The GISAID Initiative

Real Time Global Communication in Disease Prevention

Bioinformatics Workshop
26-27 August 2019
Singapore

WHO Global Influenza Programme (GIP)
National Centre for Infectious Diseases (NCID) Singapore
Bioinformatics Institute (BII)
Global Initiative on Sharing All Influenza Data (GISAID)

Database Technical Group

hosted by the Federal Republic of Germany

Naomi.Komadina@influenzacentre.org

Challenges in 2006-2007

- Access to the latest genetic data for highly pathogenic H5N1 influenza zoonotic infections was often restricted, in part due to the hesitancy by MS to share their information.
- Scientists' reticence to share data pre-publication, out of fear of being scooped[1] delayed sharing.
- Public-domain archives (anonymous access - use of data) offered:
  - no protection of owners' interests (intellectual property rights);
  - limited incentive to share data;
  - no transparency on the use of data;
  - no effective mechanisms to ensure acknowledgement of labs providing viruses or contributing data.

[1] **Pearson H. - Competition in biology: It's a scoop! *Nature* 2003; DOI: 10.1038/news031124-9**

GISAID data access and sharing principles

- Data Submitters grant Data Users an irrevocable licence for the use of their data, providing legal certainty for use in research and publications and for the development, testing and dissemination of interventions such vaccines, diagnostics and therapeutics
- Access to the GISAID EpiFlu™ Database is free of charge and open to everyone provided they identify themselves in order to foster collaboration and permit an effective oversight to uphold the enshrined sharing principles
- The guiding principle for those who use the Data in particular in publications is the need to acknowledge the contribution of both the suppliers and the submitters of the data
- Whilst all Data are publicly accessible, Submitters do not forfeit their rights (IPR) to the data they deposit in GISAID

See: GISAID EpiFlu™ Database Access Agreement

Naomi.Komadina@influenzacentre.org

EpiFlu™ Database

- The most complete set of influenza sequences (285K isolates from 1.2 million sequences)
- Includes patient metadata & geographical locations
- Submitters retain rights over their data
- Easy upload of data
- Isolate based entry rather than sequence only
- Isolates can be updated by submitter in real-time
- Automated sequence annotation on upload
- Customized search function outputs
- Customized FASTA headers with sequence downloads
- Metadata can be downloaded in spreadsheet
- Workset function, worksets can be shared amongst users
- Analysis tools

Workflow of EpiFlu™ Database

Data

Sequence data is organized into unique isolates assigned an EPI_ISL Accession (defined by virus name and passage history), and segment EPI Accession metadata fields include: Isolate name, virus type, HA and NA subtype, lineage (B; H1N1), host, collection date, location, patient data i.e.: age, gender, vax status.

Originating Laboratory - collected the specimen (Lab ID)

Submitting Laboratory - sequenced the virus

Upload

Single Upload – using the webpage interface (5-10 min)

Batch Upload – using a macro-enabled excel spreadsheet (necessary fields; time dependent on amount of data; automatic)

Complimented with data from public-domain archives routinely imported

Curation

Checks the correctness of sequence and metadata

Flags errors

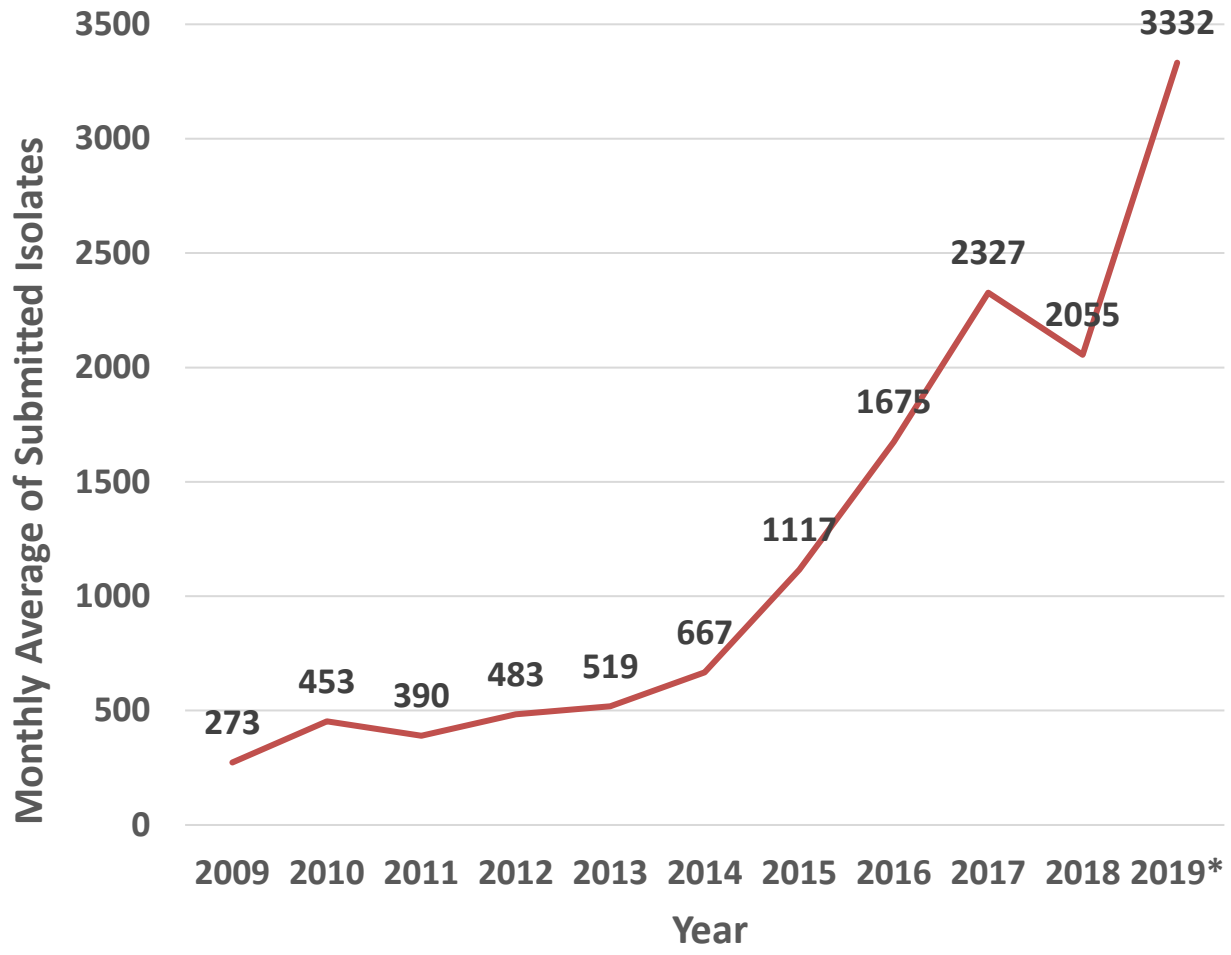
Data to Unreleased Files

Pop-up message – errors, where and what type

Release of corrected data

Naomi.Komadina@influenzacentre.org

Increase in genetic sequence data published via GISAID

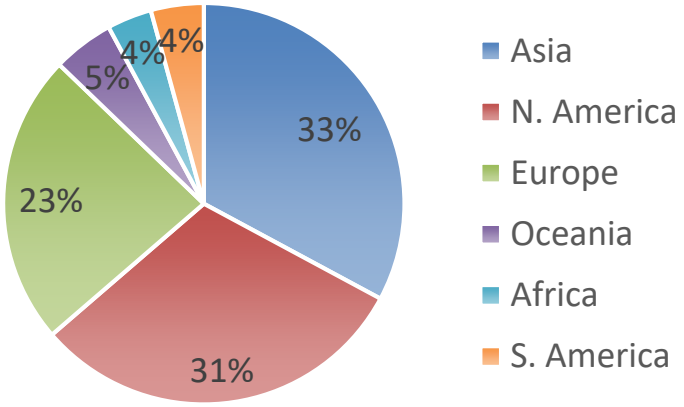


Geographical representativeness

- ~1,2m sequences of ~283,000 influenza viruses (70% human; 30% animal)
- Data from >1,200 labs (incl. all 152 GISRS)
- Contains data from >2,000 influenza viruses with pandemic potential (IVPP)
    - >30,000 animal viruses (H4-H11)
    - Candidate vaccine viruses
- 8,700 active participants
- Human and animal health
- 2017-18: sequence data of ~3,000 viruses (avg) uploaded monthly (70% directly to GISAID)

Naomi.Komadina@influenzacentre.org

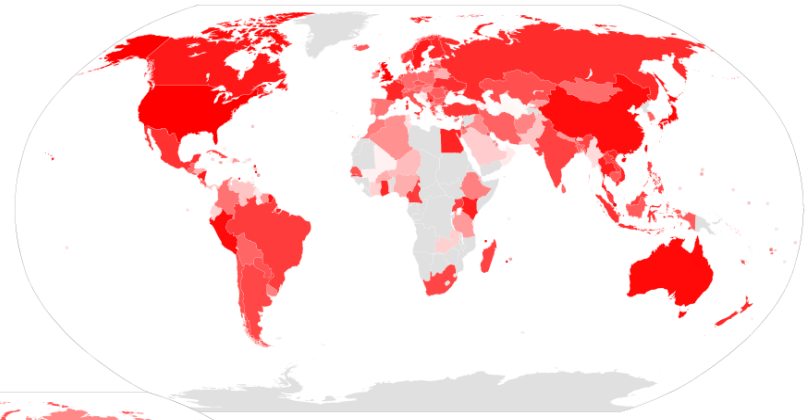
- Timely submission directly to GISAID: 94% of total within 3 months of collection; 83% (within 6 mo); 81% (within 12 mo); 77% (within 24 mo)

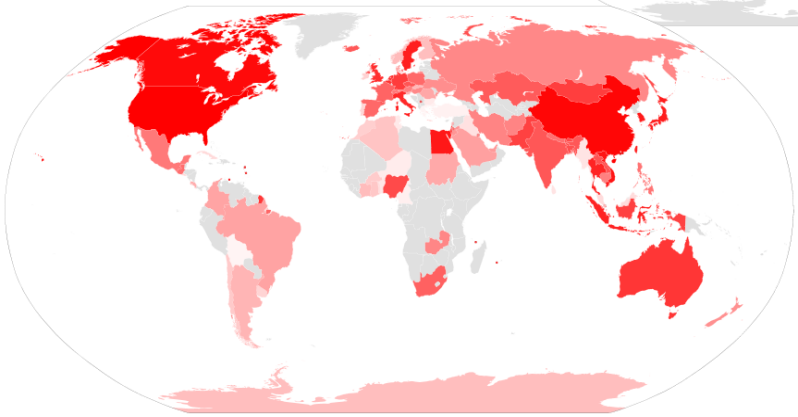


(Hemispheres, Countries, Regions)

(282,481 viruses / 1,175,910 sequences; 9th Aug 2019)

Human-derived influenza virus





Animal-derived influenza virus

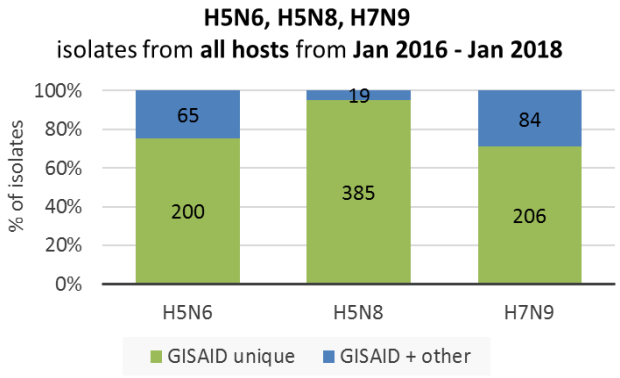
Data in GISAID from >1,200 labs worldwide

Countries ranked by number of HA sequences in GISAID, colored by rank percentile from red (most) to white (least), gray (none)

(Analyses compiled by J. Büch (MPII, Germany) & S. Maurer-Stroh (A*STAR BII, Singapore)

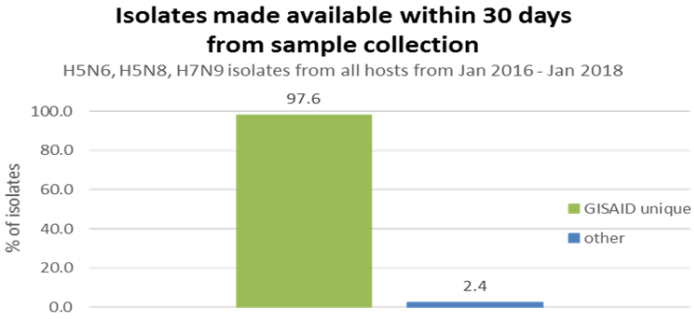
Naomi.Komadina@influenzacentre.org

Rapid sharing of new avian influenza viruses with pandemic potential via GISAID

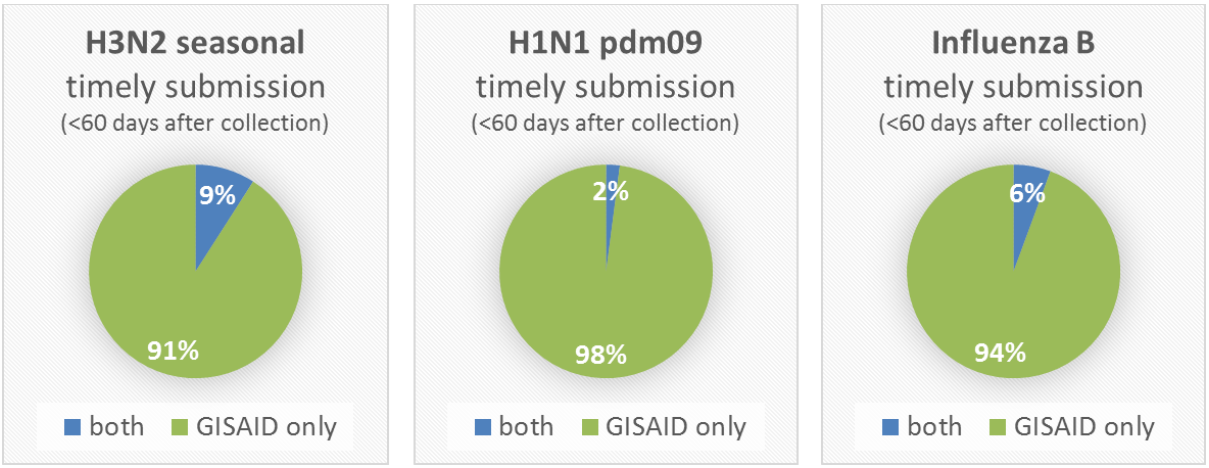


The majority of isolates from new avian influenza subtypes with pandemic potential are **unique to GISAID.**

Timeliness:
98% of the strains submitted within 30 days from sample collection are **unique to GISAID**



Timely submission of data to GISAID-human seasonal viruses



both … public-domain and GISAID

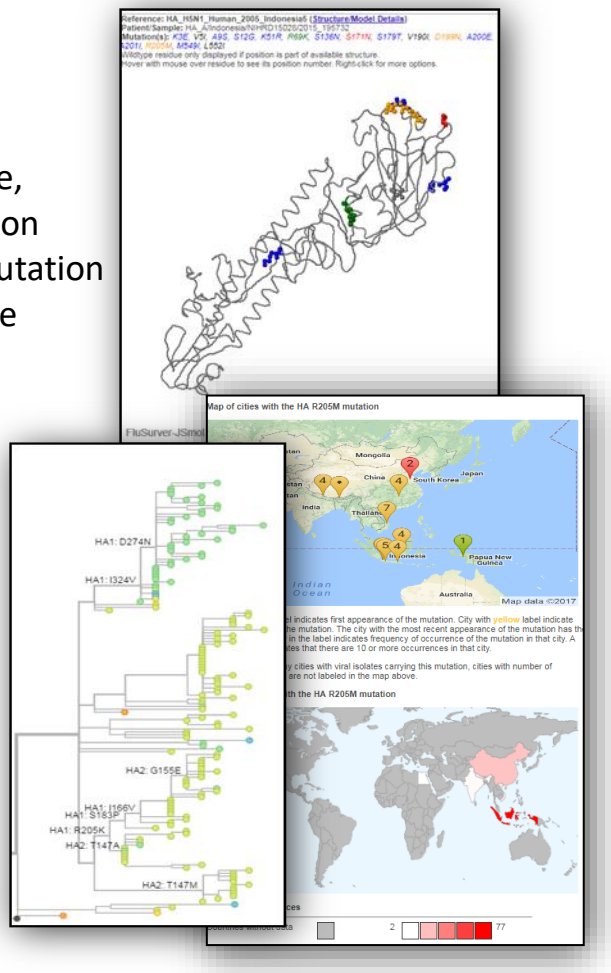
Proportion of unique sequence information in GISAID

Naomi.Komadina@influenzacentre.org

Near real time data analysis

FluSurver: developed at the BII A*STAR, Singapore, assists the identification, analysis and interpretation of mutations in influenza sequences, providing mutation frequencies automatically kept up to date with the latest GISAID data.

