

Genome Assembly - INSaFLU



WHO Collaborating Centre
for Reference and
Research on Influenza
VIDRL



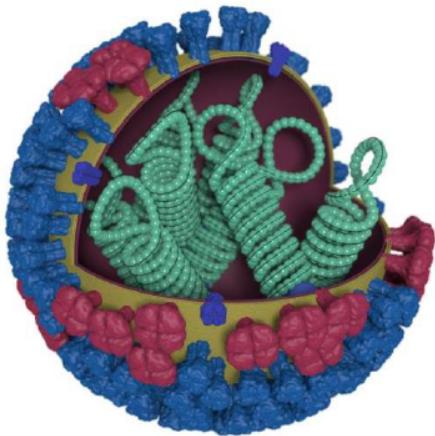
A joint venture between The University of Melbourne and The Royal Melbourne Hospital



Objectives

- To assemble the genome of influenza using INSaFLU

Influenza virus genome

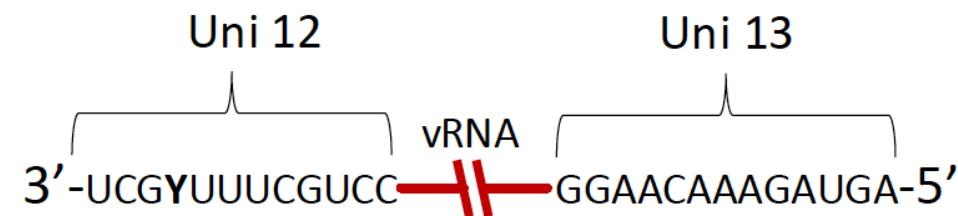


Negative-sense RNA: 8 Segments

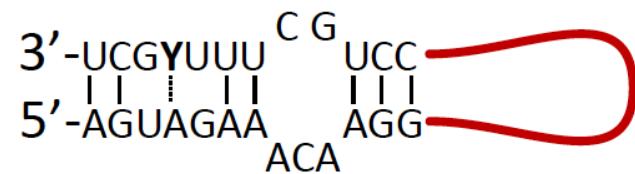
The diagram illustrates the size of each influenza A virus gene segment. The segments are:

- PB2: 2.3 kb
- PB1: 2.3 kb
- PA: 2.2 kb
- HA: 1.8 kb
- NP: 1.6 kb
- NA: 1.4 kb
- M: 1.0 kb
- NS: 0.9 kb

Conserved Terminal Ends



Panhandle Structure





From “reads” to viral metagenomics detection and routine genomic surveillance

INSAFLU (“INSide the FLU”) is a free bioinformatics web-based (also locally installable) suite that deals with primary sequencing data (Illumina, Ion Torrent and Oxford Nanopore Technologies reads) towards:

- **metagenomics virus detection** (from reads to virus detection).
- **routine genomic surveillance** (from reads to mutation detection, consensus generation, virus classification, alignments, “genotype-phenotype”, screening, phylogenetics, integrative Nextstrain phylogeographical and temporal analysis etc).

INSAFLU-TELEVIR versatility and functionality is expected to supply public health laboratories and researchers with a user-oriented “start-to-end” bioinformatics framework that can potentiate a strengthened and timely detection and monitoring of viral (emerging) threats.



News

INSAFLU now also includes a pathogen detection module.

