# Sequencing and the GISAID EpiFlu<sup>TM</sup> Database

Naomi Komadina

www.influenzacentre.org



WHO Collaborating Centre for Reference and Research on Influenza VIDRL

- Introduction
  - GISAID data access and sharing principles
  - GISAID EpiFlu<sup>™</sup> features
- The new GISAID EpiFlu<sup>™</sup> 2.0 database
  - Browsing and searching EpiFlu<sup>™</sup>
  - Upload of sequences and new embargo and IVHPP functions
- Analysis of sequences
  - BLAST, Alignment and Phylogenetic tree
  - FluSurver mutation interpretation
- Outlook and conclusions

#### GISAID data access and sharing principles

Global Initiative on Sharing All Influenza Data is a non profit association

- Operates the publicly accessible EpiFlu™ database (hosted by German Gov't)
- Data must acknowledge the contribution of those providing the both the original isolate & the sequence data.
- Access to GISAID is free of charge and open to everyone, however registration is required.
- Data submitted to GISAID is publicly accessible without any loss of ownership (in contrast to deposits made to Public Domain archives e.g. GenBank).
- GISAID is the only publicly accessible database with a fair sharing principle is second-to-none for timely sharing of influenza sequences of public interest ie pdmH1N1 & H7N9.

#### **Features**

- Has the most complete set of influenza sequences
- Includes patient meta data & geographical locations
- Submitters retain editing rights
- Easy upload of data, single or batch upload templates
- Isolate based entry rather than sequence only
- Can add sequences to an isolate entry
- Automated sequence annotation on upload
- Customised search function outputs
- Customised fasta headers with sequence downloads
- Can download meta data in spreadsheet format
- Worksets function, can be shared between collaborators
- Analysis tools

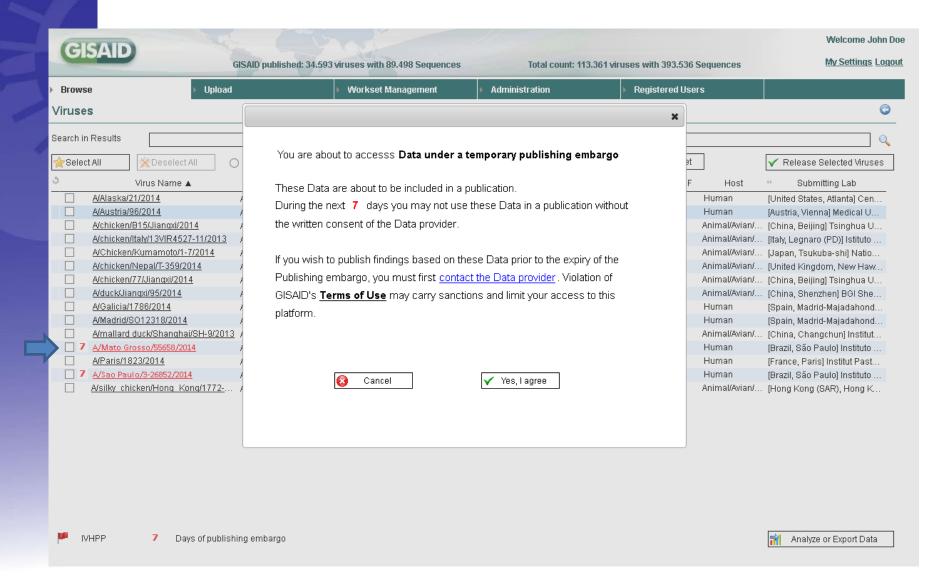
# GISAID EpiFlu<sup>TM</sup> V2.0 database

