worksheet

Influenza – an evolving problem!

Many vaccinations provide immunity for a number of years. Some even protect you for life.

So why do doctors recommend that you get an influenza vaccination each year?

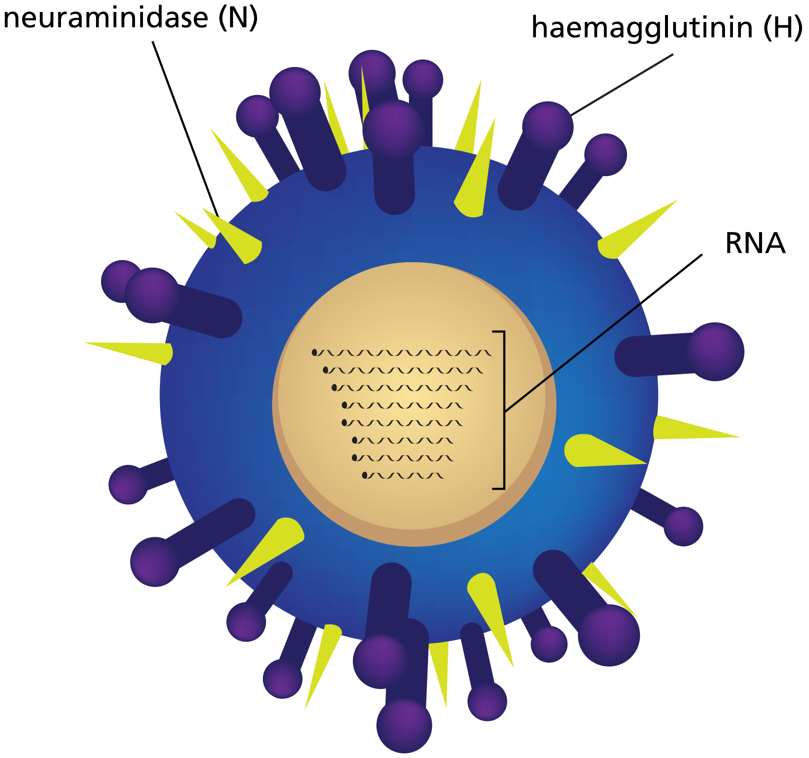
Well, your immune system has a memory of past infections; some memory is long-lived and prevents you from catching a disease more than once. If reinfected by a pathogen that has previously made you ill, your immune system will usually act quickly to attack it, often preventing onset of any major symptoms. However, the immune memory, whilst effective, is specific, so changes that occur to pathogens over time may make them unrecognisable to your immune system.

Influenza is a virus that multiplies quickly and is prone to replication errors or mutations in its RNA sequence.

If mutations occur in the segment that codes for two surface proteins (H and N) on the virus, then your existing antibodies may not recognise the virus and will not bind to it. Thus the virus will more easily infect your cells.

Changes that occur to influenza viruses occur rapidly as they move around the world. Strains that develop in a local population one year may be different from those found the following year. This explains why doctors recommend some people should be vaccinated for influenza every year.

Each year a panel of scientists reviews influenza strains currently infecting people around the world, then makes recommendations for strains that should be included in vaccines.



structure of influenza A virus

The Influenza Research Database (IRD)

Sometimes, when you go to the doctor with influenza-like symptoms, a sample of mucus or blood may be taken and sent away for testing. This is done to determine exactly what's making you ill, so you can be treated appropriately. Sometimes this information may be added to a database (without your identifying details) to allow scientists to keep track of strains of viruses and bacteria present in populations. Over time, these data build a picture about what is happening to pathogen populations around the world.

The Influenza Research Database contains information about many different strains of influenza, from many different populations. In this activity you will focus on influenza virus H1N1 which caused the swine 'flu pandemic in humans in 2009. You will see how this strain of influenza was different from those in previous years. You'll compare strains, using bioinformatics tools such as multiple sequence alignment and phylogenetic trees, to examine how mutations change the relatedness of some strains of influenza.

To use the database follow the steps below. Screenshots are provided to help you navigate the website.

Before you start

Check that the browser you are using supports Java by going to http://www.java.com/testjava/
More information is available in the teachers guide (**Technical requirements**).

Overview of database use

Ⓐ Go to http://fludb.org, log-in and get a workbench space. You need a workbench space to store your searches.

↓

Ⓑ Search for sequence segments of H1N1.

↓

Ⓒ Find information about the virus strain, H1N1.

↓

Ⓓ Align segment sequences.

↓

Ⓔ Investigate mutations (nucleotide differences) between aligned segments.

↓

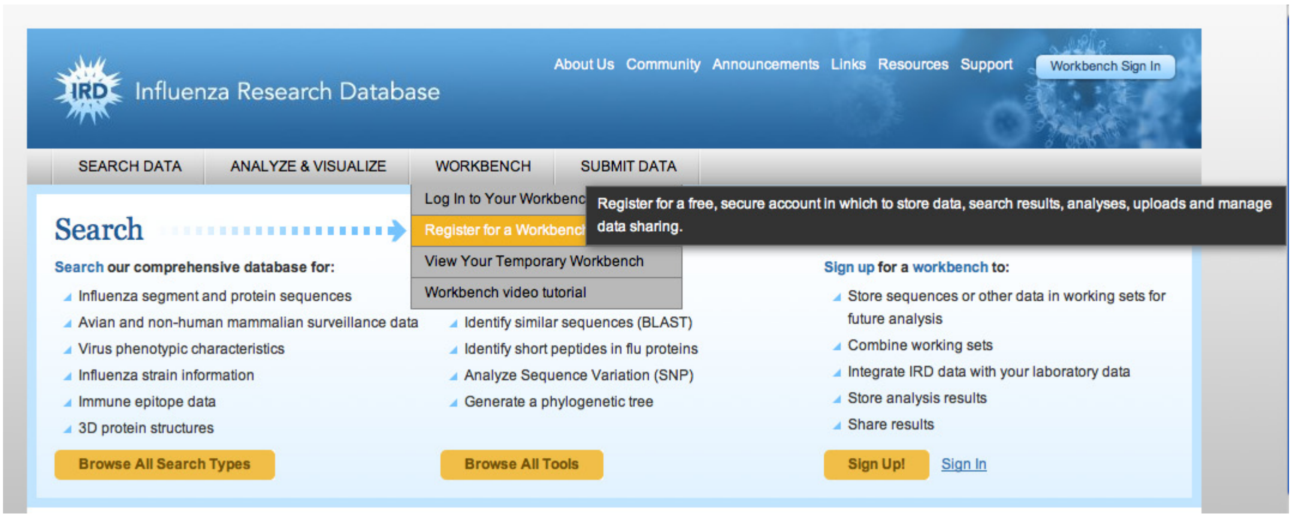
Ⓕ Visualise mutations using a phylogenetic tree.

Ⓐ Register for a workbench

The workbench enables you to save searches and recall data. To produce a meaningful phylogenetic tree you may need to run several searches and save your data as you go. To register for a workbench, select **Register for a workbench** and follow the instructions. There are also various tutorials about the IRD that you can access from the blue banner, by selecting **Support**.

Record your login details as you register for a workbench.

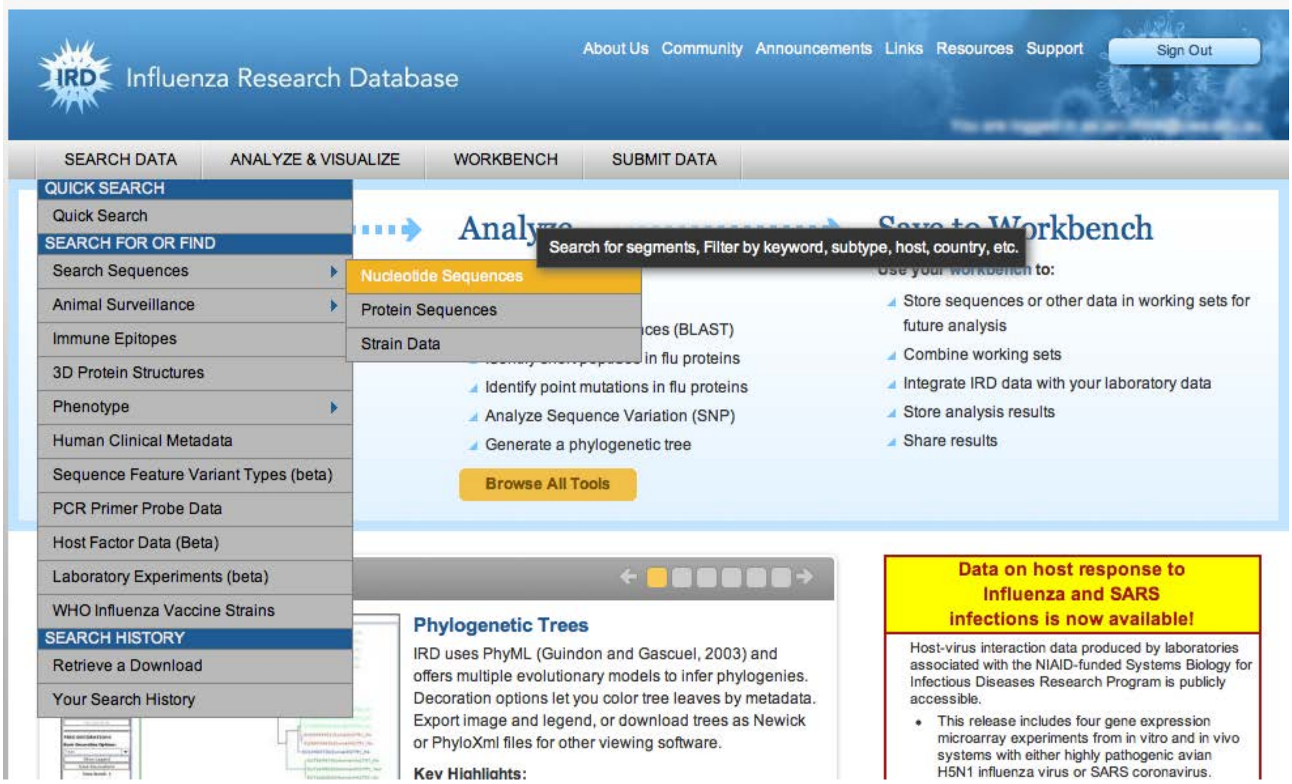
Login details required: email address and password.



Register for a free, secure account in which to store data, search results, analyses, uploads and manage data sharing.

Ⓑ Search for H1N1

From the grey navigation bar roll over **SEARCH DATA**, then **Search Sequences** and select **Nucleotide Sequences**.



Search for segments, Filter by keyword, subtype, host, country, etc.

Phylogenetic Trees

IRD uses PhyML (Guindon and Gascuel, 2003) and offers multiple evolutionary models to infer phylogenies. Decoration options let you color tree leaves by metadata. Export image and legend, or download trees as Newick or PhyloXml files for other viewing software.

Key Highlights:

Data on host response to Influenza and SARS infections is now available!

Host-virus interaction data produced by laboratories associated with the NIAID-funded Systems Biology for Infectious Diseases Research Program is publicly accessible.

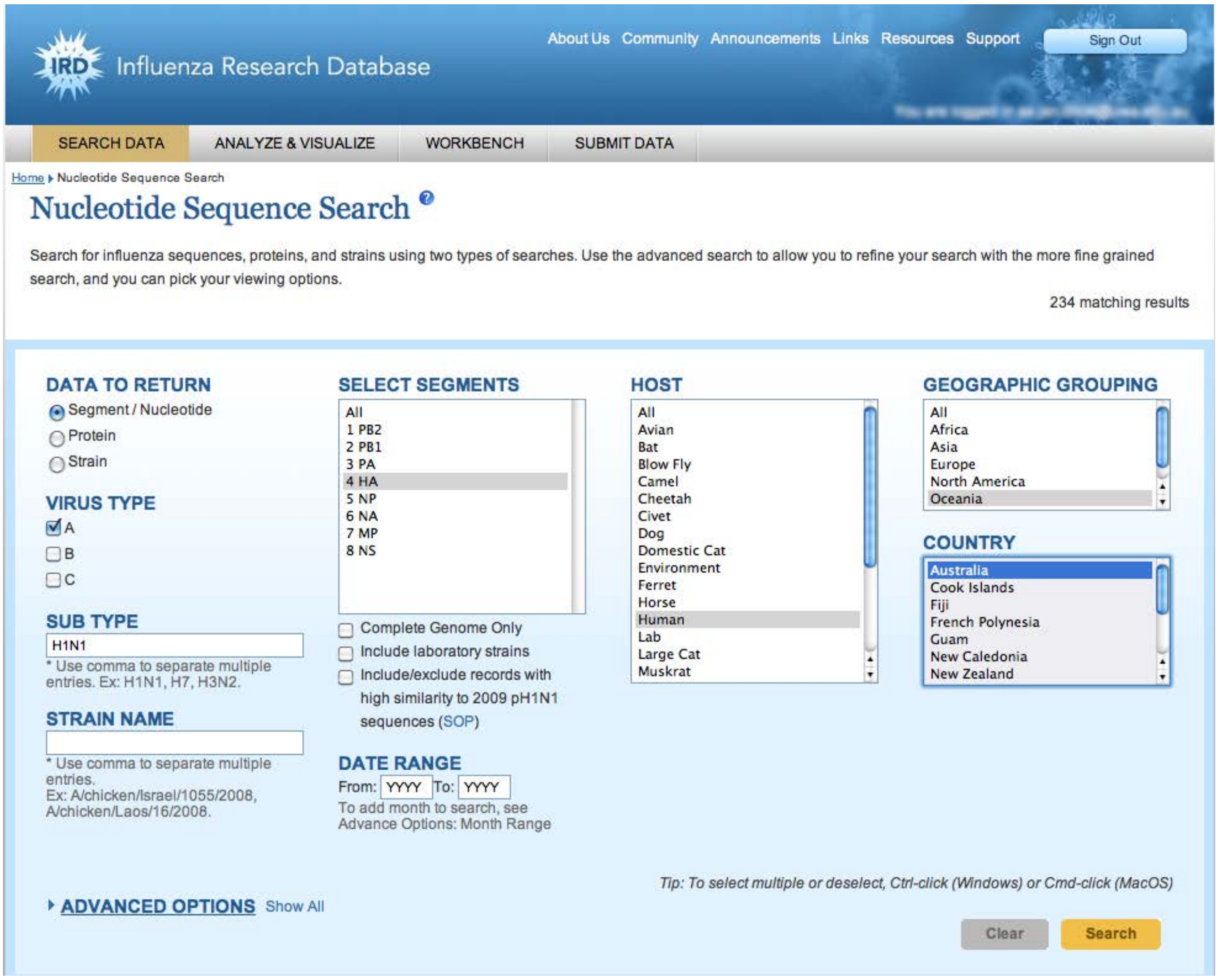
- This release includes four gene expression microarray experiments from in vitro and in vivo systems with either highly pathogenic avian H5N1 influenza virus or SARS coronavirus.

Nucleotide sequence search

Select or enter the following settings:

VIRUS TYPE	A
SELECT SEGMENTS	4 HA
SUB TYPE	H1N1
HOST	Human
GEOGRAPHIC GROUPING	Oceania
COUNTRY	Australia

Leave the other fields (**STRAIN NAME** and **DATE RANGE**) blank.