Annotated Tree Tool via nextflu: developed at the University of Basel, displays phylogenetic trees, along with alignments of nucleotide and amino acid sequences, showing the emergence of new clades in a temporal & geographical context



Projects

- Global Server Resilience Strategy
- Expansion of Tools and Next Generation Sequencing
- Capacity building through educational program via GISAID's series of workshops, webinars and seminars
- Realization of GISAID scientific retreat and school
- Application of GISAID sharing mechanism to other pathogens

- European Commission PREDEMICS

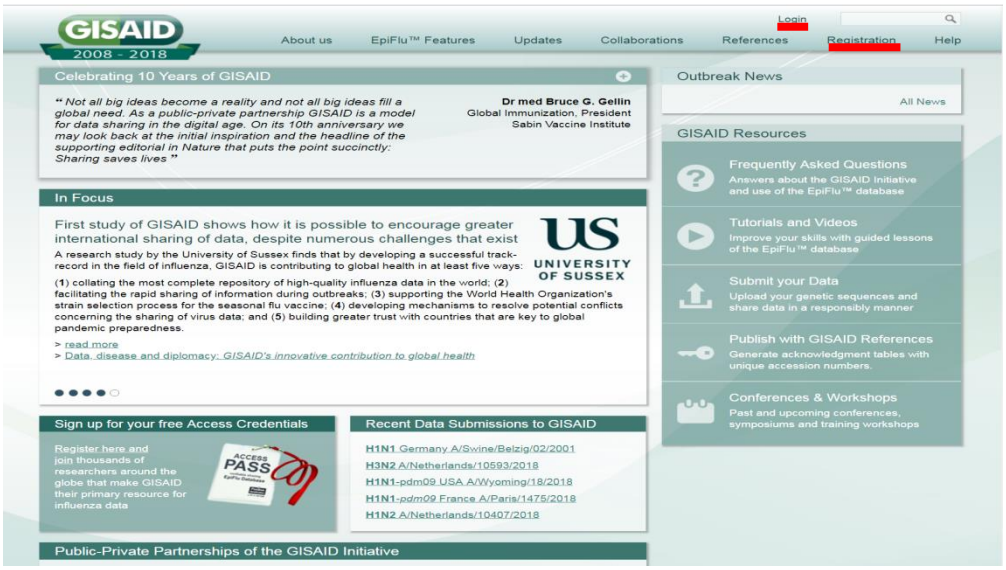
Naomi.Komadina@influenzacentre.org

GISAID

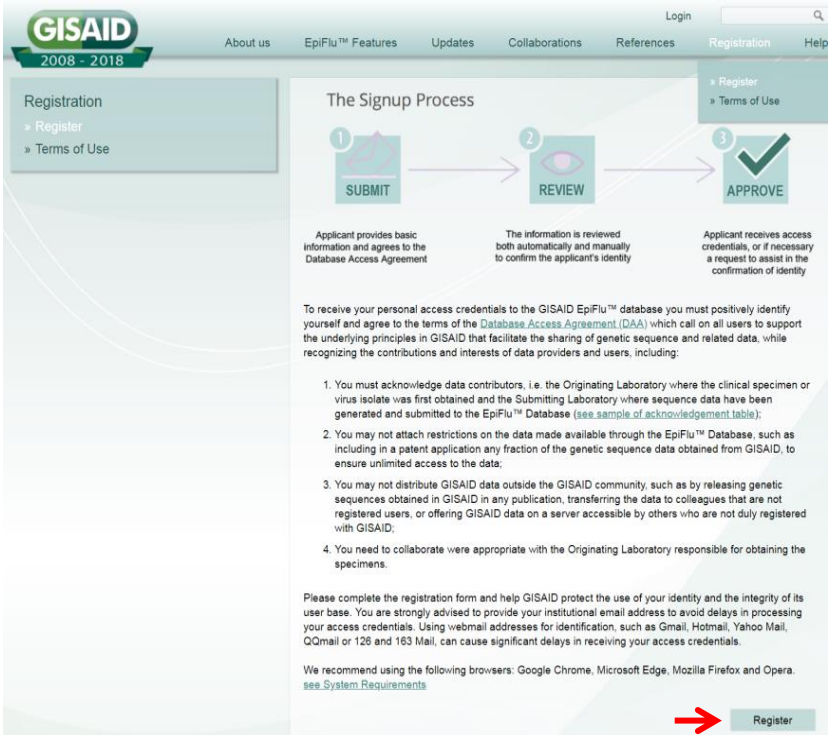
How to Register and obtain Access Credentials

Access the GISAID homepage https://gisaid.org

Click the 'Registration' button on the top RHS and select 'Register" in the drop down menu

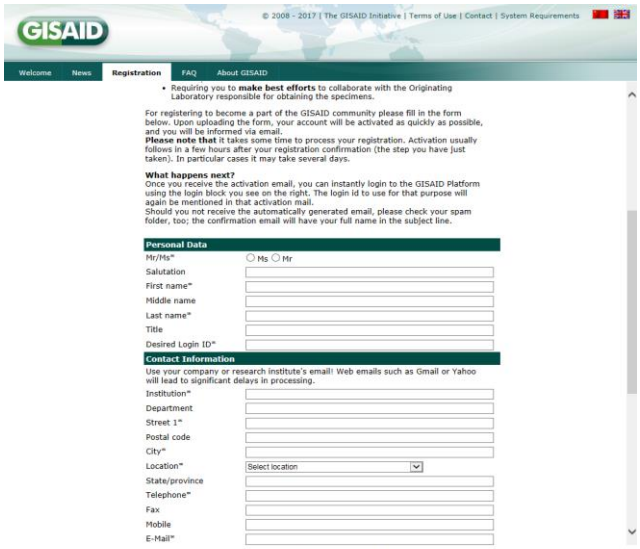


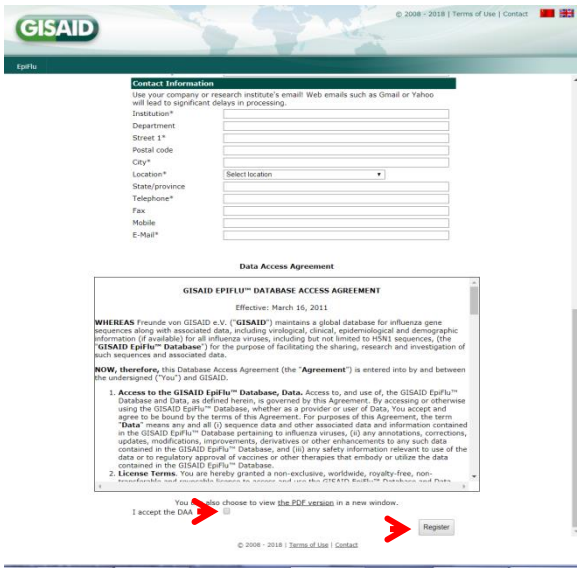
Be sure to familiarize yourself with the terms of access summarized in the Sign Process screen before you select the 'Register' button at the bottom RHS of the page.



Naomi.Komadina@influenzacentre.org

Part of the registration page





Follow the instructions, and complete the form with accurate information.

Please make sure that you use an institutional email address and not a personal email address, such as Gmail or Yahoo, as this will delay in the issuance of your access credentials

For those without access to an institutional email service, please ensure to provide your mobile phone and the institute's phone number so the registration team can get in touch with you to positively establish your identity. This is necessary to protect the use of your identity, and to uphold the GISAID sharing mechanism and the integrity of its userbase.

Once you have read the conditons of Database Access Agreement (DAA) which outlines the terms of your access to and use of the EpiFlu™ database, you will need to agree to the DAA by ticking the box at the bottom where it states "I accept the DAA" before you select the Register button, to submit your request.

This will be followed by a message asking you to look for an email the system automatically generats. Be sure to look for this email in your inbox, or if need be in your SPAM folder. This email contains a verification link that you are asked to click on, so the GISAID system knows the email address you provided is accurate.

After successful verification of the information you provided, you will on average <12 hours receive an email with your personal access credentials (Username & Password), which are yours to keep even if you should change insitutions down the road.

As with any passwords, be sure to change the password to something you can remember. To change the password, logged in first and select the "Settings" tab, see figure at the bottom of Page 9

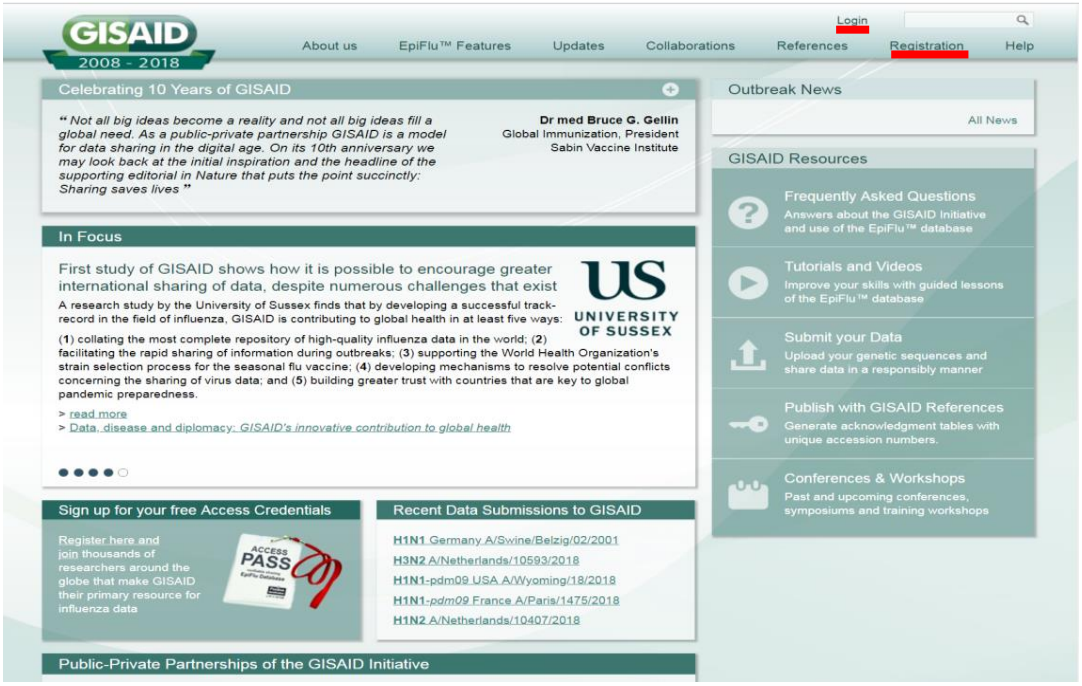
As a reminder: Never ever share your username or password, as this might cause the loss of your access credentials.

For those of you who are already registered but forgot their username or password, select “Forgot your password or Login?” . This can be found on the front page directly beneath the ‘Login’ button, refer to the first figure of this document.

How to Login

Open the GISAID website

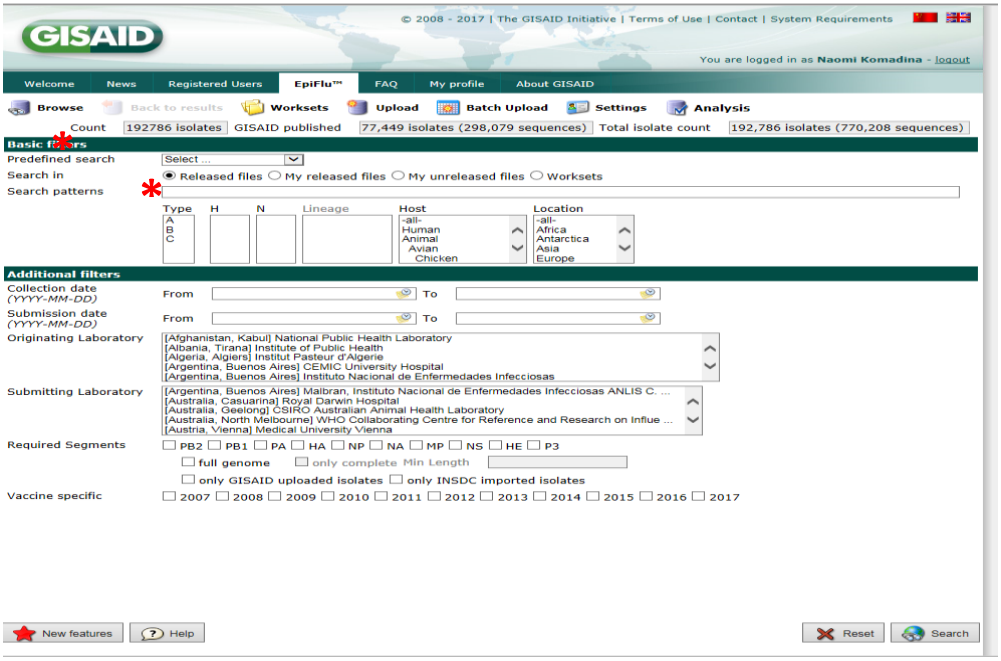
Click the ‘Login’ button on the top RHS and login with your Username and Password on the Login page



Searching for isolates in GISAID

Naomi.Komadina@influenzacentre.org

EpiFlu™ Start Page



Select which database you want to search

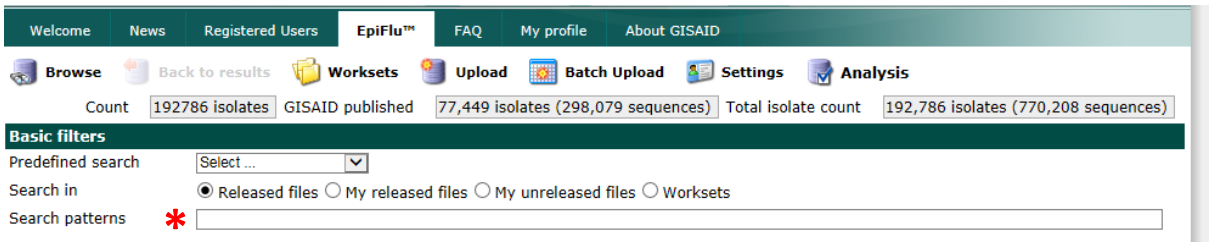
- Released Files
  - This section contains all available sequences in GISAID
- My released files
  - If you have uploaded data to GISAID all your data is contained here and can be searched
- My unreleased files
  - All files which you have uploaded to GISAID but are yet to be released to the main data base are held here. Once checked they can be released & will transfer to 'My released files' and 'Released files'

If you have never uploaded to GISAID only the Released Files databses is available for searching

Multiple search options

Count Button:- gives a running total of how many seqeunces your search has located

Searching using 'Search patterns' field



Free text can be used here and different types and subtypes can be seached for at the same time.

