National Institute of Health (NIH) Dr. Ricardo Jorge, Lisbon, Portugal

INSaFLU-TELEVIR: an open web-based bioinformatics suite for metagenomic virus detection and routine genomic surveillance

https://insaflu.insa.pt

Joana Isidro
National Institute of Health Dr. Ricardo Jorge, Portugal
Department of Infectious Diseases
Bioinformatics Unit

joana.isidro@insa.min-saude.pt

New era of virus surveillance *Targeting the whole-genome*

Many laboratories do not have bioinformatics capabilities and/or staff needed to analyze raw NGS data, which is an obstacle in the transition to whole-genome sequencing

Next-generation Sequencing

illumina

iontorrent

Oxford NANOPORE Technologies

From "reads" to WGS-based first-line influenza laboratory surveillance

INSaFLU ("INSide the FLU") is an influenza-specific bioinformatics free web-based suite that deals with primary data (reads) towards the automatic generation of the output data that are actually the core first-line "genetic requests" for effective and timely influenza laboratory surveillance (e.g., type and sub-type, gene and whole-genome consensus sequences, variants annotation, alignments and phylogenetic trees). Data integration is continuously scalable, fitting the need for a real-time epidemiological surveillance during the flu epidemics.

Future Directions

How to cite

Contacts

News

Useful links

https://insaflu.insa.pt

Borges et al, 2018
Genome Medicine 10: 46

Which are the main features of INSaFLU?

https://insaflu.insa.pt

- Online platform (free, confidential accounts) but it can be installed locally
- User-oriented / no advanced expertise in bioinformatics is needed
- Applicable to NGS data (single / paired-end data; Illumina / Ion Torrent / ONT) collected from diverse amplicon-based schemas

Which are the main features of INSaFLU?

INSaFLU

https://insaflu.insa.pt

- Tailored to seasonal human influenza and SARS-CoV-2 but is applicable to other pathogens
- Dynamically integrates data in a cumulative manner, fitting the continuous epidemiological surveillance during epidemics
- Outputs are standardized and compatible with other platforms for downstream data visualization/analysis

https://insaflu.insa.pt

INSaFLU-TELEVIR: an open web-based bioinformatics suite for metagenomic virus detection and **routine genomic surveillance** (influenza, SARS-CoV-2 and other pathogens)

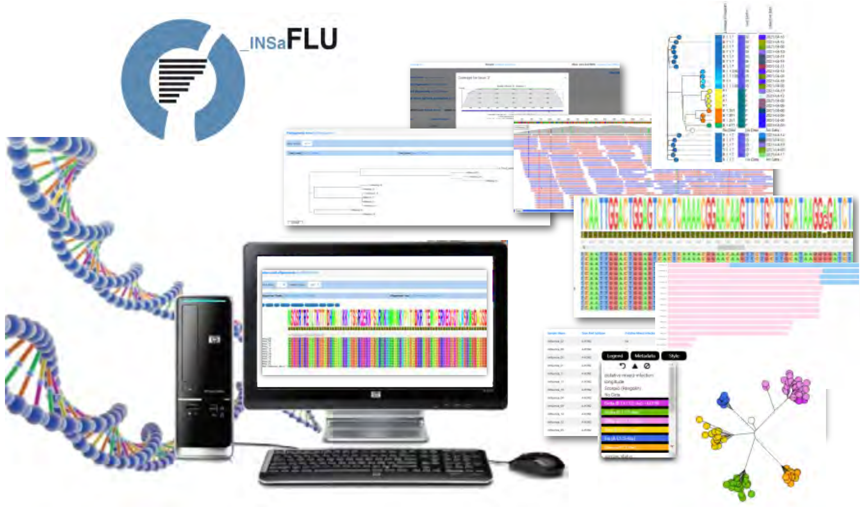
INPUTS

✓ **Sample metadata table**

+

✓ **NGS data ***

**Compatible with distinct sequencing technologies:*
Illumina, Ion Torrent and Oxford Nanopore



https://insaflu.insa.pt

OUTPUTS

ROUTINE GENOMIC SURVEILLANCE

✓ Read quality analysis and improvement
✓ Human betacoronaviruses, MPXV and influenza type/subtype classification (directly from reads)
✓ Interactive coverage analysis
✓ Mutation annotation and consensus sequence generation/curation
✓ Gene- and genome-based alignments and phylogenies
✓ Intra-host minor variant detection (including flagging of putative mixed infections)
✓ SARS-CoV-2 “Pango lineage” classification
✓ Direct link to “Nextclade” navigation
✓ Genotype-phenotype screening

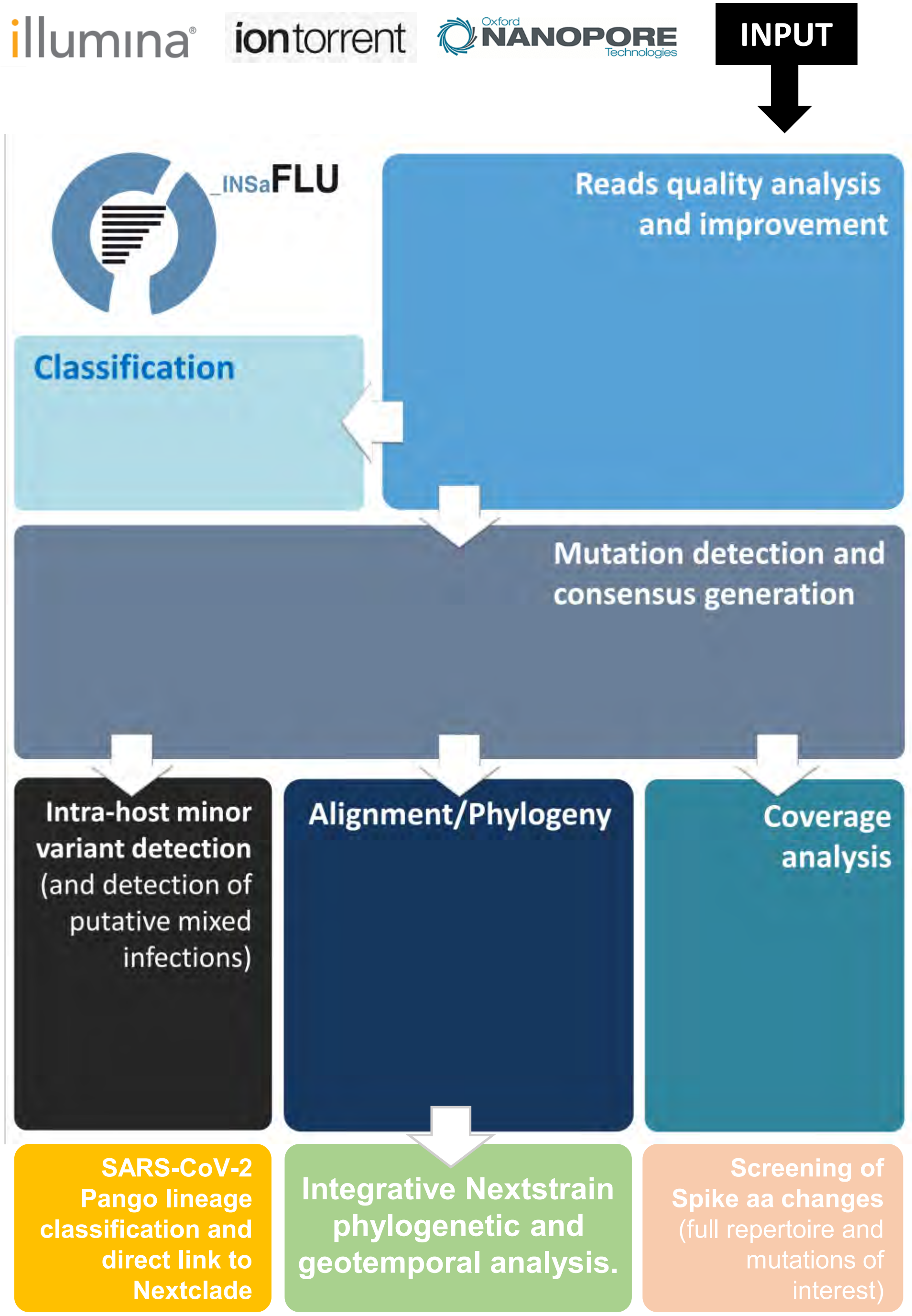
INSaFLU-TELEVIR development is being co-funded by the European Commission on behalf of OneHealth EJP

TELEVIR project: https://onehealthejp.eu/jrp-tele-vir/

illumina iontorrent Oxford NANOPORE Technologies

SAMPLE TABLE – sample_ID, reads_ID + METADATA

READS – DRAG & DROP (fastq.gz format)



INPUT

INSaFLU

Reads quality analysis and improvement

Classification

Mutation detection and consensus generation

Intra-host minor variant detection (and detection of putative mixed infections)

Alignment/Phylogeny

Coverage analysis

SARS-CoV-2 Pango lineage classification and direct link to Nextclade

Integrative Nextstrain phylogenetic and geotemporal analysis.

Screening of Spike aa changes (full repertoire and mutations of interest)

Influenza A and B type/sub-types
5 Human BetaCoronaviruses and MPXV
(SPAdes, ABRIcate)

MINOR VARIANTS DETECTION AND ANNOTATION
(Freebayes, bcftools, SnpEff)
illumina

PANGO LINEAGE CLASSIFICATION AND NEXTCLADE LINKAGES

READ QC
(Trimmomatic / NanoFilt)

- GRAPHICAL QUALITY REPORTS (pre- and post-QC)
(FastQC, NanoStat, RabbitQC)

- QUALITY-PROCESSED READS

REFERENCE-based MAPPING
(BWA / Medaka)

MUTATION DETECTION AND ANNOTATION
(Freebayes / Medaka, bcftools / SnpEff)

CONSENSUS GENERATION / CURATION
(automatic masking of low coverage regions; MSA_masker)

GRAPHICAL MAPPING PROFILE *(IGV)*

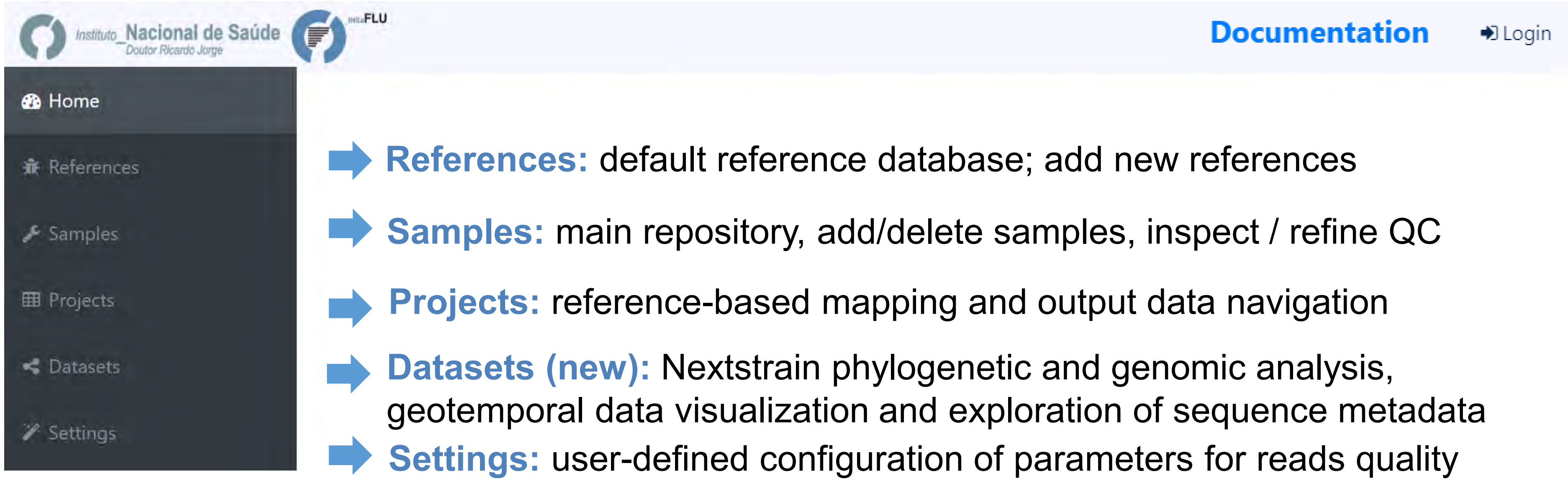
COLOR-CODES FOR COVERAGE

GENOME AND GENE/PROTEIN ALIGNMENTS
(MAFFT, MSAViewer)