December 21, 2022 by João Santos

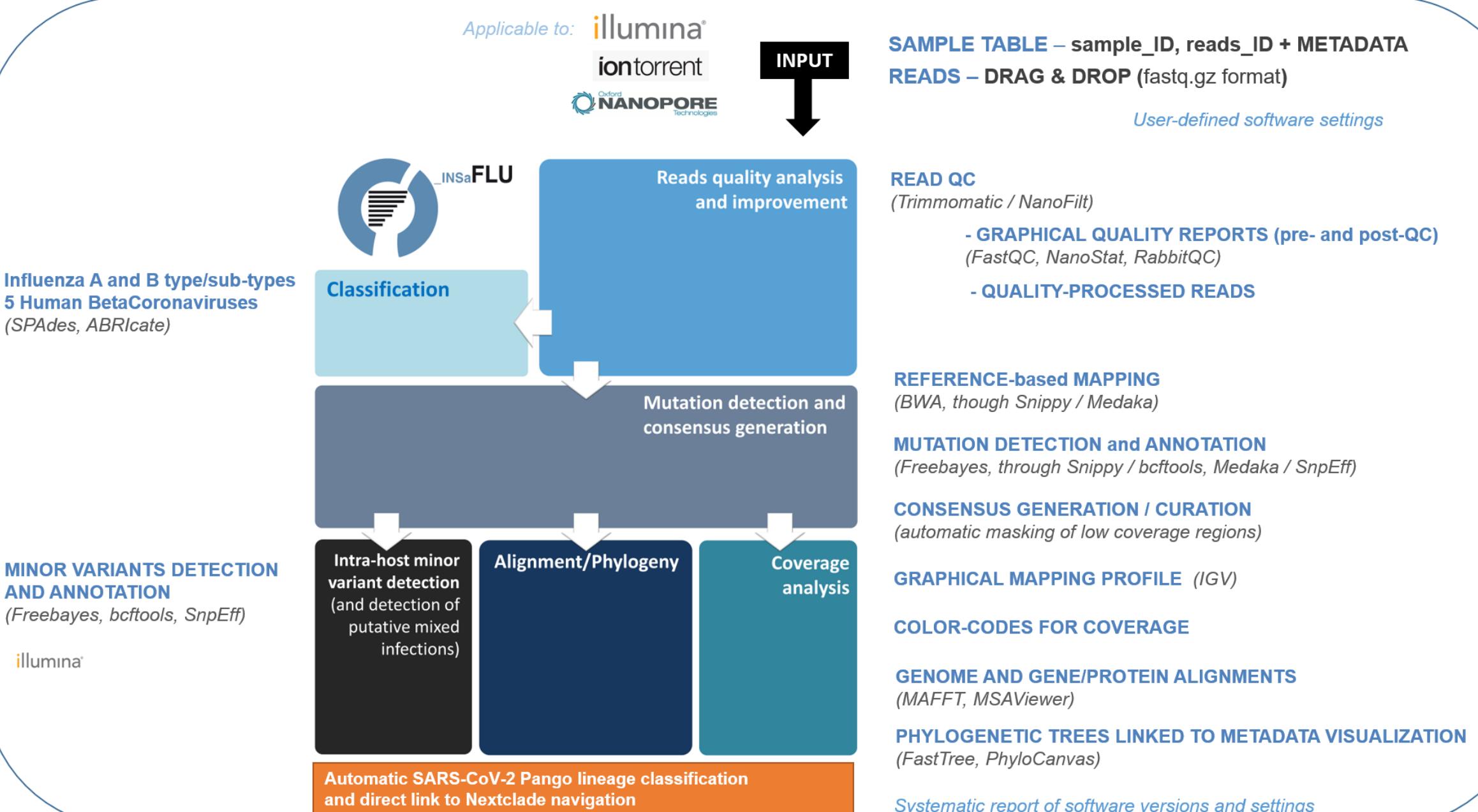
This upgrade is available both in INSAFLU free online and locally installable

(<https://github.com/INSAFLU/docker>) versions. Please check all details in INSAFLU documentation (<https://insaflu.readthedocs.io/en/latest/>)

INSAFLU now also enables Nextstrain phylogenetic and geotemporal analysis.

October 27, 2022 by Daniel Sobral

Overview of INSaFLU pipeline



Upload sample metadata and NGS data

The screenshot shows the INSAFLU web application interface. At the top left is the logo of the Instituto Nacional de Saúde Doutor Ricardo Jorge. Next to it is the INSAFLU logo. On the right side, there are links for 'Documentation' and 'clydedapat - Logout'. A vertical sidebar on the far left contains several icons: a gear, a magnifying glass, a wrench (highlighted with a red box), a grid, and a pencil.