Naomi.Komadina@influenzacentre.org

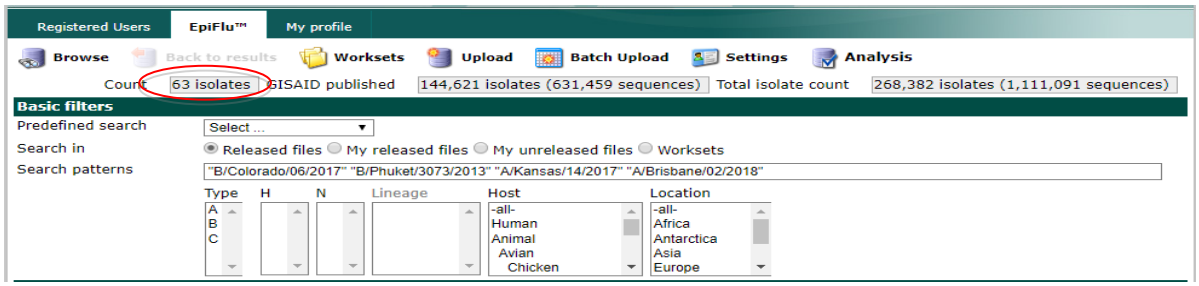
Search here using virus name or ISL_ID or segment number

All virus names must be within inverted commas is “A/Sydney/5/1997”

Search using wild cards ie “A/Hong Kong/27%%/2019”

Each name or ID number must also be separted by a space.

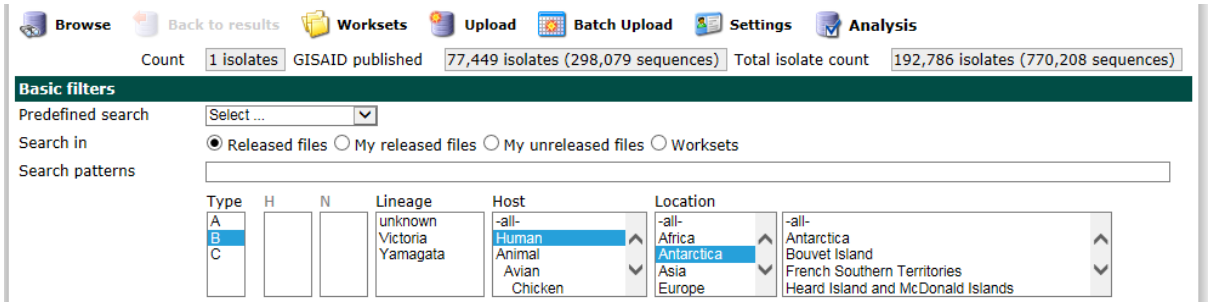
An example of a multi type or subtype search pattern is:-



63 Isolates have been located for these 4 viruses.

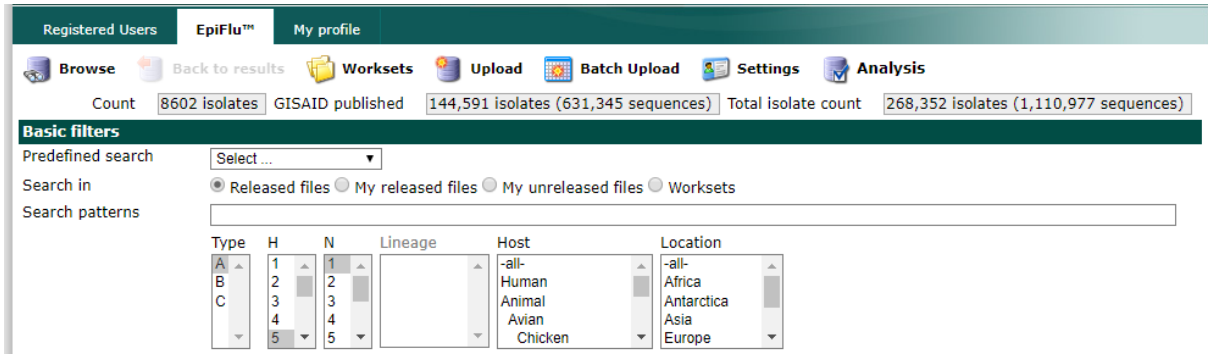
If you wish to search for a particular type or subtype from a broad area, such as a continent:-

ie:- Type B viruses from Antarctica, 1 virus found.



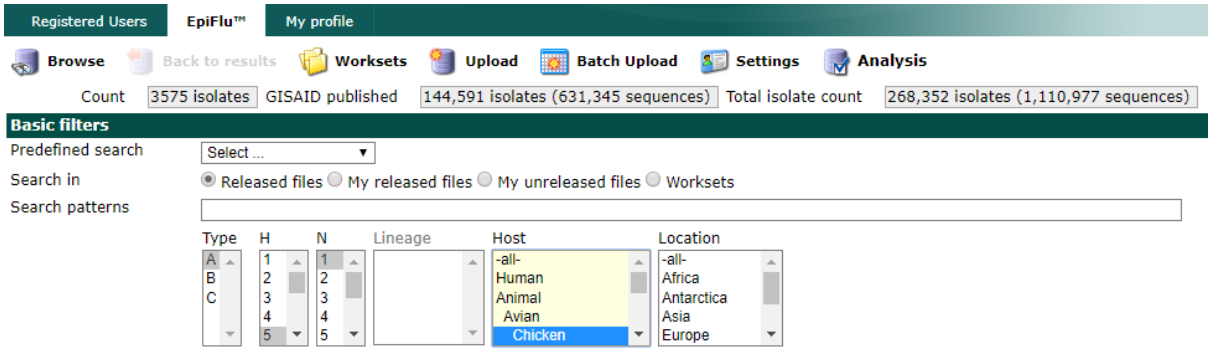
Searching by subtype ie:-H5N1 viruses

Naomi.Komadina@influenzacentre.org

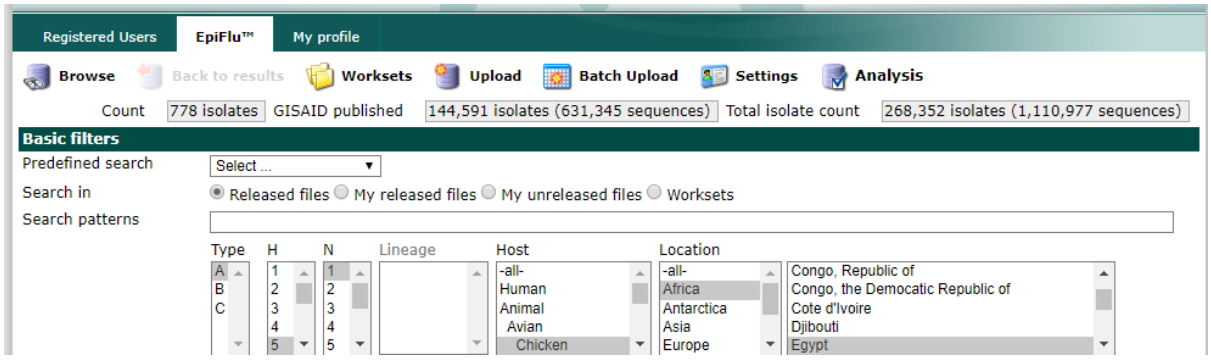


A total of 8602 isolates have been located

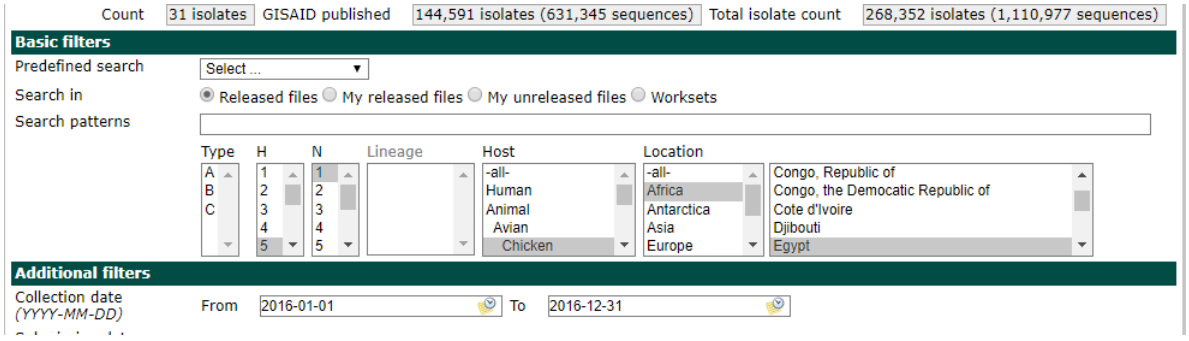
By further refining the search, ie:- chickens, the number of isolates has dropped to 3575.



Search can be even further refined by selecting a location



778 isolates sourced from chicken in Egypt have been located and this can be further refined by selecting a time period



Egypt had 31 H5N1 isolates from chicken with specimen dates in 2016 submitted to GISAID

Naomi.Komadina@influenzacentre.org

This search can be further narrowed by selecting a laboratory who provided the samples, a laboratory which submitted the sequences, a time period when the 2016 were submitted to GISAID and which genes you which gene you wish to search for as well as gene length and full genome sets as well.

Searches can be restricted to what was uploaded to GISAID only and therefore is not available on other databases, or those isolates which were uploaded to other databases first then imported to GISAID.

Search parameters are saved and not lost when you move onto another page and can be added to, however if you then wish to search for a different subtype you need to reset the search and start again.

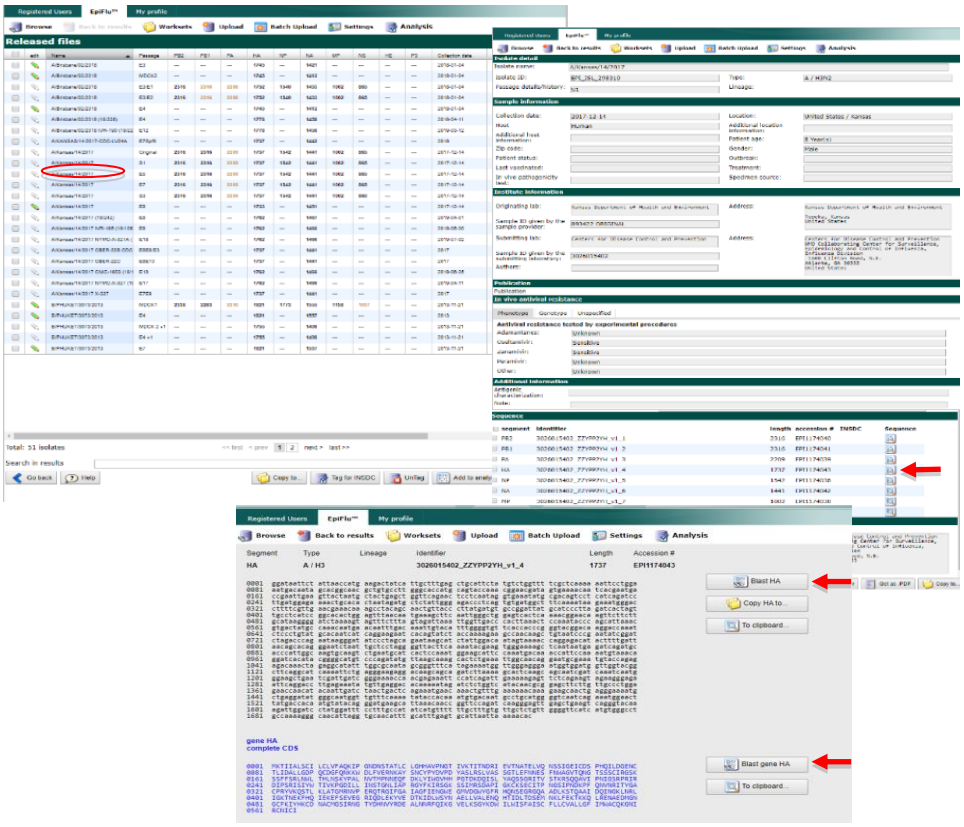
To view sequences, select the Search function at bottom RHS of page and the 34 isolates are displayed

Analysis functions include alignments, BLAST, mutation analysis using FluSurver & nextflu

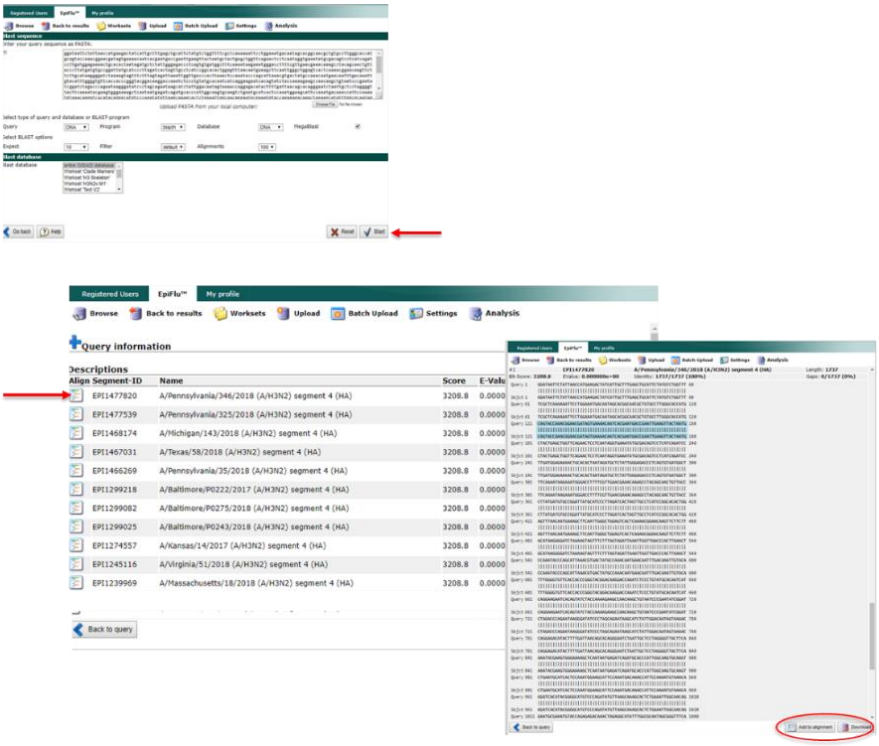
Data downloads include nucleotide and protein sequences, sequence metadata, sequence acknowledgement tables, BLAST results

BLAST function

From search results, click on virus name



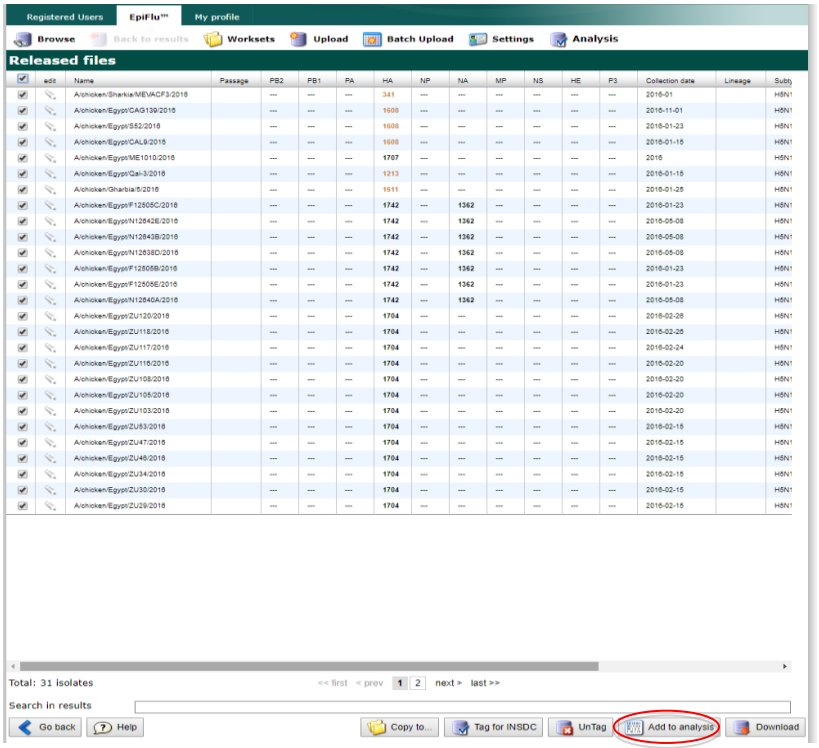
Naomi.Komadina@influenzacentre.org



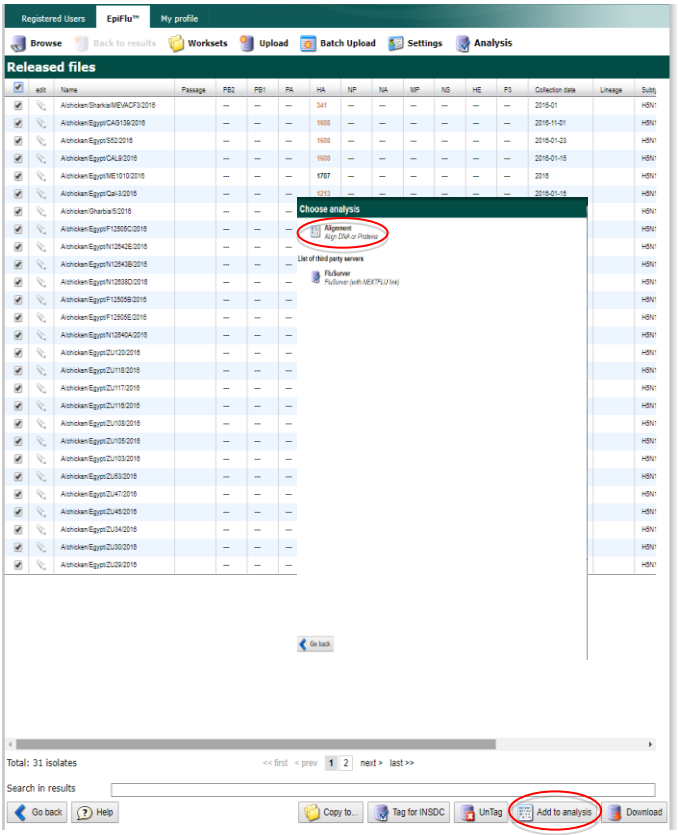
BLAST searches can be carried out against the entire database (Released files) or My released files or any workset that you have created.

