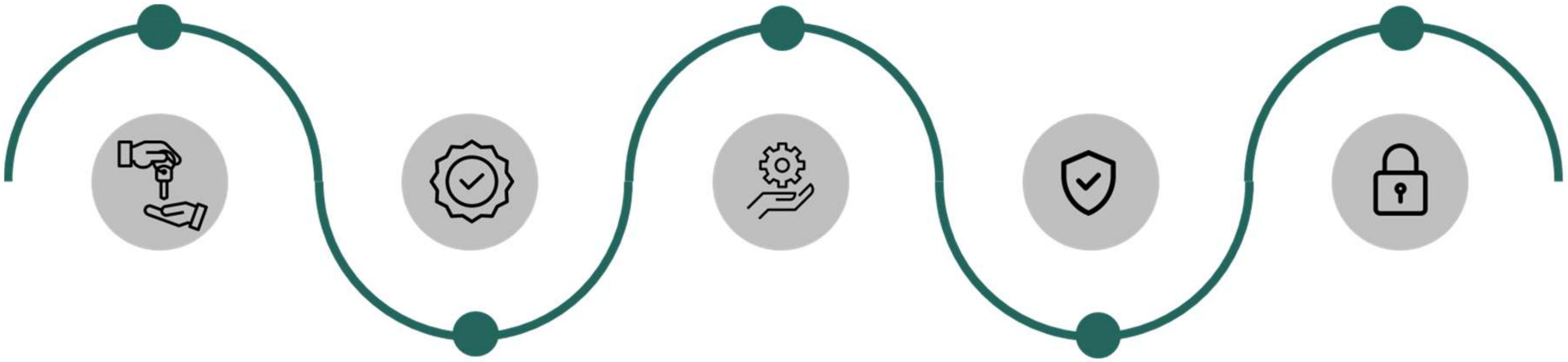
Enabling Insights from Mpox data

Dr. Sebastian Maurer-Stroh

GISAID Data Science Centre, Singapore

GISAID – Trust is respecting submitters' rights



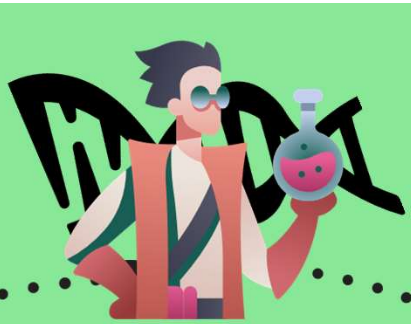
Ownership
Data submitters retain ownership of their data and intellectual property

Quality
GISAID curation team works with submitters to ensure highest standard of quality for genomic data and metadata

Stewardship
Trust built over a decade of partnerships.
GISAID provides consistent reliability and accountability for data management and service

Security
Protecting digital content and infrastructure
Reliability of systems

Privacy
GISAID commits to data privacy for users and submitters.
GISAID strives to ensure case patient anonymity through robust checks, policies and systems.



Alchemist

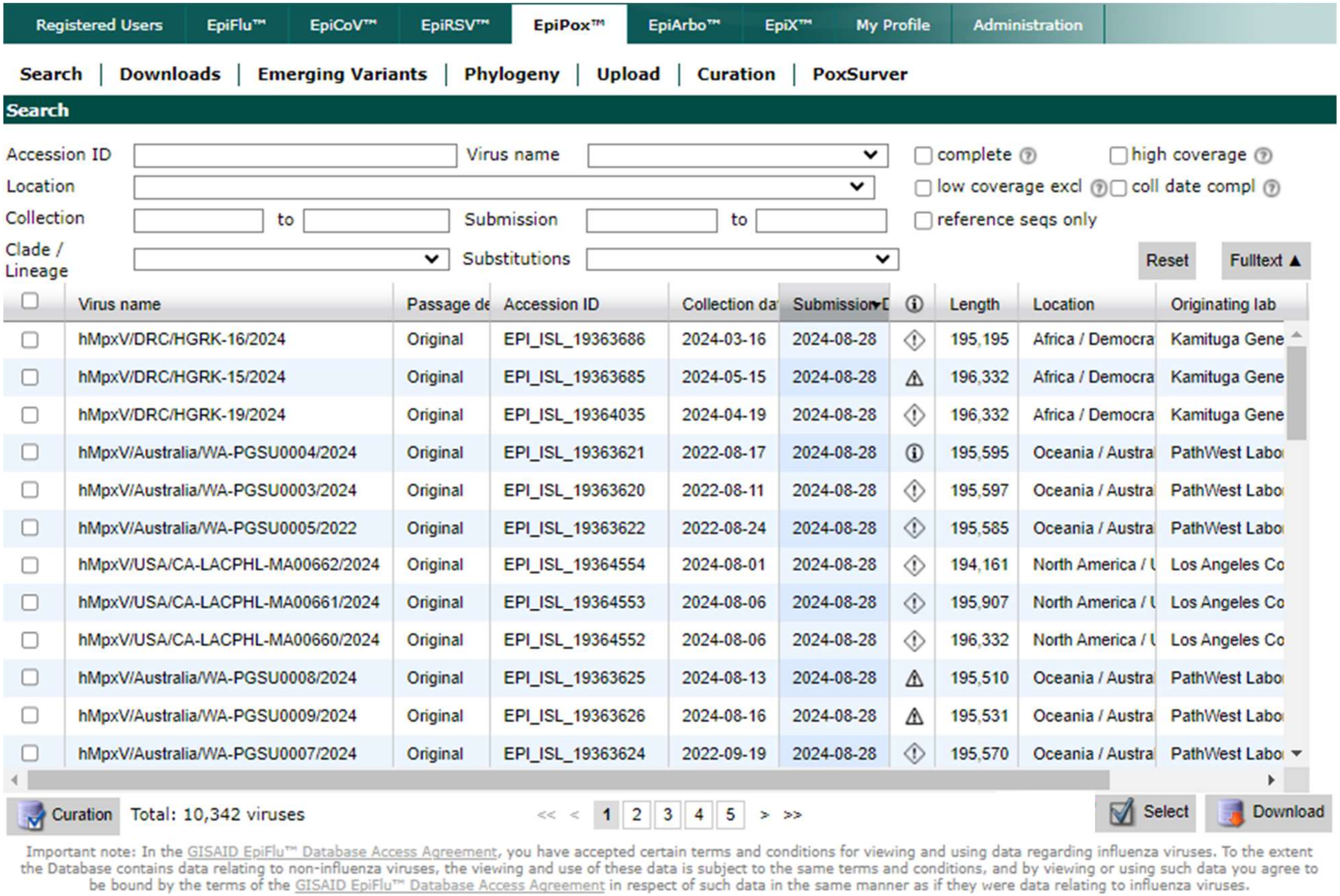
INTELLIGENT ASSEMBLY

human-reads-removed genome data

... so you can focus on more
important things

Upload raw sequencing reads (FASTQ)

Consensus sequences (FASTA) on GISAID



Search

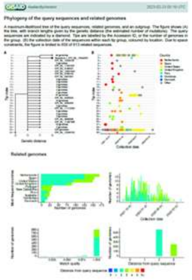
Accession ID ____ Virus name ____ ☐ complete ☐ high coverage
Location ____ ☐ low coverage excl ☐ coll date compl
Collection ____ to ____ Submission ____ to ____ ☐ reference seqs only
Clade / Lineage ____ Substitutions ____

