



EpiFlu™ Database v1.1

www.gisaid.org

Joachim Büch^{1*}, Kirsten Roomp^{2*}, Trevor Bedford³, Naomi Komadina⁴, Rebecca J. Kondor⁵, Richard Neher⁶, Catherine Smith⁵, Sebastian Maurer-Stroh^{7**}, Gunter Bach^{8**}

¹ Max-Planck-Institute for Informatics, Saarbrücken, Germany, ² Luxembourg Centre for Systems Biomedicine, University of Luxembourg
³ Fred Hutchinson Cancer Research Center, Seattle, USA, ⁴ WHO Collaborating Centre for Reference & Research on Influenza, Melbourne, Australia
⁵ Centers for Disease Control and Prevention, Atlanta, USA, ⁶ Biozentrum, University of Basel, Switzerland, ⁷ Bioinformatics Institute, Agency for Science, Technology and Research, Singapore, ⁸ The GISAID Initiative, Munich, Germany. *Joint first authors, ** Joint last authors

BACKGROUND

The Global Initiative on Sharing All Influenza Data (GISAID) provides a sharing mechanism for its publicly accessible EpiFlu™ database that incentivizes the rapid exchange of influenza virus data, by providing open-access to researchers and the development of medical interventions in a transparent manner, while protecting inherent rights of data submitters.

GISAID fosters benefit sharing and collaboration between submitters and users of data. Scientists' reticence to share data prior to publication via public-domain archives e.g. GenBank, where use of data takes place anonymously and void of any enforceable conditions to safeguard contributors' rights, prompted the creation of GISAID in 2008.

OVERVIEW

The EpiFlu™ database application, developed by the Max-Planck-Institute for Informatics, enables the analysis of the world's most complete collection of the latest seasonal to new animal influenza viruses. The application is based on a proprietary software code and Oracle software. Extensive metadata are also collected for most isolates. EpiFlu™ also provides features for searching, filtering specific datasets for download and upload functionality.

All users of EpiFlu™ have agreed to positively identify themselves. While access is free of charge, all users also agreed that they will not attach any restrictions on the data, but will acknowledge both the originator of the specimen and the submitter of the data, and seek to collaborate with the Originating Laboratory.

FEATURES IN DETAIL

- Contains influenza sequences and associated meta data with each isolate
- Genetic, clinical, epidemiological & geographical data for human isolates plus species specific data associated with non-human isolates
- Batch and single upload functions
- Ability of submitter to edit data submitted to the platform
- Each isolate accompanied by an audit trail with a history of edits
- Submitted data available for view to other users immediately
- Download facility of meta data associated with isolates in an Excel format
- Sequence download in FASTA format with user defined headers
- BLAST, alignment tool, FluSurver, nextflu phylogenetic trees (Figure 3)