PHYLOGENETIC TREES LINKED TO METADATA VISUALIZATION
(FastTree, PhyloCanvas)

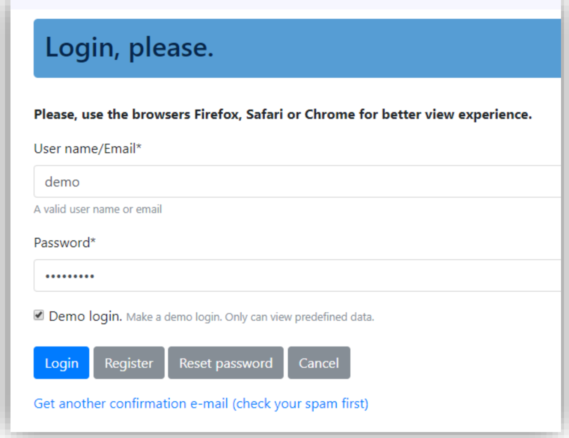
GENOTYPE-PHENOTYPE SCREENING

NEXTSTRAIN PHYLOGENETIC AND GEOTEMPORAL ANALYSIS.
(Seasonal Influenza, SARS-CoV-2, Monkeypox, and a "generic" build other viruses)



- **References:** default reference database; add new references
- **Samples:** main repository, add/delete samples, inspect / refine QC
- **Projects:** reference-based mapping and output data navigation
- **Datasets (new):** Nextstrain phylogenetic and genomic analysis, geotemporal data visualization and exploration of sequence metadata
- **Settings:** user-defined configuration of parameters for reads quality analysis, mapping, mutation validation and consensus generation/masking

Register and **Login** to a user restricted account



Login, please.

Please, use the browsers Firefox, Safari or Chrome for better view experience.

User name/Email*

demo

A valid user name or email

Password*

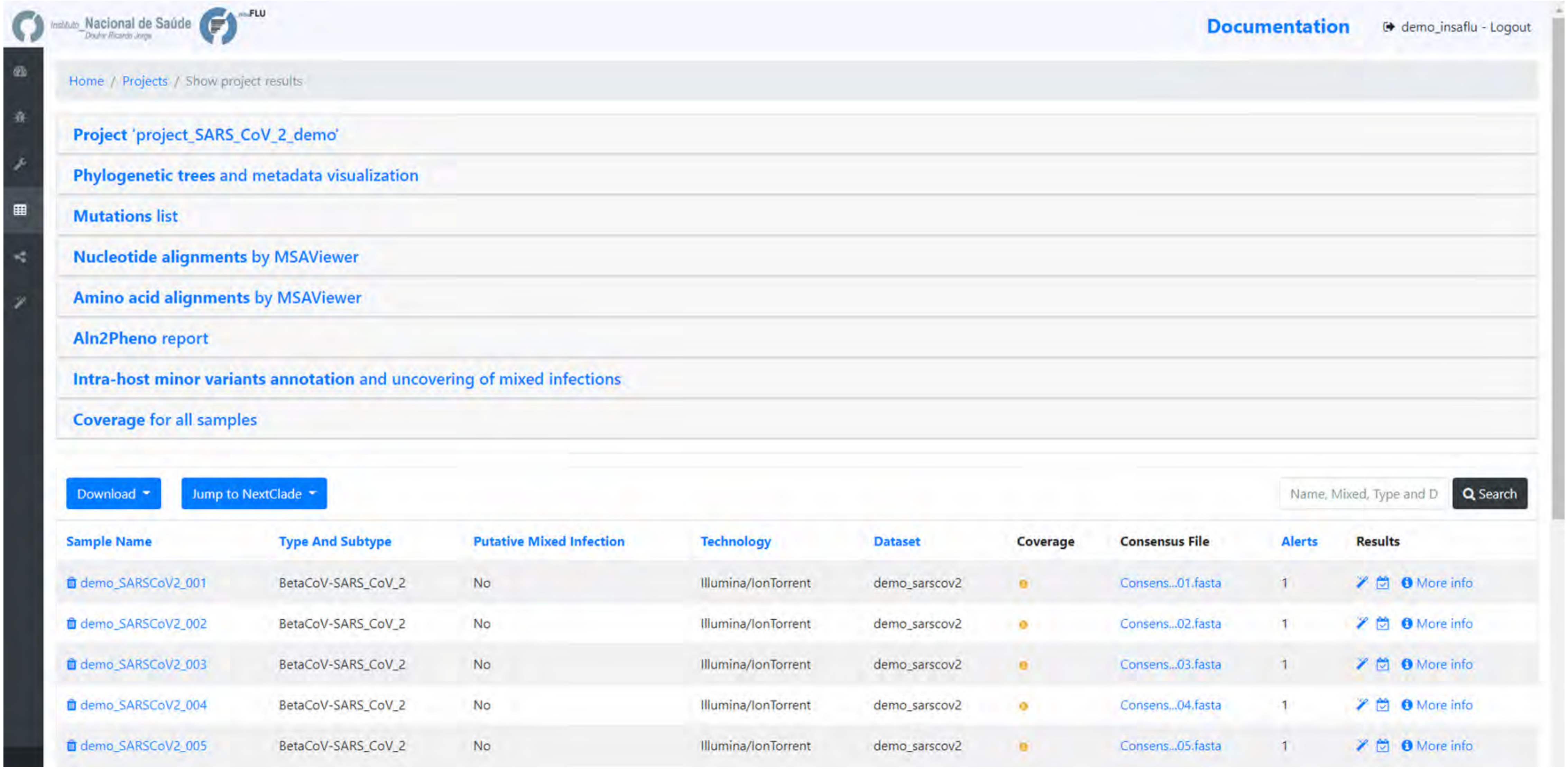
•••••••••

☑ Demo login. Make a demo login. Only can view predefined data.

Applicable to: illumina® iontorrent Oxford NANOPORE Technologies

https://insaflu.insa.pt

INSaFLU projects - reference-based mapping and output data navigation



Home / Projects / Show project results

Project 'project_SARS_CoV_2_demo'

Phylogenetic trees and metadata visualization

Mutations list

Nucleotide alignments by MSAViewer

Amino acid alignments by MSAViewer

Aln2Pheno report

Intra-host minor variants annotation and uncovering of mixed infections

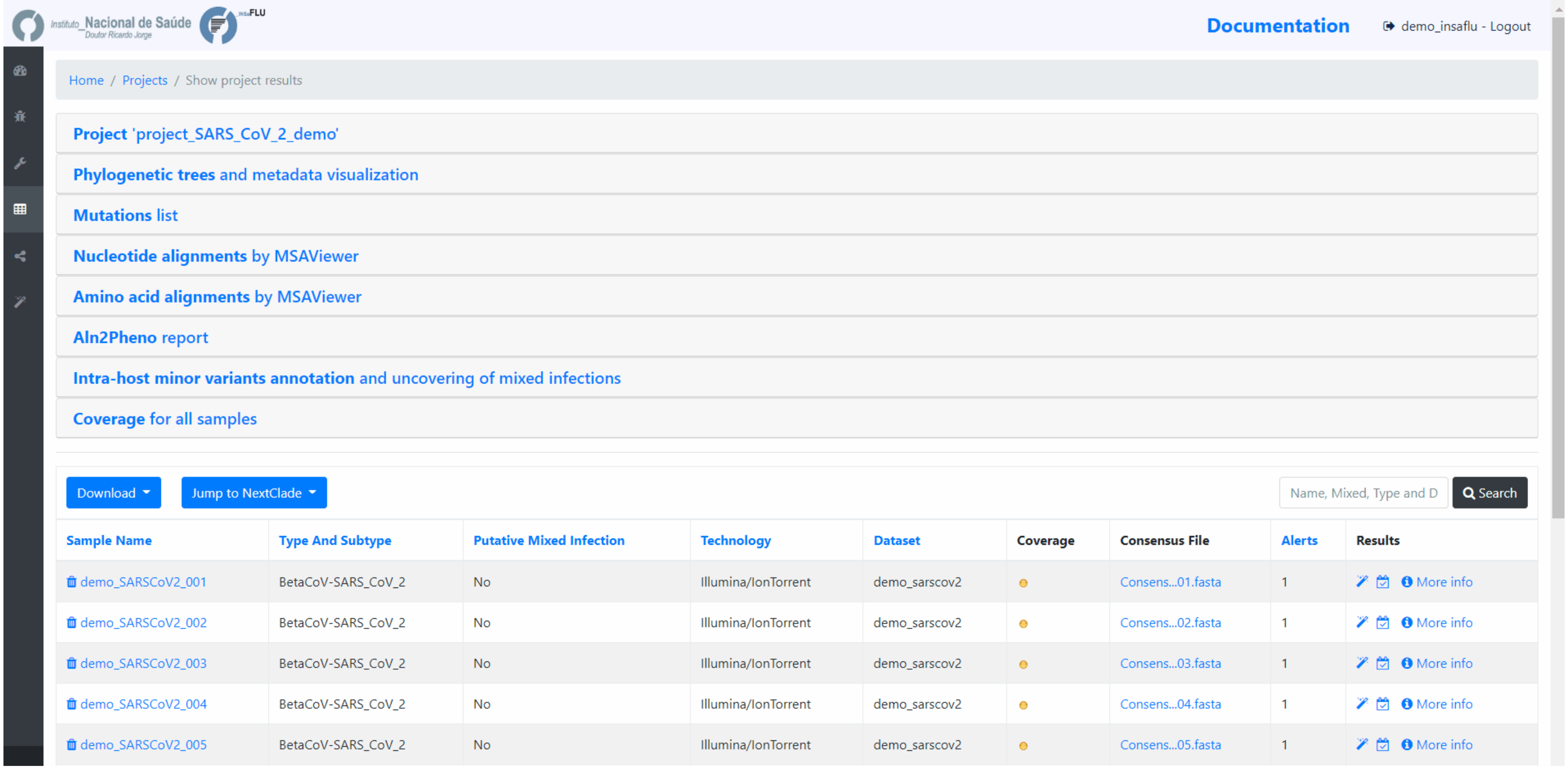
Coverage for all samples

Sample Name	Type And Subtype	Putative Mixed Infection	Technology	Dataset	Coverage	Consensus File	Alerts	Results
demo_SARSCoV2_001	BetaCoV-SARS_CoV_2	No	Illumina/IonTorrent	demo_sarscov2	●	Consens...01.fasta	1	More info
demo_SARSCoV2_002	BetaCoV-SARS_CoV_2	No	Illumina/IonTorrent	demo_sarscov2	●	Consens...02.fasta	1	More info
demo_SARSCoV2_003	BetaCoV-SARS_CoV_2	No	Illumina/IonTorrent	demo_sarscov2	●	Consens...03.fasta	1	More info
demo_SARSCoV2_004	BetaCoV-SARS_CoV_2	No	Illumina/IonTorrent	demo_sarscov2	●	Consens...04.fasta	1	More info
demo_SARSCoV2_005	BetaCoV-SARS_CoV_2	No	Illumina/IonTorrent	demo_sarscov2	●	Consens...05.fasta	1	More info

Multiple outputs in nomenclature-stable and standardized formats that can be downloaded or visually explored *in situ.*

https://insaflu.insa.pt

Examples of INSaFLU projects outputs



Home / Projects / Show project results

Project 'project_SARS_CoV_2_demo'

Phylogenetic trees and metadata visualization

Mutations list

Nucleotide alignments by MSAViewer

Amino acid alignments by MSAViewer

Aln2Pheno report

Intra-host minor variants annotation and uncovering of mixed infections

Coverage for all samples

Sample Name	Type And Subtype	Putative Mixed Infection	Technology	Dataset	Coverage	Consensus File	Alerts	Results
demo_SARSCoV2_001	BetaCoV-SARS_CoV_2	No	Illumina/IonTorrent	demo_sarscov2	●	Consens...01.fasta	1	More info
demo_SARSCoV2_002	BetaCoV-SARS_CoV_2	No	Illumina/IonTorrent	demo_sarscov2	●	Consens...02.fasta	1	More info
demo_SARSCoV2_003	BetaCoV-SARS_CoV_2	No	Illumina/IonTorrent	demo_sarscov2	●	Consens...03.fasta	1	More info
demo_SARSCoV2_004	BetaCoV-SARS_CoV_2	No	Illumina/IonTorrent	demo_sarscov2	●	Consens...04.fasta	1	More info
demo_SARSCoV2_005	BetaCoV-SARS_CoV_2	No	Illumina/IonTorrent	demo_sarscov2	●	Consens...05.fasta	1	More info