☐	Virus name	Passage de	Accession ID	Collection da	Submission▾	ⓘ	Length	Location	Originating lab
☐	hMpxV/DRC/HGRK-16/2024	Original	EPI_ISL_19363686	2024-03-16	2024-08-28	◇	195,195	Africa / Democra	Kamituga Gene
☐	hMpxV/DRC/HGRK-15/2024	Original	EPI_ISL_19363685	2024-05-15	2024-08-28	⚠	196,332	Africa / Democra	Kamituga Gene
☐	hMpxV/DRC/HGRK-19/2024	Original	EPI_ISL_19364035	2024-04-19	2024-08-28	◇	196,332	Africa / Democra	Kamituga Gene
☐	hMpxV/Australia/WA-PGSU0004/2024	Original	EPI_ISL_19363621	2022-08-17	2024-08-28	ⓘ	195,595	Oceania / Austra	PathWest Labo
☐	hMpxV/Australia/WA-PGSU0003/2024	Original	EPI_ISL_19363620	2022-08-11	2024-08-28	◇	195,597	Oceania / Austra	PathWest Labo
☐	hMpxV/Australia/WA-PGSU0005/2022	Original	EPI_ISL_19363622	2022-08-24	2024-08-28	◇	195,585	Oceania / Austra	PathWest Labo
☐	hMpxV/USA/CA-LACPHL-MA00662/2024	Original	EPI_ISL_19364554	2024-08-01	2024-08-28	◇	194,161	North America /	Los Angeles Co
☐	hMpxV/USA/CA-LACPHL-MA00661/2024	Original	EPI_ISL_19364553	2024-08-06	2024-08-28	◇	195,907	North America /	Los Angeles Co
☐	hMpxV/USA/CA-LACPHL-MA00660/2024	Original	EPI_ISL_19364552	2024-08-06	2024-08-28	◇	196,332	North America /	Los Angeles Co
☐	hMpxV/Australia/WA-PGSU0008/2024	Original	EPI_ISL_19363625	2024-08-13	2024-08-28	⚠	195,510	Oceania / Austra	PathWest Labo
☐	hMpxV/Australia/WA-PGSU0009/2024	Original	EPI_ISL_19363626	2024-08-16	2024-08-28	⚠	195,531	Oceania / Austra	PathWest Labo
☐	hMpxV/Australia/WA-PGSU0007/2024	Original	EPI_ISL_19363624	2022-09-19	2024-08-28	◇	195,570	Oceania / Austra	PathWest Labo

Total: 10,342 viruses

Important note: In the GISAID EpiFlu™ Database Access Agreement, you have accepted certain terms and conditions for viewing and using data regarding influenza viruses. To the extent the Database contains data relating to non-influenza viruses, the viewing and use of these data is subject to the same terms and conditions, and by viewing or using such data you agree to be bound by the terms of the GISAID EpiFlu™ Database Access Agreement in respect of such data in the same manner as if they were data relating to influenza viruses.

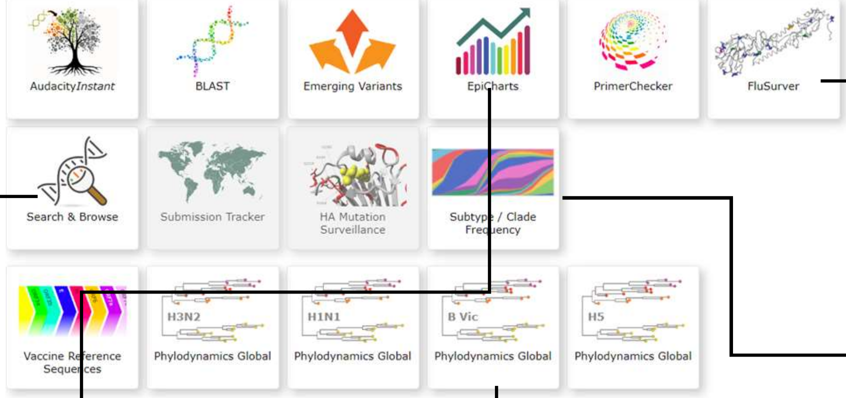
Empowering Actionable Data– Answering YOUR Questions



Question:
Which sequences in the database are most similar to my own sequence?

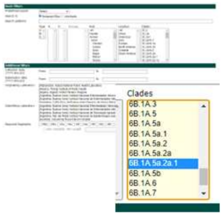
Question:
Any emerging variant with unique constellation of mutations?

Question:
Any mutation affecting common or my primers?

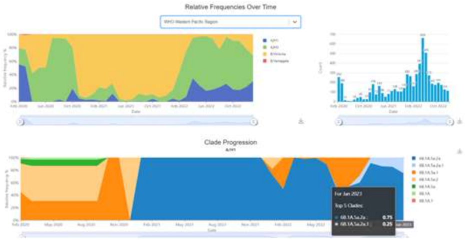


Question:
Any phenotypically important mutation in my sequences (host specificity, antibody binding, drug resistance, glycosylatio
virulence)?

Question:
Can I browse, search and filter entries in the database?

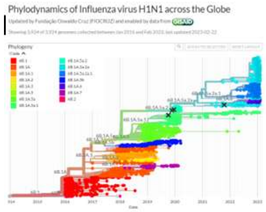


Question:
Which clades are currently dominant or rising in my country/region?