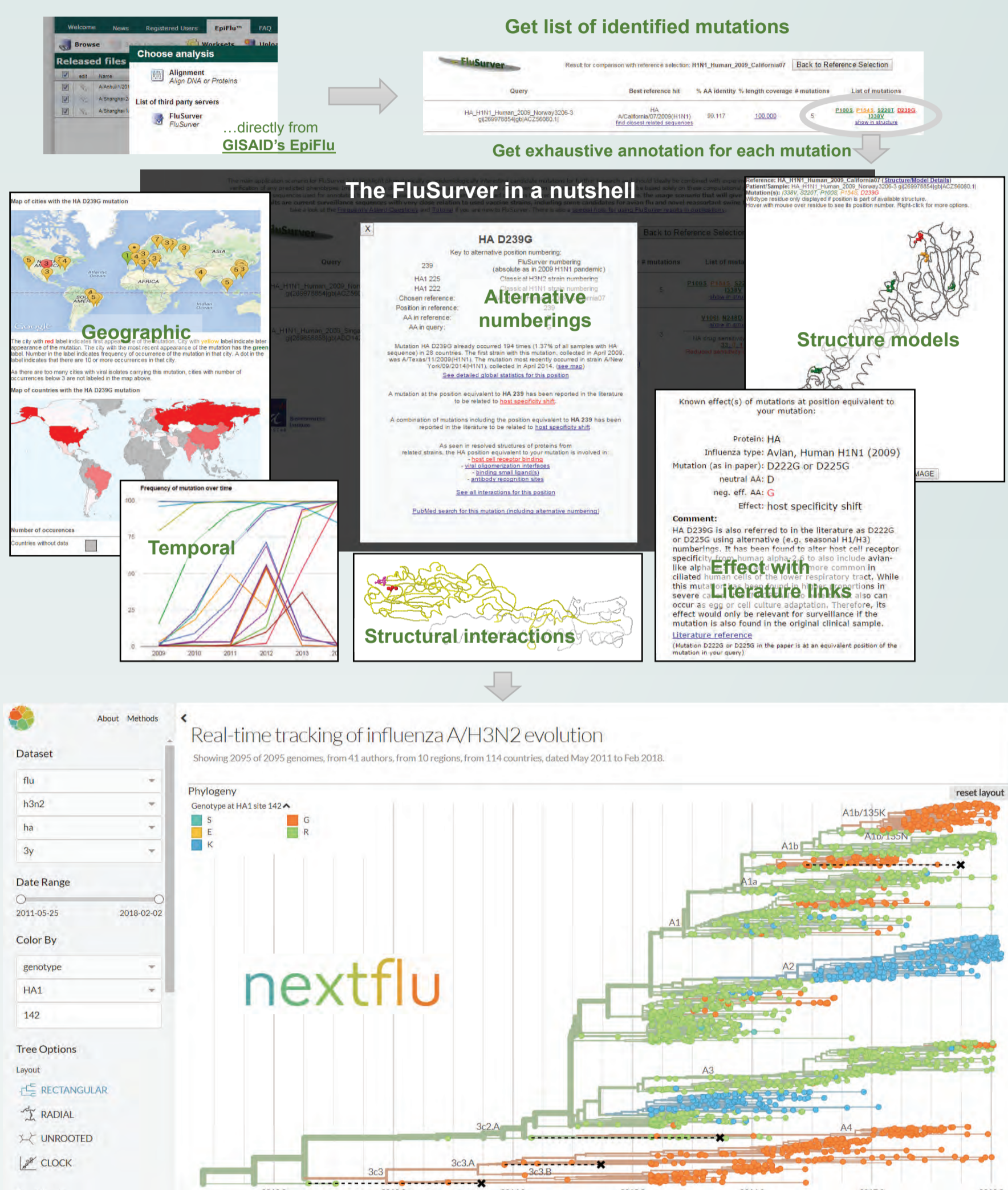


Figure 3. FluSurver mutation analysis connected to nextflu phylogenetic trees

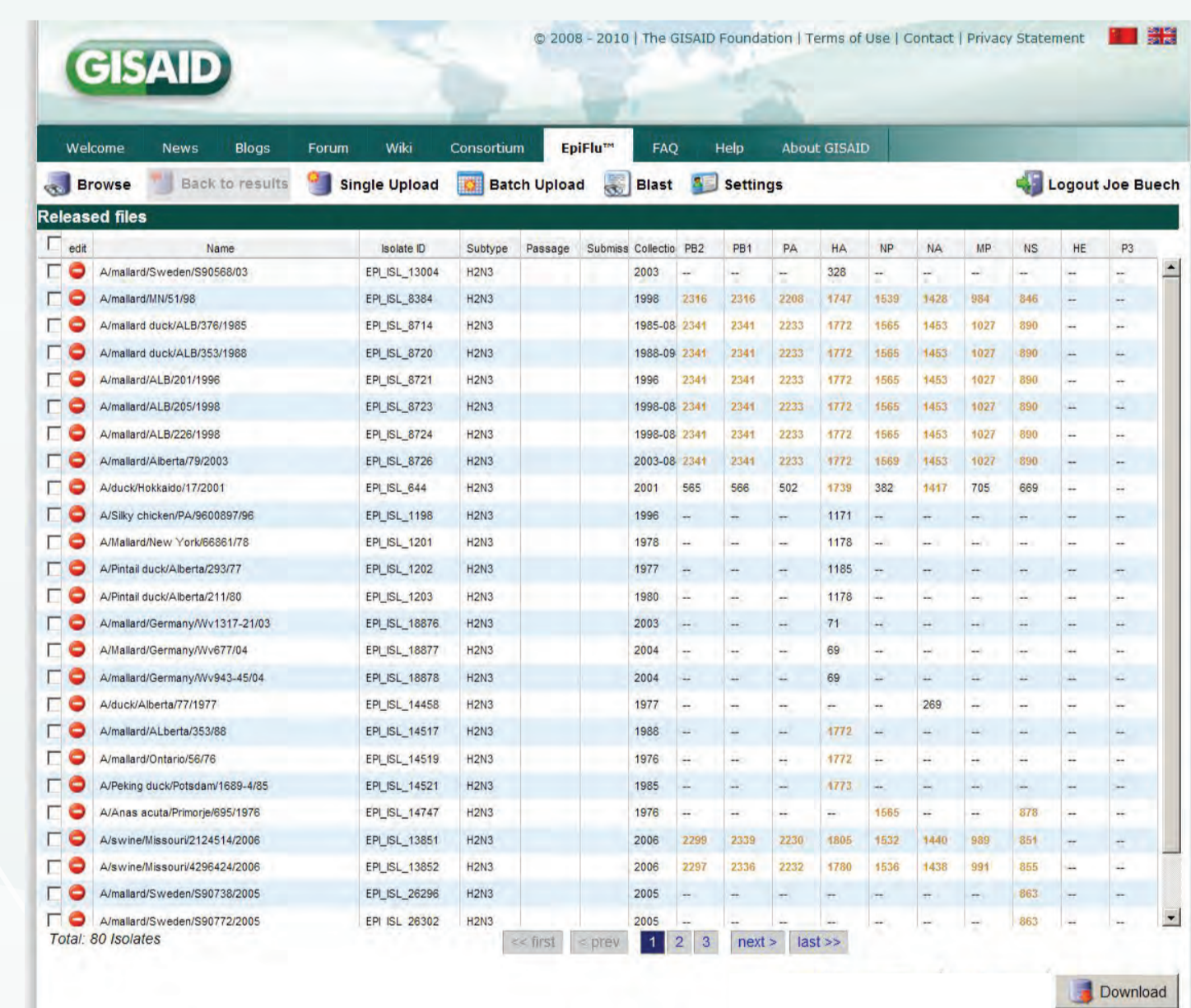


Figure 2. The search results page

Browse Functions

- Can choose to browse all isolates in the database, user submitted isolates, or isolates not available on other databases (Figure 1)
- Can browse using one or more filters for:
 - Type, subtype, isolate name, genes, species, region, country of origin
 - Lineage swl or seasonal for H1N1, Yamagata or Victoria for Type B influenza viruses
 - Isolate Originating laboratory
 - Isolate sequence Submitting laboratory
 - Isolate specimen date
 - Isolate submission date
 - Can define fields to be displayed