Expand-and-collapse panels

Multiple outputs in nomenclature-stable and standardized formats that can be downloaded or visually explored *in situ*.

https://insaflu.insa.pt

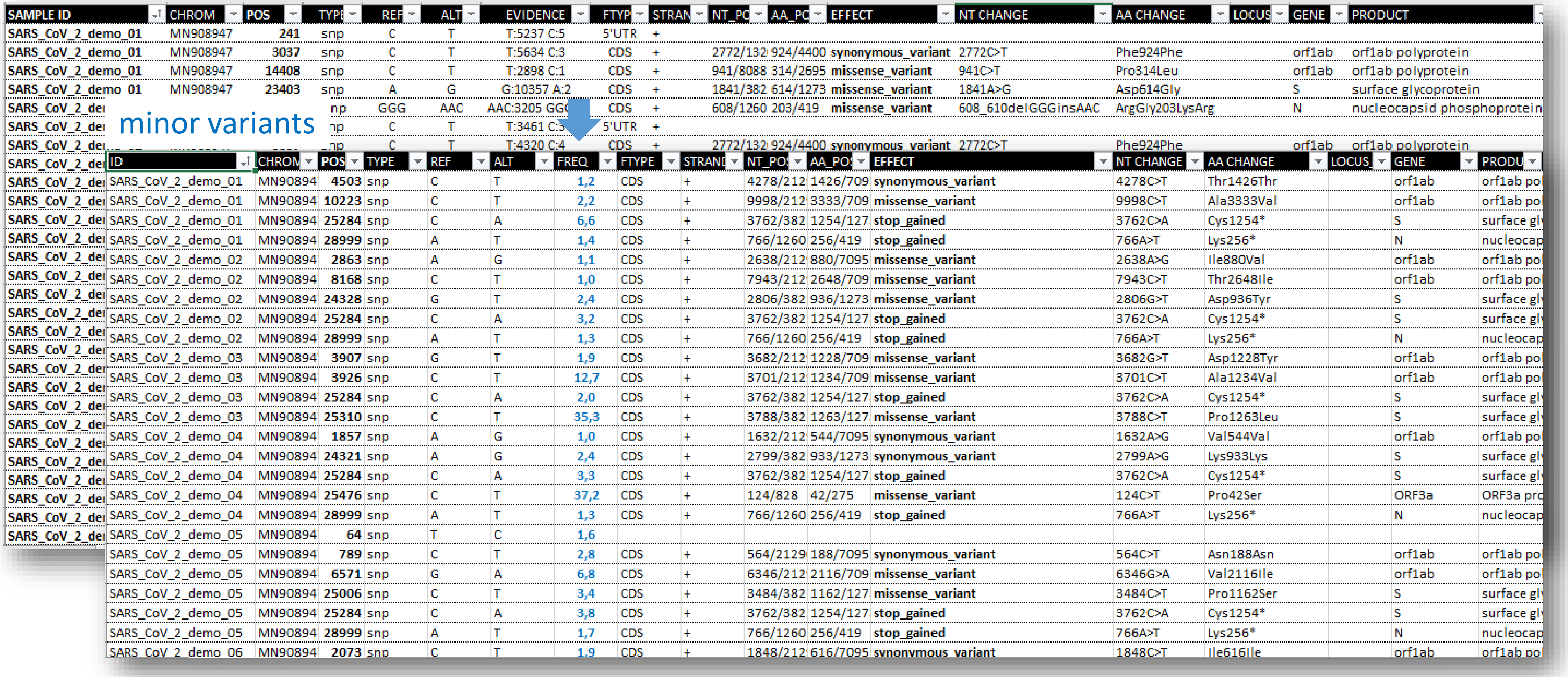
Examples of “Projects” outputs

INSaFLU

https://insaflu.insa.pt

Mutations in consensus

minor variants



SAMPLE ID	CHROM	POS	TYP	REF	ALT	EVIDENCE	FTYP	STRAN	NT_PO	AA_PO	EFFECT	NT CHANGE	AA CHANGE	LOCUS	GENE	PRODUCT
SARS_CoV_2_demo_01	MN908947	241	snp	C	T	T:5237 C:5	5'UTR	+								
SARS_CoV_2_demo_01	MN908947	3037	snp	C	T	T:5634 C:3	CDS	+	2772/132	924/4400	synonymous_variant	2772C>T	Phe924Phe		orf1ab	orf1ab polyprotein
SARS_CoV_2_demo_01	MN908947	14408	snp	C	T	T:2898 C:1	CDS	+	941/8088	314/2695	missense_variant	941C>T	Pro314Leu		orf1ab	orf1ab polyprotein
SARS_CoV_2_demo_01	MN908947	23403	snp	A	G	G:10357 A:2	CDS	+	1841/382	614/1273	missense_variant	1841A>G	Asp614Gly		S	surface glycoprotein
SARS_CoV_2_de			np	GGG	AAC	AAC:3205 GG	CDS	+	608/1260	203/419	missense_variant	608_610delGGGinsAAC	ArgGly203LysArg		N	nucleocapsid phosphoprotein
SARS_CoV_2_de			np	C	T	T:3461 C:3	5'UTR	+								
SARS_CoV_2_de			np	C	T	T:4320 C:4	CDS	+	2772/132	924/4400	synonymous_variant	2772C>T	Phe924Phe		orf1ab	orf1ab polyprotein
SARS_CoV_2_de																
SARS_CoV_2_de																
SARS_CoV_2_de																
SARS_CoV_2_de																
SARS_CoV_2_de																
SARS_CoV_2_de																
SARS_CoV_2_de																
SARS_CoV_2_de																
SARS_CoV_2_de																
SARS_CoV_2_de																
SARS_CoV_2_de																
SARS_CoV_2_de																
SARS_CoV_2_de																
SARS_CoV_2_de																
SARS_CoV_2_de																
SARS_CoV_2_de																
SARS_CoV_2_de																
SARS_CoV_2_de																
SARS_CoV_2_de																
SARS_CoV_2_de																
SARS_CoV_2_de																

ID	CHROM	POS	TYPE	REF	ALT	FREQ	FTYPE	STRAN	NT_PO	AA_PO	EFFECT	NT CHANGE	AA CHANGE	LOCUS	GENE	PRODU
SARS_CoV_2_demo_01	MN90894	4503	snp	C	T	1,2	CDS	+	4278/212	1426/709	synonymous_variant	4278C>T	Thr1426Thr		orf1ab	orf1ab po
SARS_CoV_2_demo_01	MN90894	10223	snp	C	T	2,2	CDS	+	9998/212	3333/709	missense_variant	9998C>T	Ala3333Val		orf1ab	orf1ab po
SARS_CoV_2_demo_01	MN90894	25284	snp	C	A	6,6	CDS	+	3762/382	1254/127	stop_gained	3762C>A	Cys1254*		S	surface gl
SARS_CoV_2_demo_01	MN90894	28999	snp	A	T	1,4	CDS	+	766/1260	256/419	stop_gained	766A>T	Lys256*		N	nucleocap
SARS_CoV_2_demo_02	MN90894	2863	snp	A	G	1,1	CDS	+	2638/212	880/7095	missense_variant	2638A>G	Ile880Val		orf1ab	orf1ab po
SARS_CoV_2_demo_02	MN90894	8168	snp	C	T	1,0	CDS	+	7943/212	2648/709	missense_variant	7943C>T	Thr2648Ile		orf1ab	orf1ab po
SARS_CoV_2_demo_02	MN90894	24328	snp	G	T	2,4	CDS	+	2806/382	936/1273	missense_variant	2806G>T	Asp936Tyr		S	surface gl
SARS_CoV_2_demo_02	MN90894	25284	snp	C	A	3,2	CDS	+	3762/382	1254/127	stop_gained	3762C>A	Cys1254*		S	surface gl
SARS_CoV_2_demo_02	MN90894	28999	snp	A	T	1,3	CDS	+	766/1260	256/419	stop_gained	766A>T	Lys256*		N	nucleocap
SARS_CoV_2_demo_03	MN90894	3907	snp	G	T	1,9	CDS	+	3682/212	1228/709	missense_variant	3682G>T	Asp1228Tyr		orf1ab	orf1ab po
SARS_CoV_2_demo_03	MN90894	3926	snp	C	T	12,7	CDS	+	3701/212	1234/709	missense_variant	3701C>T	Ala1234Val		orf1ab	orf1ab po
SARS_CoV_2_demo_03	MN90894	25284	snp	C	A	2,0	CDS	+	3762/382	1254/127	stop_gained	3762C>A	Cys1254*		S	surface gl
SARS_CoV_2_demo_03	MN90894	25310	snp	C	T	35,3	CDS	+	3788/382	1263/127	missense_variant	3788C>T	Pro1263Leu		S	surface gl
SARS_CoV_2_demo_04	MN90894	1857	snp	A	G	1,0	CDS	+	1632/212	544/7095	synonymous_variant	1632A>G	Val544Val		orf1ab	orf1ab po
SARS_CoV_2_demo_04	MN90894	24321	snp	A	G	2,4	CDS	+	2799/382	933/1273	synonymous_variant	2799A>G	Lys933Lys		S	surface gl
SARS_CoV_2_demo_04	MN90894	25284	snp	C	A	3,3	CDS	+	3762/382	1254/127	stop_gained	3762C>A	Cys1254*		S	surface gl
SARS_CoV_2_demo_04	MN90894	25476	snp	C	T	37,2	CDS	+	124/828	42/275	missense_variant	124C>T	Pro42Ser		ORF3a	ORF3a pro
SARS_CoV_2_demo_04	MN90894	28999	snp	A	T	1,3	CDS	+	766/1260	256/419	stop_gained	766A>T	Lys256*		N	nucleocap
SARS_CoV_2_demo_05	MN90894	64	snp	T	C	1,6										
SARS_CoV_2_demo_05	MN90894	789	snp	C	T	2,8	CDS	+	564/2129	188/7095	synonymous_variant	564C>T	Asn188Asn		orf1ab	orf1ab po
SARS_CoV_2_demo_05	MN90894	6571	snp	G	A	6,8	CDS	+	6346/212	2116/709	missense_variant	6346G>A	Val2116Ile		orf1ab	orf1ab po
SARS_CoV_2_demo_05	MN90894	25006	snp	C	T	3,4	CDS	+	3484/382	1162/127	missense_variant	3484C>T	Pro1162Ser		S	surface gl
SARS_CoV_2_demo_05	MN90894	25284	snp	C	A	3,8	CDS	+	3762/382	1254/127	stop_gained	3762C>A	Cys1254*		S	surface gl
SARS_CoV_2_demo_05	MN90894	28999	snp	A	T	1,7	CDS	+	766/1260	256/419	stop_gained	766A>T	Lys256*		N	nucleocap
SARS_CoV_2_demo_06	MN90894	2073	snp	C	T	1,9	CDS	+	1848/212	616/7095	synonymous_variant	1848C>T	Ile616Ile		orf1ab	orf1ab po