Nucleotide Sequence Search

Search for influenza sequences, proteins, and strains using two types of searches. Use the advanced search to allow you to refine your search with the more fine grained search, and you can pick your viewing options.

234 matching results

DATA TO RETURN

● Segment / Nucleotide

○ Protein

○ Strain

VIRUS TYPE

☑ A

☐ B

☐ C

SUB TYPE

H1N1

* Use comma to separate multiple entries. Ex: H1N1, H7, H3N2.

STRAIN NAME

* Use comma to separate multiple entries. Ex: A/chicken/Israel/1055/2008, A/chicken/Laos/16/2008.

SELECT SEGMENTS

4 HA

☐ Complete Genome Only

☐ Include laboratory strains

☐ Include/exclude records with high similarity to 2009 pH1N1 sequences (SOP)

DATE RANGE

From: YYYY To: YYYY

To add month to search, see Advance Options: Month Range

HOST

Human

GEOGRAPHIC GROUPING

Oceania

COUNTRY

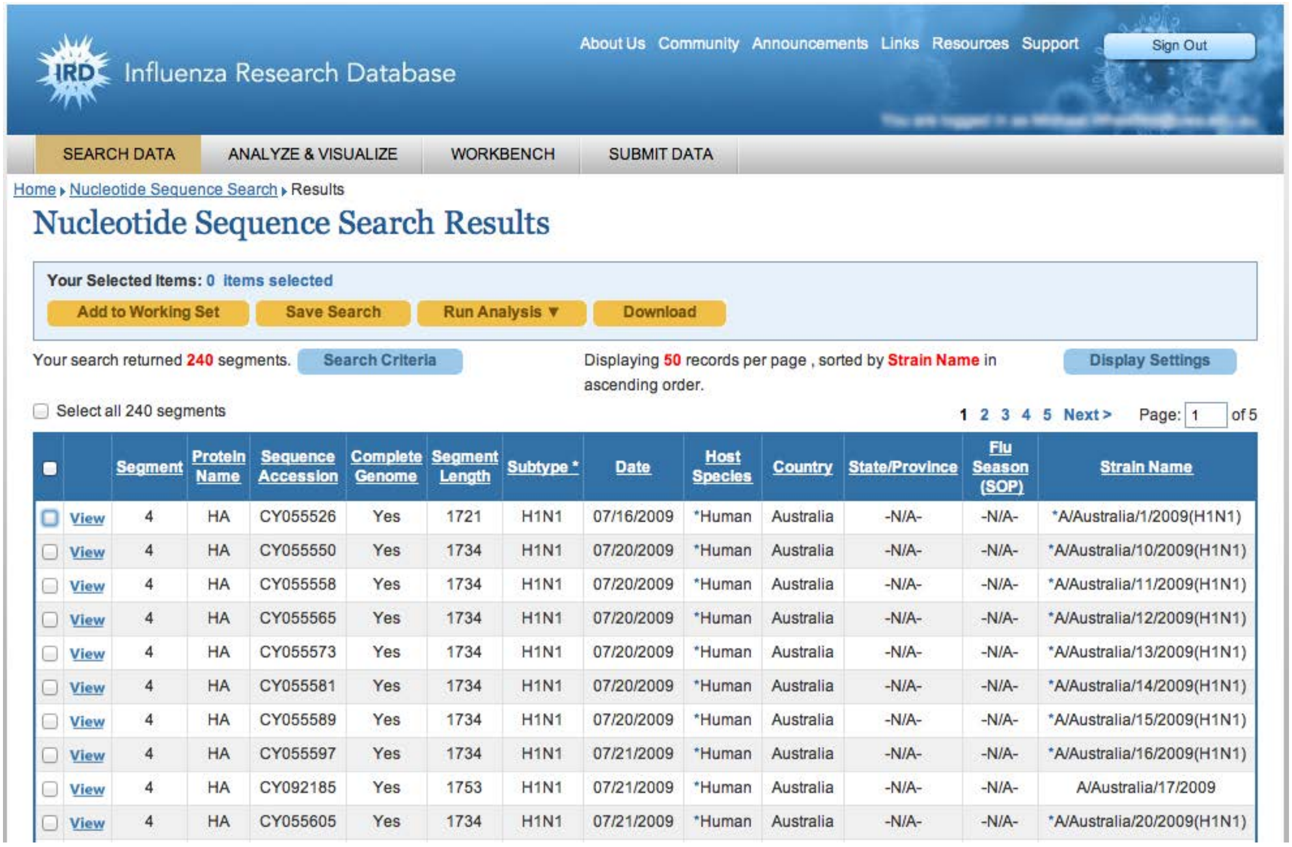
Australia

Tip: To select multiple or deselect, Ctrl-click (Windows) or Cmd-click (MacOS)

Select **Search**.

You're searching for H1N1 influenza viruses in Australia, a common type of influenza that affects humans and the one responsible for the global 'swine flu' outbreak. You'll get a list of many strains!

Note: Your results may differ from the screenshots that appear in this worksheet as the database is constantly updated.



Nucleotide Sequence Search Results

Your Selected Items: 0 items selected

Your search returned **240** segments.

Displaying **50** records per page , sorted by **Strain Name** in ascending order.

☐ Select all 240 segments

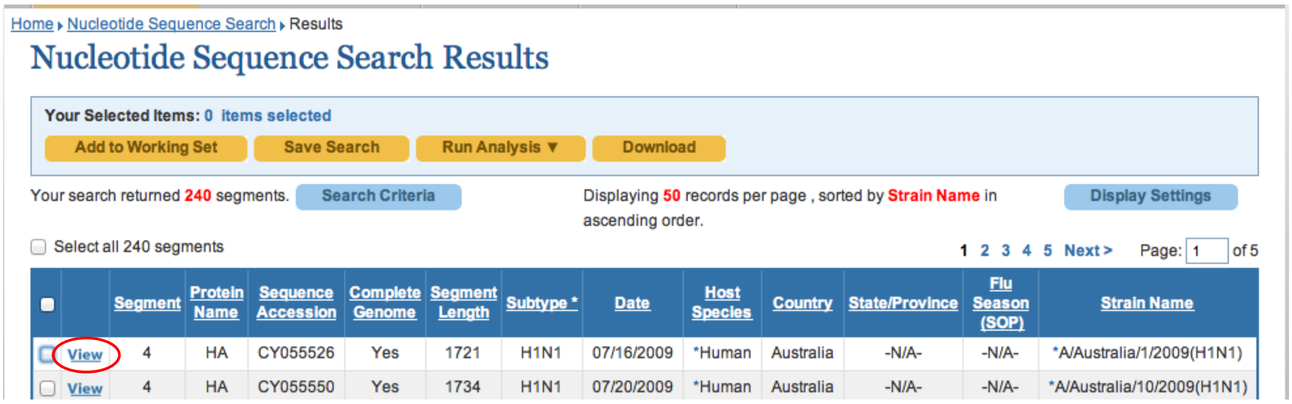
1 2 3 4 5 Next > Page: 1 of 5

☐		Segment	Protein Name	Sequence Accession	Complete Genome	Segment Length	Subtype *	Date	Host Species	Country	State/Province	Flu Season (SOP)	Strain Name
☐	View	4	HA	CY055526	Yes	1721	H1N1	07/16/2009	*Human	Australia	-N/A-	-N/A-	*A/Australia/1/2009(H1N1)
☐	View	4	HA	CY055550	Yes	1734	H1N1	07/20/2009	*Human	Australia	-N/A-	-N/A-	*A/Australia/10/2009(H1N1)
☐	View	4	HA	CY055558	Yes	1734	H1N1	07/20/2009	*Human	Australia	-N/A-	-N/A-	*A/Australia/11/2009(H1N1)
☐	View	4	HA	CY055565	Yes	1734	H1N1	07/20/2009	*Human	Australia	-N/A-	-N/A-	*A/Australia/12/2009(H1N1)
☐	View	4	HA	CY055573	Yes	1734	H1N1	07/20/2009	*Human	Australia	-N/A-	-N/A-	*A/Australia/13/2009(H1N1)
☐	View	4	HA	CY055581	Yes	1734	H1N1	07/20/2009	*Human	Australia	-N/A-	-N/A-	*A/Australia/14/2009(H1N1)
☐	View	4	HA	CY055589	Yes	1734	H1N1	07/20/2009	*Human	Australia	-N/A-	-N/A-	*A/Australia/15/2009(H1N1)
☐	View	4	HA	CY055597	Yes	1734	H1N1	07/21/2009	*Human	Australia	-N/A-	-N/A-	*A/Australia/16/2009(H1N1)
☐	View	4	HA	CY092185	Yes	1753	H1N1	07/21/2009	*Human	Australia	-N/A-	-N/A-	A/Australia/17/2009
☐	View	4	HA	CY055605	Yes	1734	H1N1	07/21/2009	*Human	Australia	-N/A-	-N/A-	*A/Australia/20/2009(H1N1)

Ⓒ Segment information

When samples are collected from people, additional information is gathered to build up a profile of how a virus is establishing itself in a population. Select **View** for one segment to view details of the strain.

Using strain and segment/protein information provided for your selected segment, answer the questions on the next page. There's a lot of information so you'll need to search carefully, and scroll down the page.



Nucleotide Sequence Search Results

Your Selected Items: 0 items selected

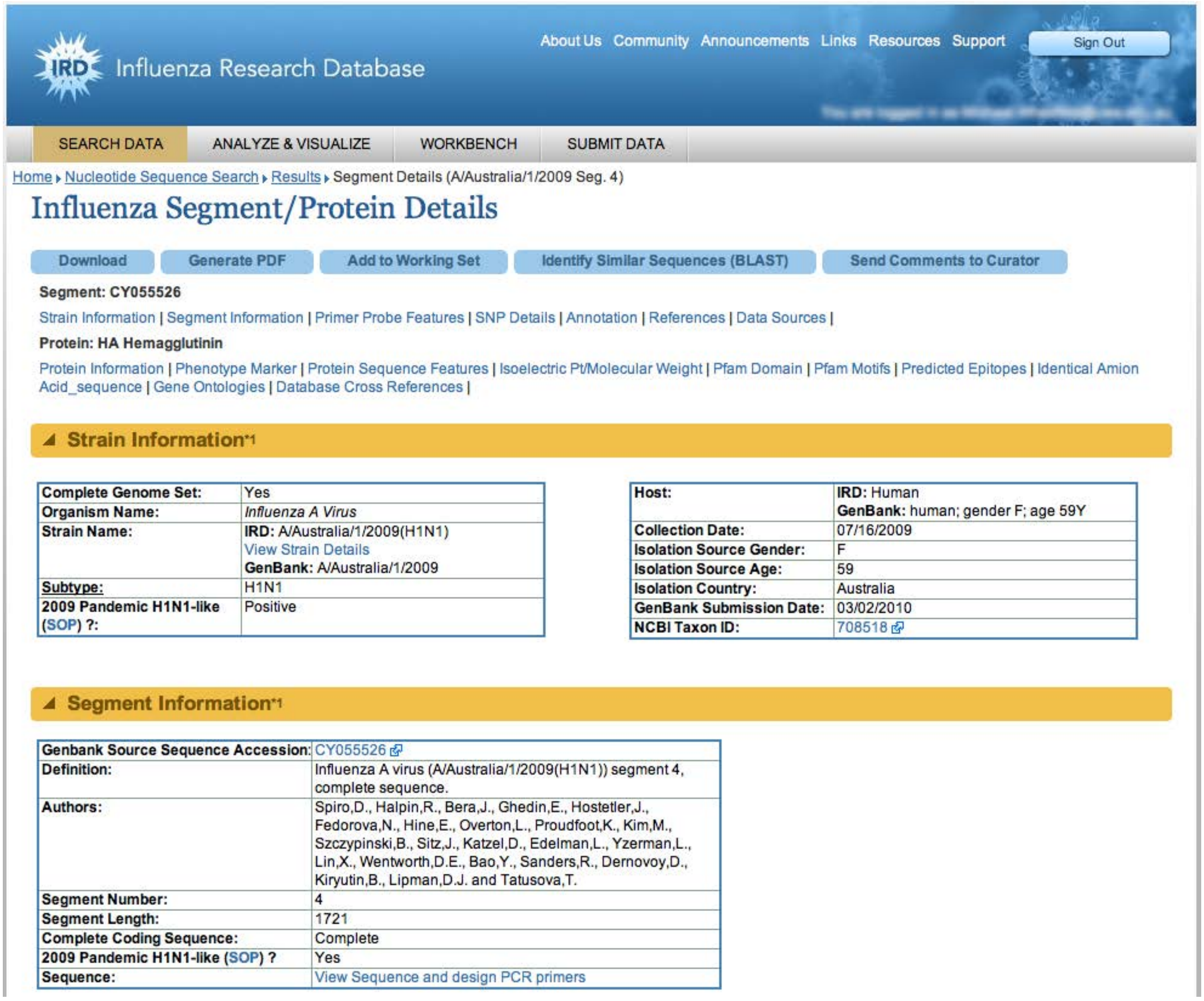
Your search returned **240** segments.

Displaying **50** records per page , sorted by **Strain Name** in ascending order.

☐ Select all 240 segments

1 2 3 4 5 Next > Page: 1 of 5

☐		Segment	Protein Name	Sequence Accession	Complete Genome	Segment Length	Subtype *	Date	Host Species	Country	State/Province	Flu Season (SOP)	Strain Name
☐	View	4	HA	CY055526	Yes	1721	H1N1	07/16/2009	*Human	Australia	-N/A-	-N/A-	*A/Australia/1/2009(H1N1)
☐	View	4	HA	CY055550	Yes	1734	H1N1	07/20/2009	*Human	Australia	-N/A-	-N/A-	*A/Australia/10/2009(H1N1)



Influenza Segment/Protein Details

Segment: CY055526

Protein: HA Hemagglutinin

Strain Information*1

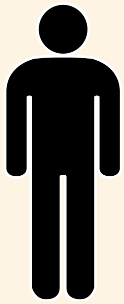
Complete Genome Set:	Yes
Organism Name:	*Influenza A Virus*
Strain Name:	**IRD:** A/Australia/1/2009(H1N1) View Strain Details **GenBank:** A/Australia/1/2009
Subtype:	H1N1
2009 Pandemic H1N1-like (SOP) ?:	Positive

Host:	**IRD:** Human **GenBank:** human; gender F; age 59Y
Collection Date:	07/16/2009
Isolation Source Gender:	F
Isolation Source Age:	59
Isolation Country:	Australia
GenBank Submission Date:	03/02/2010
NCBI Taxon ID:	708518

Segment Information*1

Genbank Source Sequence Accession:	CY055526
Definition:	Influenza A virus (A/Australia/1/2009(H1N1)) segment 4, complete sequence.
Authors:	Spiro,D., Halpin,R., Bera,J., Ghedin,E., Hostetler,J., Fedorova,N., Hine,E., Overton,L., Proudfoot,K., Kim,M., Szczypinski,B., Sitz,J., Katzel,D., Edelman,L., Yzerman,L., Lin,X., Wentworth,D.E., Bao,Y., Sanders,R., Dernovoy,D., Kiryutin,B., Lipman,D.J. and Tatusova,T.
Segment Number:	4
Segment Length:	1721
Complete Coding Sequence:	Complete
2009 Pandemic H1N1-like (SOP) ?	Yes
Sequence:	View Sequence and design PCR primers

1. Complete the following details:



Name of the strain (look under the yellow banner - **Strain Information**): ....