Analysis Tools

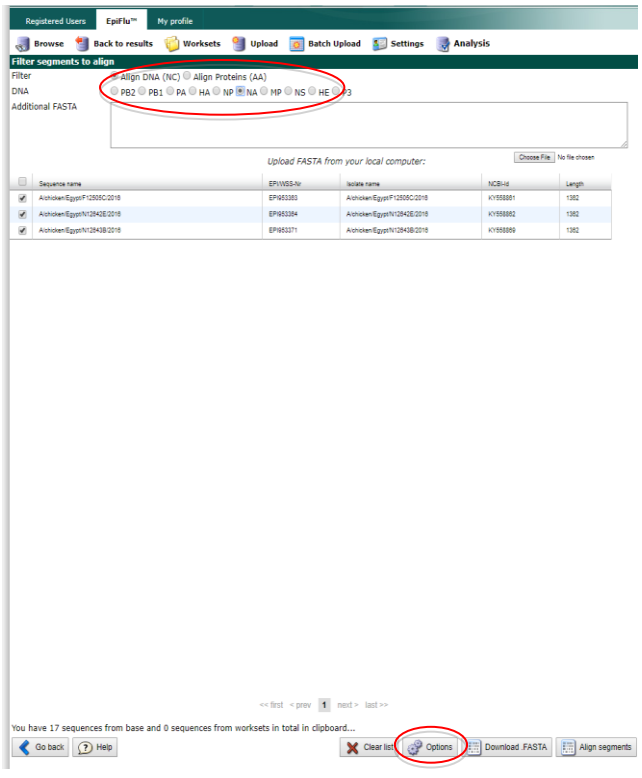
To add to Analysis, select viruses of interest, then select Add to Analysis Button



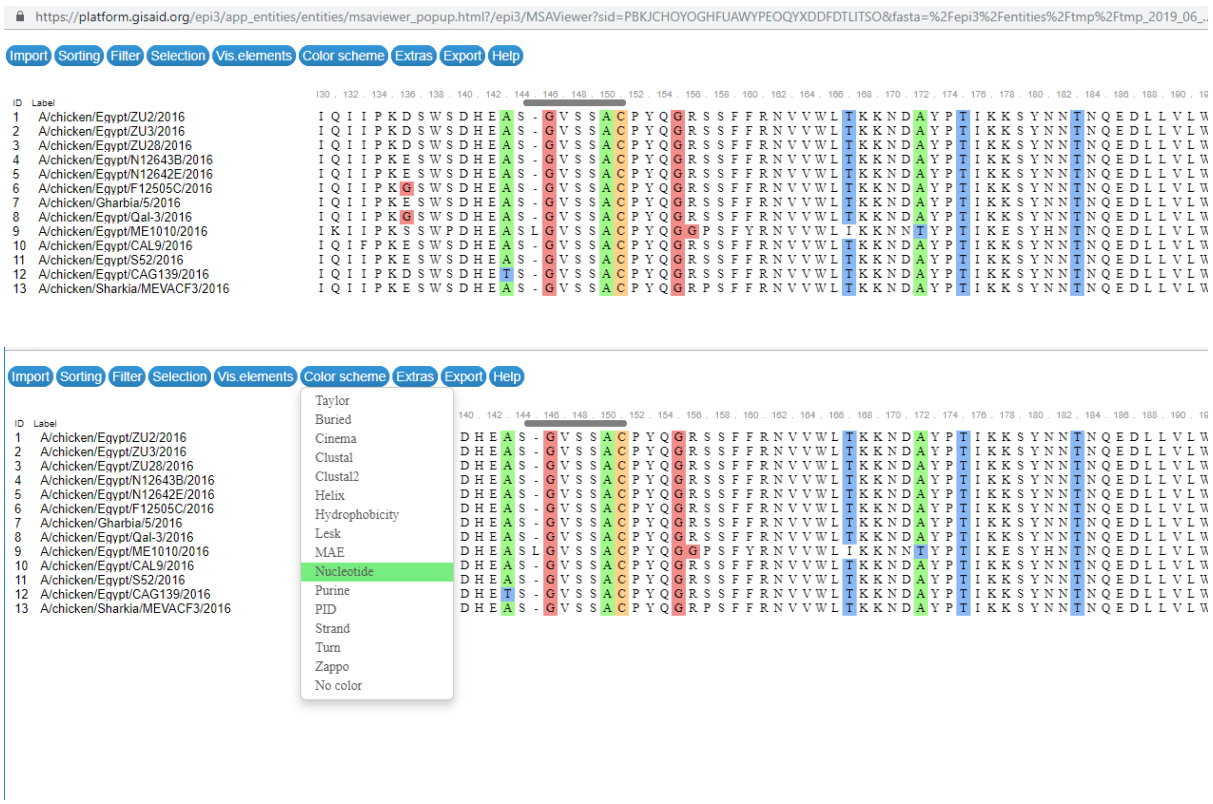
Naomi.Komadina@influenzacentre.org



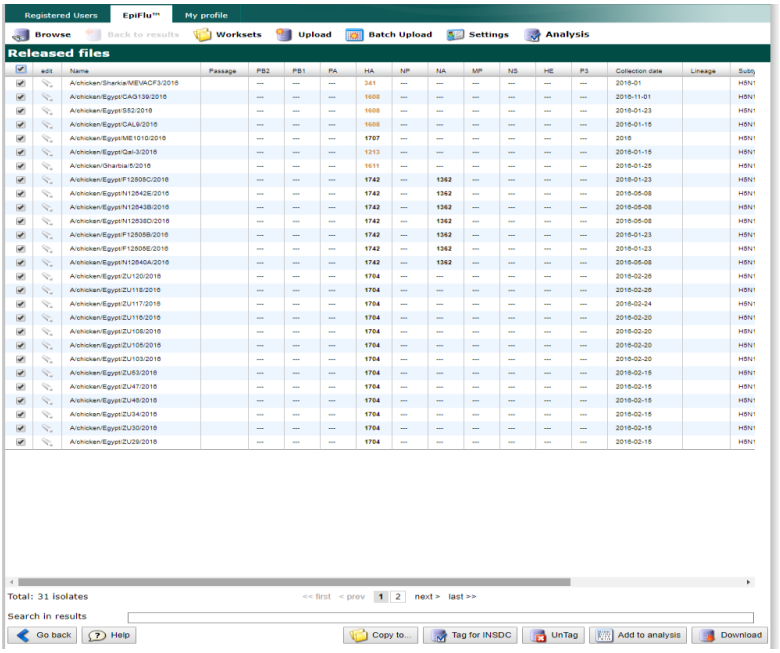
Select the gene of interest. Select ‘Options’ to vary analysis parameter. Select Align segments.



Naomi.Komadina@influenzacentre.org



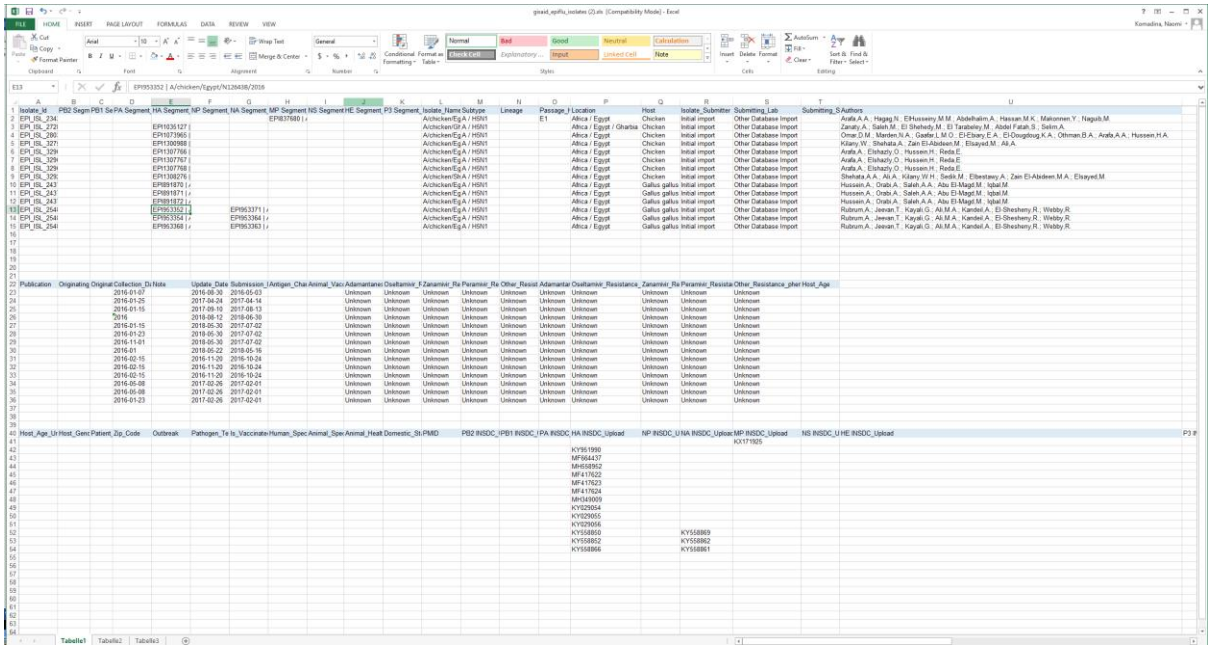
To download select isolates of interest by selecting check box next to each sequence of interest, for all select check box above the line, then select download



Metadata can be downloaded in an excel sheet or sequences as either DNA or Proteins and an acknowledgment table for inclusion in publication supplementary material

Naomi.Komadina@influenzacentre.org

For metadata select “Isolates as XLS”

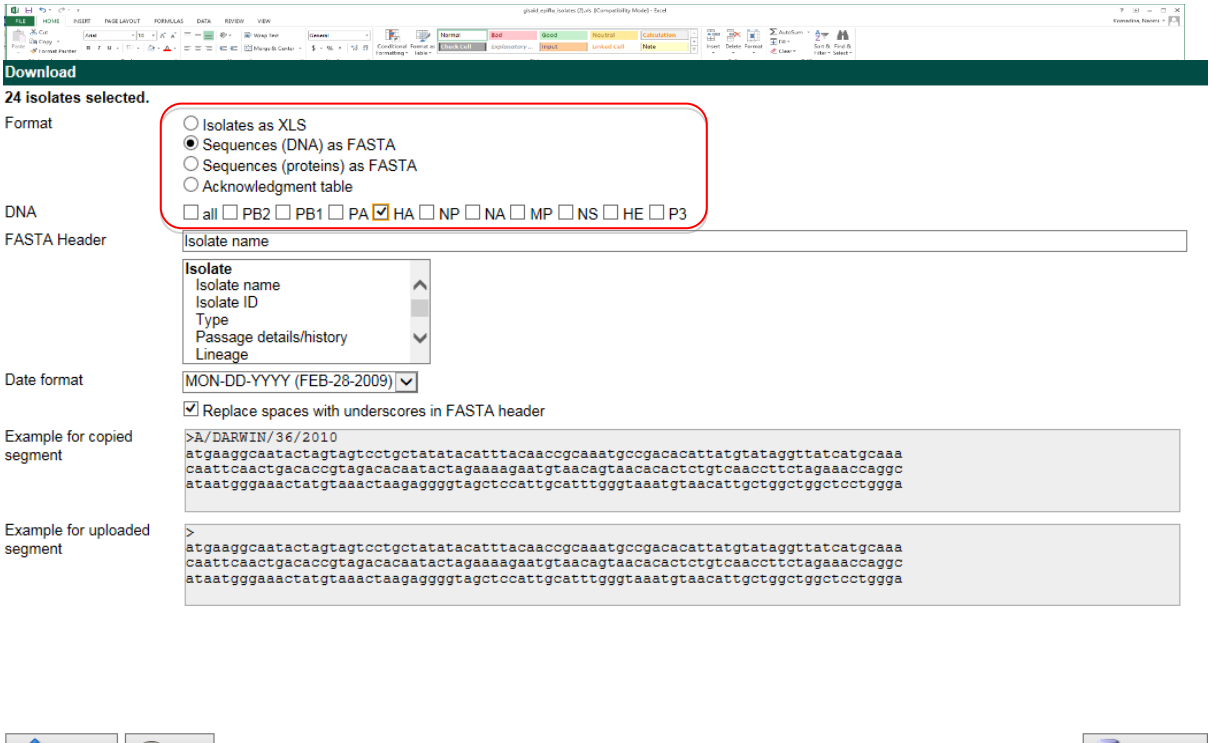


To download sequences, select genes of interest

Select other information of interest, these choices remain saved for other downloads

Select date format, this remains saved for other downloads

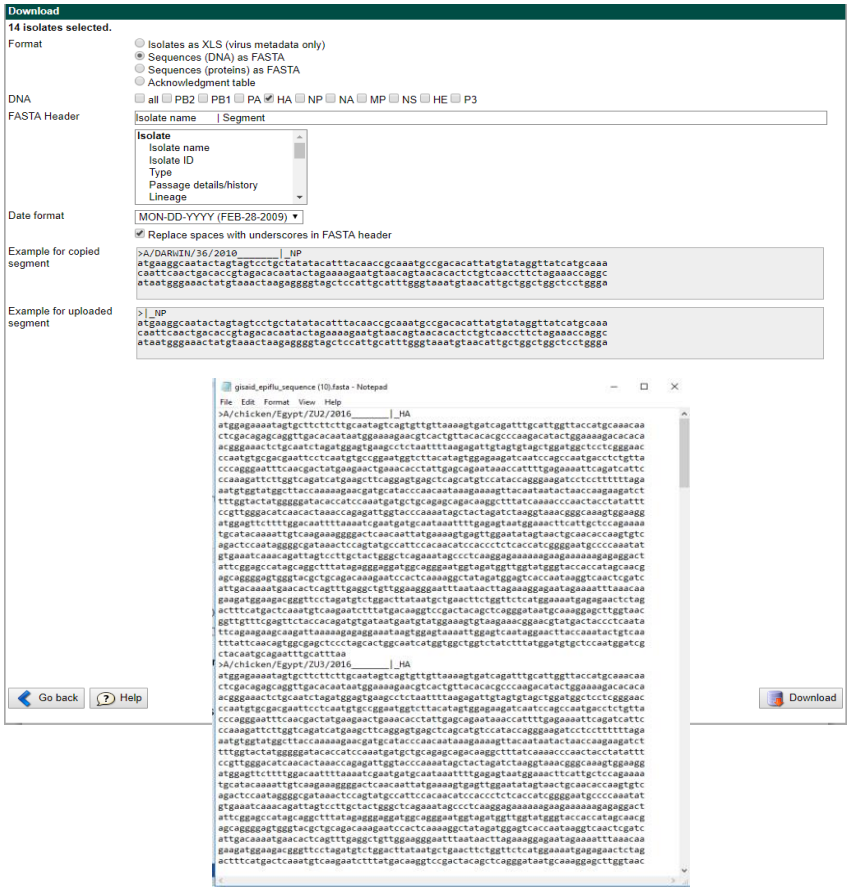
Make sure the ‘Replace spaces with underscores in FASTA header’ box has been ticked, this remains saved



Naomi.Komadina@influenzacentre.org

For DNA or protein sequence data select either ‘sequences as (DNA) FASTA’ or sequences (proteins) as FASTA.