Examples of “Projects” outputs

Genome alignment

Nucleotide alignments by MSAViewer

Select a locus: All

Alignment (.fasta): Alignment_nt_All.fasta

Alignment (.nex): Alignment_nt_All.nex

Protein alignments

e.g. ORF1ab

Amino acid alignments by MSAViewer

Select a locus: SARS_CoV_2

Select a gene: orf1ab

Alignment 'fasta': Alignment_aa_SARS_CoV_2_orf1ab.fasta

Alignment 'nex': Alignment_aa_SARS_CoV_2_orf1ab.nex

e.g. Spike

Amino acid alignments by MSAViewer

Select a locus: SARS_CoV_2

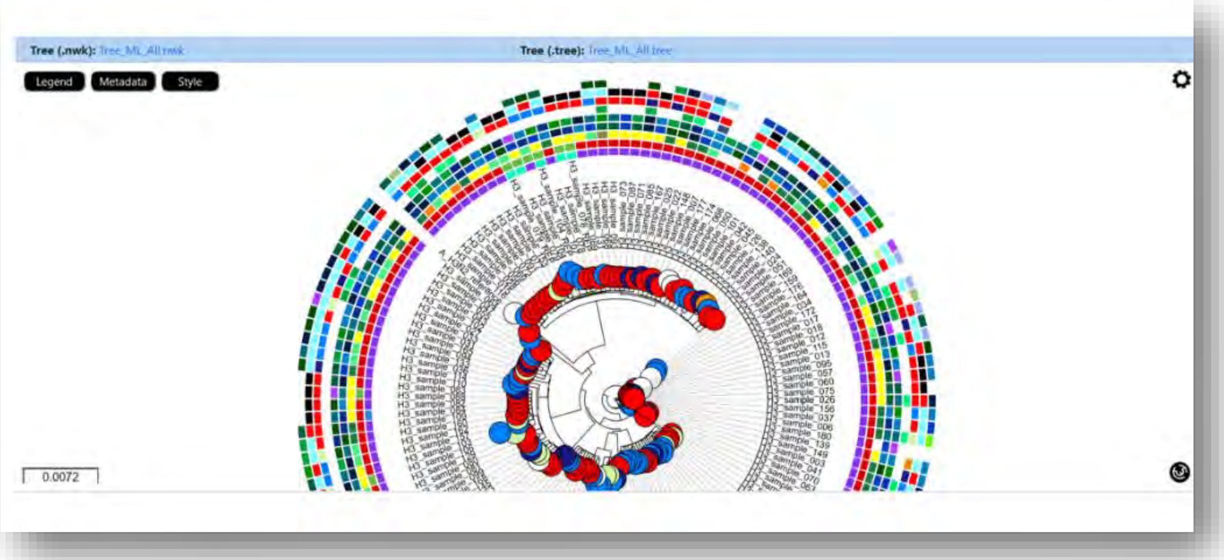
Select a gene: S

Alignment 'fasta': Alignment_aa_SARS_CoV_2_S.fasta

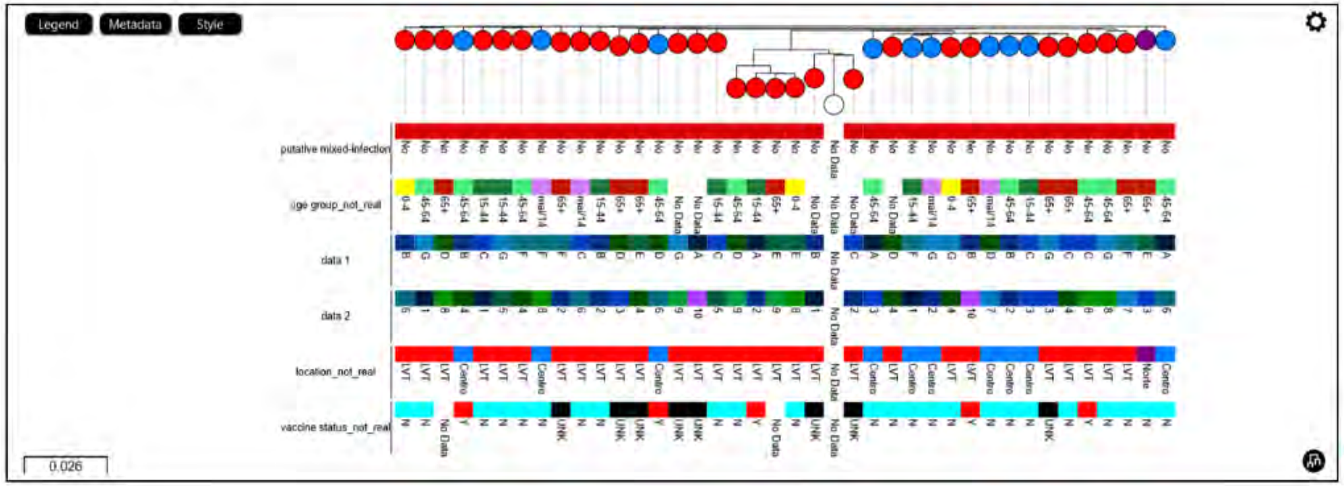
Alignment 'nex': Alignment_aa_SARS_CoV_2_S.nex

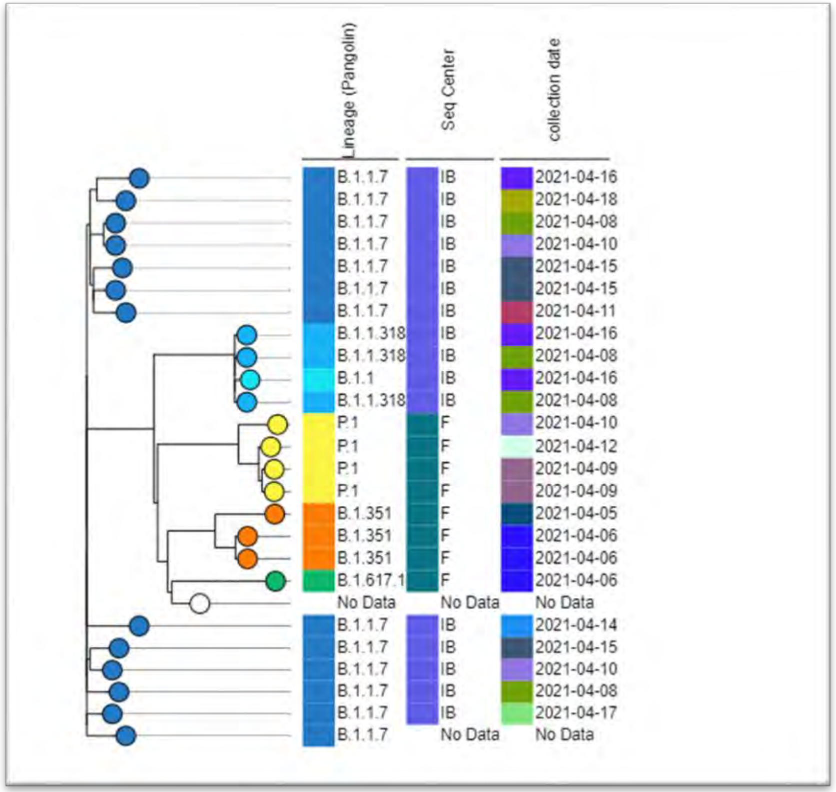
Examples of “Projects” outputs

https://insaflu.insa.pt



Phylogenetic trees and metadata visualization

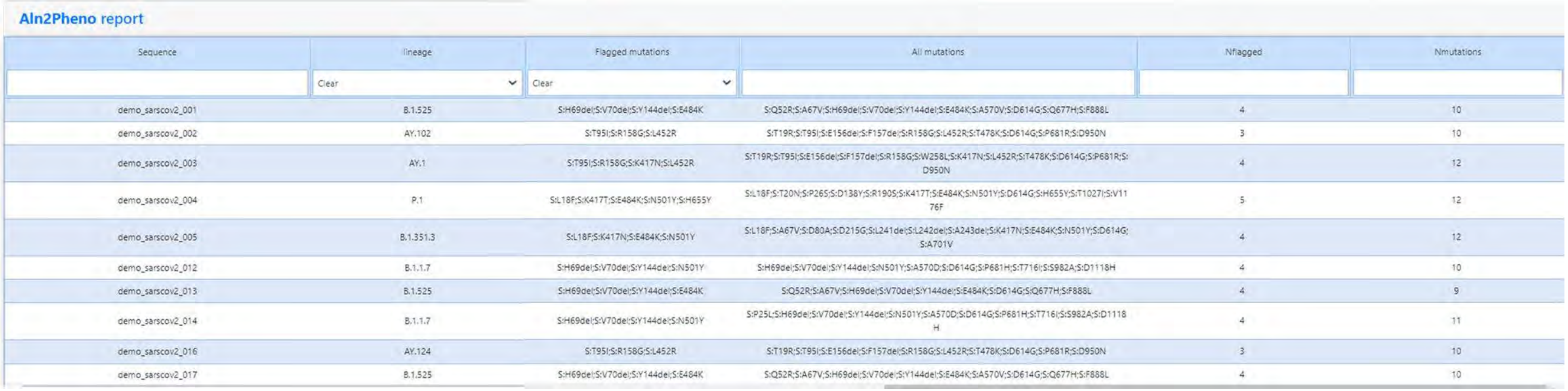




Lineage (Pangolin)	Seq Center	collection date
B.1.1.7	IB	2021-04-16
B.1.1.7	IB	2021-04-18
B.1.1.7	IB	2021-04-08
B.1.1.7	IB	2021-04-10
B.1.1.7	IB	2021-04-15
B.1.1.7	IB	2021-04-15
B.1.1.7	IB	2021-04-11
B.1.1.318	IB	2021-04-16
B.1.1.318	IB	2021-04-08
B.1.1	IB	2021-04-16
B.1.1.318	IB	2021-04-08
P.1	F	2021-04-10
P.1	F	2021-04-12
P.1	F	2021-04-09
P.1	F	2021-04-09
B.1.351	F	2021-04-05
B.1.351	F	2021-04-06
B.1.351	F	2021-04-06
B.1.617.1	F	2021-04-06
No Data	No Data	No Data
B.1.1.7	IB	2021-04-14
B.1.1.7	IB	2021-04-15
B.1.1.7	IB	2021-04-10
B.1.1.7	IB	2021-04-08
B.1.1.7	IB	2021-04-17
B.1.1.7	No Data	No Data

Examples of “Projects” outputs

align2pheno tool (pilot module for SARS-CoV-2) screens and reports **mutations of interest in Spike.**



Aln2Pheno report

Sequence	lineage	Flagged mutations	All mutations	Nflagged	Nmutations
	Clear	Clear			
demo_sarscov2_001	B.1.525	S:H69del;S:V70del;S:Y144del;S:E484K	S:Q52R;S:A67V;S:H69del;S:V70del;S:Y144del;S:E484K;S:A570V;S:D614G;S:Q677H;S:F888L	4	10
demo_sarscov2_002	AY.102	S:T95I;S:R158G;S:L452R	S:T19R;S:T95I;S:E156del;S:F157del;S:R158G;S:L452R;S:T478K;S:D614G;S:P681R;S:D950N	3	10
demo_sarscov2_003	AY.1	S:T95I;S:R158G;S:K417N;S:L452R	S:T19R;S:T95I;S:E156del;S:F157del;S:R158G;S:W258L;S:K417N;S:L452R;S:T478K;S:D614G;S:P681R;S:D950N	4	12
demo_sarscov2_004	P.1	S:L18F;S:K417T;S:E484K;S:N501Y;S:H655Y	S:L18F;S:T20N;S:P26S;S:D138Y;S:R190S;S:K417T;S:E484K;S:N501Y;S:D614G;S:H655Y;S:T1027I;S:V1176F	5	12
demo_sarscov2_005	B.1.351.3	S:L18F;S:K417N;S:E484K;S:N501Y	S:L18F;S:A67V;S:D80A;S:D215G;S:L241del;S:L242del;S:A243del;S:K417N;S:E484K;S:N501Y;S:D614G;S:A701V	4	12
demo_sarscov2_012	B.1.1.7	S:H69del;S:V70del;S:Y144del;S:N501Y	S:H69del;S:V70del;S:Y144del;S:N501Y;S:A570D;S:D614G;S:P681H;S:T716I;S:S982A;S:D1118H	4	10
demo_sarscov2_013	B.1.525	S:H69del;S:V70del;S:Y144del;S:E484K	S:Q52R;S:A67V;S:H69del;S:V70del;S:Y144del;S:E484K;S:D614G;S:Q677H;S:F888L	4	9
demo_sarscov2_014	B.1.1.7	S:H69del;S:V70del;S:Y144del;S:N501Y	S:P25L;S:H69del;S:V70del;S:Y144del;S:N501Y;S:A570D;S:D614G;S:P681H;S:T716I;S:S982A;S:D1118H	4	11
demo_sarscov2_016	AY.124	S:T95I;S:R158G;S:L452R	S:T19R;S:T95I;S:E156del;S:F157del;S:R158G;S:L452R;S:T478K;S:D614G;S:P681R;S:D950N	3	10
demo_sarscov2_017	B.1.525	S:H69del;S:V70del;S:Y144del;S:E484K	S:Q52R;S:A67V;S:H69del;S:V70del;S:Y144del;S:E484K;S:A570V;S:D614G;S:Q677H;S:F888L	4	10

Screening of SARS-CoV-2 Spike amino acid alignments against two “genotype-phenotype” databases: the COG-UK Antigenic mutations (https://sars2.cvr.gla.ac.uk/cog-uk/) and the Pokay Database (https://github.com/nodrogluap/pokay/tree/master/data).