GISAID		GISAID published: 34.5	93 viruse:	riruses with 89.498 Sequences Total count: 113.361 viruses with 393.536 Sequences						equences	Welcome John Doe		
▶ Browse ▶ Up		Upload	▶ Wo	rkset Managemer	ent > Administration >			Registered Users					
Number of results: 47 Vi	ruses			<ul><li>Released</li></ul>	O My released	⊝ My u	ınreleased 🞇	Reset F	ilters	23 Count Viruses	Q Search	Results	
Search pattern													
Туре▲ - Н⊿	▲ - N ▲ - Lineage ▲				Geographic Grouping ▲		<b>A</b>		Country Name		ı		
A 3 B 4 C 5 6 7		5		д д Е П С	ifrica Intarctica Isia Iurope Iorth America Iceania				Qatar Saudi Ara Singapori Sri Lanka Syrian Ara Taiwan			^ 	
Passage Category ▲ Cell Egg Mixed Other		Passage History		8	South America				Tajikistan Thailand Turkey Turkmeni United Ara Uzbekista Vietnam Yemen	ab Emirates			
Host Group ▲		Host Subgroup ▲		Hos	t Family ▲	-	Host Spe	cies 🛦		Host Si	ubspecies		
Animal Environment Human Laboratory derived Unknown		Avian mammals		Chicken Duck Eagle Falcon Goose		* * * * * * * * * * * * * * * * * * *			<b>V</b>				
Collection Date from	2013-03-	01 iii to		-			Orig	inating La	ab ▲				
Submission Date from to Harvest Date from to				]	[Argentina, Buenos Aires] Instituto Nacional de Enfermedades Infecciosas [Argentina, Buenos Aires] Malbran [Argentina, CABA Pcia. de Buenos Aires] Servicio de Virosis Respiratorias INEI ANLIS Carlos G. Malbran [Argentina, Capital Federal- Buenos Aires] Servicio de Virosis Respiratorias INEI-ANLIS "Carlos G. Malb [Argentina, Mar del Plata] Intituto Nacional de Epidemiologia [Armenia, Yerevan] Centre for Diseases Control and Prevention						<u> </u>		
GISAID Viruses only													

# EpiFlu<sup>TM</sup> 2.0 – Isolate detail

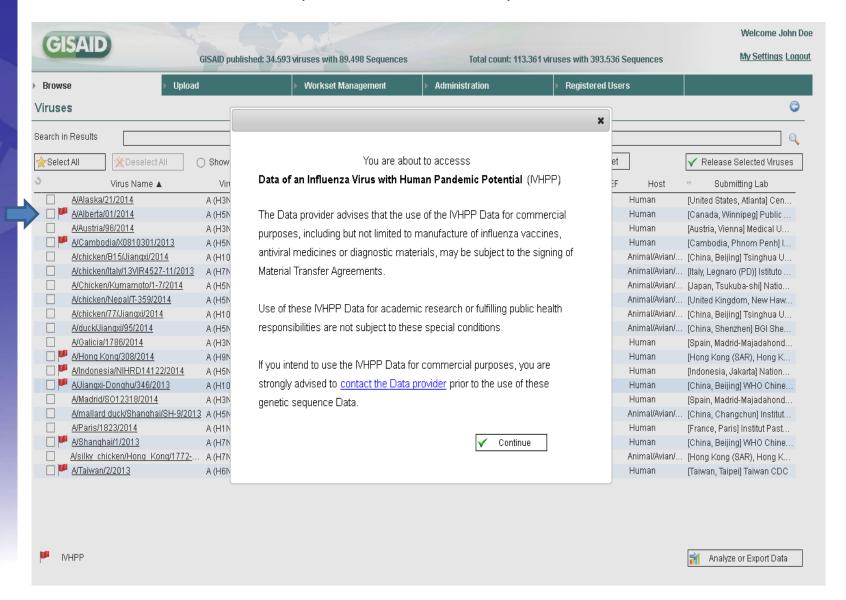
GISAID									
GISAID	GISAID p	98 Sequences	equences Total count: 113.361 viruses with 393.536 Sequences						
▶ Browse	▶ Upload	▶ Workset Mana	gement > A	dministration		Registered Users			
Virus								0	A
Virus Name	A/Texas/50/2012			V	irus Type	A / H3N2			
Virus ID	EPI_ISL_122006				Lineage				
Passage Category	Passa	ge History M1/C1		Han	vest date				
Sample Information —									
Collection Date	2012-04-15				Location	United States/Texas			
Host	Human			City	or Town				
Additional Host Information				:	Zip Code				
Patient Age		Gender Male		Coordinat	es (N, E)			Q-	
Patient Status		Last Vaccinated							
Outbreak				Antigenic Charac	terization	A/PERTH/16/2009-LIKE (H3N2	) GP		
Specimen Source					Note				
In-Vivo Pathogenicity Test									
Sequence -									Ε
	Segment	Identifier		Length		AID Accession Number	INSDC Accession Number		
		94_217304_v1_4 94_217304_v1_6		1701 1410	EPI37749				
		94_217304_v1_7		984	EPI37749				
Antiviral Susceptibility	/Antiviral ▲	Susceptibility Level							
	Laninamivir	Normal							
	Oseltamivir Peramivir	Normal Normal							
	Zanamivir	Normal							
WHO Reference Inform	mation —								
		e Strain NH Vaccine Strain SH							
	2014 2013	<b>▽</b>	<b>▽</b>	<b>√</b>					
	2013	V	(V)						
Institute Information				Odininakia a Lab	Address				
Originating Lab Originating Sample ID		Health Services-Laboratory Service	es	Originating Lab	Address	1100 W. 49th St. 78756-9987 Austin, United Sta	tes		
Submitting Lab		rol and Prevention		Submitting Lab	Address	1600 Clifton Road, N.E.			
Submitting Sample ID						30333 Atlanta, United States			+