Sequences are downloaded in a FASTA file, which can then be used in a number of analysis programs



Download Acknowledgement Table for inclusions in Supplementary Material of Publications

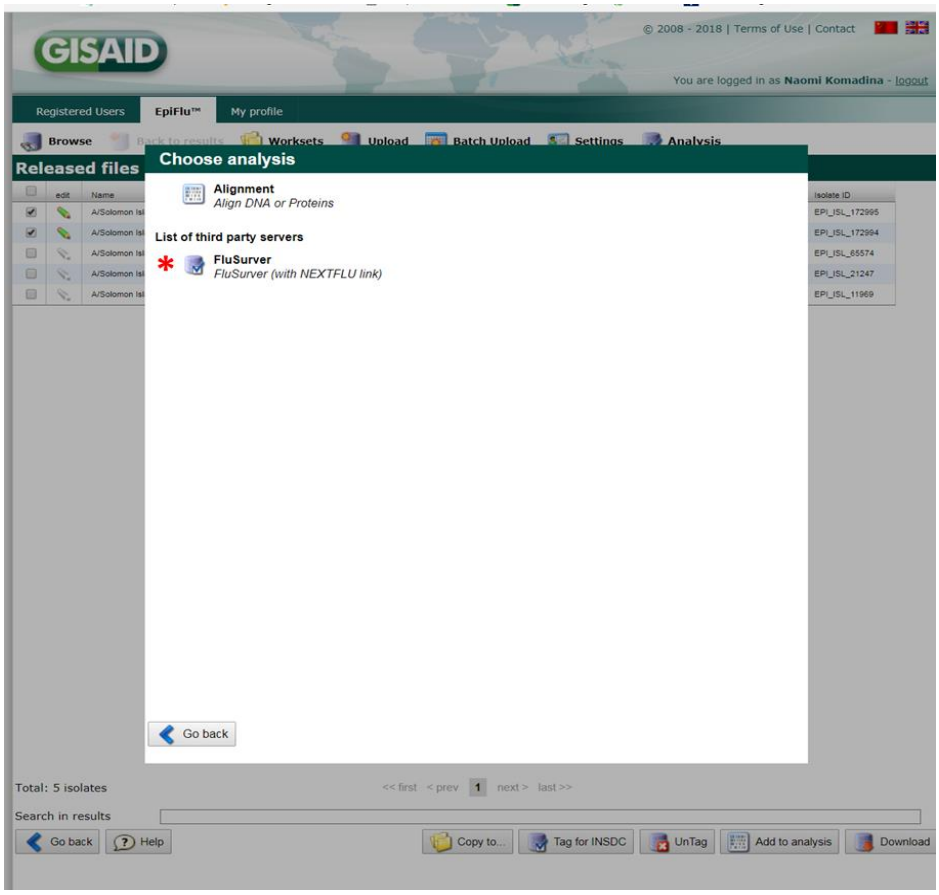
We acknowledge the authors, originating and submitting laboratories of the sequences from GISAID's EpiFlu™ Database on which this research is based. The list is detailed below.

All submitters of data may be contacted directly via the GISAID website ***www.gisaid.org***

Segment ID	Segment	Country	Collection date	Isolate name	Originating Lab	Submitting Lab	Authors
EPI1035127	HA	Egypt	2016-Jan-25	A/chicken/Gharbia/5/2016		Other Database Import	Zanaty,A.; Saleh,M.; El Shehedy,M.; El Tarabeley,M.; Abdel Fatah,S.; Selim,A.
EPI1073965	HA	Egypt	2016-Jan-15	A/chicken/Egypt/Qal-3/2016		Other Database Import	Omar,D.M.; Marden,N.A.; Gaafar,L.M.O.; El-Ebiary,E.A.; El-Dougdoug,K.A.; Othman,B.A.; Arafa,A.A.; Hussein,H.A.
EPI1300988	HA	Egypt	2016-Sep-01	A/chicken/Egypt/ME1010/2016		Other Database Import	Kilany,W.; Shehata,A.; Zain El-Abideen,M.; Elsayed,M.; Ali,A.
EPI1307766	HA	Egypt	2016-Jan-15	A/chicken/Egypt/CAL9/2016		Other Database Import	Arafa,A.; Elshazly,O.; Hussein,H.; Reda,E.
EPI1307767	HA	Egypt	2016-Jan-23	A/chicken/Egypt/S52/2016		Other Database Import	Arafa,A.; Elshazly,O.; Hussein,H.; Reda,E.
EPI1307768	HA	Egypt	2016-Nov-01	A/chicken/Egypt/CAG139/2016		Other Database Import	Arafa,A.; Elshazly,O.; Hussein,H.; Reda,E.
EPI1308276	HA	Egypt	2016-Jan-01	A/chicken/Sharkia/MEVACF3/2016		Other Database Import	Shehata,A.A.; Ali,A.; Kilany,W.H.; Sedik,M.; Elbestawy,A.; Zain El-Abideen,M.A.; Elsayed,M.
EPI891870	HA	Egypt	2016-Feb-15	A/chicken/Egypt/ZU2/2016		Other Database Import	Hussein,A.; Orabi,A.; Saleh,A.A.; Abu El-Magd,M.; Iqbal,M.
EPI891871	HA	Egypt	2016-Feb-15	A/chicken/Egypt/ZU3/2016		Other Database Import	Hussein,A.; Orabi,A.; Saleh,A.A.; Abu El-Magd,M.; Iqbal,M.
EPI891872	HA	Egypt	2016-Feb-15	A/chicken/Egypt/ZU28/2016		Other Database Import	Hussein,A.; Orabi,A.; Saleh,A.A.; Abu El-Magd,M.; Iqbal,M.
EPI953352	HA	Egypt	2016-May-08	A/chicken/Egypt/N12643B/2016		Other Database Import	Rubrum,A.; Jeevan,T.; Kayali,G.; Ali,M.A.; Kandeil,A.; El-Shesheny,R.; Webby,R.
EPI953354	HA	Egypt	2016-May-08	A/chicken/Egypt/N12642E/2016		Other Database Import	Rubrum,A.; Jeevan,T.; Kayali,G.; Ali,M.A.; Kandeil,A.; El-Shesheny,R.; Webby,R.
EPI953368	HA	Egypt	2016-Jan-23	A/chicken/Egypt/F12505C/2016		Other Database Import	Rubrum,A.; Jeevan,T.; Kayali,G.; Ali,M.A.; Kandeil,A.; El-Shesheny,R.; Webby,R.

Data can be added to a third party Server for analysis

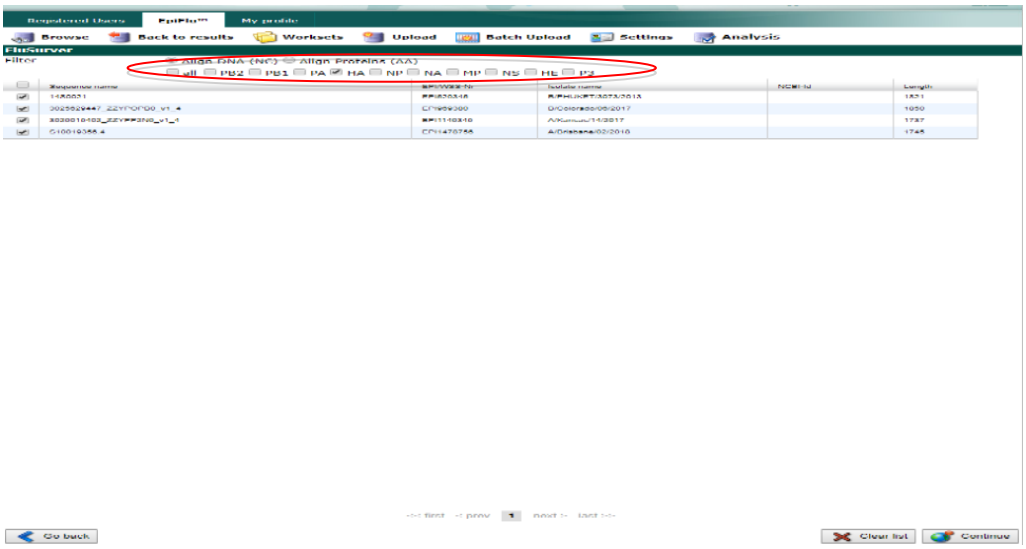
Naomi.Komadina@influenzacentre.org



FluSurver compares your sequences of interest and summarises amino acid changes to a reference strain.

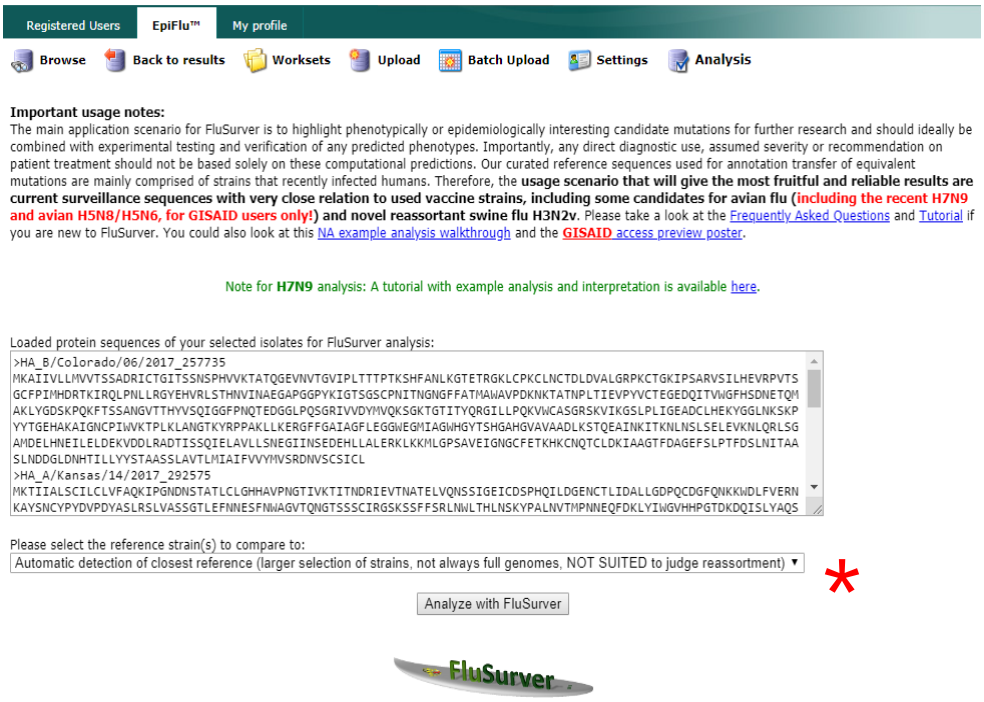
Can be used to for sites which are known to have amino acid changes associated with neuraminidase inhibition activity

To add data for analysis in FluSurver, check the box for the gene of interest

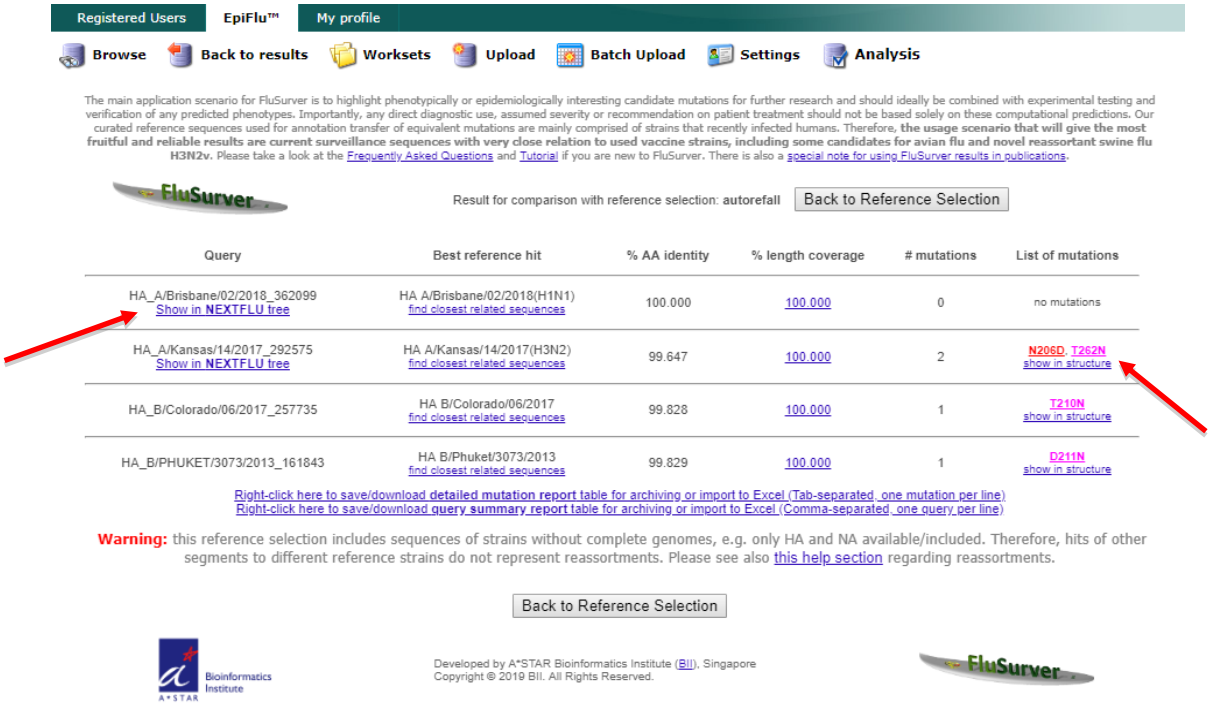


Naomi.Komadina@influenzacentre.org

Select continue and the genes of interest are inserted into the FluSurver program