Isolation year: ....

Age and gender of the individual (look under the box **Host**): ....

How long is the segment? (yellow banner **Segment Information**): ....

How many amino acids does it encode? (scroll down to **Segment Annotation** – protein length)

....

2. What does H1N1 refer to?

....

3. Your search has yielded hundreds of strains. In terms of virus evolution, explain this.

....

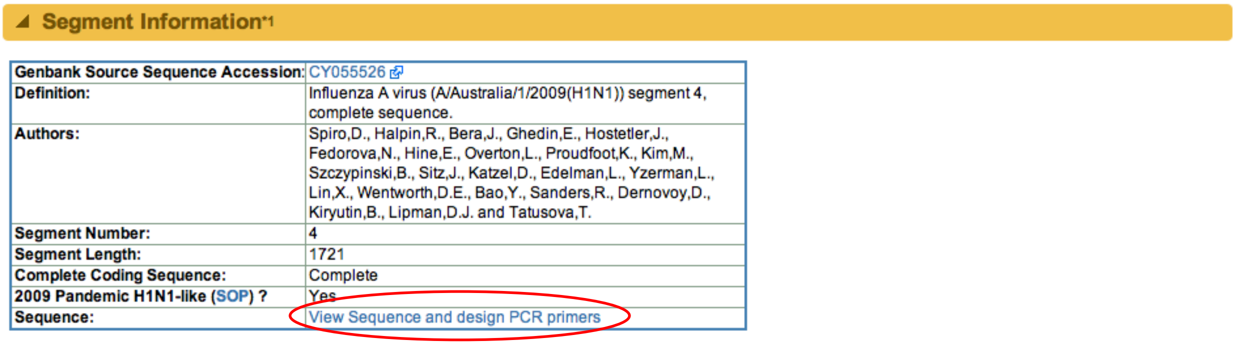
....

View sequence

Stay on the current page (**Influenza Segment / Protein Details**).

Some collected samples are sequenced, which means that genetic material is analysed and nucleotide order determined. This precise analysis enables comparison of strains to determine where mutations have occurred.

Under **Segment Information** select **View Sequence and design PCR primers** to retrieve the sequence.



Segment Information*1

Genbank Source Sequence Accession:	CY055526
Definition:	Influenza A virus (A/Australia/1/2009(H1N1)) segment 4, complete sequence.
Authors:	Spiro,D., Halpin,R., Bera,J., Ghedin,E., Hostetler,J., Fedorova,N., Hine,E., Overton,L., Proudfoot,K., Kim,M., Szczypinski,B., Sitz,J., Katzel,D., Edelman,L., Yzerman,L., Lin,X., Wentworth,D.E., Bao,Y., Sanders,R., Dernovoy,D., Kiryutin,B., Lipman,D.J. and Tatusova,T.
Segment Number:	4
Segment Length:	1721
Complete Coding Sequence:	Complete
2009 Pandemic H1N1-like (SOP) ?	Yes
Sequence:	View Sequence and design PCR primers

You should get a large table of nucleotides showing your results. This is just a portion of the whole strain; it may be the beginning, middle or end of a viral sequence.

An example of a genomic sequence follows on the next page.

4. What do letters in the sequence represent?

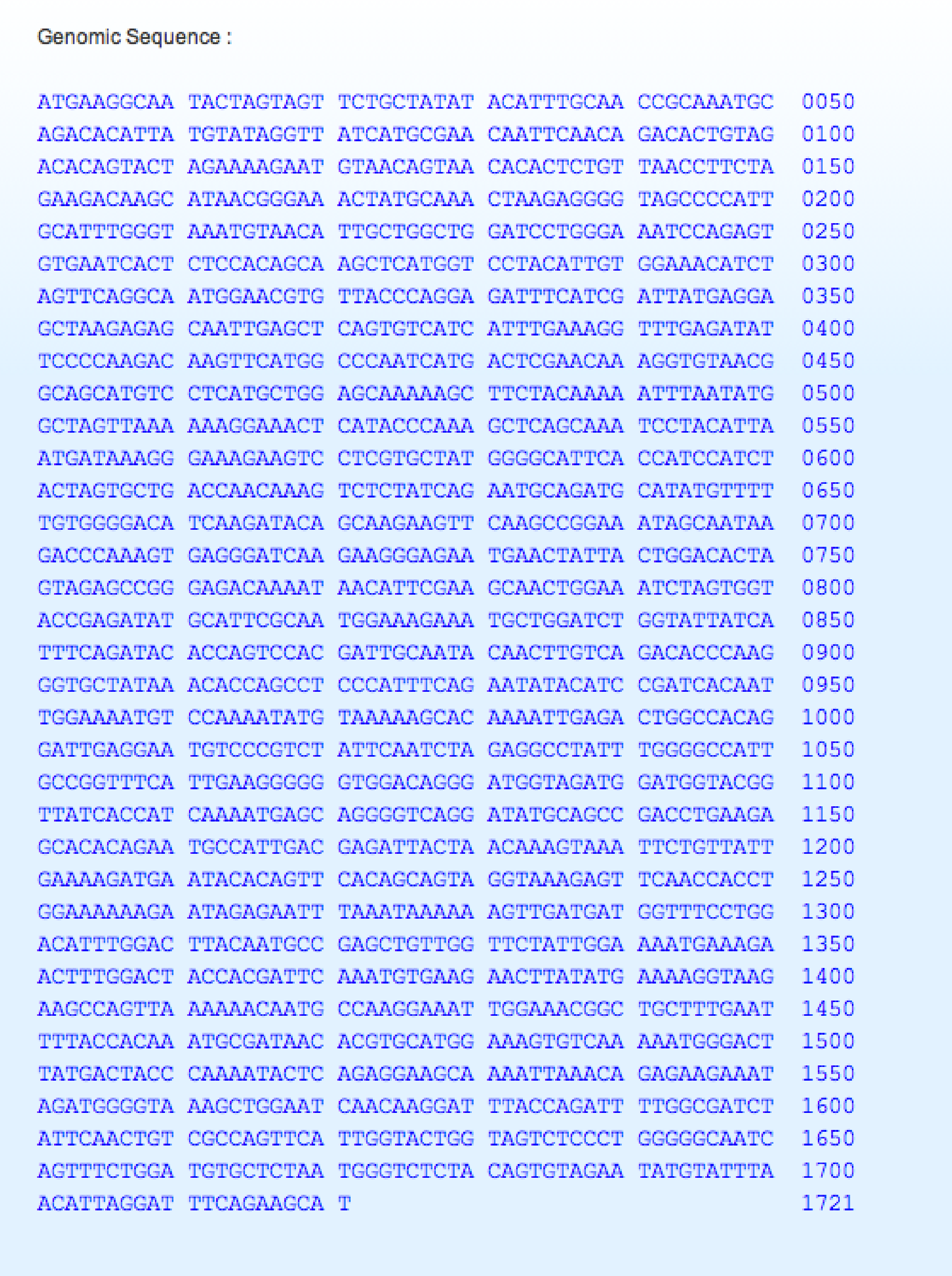
....

....

5. Why is the order of letters important?

....

....



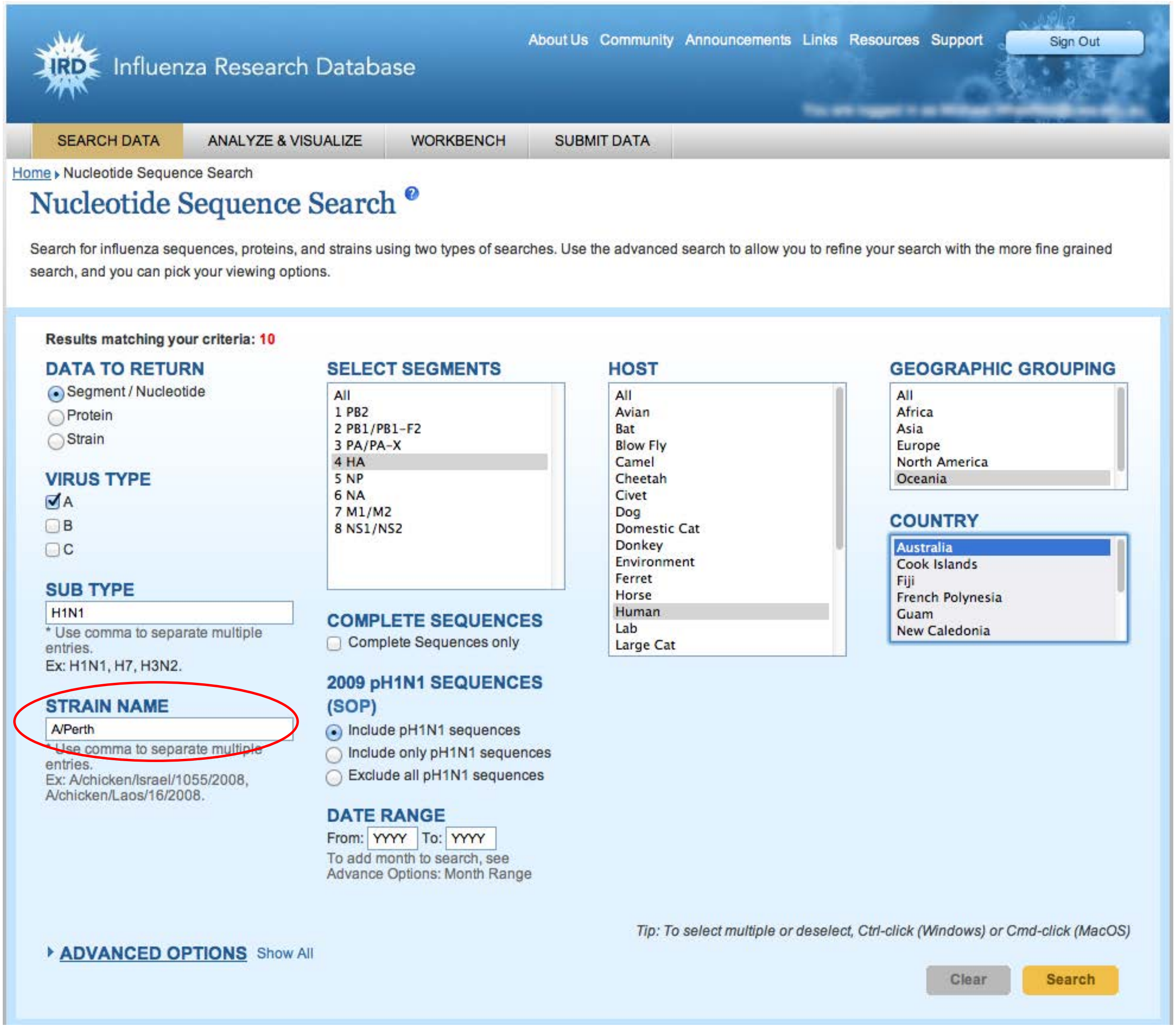
Genomic Sequence :

```
ATGAAGGCAA TACTAGTAGT TCTGCTATAT ACATTTGCAA CCGCAAATGC  0050
AGACACATTA TGTATAGGTT ATCATGCGAA CAATTCAACA GACACTGTAG  0100
ACACAGTACT AGAAAAGAAT GTAACAGTAA CACACTCTGT TAACCTTCTA  0150
GAAGACAAGC ATAACGGGAA ACTATGCAAA CTAAGAGGGG TAGCCCCATT  0200
GCATTTGGGT AAATGTAACA TTGCTGGCTG GATCCTGGGA AATCCAGAGT  0250
GTGAATCACT CTCCACAGCA AGCTCATGGT CCTACATTGT GGAAACATCT  0300
AGTTCAGGCA ATGGAACGTG TTACCCAGGA GATTTCATCG ATTATGAGGA  0350
GCTAAGAGAG CAATTGAGCT CAGTGTCATC ATTTGAAAGG TTTGAGATAT  0400
TCCCCAAGAC AAGTTCATGG CCCAATCATG ACTCGAACAA AGGTGTAACG  0450
GCAGCATGTC CTCATGCTGG AGCAAAAAGC TTCTACAAAA ATTTAATATG  0500
GCTAGTTAAA AAAGGAAACT CATACCCAAA GCTCAGCAAA TCCTACATTA  0550
ATGATAAAGG GAAAGAAGTC CTCGTGCTAT GGGGCATTCA CCATCCATCT  0600
ACTAGTGCTG ACCAACAAAG TCTCTATCAG AATGCAGATG CATATGTTTT  0650
TGTGGGGACA TCAAGATACA GCAAGAAGTT CAAGCCGGAA ATAGCAATAA  0700
GACCCAAAGT GAGGGATCAA GAAGGGAGAA TGAACTATTA CTGGACACTA  0750
GTAGAGCCGG GAGACAAAAT AACATTCGAA GCAACTGGAA ATCTAGTGGT  0800
ACCGAGATAT GCATTCGCAA TGGAAAGAAA TGCTGGATCT GGTATTATCA  0850
TTTCAGATAC ACCAGTCCAC GATTGCAATA CAACTTGTCA GACACCCAAG  0900
GGTGCTATAA ACACCAGCCT CCCATTTCAG AATATACATC CGATCACAAT  0950
TGGAAAATGT CCAAAATATG TAAAAAGCAC AAAATTGAGA CTGGCCACAG  1000
GATTGAGGAA TGTCCCGTCT ATTCAATCTA GAGGCCTATT TGGGGCCATT  1050
GCCGGTTTCA TTGAAGGGGG GTGGACAGGG ATGGTAGATG GATGGTACGG  1100
TTATCACCAT CAAAATGAGC AGGGGTCAGG ATATGCAGCC GACCTGAAGA  1150
GCACACAGAA TGCCATTGAC GAGATTACTA ACAAAGTAAA TTCTGTTATT  1200
GAAAAGATGA ATACACAGTT CACAGCAGTA GGTAAAGAGT TCAACCACCT  1250
GGAAAAAAGA ATAGAGAATT TAAATAAAAA AGTTGATGAT GGTTTCCTGG  1300
ACATTTGGAC TTACAATGCC GAGCTGTTGG TTCTATTGGA AAATGAAAGA  1350
ACTTTGGACT ACCACGATTC AAATGTGAAG AACTTATATG AAAAGGTAAG  1400
AAGCCAGTTA AAAAACAATG CCAAGGAAAT TGGAAACGGC TGCTTTGAAT  1450
TTTACCACAA ATGCGATAAC ACGTGCATGG AAAGTGTCAA AAATGGGACT  1500
TATGACTACC CAAAATACTC AGAGGAAGCA AAATTAAACA GAGAAGAAAT  1550
AGATGGGGTA AAGCTGGAAT CAACAAGGAT TTACCAGATT TTGGCGATCT  1600
ATTCAACTGT CGCCAGTTCA TTGGTACTGG TAGTCTCCCT GGGGGCAATC  1650
AGTTTCTGGA TGTGCTCTAA TGGGTCTCTA CAGTGTAGAA TATGTATTTA  1700
ACATTAGGAT TTCAGAAGCA T                                 1721
```

Perth H1N1 pandemic 2009

Search for H1N1 viruses identified in Perth.

