Comparative Sequence Analysis of Circulating and Pandemic Flu Viruses in Swine and Humans

Alicia Lee

CNIC, Chinese Academy of Sciences
Beijing, China

Final Report

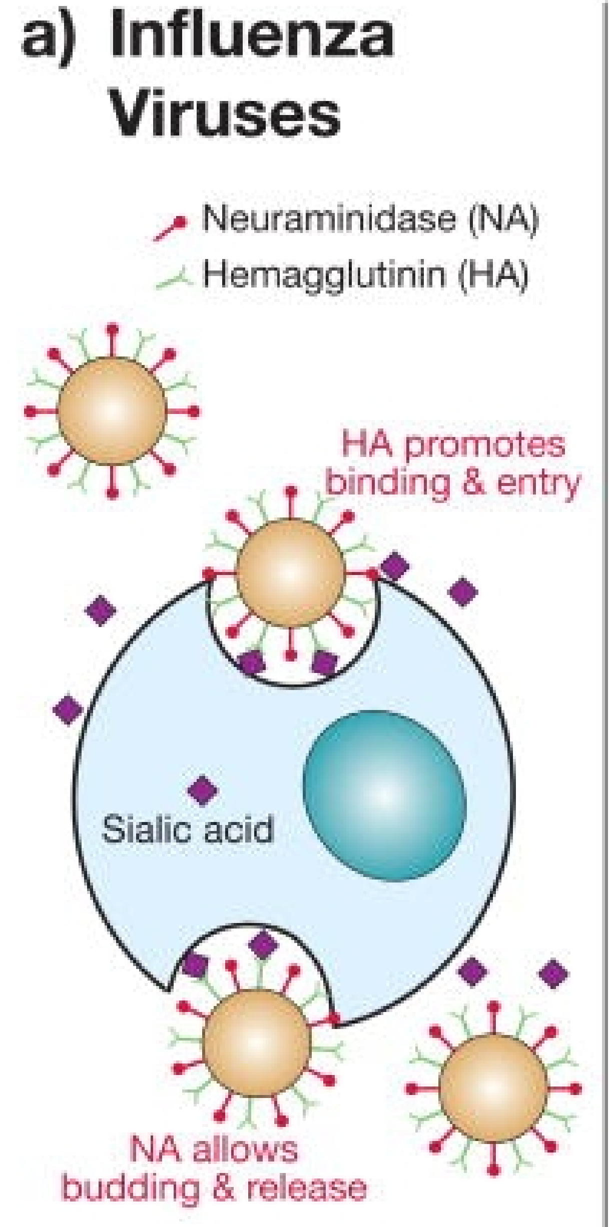
August 25, 2010

Influenza

- Influenza, more commonly known as the flu, is an RNA virus that primarily affects the respiratory system.
- Influenza viruses are named for their surface proteins hemagglutinin (HA) and neuraminidase (NA).

Proteins

- Hemagglutinin (HA)
- Neuraminidase (NA)
- Receptor Proteins responsible for entry and exit of virus cells
- HA – binding and entry
- NA – budding and release



http://www.ncbi.nlm.nih.gov/bookshelf/br.fcgi?book=glyco2&part=ch39

Mutations

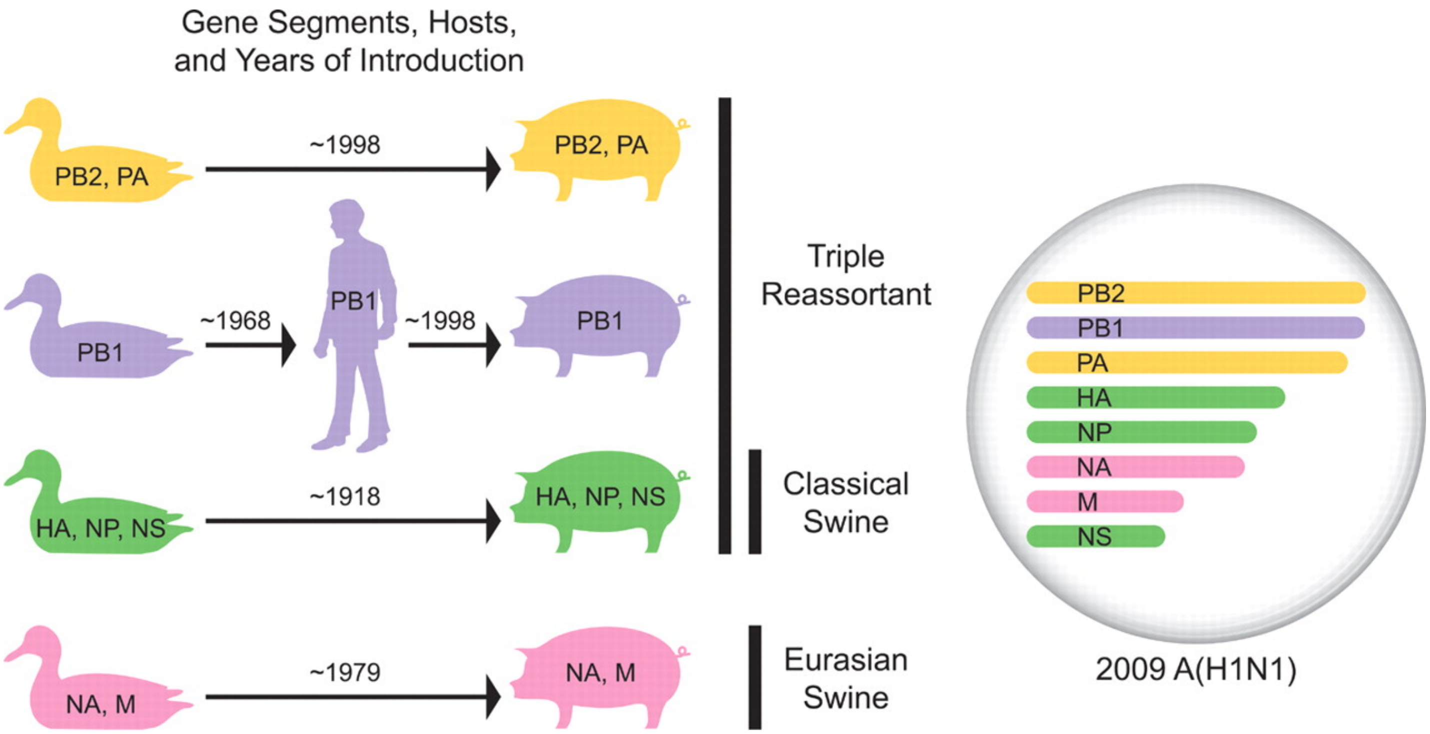
- **Antigenic drift** – RNA sequence has mutated the HA or NA protein enough to become immune to previous vaccinations.
- **Antigenic shift** – two or more different influenza viruses affect the same host, leading to a reassortment of genomic segments, creating a new virus against which humans have no pre-existing immunity.

Pandemic vs Seasonal viruses

- Pandemic strain
  - Infects many people around the world
  - Easily to spread; more contagious
  - 4 pandemics
    - 1918 Spanish Flu (H1N1)
    - 1957 Avian Flu (H2N2)
    - 1968 Hong Kong Flu (H3N2)
    - 2009 Swine Flu (H1N1)
- Seasonal (non-pandemic) strain
  - Localized, more regular occurrences

2009 Swine Flu Strain

- The 2009 swine flu strain (SO-IAV) is a ‘triple reassortant’, with its hemagglutinin gene from the swine lineage.



http://www.sciencemag.org/cgi/content/full/325/5937/197

- Important to monitor currently circulating swine strains

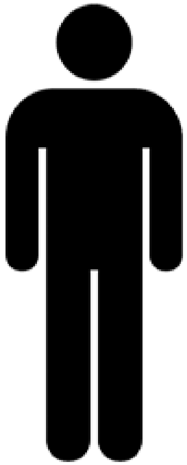
Specific Aims

1. Identify signature motifs found in H1N1 pandemic strains affecting humans
   - compare them to sequences found in current swine flu strains circulating among swine using protein sequence analysis

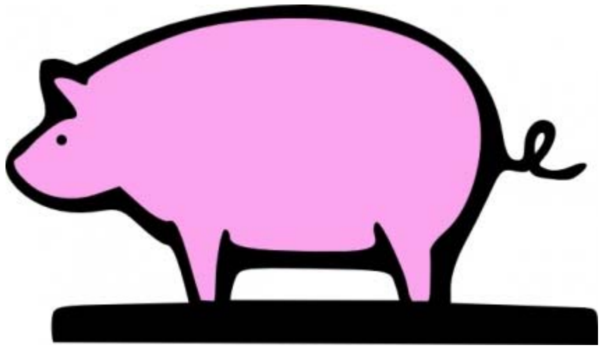
2. Model these mutations using VMD or Chimera
   - visualize protein-ligand interactions and measure binding affinities to human receptor analogs using AutoDock suite tools.

Modified Aims

1. Identify signature motifs, or highly conserved regions, found in H1N1 pandemic strains affecting humans
   - Compare to currently circulating swine strains and seasonal strains focusing on immune epitopes

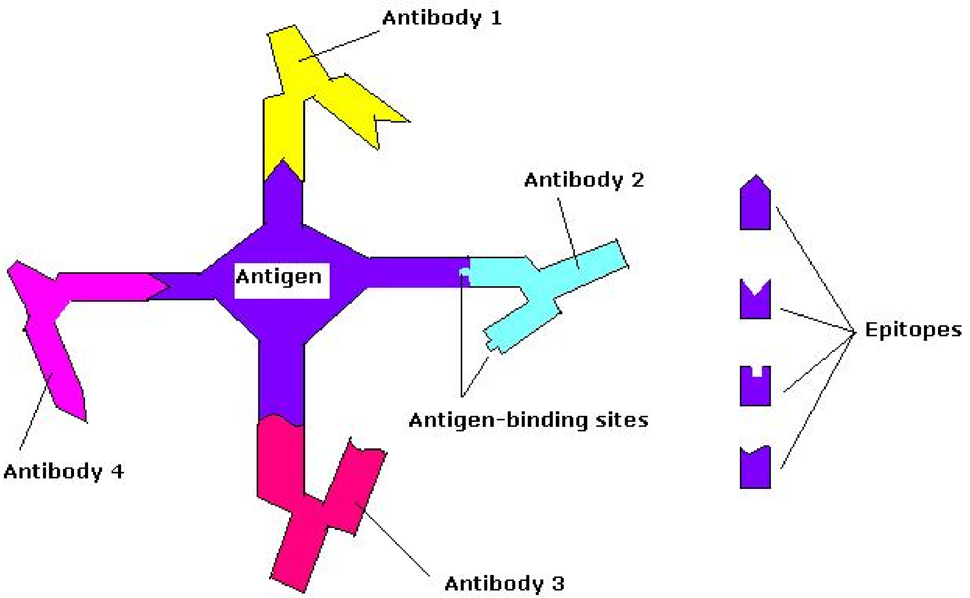


vs



Epitopes

- Parts of the virus which are recognized by antibodies or T cells
- Important mutations occur at the epitopes

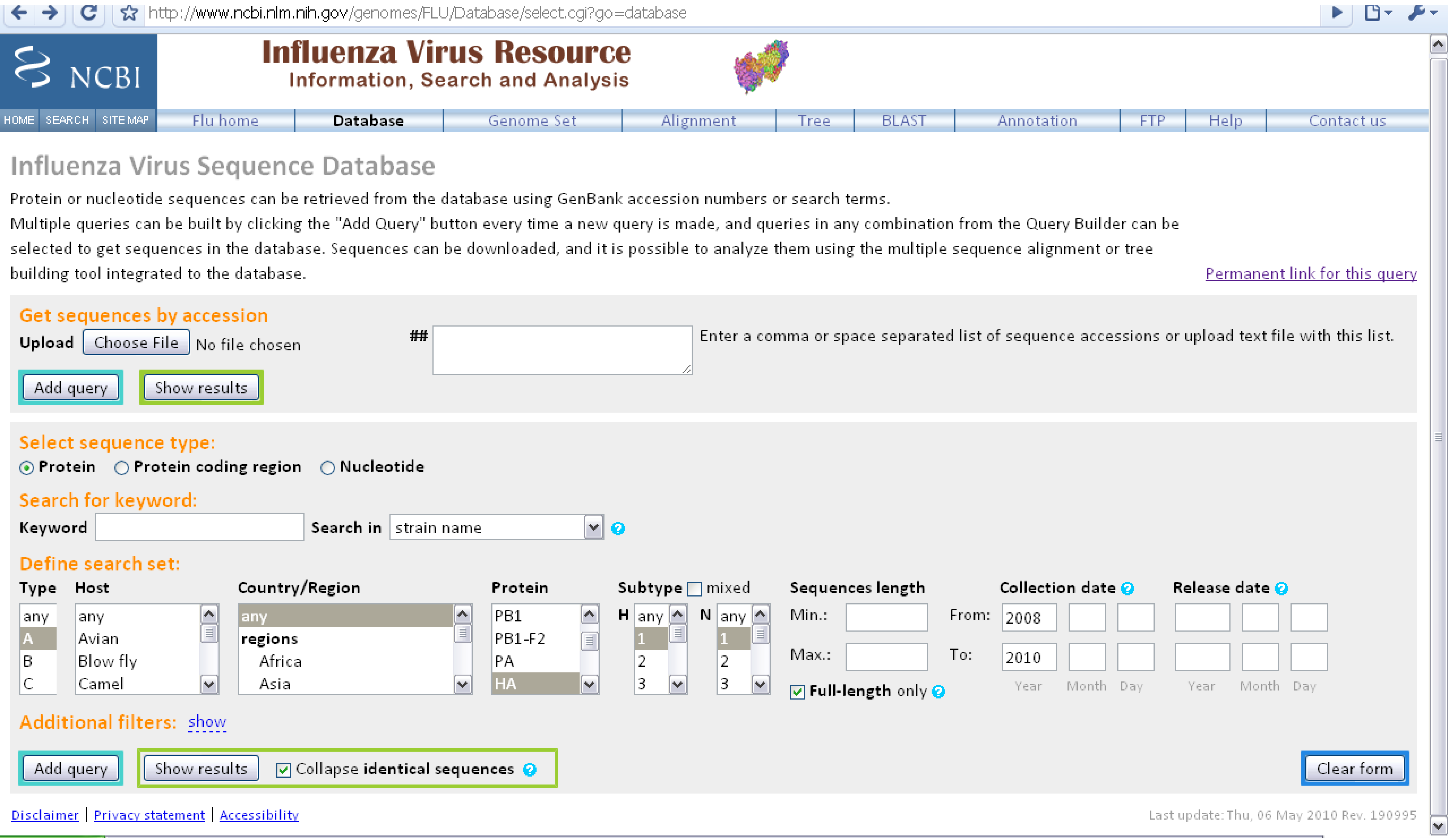


http://en.wikibooks.org/wiki/Structural_Biochemistry/Protein_function/Epitope

Aim 1: Find motifs characterized by pandemic strains

- Tools:
  - Influenza Virus Resource (IVR) at NCBI
  - Basic Local Alignment Search Tool (BLAST)
  - MEME suite
    - Multiple Em for Motif Elicitation (MEME)
    - Motif Alignment and Search Tool (MAST)
  - Immune Epitope Database (IEDB)
  - Chimera

Influenza Virus Resource



Influenza Virus Sequence Database

Protein or nucleotide sequences can be retrieved from the database using GenBank accession numbers or search terms.
Multiple queries can be built by clicking the "Add Query" button every time a new query is made, and queries in any combination from the Query Builder can be selected to get sequences in the database. Sequences can be downloaded, and it is possible to analyze them using the multiple sequence alignment or tree building tool integrated to the database.