See more details in INSaFLU documentation: https://insaflu.readthedocs.io/en/latest/data_analysis.html#algn2pheno and https://insaflu.readthedocs.io/en/latest/output_visualization.html#h-explore-the-algn2pheno-report-panel-and-results https://cov-lineages.org/

Automate (and up-to-date)

Pango lineage classification

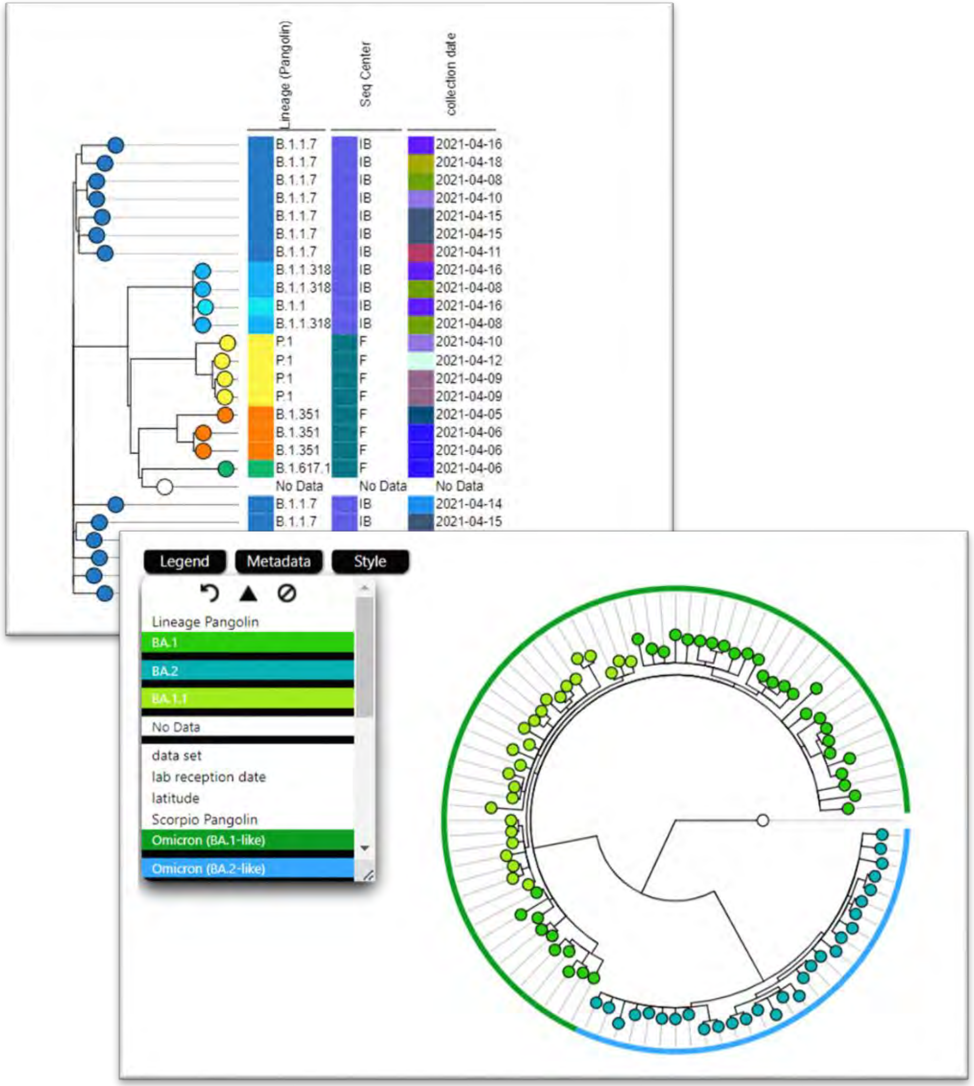
Direct link to

Nextclade navigation

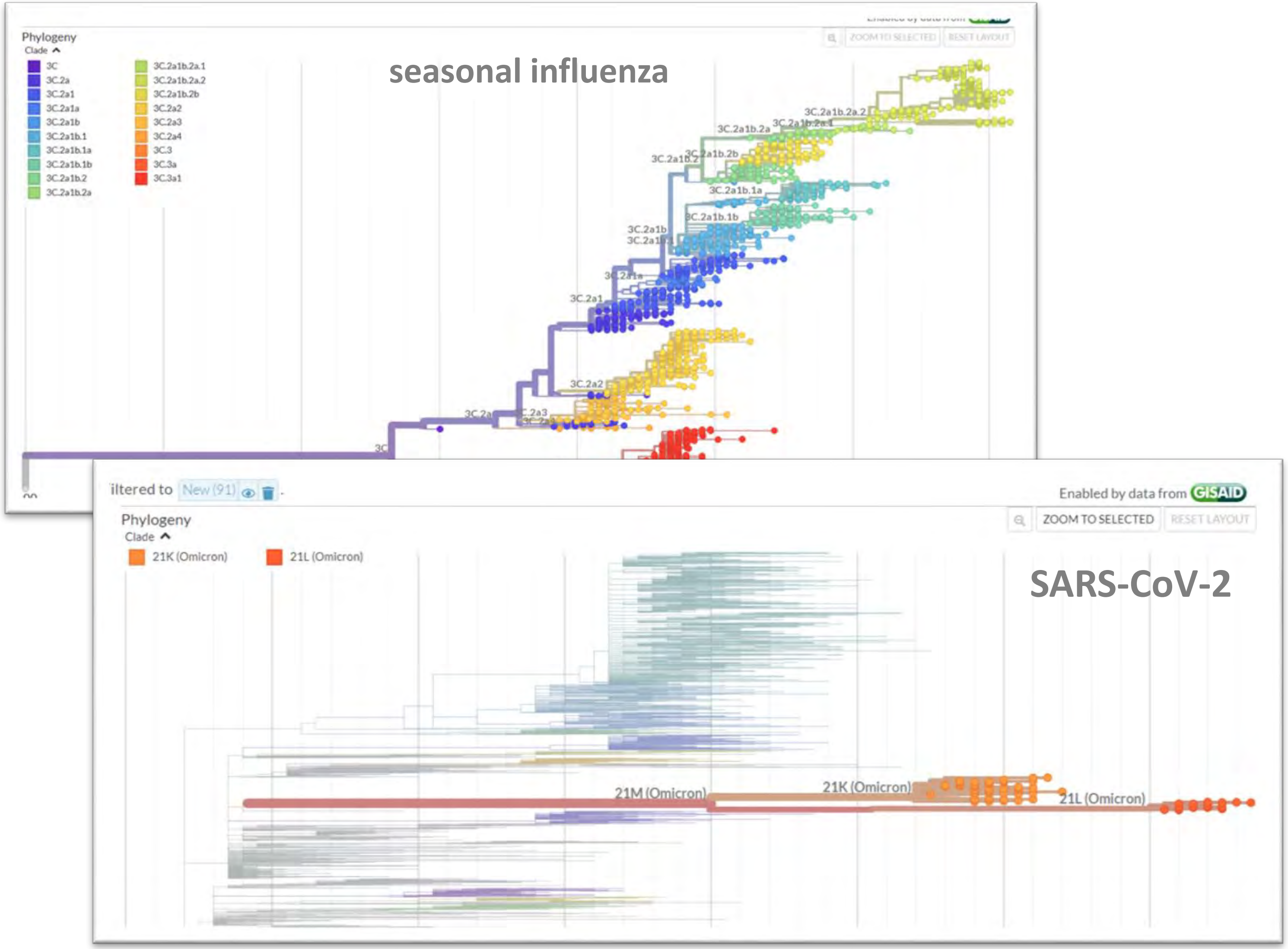
Nextclade

Clade assignment, mutation calling, and sequence quality checks

https://clades.nextstrain.org/



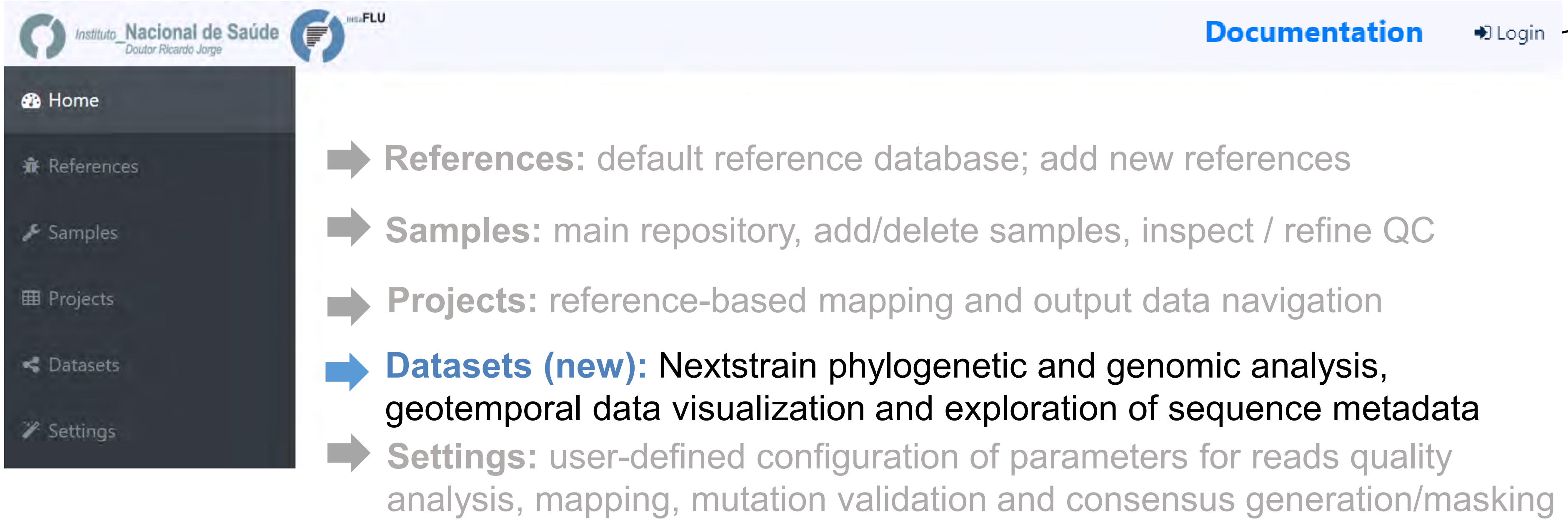
Lineage (Pangolin)	Seq Center	collection date
B.1.1.7	IB	2021-04-16
B.1.1.7	IB	2021-04-18
B.1.1.7	IB	2021-04-08
B.1.1.7	IB	2021-04-10
B.1.1.7	IB	2021-04-15
B.1.1.7	IB	2021-04-15
B.1.1.7	IB	2021-04-11
B.1.1.318	IB	2021-04-16
B.1.1.318	IB	2021-04-08
B.1.1	IB	2021-04-16
B.1.1.318	IB	2021-04-08
P.1	F	2021-04-10
P.1	F	2021-04-12
P.1	F	2021-04-09
P.1	F	2021-04-09
B.1.351	F	2021-04-05
B.1.351	F	2021-04-06
B.1.351	F	2021-04-06
B.1.617.1	F	2021-04-06
No Data	No Data	No Data
B.1.1.7	IB	2021-04-14
B.1.1.7	IB	2021-04-15



seasonal influenza

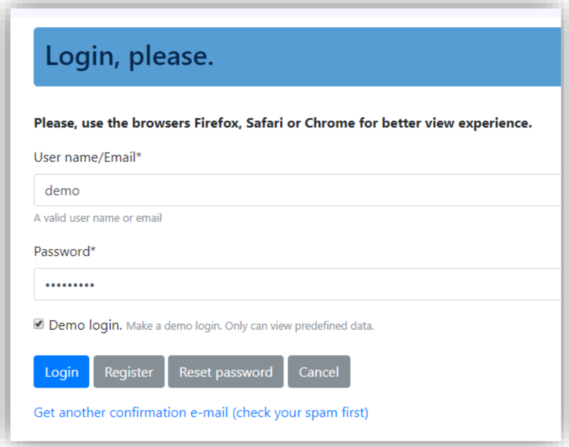
SARS-CoV-2

DATASETS – New module



- **References:** default reference database; add new references
- **Samples:** main repository, add/delete samples, inspect / refine QC
- **Projects:** reference-based mapping and output data navigation
- **Datasets (new):** Nextstrain phylogenetic and genomic analysis, geotemporal data visualization and exploration of sequence metadata
- **Settings:** user-defined configuration of parameters for reads quality analysis, mapping, mutation validation and consensus generation/masking

Register and **Login**
to a user restricted account



Login, please.