The main content area has a header with 'Home / Samples'. Below this is a table with the following columns: Sample Name, Uploaded Date, #Fastq Files, Technology, Classification, Data Set, Alerts, #Quality Seq. (Fastq1)-(Fastq2), and Extra Info. A blue button labeled 'Add Samples ▾' is located at the top left of the table. To the right of the table is a search bar with the placeholder 'Name, Type, Data set, T...' and a 'Search' button. At the bottom left is a 'Refresh' button, and at the bottom right is the text 'Total samples: 0'.

Sample Name	Uploaded Date	#Fastq Files	Technology	Classification	Data Set	Alerts	#Quality Seq. (Fastq1)-(Fastq2)	Extra Info
There are no Samples to show...								

INSAFLU needs:

1. NGS data (fastq reads)
2. Sample metadata (to link each sample to corresponding NGS data)



Home / Samples

Add Samples ▾

Add One Sample

Add one sample and metadata at a time

fq Files

Technology	Classification	Data Set	Alerts	#Quality Seq. (Fastq1)-(Fastq2)	Extra Info
------------	----------------	----------	--------	---------------------------------	------------

Add Multiple Samples

Add Fastq Files

Update metadata



Total samples: 0

[Home](#) / [Samples](#) / Add a single sample

General data

Name*

SA32

Unique identifier for this sample. Only letters, numbers and underscores are allowed.

 Dataset

Generic

Specific dataset (useful for grouping sets of samples)

 Vaccine status

Discrimination of vaccination status

Dates

Choose a date

- Onset date
- Collection date
- Lab reception date

Choose the option you want to be used to calculate the calendar week number.

Onset date

dd/mm/yyyy



Date of onset

Collection date

08/02/2023



Date of collection

Lab reception date

dd/mm/yyyy



Date receipt on the lab.

Choose a date

Onset date

Collection date

Lab reception date

Choose the option you want to be used to calculate the calendar week number.

Onset date

dd/mm/yyyy

Collection date

08/02/2023

Lab reception date

dd/mm/yyyy

Global position

Latitude

Geolocation where the sample was collected. Decimal Degrees (DD) format.

Longitude

Geolocation where the sample was collected. Decimal Degrees (DD) format.

Fastq files

Raw fastq.gz (R1)*

N10071616_SA32.fastq.gz

Max raw file R1 with fastq gzip file (< 410.0 MB).

Raw fastq.gz (R2)

No file chosen

Max raw file R2 with fastq gzip file (< 410.0 MB).

 Instituto Nacional de Saúde Doutor Ricardo Jorge INSAFLU

Documentation  clydedapat - Logout

Sample 'SA32' was created successfully 

Home / Samples

Add Samples ▾ 

Sample Name	Uploaded Date	#Fastq Files	Technology	Classification	Data Set	Alerts	#Quality Seq. (Fastq1)-(Fastq2)	Extra Info
 SA32	07-06-2023 07:24	1	ONT	Not yet	Generic	0	Not yet	Processing

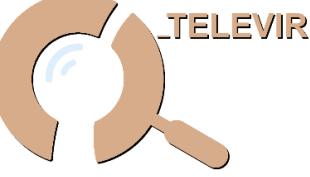
 Refresh Total samples: 1

Create a project

INSaFLU

Documentation clydedapat - Logout

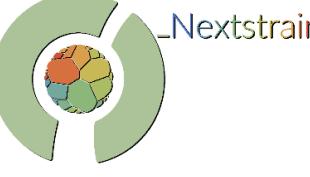
Home / Project Index

 **TELEVIR**

Virus Detection
From Reads to viral detection

 **INSaFLU**

Reference-based genome assembly
From reads to consensus sequence, mutations list, alignments, Pango lineage classification, etc.

 **Nextstrain**

Nextstrain
From consensus sequences to phylogeographic and temporal analysis and metadata navigation

 **Projects**

 **Projects**

Datasets

[Home](#) / [Project Index](#) / [Projects](#)

 Create project

Name, Reference, Samp

 Search

Project Name	Reference	Last Change Date	Creation Date	#Samples (P/W/E)	Results
--------------	-----------	------------------	---------------	------------------	---------

There are no Projects to show...