Permanent link for this query

Get sequences by accession

Upload Choose File No file chosen ## Enter a comma or space separated list of sequence accessions or upload text file with this list.

Add query | Show results

Select sequence type:

● **Protein** ○ **Protein coding region** ○ **Nucleotide**

Search for keyword:

Keyword ______ **Search in** strain name

Define search set:

Type	Host	Country/Region	Protein	Subtype □ mixed		Sequences length	Collection date	Release date
any A B C	any Avian Blow fly Camel	any regions Africa Asia	PB1 PB1-F2 PA HA	H: any 1 2 3	N: any 1 2 3	Min.: ______ Max.: ______ ☑ Full-length only	From: 2008 (Year) (Month) (Day) To: 2010 (Year) (Month) (Day)	(Year) (Month) (Day) (Year) (Month) (Day)

Selected: Type A; Country/Region any; Protein HA; H 1; N 1.

Additional filters: show

Add query | Show results | ☑ Collapse **identical sequences** | Clear form

Disclaimer | Privacy statement | Accessibility

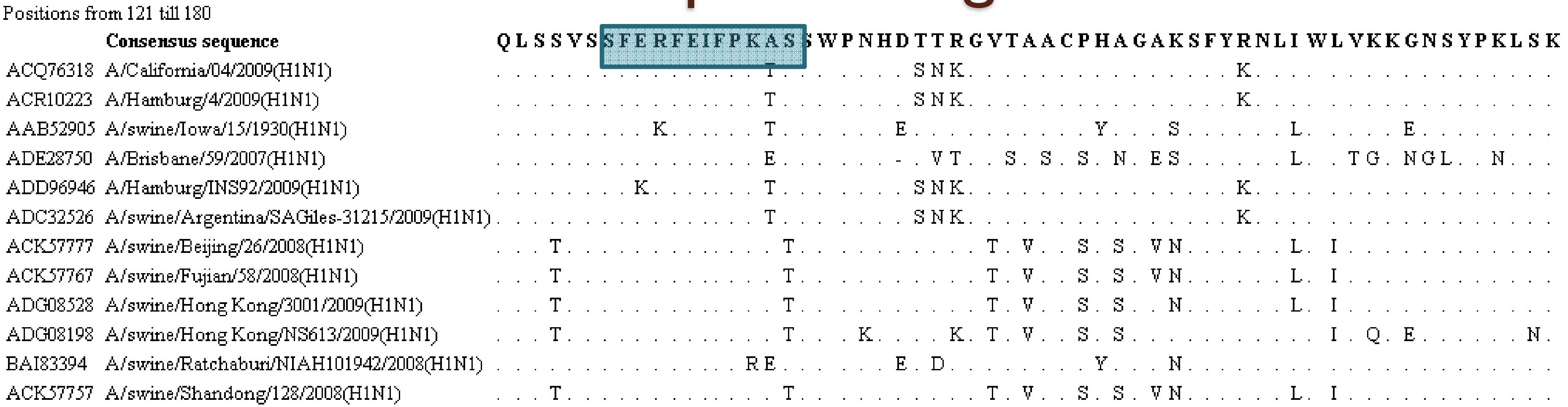
Last update: Thu, 06 May 2010 Rev. 190995

Used IVR to find strains to analyze

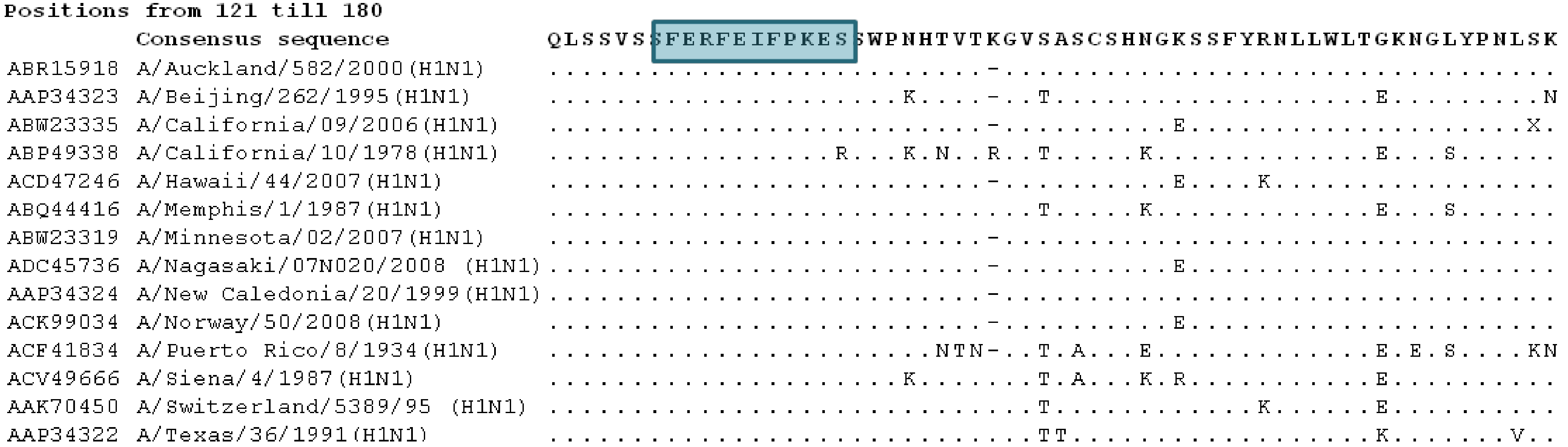
Some strains

- Seasonal H1N1 (total of 15)
  - A/Brisbane/59/2007
- Pandemic H1N1 (total of 97)
  - A/California/04/2009 (3LZG)
  - A/New York/1682/2009 2009/04/27 HA
- Previously circulating swine H1N1 strain (1)
  - A/swine/Iowa/15/30(H1N1) (1RUY)
- Currently circulating H1N1 swine strains (26)
  - A/swine/Minnesota/03000/2010(H1N1)
  - A/swine/Beijing/26/2008(H1N1)
  - A/swine/Hong Kong/NS29/2009(H1N1)

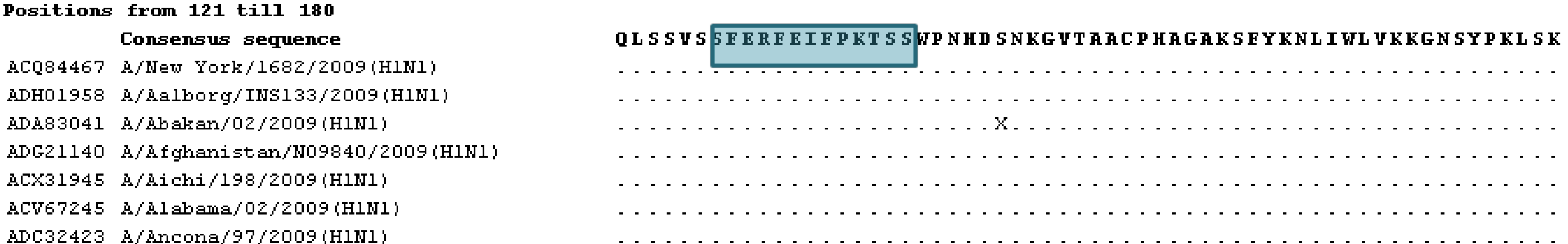
Protein Sequence Alignment



```
Positions from 121 till 180
          Consensus sequence                              QLSSVSSFERFEIFPKASSWPNHDTTRGVTAACPHAGAKSFYRNLIWLVKKGNSYPKLSK
ACQ76318  A/California/04/2009(H1N1)                      ................T.......SNK...............K.................
ACR10223  A/Hamburg/4/2009(H1N1)                          ................T.......SNK...............K.................
AAB52905  A/swine/Iowa/15/1930(H1N1)                      .........K......T......E..........Y...S......L.....E........
ADE28750  A/Brisbane/59/2007(H1N1)                        ................E......-.VT..S.S.S.N.ES......L..TG.NGL..N...
ADD96946  A/Hamburg/INS92/2009(H1N1)                      ........K.......T.......SNK...............K.................
ADC32526  A/swine/Argentina/SAGiles-31215/2009(H1N1)      ................T.......SNK...............K.................
ACK57777  A/swine/Beijing/26/2008(H1N1)                   ...T.............T..........T.V..S.S.VN......L.I............
ACK57767  A/swine/Fujian/58/2008(H1N1)                    ...T.............T..........T.V..S.S.VN......L.I............
ADG08528  A/swine/Hong Kong/3001/2009(H1N1)               ...T.............T..........T.V..S.S..N......L.I............
ADG08198  A/swine/Hong Kong/NS613/2009(H1N1)              ...T.............T...K....K.T.V..S.S...........I.Q.E......N.
BAI83394  A/swine/Ratchaburi/NIAH101942/2008(H1N1)        ...............RE......E.D........Y...N.....................
ACK57757  A/swine/Shandong/128/2008(H1N1)                 ...T.............T..........T.V..S.S.VN......L.I............
```

