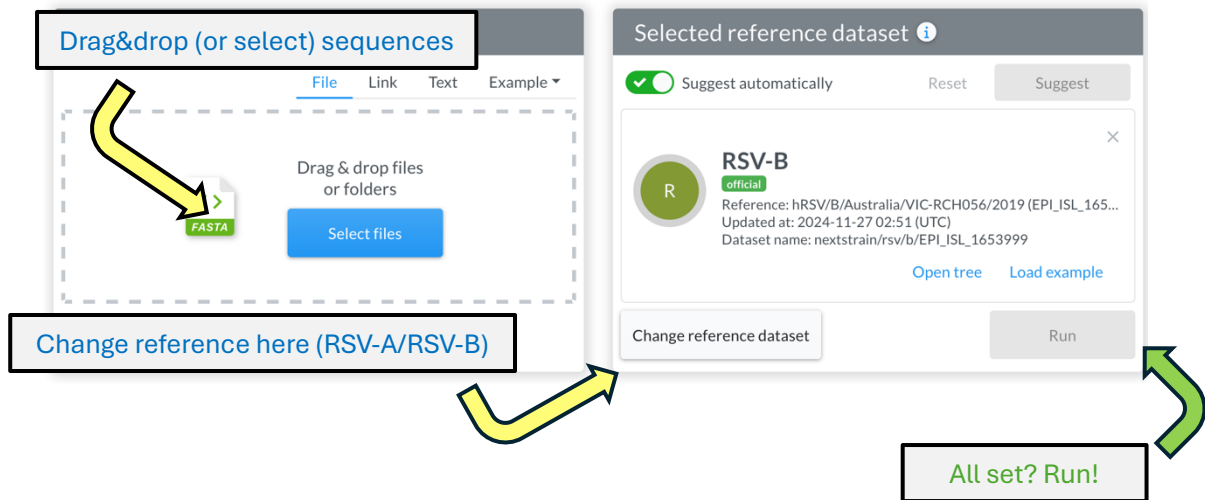
### **Part 1: Contextual datasets analysis**

The first datasets correspond to fasta files **RSVA/B\_contextual.fasta** containing RSVA/B sequences. Your task here is to select sequences that are suitable for further analysis. You should use the online tool to **Nextclade** (<https://www.nextclade.org>) to do this:

- 1) Overall quality such as coverage of the whole genome and of the G / F genes, frameshifts, stop codons, accumulation of private mutations, high levels of ambiguous bases, ...
- 2) Determine clades and verify that your dataset is a good representation of RSV diversity.
- 3) Prepare your dataset for the next step.

#### **1. Upload the Dataset:**

- Go to **Nextclade** (<https://www.nextclade.org>).
- Upload the provided **RSVA/B\_contextual.fasta**.
- Since Nextclade proposes different references for each RSV, you must launch the analysis multiple times (one time for each reference).



## 2. Sequence Quality Control:

- Review the sequence quality scores. Identify any sequences with significant issues (e.g., high number of missing bases, ambiguous bases). Sequences with an overall score “bad” (red ones) should be rejected.
- Check for non-covered regions, focusing on the **G** and **F** genes. Document sequences with significant loss of information on these regions. A sequence without the F gene should not be kept.

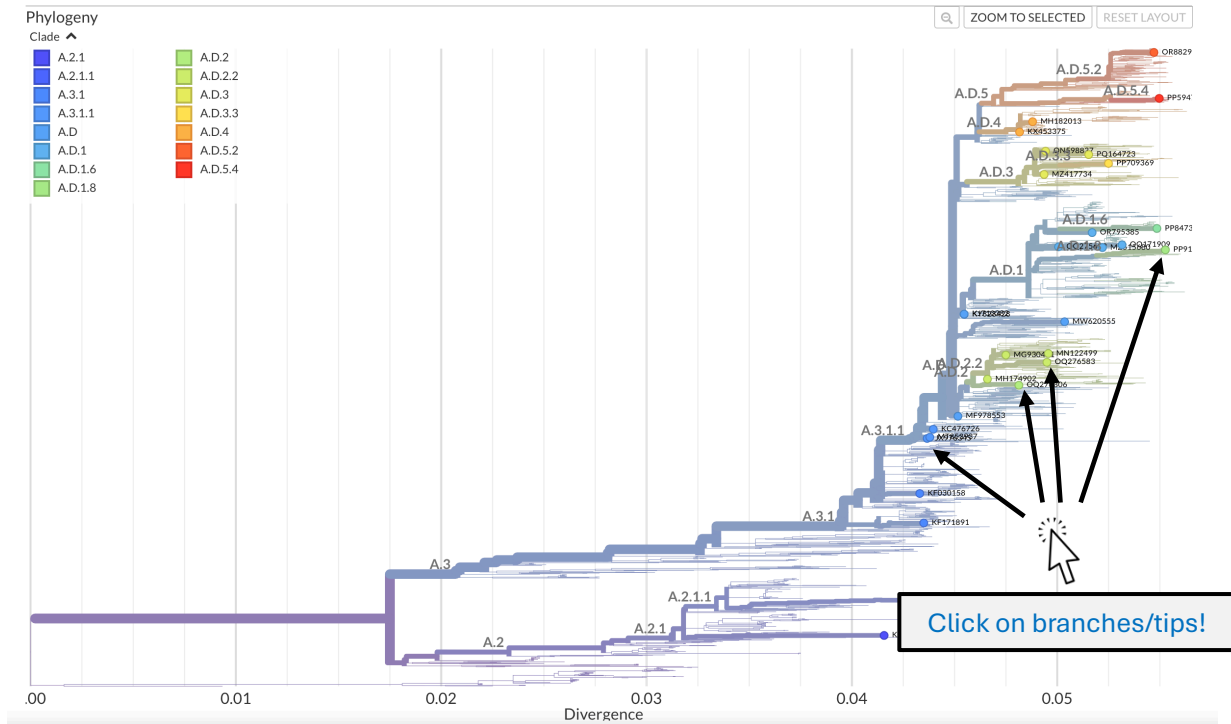
#	i	Sequence name	QC	Clade	G clades	Mut.	non-ACGTN	Ns	Cov.	Gaps	Ins.	FS	SC	Genetic feature	Relative to
6	6	▲ PQ164723	M P F S	A.D.3	GA2.3.5	22	0	0	3.6%	0	0	0	0		
7	7	▲ ON598837	M P F S	A.D.3	GA2.3.5	15	0	0	2.5%	0	0	0	0	Gene 'F' is missing	
8	9	▲ OQ275639	M P F S	A.D.1	GA2.3.5	16	0	0	11.3%	0	0	0	0		
9	10	▲ OQ276583	M P F S	A.D.2.2	GA2.3.5	10	0	0	11.3%	0	0	0	0		
10	11	▲ OQ276806	M P F S	A.D.2	GA2.3.5	14	0	0	11.3%	0	0	0	0		