 Refresh

Total projects: 0

Project Name: H1N1pdm09



Select One	Reference Name	Isolate Name	Uploaded Date	Owner	#Locus
<input type="checkbox"/>	hRSV_B_USA_96I_065A_01_01_1996	hRSV_B_USA_96I_065A_01_01_1996	06-03-2023 16:16	system	1
<input type="checkbox"/>	hRSV_B_USA_MCRSV_282_1996	hRSV_B_USA_MCRSV_282_1996	06-03-2023 16:16	system	1
<input type="checkbox"/>	hRSV_B_USA_NH1161_2002	hRSV_B_USA_NH1161_2002	06-03-2023 16:16	system	1
<input type="checkbox"/>	hRSV_B_USA_TH_10526_2014	hRSV_B_USA_TH_10526_2014	06-03-2023 16:16	system	1
<input type="checkbox"/>	hRSV_B_wild_type_strain_B1_WV_14617_1985	hRSV_B_wild_type_strain_B1_WV_14617_1985	06-03-2023 16:16	system	1
<input type="checkbox"/>	A_H3N2_A_Darwin_6_2021	A_H3N2_A_Darwin_6_2021	08-05-2023 09:08	system	8
<input type="checkbox"/>	A_H1N1_A_Wisconsin_67_2022	A_H1N1_A_Wisconsin_67_2022	08-05-2023 09:08	system	8
<input checked="" type="checkbox"/>	A_H1N1pdm09_A_California_07_2009	A_H1N1pdm09_A_California_07_2009	08-05-2023 09:10	system	8
<input type="checkbox"/>	B_Vic_B_Brisbane_60_2008	B_Vic_B_Brisbane_60_2008	08-05-2023 09:10	system	8
<input type="checkbox"/>	A_H1N1pdm09_A_Michigan_45_2015	A_H1N1pdm09_A_Michigan_45_2015	08-05-2023 09:10	system	8
<input type="checkbox"/>	A_H3N2_A_Victoria_361_2011	A_H3N2_A_Victoria_361_2011	08-05-2023 09:10	system	8
<input type="checkbox"/>	A_H3N2_A_Perth_16_2009	A_H3N2_A_Perth_16_2009	08-05-2023 09:10	system	8

Choose the software parameters for this project

Instituto Nacional de Saúde Doutor Ricardo Jorge  Documentation  clydedapat - Logout

Project 'H1N1pdm09' was created successfully 

Home / Project Index / Projects

[+ Create project](#) 

Project Name	Reference	Last Change Date	Creation Date	#Samples (P/W/E)	Results
H1N1pdm09	A_H1N1pdm09_A_California_07_2009	Not set yet	07-06-2023 07:44	 (0/0/0) 	

 Total projects: 1



Please choose your software settings

 Extra  Illumina/IonTorrent  ONT

 Mutation detection and consensus generation

Software parameters for project: 'H1N1pdm09'

To Run	Software	Technology	Version	Parameters	Options
<input checked="" type="checkbox"/>	Mask consensus by sites	Extra	1.0	Not applicable	 

Note: you should fit the parameters to the conditions you are working with, e.g., default end's trimming size for ONT may be too restrictive depending on design of your tiling amplicon multiplex PCR method and on your sequencing protocol (if reads are or not already trimmed/clipped).



Please choose your software settings

Extra

Illumina/IonTorrent

ONT

Mutation detection and consensus generation

Coverage analysis

Software parameters for project: 'H1N1pdm09'

To Run	Software	Technology	Version	Parameters	Options
<input checked="" type="checkbox"/>	Consensus Generation (Medaka)	ONT	1.2.1	-m r941_min_high_g360	
<input checked="" type="checkbox"/>	Minimum depth of coverage per site to validate the sequence (mincov)	ONT	1.0	Threshold:30	
<input checked="" type="checkbox"/>	Minimum proportion for variant evidence (minfrac)	ONT	1.0	Threshold:0.80	



Please choose your software settings

☰ Extra

☰ Illumina/IonTorrent

☰ ONT

☰ Mutation detection and consensus generation

☰ Coverage analysis

Software parameters for project: 'H1N1pdm09'