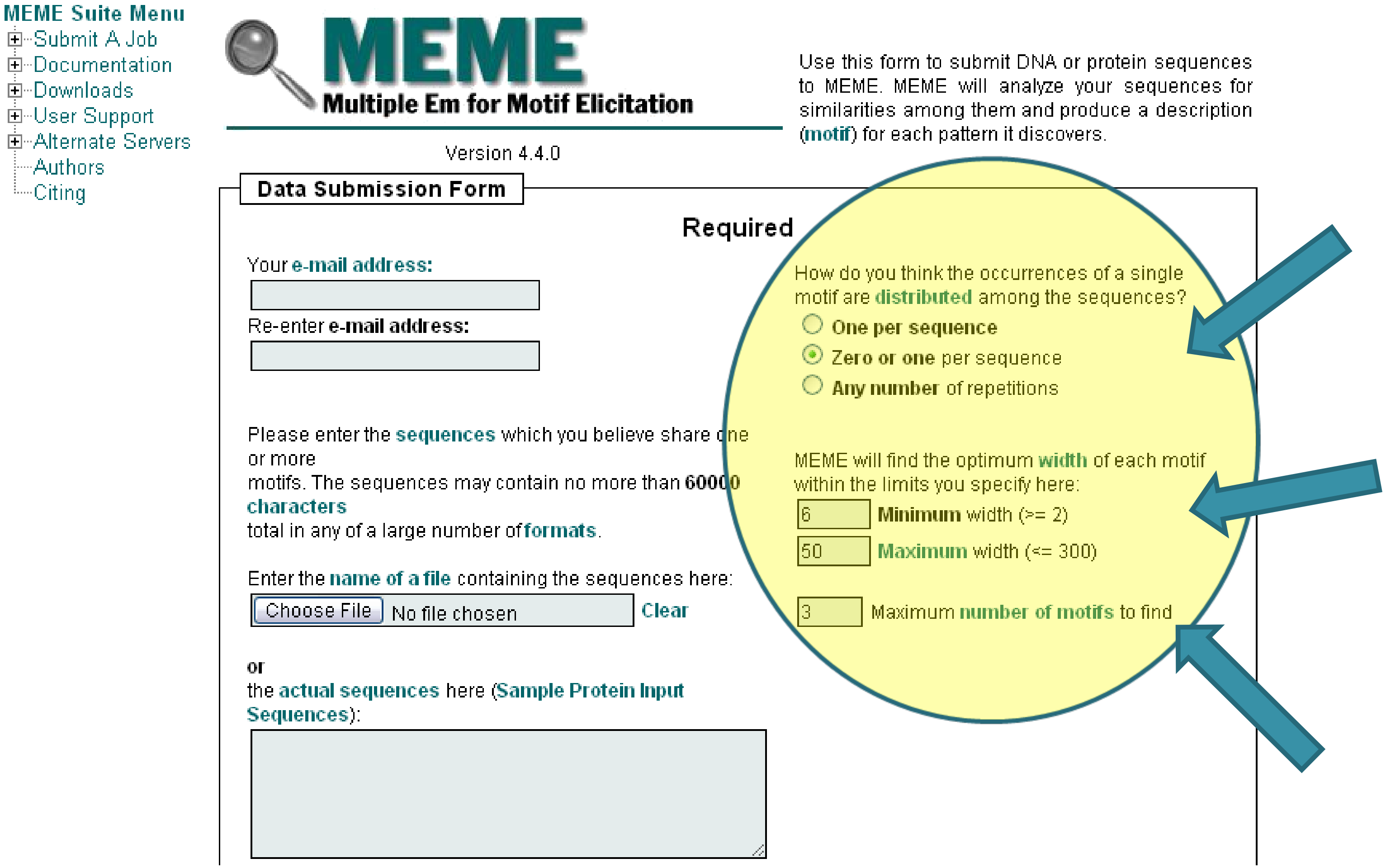


```
Positions from 121 till 180
          Consensus sequence                   QLSSVSSFERFEIFPKESSWPNHTVTKGVSASCSHNGKSSFYRNLLWLTGKNGLYPNLSK
ABR15918  A/Auckland/582/2000(H1N1)            ..........................-.................................
AAP34323  A/Beijing/262/1995(H1N1)             .....................K....-..T...................E.........N
ABW23335  A/California/09/2006(H1N1)           ..........................-..........E....................X.
ABP49338  A/California/10/1978(H1N1)           .................R...K.N..R..T.....K.............E...S......
ACD47246  A/Hawaii/44/2007(H1N1)               ..........................-..........E....K.................
ABQ44416  A/Memphis/1/1987(H1N1)               .............................T.....K.............E...S......
ABW23319  A/Minnesota/02/2007(H1N1)            ..........................-.................................
ADC45736  A/Nagasaki/07N020/2008 (H1N1)        ..........................-..........E......................
AAP34324  A/New Caledonia/20/1999(H1N1)        ..........................-.................................
ACK99034  A/Norway/50/2008(H1N1)               ..........................-..........E......................
ACF41834  A/Puerto Rico/8/1934(H1N1)           .......................NTN-..T.A...E.............E.E.S....KN
ACV49666  A/Siena/4/1987(H1N1)                 .....................K.......T.A...K.R...........E..........
AAK70450  A/Switzerland/5389/95 (H1N1)         .............................T............K......E..........
AAP34322  A/Texas/36/1991(H1N1)                .............................TT..................K.......V..
```



```
Positions from 121 till 180
          Consensus sequence                QLSSVSSFERFEIFPKTSSWPNHDSNKGVTAACPHAGAKSFYKNLIWLVKKGNSYPKLSK
ACQ84467  A/New York/1682/2009(H1N1)        ............................................................
ADH01958  A/Aalborg/INS133/2009(H1N1)       ............................................................
ADA83041  A/Abakan/02/2009(H1N1)            ........................X...................................
ADG21140  A/Afghanistan/N09840/2009(H1N1)   ............................................................
ACX31945  A/Aichi/198/2009(H1N1)            ............................................................
ACV67245  A/Alabama/02/2009(H1N1)           ............................................................
ADC32423  A/Ancona/97/2009(H1N1)            ............................................................
```

MEME



Version 4.4.0

Use this form to submit DNA or protein sequences to MEME. MEME will analyze your sequences for similarities among them and produce a description (**motif**) for each pattern it discovers.

Data Submission Form

Required

Your **e-mail address:**

Re-enter **e-mail address:**

Please enter the **sequences** which you believe share one or more
motifs. The sequences may contain no more than **60000 characters**
total in any of a large number of **formats**.

Enter the **name of a file** containing the sequences here:

or
the **actual sequences** here (**Sample Protein Input Sequences**):

How do you think the occurrences of a single motif are **distributed** among the sequences?

○ **One per sequence**
● **Zero or one** per sequence
○ **Any number** of repetitions

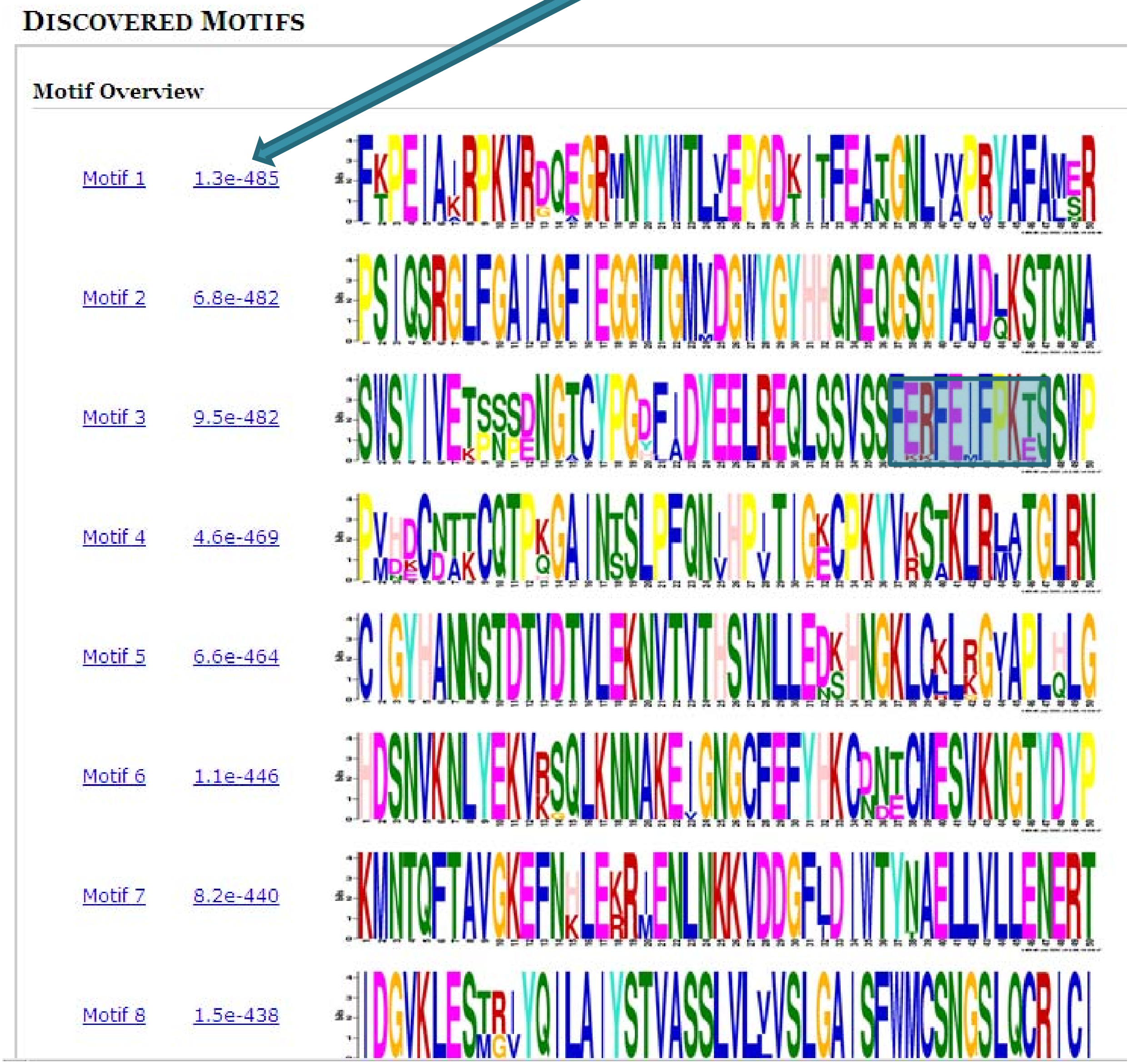
MEME will find the optimum **width** of each motif within the limits you specify here:

6 **Minimum** width (>= 2)
50 **Maximum** width (<= 300)

3 Maximum **number of motifs** to find

MEME results

E-value: the probability of finding an equally well-conserved pattern in random sequences



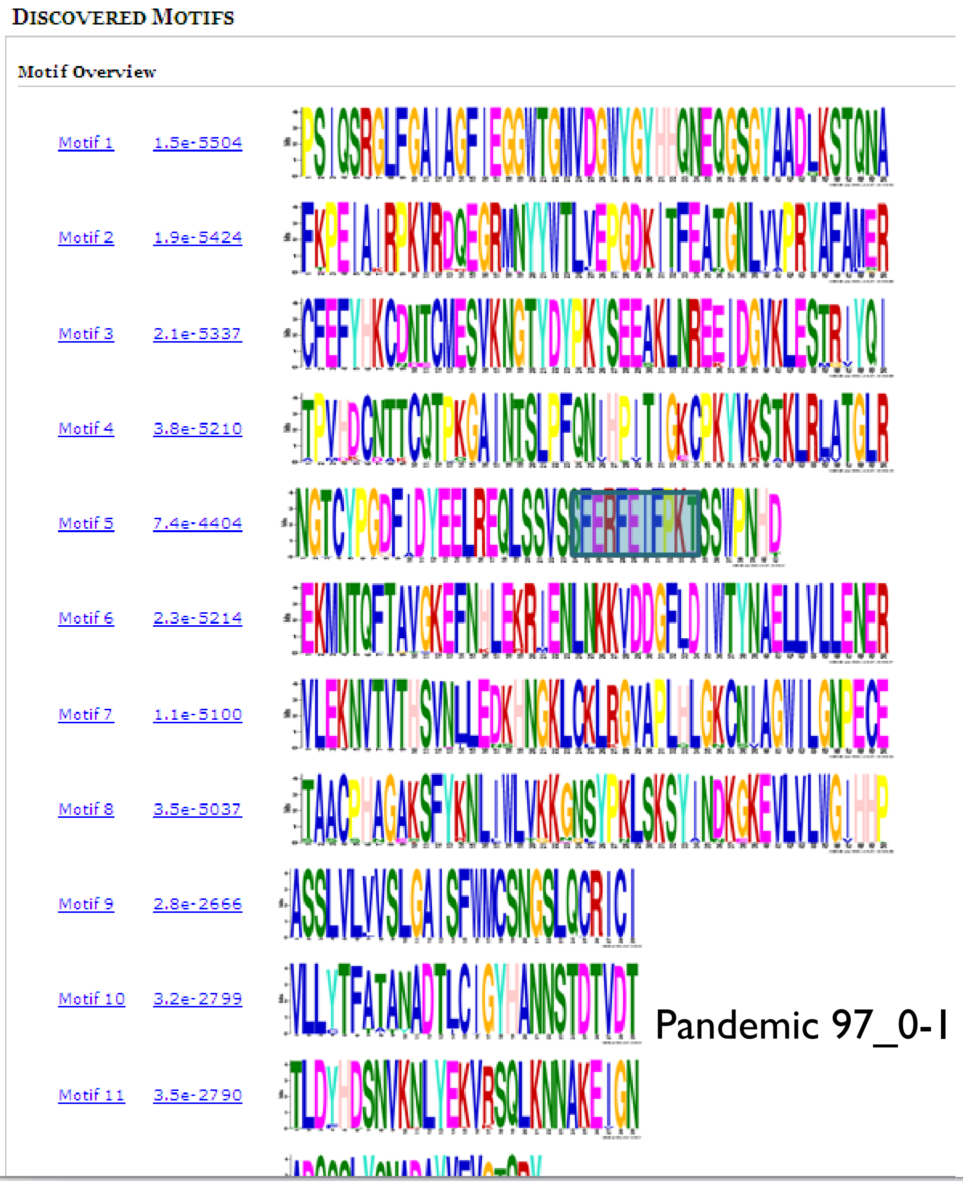
Discovered Motifs

Motif Overview

Motif 1	1.3e-485
Motif 2	6.8e-482
Motif 3	9.5e-482
Motif 4	4.6e-469
Motif 5	6.6e-464
Motif 6	1.1e-446
Motif 7	8.2e-440
Motif 8	1.5e-438

- Input: Pandemic, Swine and Seasonal strains
- 17 motifs found
- Width: 6-50
- Zero or one occurrence of motifs per sequence
- 12 sequences:
  - 4 swine
  - 4 seasonal
  - 4 pandemic