Annotations below the table:

- Quality (points to QC column)
- Clades (points to Clade and G clades columns)
- Overall description (points to Mut., non-ACGTN, Ns, Cov., Gaps, Ins., FS, SC columns)
- Gene / genome view (points to Genetic feature and Relative to columns)

## 3. Determine Clades:

- Analyse the clades assigned to the sequences and check that all major clades are well represented. You can do this by looking at the assigned clades in the Results tab, but you can also visualise the phylogenetic placement in the Tree tab to check for diversity. You can click on tips / branches to get more information.
- If possible, document missing clades.



#### 4. Save Curated Contextual Datasets:

For each dataset, save both **nextclade.tsv** and **nextclade.aligned.fasta** files.

- **nextclade.tsv** contains all results that you can see in the “results” tab.
- **nextclade.aligned.fasta** contains the sequences that have been aligned to the reference (nextalign).
- Save these files for both RSVA and RSVB analyses.



#### Download output files

Files Column config

Configure columns

nextclade.tsv  
Summarized results  
Contains summarized metrics etc., in tabular using spreadsheets

Download!

Configure columns

nextclade.auspice.json  
Phylogenetic tree with sequences placed onto it, in Auspice JSON v2 format.  
Can be viewed locally with Nextstrain Auspice or in [auspice.us](https://auspice.us).

nextclade.nwk  
Phylogenetic tree with sequences placed onto it, in Newick format.  
Can be viewed in most tree viewers, including: [icytree.org](https://icytree.org) or [auspice.us](https://auspice.us).

nextclade.aligned.fasta  
Aligned sequences in FASTA format  
Contains aligned sequences

Download!

## Questions

- Did both context files have any "bad" sequences? If so, which ones? What were the reasons for classifying them as such?
- Do you think the data sets represent the diversity of RSVs well?
- Are there any pre-duplication individuals in the data set? How do you recognise them?

## Part 2: Clade placement and DRMs for New Sequences

In this exercise, you will determine the clades of few sequences contained in the files **RSVA/B\_queries.fasta**. But this time, instead of using Nextclade, you will use [NGPhylogeny.fr](https://ngphylogeny.fr/) (<https://ngphylogeny.fr/>) webserver.

In addition, you will determine the presence of drug-resistance mutations in these individuals by using the tool [RSVSurver](https://rsvsurver.bii.a-star.edu.sg/) (<https://rsvsurver.bii.a-star.edu.sg/>) as well as the [Virus French Resistance database](https://virusfrenchresistance.org/virus-french-resistance-rsv/) <https://virusfrenchresistance.org/virus-french-resistance-rsv/>.

### 1. Construct Phylogenetic Trees:

- Upload the merged files (contextual + queries) to [NGPhylogeny.fr](https://ngphylogeny.fr/) (<https://ngphylogeny.fr/>) to align these sequences and build phylogenetic trees.
  - Perform a "One-click" analysis.
  - Use the default settings for tree construction. In this step, sequences are aligned and then used to infer a phylogeny.
- Analyse the phylogenetic trees to determine the clades for the new sequences. To do this, check where your query sequences sit in relation to contextual sequences (for which you already know the clade).

Robust phylogenetic analysis for everyone.

> Free, simple to use web service dedicated to reconstructing and analysing phylogenetic relationships between molecular sequences.