To Run	Software	Technology	Version	Parameters	Options
<input checked="" type="checkbox"/>	Coverage calculation (Samtools)	ONT	1.3	-aa	
<input checked="" type="checkbox"/>	Minimum percentage of horizontal coverage to generate consensus	ONT	1.0	Threshold:70	

Add samples to be included in the project

Instituto Nacional de Saúde Doutor Ricardo Jorge  Documentation  clydedapat - Logout

Project 'H1N1pdm09' was created successfully 

Home / Project Index / Projects

[+ Create project](#) 

Project Name	Reference	Last Change Date	Creation Date	#Samples (P/W/E)	Results
H1N1pdm09	A_H1N1pdm09_A_California_07_2009	Not set yet	07-06-2023 07:44	 (0/0/0)  Add	

 Refresh Total projects: 1

[Home](#) / [Project Index](#) / [Projects](#) / Add samples to project

Project name -> H1N1pdm09

 Add all selected samples Add 1 sample

Name, Type, Week and C

 Search

<input type="checkbox"/>	Sample Name	Uploaded Date	Type-Subtype	Week	Data Set
<input checked="" type="checkbox"/>	SA32	07-06-2023 07:24	A-H1N1	7	Generic

 Go back

Total samples: 1

One sample was added to your project.



Home / Project Index / Projects

 Create project

Download ▾

Name, Reference, Samp

 Search

Project Name	Reference	Last Change Date	Creation Date	#Samples (P/W/E)	Results
 H1N1pdm09	A_H1N1pdm09_A_California_07_2009	Not set yet	07-06-2023 07:44	 (0/1/0)  Add	1 processing

 Refresh

Total projects: 1

Monitor Project's progress

Instituto Nacional de Saúde Doutor Ricardo Jorge 

Documentation  clydedapat - Logout

One sample was added to your project. 

Check Project status:
P: PROCESSED
W: WAITING
E: ERROR

Home / Project Index / Projects

Create project Download ▾

Project Name	Reference	Last Change Date	Creation Date	#Samples (P/W/E)	Results
H1N1pdm09	A_H1N1pdm09_A_California_07_2009	Not set yet	07-06-2023 07:44	 (0/1/0)  Add	1 processing

 Total projects: 1

Processed: 0
Waiting: 1
Error: 0

INSAFLU projects are automatically run upon creation. Users may start monitoring the Project progress by checking the number of samples in the following status: Processed (P); Waiting (W); and Error (E)

View Results

Instituto Nacional de Saúde Doutor Ricardo Jorge 

[Documentation](#)  clydedapat - Logout

Home / Project Index / Projects

		Create project	Download ▾		<input type="text" value="Name, Reference, Samp"/>	Search
Project Name	Reference	Last Change Date	Creation Date	#Samples (P/W/E)	Results	
 H1N1pdm09	A_H1N1pdm09_A_California_07_2009	07-06-2023 08:00	07-06-2023 07:44	 (1/0/0)  Add	  Total projects: 1	

[!\[\]\(587df458d51bc842b910a8c21f247a9c_img.jpg\) Refresh](#)

[Home](#) / [Projects](#) / Show project results

Project 'H1N1pdm09'

Project Name: H1N1pdm09**Reference Name:** A_H1N1pdm0...2009.fasta**#Samples:** 1**#Samples with Alerts:** 1**#Samples in process:** 0**#Samples with errors:** 0**Coverage:** [coverage.tsv](#)**Variants:** [validated_variants.tsv](#)**Minor intra-host SNVs:** File not available yet.

[Phylogenetic trees and metadata visualization](#)

[Mutations list](#)

[Nucleotide alignments by MSAViewer](#)

[Amino acid alignments by MSAViewer](#)

[Coverage for all samples](#)


Project Name: H1N1pdm09

Reference Name: A_H1N1pdm0...2009.fasta

#Samples: 1

#Samples with Alerts: 1

#Samples in process: 0

#Samples with errors: 0

Coverage: coverage.tsv

Variants: validated_variants.tsv

Minor intra-host SNVs: File not available yet.