To do this, return to the **HOME** page and repeat your search, only this time specify **STRAIN NAME** as A/Perth. This should yield ten segments, from 2006 onwards (remember to specify **SUB TYPE** as H1N1).



Home ▸ Nucleotide Sequence Search

Nucleotide Sequence Search

Search for influenza sequences, proteins, and strains using two types of searches. Use the advanced search to allow you to refine your search with the more fine grained search, and you can pick your viewing options.

Results matching your criteria: 10

DATA TO RETURN
● Segment / Nucleotide
○ Protein
○ Strain

VIRUS TYPE
☑ A
☐ B
☐ C

SUB TYPE: H1N1
* Use comma to separate multiple entries.
Ex: H1N1, H7, H3N2.

STRAIN NAME: A/Perth
* Use comma to separate multiple entries.
Ex: A/chicken/Israel/1055/2008, A/chicken/Laos/16/2008.

SELECT SEGMENTS: All, 1 PB2, 2 PB1/PB1-F2, 3 PA/PA-X, 4 HA (selected), 5 NP, 6 NA, 7 M1/M2, 8 NS1/NS2

COMPLETE SEQUENCES
☐ Complete Sequences only

2009 pH1N1 SEQUENCES (SOP)
● Include pH1N1 sequences
○ Include only pH1N1 sequences
○ Exclude all pH1N1 sequences

DATE RANGE
From: YYYY To: YYYY
To add month to search, see Advance Options: Month Range

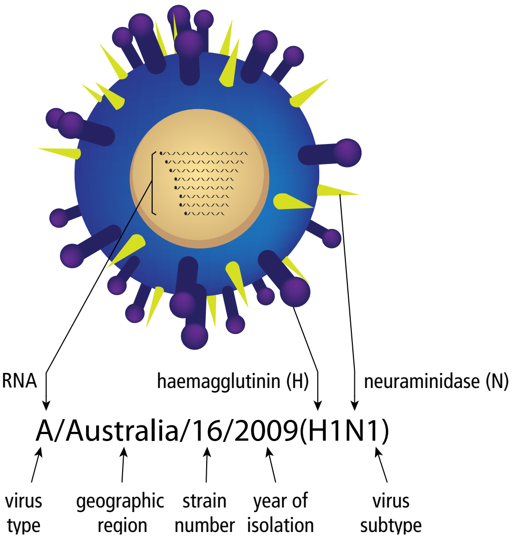
HOST: All, Avian, Bat, Blow Fly, Camel, Cheetah, Civet, Dog, Domestic Cat, Donkey, Environment, Ferret, Horse, Human (selected), Lab, Large Cat

GEOGRAPHIC GROUPING: All, Africa, Asia, Europe, North America, Oceania (selected)

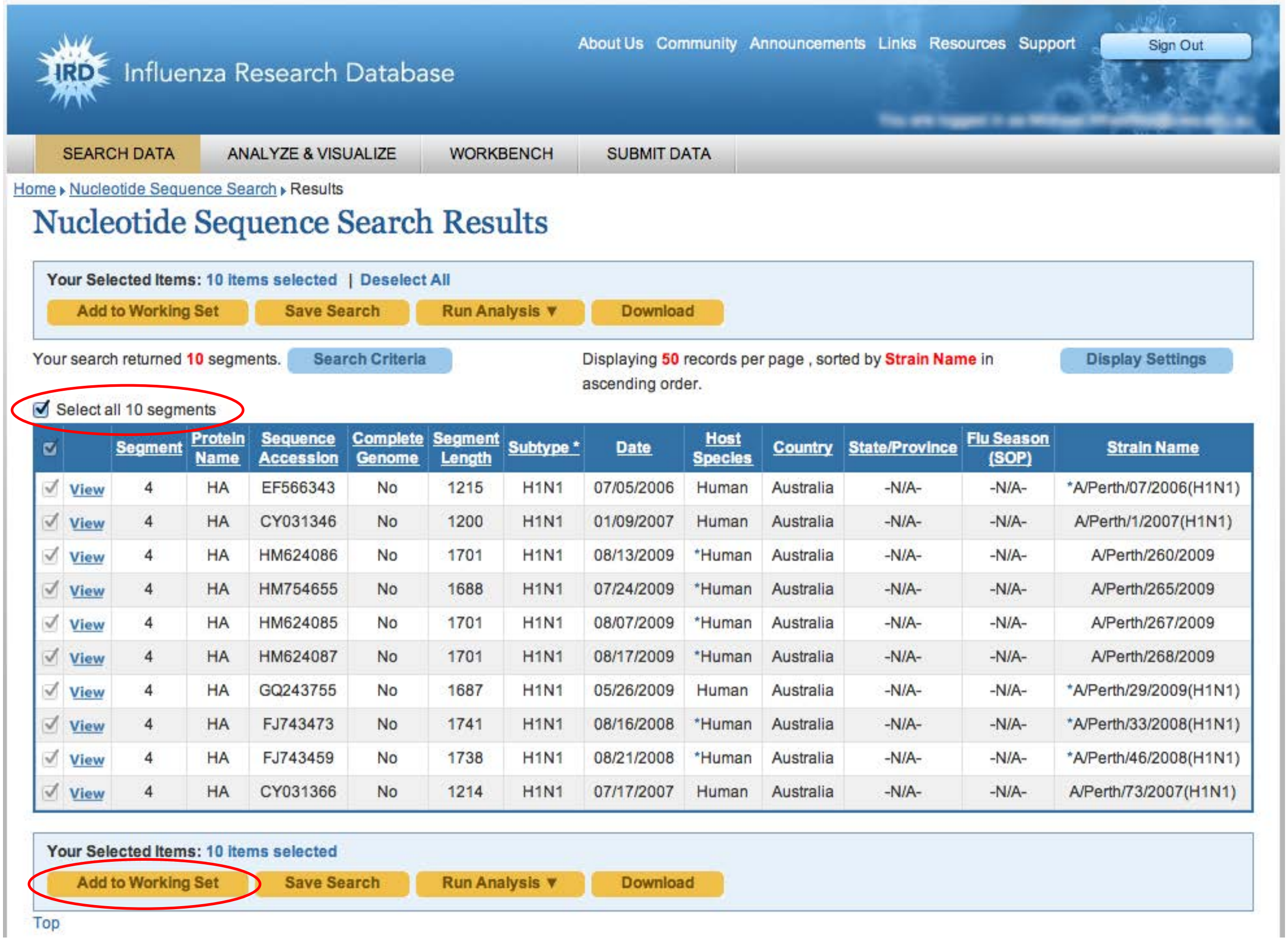
COUNTRY: Australia (selected), Cook Islands, Fiji, French Polynesia, Guam, New Caledonia

Tip: To select multiple or deselect, Ctrl-click (Windows) or Cmd-click (MacOS)

The diagram on the right shows the convention for naming human influenza strains.



Save this search to your workbench by selecting **Select all 10 segments**, then **Add to Working Set**. You will be prompted for a name for this set: you may want to call it ‘Perth H1N1’.



Nucleotide Sequence Search Results

Your Selected Items: 10 items selected | Deselect All

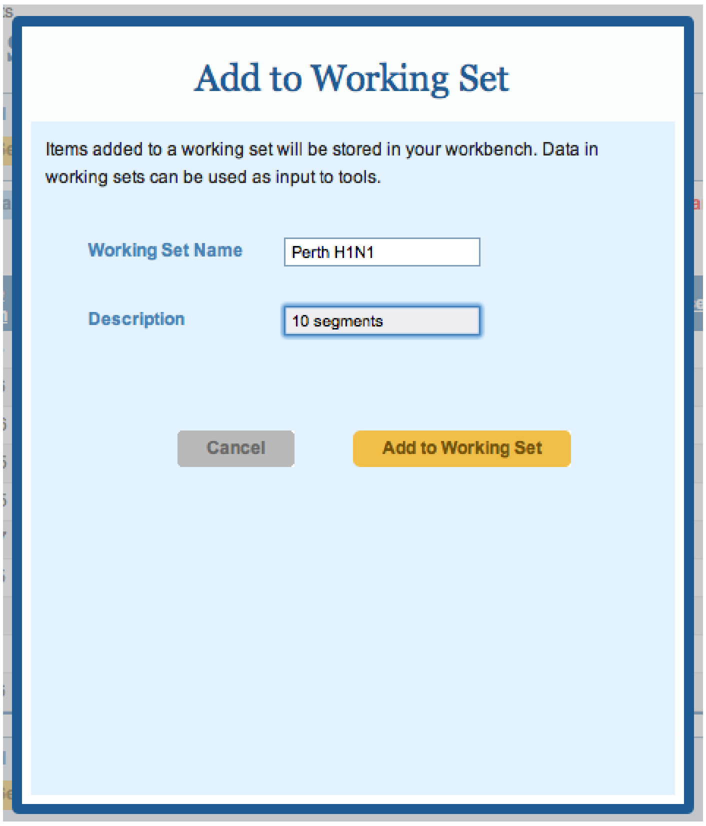
Your search returned 10 segments.

Displaying 50 records per page , sorted by Strain Name in ascending order.

☑ Select all 10 segments

☑		Segment	Protein Name	Sequence Accession	Complete Genome	Segment Length	Subtype *	Date	Host Species	Country	State/Province	Flu Season (SOP)	Strain Name
☑	View	4	HA	EF566343	No	1215	H1N1	07/05/2006	Human	Australia	-N/A-	-N/A-	*A/Perth/07/2006(H1N1)
☑	View	4	HA	CY031346	No	1200	H1N1	01/09/2007	Human	Australia	-N/A-	-N/A-	A/Perth/1/2007(H1N1)
☑	View	4	HA	HM624086	No	1701	H1N1	08/13/2009	*Human	Australia	-N/A-	-N/A-	A/Perth/260/2009
☑	View	4	HA	HM754655	No	1688	H1N1	07/24/2009	*Human	Australia	-N/A-	-N/A-	A/Perth/265/2009
☑	View	4	HA	HM624085	No	1701	H1N1	08/07/2009	*Human	Australia	-N/A-	-N/A-	A/Perth/267/2009
☑	View	4	HA	HM624087	No	1701	H1N1	08/17/2009	*Human	Australia	-N/A-	-N/A-	A/Perth/268/2009
☑	View	4	HA	GQ243755	No	1687	H1N1	05/26/2009	Human	Australia	-N/A-	-N/A-	*A/Perth/29/2009(H1N1)
☑	View	4	HA	FJ743473	No	1741	H1N1	08/16/2008	*Human	Australia	-N/A-	-N/A-	*A/Perth/33/2008(H1N1)
☑	View	4	HA	FJ743459	No	1738	H1N1	08/21/2008	*Human	Australia	-N/A-	-N/A-	*A/Perth/46/2008(H1N1)
☑	View	4	HA	CY031366	No	1214	H1N1	07/17/2007	Human	Australia	-N/A-	-N/A-	A/Perth/73/2007(H1N1)

Your Selected Items: 10 items selected



Add to Working Set

Items added to a working set will be stored in your workbench. Data in working sets can be used as input to tools.

Working Set Name: Perth H1N1

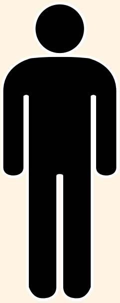
Description: 10 segments

After naming the working set, select **Add to Working Set**, then **Close**.

These strains were identified through blood and mucus samples from influenza patients. It is possible to find out more information about strains by looking at segment details.

Look specifically at segment details for the strain **A/Perth/265/2009** and complete question 6. Remember to select **View** and scroll through information provided.

6. Add information from the database for strain A/ Perth/265/2009 to the table below.



When was this virus collected? ....

What was the age and gender of the individual? ....

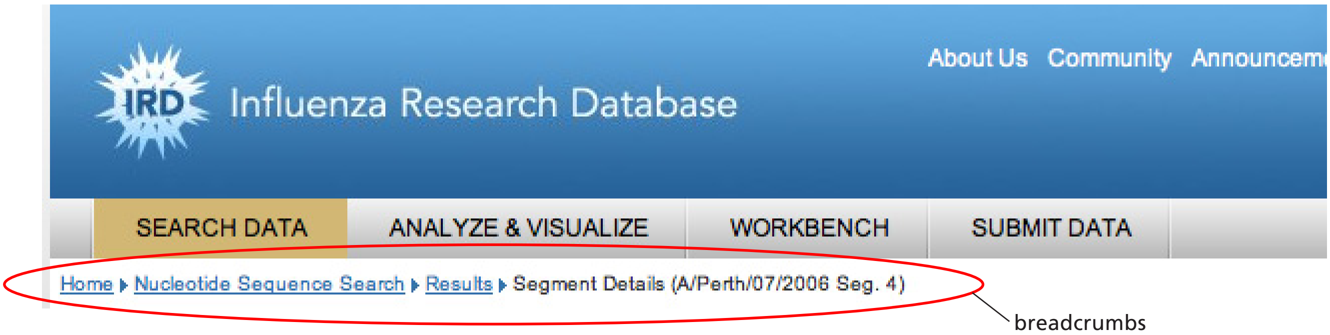
At what location was the specimen collected? ....

(see GenBank header notes)

How long is the protein that this sequence encodes? ....

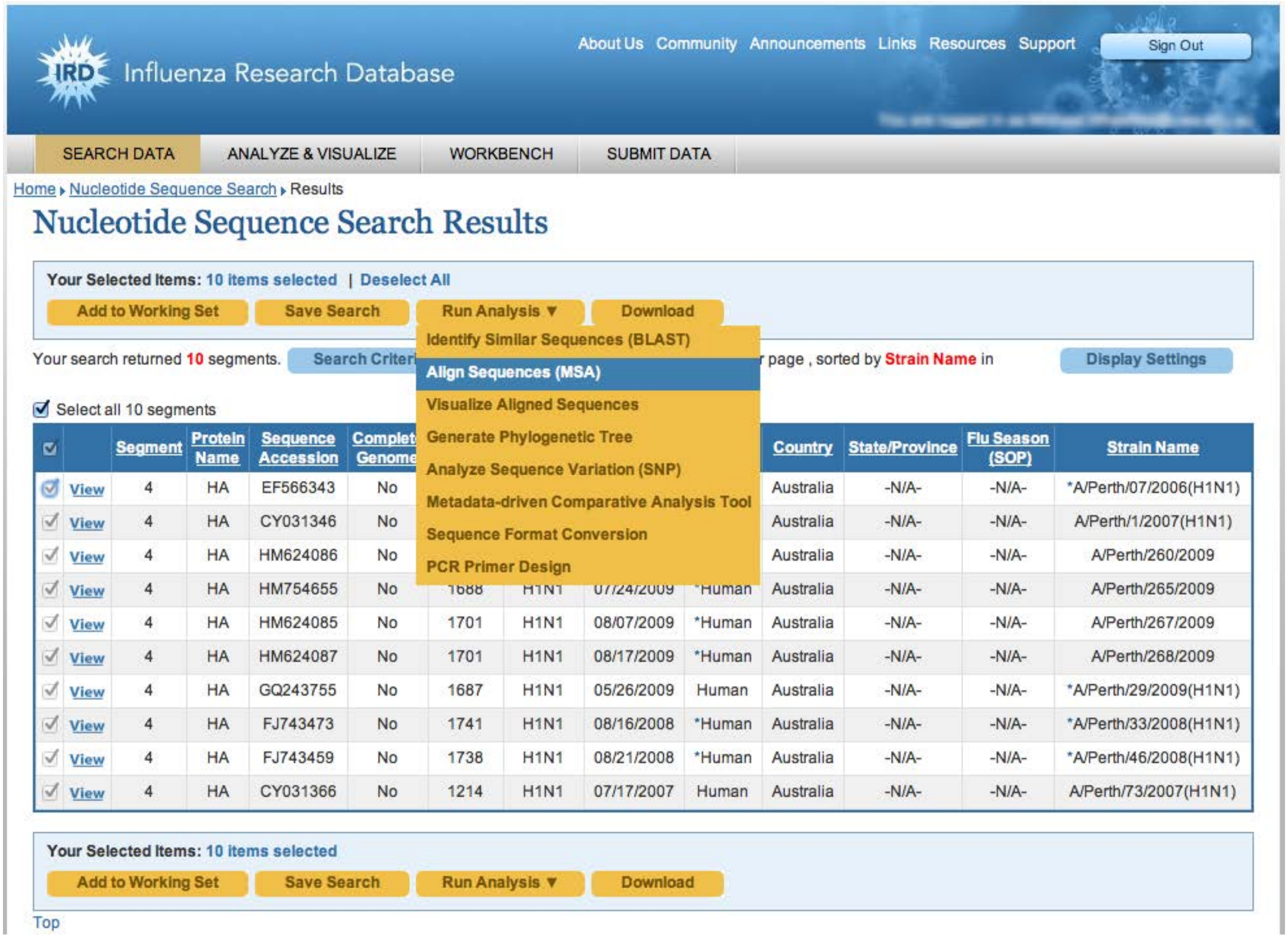
Ⓓ Analyse sequences

Go back to **Results** (use the ‘breadcrumbs’ toolbar, under dropdown menus).



breadcrumbs

Examine Perth strains more closely, by performing a multiple sequence alignment. To do this, select **Select all 10 segments**, then **Run Analysis**, then **Align Sequences (MSA)**. This will place sequences in rows. At certain positions mutations (nucleotide differences) will be visible.