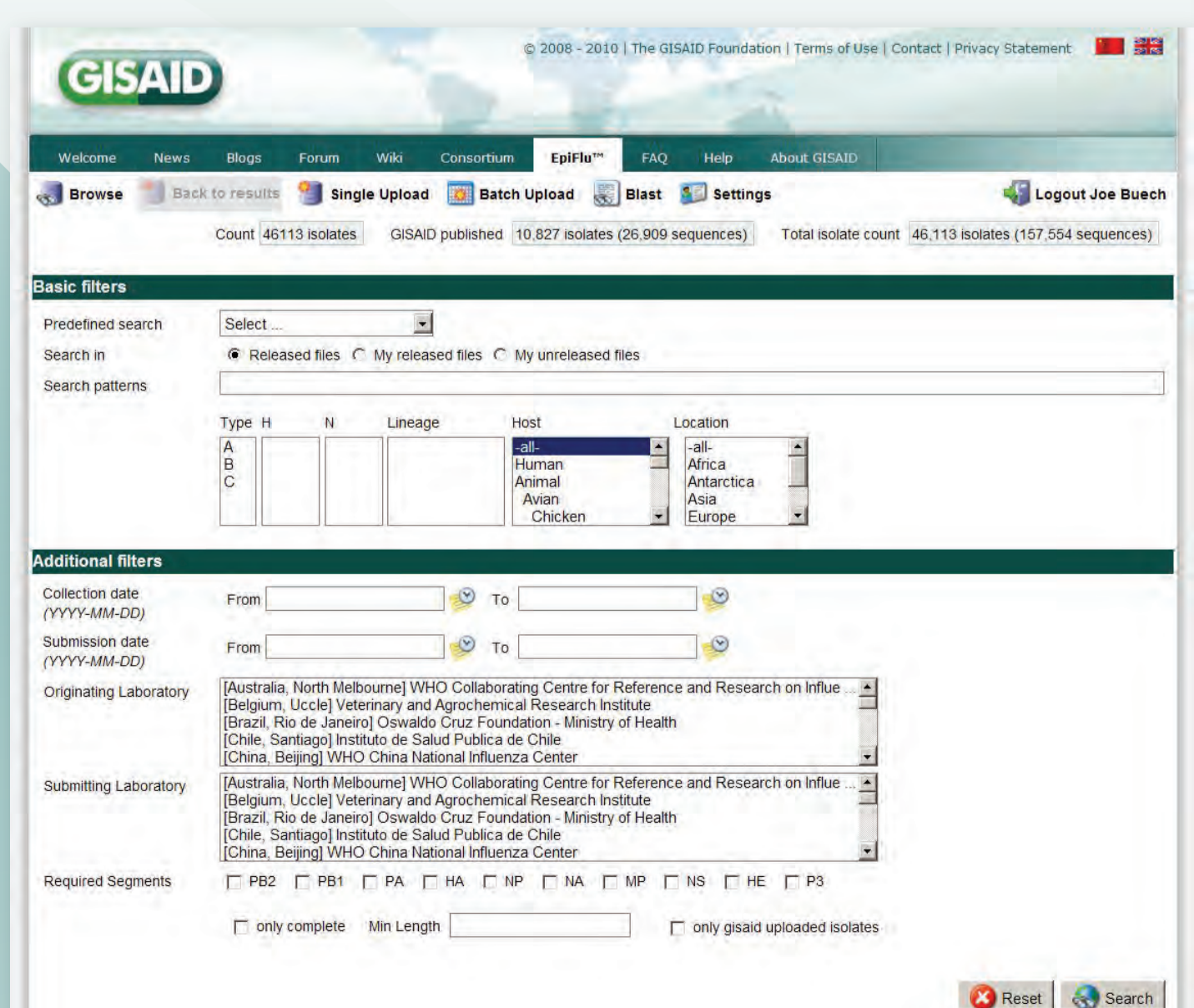


Figure 1. The browser menu

RESULTS

As of August 2019, 9,000 participants rely on data from 1,200 laboratories entrusted to GISAID from 198 nations.

EpiFlu™ database remains essential for the Global Influenza Surveillance and Response System (GISRS) and WHO biannual vaccine strain selection. Data in EpiFlu™ is comprised of 1,2 million nucleotide sequences from nearly 300,000 influenza virus strains.

Among GISAID's contributors are OIE and FAO Reference Laboratories for Avian Influenza, as well as all WHO National Influenza Centers and WHO Collaborating Centers for Surveillance, Epidemiology and Control of Influenza. To provide a complete picture of circulating influenza strains, data from public-domain archives are routinely imported.

OUTLOOK & CONCLUSION

Since in 2010, the Federal Republic of Germany constitutes an important pillar for the sustainability of GISAID through its ongoing commitment to host the EpiFlu™ database. GISAID is in the process of developing a new and advanced database application software (v3.0) to address the needs of the user community for advanced bioinformatics capabilities.

This development will also extend the spectrum of data analysis tools. The functionality of the database will also be expanded to include more data types.

GISAID EpiFlu™ has not only become a trusted, but indispensable resource for the global scientific community of influenza researchers in addition to public and veterinary health officials.

Upload Functions

- Can upload data via single upload function or batch upload sheet (Figure 4) for multiple isolates
- Metadata only needs to be entered once for each isolate
- Once isolates uploaded, extra gene sequences can be added via the edit function
- For isolates added via the batch upload sheet multiple extra gene sequences can be added by reusing the initial batch upload sheet
- Batch upload occurs in real time, with an immediate response
- Error messages are displayed for isolates & sequences not uploaded
- Duplicate sequences & isolates are flagged and not added
- Successful uploads are flagged and isolate accession and segment accession numbers are added to the return batch upload sheet