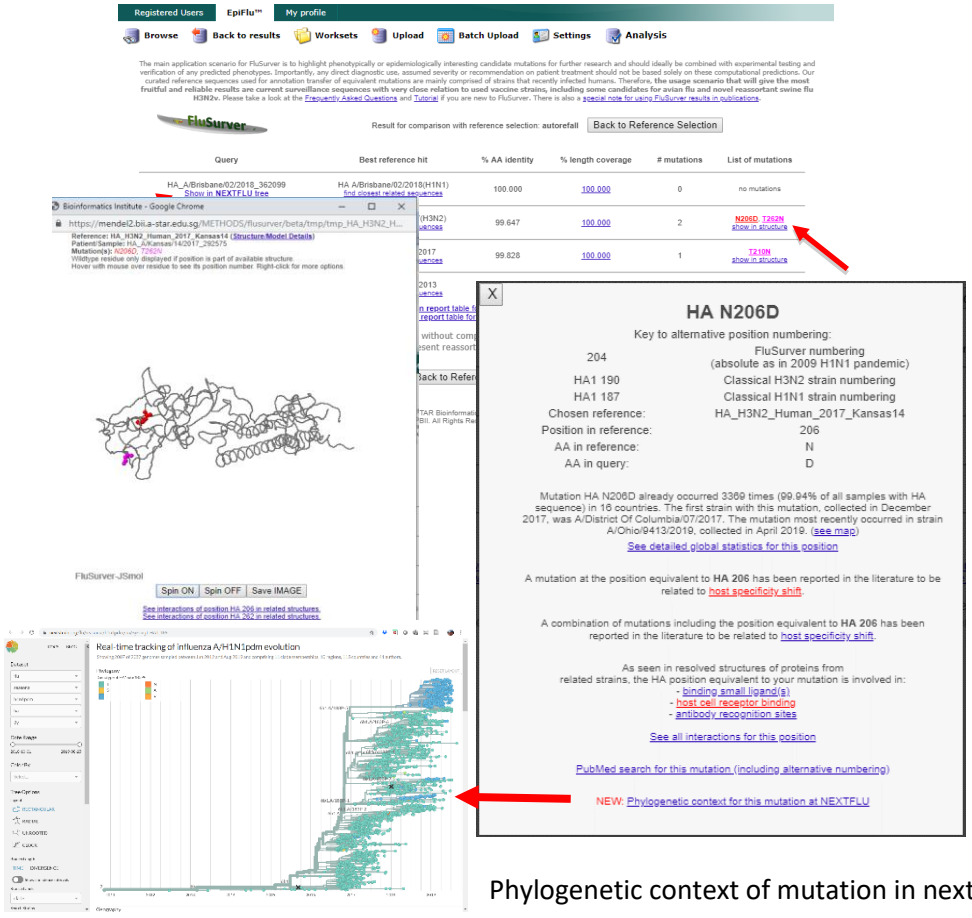


Select Analyse with FluSurver, can also select which reference strain you wish to compare to



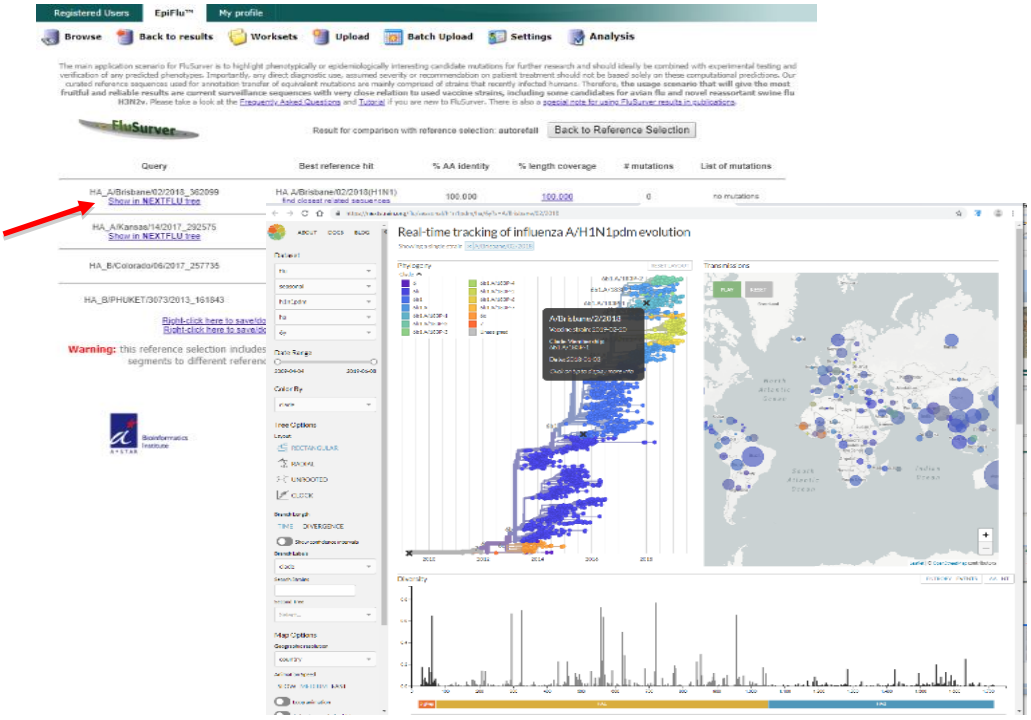
Naomi.Komadina@influenzacentre.org

Click on the underlined fields to find information, view the structure in 3D and to see a list of important sites related to drug resistance.

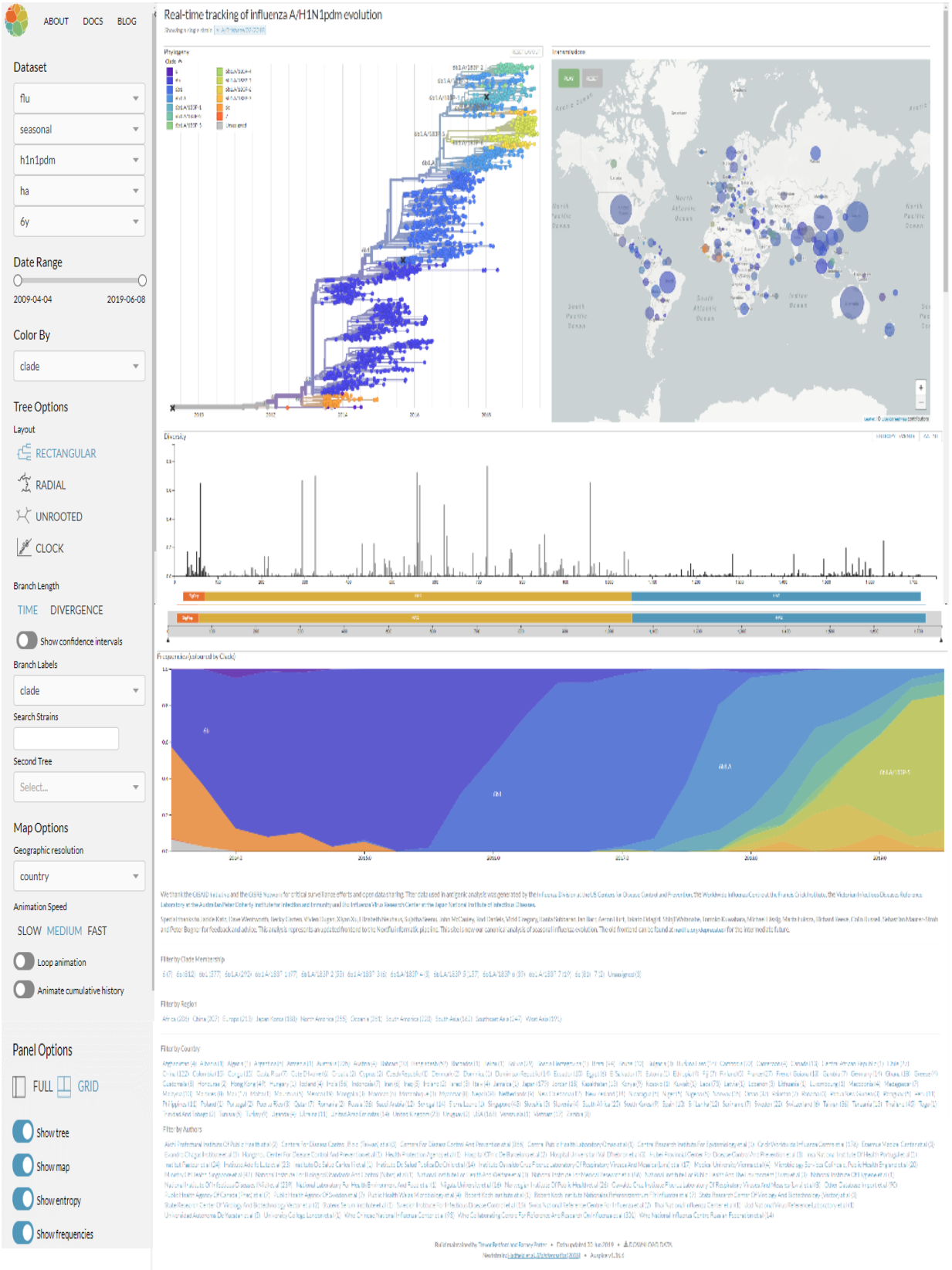


Phylogenetic context of mutation in nextflu

Select ‘Show in nextflu’ to see virus in phylogenetic tree



Naomi.Komadina@influenzacentre.org



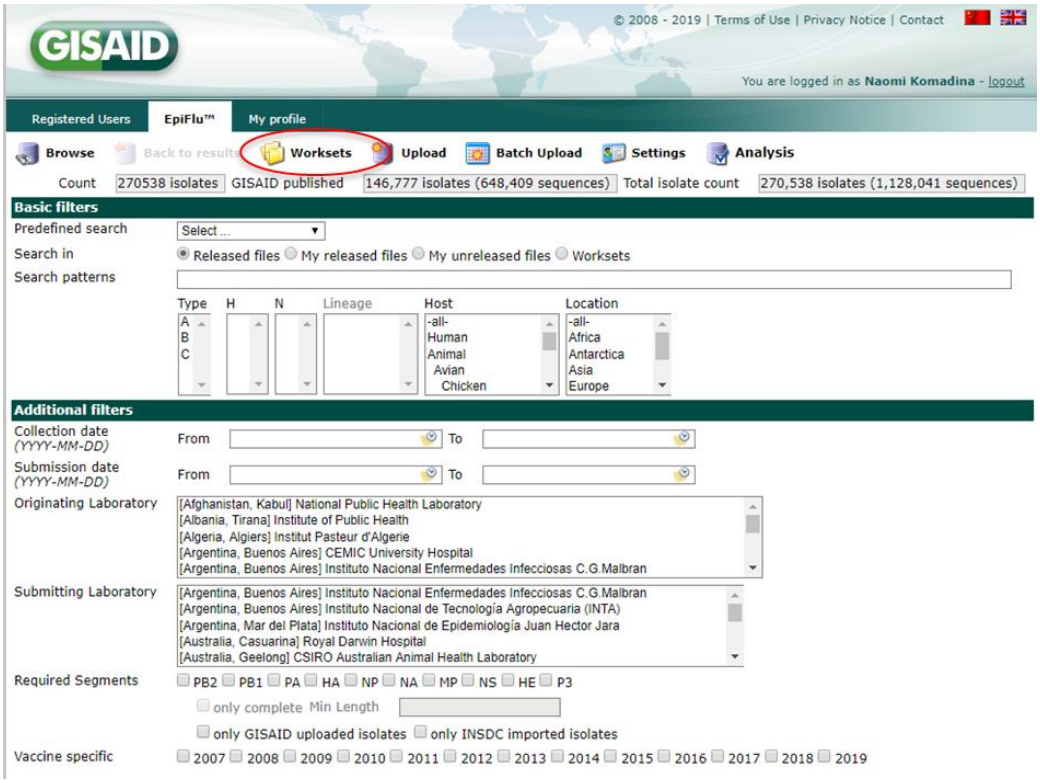
Naomi.Komadina@influenzacentre.org

How to Create Work-sets

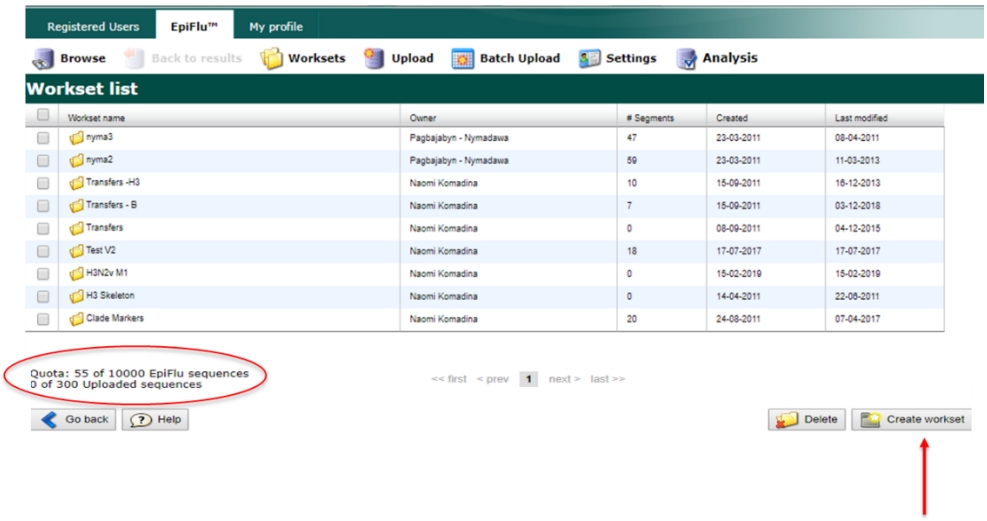
Work-sets can be shared between both Institute colleagues or between colleagues in other Institutes within the country or internationally.

Useful for storing frequently used datasets and inter-institution collaborative work. Up to 10,000 sequences can be stored in the Workset folder.

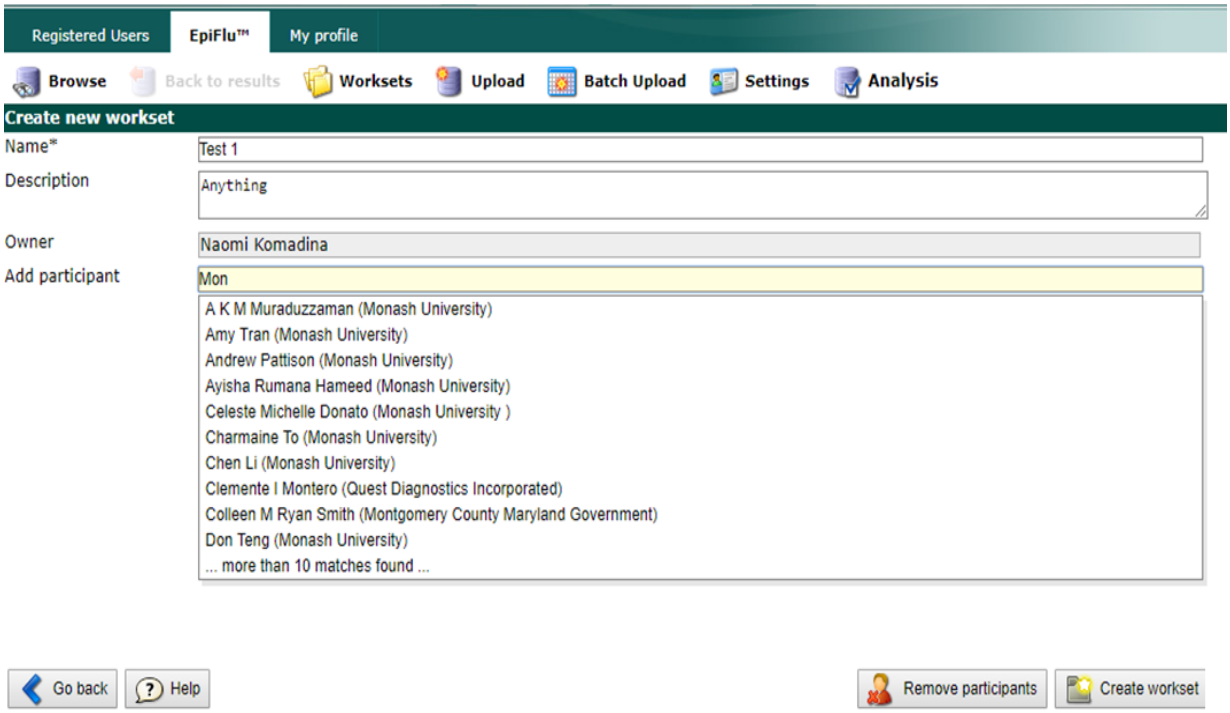
On front page click on the Worksets option



Select 'Create Workset"



Naomi.Komadina@influenzacentre.org



Name the workset and add a description if you wish.