MEME Swine and Pandemic motifs

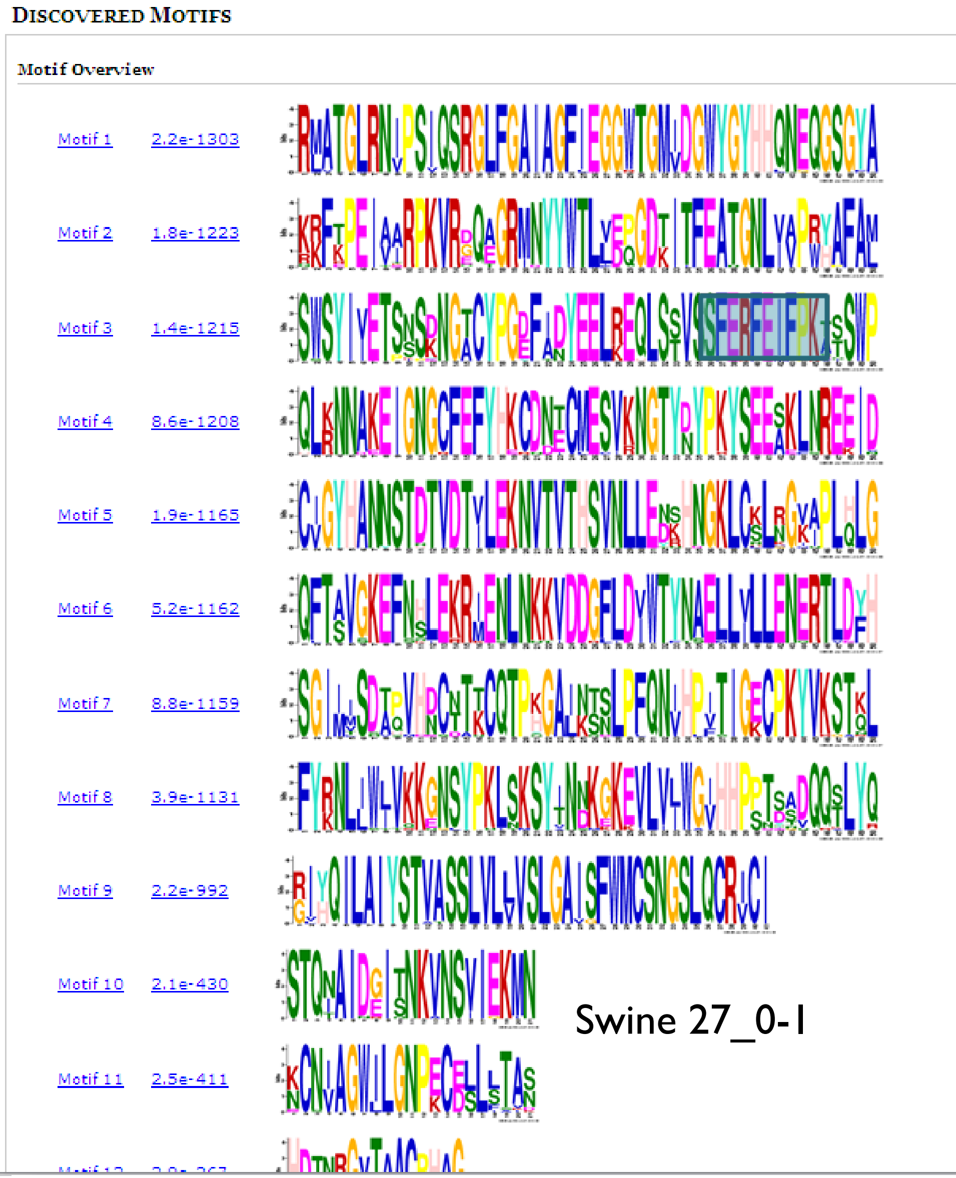


Discovered Motifs

Motif Overview

Motif 1	1.5e-5504
Motif 2	1.9e-5424
Motif 3	2.1e-5337
Motif 4	3.8e-5210
Motif 5	7.4e-4404
Motif 6	2.3e-5214
Motif 7	1.1e-5100
Motif 8	3.5e-5037
Motif 9	2.8e-2666
Motif 10	3.2e-2799
Motif 11	3.5e-2790

Pandemic 97_0-1



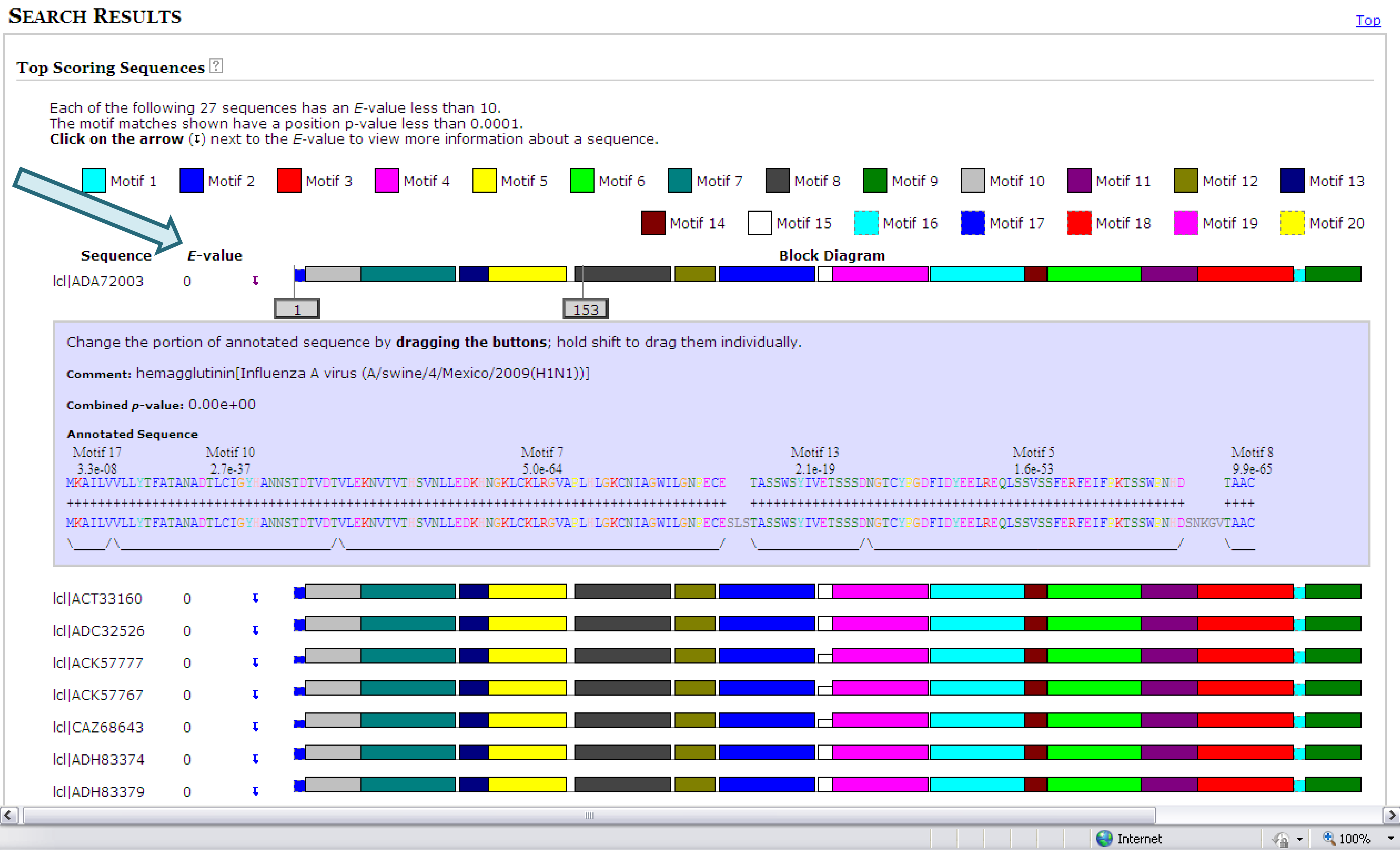
Discovered Motifs

Motif Overview

Motif 1	2.2e-1303
Motif 2	1.8e-1223
Motif 3	1.4e-1215
Motif 4	8.6e-1208
Motif 5	1.9e-1165
Motif 6	5.2e-1162
Motif 7	8.8e-1159
Motif 8	3.9e-1131
Motif 9	2.2e-992
Motif 10	2.1e-430
Motif 11	2.5e-411
Motif 12	[illegible]

Swine 27_0-1

MAST results



SEARCH RESULTS

Top

Top Scoring Sequences ?

Each of the following 27 sequences has an *E*-value less than 10.
The motif matches shown have a position p-value less than 0.0001.
Click on the arrow (↧) next to the *E*-value to view more information about a sequence.

Sequence	*E*-value
lcl|ADA72003	0

Change the portion of annotated sequence by **dragging the buttons**; hold shift to drag them individually.

Comment: hemagglutinin[Influenza A virus (A/swine/4/Mexico/2009(H1N1))]

Combined *p*-value: 0.00e+00

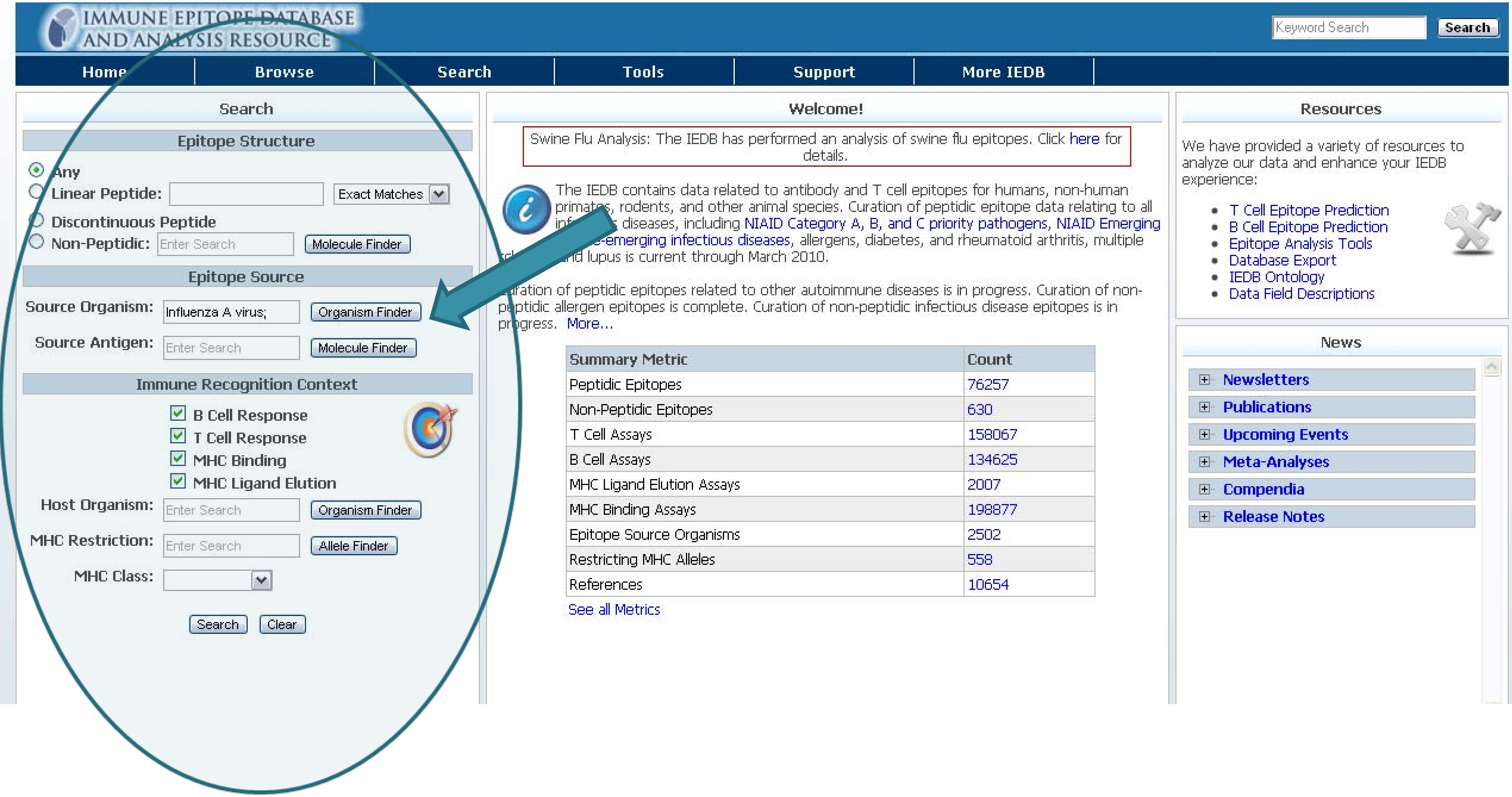
Annotated Sequence

```
 Motif 17              Motif 10                                                   Motif 7                                                    Motif 13                                         Motif 5                                         Motif 8
  3.3e-08               2.7e-37                                                    5.0e-64                                                   2.1e-19                                          1.6e-53                                        9.9e-65
MKAILVVLLYTFATANADTLCIGYHANNSTDTVDTVLEKNVTVTHSVNLLEDKHNGKLCKLRGVAPLHLGKCNIAGWILGNPECE   TASSWSYIVETSSSDNGTCYPGDFIDYEELREQLSSVSSFERFEIFPKTSSWPNHD     TAAC
+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++   ++++++++++++++++++++++++++++++++++++++++++++++++++++++++     ++++
MKAILVVLLYTFATANADTLCIGYHANNSTDTVDTVLEKNVTVTHSVNLLEDKHNGKLCKLRGVAPLHLGKCNIAGWILGNPECESLSTASSWSYIVETSSSDNGTCYPGDFIDYEELREQLSSVSSFERFEIFPKTSSWPNHDSNKGVTAAC
\____/\___________________________/\__________________________________________/   \______________/\_____________________________________/     \___
```

Sequence	*E*-value
lcl|ACT33160	0
lcl|ADC32526	0
lcl|ACK57777	0
lcl|ACK57767	0
lcl|CAZ68643	0
lcl|ADH83374	0
lcl|ADH83379	0

Pandemic 97 0-1 vs Swine 27

Immune Epitope Database



IMMUNE EPITOPE DATABASE AND ANALYSIS RESOURCE

Keyword Search | Search

Home | Browse | Search | Tools | Support | More IEDB

Search

Epitope Structure

● Any

○ Linear Peptide: [ ] Exact Matches

○ Discontinuous Peptide

○ Non-Peptidic: Enter Search | Molecule Finder

Epitope Source

Source Organism: Influenza A virus; | Organism Finder

Source Antigen: Enter Search | Molecule Finder

Immune Recognition Context

☑ B Cell Response

☑ T Cell Response

☑ MHC Binding

☑ MHC Ligand Elution

Host Organism: Enter Search | Organism Finder

MHC Restriction: Enter Search | Allele Finder

MHC Class:

Search | Clear

Welcome!