Please, use the browsers Firefox, Safari or Chrome for better view experience.

User name/Email*: demo

A valid user name or email

Password*: •••••••••

☑ Demo login. Make a demo login. Only can view predefined data.

Get another confirmation e-mail (check your spam first)

Applicable to:

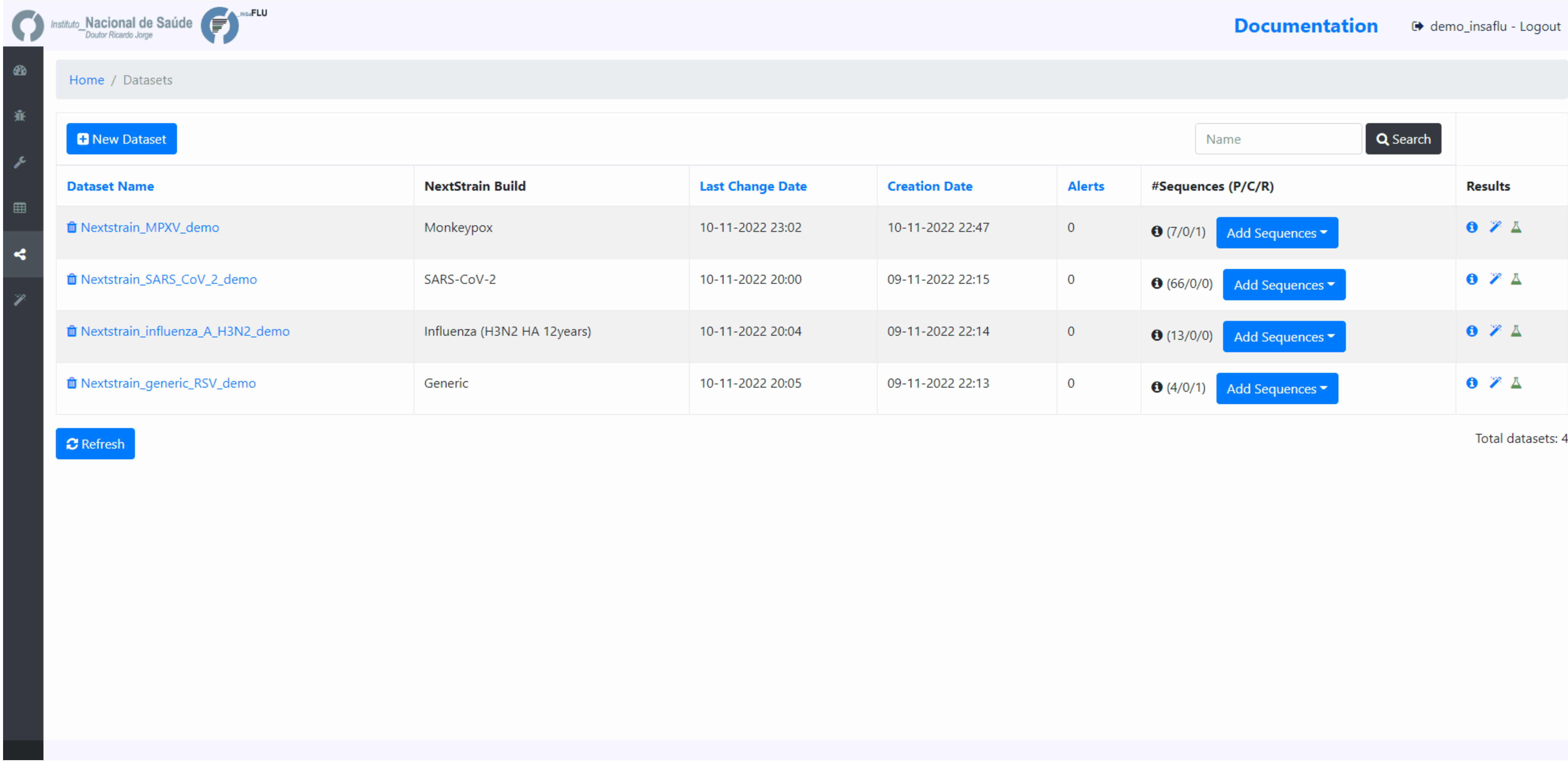
illumina

iontorrent

Oxford NANOPORE Technologies

INSaFLU

https://insaflu.insa.pt



Dataset Name	NextStrain Build	Last Change Date	Creation Date	Alerts	#Sequences (P/C/R)	Results
Nextstrain_MPXV_demo	Monkeypox	10-11-2022 23:02	10-11-2022 22:47	0	(7/0/1) Add Sequences	
Nextstrain_SARS_CoV_2_demo	SARS-CoV-2	10-11-2022 20:00	09-11-2022 22:15	0	(66/0/0) Add Sequences	
Nextstrain_influenza_A_H3N2_demo	Influenza (H3N2 HA 12years)	10-11-2022 20:04	09-11-2022 22:14	0	(13/0/0) Add Sequences	
Nextstrain_generic_RSV_demo	Generic	10-11-2022 20:05	09-11-2022 22:13	0	(4/0/1) Add Sequences	

Total datasets: 4

Available builds: Seasonal Influenza, SARS-CoV-2, Monkeypox and Generic

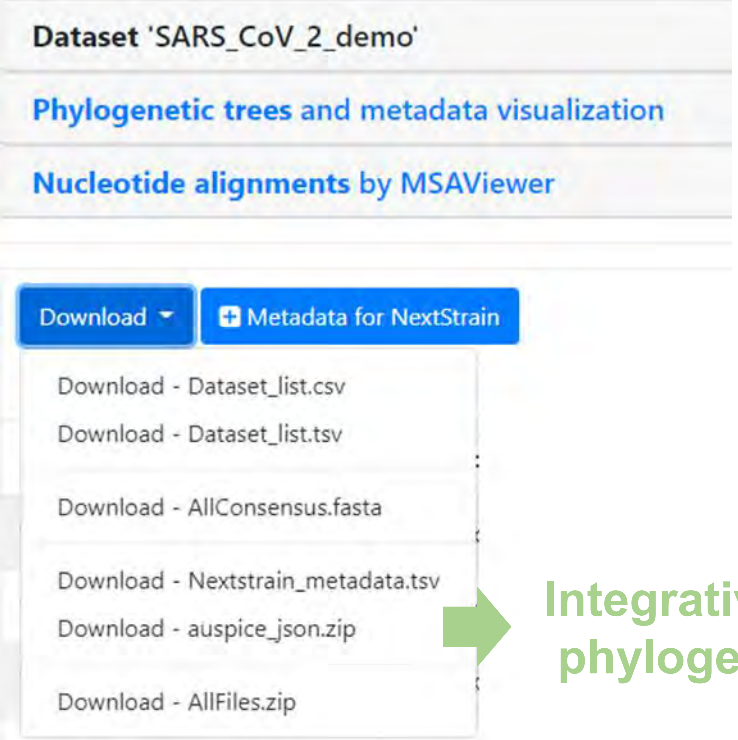
https://insaflu.insa.pt

Examples of “Datasets” outputs

https://insaflu.insa.pt

Nextstrain (https://nextstrain.org/) phylogenetic and geotemporal analysis.

Powered by Nextstrain



Dataset 'SARS_CoV_2_demo'

Phylogenetic trees and metadata visualization

Nucleotide alignments by MSAViewer

Download - Dataset_list.csv

Download - Dataset_list.tsv

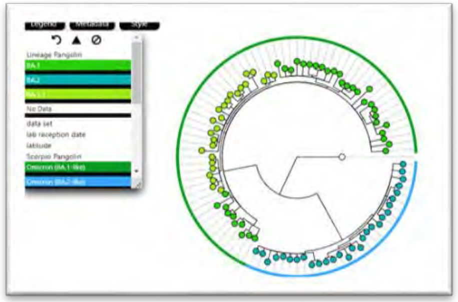
Download - AllConsensus.fasta

Download - Nextstrain_metadata.tsv

Download - auspice_json.zip

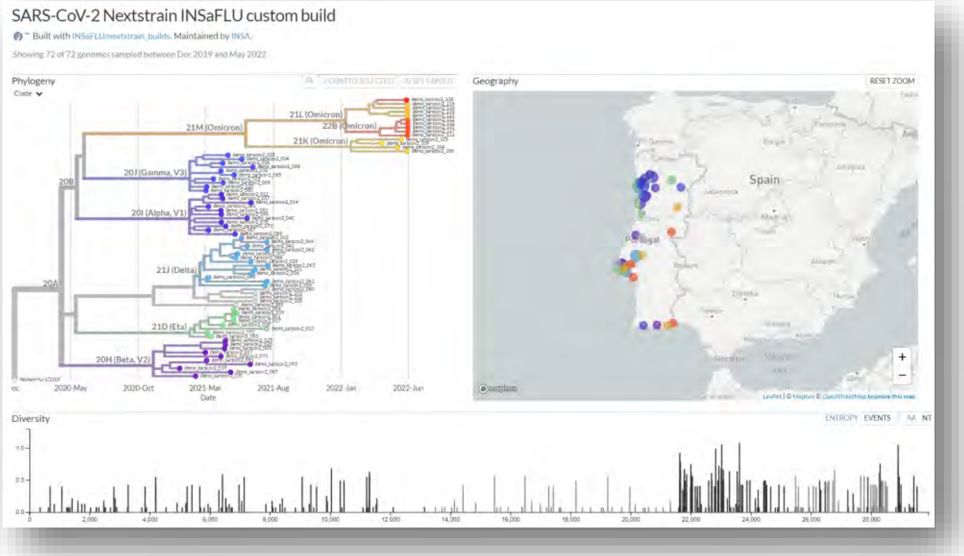
Download - AllFiles.zip

Interactive exploration of Nextstrain alignments and phylogenetic trees using PhyloCanvas



Integrative and interactive Nextstrain phylogenetic and geotemporal analysis.