#### EpiFlu<sup>TM</sup> 2.0 – New publication embargo function

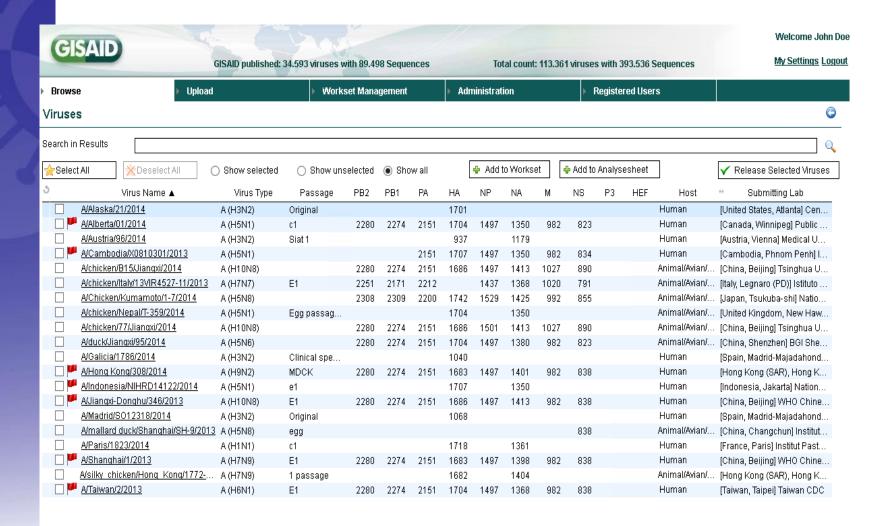


## EpiFlu<sup>™</sup> 2.0 – New IVHPP function

(WHO PIP framework)