Nucleotide Sequence Search Results

Your Selected Items: 10 items selected | Deselect All

Your search returned 10 segments. … page , sorted by Strain Name in

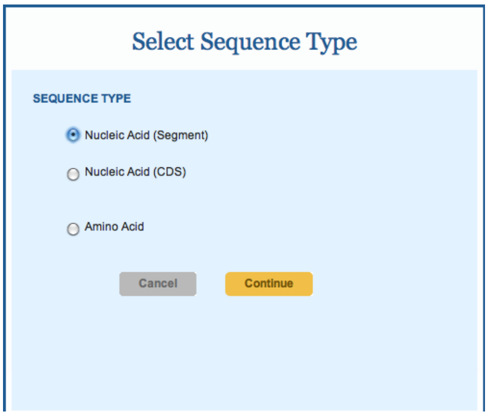
☑ Select all 10 segments

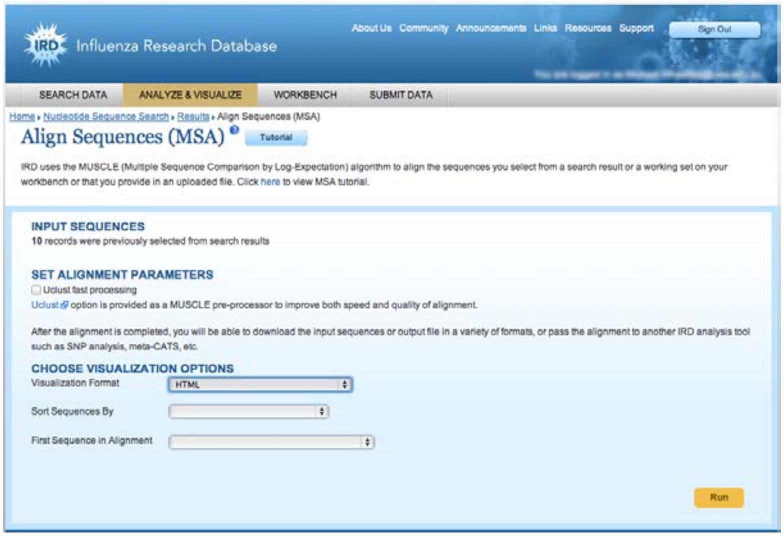
☑		Segment	Protein Name	Sequence Accession	Complet Genome					Country	State/Province	Flu Season (SOP)	Strain Name
☑	View	4	HA	EF566343	No					Australia	-N/A-	-N/A-	*A/Perth/07/2006(H1N1)
☑	View	4	HA	CY031346	No					Australia	-N/A-	-N/A-	A/Perth/1/2007(H1N1)
☑	View	4	HA	HM624086	No					Australia	-N/A-	-N/A-	A/Perth/260/2009
☑	View	4	HA	HM754655	No	1688	H1N1	07/24/2009	*Human	Australia	-N/A-	-N/A-	A/Perth/265/2009
☑	View	4	HA	HM624085	No	1701	H1N1	08/07/2009	*Human	Australia	-N/A-	-N/A-	A/Perth/267/2009
☑	View	4	HA	HM624087	No	1701	H1N1	08/17/2009	*Human	Australia	-N/A-	-N/A-	A/Perth/268/2009
☑	View	4	HA	GQ243755	No	1687	H1N1	05/26/2009	Human	Australia	-N/A-	-N/A-	*A/Perth/29/2009(H1N1)
☑	View	4	HA	FJ743473	No	1741	H1N1	08/16/2008	*Human	Australia	-N/A-	-N/A-	*A/Perth/33/2008(H1N1)
☑	View	4	HA	FJ743459	No	1738	H1N1	08/21/2008	*Human	Australia	-N/A-	-N/A-	*A/Perth/46/2008(H1N1)
☑	View	4	HA	CY031366	No	1214	H1N1	07/17/2007	Human	Australia	-N/A-	-N/A-	A/Perth/73/2007(H1N1)

Your Selected Items: 10 items selected

A window will pop up, select **Nucleic Acid (Segment)**, then **Continue**.

On the next screen, without changing any parameters, select **Run**.





IRD uses the MUSCLE (Multiple Sequence Comparison by Log-Expectation) algorithm to align the sequences you select from a search result or a working set on your workbench or that you provide in an uploaded file. Click here to view MSA tutorial.

INPUT SEQUENCES

10 records were previously selected from search results

SET ALIGNMENT PARAMETERS

☐ Uclust fast processing

Uclust option is provided as a MUSCLE pre-processor to improve both speed and quality of alignment.

After the alignment is completed, you will be able to download the input sequences or output file in a variety of formats, or pass the alignment to another IRD analysis tool such as SNP analysis, meta-CATS, etc.

CHOOSE VISUALIZATION OPTIONS

Visualization Format: HTML

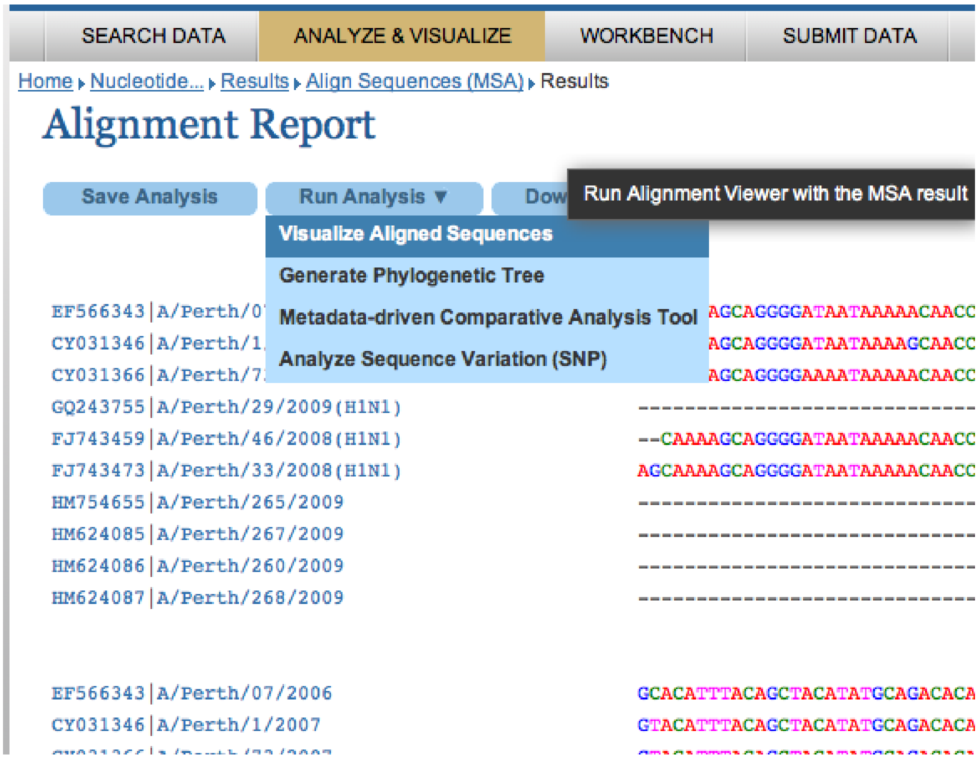
Sort Sequences By:

First Sequence in Alignment:

It takes some time for the analysis to run; a message is displayed while it is processing.

Ⓔ Compare sequences

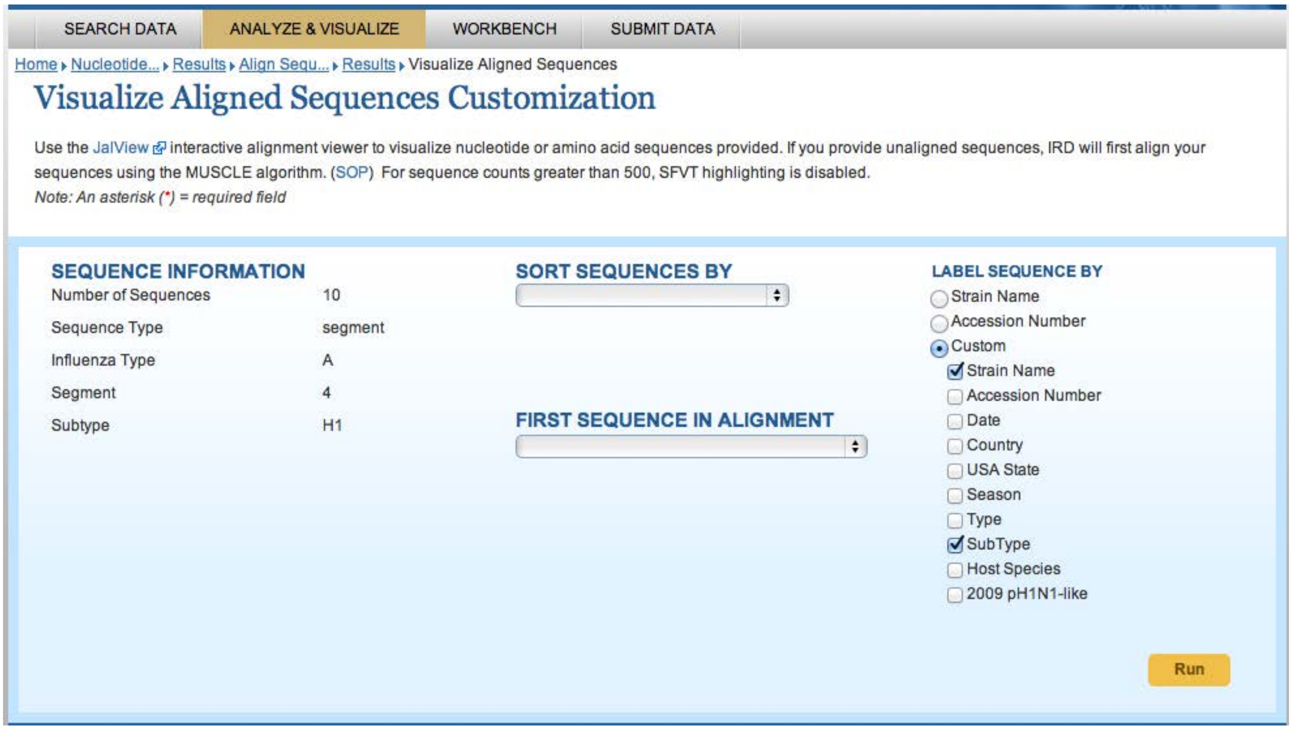
Once the analysis finishes, roll over the blue **Run Analysis** button and select **Visualise Aligned Sequences** from the dropdown menu.



```
EF566343|A/Perth/0                                  AGCAGGGGATAATAAAAACAACC
CY031346|A/Perth/1,                                 AGCAGGGGATAATAAAAGCAACC
CY031366|A/Perth/7:                                 AGCAGGGGAAAATAAAAACAACC
GQ243755|A/Perth/29/2009(H1N1)          ----------------------------
FJ743459|A/Perth/46/2008(H1N1)          --CAAAAGCAGGGGATAATAAAAACAACC
FJ743473|A/Perth/33/2008(H1N1)          AGCAAAAGCAGGGGATAATAAAAACAACC
HM754655|A/Perth/265/2009               ----------------------------
HM624085|A/Perth/267/2009               ----------------------------
HM624086|A/Perth/260/2009               ----------------------------
HM624087|A/Perth/268/2009               ----------------------------


EF566343|A/Perth/07/2006                GCACATTTACAGCTACATATGCAGACACA
CY031346|A/Perth/1/2007                 GTACATTTACAGCTACATATGCAGACACA
```

Customise **LABEL SEQUENCE BY** in the alignment: select **Custom**, and underneath it, a submenu will appear, select **Strain Name** and **SubType**, then select **Run**.



Visualize Aligned Sequences Customization

Use the JalView interactive alignment viewer to visualize nucleotide or amino acid sequences provided. If you provide unaligned sequences, IRD will first align your sequences using the MUSCLE algorithm. (SOP) For sequence counts greater than 500, SFVT highlighting is disabled.