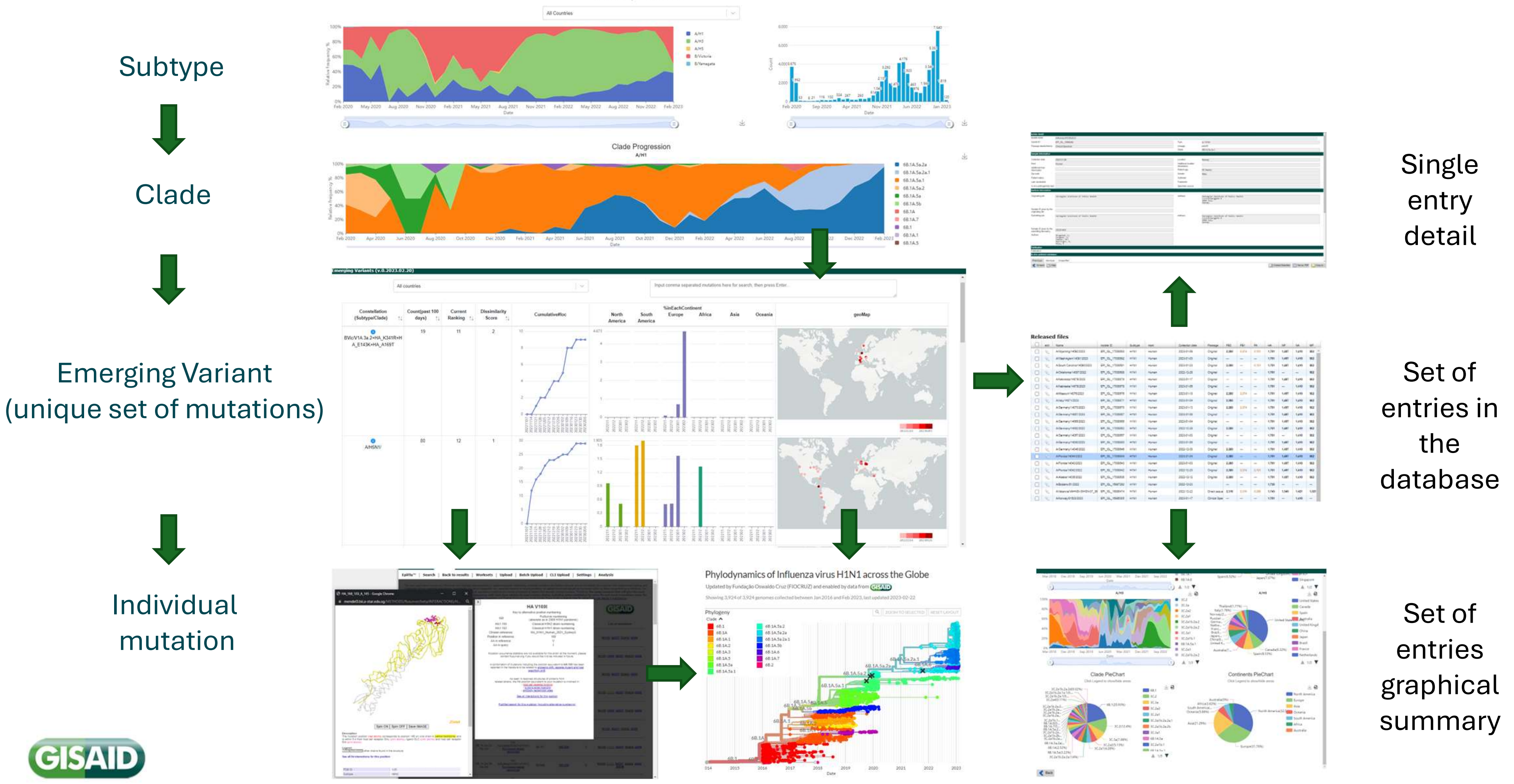
Question:
How can I create summary graphs for hits from a custom database search?

Question:
Summary of phylogenetic evolution in subsam



GISAID

An Ecosystem of integrated tools supporting scale-free view of virus evolution



Subtype

Clade

Emerging Variant
(unique set of mutations)

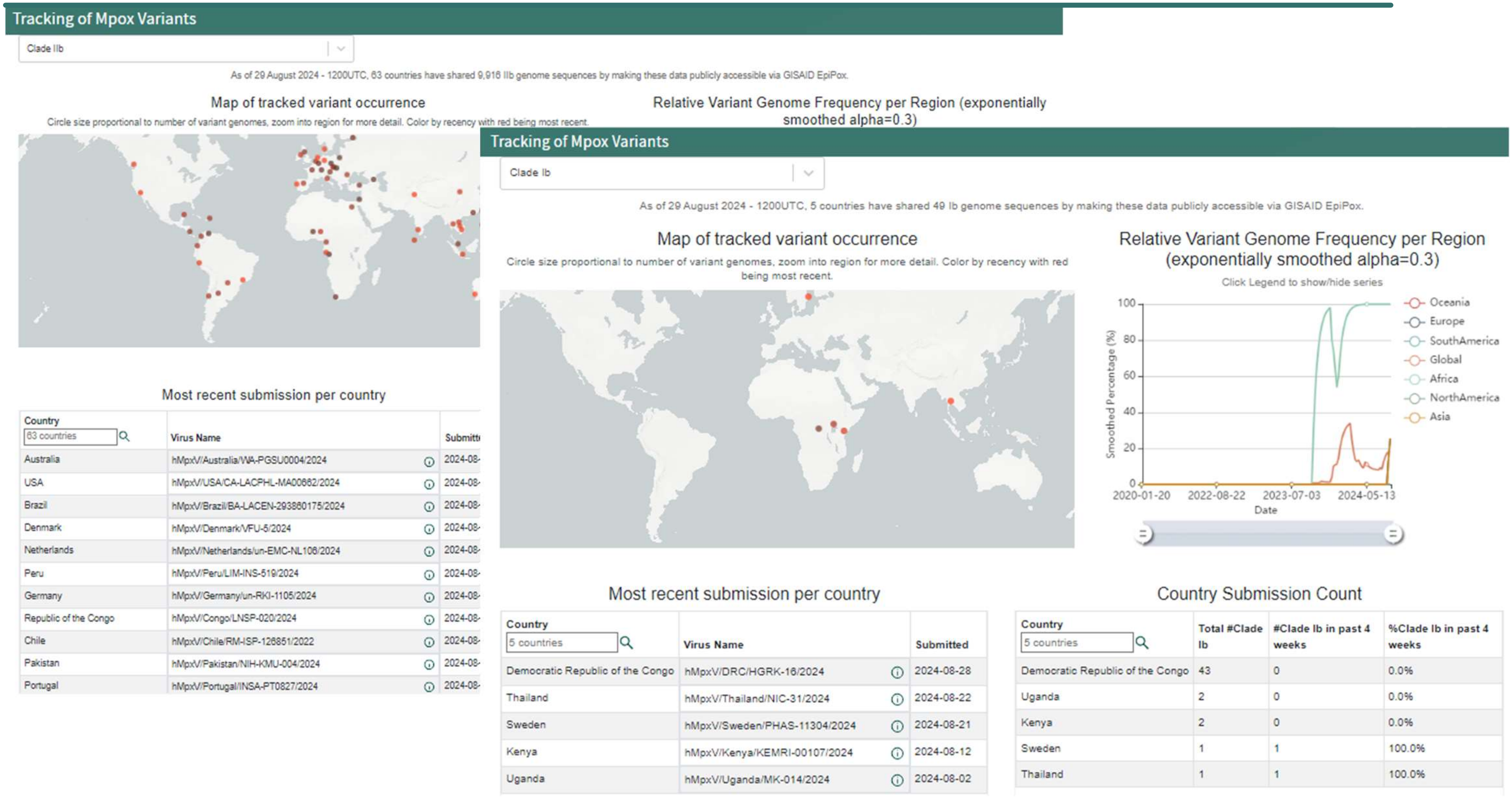
Individual
mutation

Single
entry
detail

Set of
entries in
the
database

Set of
entries
graphical
summary

EpiPox – Tracking Variants



Tracking of Mpox Variants

Clade IIb

As of 29 August 2024 - 1200UTC, 63 countries have shared 9,916 IIb genome sequences by making these data publicly accessible via GISAID EpiPox.

Map of tracked variant occurrence

Circle size proportional to number of variant genomes, zoom into region for more detail. Color by recency with red being most recent.

Relative Variant Genome Frequency per Region (exponentially smoothed alpha=0.3)

Most recent submission per country

Country 63 countries	Virus Name	Submitt
Australia	hMpxV/Australia/WA-PGSU0004/2024	2024-08-
USA	hMpxV/USA/CA-LACPHL-MA00662/2024	2024-08-
Brazil	hMpxV/Brazil/BA-LACEN-293860175/2024	2024-08-
Denmark	hMpxV/Denmark/VFU-5/2024	2024-08-
Netherlands	hMpxV/Netherlands/un-EMC-NL106/2024	2024-08-
Peru	hMpxV/Peru/LIM-INS-519/2024	2024-08-
Germany	hMpxV/Germany/un-RKI-1105/2024	2024-08-
Republic of the Congo	hMpxV/Congo/LNSP-020/2024	2024-08-
Chile	hMpxV/Chile/RM-ISP-126851/2022	2024-08-
Pakistan	hMpxV/Pakistan/NIH-KMU-004/2024	2024-08-
Portugal	hMpxV/Portugal/INSA-PT0827/2024	2024-08-

Tracking of Mpox Variants

Clade Ib

As of 29 August 2024 - 1200UTC, 5 countries have shared 49 Ib genome sequences by making these data publicly accessible via GISAID EpiPox.