Phylogenetic trees and metadata visualization

[Mutations list](#)
[Nucleotide alignments by MSAViewer](#)
[Amino acid alignments by MSAViewer](#)
[Coverage for all samples](#)
[Download ▾](#)
[Jump to NextClade ▾](#)

[Q Search](#)

Sample Name	Type And Subtype	Putative Mixed Infection	Technology	Dataset	Coverage	Consensus File	Alerts	Results
SA32	A-H1N1	NA (not applicable)	ONT	Generic		Consens...32.fasta	4	  

[Go back](#)
[Refresh](#)

Total samples: 1

Project INSAFLU
Phylogenetic trees and metadata visualization
Mutations list

ID	CHROM	POS	TYPE	REF	ALT	FREQ	COVERAGE	EVIDENCE	FTYPE	STRAND	NT_POS	AA_POS	EFFECT	NT_CHANGE	AA_CHANGE	AA_CHANGE_ALT	Locus
	Clear	▼		Cle	▼				Clea	▼	Cle	▼		Clear	▼		Clear
SA32	4	22	snp	C	A	98	147/242	A:144 C:3	CDS	+	22/51	8/16	missense_variant	c.22C>A	p.Leu8Met	p.L8M	locus
SA32	4	27	snp	A	G	96.6	145/233	G:140 A:5	CDS	+	27/51	9/16	synonymous_variant	c.27A>G	p.Leu9Leu	p.L9L	locus
SA32	4	37	snp	G	A	97.3	221/243	A:215 G:6	CDS	+	37/51	13/16	missense_variant	c.37G>A	p.Ala13Thr	p.A13T	locus
SA32	4	99	snp	A	G	98.3	237/234	G:233 A:4	CDS	+	48/981	16/326	synonymous_variant	c.48A>G	p.Val16Val	p.V16V	locus
SA32	4	141	snp	T	C	99.6	258/258	C:257 T:1	CDS	+	90/981	30/326	synonymous_variant	c.90T>C	p.Val30Val	p.V30V	locus
SA32	4	144	snp	C	T	98.4	249/251	T:245 C:4	CDS	+	93/981	31/326	synonymous_variant	c.93C>T	p.Asn31Asn	p.N31N	locus
SA32	4	168	snp	G	A	98.4	244/260	A:240 G:4	CDS	+	117/981	39/326	synonymous_variant	c.117G>A	p.Gly39Gly	p.G39G	locus
SA32	4	189	snp	G	A	97.9	240/254	A:235 G:5	CDS	+	138/981	46/326	synonymous_variant	c.138G>A	p.Gly46Gly	p.G46G	locus

Variants: 233



Download as CSV

Nucleotide alignments by MSAViewer

Phylogenetic trees and metadata visualization

Mutations list

Nucleotide alignments by MSAViewer

Amino acid alignments by MSAViewer

Coverage for all samples

Coverage file: [coverage.tsv](#)

Samples	1	2	3	4	5	6	7	8
SA32	6.5 - 0.0% (30x)	10.2 - 0.0% (30x)	1.9 - 0.0% (30x)	58.5 - 100.0% (30x)	55.0 - 82.0% (30x)	12.3 - 100.0% (30x)	72.7 - 100.0% (30x)	78.6 - 100.0% (30x)