Note: An asterisk (*) = required field

SEQUENCE INFORMATION

Number of Sequences	10
Sequence Type	segment
Influenza Type	A
Segment	4
Subtype	H1

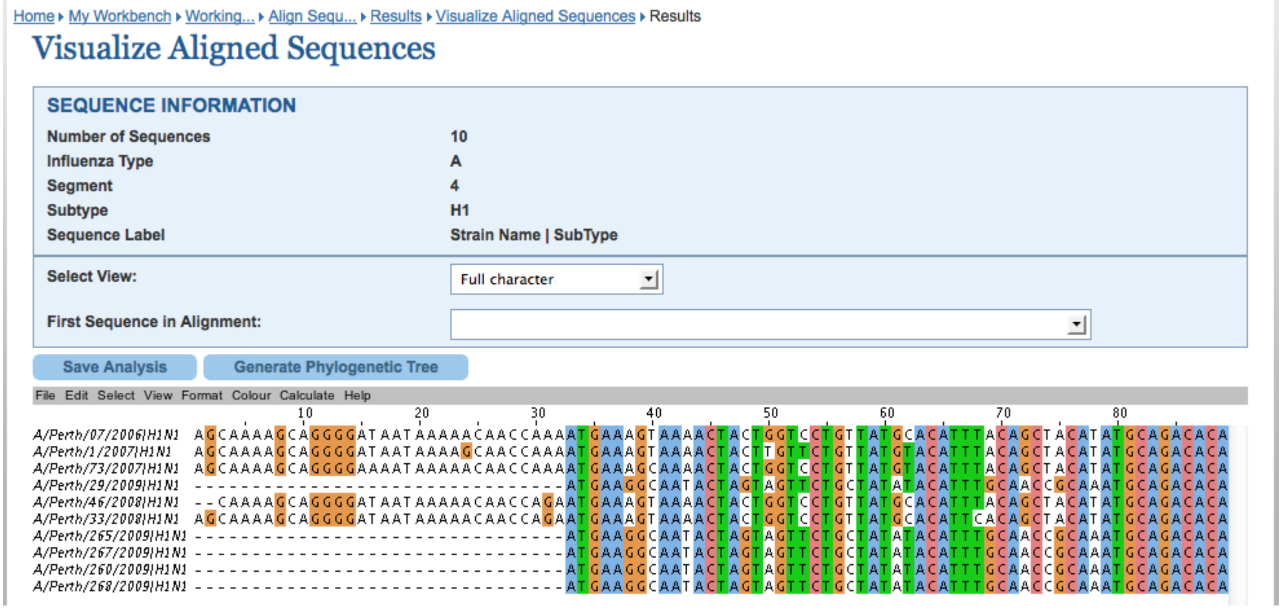
SORT SEQUENCES BY

FIRST SEQUENCE IN ALIGNMENT

LABEL SEQUENCE BY
○ Strain Name
○ Accession Number
● Custom
  ☑ Strain Name
  ☐ Accession Number
  ☐ Date
  ☐ Country
  ☐ USA State
  ☐ Season
  ☐ Type
  ☑ SubType
  ☐ Host Species
  ☐ 2009 pH1N1-like

You can now see the visually aligned sequences.

Note: You may need to 'allow' a Java applet to run in order to display aligned sequences. A blank space on the page or 'missing plug-in' message may indicate problems with the installation of Java or its security settings on your computer.



Visualize Aligned Sequences

SEQUENCE INFORMATION

Number of Sequences	**10**
Influenza Type	**A**
Segment	**4**
Subtype	**H1**
Sequence Label	**Strain Name | SubType**

Select View: Full character

First Sequence in Alignment:

```
                             10        20        30        40        50        60        70        80
A/Perth/07/2006|H1N1   AGCAAAAGCAGGGGATAATAAAAACAACCAAAATGAAAGTAAAACTACTGGTCCTGTTATGCACATTTACAGCTACATATGCAGACACA
A/Perth/1/2007|H1N1    AGCAAAAGCAGGGGATAATAAAAGCAACCAAAATGAAAGTAAAACTACTTGTTCTGTTATGTACATTTACAGCTACATATGCAGACACA
A/Perth/73/2007|H1N1   AGCAAAAGCAGGGGAAAATAAAAACAACCAAAATGAAAGCAAAACTACTGGTCCTGTTATGTACATTTACAGCTACATATGCAGACACA
A/Perth/29/2009|H1N1   --------------------------------ATGAAGGCAATACTAGTAGTTCTGCTATATACATTTGCAACCGCAAATGCAGACACA
A/Perth/46/2008|H1N1   --CAAAAGCAGGGGATAATAAAAACAACCAGAATGAAAGTAAAACTACTGGTCCTGTTATGCACATTTACAGCTACATATGCAGACACA
A/Perth/33/2008|H1N1   AGCAAAAGCAGGGGATAATAAAAACAACCAGAATGAAAGTAAAACTACTGGTCCTGTTATGCACATTCACAGCTACATATGCAGACACA
A/Perth/265/2009|H1N1  --------------------------------ATGAAGGCAATACTAGTAGTTCTGCTATATACATTTGCAACCGCAAATGCAGACACA
A/Perth/267/2009|H1N1  --------------------------------ATGAAGGCAATACTAGTAGTTCTGCTATATACATTTGCAACCGCAAATGCAGACACA
A/Perth/260/2009|H1N1  --------------------------------ATGAAGGCAATACTAGTAGTTCTGCTATATACATTTGCAACCGCAAATGCAGACACA
A/Perth/268/2009|H1N1  --------------------------------ATGAAGGCAATACTAGTAGTTCTGCTATATACATTTGCAACCGCAAATGCAGACACA
```

Scroll left and right to look at the full alignment. Solid columns of colour (including white) indicate each sequence that has the same nucleotide, in a particular position. Colour changes indicate where mutations have occurred.

7. What do you think dashes in the sequence mean?

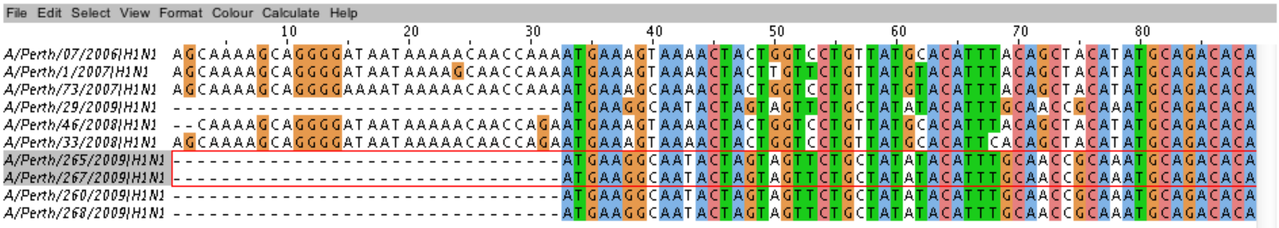
………

………

From these sequences it's possible to determine the degree of alignment. 100% means there are no mutations between strains (perfect alignment).

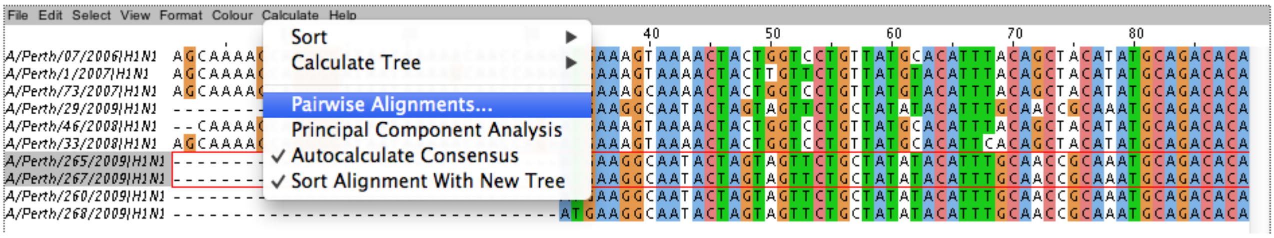
Select two 2009 sequences from Perth by clicking on them (eg 265/2009 and 267/2009).

Note: to select sequences that are not adjacent, on a Mac, hold down the **control** key. For those that are adjacent, hold down the **shift** key, as you select them.



```
                               10        20        30        40        50        60        70        80
A/Perth/07/2006|H1N1  AGCAAAAGCAGGGGATAATAAAAACAACCAAAATGAAAGTAAAACTACTGGTCCTGTTATGCACATTTACAGCTACATATGCAGACACA
A/Perth/1/2007|H1N1   AGCAAAAGCAGGGGATAATAAAAGCAACCAAAATGAAAGTAAAACTACTTGTTCTGTTATGTACATTTACAGCTACATATGCAGACACA
A/Perth/73/2007|H1N1  AGCAAAAGCAGGGGAAAATAAAAACAACCAAAATGAAAGCAAAACTACTGGTCCTGTTATGTACATTTACAGCTACATATGCAGACACA
A/Perth/29/2009|H1N1  --------------------------------ATGAAGGCAATACTAGTAGTTCTGCTATATACATTTGCAACCGCAAATGCAGACACA
A/Perth/46/2008|H1N1  --CAAAAGCAGGGGATAATAAAAACAACCAGAATGAAAGTAAAACTACTGGTCCTGTTATGCACATTTACAGCTACATATGCAGACACA
A/Perth/33/2008|H1N1  AGCAAAAGCAGGGGATAATAAAAACAACCAGAATGAAAGTAAAACTACTGGTCCTGTTATGCACATTCACAGCTACATATGCAGACACA
A/Perth/265/2009|H1N1 --------------------------------ATGAAGGCAATACTAGTAGTTCTGCTATATACATTTGCAACCGCAAATGCAGACACA
A/Perth/267/2009|H1N1 --------------------------------ATGAAGGCAATACTAGTAGTTCTGCTATATACATTTGCAACCGCAAATGCAGACACA
A/Perth/260/2009|H1N1 --------------------------------ATGAAGGCAATACTAGTAGTTCTGCTATATACATTTGCAACCGCAAATGCAGACACA
A/Perth/268/2009|H1N1 --------------------------------ATGAAGGCAATACTAGTAGTTCTGCTATATACATTTGCAACCGCAAATGCAGACACA
```

From the grey submenu bar select **Calculate**, then **Pairwise Alignments…** (from the dropdown menu).



A separate page will pop up, scroll down it to find **Percentage ID**.

These sequences are 99.82% similar (you may get a different value, depending on the sequences you choose). Close the window.

Now select sequences from two different years, eg 268/2009 and 1/2007. These sequences are 75.56% similar.

Select sequences from 2006 and 2008 (eg 07/2006 and 46/2008). They are 97.04% similar.

8. Complete this table based on your calculations

Table 1: Percentage similarity between strains

Strain/year					
	29/2009	265/2009	268/2009	1/2007	07/2006
29/2009	-	-	-	-	-
265/2009	99.59	-	-	-	-
268/2009	99.76	99.82	-	-	-
1/2007	75.58	75.64	75.56	-	-
07/2006	75.67			96.33	-
46/2008	76.72				

9. Using percentage similarity data (table 1), explain how closely related these strains are:

a) 2009 strains; and

....

....

b) 2006, 2007 and 2008 strains.

....

....

10. How do 2009 viruses compare to other viruses in Table 1?

....

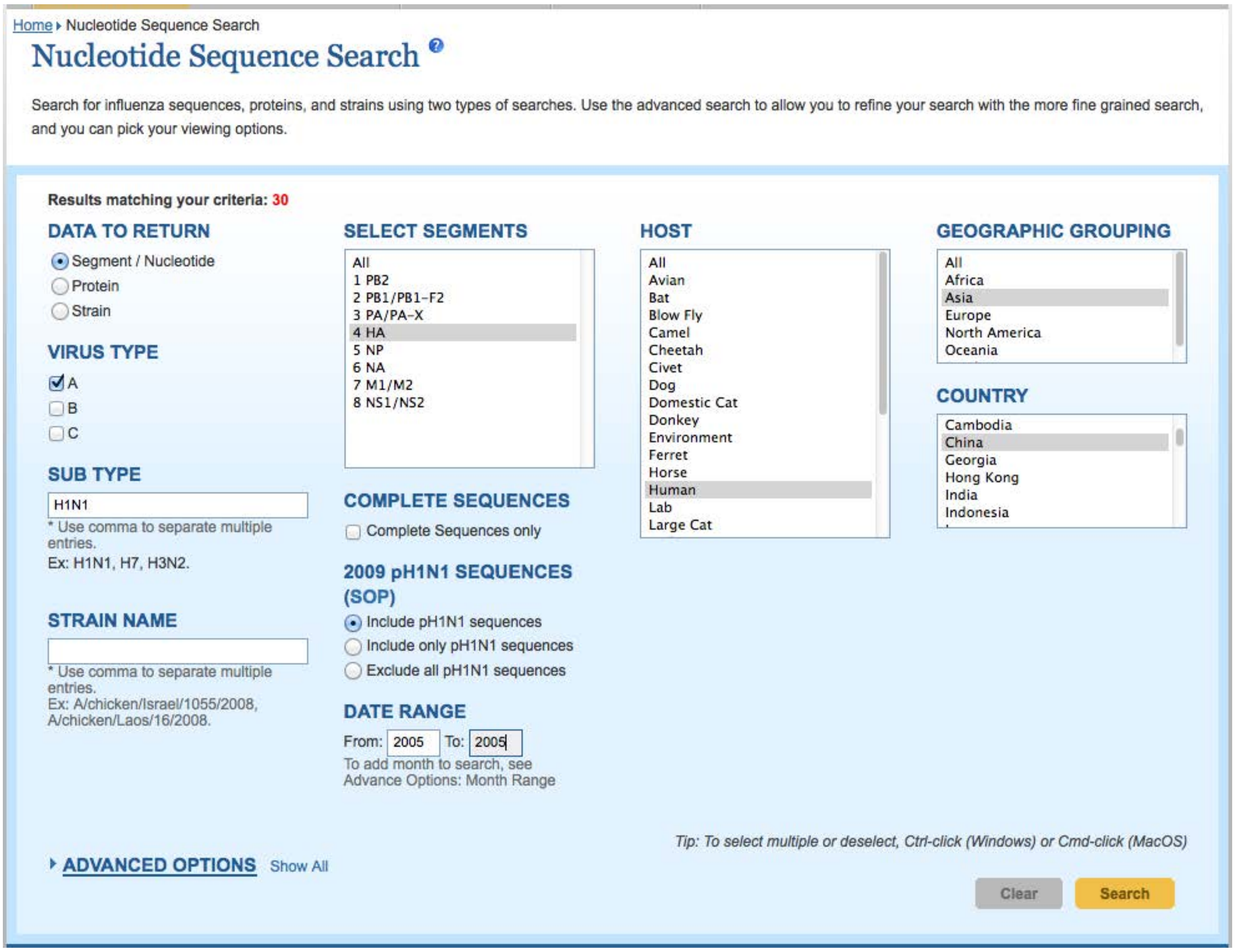
....

....

Ⓕ Phylogenetic trees

The database can be used to create a phylogenetic tree that shows how H1N1 strains from Perth are related to each other, and how they diverge over time. To create a meaningful phylogenetic tree you must choose an outgroup. This is a strain that shares a common ancestor with Perth H1N1 strains, but is genetically distinct from them.

Return to **Nucleotide Sequence Search** and search for a 2005 H1N1 strain from China as your outgroup (set **Date Range** as 2005 – 2005).



Home ▸ Nucleotide Sequence Search

Nucleotide Sequence Search

Search for influenza sequences, proteins, and strains using two types of searches. Use the advanced search to allow you to refine your search with the more fine grained search, and you can pick your viewing options.

Results matching your criteria: 30

DATA TO RETURN
● Segment / Nucleotide
○ Protein
○ Strain

VIRUS TYPE
☑ A
☐ B
☐ C

SUB TYPE
H1N1
* Use comma to separate multiple entries.
Ex: H1N1, H7, H3N2.

STRAIN NAME
* Use comma to separate multiple entries.
Ex: A/chicken/Israel/1055/2008, A/chicken/Laos/16/2008.