Map of tracked variant occurrence

Circle size proportional to number of variant genomes, zoom into region for more detail. Color by recency with red being most recent.

Relative Variant Genome Frequency per Region (exponentially smoothed alpha=0.3)

Click Legend to show/hide series

Most recent submission per country

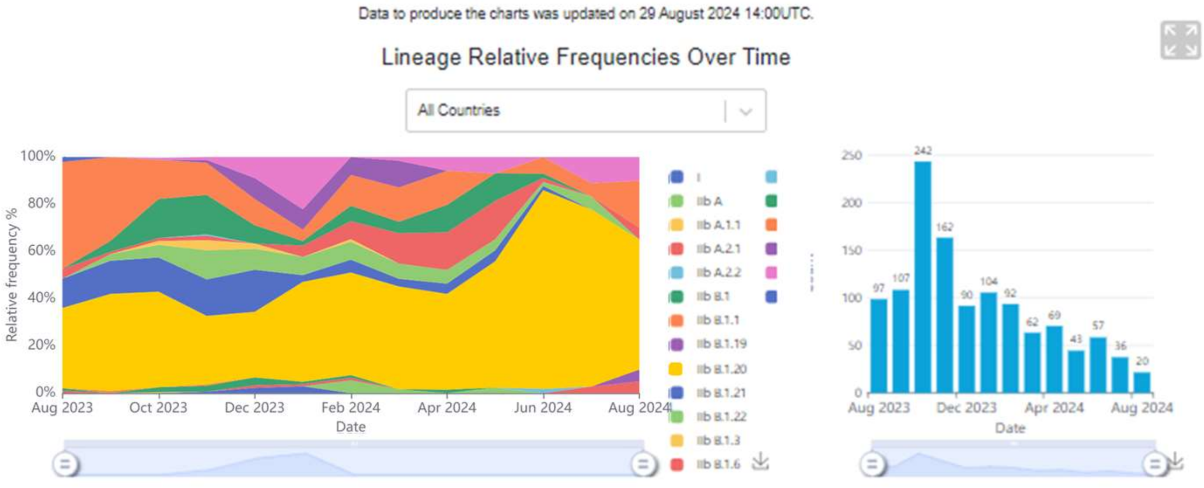
Country 5 countries	Virus Name	Submitted
Democratic Republic of the Congo	hMpxV/DRC/HGRK-16/2024	2024-08-28
Thailand	hMpxV/Thailand/NIC-31/2024	2024-08-22
Sweden	hMpxV/Sweden/PHAS-11304/2024	2024-08-21
Kenya	hMpxV/Kenya/KEMRI-00107/2024	2024-08-12
Uganda	hMpxV/Uganda/MK-014/2024	2024-08-02

Country Submission Count

Country 5 countries	Total #Clade Ib	#Clade Ib in past 4 weeks	%Clade Ib in past 4 weeks
Democratic Republic of the Congo	43	0	0.0%
Uganda	2	0	0.0%
Kenya	2	0	0.0%
Sweden	1	1	100.0%
Thailand	1	1	100.0%

Submission Volumes and Lineage Frequencies Over Time

Global

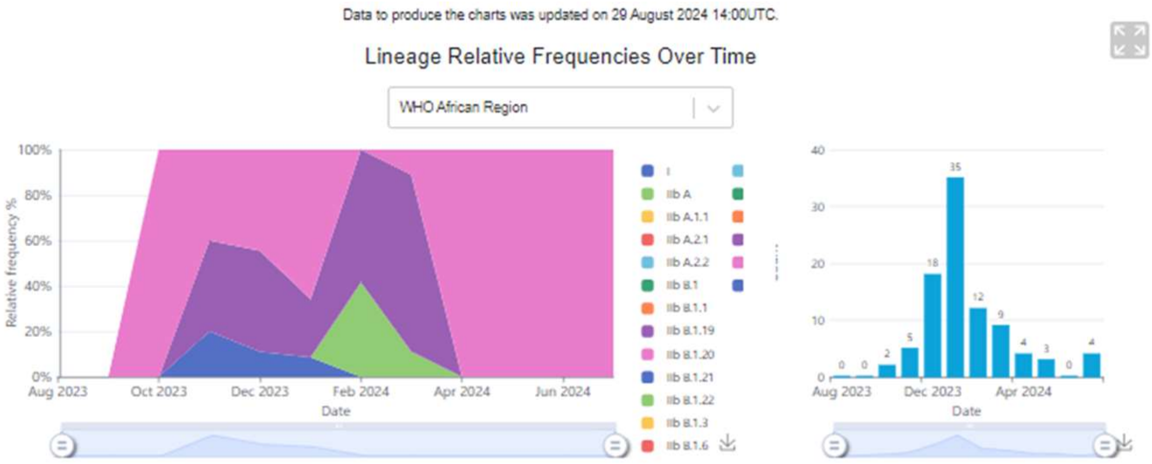


Data to produce the charts was updated on 29 August 2024 14:00UTC.