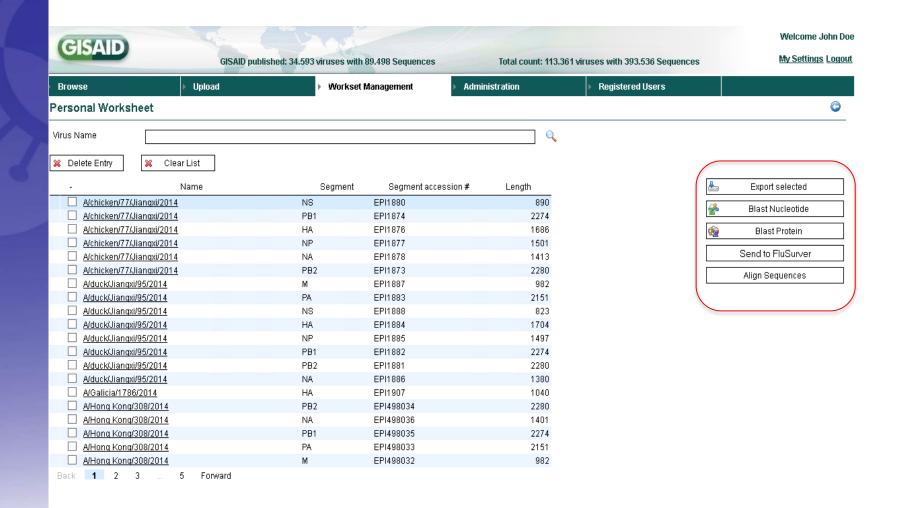


#### EpiFlu<sup>™</sup> 2.0 – Worksets for Sharing & Analysis

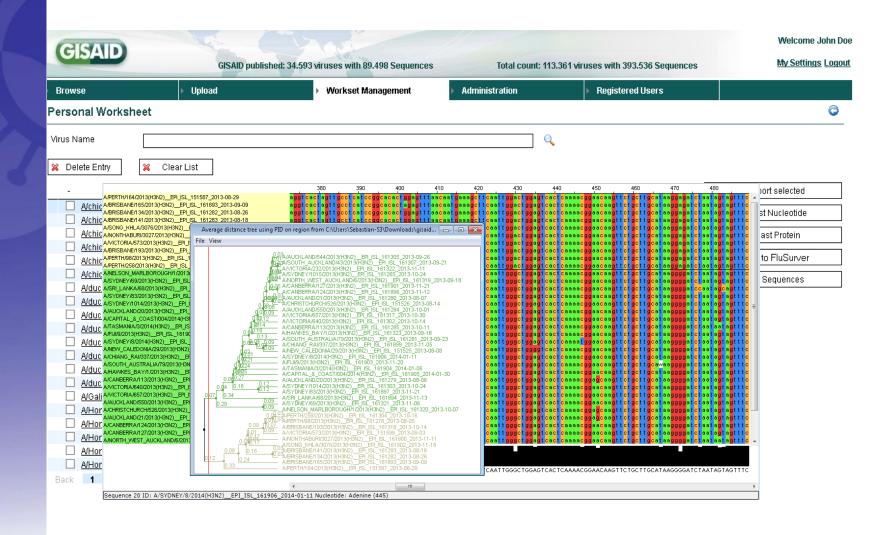




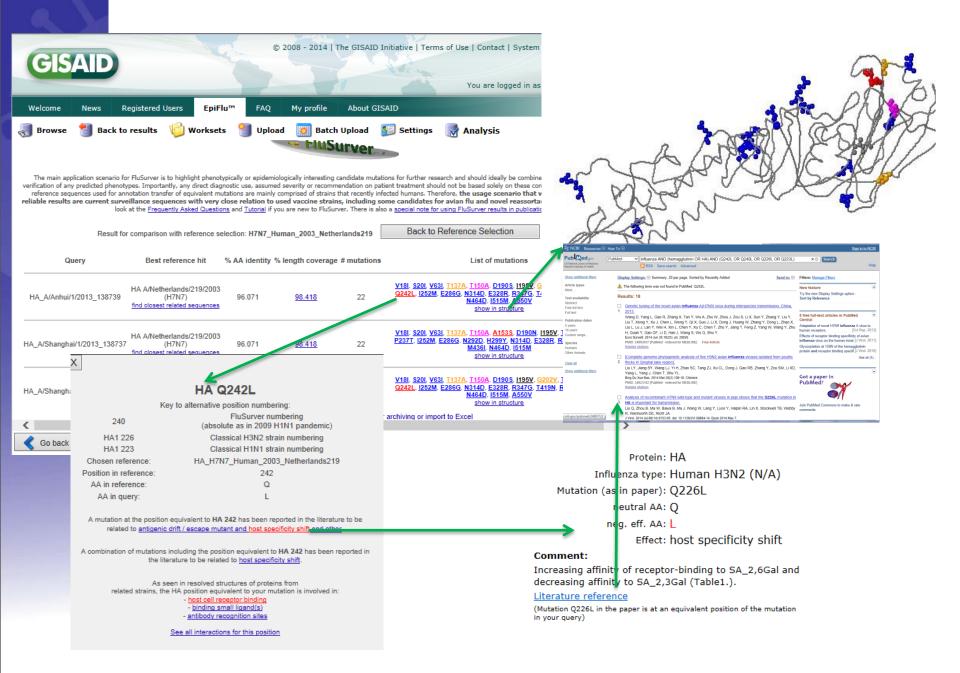
## EpiFlu<sup>™</sup> 2.0 – Analysis Tools



#### Analysis - BLAST, Alignment, Phylogenetic tree



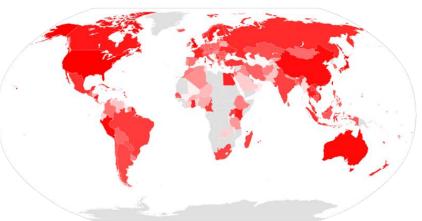
#### Analysis – FluSurver for Mutation Interpretation

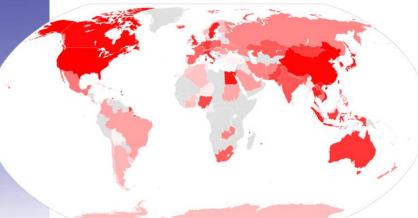


#### **GISAID** Outlook and Future

Strong and rising contributions from Asia

Global distribution of human host derived isolates in GISAID





Global distribution of animal host derived isolates in GISAID

Countries ranked by number of HA sequences in GISAID in August 2014, colored by rank percentile from red (most) to white (least), gray (none)

#### **GISAID** Outlook and Future

- New sequencing technologies in influenza surveillance will give us:
  - More sequences, more detail, more to interpret...
- New features
  - Annotation to include H5 clade designation on upload
  - Updated worksets with more options
  - New search functions
  - Biannual vaccine recommendations
  - GPS Coordinates
  - User tracking on sensitive pandemic potential viruses
- GISAID allows for easy, safe and timely sharing even of highly sensitive new outbreak sequences

# Acknowledgements



**GISAID** Database Technical Group



Bioinformatics Institute, A\*STAR, Singapore



Berlin, Germany

hosted by Germany's





WHO Collaborating Centre for Reference and Research on Influenza VIDRL