Isolate ID	Segment IDs	Isolate Name	Subtype	Lineage	Passage	History	Country	Province	sub_province	Location	Additional Info	Host	Host Additional Info	Seq ID (HA)	Seq ID (NA)
66761 HA	EPIC23047 / NA	EPIC23046	H1N1	swl	Clinical Specimen	Australia	Victoria			Devonlin		Human	deceased	0954280	0954291
66762 HA	EPIC23048 / NA	EPIC23049	H1N1	swl	Clinical Specimen	Australia	Victoria			Lansfield		Human	deceased	0954291	0954292
66763 HA	EPIC23051 / NA	EPIC23050	H1N1	swl	Clinical Specimen	Australia	Victoria			Richmond		Human	deceased	0954292	0954293
66764 HA	EPIC23053 / NA	EPIC23052	H1N1	swl	Clinical Specimen	Australia	Victoria			Frankston		Human	deceased	0954293	0954295
66765 HA	EPIC23055 / NA	EPIC23054	H1N1	swl	Clinical Specimen	Australia	Tasmania			Hobart		Human	deceased	0954295	0954296
66766 HA	EPIC23057 / NA	EPIC23056	H1N1	swl	Clinical Specimen	Australia	Victoria					Human	deceased	0954296	0954297
66767 HA	EPIC23059 / NA	EPIC23058	H1N1	swl	MDOXX	Australia	Western Australia			Bussellton		Human	deceased	0954297	0954298
66768 HA	EPIC23061 / NA	EPIC23060	H1N1	swl	MDOXX	Australia	Western Australia			Ardfield		Human	deceased	0954298	0954299
66769 HA	EPIC23062		H1N1	swl	Clinical Specimen	French Polynesia	Tahiti					Human	deceased	0954299	0954300
66770 HA	EPIC23044 / NA	EPIC23043	MP	EPIC23042	seasonal E2	Australia	Victoria			Melbourne		Human		0954300	0954328
66771 HA	EPIC23044 / NA	EPIC23043	MP	EPIC23042	seasonal E2	Australia	Victoria			Melbourne		Human		0954328	0954329
66772 HA	EPIC23066 / NA	EPIC23065	H1N1	swl	Clinical Specimen	New Zealand	North Island			Waikato		Human		0954479	0954506
66773 HA	EPIC23067 / NA	EPIC23066	H1N1	swl	ndck-ndck-swl3	New Zealand	North Island			Waikato		Human		0954511	0954518
66774 HA	EPIC23068 / NA	EPIC23067	H1N1	swl	ndck-ndck-swl3	New Zealand	North Island			Waikato		Human		0954512	0954519
66775 HA	EPIC23070 / NA	EPIC23069	H1N1	swl	ndck-ndck-swl3	New Zealand	South Island			Dunedin		Human		0954513	0954520
66776 HA	EPIC23072 / NA	EPIC23071	H1N1	swl	ndck-ndck-swl3	New Zealand	South Island			Dunedin		Human		0954514	0954521
66777 HA	EPIC23074 / NA	EPIC23073	H1N1	swl	MDOXX	Australia	Victoria			Frankston		Human		0954515	0954522
66778 HA	EPIC23076 / NA	EPIC23075	H1N1	swl	Clinical Specimen	Australia	Victoria			Melbourne		Human		0954516	0954540
66779 HA	EPIC23078 / NA	EPIC23077	H1N1	swl	Clinical Specimen	Australia	Victoria			Melbourne		Human		0954517	0954541
66780 HA	EPIC23079		H1N1	swl	Clinical Specimen	Australia	Victoria			Melbourne		Human		0954518	0954542
66781 HA	EPIC23081 / NA	EPIC23080	H1N1	swl	Clinical Specimen	Australia	Victoria			Melbourne		Human		0954519	0954544
66782 HA	EPIC23083 / NA	EPIC23082	H1N1	swl	Clinical Specimen	Australia	Victoria			Melbourne		Human		0954520	0954545
66783 HA	EPIC23085 / NA	EPIC23084	H1N1	swl	Clinical Specimen	Australia	Victoria			Melbourne		Human		0954521	0954546
66784 HA	EPIC23087 / NA	EPIC23086	H2N2	swl	Clinical Specimen	Australia	Victoria			Lysterfield		Human		0954522	0954548
66785 HA	EPIC23089 / NA	EPIC23088	H1N1	swl	MDOXX/H2N2	Malaysia						Human		0954523	0954549
66786 HA	EPIC23091 / NA	EPIC23090	H1N1	swl	swl	New Zealand	North Island					Human		0954524	0954590

Figure 4: The batch upload facility, successfully uploaded isolates