Swine Flu Analysis: The IEDB has performed an analysis of swine flu epitopes. Click here for details.

The IEDB contains data related to antibody and T cell epitopes for humans, non-human primates, rodents, and other animal species. Curation of peptidic epitope data relating to all inf s diseases, including NIAID Category A, B, and C priority pathogens, NIAID Emerging -emerging infectious diseases, allergens, diabetes, and rheumatoid arthritis, multiple scl nd lupus is current through March 2010.

ration of peptidic epitopes related to other autoimmune diseases is in progress. Curation of non-peptidic allergen epitopes is complete. Curation of non-peptidic infectious disease epitopes is in progress. More...

Summary Metric	Count
Peptidic Epitopes	76257
Non-Peptidic Epitopes	630
T Cell Assays	158067
B Cell Assays	134625
MHC Ligand Elution Assays	2007
MHC Binding Assays	198877
Epitope Source Organisms	2502
Restricting MHC Alleles	558
References	10654

See all Metrics

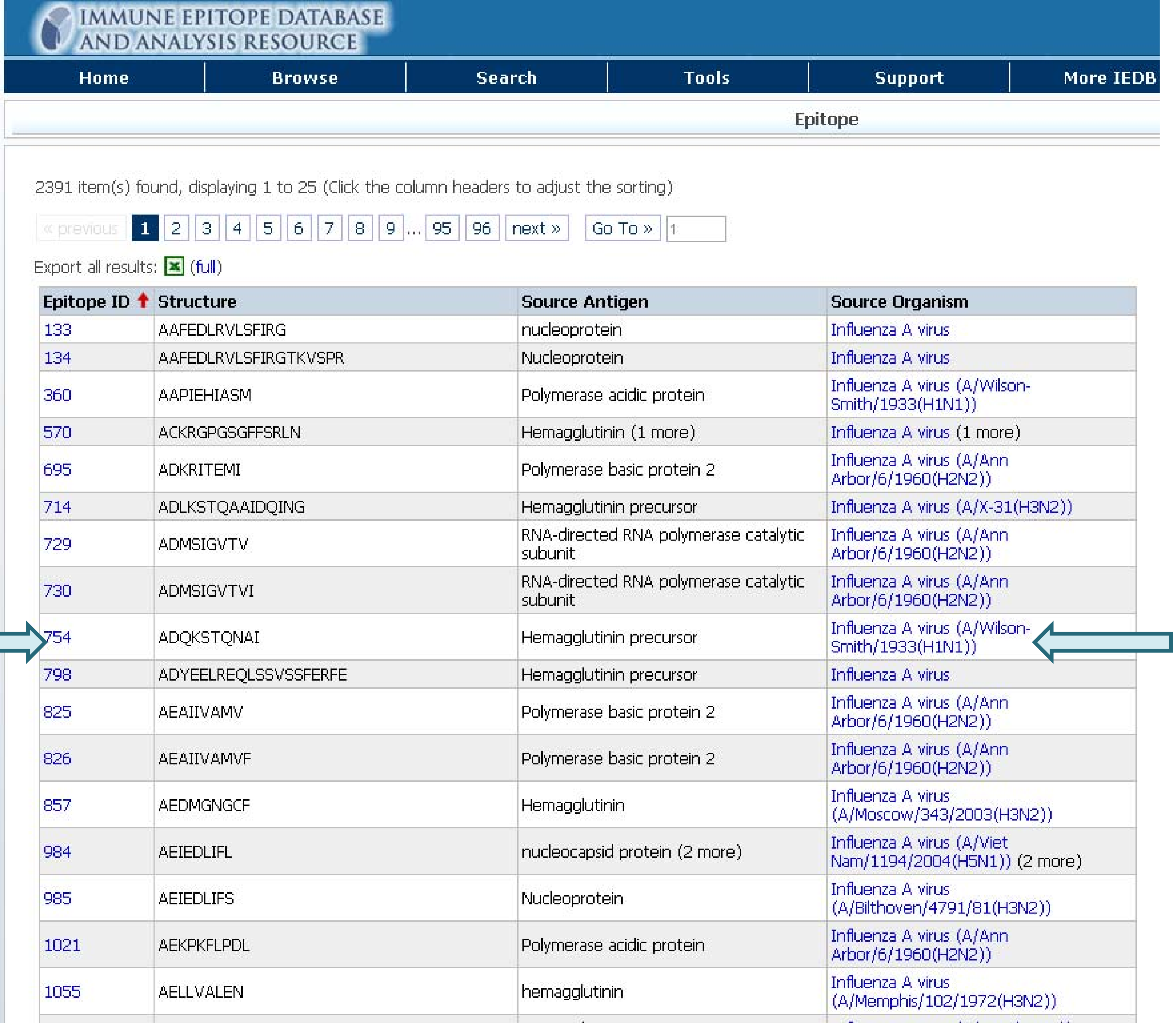
Resources

We have provided a variety of resources to analyze our data and enhance your IEDB experience:

- T Cell Epitope Prediction
- B Cell Epitope Prediction
- Epitope Analysis Tools
- Database Export
- IEDB Ontology
- Data Field Descriptions

News

- Newsletters
- Publications
- Upcoming Events
- Meta-Analyses
- Compendia
- Release Notes



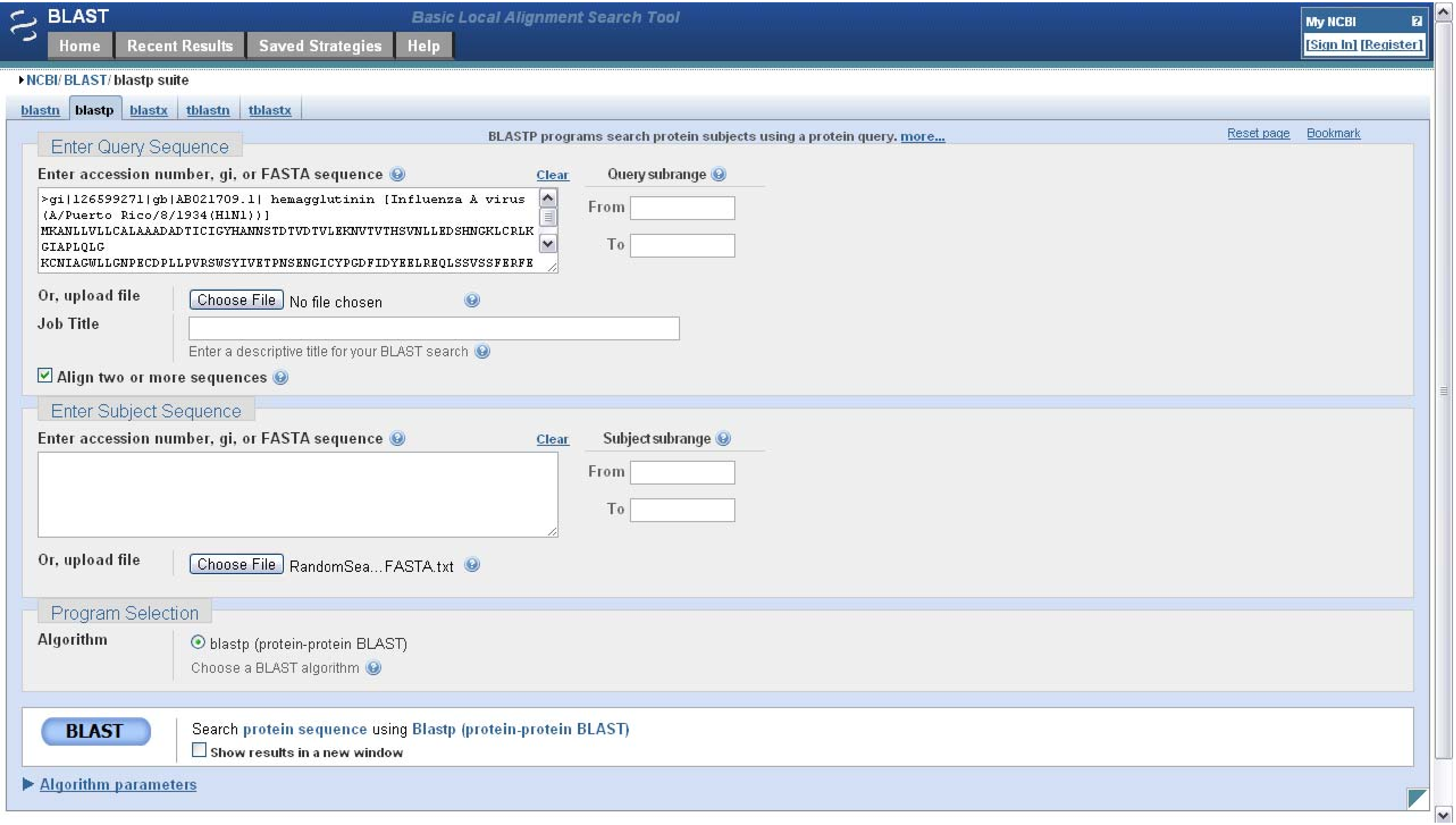
Epitope

2391 item(s) found, displaying 1 to 25 (Click the column headers to adjust the sorting)

Export all results: (full)

Epitope ID ↑	Structure	Source Antigen	Source Organism
133	AAFEDLRVLSFIRG	nucleoprotein	Influenza A virus
134	AAFEDLRVLSFIRGTKVSPR	Nucleoprotein	Influenza A virus
360	AAPIEHIASM	Polymerase acidic protein	Influenza A virus (A/Wilson-Smith/1933(H1N1))
570	ACKRGPGSGFFSRLN	Hemagglutinin (1 more)	Influenza A virus (1 more)
695	ADKRITEMI	Polymerase basic protein 2	Influenza A virus (A/Ann Arbor/6/1960(H2N2))
714	ADLKSTQAAIDQING	Hemagglutinin precursor	Influenza A virus (A/X-31(H3N2))
729	ADMSIGVTV	RNA-directed RNA polymerase catalytic subunit	Influenza A virus (A/Ann Arbor/6/1960(H2N2))
730	ADMSIGVTVI	RNA-directed RNA polymerase catalytic subunit	Influenza A virus (A/Ann Arbor/6/1960(H2N2))
754	ADQKSTQNAI	Hemagglutinin precursor	Influenza A virus (A/Wilson-Smith/1933(H1N1))
798	ADYEELREQLSSVSSFERFE	Hemagglutinin precursor	Influenza A virus
825	AEAIIVAMV	Polymerase basic protein 2	Influenza A virus (A/Ann Arbor/6/1960(H2N2))
826	AEAIIVAMVF	Polymerase basic protein 2	Influenza A virus (A/Ann Arbor/6/1960(H2N2))
857	AEDMGNGCF	Hemagglutinin	Influenza A virus (A/Moscow/343/2003(H3N2))
984	AEIEDLIFL	nucleocapsid protein (2 more)	Influenza A virus (A/Viet Nam/1194/2004(H5N1)) (2 more)
985	AEIEDLIFS	Nucleoprotein	Influenza A virus (A/Bilthoven/4791/81(H3N2))
1021	AEKPKFLPDL	Polymerase acidic protein	Influenza A virus (A/Ann Arbor/6/1960(H2N2))
1055	AELLVALEN	hemagglutinin	Influenza A virus (A/Memphis/102/1972(H3N2))

BLAST



NCBI/ BLAST/ blastp suite

BLASTP programs search protein subjects using a protein query. more...

Enter Query Sequence

Enter accession number, gi, or FASTA sequence

```
>gi|126599271|gb|AB021709.1| hemagglutinin [Influenza A virus
(A/Puerto Rico/8/1934(H1N1))]
MKANLLVLLCALAAADADTICIGYHANNSTDTVDTVLEKNVTVTHSVNLLEDSHNGKLCRLK
GIAPLQLG
KCNIAGWLLGNPECDPLLPVRSWSYIVETPNSENGICYPGDFIDYEELREQLSSVSSFERFE
```