Workflow

Select this one!

> **One Click**  
Fully automatic workflow  
Default tools + default parameters.

> **Advanced**  
Semi automatic workflow  
Default tools + custom parameters

> **A la Carte**  
Custom workflow  
Custom tools + Custom parameters.

Institut Pasteur CSE-USK 3756 LIRMM ATGC ifb

- PhyML/OneClick
- PhyML+SMS/OneClick
- FastTree/OneClick

Give as input aligned sequences

Submit the analysis

### Choose input data

Input Data (Fasta format with more than 3 sequences)

**Input file**

Choose file To Upload
Choose File

**Pasted text**

**Blast run\***

--

**Galaxyfile\***

--

Submit
Example
🗑️



Tool	Step	File Name	Status	
Newick Display	15.	All tree images	✓	<a href="#">+</a> <a href="#">⚙️</a> <a href="#">➡️</a> <a href="#">⬇️</a> <a href="#">👁️.tar</a>
	14.	Tree image	✓	<a href="#">+</a> <a href="#">⚙️</a> <a href="#">➡️</a> <a href="#">⬇️</a> <a href="#">📄.svg</a>
FastME	13.	Mapping between short sequence id and names (useful to interpret some bootstrap log files if any)	✓	<a href="#">+</a> <a href="#">⚙️</a> <a href="#">➡️</a> <a href="#">⬇️</a> <a href="#">📄.txt</a>
	12.	Output Tree	✓	<a href="#">+</a> <a href="#">⚙️</a> <a href="#">➡️</a> <a href="#">⬇️</a> <a href="#">👁️.nhx</a> <a href="#">📄 Viewer</a> <a href="#">🗑️ iTol</a>
	11.	FastME Distance matrix	✓	<a href="#">+</a> <a href="#">⚙️</a> <a href="#">➡️</a> <a href="#">⬇️</a> <a href="#">👁️.txt</a>
	10.	FastME Information	✓	<a href="#">+</a> <a href="#">⚙️</a> <a href="#">➡️</a> <a href="#">⬇️</a> <a href="#">👁️.txt</a>

Vizualise with iTol



## 2. Check for Resistance Mutations:

- Focus on key resistance-related mutations in the F gene. Look at the clade assignment and specific mutation annotations (if available in the tree or sequence metadata). Use the [RSVsurver](https://rsvsurver.bii.a-star.edu.sg/) software (<https://rsvsurver.bii.a-star.edu.sg/>).
- You'll find a list of mutations of Resistance in the following webpage: <https://virusfrenchresistance.org/virus-french-resistance-rsv/>

# RSVsurver



Paste your protein or nucleotide FASTA sequence(s) into the text area below. (Sample FASTA sequence: [Example hCoV-19 genome](#))

```
>hRSV-A_example_genome
CCAAAAAATGGGGCAAATAAGAATTTGATAAGTACCACTTAAATTTAACTCCTTTGGTTAGAGATGGGCAGCAACTCATTGAGT
ATGATAAAAGTTAGATTGCAAAATCTGTTTGACAATGATGAAGTAGCATTGTTAAAAATAACATGCTATGCTGACAAATTAATACA
GTTAACTAATGCTTTGGCTAAGGCAGTTATACATACAATCAAATTTGAATGGCATTGTATTTGTGCATGTTATTACAAGTAGTGATA
TTGGCCCTAATAATAATATTGTAGTGAAATCCAATTTCAACAATGCCAGTATTACAAAATGGAGTTATATATGGGAAATGATG
GAAATTTACACTGCTCCCAACCTAATGGCCTAATAGATGACAATTTGAAATTAATTTCTCCAAAAAACTAAGTGATTCAACAA
TGACTAATTATATGAATCAATTATCTGAATTAATTTGACCTCAATCCATAAATCATAATAAATATCAACTAGCAAATCAATGT
CACTAACACCATTAGTTAATATAAACTTGACAGAAGATAAAAAATGGGGCAAATAAATCAATTCAGCCGACCAACCATGGACA
```

Give sequences as input

OR upload your protein or nucleotide sequences in a FASTA file

Parcourir... Aucun fichier sélectionné.

The server can **automatically** determine the type of input (either protein or nucleotide) and the closest reference sequence among current strains to compare.

To compare with more remotely related sequences/strains, it is possible to select a specific reference strain by choosing below.

Compare with:  
Automatic detection of closest reference

Submit the analysis

Submit Reset (estimated time needed: ~10 seconds per sequence in automatic mode)

**LS935 R419E**

Query	Clade	Best reference hit	%id	%coverage	#mut:	List of mutations
	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				M194I#o, V352A
	P	hRSV/A/England/397/2017				L55P
	M	hRSV/A/England/397/2017				no mutations
	SH	hRSV/A/England/397/2017				I21V
	G	hRSV/A/England/397/2017	96.0%	100%	13	P71L, P101F, T118P, I134K, G224E, S243I, K262E, 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

Check for mutations on F

Click on mutations!

## Questions

- To which clades these query RSVs belong to?
- Do you see any drug resistance mutations?