SELECT SEGMENTS: 4 HA

COMPLETE SEQUENCES
☐ Complete Sequences only

2009 pH1N1 SEQUENCES (SOP)
● Include pH1N1 sequences
○ Include only pH1N1 sequences
○ Exclude all pH1N1 sequences

DATE RANGE
From: 2005 To: 2005
To add month to search, see Advance Options: Month Range

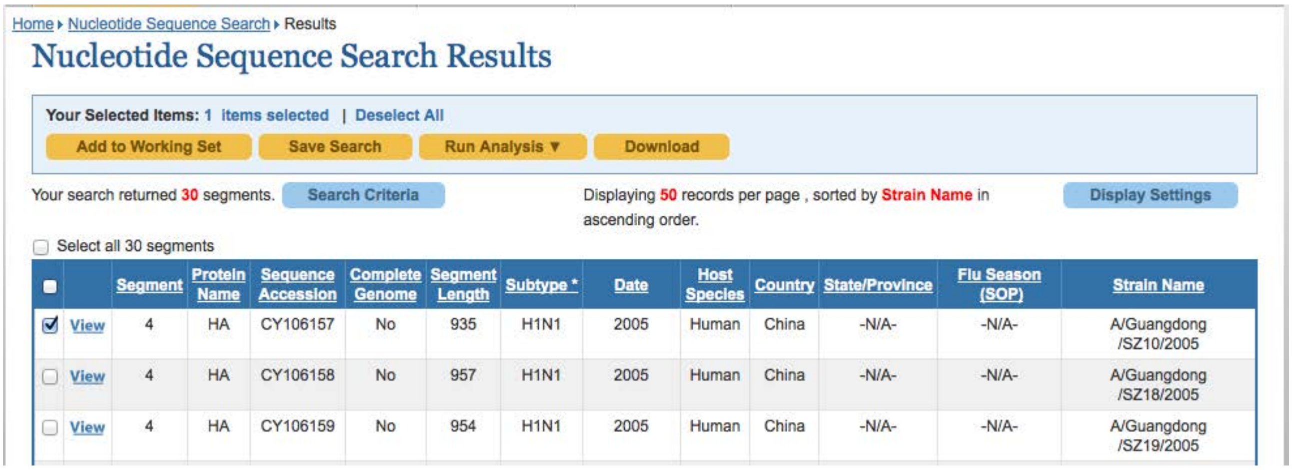
HOST: Human

GEOGRAPHIC GROUPING: Asia

COUNTRY: China

Tip: To select multiple or deselect, Ctrl-click (Windows) or Cmd-click (MacOS)

Your search will find a number of sequences. Select a sequence, then select **Add to Working Set**.



Home ▸ Nucleotide Sequence Search ▸ Results

Nucleotide Sequence Search Results

Your Selected Items: 1 items selected | Deselect All

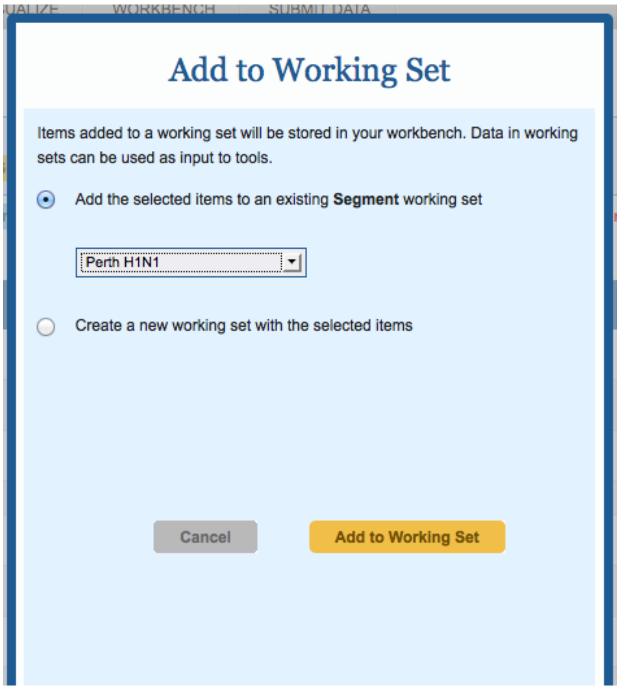
Your search returned 30 segments.

Displaying 50 records per page , sorted by Strain Name in ascending order.

☐ Select all 30 segments

☐		Segment	Protein Name	Sequence Accession	Complete Genome	Segment Length	Subtype *	Date	Host Species	Country	State/Province	Flu Season (SOP)	Strain Name
☑	View	4	HA	CY106157	No	935	H1N1	2005	Human	China	-N/A-	-N/A-	A/Guangdong /SZ10/2005
☐	View	4	HA	CY106158	No	957	H1N1	2005	Human	China	-N/A-	-N/A-	A/Guangdong /SZ18/2005
☐	View	4	HA	CY106159	No	954	H1N1	2005	Human	China	-N/A-	-N/A-	A/Guangdong /SZ19/2005

Add the selected strain to your working set (Perth H1N1, or whatever you chose to name it), select **Add to Working Set**, then **Close**.



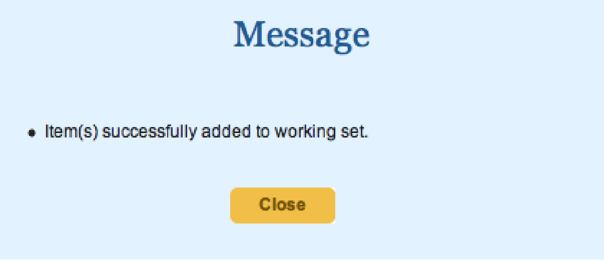
Add to Working Set

Items added to a working set will be stored in your workbench. Data in working sets can be used as input to tools.

● Add the selected items to an existing **Segment** working set

Perth H1N1

○ Create a new working set with the selected items

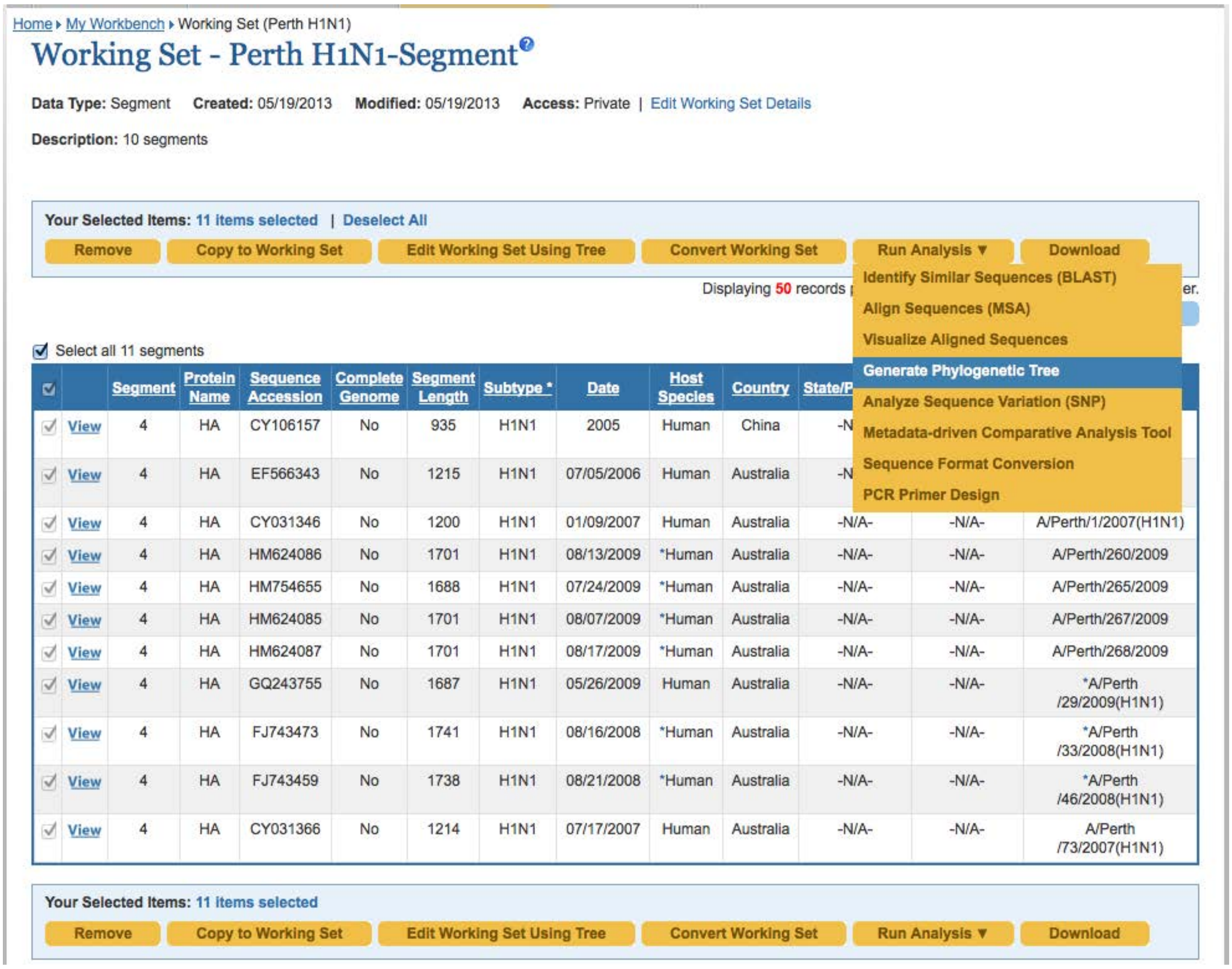


Message

- Item(s) successfully added to working set.

Under **WORKBENCH** select **View Your Workbench**, select the Perth H1N1 working set, then select **View**. Once this working set is displayed, select all strains (there should be 11).

Select **Run Analysis** then **Generate Phylogenetic Tree**. This will take some time.



Working Set - Perth H1N1-Segment

Data Type: Segment **Created:** 05/19/2013 **Modified:** 05/19/2013 **Access:** Private | Edit Working Set Details

Description: 10 segments

Your Selected Items: 11 items selected | Deselect All

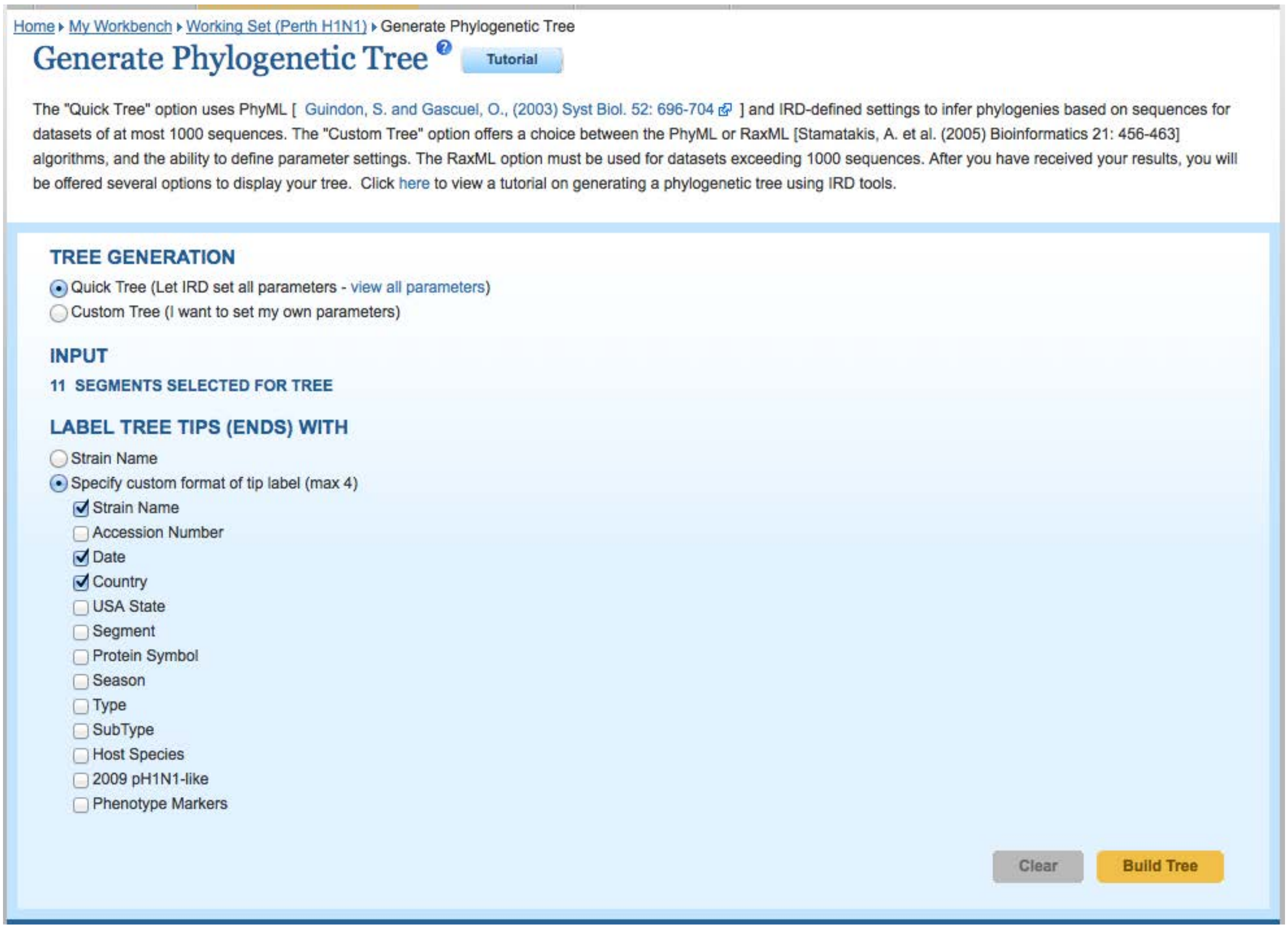
Displaying 50 records

☑ Select all 11 segments

☑		Segment	Protein Name	Sequence Accession	Complete Genome	Segment Length	Subtype *	Date	Host Species	Country	State/P		
☑	View	4	HA	CY106157	No	935	H1N1	2005	Human	China	-N		
☑	View	4	HA	EF566343	No	1215	H1N1	07/05/2006	Human	Australia	-N		
☑	View	4	HA	CY031346	No	1200	H1N1	01/09/2007	Human	Australia	-N/A-	-N/A-	A/Perth/1/2007(H1N1)
☑	View	4	HA	HM624086	No	1701	H1N1	08/13/2009	*Human	Australia	-N/A-	-N/A-	A/Perth/260/2009
☑	View	4	HA	HM754655	No	1688	H1N1	07/24/2009	*Human	Australia	-N/A-	-N/A-	A/Perth/265/2009
☑	View	4	HA	HM624085	No	1701	H1N1	08/07/2009	*Human	Australia	-N/A-	-N/A-	A/Perth/267/2009
☑	View	4	HA	HM624087	No	1701	H1N1	08/17/2009	*Human	Australia	-N/A-	-N/A-	A/Perth/268/2009
☑	View	4	HA	GQ243755	No	1687	H1N1	05/26/2009	Human	Australia	-N/A-	-N/A-	*A/Perth /29/2009(H1N1)
☑	View	4	HA	FJ743473	No	1741	H1N1	08/16/2008	*Human	Australia	-N/A-	-N/A-	*A/Perth /33/2008(H1N1)
☑	View	4	HA	FJ743459	No	1738	H1N1	08/21/2008	*Human	Australia	-N/A-	-N/A-	*A/Perth /46/2008(H1N1)
☑	View	4	HA	CY031366	No	1214	H1N1	07/17/2007	Human	Australia	-N/A-	-N/A-	A/Perth /73/2007(H1N1)

Your Selected Items: 11 items selected

On the **Generate Phylogenetic Tree** page, select **Quick Tree**, then **Specify custom format of tip label**. Select **Strain Name**, **Date** and **Country**, then **Build Tree**.