Lineage Relative Frequencies Over Time

All Countries

WHO AFRO

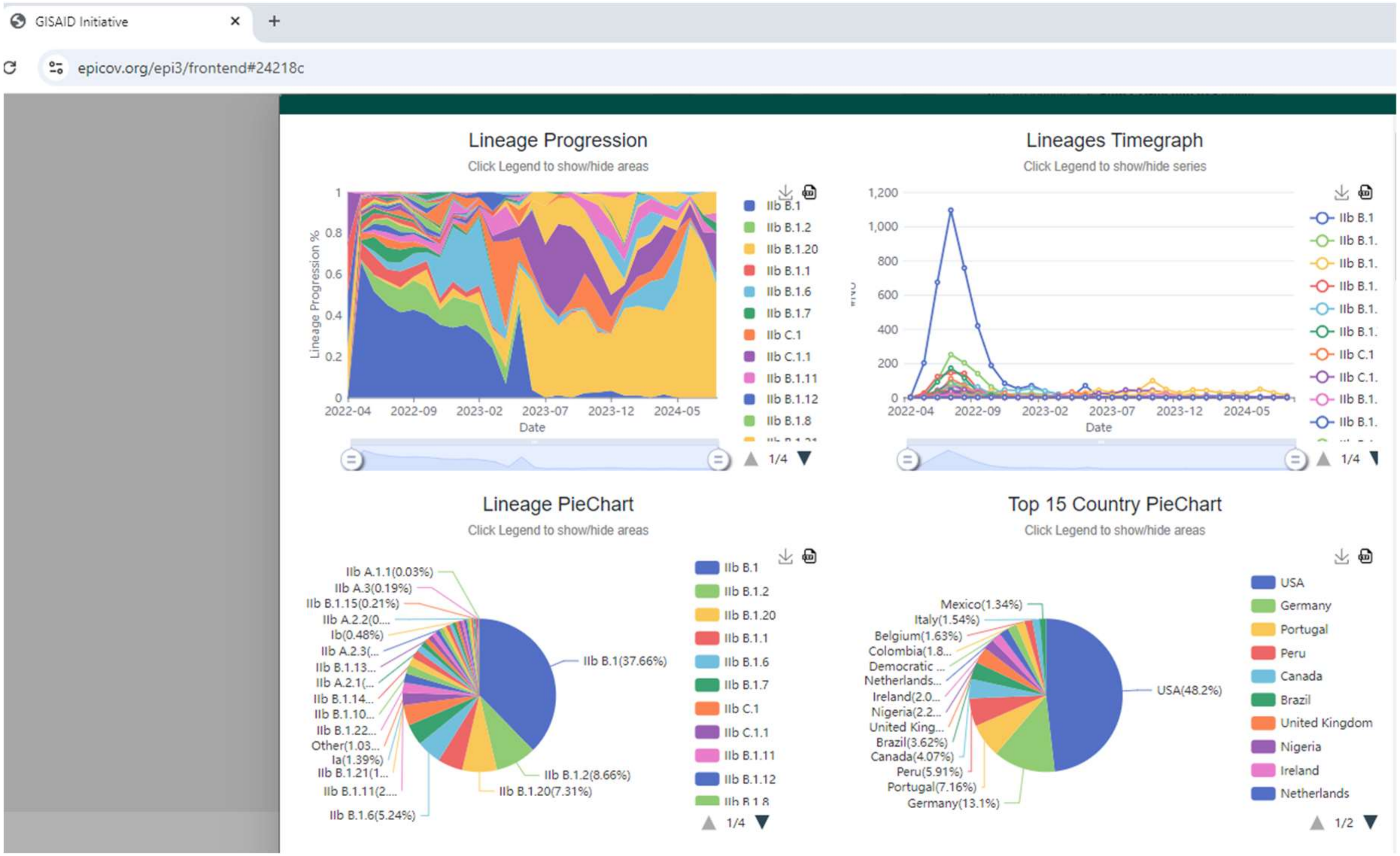


Data to produce the charts was updated on 29 August 2024 14:00UTC.