Query subrange

From

To

Or, upload file: No file chosen

Job Title

Enter a descriptive title for your BLAST search

☑ Align two or more sequences

Enter Subject Sequence

Enter accession number, gi, or FASTA sequence

Subject subrange

From

To

Or, upload file: RandomSea...FASTA.txt

Program Selection

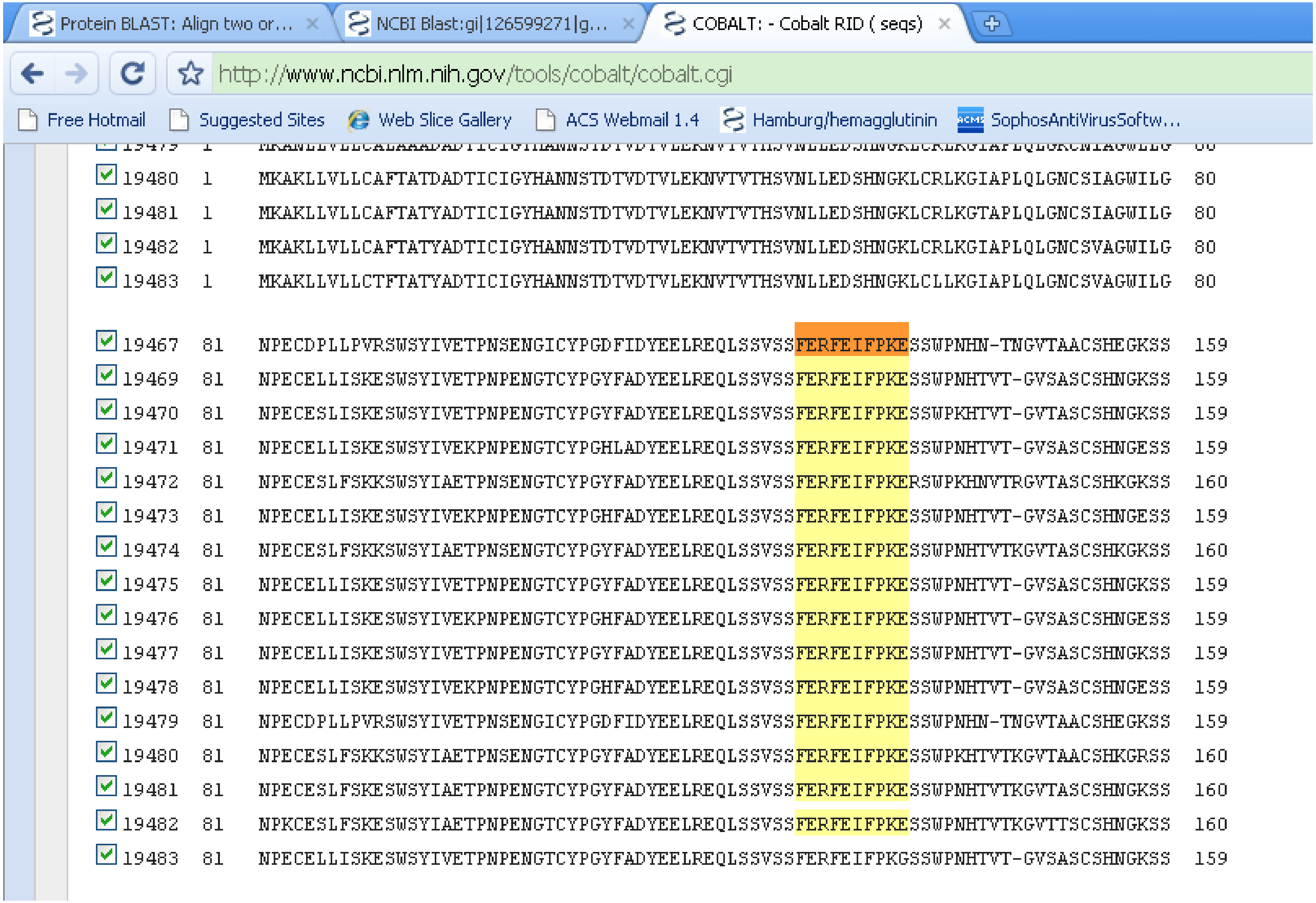
Algorithm: ● blastp (protein-protein BLAST)

Choose a BLAST algorithm

Search protein sequence using Blastp (protein-protein BLAST)

□ Show results in a new window

Multiple alignment

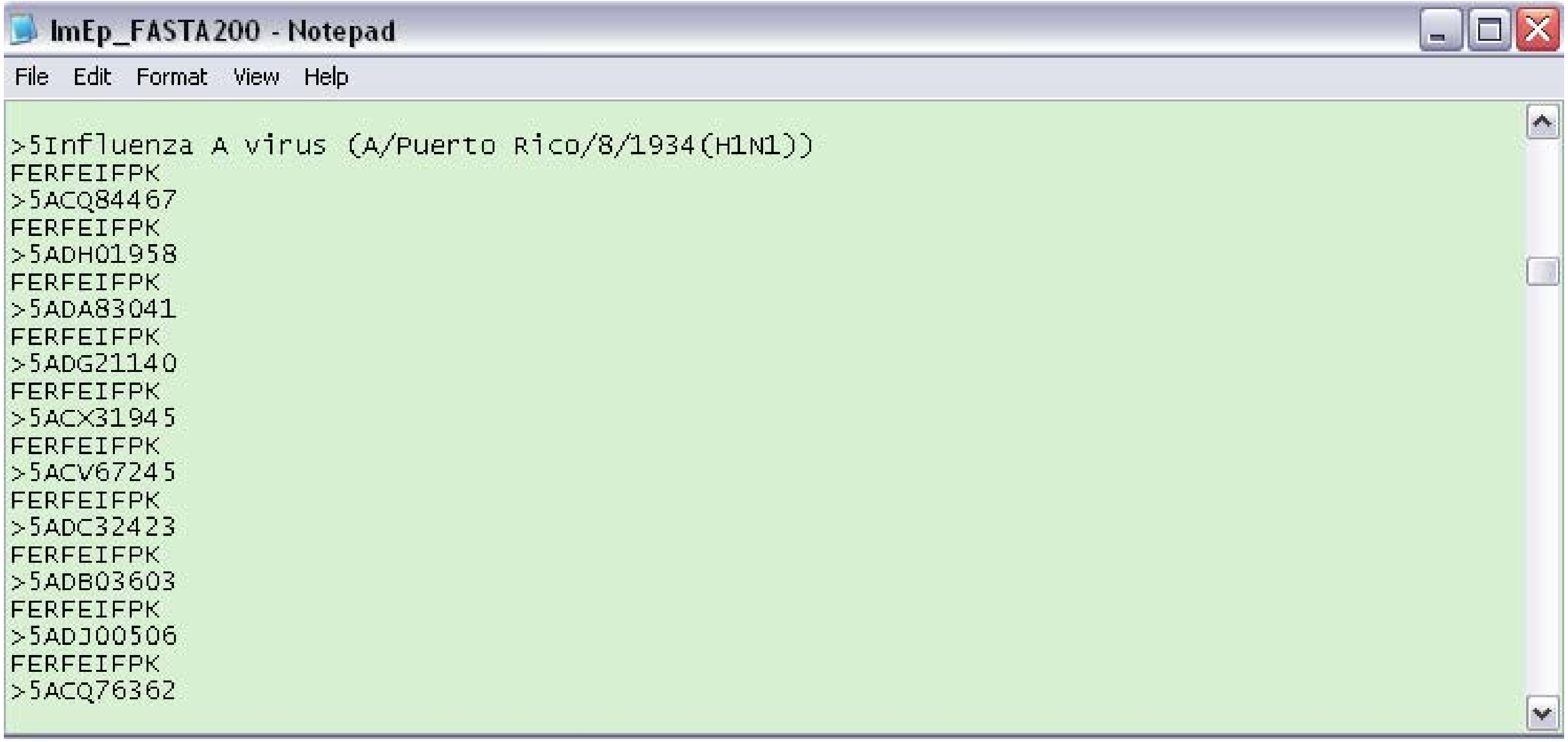

Protein BLAST: Align two or...
NCBI Blast:gi|126599271|g...
COBALT: - Cobalt RID (seqs)
http://www.ncbi.nlm.nih.gov/tools/cobalt/cobalt.cgi
Free Hotmail
Suggested Sites
Web Slice Gallery
ACS Webmail 1.4
Hamburg/hemagglutinin
SophosAntiVirusSoftw...

```
☑ 19479   1    MKANLLVLLCALAAADADTICIGYHANNSTDTVDTVLEKNVTVTHSVNLLEDSHNGKLCRLKGIAPLQLGKCNIAGWLLG  80
☑ 19480   1    MKAKLLVLLCAFTATDADTICIGYHANNSTDTVDTVLEKNVTVTHSVNLLEDSHNGKLCRLKGIAPLQLGNCSIAGWILG  80
☑ 19481   1    MKAKLLVLLCAFTATYADTICIGYHANNSTDTVDTVLEKNVTVTHSVNLLEDSHNGKLCRLKGTAPLQLGNCSIAGWILG  80
☑ 19482   1    MKAKLLVLLCAFTATYADTICIGYHANNSTDTVDTVLEKNVTVTHSVNLLEDSHNGKLCRLKGIAPLQLGNCSVAGWILG  80
☑ 19483   1    MKAKLLVLLCTFTATYADTICIGYHANNSTDTVDTVLEKNVTVTHSVNLLEDSHNGKLCLLKGIAPLQLGNCSVAGWILG  80

☑ 19467   81   NPECDPLLPVRSWSYIVETPNSENGICYPGDFIDYEELREQLSSVSSFERFEIFPKESSWPNHN-TNGVTAACSHEGKSS  159
☑ 19469   81   NPECELLISKESWSYIVETPNPENGTCYPGYFADYEELREQLSSVSSFERFEIFPKESSWPNHTVT-GVSASCSHNGKSS  159
☑ 19470   81   NPECESLISKESWSYIVETPNPENGTCYPGYFADYEELREQLSSVSSFERFEIFPKESSWPKHTVT-GVTASCSHNGKSS  159
☑ 19471   81   NPECELLISKESWSYIVEKPNPENGTCYPGHLADYEELREQLSSVSSFERFEIFPKESSWPNHTVT-GVSASCSHNGESS  159
☑ 19472   81   NPECESLFSKKSWSYIAETPNSENGTCYPGYFADYEELREQLSSVSSFERFEIFPKERSWPKHNVTRGVTASCSHKGKSS  160
☑ 19473   81   NPECELLISKESWSYIVEKPNPENGTCYPGHFADYEELREQLSSVSSFERFEIFPKESSWPNHTVT-GVSASCSHNGESS  159
☑ 19474   81   NPECESLFSKKSWSYIAETPNSENGTCYPGYFADYEELREQLSSVSSFERFEIFPKESSWPNHTVTKGVTASCSHKGKSS  160
☑ 19475   81   NPECELLISKESWSYIVETPNPENGTCYPGYFADYEELREQLSSVSSFERFEIFPKESSWPNHTVT-GVSASCSHNGKSS  159
☑ 19476   81   NPECELLISKESWSYIVEKPNPENGTCYPGHFADYEELREQLSSVSSFERFEIFPKESSWPNHTVT-GVSASCSHNGESS  159
☑ 19477   81   NPECELLISKESWSYIVETPNPENGTCYPGYFADYEELREQLSSVSSFERFEIFPKESSWPNHTVT-GVSASCSHNGKSS  159
☑ 19478   81   NPECELLISKESWSYIVEKPNPENGTCYPGHFADYEELREQLSSVSSFERFEIFPKESSWPNHTVT-GVSASCSHNGESS  159
☑ 19479   81   NPECDPLLPVRSWSYIVETPNSENGICYPGDFIDYEELREQLSSVSSFERFEIFPKESSWPNHN-TNGVTAACSHEGKSS  159
☑ 19480   81   NPECESLFSKKSWSYIAETPNSENGTCYPGYFADYEELREQLSSVSSFERFEIFPKESSWPKHTVTKGVTAACSHKGRSS  160
☑ 19481   81   NPECESLFSKESWSYIAETPNPENGTCYPGYFADYEELREQLSSVSSFERFEIFPKESSWPNHTVTKGVTASCSHNGKSS  160
☑ 19482   81   NPKCESLFSKESWSYIAETPNPENGTCYPGYFADYEELREQLSSVSSFERFEIFPKESSWPNHTVTKGVTTSCSHNGKSS  160
☑ 19483   81   NPECELLISKESWSYIVETPNPENGTCYPGYFADYEELREQLSSVSSFERFEIFPKGSSWPNHTVT-GVSASCSHNGKSS  159
```

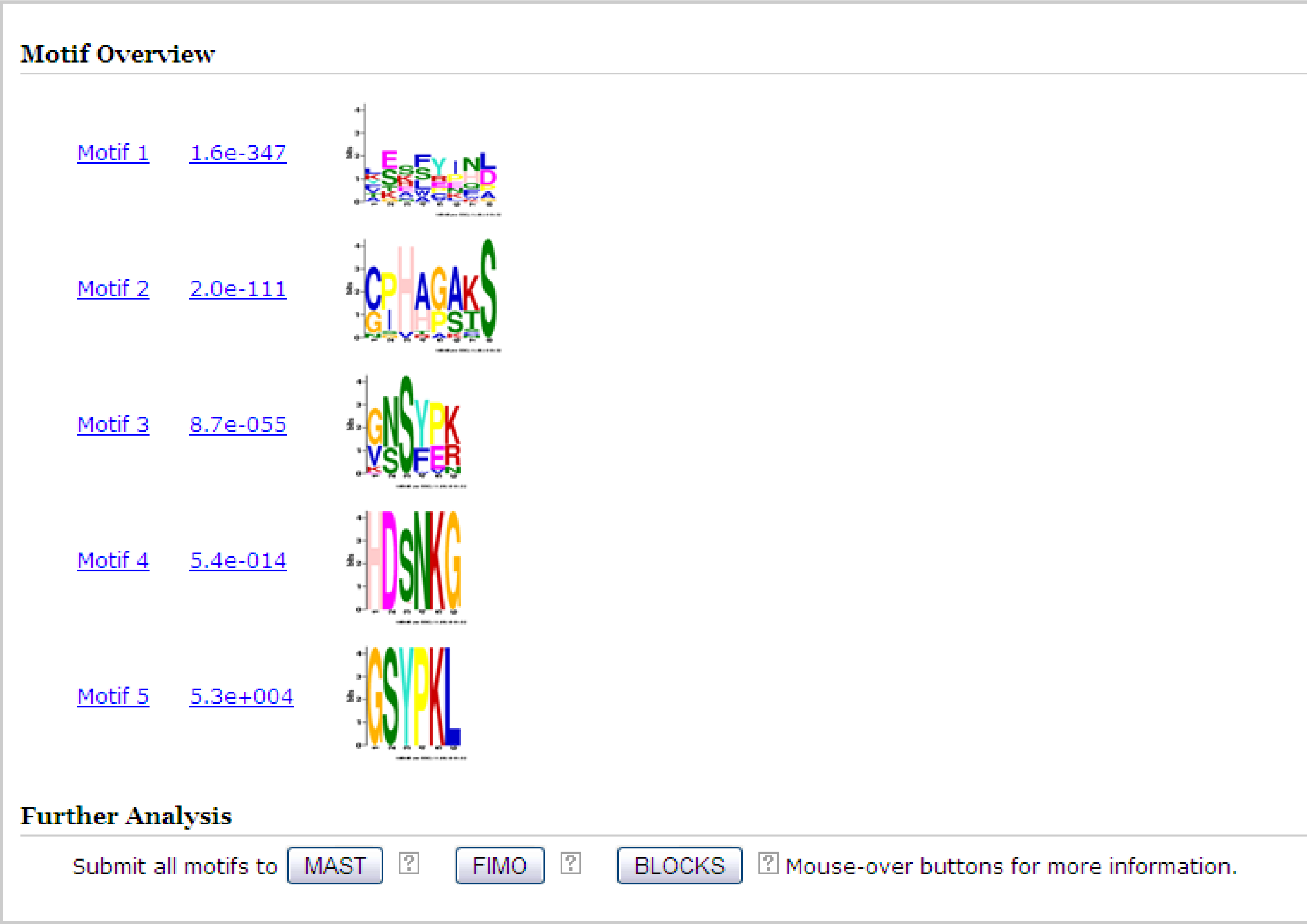
Creating FASTA files

- Created FASTA files from BLAST results to get MEME results



```
>5Influenza A virus (A/Puerto Rico/8/1934(H1N1))
FERFEIFPK
>5ACQ84467
FERFEIFPK
>5ADH01958
FERFEIFPK
>5ADA83041
FERFEIFPK
>5ADG21140
FERFEIFPK
>5ACX31945
FERFEIFPK
>5ACV67245
FERFEIFPK
>5ADC32423
FERFEIFPK
>5ADB03603
FERFEIFPK
>5ADJ00506
FERFEIFPK
>5ACQ76362
```