To invite collaborators, begin typing in a name and select the name of your collaborator.

Only registered GISAID users can be selected.

The person who created the workset is the owner and can add or delete collaborators or delete the entire workset.

Collaborators can add sequences to the workset.

Sequences can be uploaded to the workset from your PC

Uploading sequences to GISAID.

Sequences can be uploaded using the Single Upload or the Batch Upload function.

A single upload must be performed before the Batch Upload sheet is used. This creates your “My unreleased files” folder.

When you login to EpiFlu™ to upload data, your details as the submitter are automatically entered.

GISAID users are able to contact isolate submitters by using the ‘Contact Submitter’ function at the bottom of the Isolate information page. Your email is hidden from view and not visible to those who contact you through this button.

Duplicates Isolates i.e. Have the same name, the same collection date and the same passage will not be uploaded.

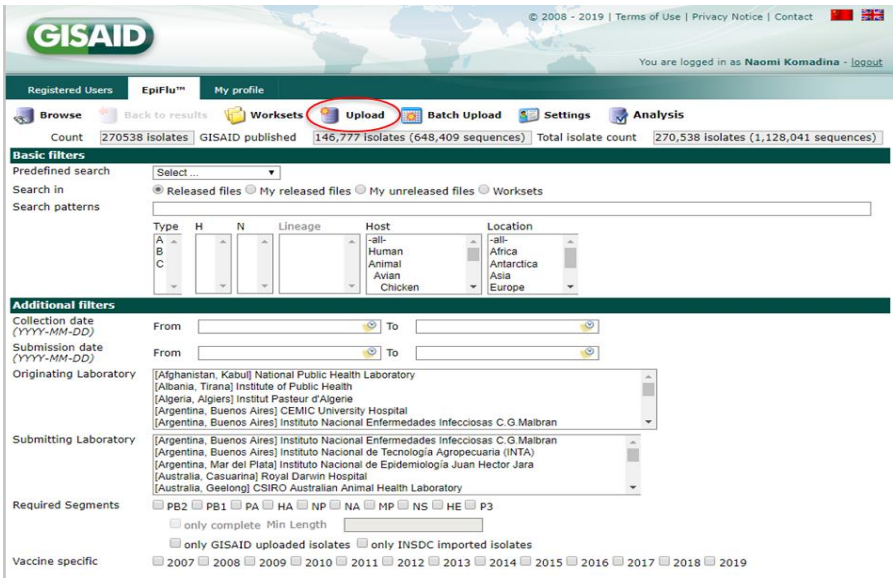
Isolate data is added in only once, sequences are attached to the isolate

Your isolates will be flagged with a green pencil on the LHS of search results.

The green pencil is the 'edit' function allowing you to add further gene sequences & metadata.

Single Uploads

Use this function if you are uploading a very small number of isolates to EpiFlu™



Click on the upload button

The Isolate ID is automatically generated

Begin filling out the form, use the drop down menus where available.

Fill in as much information that you have. Metadata is valuable to researchers.

Fields that have * are mandatory.

When you select a region i.e. Africa a second menu opens up for Country selection

In Institute information, select the region & country then open the 'Originating Lab' filed and a list of laboratories in that country opens, select the correct submitting laboratory.

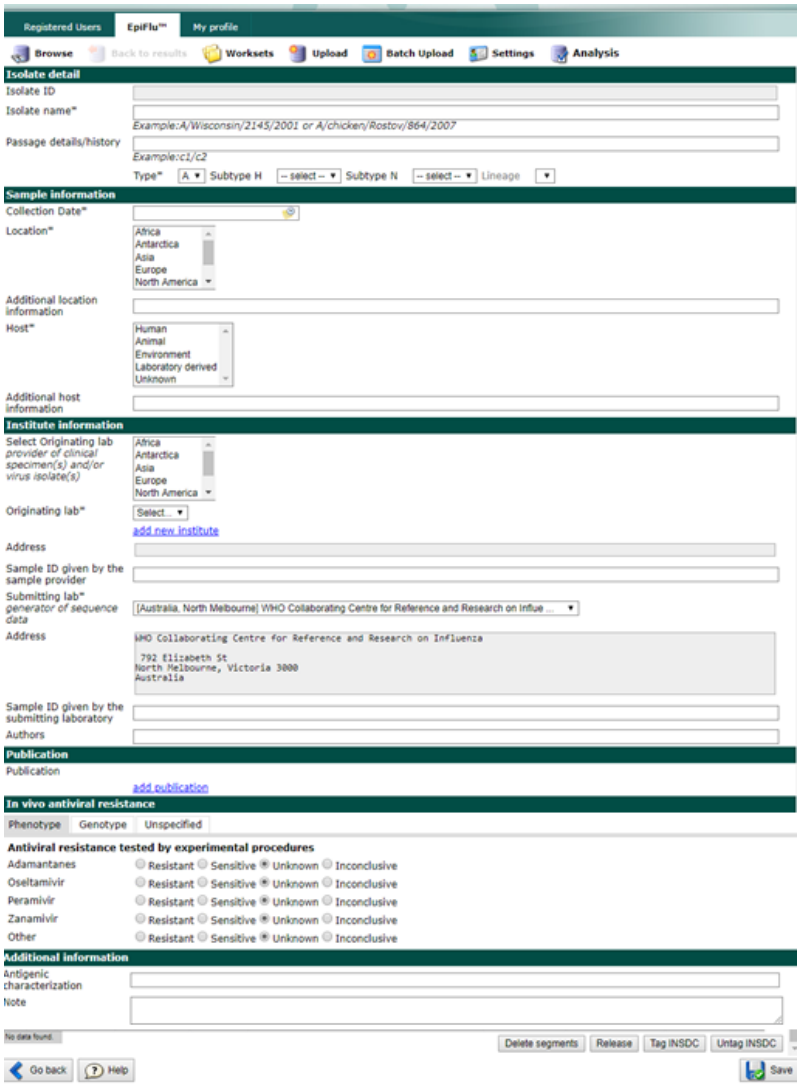
You will notice that your Institute details as the submitter have been already filled in.

Once you have completed all the Isolate information, select the save button at the bottom RHS.

If any mandatory information has been missed it will be flagged and must be filled in before you can move to the next section.

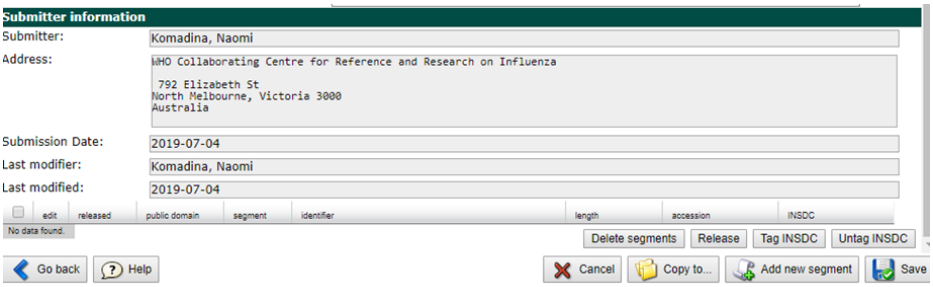
This page creates the 'Isolate information' to which sequences can be added to.

Naomi.Komadina@influenzacentre.org



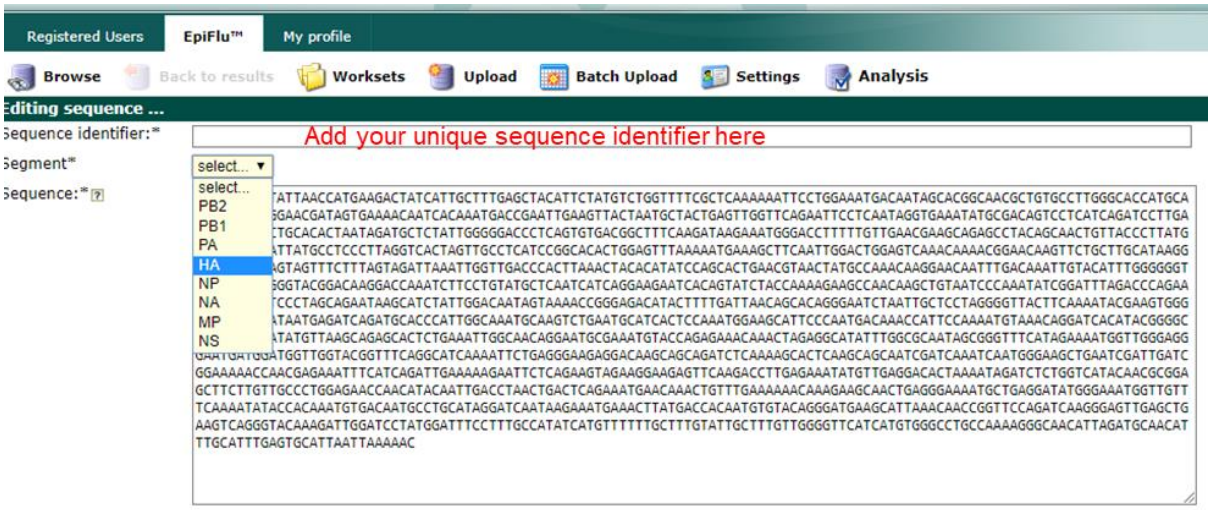
Once the isolate has been saved, this panel appears at the bottom of the “Isolate information”

Select the “Add new segment” button at the bottom of the RHS to add sequences to the isolate.



Add sequence to the following page

Naomi.Komadina@influenzacentre.org



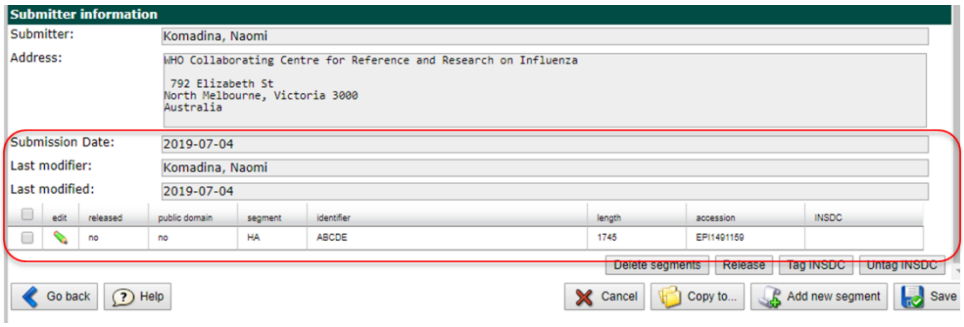
Add your ID in the Sequence identifier field

Select the segment you are adding, then paste in the sequence and select “Save’ at the bottom RHS.

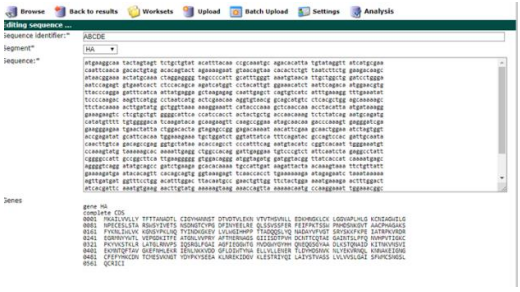
If you have selected HA and pasted in the NA sequence this will be flagged as an error which needs to be fixed.

If there is an error in the sequence which places a stop codon in the wrong place, this is also flagged and can be amended immediately.

Once saved the Isolate page reopens ready to add a new segment. The bottom of the page now has the segments available added as well as the upload date and the date it was last amended.

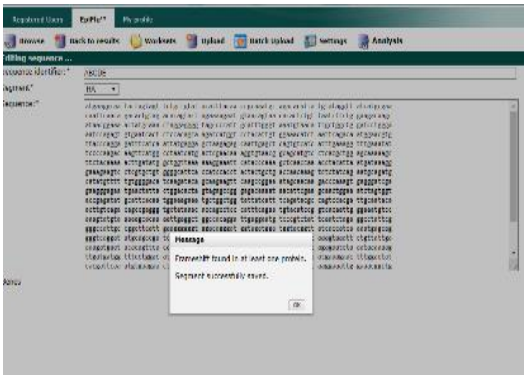


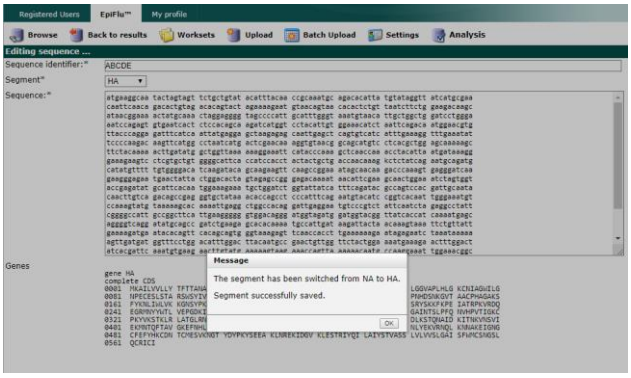
A correct sequence appears as below. The Isolate has correct data, with correct sequence and needs to be released to become publicly accessible to registered users of GISAID.

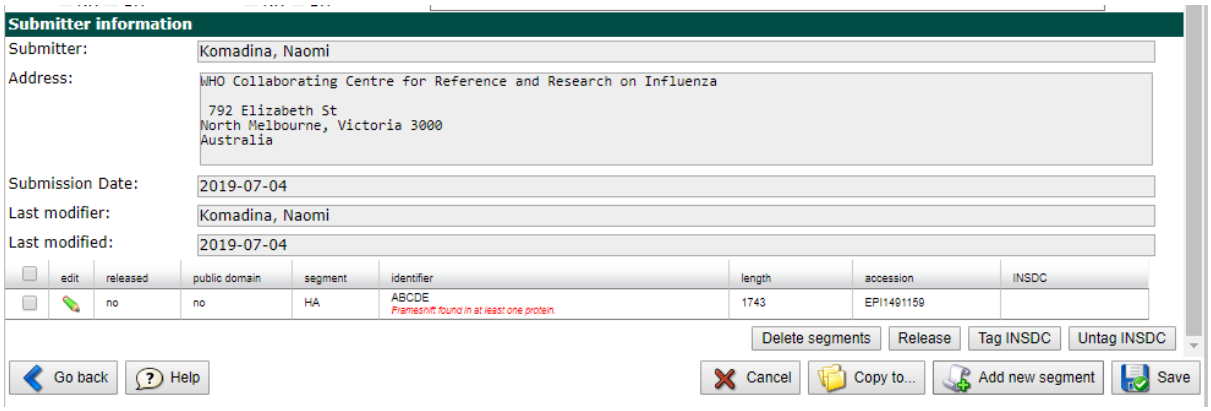


Naomi.Komadina@influenzacentre.org

Examples of flagged errors

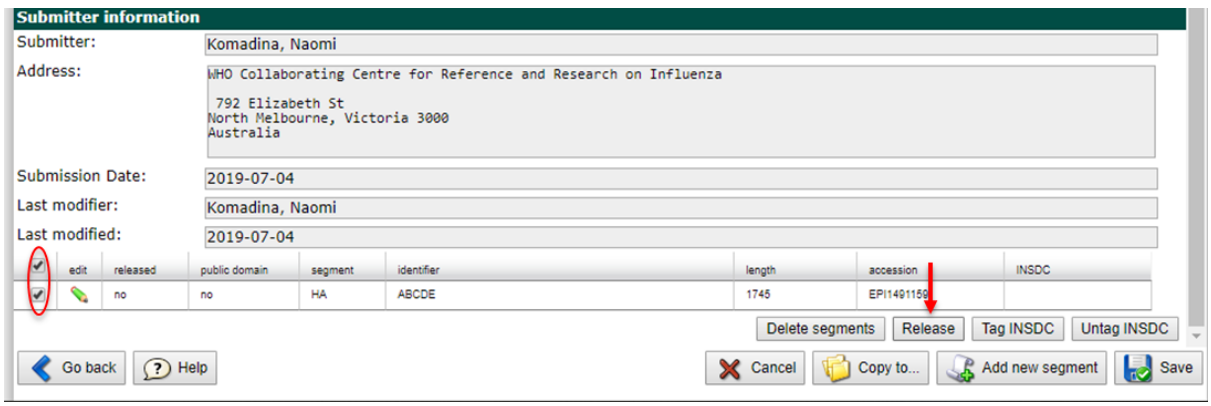






Yellow triangle indicates there is an error

How to release to EPiFlu™ Database



Naomi.Komadina@influenzacentre.org

Select the tick box on LH, either for each virus or for all, select the Release button on the RHS.