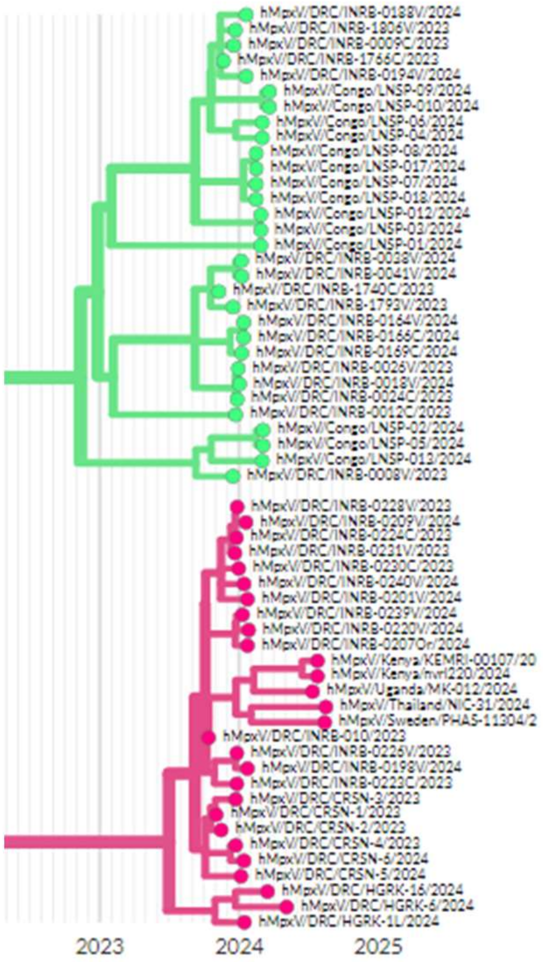
Lineage Relative Frequencies Over Time

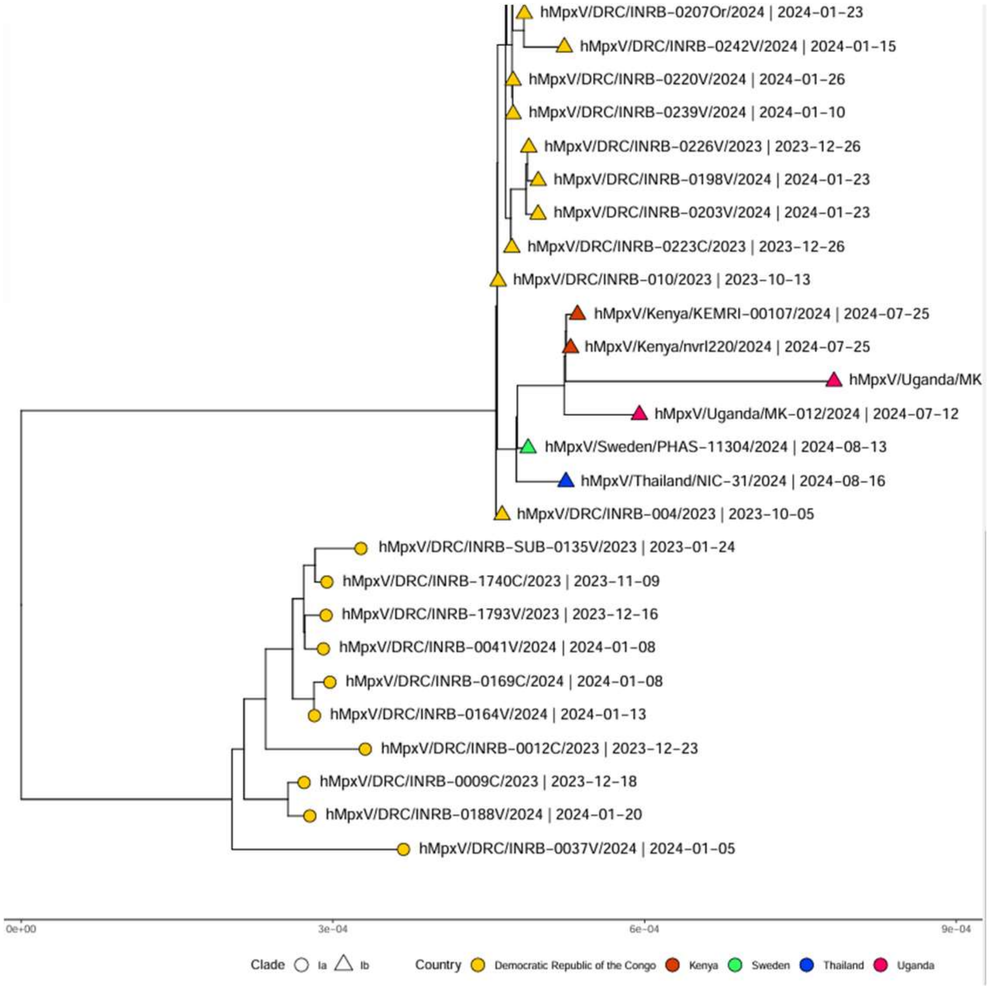
WHO African Region

EpiCharts – beta version to create custom charts

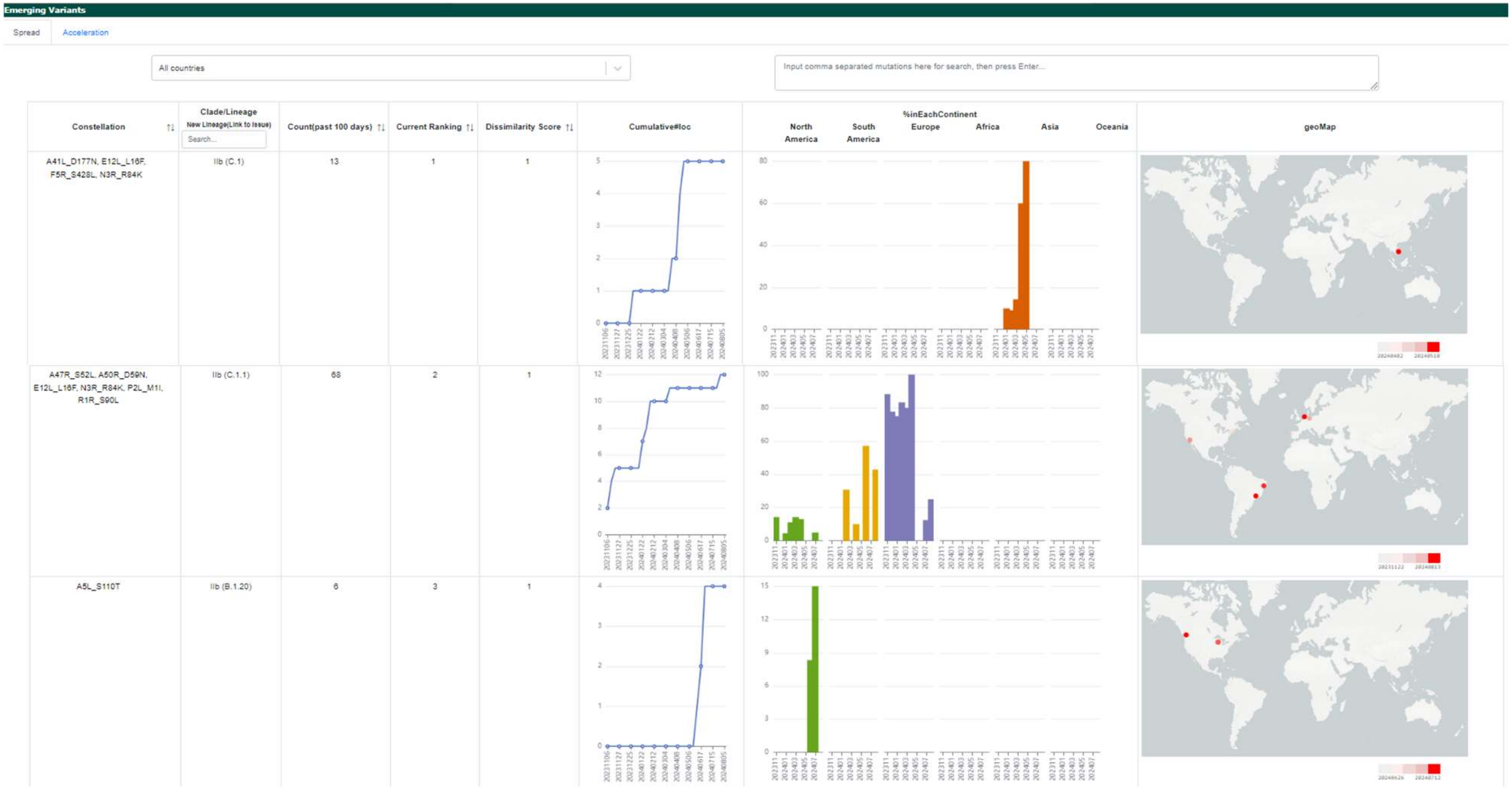


Phylogenetic tree Clade I





Tracking Emerging Variants

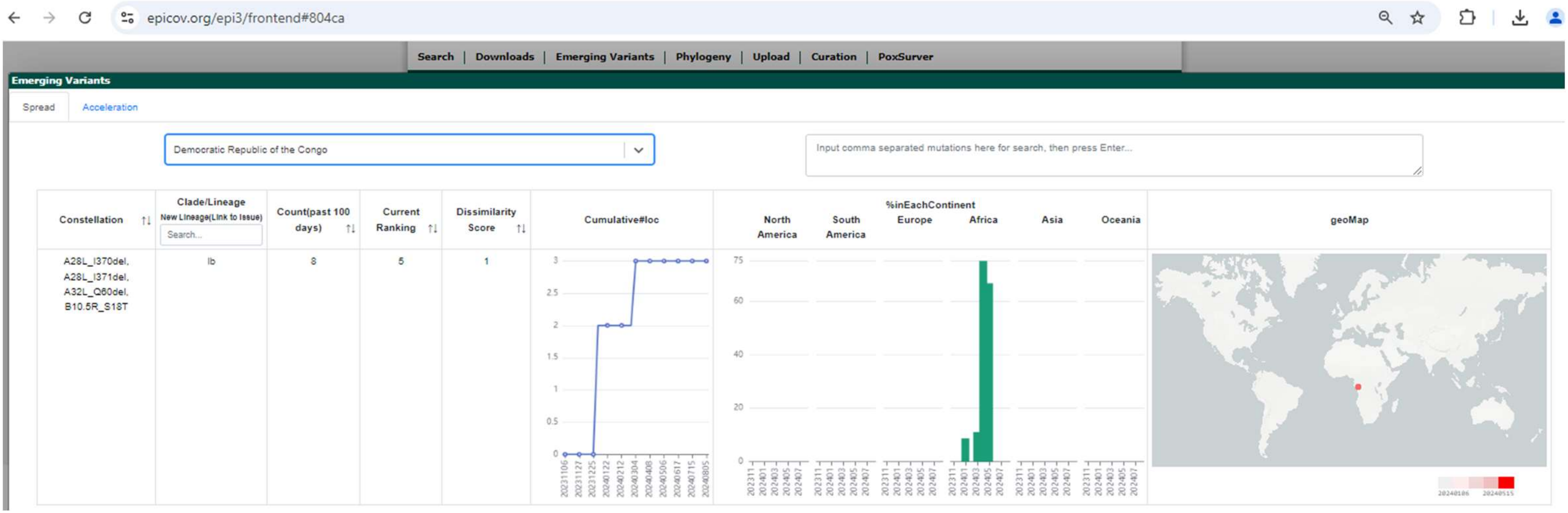


All countries

Input comma separated mutations here for search, then press Enter...