[Download](#) ▾

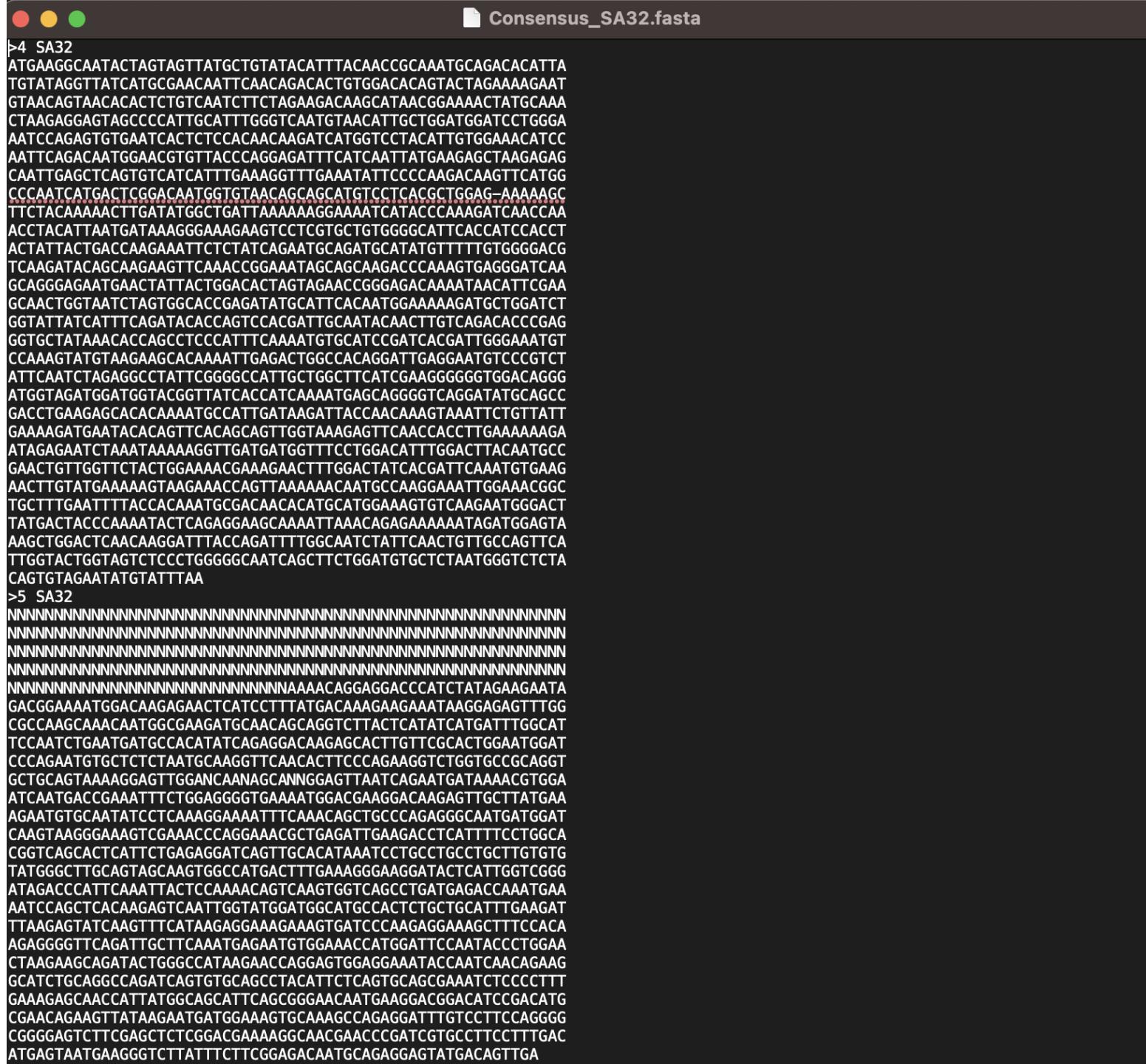
[Jump to NextClade](#) ▾

[Search](#)

Sample Name	Type And Subtype	Putative Mixed Infection	Technology	Dataset	Coverage	Consensus File	Alerts	Results
SA32	A-H1N1	NA (not applicable)	ONT	Generic		Consens...32.fasta	4	   More info

[Go back](#)
[Refresh](#)

Total samples: 1



[Home](#) / [Projects](#) / Show project results[Download - Sample_list.csv](#)[Download - Sample_list.tsv](#)[Download - Sample_list_settings.csv](#)[Download - Sample_list_settings.tsv](#)[Download - AllConsensus.fasta](#)[Download - coverage.tsv](#)[Download - validated_variants.tsv](#)[Download - AllFiles.zip](#)[Download ▾](#)[Jump to NextClade ▾](#) Name, Mixed, Type and I[Search](#)

Sample Name	Type And Subtype	Putative Mixed Infection	Technology	Dataset	Coverage	Consensus File	Alerts	Results
 SA32	A-H1N1	NA (not applicable)	ONT	Generic		Consens...32.fasta	4	More info

[← Go back](#)[↻ Refresh](#)