Home ▸ My Workbench ▸ Working Set (Perth H1N1) ▸ Generate Phylogenetic Tree

Generate Phylogenetic Tree ? Tutorial

The "Quick Tree" option uses PhyML [Guindon, S. and Gascuel, O., (2003) Syst Biol. 52: 696-704] and IRD-defined settings to infer phylogenies based on sequences for datasets of at most 1000 sequences. The "Custom Tree" option offers a choice between the PhyML or RaxML [Stamatakis, A. et al. (2005) Bioinformatics 21: 456-463] algorithms, and the ability to define parameter settings. The RaxML option must be used for datasets exceeding 1000 sequences. After you have received your results, you will be offered several options to display your tree. Click here to view a tutorial on generating a phylogenetic tree using IRD tools.

TREE GENERATION

● Quick Tree (Let IRD set all parameters - view all parameters)
○ Custom Tree (I want to set my own parameters)

INPUT

11 SEGMENTS SELECTED FOR TREE

LABEL TREE TIPS (ENDS) WITH

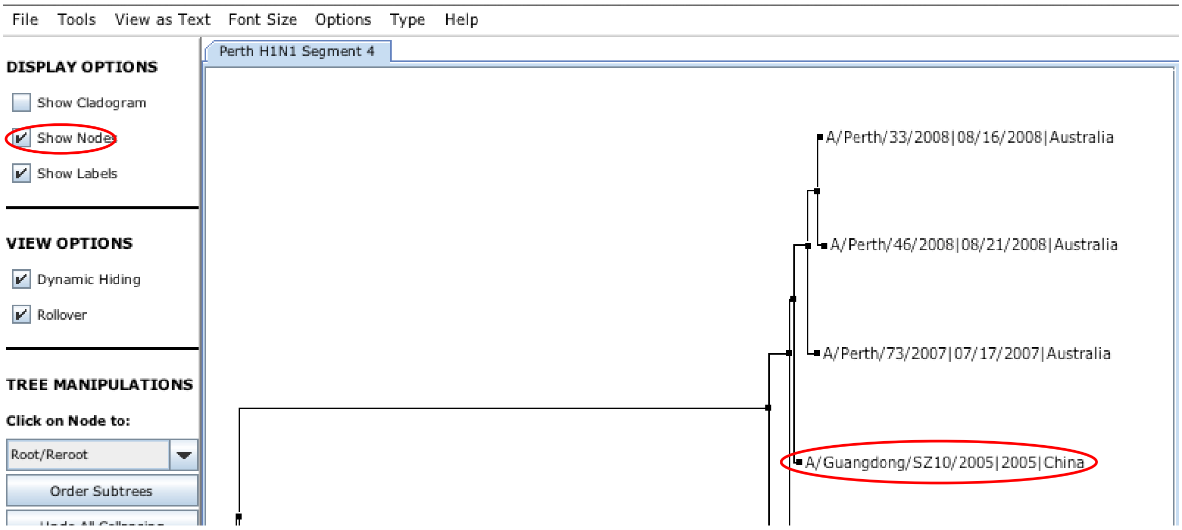
○ Strain Name
● Specify custom format of tip label (max 4)
  ☑ Strain Name
  ☐ Accession Number
  ☑ Date
  ☑ Country
  ☐ USA State
  ☐ Segment
  ☐ Protein Symbol
  ☐ Season
  ☐ Type
  ☐ SubType
  ☐ Host Species
  ☐ 2009 pH1N1-like
  ☐ Phenotype Markers

Clear Build Tree

View tree

Once the analysis is complete, select **View Tree** (note: this will launch a Java applet, *Archaeopterxy Phylogenetic Tree Viewer* — you should 'allow' this applet to run).

The China outgroup should be the 'root' of the tree, that is, a strain that branches off the main trunk almost immediately. If your outgroup is not the root of the tree then, under **DISPLAY OPTIONS**, select **Show Nodes**. Select a node (the small black square) near the outgroup segment (China 2005) to re-root the tree. (The cursor turns to a cross when you run your mouse over a node.)



DISPLAY OPTIONS
☐ Show Cladogram
☑ Show Nodes
☑ Show Labels

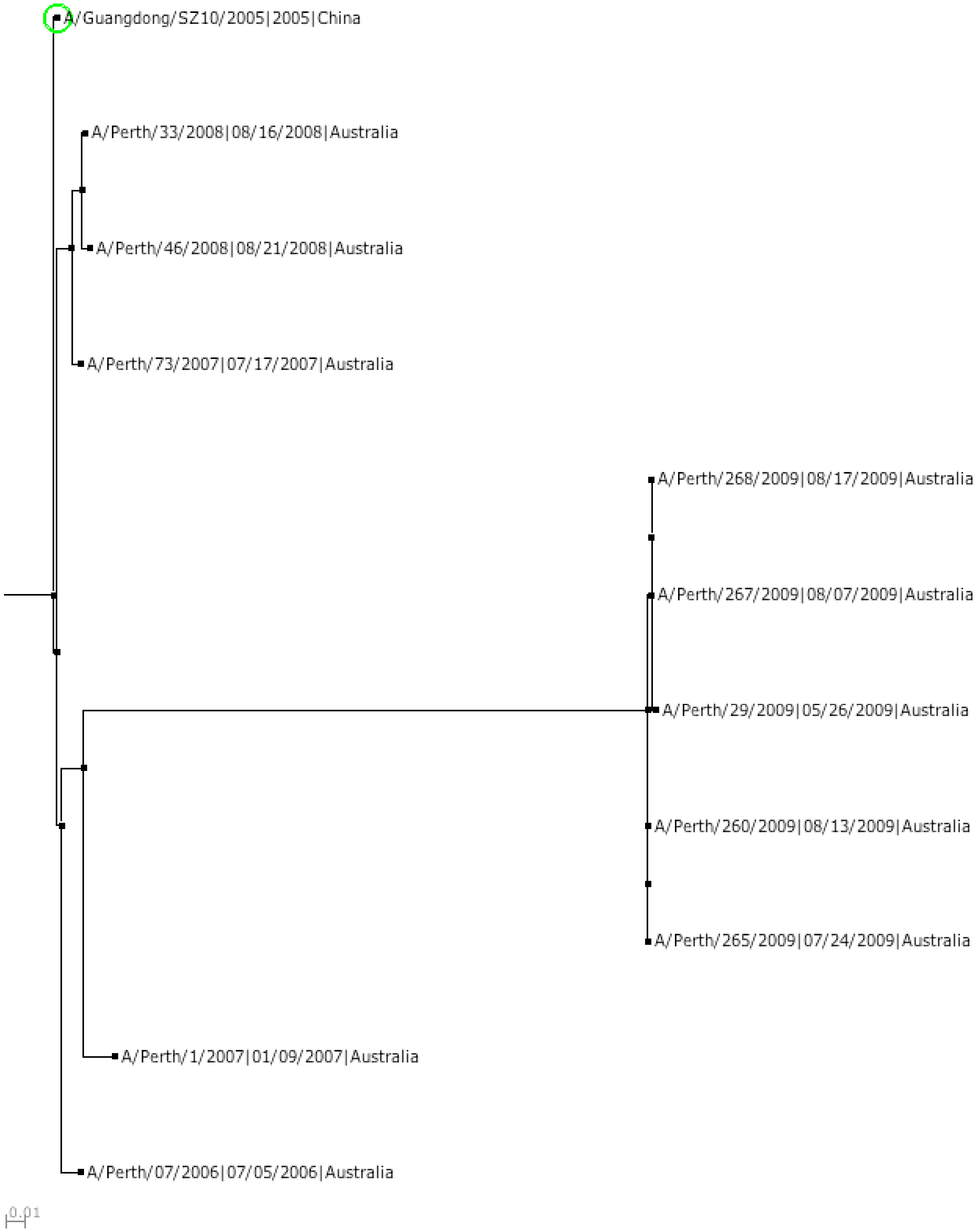
VIEW OPTIONS
☑ Dynamic Hiding
☑ Rollover

TREE MANIPULATIONS
Click on Node to:
Root/Reroot
Order Subtrees

Under **TREE DECORATIONS**, select **Year** from the drop-down menu **Basic Decoration Options**.

From the **Options** menu bar (at top of page) you can also display **Branch Length Values**.

The tree below uses an outgroup H1N1 from China, 2005. You can clearly see two distinct groups (called clades). The 2006, 2007 and 2008 segments are slightly different, however there's a major change for 2009. The 2009 strains clearly show the swine 'flu epidemic.



11. A/Perth 2006-2007 strains have a small number of mutations. What evolutionary process accounts for these differences?

....

....

....

....

12. In these sequences it appears that a number of mutations appeared in 2009. Can you describe what occurred in terms of antigenic shift?

....

....

13. How closely does the outgroup (China, 2005) relate to Perth strains of H1N1?
Consider: A/Perth/07/2006; A/Perth/46/2008 and A/Perth/265/2009.

....

....

....

14. In addition to yearly influenza vaccinations, a vaccine was offered in 2009 for swine influenza. Why do you think this was necessary?

....

....

....

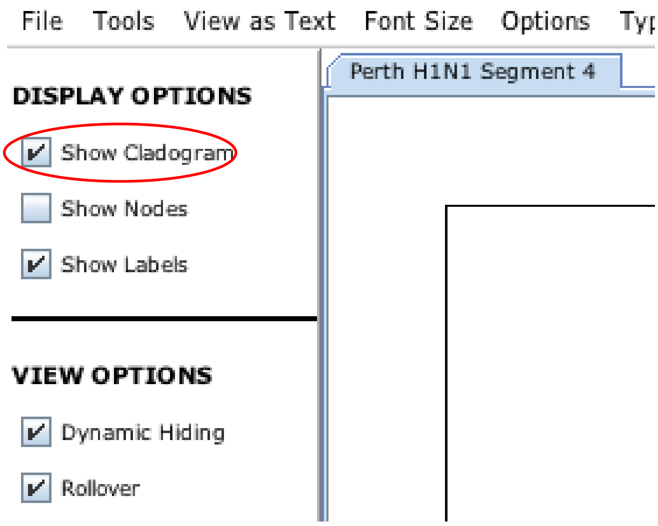
15. Look at different lengths of horizontal lines on the tree. What information can be gained from the length of horizontal branches?

....

....

....

Under **DISPLAY OPTIONS**, select **Show Cladogram**.



☑ Show Cladogram
☐ Show Nodes
☑ Show Labels
☑ Dynamic Hiding
☑ Rollover

16. The cladogram looks quite different to the phylogenetic tree. What sort of information does a cladogram show?

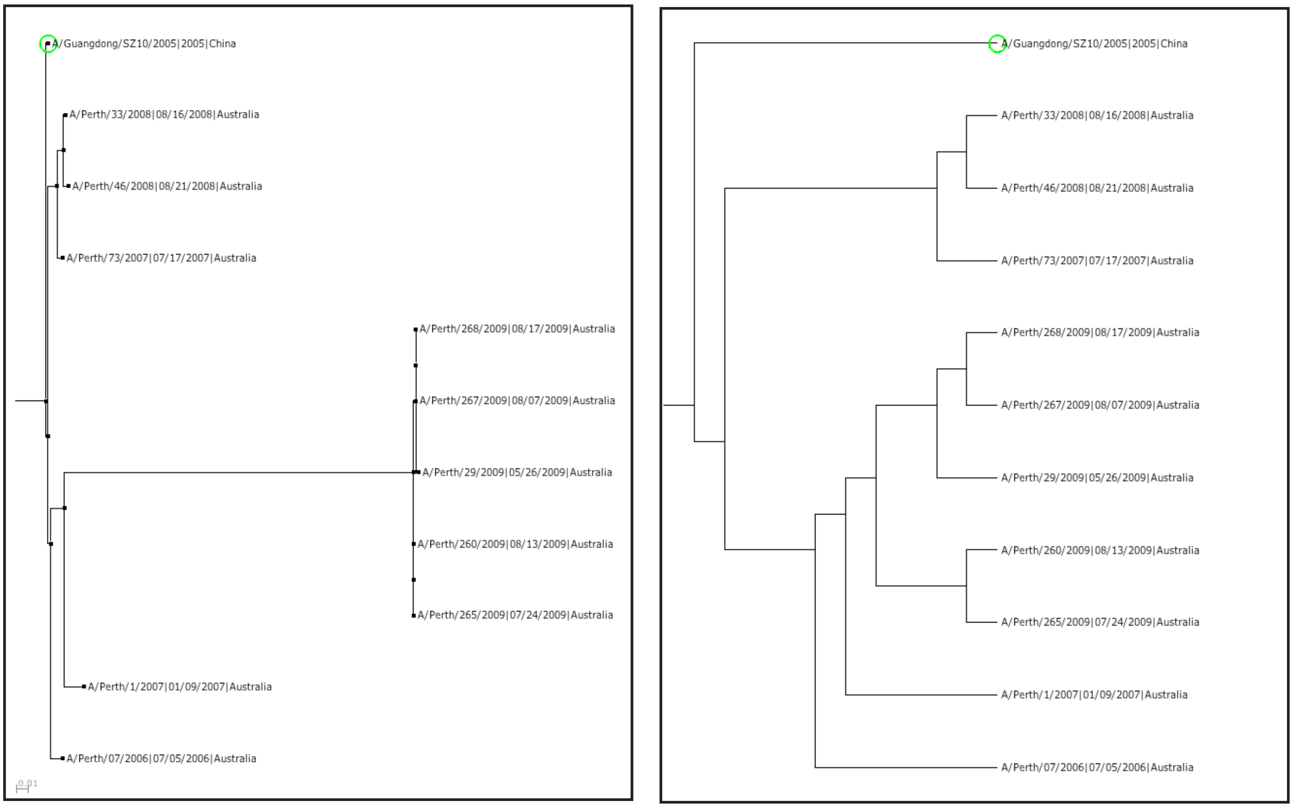
....

....

17. What information does a phylogenetic tree show that is not shown by a cladogram?

....

....



phylogenetic tree (left) and cladogram (right) for the same dataset

Other trees to create using the IRD

- Compare the same subtype H1N1, from different regions, across one year.
- Compare different human influenza viruses, eg H1N1 and H3N2 subtypes.
- Create a tree using segment 6, or NA. This segment of RNA codes for neuraminidase protein, another important protein on the surface of influenza virus. Does it suggest a similar pattern of evolutionary relationships?
- H5N1 is an avian (bird) flu virus that has caused outbreaks in domestic poultry in parts of Asia and the Middle East. This virus can also jump from birds to other species, including humans. Use the Influenza Research Database to investigate occurrence of H5N1 in China in 2007. Is there any evidence to indicate how humans might have contracted the virus? (Note: include host species in your tree label. To root your tree, add a single occurrence of human H3N2 flu to your dataset. When you process this dataset you may get a warning that you are working with two different subtypes (H5N1 and H3N2) but this will not be a problem here.)

Reference: US Dept of Health & Human Services. (n.d.). *H5N1 Avian Flu (H5N1 Bird Flu)*. Retrieved 21 May 2013 from http://www.flu.gov/about_the_flu/h5n1/