Constellation	Clade/Lineage New Lineage(Link to Issue)	Count(past 100 days)	Current Ranking	Dissimilarity Score	Cumulative#loc	%inEachContinent	geoMap
A41L_D177N, E12L_L16F, F5R_S428L, N3R_R84K	IIb (C.1)	13	1	1			
A47R_S52L, A50R_D59N, E12L_L16F, N3R_R84K, P2L_M1I, R1R_S90L	IIb (C.1.1)	68	2	1			
A5L_S110T	IIb (B.1.20)	6	3	1			

Tracking Emerging Variants



epicov.org/epi3/frontend#804ca

Search | Downloads | Emerging Variants | Phylogeny | Upload | Curation | PoxSurver

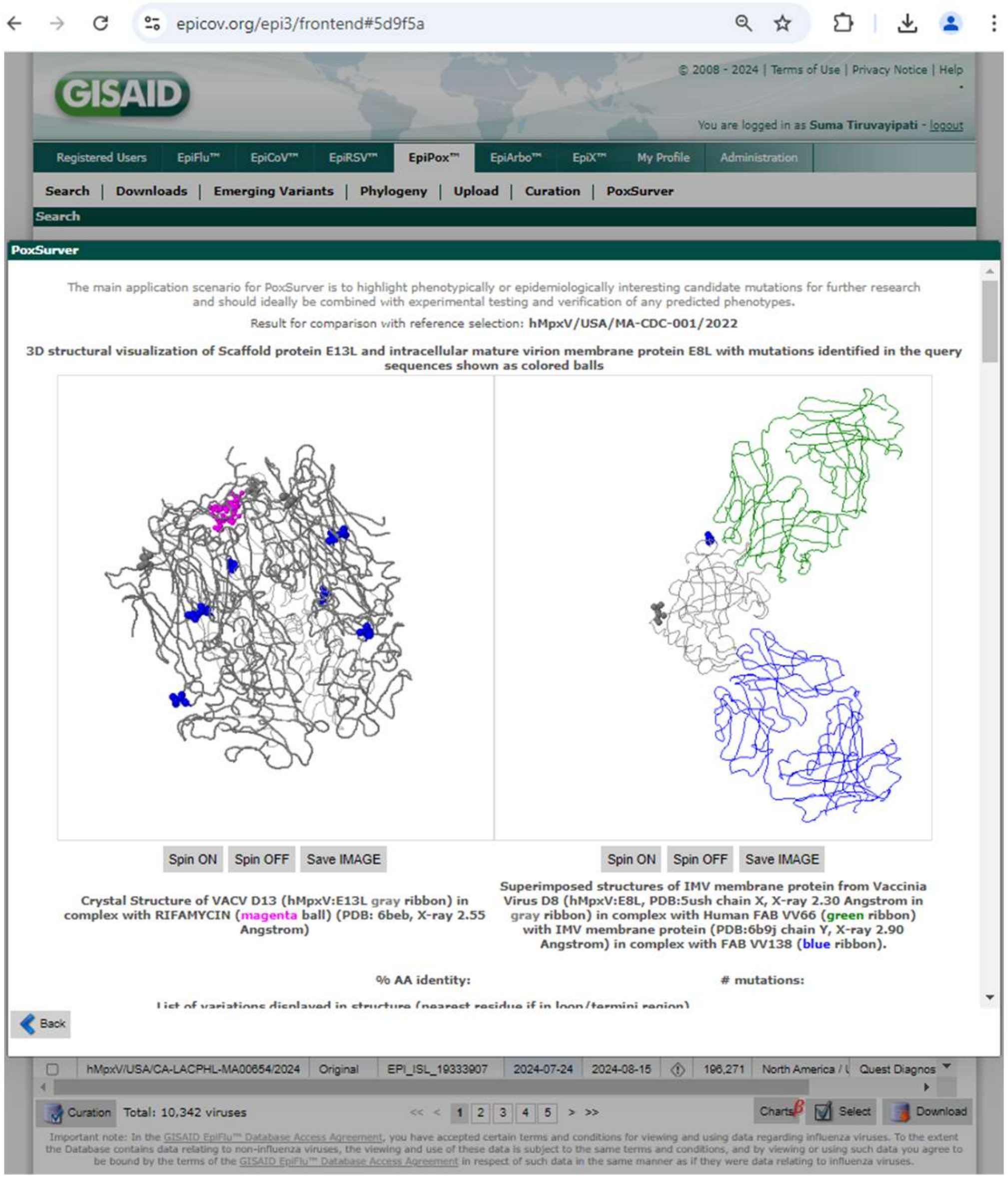
Emerging Variants

Spread | Acceleration

Democratic Republic of the Congo

Input comma separated mutations here for search, then press Enter...

Constellation ↑↓	Clade/Lineage New Lineage(Link to Issue) Search...	Count(past 100 days) ↑↓	Current Ranking ↑↓	Dissimilarity Score ↑↓	Cumulative#loc	%inEachContinent						geoMap
						North America	South America	Europe	Africa	Asia	Oceania	
A28L_I370del, A28L_I371del, A32L_Q60del, B10.5R_S18T	Ib	8	5	1								



The main application scenario for PoxSurver is to highlight phenotypically or epidemiologically interesting candidate mutations for further research and should ideally be combined with experimental testing and verification of any predicted phenotypes.

Result for comparison with reference selection: **hMpxV/USA/MA-CDC-001/2022**

3D structural visualization of Scaffold protein E13L and intracellular mature virion membrane protein E8L with mutations identified in the query sequences shown as colored balls

Crystal Structure of VACV D13 (hMpxV:E13L gray ribbon) in complex with RIFAMYCIN (magenta ball) (PDB: 6beb, X-ray 2.55 Angstrom)

Superimposed structures of IMV membrane protein from Vaccinia Virus D8 (hMpxV:E8L, PDB:5ush chain X, X-ray 2.30 Angstrom in gray ribbon) in complex with Human FAB VV66 (green ribbon) with IMV membrane protein (PDB:6b9j chain Y, X-ray 2.90 Angstrom) in complex with FAB VV138 (blue ribbon).