DRAG&DROP onto https://auspice.us/



SARS-CoV-2 Nextstrain INSaFLU custom build

See more details in INSaFLU documentation: https://insaflu.readthedocs.io/en/latest/data_analysis.html#nextstrain-datasets and https://insaflu.readthedocs.io/en/latest/output_visualization.html#navigate-through-nextstrain-datasets and https://github.com/INSaFLU/nextstrain_builds

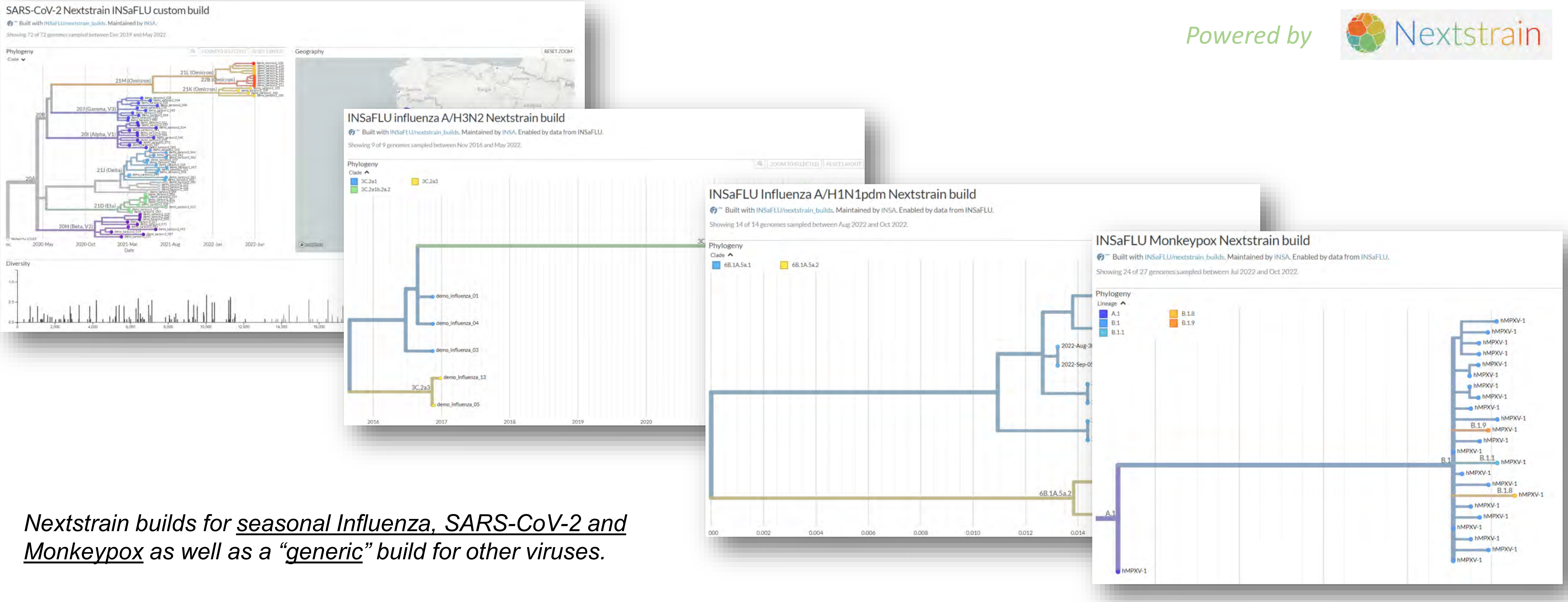
Examples of “Datasets” outputs

INSaFLU

https://insaflu.insa.pt

Nextstrain ([https://nextstrain.org/)](https://nextstrain.org/) phylogenetic and geotemporal analysis.

Powered by Nextstrain



SARS-CoV-2 Nextstrain INSaFLU custom build

Built with INSaFLU/nextstrain_builds. Maintained by INSA.

Showing 72 of 72 genomes sampled between Dec 2019 and May 2022.

INSaFLU influenza A/H3N2 Nextstrain build

Built with INSaFLU/nextstrain_builds. Maintained by INSA. Enabled by data from INSaFLU.

Showing 9 of 9 genomes sampled between Nov 2016 and May 2022.

INSaFLU Influenza A/H1N1pdm Nextstrain build

Built with INSaFLU/nextstrain_builds. Maintained by INSA. Enabled by data from INSaFLU.

Showing 14 of 14 genomes sampled between Aug 2022 and Oct 2022.

INSaFLU Monkeypox Nextstrain build

Built with INSaFLU/nextstrain_builds. Maintained by INSA. Enabled by data from INSaFLU.

Showing 24 of 27 genomes sampled between Jul 2022 and Oct 2022.

Nextstrain builds for <u>seasonal Influenza, SARS-CoV-2 and Monkeypox</u> as well as a “<u>generic</u>” build for other viruses.

See more details in INSaFLU documentation: https://insaflu.readthedocs.io/en/latest/data_analysis.html#nextstrain-datasets and https://insaflu.readthedocs.io/en/latest/output_visualization.html#navigate-through-nextstrain-datasets and https://github.com/INSaFLU/nextstrain_builds

INSaFLU-TELEVIR has been continuously updated

- Firstly tailored for **influenza surveillance**
- **SARS-CoV-2 –** adapted at beginning of pandemic
- **2022 Monkeypox virus outbreak**

Virus-specific routine genomic surveillance

New challenges - viral metagenomics

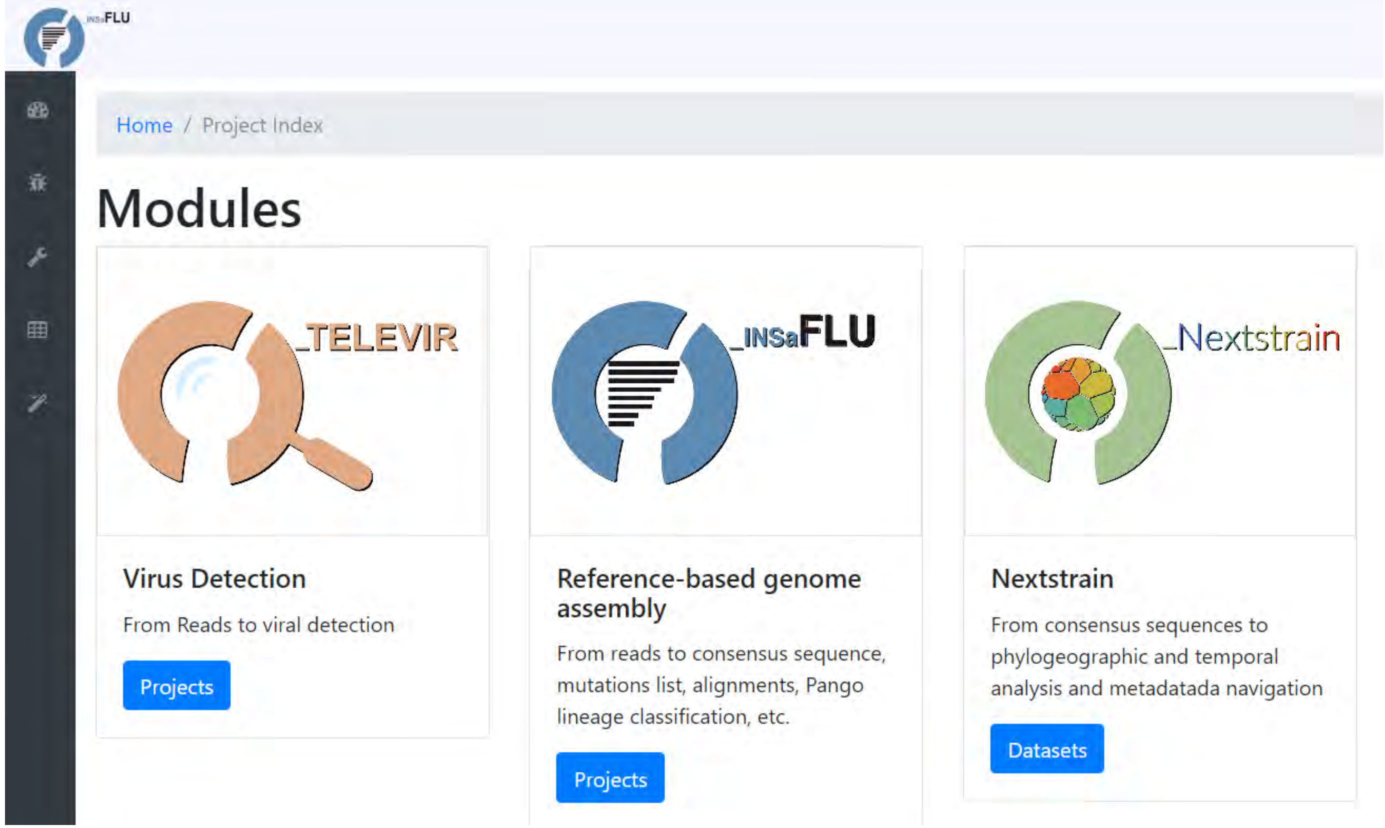
- **Virus detection in complex biological samples**
- **Crucial for early virus detection** (especially, novel threats)
- Bioinformatics analysis is a major obstacle/need in the field

Virus detection in complex samples

SOURCE: https://www.insideprecisionmedicine.com/magazine-editions/volume-6-issue-number-4-july-august-2019/mining-the-microbiome-microbiome-startups-are-proliferating-exploring-new-markets-and-colonizing-diverse-therapeutic-niches/

TELEVIR Module - Project TELEVIR (OneHealth EJP)

- Implementation of a new module for virus detection from reads



Home / Project Index

Modules

Virus Detection

From Reads to viral detection

Reference-based genome assembly

From reads to consensus sequence, mutations list, alignments, Pango lineage classification, etc.

Nextstrain

From consensus sequences to phylogeographic and temporal analysis and metadatada navigation

New virus detection module

Genome-based surveillance-oriented modules

https://insaflu.insa.pt

INSaFLU-TELEVIR: an open web-based bioinformatics suite for **metagenomic virus detection** and routine genomic surveillance (influenza, SARS-CoV-2 and other pathogens)

New virus detection module SOON !

INPUTS

- ✓ **Sample metadata table**

+

- ✓ **NGS data ***

**Compatible with distinct sequencing technologies:*
Illumina, Ion Torrent and Oxford Nanopore

Description	Taxid	accID	Cov (%)	Depth	DepthC	Mapped reads	ref_prop (%)	mapped_prop (%)	Gaps	success
Bubaline alphaherpesvirus 1 strain b6, complete genome	202910	NC_043054_1	89.83	35.84	40.3	1271	4.51077	4.51109	159	1

Database: refseq_viral.genome.fna NCBI length: 137452 contig string: NC_043054.1

OUTPUTS

VIRUS DETECTION *(TELEVIR module)*

- ✓ Read quality analysis and improvement

- ✓ **State of the art software**
- ✓ Modular Pipeline

- ✓ **Multiple Classification methods**
- ✓ **Multiple Viral databases**

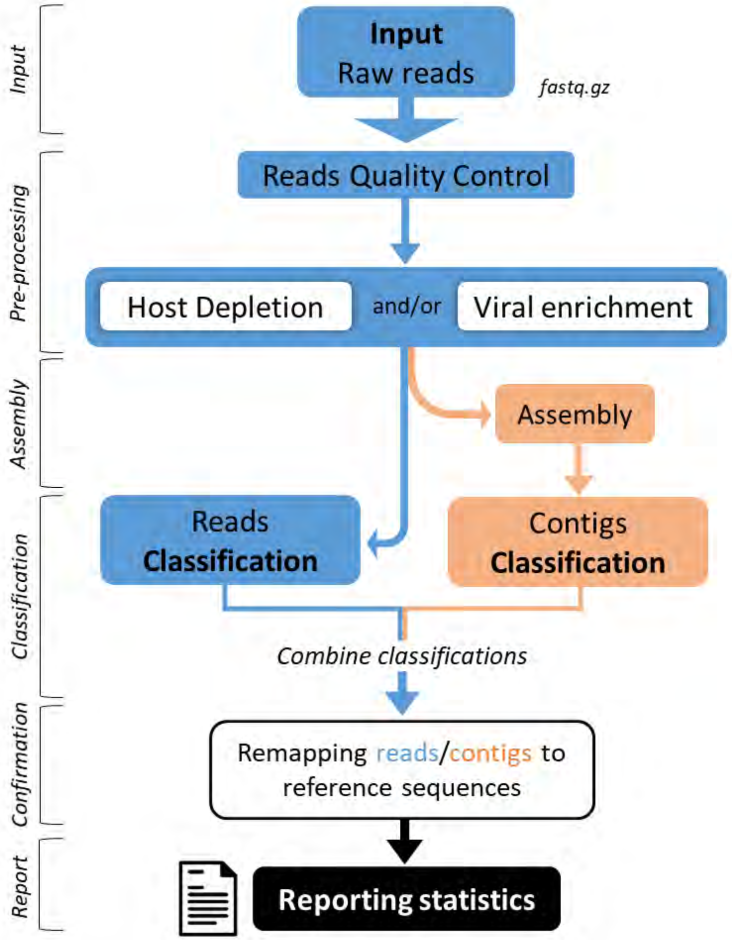
- ✓ **Summary statistics and intuitive and interactive end-user reports**

INSaFLU-TELEVIR development is being co-funded by the European Commission on behalf of OneHealth EJP

TELE-Vir project: https://onehealthejp.eu/jrp-tele-vir/

TELEVIR Module - Project TELEVIR (OneHealth EJP)