Total samples: 1

Check FASTQC result

Instituto Nacional de Saúde Doutor Ricardo Jorge INS-FLU

Documentation clydedapat - Logout

Home / Samples

Add Samples ▾ Download ▾ Name, Type, Data set, Te Q Search

Sample Name	Uploaded Date	#Fastq Files	Technology	Classification	Data Set	Alerts	#Quality Seq. (Fastq1)-(Fastq2)	Extra Info
SA32	07-06-2023 07:24	1	ONT	A-H1N1	Generic	0	ⓘ (51999/727.5) (51999)	More Info

Refresh Total samples: 1



[Home](#) / [Samples](#) / Sample details**Name:** SA32**Onset Date:****Collection Date:** Feb. 8, 2023**Reception Date:****Week:** 7**Vaccine status:****Data set:** Generic**Alerts:** 0**Upload date:** June 7, 2023**Geolocation pos.(Lat, Lng):****Type and subtype/lineage:** A-H1N1**Owner:** clydedapat

Original fastq

Quality processed reads

Filter Stat.

Type and subtype/lineage

Alerts

Original fastq files and quality result

Original fastq: [N10071616_SA32.fastq.gz](#)**Original fastq quality:** [N10071616_SA32.fastq.gz.html](#)**Software version:** NanoStat-1.4.0 [Go back](#)



Home / Samples / Sample details

Name: SA32

Onset Date:

Collection Date: Feb. 8, 2023

Reception Date:

Week: 7

Vaccine status:

Data set: Generic

Alerts: 0

Upload date: June 7, 2023

Geolocation pos.(Lat, Lng):

Type and subtype/lineage: A-H1N1

Owner: clydedapat

Original fastq

Quality processed reads

Filter Stat.

Type and subtype/lineage

Alerts

Processed reads and quality results

NanoFilt fastq: [SA32.fastq.gz](#)

RabbitQC fastq quality: [SA32_rabbitQC.html](#)

Software version: NanoFilt-2.6.0; (-q 8 -l 50 --headcrop 30 --tailcrop 30 --maxlength 50000)/NanoStat-1.4.0

Go back



Name: SA32 **Onset Date:** **Collection Date:** Feb. 8, 2023 **Reception Date:**

Week: 7 **Vaccine status:** **Data set:** Generic **Alerts:** 0

Upload date: June 7, 2023 **Geolocation pos.(Lat, Lng):** **Type and subtype/lineage:** A-H1N1 **Owner:** clydedapat

 [Original fastq](#) [Quality processed reads](#) [Filter Stat.](#) [Type and subtype/lineage](#) [Alerts](#)

NanoStat results

	Original	Processed	Difference	Remain %
Mean read length	787.4	727.5	-59.9	92.4
Mean read quality	12.0	12.3	0.3	--
Median read length	756.0	696.0	-60.0	92.1
Median read quality	11.8	12.1	0.3	--
Number of reads	52,000.0	51,999.0	-1.0	100.0
Read length N50	803.0	748.0	-55.0	--
STDEV read length	311.1	311.1	0.0	--
Total bases	40,947,363.0	37,827,326.0	-3,120,037.0	92.4

Software version: NanoFilt-2.6.0; (-q 8 -l 50 --headcrop 30 --tailcrop 30 --maxlength 50000)/NanoStat-1.4.0

[Home](#) / [Samples](#) / Sample details**Name:** SA32**Onset Date:****Collection Date:** Feb. 8, 2023**Reception Date:****Week:** 7**Vaccine status:****Data set:** Generic**Alerts:** 0**Upload date:** June 7, 2023**Geolocation pos.(Lat, Lng):****Type and subtype/lineage:** A-H1N1**Owner:** clydedapat [Original fastq](#) [Quality processed reads](#) [Filter Stat.](#) [Type and subtype/lineage](#) [Alerts](#)

Type and subtype/lineage

Type: Subtype**Name:** H1**Type:** Subtype**Name:** N1**Type:** Type**Name:** A**Draft assembly:****Seg./Ref. to reads:** [SA32_segments_to_reads.tsv](#)[How to take advantage of draft assemblies](#)