MEME



Motif Overview

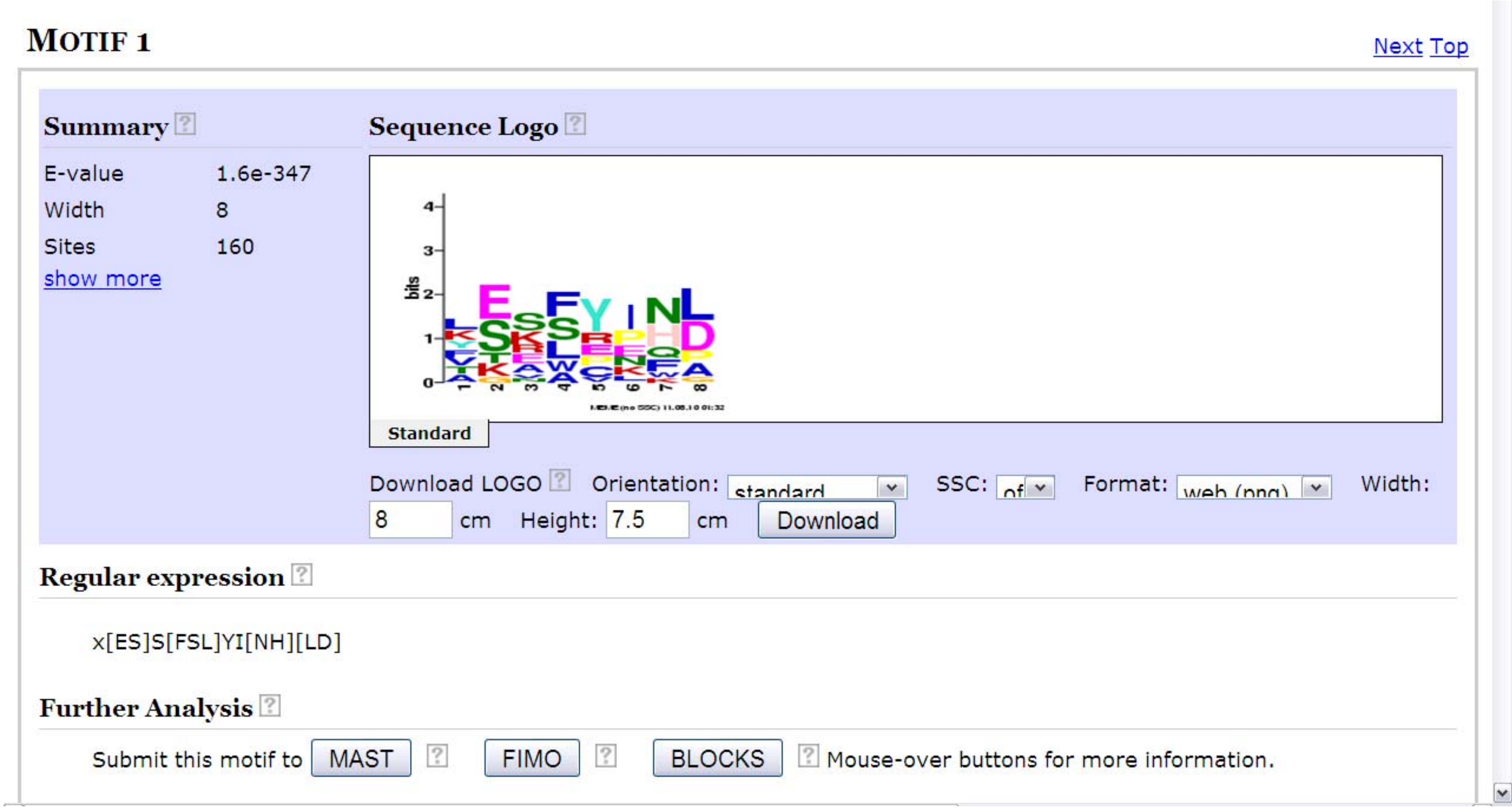
Motif 1	1.6e-347
Motif 2	2.0e-111
Motif 3	8.7e-055
Motif 4	5.4e-014
Motif 5	5.3e+004

Further Analysis

Submit all motifs to MAST ? FIMO ? BLOCKS ? Mouse-over buttons for more information.

MEME_ImEplessthan200 (pandemic)

MEME



Motif 1

Next Top

Summary ?

E-value	1.6e-347
Width	8
Sites	160

show more

Sequence Logo ?

Standard

Download LOGO ? Orientation: standard SSC: off Format: web (png) Width: 8 cm Height: 7.5 cm Download

Regular expression ?

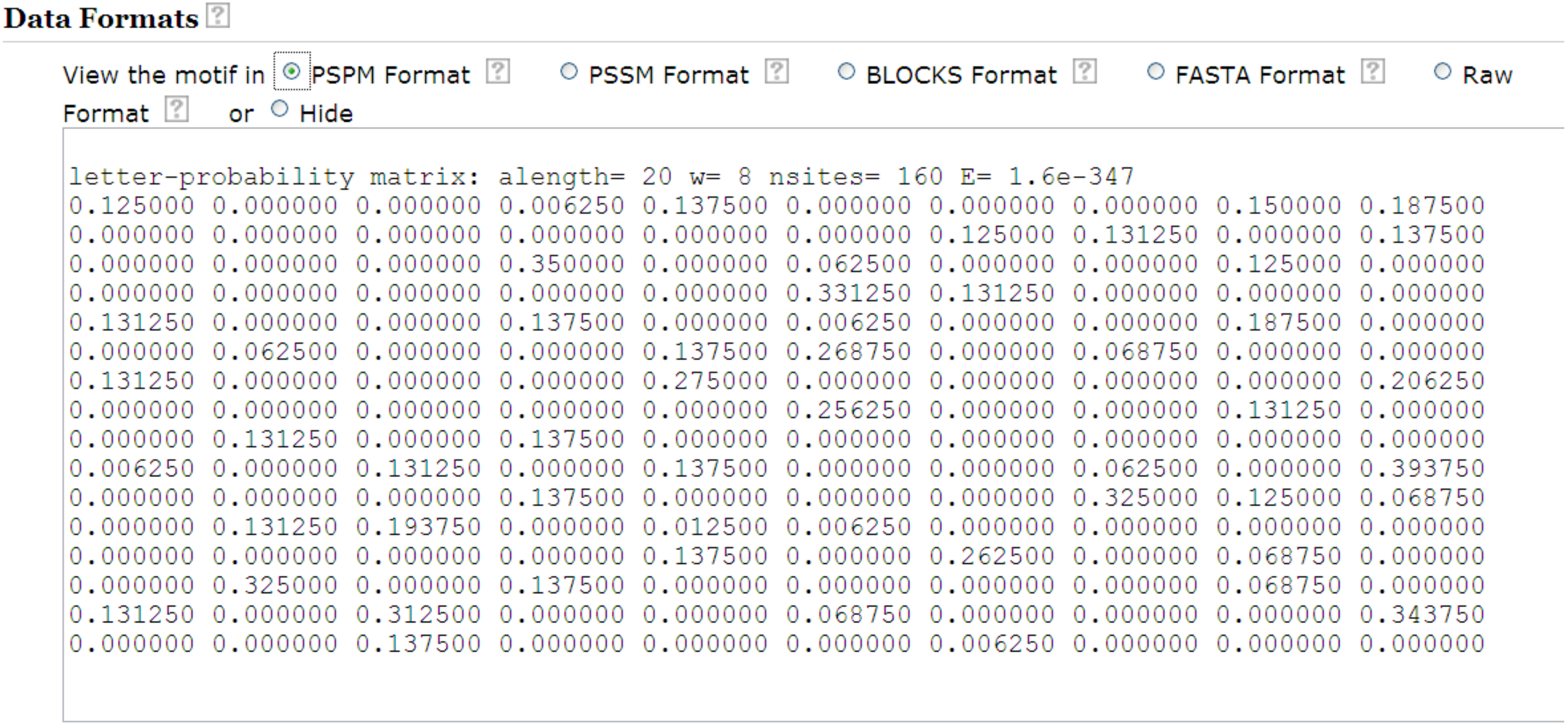
x[ES]S[FSL]YI[NH][LD]

Further Analysis ?

Submit this motif to MAST ? FIMO ? BLOCKS ? Mouse-over buttons for more information.