Isolates disappear from ‘My unreleased files’ and are transferred to both the ‘Released files’ and ‘My released files”

If an isolate has been released and you find that there is a problem with the isolate it can be removed from the Released Files back to My unreleased files by contacting service@gisaid.org and requesting that the isolate be ‘unreleased’ , supply the name & the EPI_ISL number of the isolate in question.

Batch Uploads

To use the Batch Upload Sheet, select Batch upload sheet, accept the T&Cs then download the sheet.

When the sheet opens up

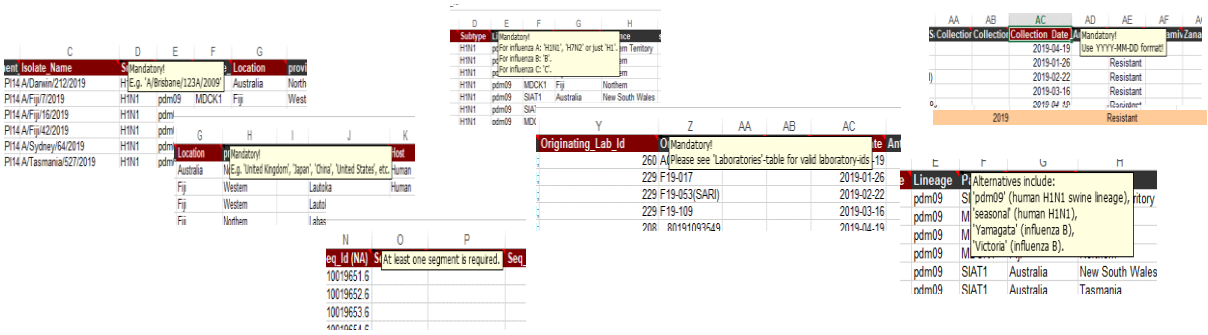
Naomi.Komadina@influenzacentre.org

Fill out the information for each isolate per line as per the example above. The first 2 columns will remain blank until they data has been uploaded when the EPI-ISL numbers will have been inserted.

All fields with a red coloured header are mandatory and must be completed

Fields with a grey header are optional, however filling out as many fields as you have information allows for better research opportunities

Hover over the red triangle in each field for pop-up with information of the format and data required, see examples below



Laboratory number for Originating laboratory are available in Alphabetical order by selecting Tab marked “Laboratories”

> Adding the Unique Originating Laboratory number to The Field named “Originating_Lab_ID” autofills the information on the Isolate page

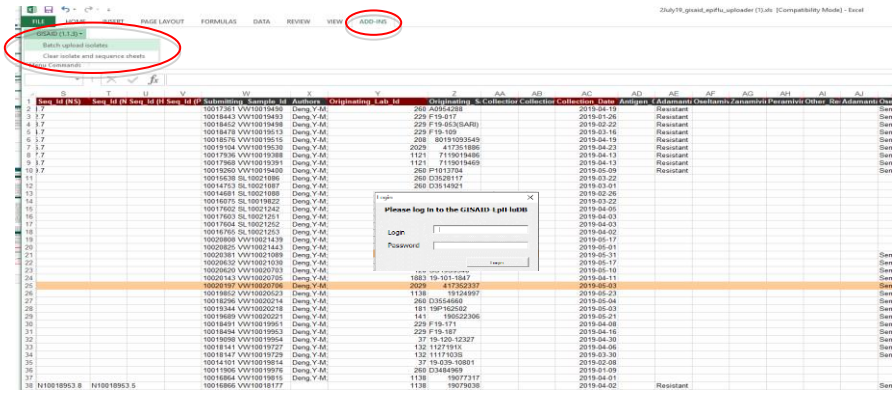
To add sequences click on Tab marked “Sequences”

> Paste in FASTA file, the page turns purple when the sequences have been pasted in

Once completed, select the “Add-Ins” tab at the top of the page, the GISAID menu at the top of the LHS of the page drops down.

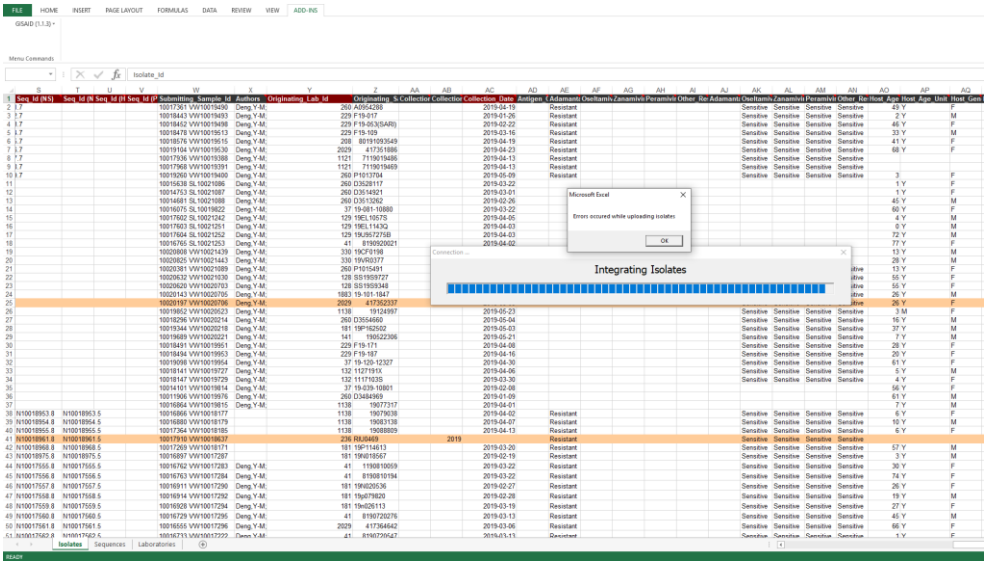
Select “Batch upload isolates” in the drop down menu and the GISAID Login feature opens up.

Enter you Login and password details and select the Login button

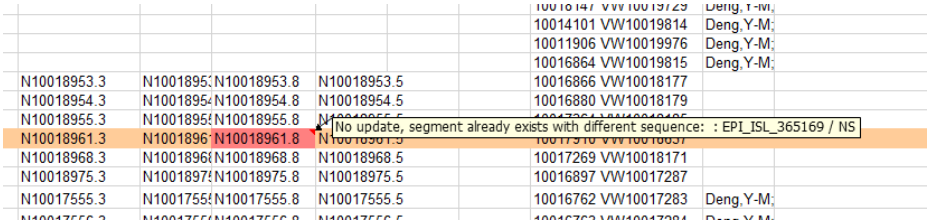


Naomi.Komadina@influenzacentre.org

If the isolates have been successfully uploaded, a message indicating that the ere has been a succesful upload, if there has been an error the message below is received.



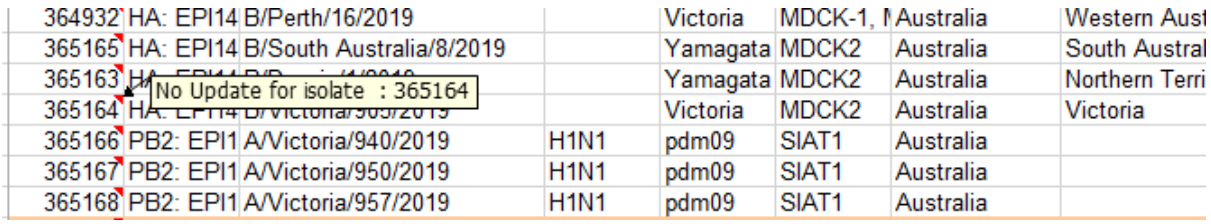
Isoaltes which have the error are flaggedin orange, the field with the error is flagged in red



Click on the red triangle in the cell top RHS and an error message appears. In the example the message relates to the fact that particular segment was not updated, this message can occur if you have uploaded a segment then re-use the same Batch-Upload sheet to add further segments.

If the sequence is missing then that is flagged in the error message.

If you are re-using a Batch Upload sheet to add extra information or sequence to incomplete data sets you will see the message below for isolates that have not been updated.



When you login and to batch upload a sheet and receive a message ‘You have no internet connection” save the Batch Upload sheet, close it, then re-open the sheet, select “Enable Content” in the yellow security warning, select Yes for making the file a “Trusted document”, select OK in the GISAID Batch Upload box, then proceed with selecting Add-Ins & uploading to GISAID as above.

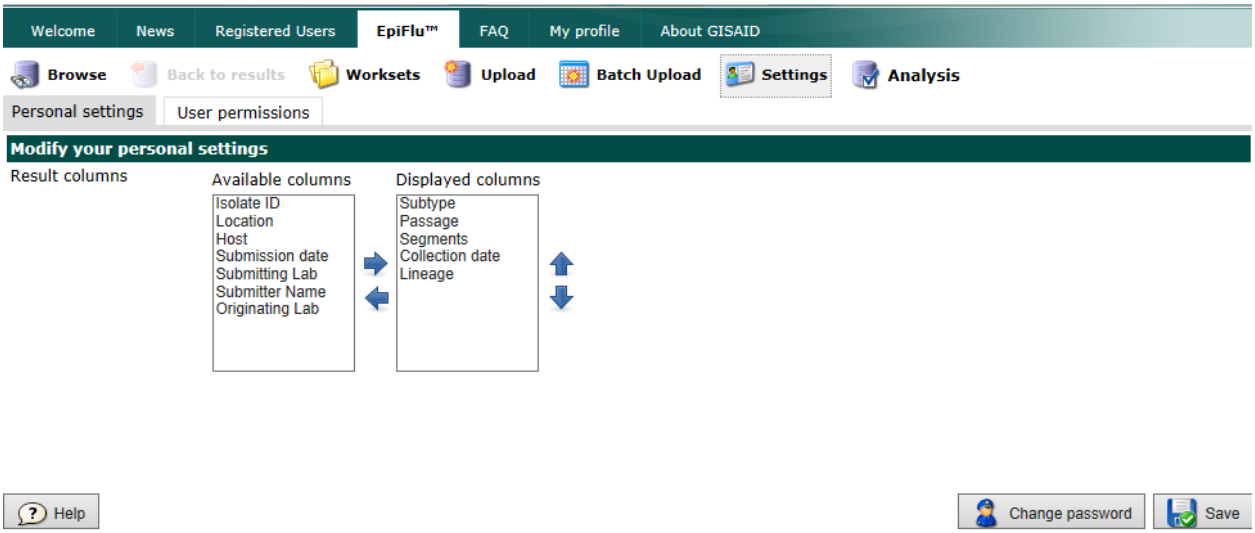
Naomi.Komadina@influenzacentre.org

Once all the isolates have been uploaded, open you “Unreleased files”, check that all is OK, then release the data to the EpiFlu™ database as for a single upload.

Settings

Select the Settings button to change your settings for viewing search returns and passwords

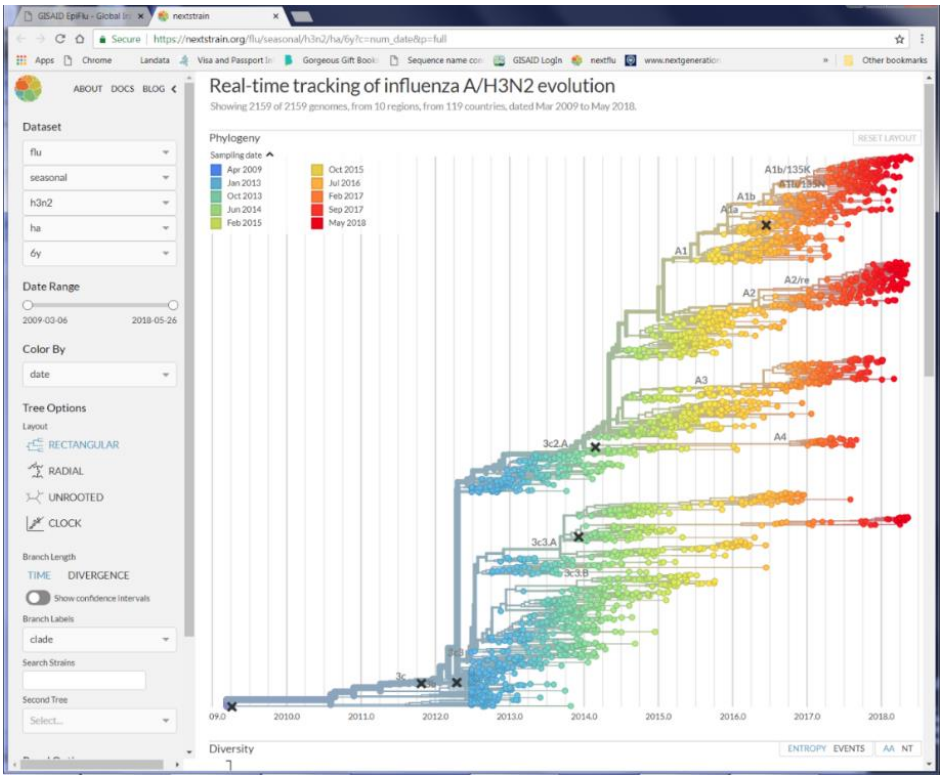
Settings can be changed any time, saved then return to what you were working on

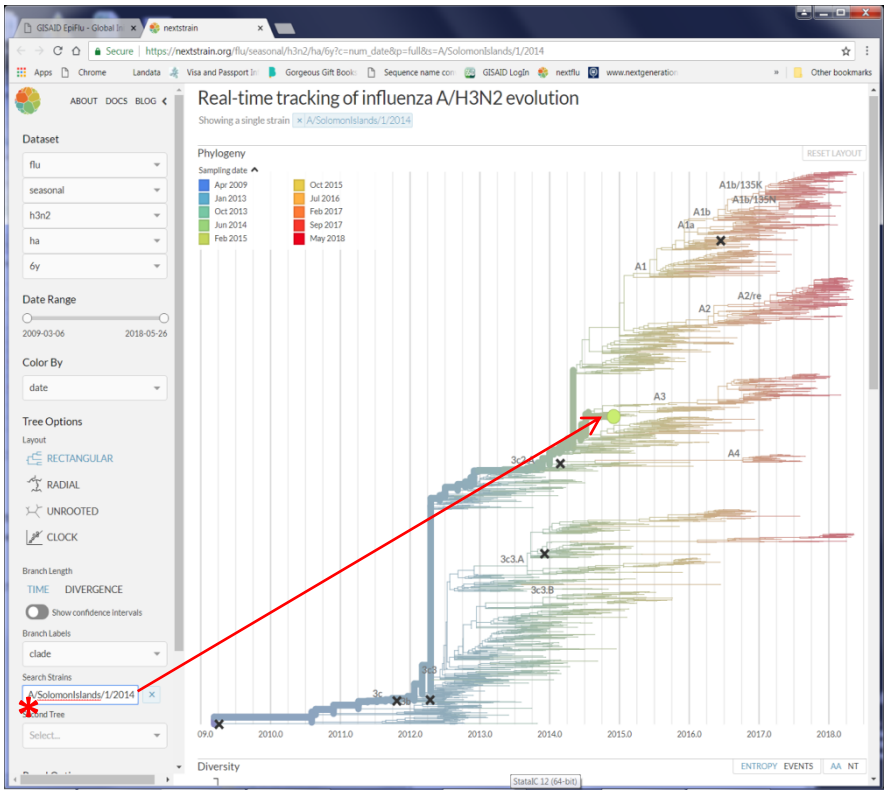


Naomi.Komadina@influenzacentre.org

nextflu

https://gisaid.org/nextflu





Naomi.Komadina@influenzacentre.org