Virus detection from reads



Remove reads from host

Keep the viral reads in the sample

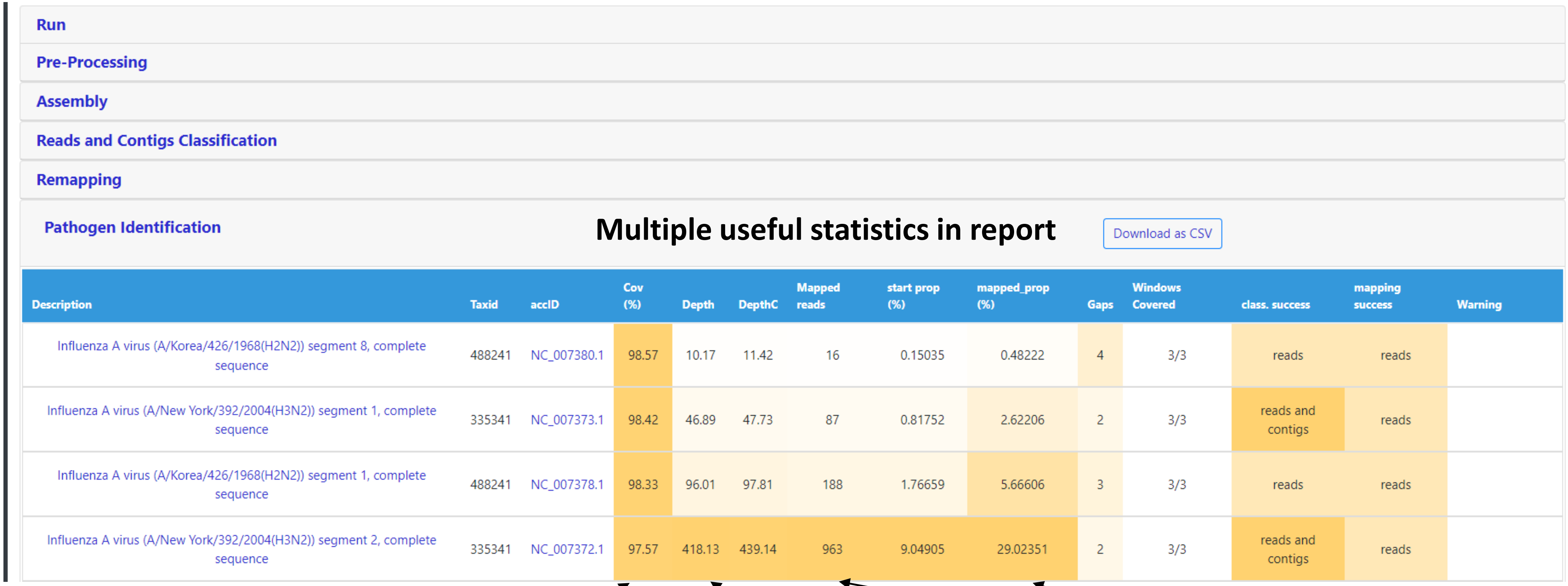
Assembly; taxonomic classification of reads and contigs

Reads/contigs are remapped against reference genomes of the identified taxa

Summary statistics of the identified hits and intuitive end-user reports

TELEVIR Module - Project TELEVIR (OneHealth EJP)

Virus detection from reads



Run

Pre-Processing

Assembly

Reads and Contigs Classification

Remapping

Pathogen Identification

Multiple useful statistics in report

Description	Taxid	accID	Cov (%)	Depth	DepthC	Mapped reads	start prop (%)	mapped_prop (%)	Gaps	Windows Covered	class. success	mapping success	Warning
Influenza A virus (A/Korea/426/1968(H2N2)) segment 8, complete sequence	488241	NC_007380.1	98.57	10.17	11.42	16	0.15035	0.48222	4	3/3	reads	reads	
Influenza A virus (A/New York/392/2004(H3N2)) segment 1, complete sequence	335341	NC_007373.1	98.42	46.89	47.73	87	0.81752	2.62206	2	3/3	reads and contigs	reads	
Influenza A virus (A/Korea/426/1968(H2N2)) segment 1, complete sequence	488241	NC_007378.1	98.33	96.01	97.81	188	1.76659	5.66606	3	3/3	reads	reads	
Influenza A virus (A/New York/392/2004(H3N2)) segment 2, complete sequence	335341	NC_007372.1	97.57	418.13	439.14	963	9.04905	29.02351	2	3/3	reads and contigs	reads	

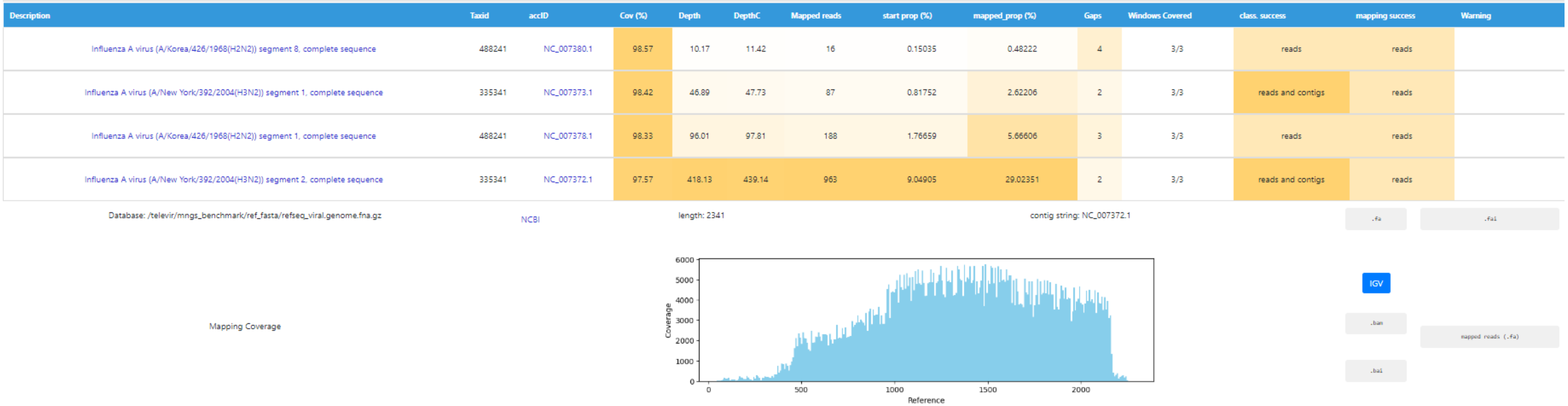
Horizontal coverage

Depth of coverage

N and % of reads mapped

TELEVIR Module - Project TELEVIR (OneHealth EJP)

Virus detection from reads



Description	Taxid	accID	Cov (%)	Depth	DepthC	Mapped reads	start prop (%)	mapped_prop (%)	Gaps	Windows Covered	class. success	mapping success	Warning
Influenza A virus (A/Korea/426/1968(H2N2)) segment 8, complete sequence	488241	NC_007380.1	98.57	10.17	11.42	16	0.15035	0.48222	4	3/3	reads	reads	
Influenza A virus (A/New York/392/2004(H3N2)) segment 1, complete sequence	335341	NC_007373.1	98.42	46.89	47.73	87	0.81752	2.62206	2	3/3	reads and contigs	reads	
Influenza A virus (A/Korea/426/1968(H2N2)) segment 1, complete sequence	488241	NC_007378.1	98.33	96.01	97.81	188	1.76659	5.66606	3	3/3	reads	reads	
Influenza A virus (A/New York/392/2004(H3N2)) segment 2, complete sequence	335341	NC_007372.1	97.57	418.13	439.14	963	9.04905	29.02351	2	3/3	reads and contigs	reads	

Database: /televir/mngs_benchmark/ref_fasta/refseq_viral.genome.fna.gz

NCBI

length: 2341

contig string: NC_007372.1

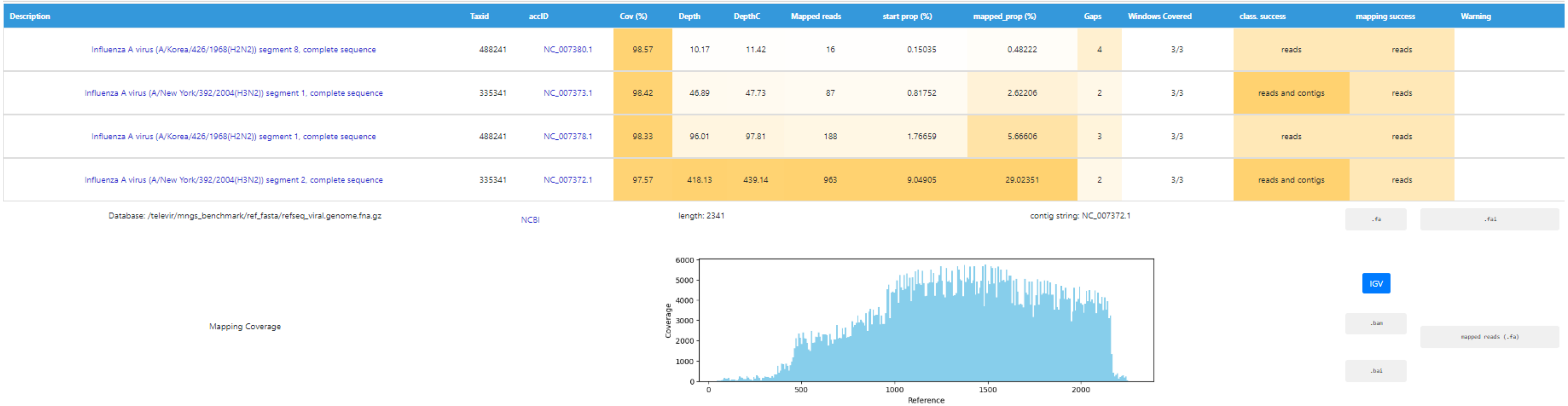
Mapping Coverage

Each hit is an "expand-and-colapse" panel that can be further vizualized and explored

- Vizualization of coverage throughout reference genome
- Mapped reads and contigs can be downloaded for further confirmations

TELEVIR Module - Project TELEVIR (OneHealth EJP)

Virus detection from reads



Description	Taxid	accID	Cov (%)	Depth	DepthC	Mapped reads	start prop (%)	mapped_prop (%)	Gaps	Windows Covered	class. success	mapping success	Warning
Influenza A virus (A/Korea/426/1968(H2N2)) segment 8, complete sequence	488241	NC_007380.1	98.57	10.17	11.42	16	0.15035	0.48222	4	3/3	reads	reads	
Influenza A virus (A/New York/392/2004(H3N2)) segment 1, complete sequence	335341	NC_007373.1	98.42	46.89	47.73	87	0.81752	2.62206	2	3/3	reads and contigs	reads	
Influenza A virus (A/Korea/426/1968(H2N2)) segment 1, complete sequence	488241	NC_007378.1	98.33	96.01	97.81	188	1.76659	5.66606	3	3/3	reads	reads	
Influenza A virus (A/New York/392/2004(H3N2)) segment 2, complete sequence	335341	NC_007372.1	97.57	418.13	439.14	963	9.04905	29.02351	2	3/3	reads and contigs	reads	

Database: /televir/mngs_benchmark/ref_fasta/refseq_viral.genome.fna.gz

length: 2341

contig string: NC_007372.1

For further information from viral genome (if there is sufficient coverage), the user can **analyze the sample with the "surveillance-oriented" modules in INSaFLU-TELEVIR** (reference-based mapping, phylogenetic trees, …)

Acknowledgements - INSaFLU-TELEVIR platform development

Instituto Nacional de Saúde
Doutor Ricardo Jorge

Bioinformatics Unit

Vítor Borges
Daniel Sobral
João Santos
Miguel Pinto
Rita Ferreira
Luis Coelho
Verónica Mixão
Ricardo Pais – former Pos-Doc
Alexandra Nunes
João Paulo Gomes

External partners

Miguel Pinheiro (iBiMED / Universidade de Aveiro)

(co-developer)

OneHealth EJP TELE-Vir project partners

Infraestrutura Nacional de Computação Distribuída (INCD), for providing computational resources for testing

National Reference Laboratory (NRL) for Influenza

Raquel Guiomar and collaborators

INSaFLU development is being co-funded by the European Commission on behalf of OneHealth EJP

TELE-Vir project: https://onehealthejp.eu/jrp-tele-vir/