PSPM –Position Specific Scoring Matrix

- Protein alphabet:
  - ACDEFGHIKLMNPQRSTVWY

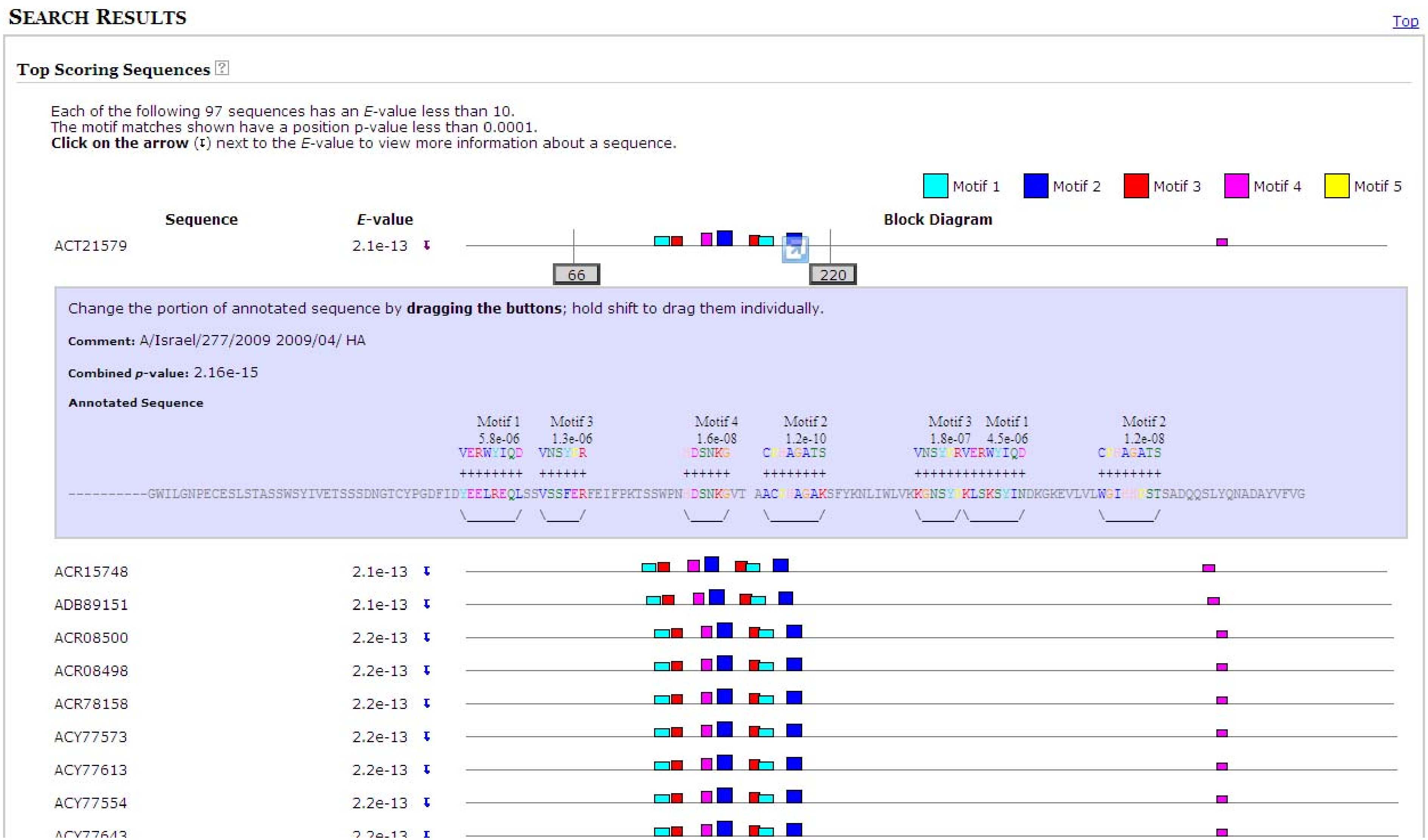


Data Formats

View the motif in ● PSPM Format ○ PSSM Format ○ BLOCKS Format ○ FASTA Format ○ Raw Format or ○ Hide

```
letter-probability matrix: alength= 20 w= 8 nsites= 160 E= 1.6e-347
0.125000 0.000000 0.000000 0.006250 0.137500 0.000000 0.000000 0.000000 0.150000 0.187500
0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.125000 0.131250 0.000000 0.137500
0.000000 0.000000 0.000000 0.350000 0.000000 0.062500 0.000000 0.000000 0.125000 0.000000
0.000000 0.000000 0.000000 0.000000 0.000000 0.331250 0.131250 0.000000 0.000000 0.000000
0.131250 0.000000 0.000000 0.137500 0.000000 0.006250 0.000000 0.000000 0.187500 0.000000
0.000000 0.062500 0.000000 0.000000 0.137500 0.268750 0.000000 0.068750 0.000000 0.000000
0.131250 0.000000 0.000000 0.000000 0.275000 0.000000 0.000000 0.000000 0.000000 0.206250
0.000000 0.000000 0.000000 0.000000 0.000000 0.256250 0.000000 0.000000 0.131250 0.000000
0.000000 0.131250 0.000000 0.137500 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000
0.006250 0.000000 0.131250 0.000000 0.137500 0.000000 0.000000 0.062500 0.000000 0.393750
0.000000 0.000000 0.000000 0.137500 0.000000 0.000000 0.000000 0.325000 0.125000 0.068750
0.000000 0.131250 0.193750 0.000000 0.012500 0.006250 0.000000 0.000000 0.000000 0.000000
0.000000 0.000000 0.000000 0.000000 0.137500 0.000000 0.262500 0.000000 0.068750 0.000000
0.000000 0.325000 0.000000 0.137500 0.000000 0.000000 0.000000 0.000000 0.068750 0.000000
0.131250 0.000000 0.312500 0.000000 0.000000 0.068750 0.000000 0.000000 0.000000 0.343750
0.000000 0.000000 0.137500 0.000000 0.000000 0.000000 0.006250 0.000000 0.000000 0.000000
```

MAST



Search Results

Top

Top Scoring Sequences ?

Each of the following 97 sequences has an *E*-value less than 10.
The motif matches shown have a position p-value less than 0.0001.
Click on the arrow (↧) next to the *E*-value to view more information about a sequence.

Change the portion of annotated sequence by **dragging the buttons**; hold shift to drag them individually.

Comment: A/Israel/277/2009 2009/04/ HA

Combined *p*-value: 2.16e-15

Annotated Sequence

```
                                                        Motif 1   Motif 3                 Motif 4     Motif 2                  Motif 3   Motif 1                   Motif 2
                                                        5.8e-06   1.3e-06                 1.6e-08     1.2e-10                  1.8e-07   4.5e-06                   1.2e-08
                                                        VERWYIQD  VNSYYR                  HDSNKG      CPHAGATS                 VNSYYRVERWYIQD                      CPHAGATS
                                                        ++++++++  ++++++                  ++++++      ++++++++                 ++++++++++++++                      ++++++++
----------GWILGNPECESLSTASSWSYIVETSSSDNGTCYPGDFIDYEELREQLSSVSSFERFEIFPKTSSWPNHDSNKGVT AACPHAGAKSFYKNLIWLVKKGNSYPKLSKSYINDKGKEVLVLWGIHHPSTSADQQSLYQNADAYVFVG
                                                        \______/  \____/                  \____/      \______/                 \____/\______/                      \______/
```

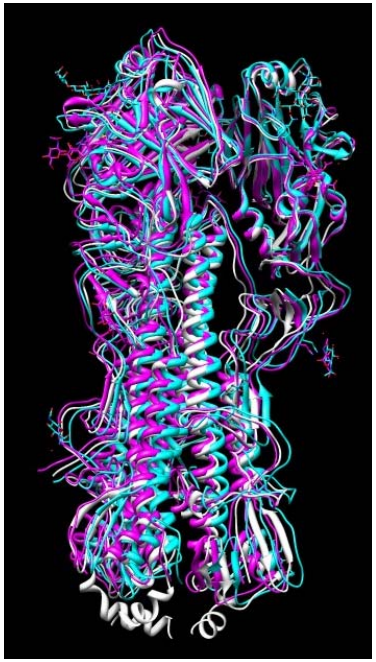
Sequence	*E*-value	Block Diagram
ACT21579	2.1e-13 ↧	
ACR15748	2.1e-13 ↧	
ADB89151	2.1e-13 ↧	
ACR08500	2.2e-13 ↧	
ACR08498	2.2e-13 ↧	
ACR78158	2.2e-13 ↧	
ACY77573	2.2e-13 ↧	
ACY77613	2.2e-13 ↧	
ACY77554	2.2e-13 ↧	
ACY77643	2.2e-13 ↧	

MAST_ImEp200vspandemic97

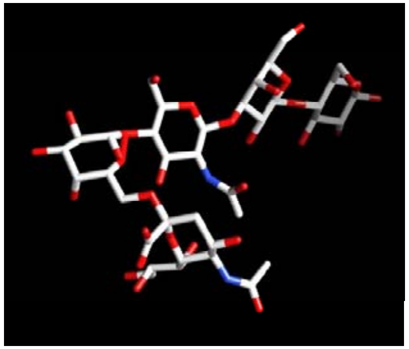
Aim 2: Model mutations known to occur in pandemic strains into the circulating swine flu strain

- Did some work on Aim 2 but could not carry it further since Aim 1 was not accomplished

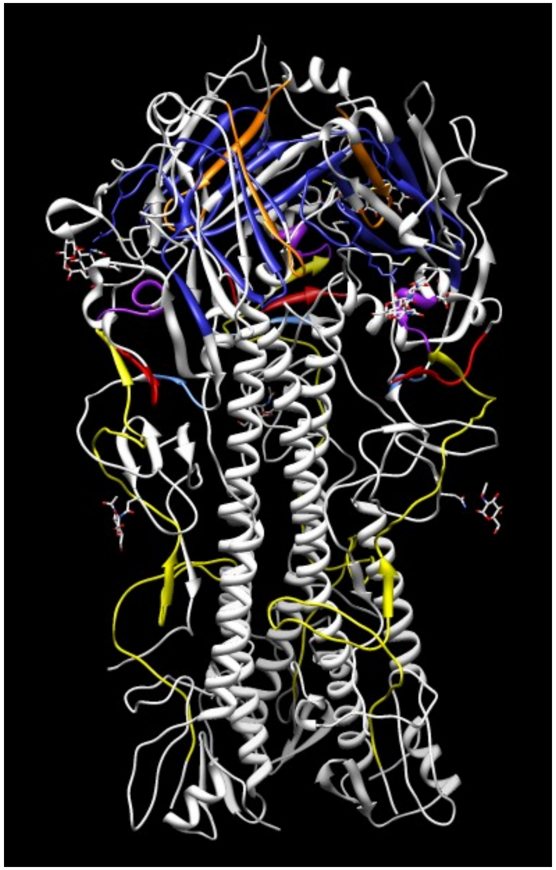
Chimera



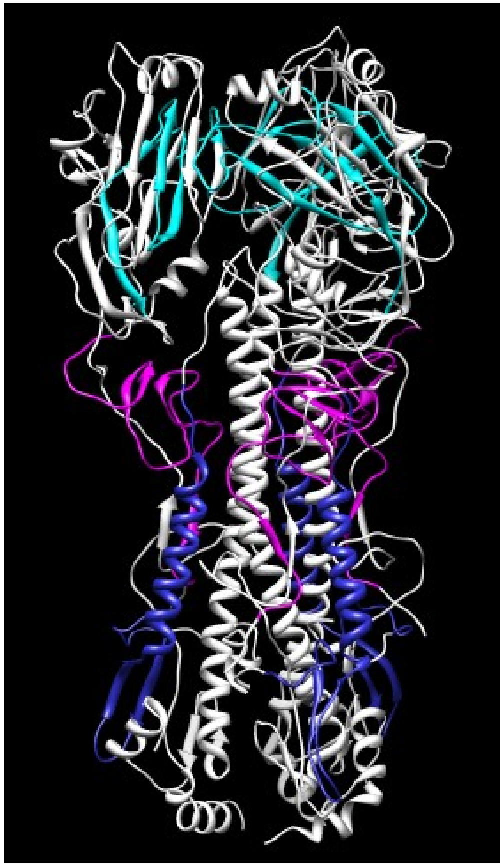
1RVX aligned with 3LZG and1RUY



1RVT LSTc



1RUY_mappedMEME swine motifs



3LZG_mapped MEME pandemic94 motifs

Results

- Found motifs using MEME
  - Changed number of expected repetitions
  - Changed length
  - Categorized by pandemic, seasonal and swine
  - Immune Epitopes
    - Overall MEME E-values showed that most motifs are statistically significant
- Submitted MEME results to MAST
  - Most pandemic motifs were found in swine and seasonal strains as well (and vice versa)
    - E-values were very small numbers signifying that the MEME motifs found in pandemic strains were conserved in swine and seasonal hemagglutinin strains as well

Results cont…

- Learned: Most hemagglutinin strains are conserved.
  - Which may explain why most papers focus more on specific amino acid changes instead of motifs
  - Difficult to tell whether a motif is significant (may result from functional, structural or evolutionary relationships between sequences)

- Next steps
  - Motifs found by MEME need to be analyzed further
    - background check through literature
    - Use other resources available through the internet to check the significance of the motifs

Significance

- Knowing signature motifs that correlate to potential virulence for cross-species infection of swine flu viruses into humans would be very useful for **diagnostic assay development**.
- It provides **markers** for monitoring the circulating swine flu viruses and whether the changes occurring would give rise to **potentially pandemic** viruses.
- Further research can also be done on **other proteins** of the virus to see if having the same signature motifs do indeed suggest potential for virulence.

References

- Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990) "Basic local alignment search tool." J. Mol. Biol. 215:403-410.
- Bailey, T.L. and Elkan, C. (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology, August In Altman, R.B., Brutlag, D.L., Karp, P.D., Lathrop, R.H., Searls, D.B. (Eds.). Menlo Park, CA AAAI Press pp. 28–36 .
- Bailey, T.L. and Gribskov, M. (1998) 'Combining evidence using P-values: application to sequence homology searches *Bioinformatics*, **14**, 48–54.
- Bailey TL, Williams N, Misleh C, Li WW (2006) MEME: discovering and analyzing DNA and protein sequence motifs. Nucleic Acids Res 34: W369–373.
- Bao Y., P. Bolotov, D. Dernovoy, B. Kiryutin, L. Zaslavsky, T. Tatusova, J. Ostell, and D. Lipman. The Influenza Virus Resource at the National Center for Biotechnology Information. J. Virol. 2008 Jan;82(2):596-601.
- Influenza A Virus From Wikipedia, the free encyclopedia; http://en.wikipedia.org/wik i/Influenza_A_virus
- Jameel, Shahid, The 2009 influenza pandemic. Current Science, Vol. 98, No. 3, 10 Feb 2010. <http://www.ias.ac.in/currsci/10feb2010/306.pdf >
- Vita R, Zarebski L, Greenbaum JA, Emami H, Hoof I, Salimi N, Damle R, Sette A, Peters B. The immune epitope database 2.0. Nucleic Acids Res. 2010 Jan;38(Database issue):D854-62. Epub 2009 Nov 11.
- Garten R. J., Davis C. T., Russell C. A., Shu B., Lindstrom S., Balish A., Sessions W. M., Xu X., Skepner E., Deyde V., Okomo-Adhiambo M., Gubareva L., Barnes J., Smith C. B., Emery S. L., Hillman M. J., Rivailler P., Smagala J., de Graaf M., Burke D. F., Fouchier R. A., Pappas C., Alpuche-Aranda C. M., Lopez-Gatell H., Olivera H., Lopez I., Myers C. A., Faix D., Blair P. J., Yu C., Keene K. M., Dotson P. D., Jr., Boxrud D., Sambol A. R., Abid S. H., George K. St., Bannerman T., Moore A. L., Stringer D. J., Blevins P., Demmler-Harrison G. J., Ginsberg M., Kriner P., Waterman S., Smole S., Guevara H. F., Belongia E. A., Clark P. A., Beatrice S. T., Donis R., Katz J., Finelli L., Bridges C. B., Shaw M., Jernigan D. B., Uyeki T. M., Smith D. J., Klimov A. I., Cox N. J. 2009. Antigenic and genetic characteristics of swine-origin 2009 A(H1N1) influenza viruses circulating in humans. Science 325:197–201
- Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem. 2004 Oct;25(13):1605-12.
- http://www.cdc.gov/flu/swineflu/key_facts.htm
- Immune Epitope Database and Analysis Resource: www.immuneepitope.org

Acknowledgements

- UCSD: National Biomedical Computation Resource (NBCR)
  - Dr.Wilfred Li
- CNIC, CAS
  - Dr. Kai Nan
  - Dr. Jianjun Yu
  - Guangyuan Liu
  - Haiyan Xu
  - Wei Chen
- PRIME
  - Dr. Gabriele Wiesenhausen
  - Dr. Peter Arzberger
  - Teri Simas
  - Jim Galvin
  - Tricia Taylor-Oliveira
- National Science Foundation, IOSE-0710726

中国科学院
CHINESE ACADEMY OF SCIENCES

UC San Diego

PRIME
PACIFIC RIM UNDERGRADUATE EXPERIENCES

中国科学院
计算机网络信息中心
Computer Network Information Center
Chinese Academy of Sciences

NSF

PACIFIC RIM APPLICATIONS AND GRID MIDDLEWARE ASSEMBLY
PRAGMA

Calit2

NBCR