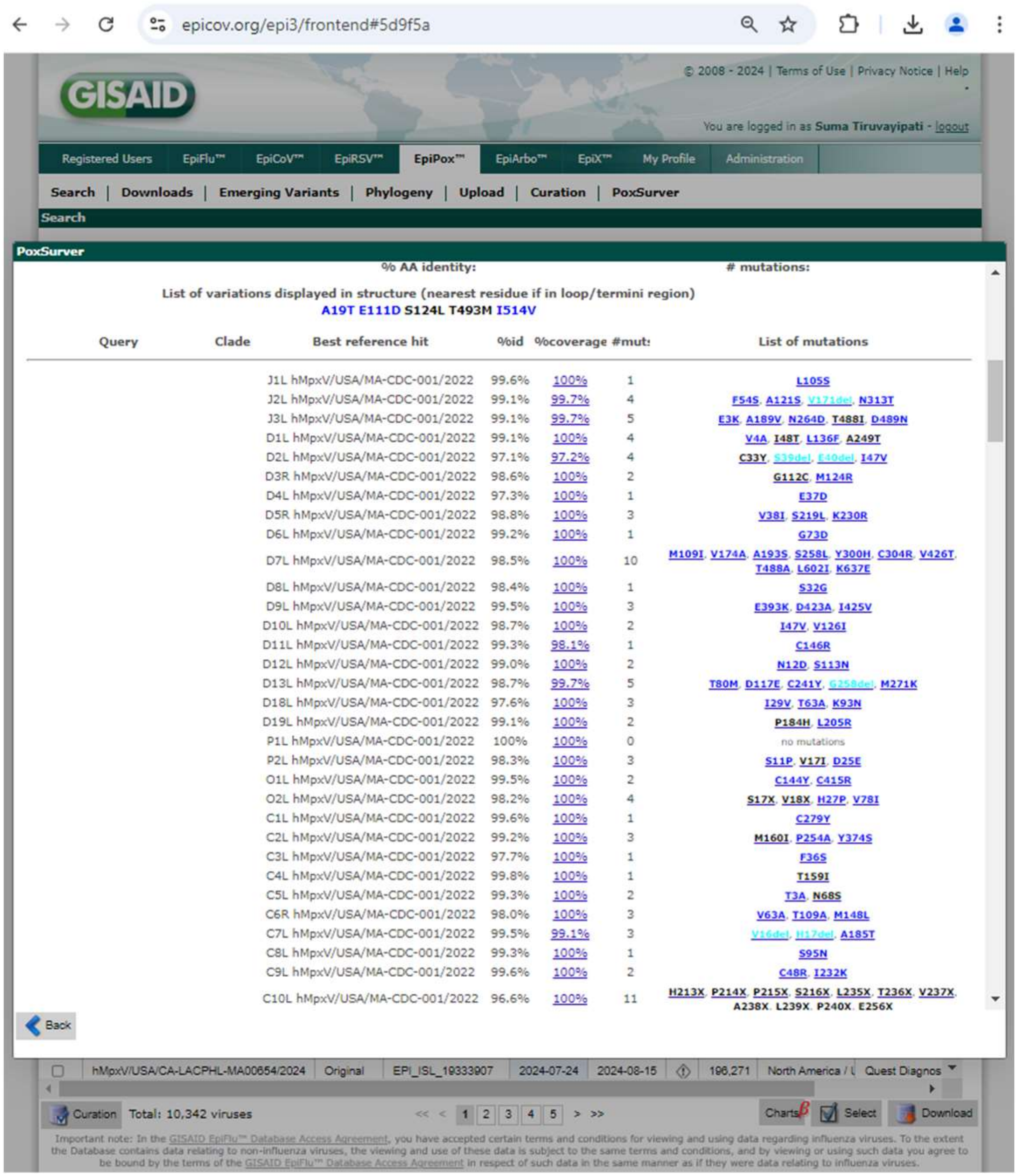
% AA identity: **# mutations:**

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

☐ hMpxV/USA/CA-LACPHL-MA00654/2024 Original EPI_ISL_19333907 2024-07-24 2024-08-15 196,271 North America / U Quest Diagnos

Total: 10,342 viruses

Important note: In the GISAID EpiFlu™ Database Access Agreement, you have accepted certain terms and conditions for viewing and using data regarding influenza viruses. To the extent the Database contains data relating to non-influenza viruses, the viewing and use of these data is subject to the same terms and conditions, and by viewing or using such data you agree to be bound by the terms of the GISAID EpiFlu™ Database Access Agreement in respect of such data in the same manner as if they were data relating to influenza viruses.



% AA identity: **# mutations:**

List of variations displayed in structure (nearest residue if in loop/termini region)
A19T E111D S124L T493M I514V

Query	Clade	Best reference hit	%id	%coverage	#mut:	List of mutations
		J1L hMpxV/USA/MA-CDC-001/2022	99.6%	100%	1	L105S
		J2L hMpxV/USA/MA-CDC-001/2022	99.1%	99.7%	4	F54S, A121S, V171del, N313T
		J3L hMpxV/USA/MA-CDC-001/2022	99.1%	99.7%	5	E3K, A189V, N264D, T488I, D489N
		D1L hMpxV/USA/MA-CDC-001/2022	99.1%	100%	4	V4A, I48T, L136F, A249T
		D2L hMpxV/USA/MA-CDC-001/2022	97.1%	97.2%	4	C33Y, S39del, E40del, I47V
		D3R hMpxV/USA/MA-CDC-001/2022	98.6%	100%	2	G112C, M124R
		D4L hMpxV/USA/MA-CDC-001/2022	97.3%	100%	1	E37D
		D5R hMpxV/USA/MA-CDC-001/2022	98.8%	100%	3	V38I, S219L, K230R
		D6L hMpxV/USA/MA-CDC-001/2022	99.2%	100%	1	G73D
		D7L hMpxV/USA/MA-CDC-001/2022	98.5%	100%	10	M109I, V174A, A193S, S258L, Y300H, C304R, V426T, T488A, L602I, K637E
		D8L hMpxV/USA/MA-CDC-001/2022	98.4%	100%	1	S32G
		D9L hMpxV/USA/MA-CDC-001/2022	99.5%	100%	3	E393K, D423A, I425V
		D10L hMpxV/USA/MA-CDC-001/2022	98.7%	100%	2	I47V, V126I
		D11L hMpxV/USA/MA-CDC-001/2022	99.3%	98.1%	1	C146R
		D12L hMpxV/USA/MA-CDC-001/2022	99.0%	100%	2	N12D, S113N
		D13L hMpxV/USA/MA-CDC-001/2022	98.7%	99.7%	5	T80M, D117E, C241Y, G258del, M271K
		D18L hMpxV/USA/MA-CDC-001/2022	97.6%	100%	3	I29V, T63A, K93N
		D19L hMpxV/USA/MA-CDC-001/2022	99.1%	100%	2	P184H, L205R
		P1L hMpxV/USA/MA-CDC-001/2022	100%	100%	0	no mutations
		P2L hMpxV/USA/MA-CDC-001/2022	98.3%	100%	3	S11P, V17I, D25E
		O1L hMpxV/USA/MA-CDC-001/2022	99.5%	100%	2	C144Y, C415R
		O2L hMpxV/USA/MA-CDC-001/2022	98.2%	100%	4	S17X, V18X, H27P, V78I
		C1L hMpxV/USA/MA-CDC-001/2022	99.6%	100%	1	C279Y
		C2L hMpxV/USA/MA-CDC-001/2022	99.2%	100%	3	M160I, P254A, Y374S
		C3L hMpxV/USA/MA-CDC-001/2022	97.7%	100%	1	F36S
		C4L hMpxV/USA/MA-CDC-001/2022	99.8%	100%	1	T159I
		C5L hMpxV/USA/MA-CDC-001/2022	99.3%	100%	2	T3A, N68S
		C6R hMpxV/USA/MA-CDC-001/2022	98.0%	100%	3	V63A, T109A, M148L
		C7L hMpxV/USA/MA-CDC-001/2022	99.5%	99.1%	3	V16del, H17del, A185T
		C8L hMpxV/USA/MA-CDC-001/2022	99.3%	100%	1	S95N
		C9L hMpxV/USA/MA-CDC-001/2022	99.6%	100%	2	C48R, I232K
		C10L hMpxV/USA/MA-CDC-001/2022	96.6%	100%	11	H213X, P214X, P215X, S216X, L235X, T236X, V237X, A238X, L239X, P240X, E256X

☐ hMpxV/USA/CA-LACPHL-MA00654/2024 Original EPI_ISL_19333907 2024-07-24 2024-08-15 196,271 North America / U Quest Diagnos

Total: 10,342 viruses

Important note: In the GISAID EpiFlu™ Database Access Agreement, you have accepted certain terms and conditions for viewing and using data regarding influenza viruses. To the extent the Database contains data relating to non-influenza viruses, the viewing and use of these data is subject to the same terms and conditions, and by viewing or using such data you agree to be bound by the terms of the GISAID EpiFlu™ Database Access Agreement in respect of such data in the same manner as if they were data relating to influenza viruses.