

PROTOCOL FOR INFLUENZA A VIRUS GLOBAL SWINE H1 CLADE CLASSIFICATION

January 23, 2017

1. Background

Swine H1 viruses have diversified into three major genetic lineages over time. Recently, Anderson et al. (2016) proposed a global classification scheme for swine H1 influenza A virus based on the phylogenetic relationships of the H1 HA segment sequences. This new global swine H1 clade classification scheme extends the previous US swine H1 classification scheme implemented in the Influenza Research Database (IRD, 2015) by incorporating additional swine sequences from outside North America, in particular Europe and Asia. To make the new classification scheme more accessible to the influenza research community, Anderson and other swine influenza experts at the United States Department of Agriculture (USDA) have collaborated with IRD and implemented the global swine H1 clade classification tool in IRD. This document describes both the global swine H1 clade classification tool and the corresponding clade classifications in IRD.

2. Automatic Assignment of Global Swine H1 Clades

The primary goal of the global swine H1 classification is to assign clades to unclassified HA/H1 sequences. An automated pipeline was developed for clade classification, whereby:

2.1 From the global swine H1 classification phylogeny developed by Anderson et al. (2016), a few representatives were selected for each defined clade. This selection led to a global swine H1 reference tree containing 239 H1 viruses of predominantly swine origin, with a few H1 viruses from human and avian hosts to represent the diversity of non-swine H1 viruses.

2.2 Input data are all HA/H1 segment sequences in IRD regardless of host and NA subtype.

2.3 A query sequence is aligned against the alignment of the sequences from the global swine H1 reference tree.

2.4 The query sequence is placed into the reference tree using pplacer (Matsen, 2010), with the reference tree serves as a “scaffold” where the query sequence is hung.

2.5 The clade for the query sequence is inferred from its local environment in the phylogeny:

- When a query sequence is placed unequivocally within the bounds of a single defined clade, this clade name is assigned to the query.
- When a query sequence does not fall within a defined clade, it is classified by the neighborhood of its placement, using a “-like” annotation. For example, the tool assigns the classification “1B.2.2-like” to viruses ancestral to both the 1B.2.2.1 and 1B.2.2.2 clades but not placed within the 1B.2.2 clade.

2.6 The output clade annotations are stored in the database and then displayed on the Sequence Search Results page and the Segment/Protein Details page (Section 3). In addition, a global swine H1 clade search option is provided under the Search Data drop-down menu (Section 3).

3. Display of H1 Clade Classifications

IRD-annotated H1 clade classifications are displayed in the Strain Information section on the Strain/Segment/Protein Details page (Figure 1), Sequence Search Results page (Figure 2). H1 clade annotations can also be used as one of the sequence search criteria on the Swine H1 Clade Sequence Search page (**Search Data > Search Sequences > Swine H1 Clade**

[https://www.fludb.org/brc/influenza_h1clade_search_segment.spg?method=ShowCleanSearch&decorator=influenza], Figure 3).

IRD Influenza Research Database

About Us Community Announcements Links Resources Support Sign Out

yun.zhang@jvri.org

SEARCH DATA ANALYZE & VISUALIZE WORKBENCH SUBMIT DATA HELP

Home > Nucleotide Sequence Search > Results

Nucleotide Sequence Search Results

Your Selected Items: 0 items selected

Add to Working Set Save Search Run Analysis Download

Your search returned **1,051** segments. Search Criteria Displaying 50 records per page, sorted by **Strain Name** in ascending order. Display Settings

☐ Select all 1,051 segments

1 2 3 4 5 6 7 Next > Page: 1 of 22


| | Segment | Protein Name | Sequence Accession | Subtype * | Collection Date | Host Species | Country | Flu Season (SOP) | Strain Name | US Swine H1 Clade | Global Swine H1 Clade | |
|--------------------------|----------------------|--------------|--------------------|-----------|-----------------|--------------|---------|------------------|-------------|---------------------------------|-----------------------|-----------|
| <input type="checkbox"/> | View | 4 | HA | CY089159 | H1N1 | 02/15/2008 | *Swine | USA | 07-08 | *A/swine//Ohio/02026/2008(H1N1) | gamma | 1A.3.3.3 |
| <input type="checkbox"/> | View | 4 | HA | KX247675 | H1N2 | 04/06/2016 | Swine | USA | 15-16 | A/swine/Alabama/A01104091/2016 | delta1 | 1B.2.2.1 |
| <input type="checkbox"/> | View | 4 | HA | CY025002 | H1N1 | 1977 | Swine | USA | -N/A- | A/swine/Arizona/148/1977 | alpha | 1A.1-like |
| <input type="checkbox"/> | View | 4 | HA | CY082359 | H1N1 | 01/31/2007 | *Swine | USA | 06-07 | A/swine/Arkansas/01460/2007 | beta | 1A.2 |
| <input type="checkbox"/> | View | 4 | HA | CY157991 | H1N1 | 01/31/2007 | *Swine | USA | 06-07 | A/swine/Arkansas/01460/2007 | beta | 1A.2 |
| <input type="checkbox"/> | View | 4 | HA | CY158929 | H1N2 | 10/10/2008 | *Swine | USA | 08-09 | A/swine/Arkansas/02404/2008 | delta1 | 1B.2.2.2 |
| <input type="checkbox"/> | View | 4 | HA | CY159915 | H1N2 | 04/06/2009 | *Swine | USA | 08-09 | A/swine/Arkansas/02720/2009 | delta2 | 1B.2.1 |

Figure 1. A section of the Nucleotide Sequence Search Results page

Strain Information[†]

| | | | |
|----------------------------------|---|--------------------------|--|
| Unique Sample Identifier: | IRD-A01104091 | Host: | Swine |
| Complete Genome Set: | Yes | Collection Date: | 04/06/2016 |
| Organism Name: | Influenza A Virus | Flu Season (SOP): | 15-16 |
| Strain Name: | A/swine/Alabama/A01104091/2016 View Strain Details | Isolation Country: | USA |
| Subtype: | H1N2 | Isolation Source: | lung |
| Global Swine H1 Clade (SOP): | 1B.2.2.1 | Passage History: | 1 |
| US Swine H1 Clade(SOP): | delta1 | GenBank Submission Date: | 06/18/2016 |
| 2009 Pandemic H1N1-like (SOP) ?: | Mixed Positive and Negative Segments | NCBI Taxon ID: | 1850173 |
| | | GenBank Notes: | passage details: 1 |
| | | GenBank Header Notes: | GenBank Accession Numbers KX247672-KX247679 represent sequences from the 8 segments of Influenza A virus (A/swine/Alabama/A01104091/2016(H1N2)). |

Figure 2. An example of the Strain Information section on the Segment/Protein Details page



Influenza Research Database

[About Us](#) [Community](#) [Announcements](#) [Links](#) [Resources](#) [Support](#)

[Sign Out](#)

SEARCH DATA
ANALYZE & VISUALIZE
WORKBENCH
SUBMIT DATA
HELP

[Home](#) > Swine H1 Clade Sequence Search

Swine H1 Clade Sequence Search ?

An IRD algorithm classifies the clade of the HA of H1 viruses, from any host and for any NA subtype, with reference to the USDA classification of Global and US swine H1 viruses. This algorithm, which is based on phylogenetic analysis, is an adaptation of that used for classifying HA(H5) sequences; it was developed by IRD team member Catherine Macken, in conjunction with Tavis Anderson and other swine influenza experts at the USDA. It has been verified as highly accurate (> 99%) for sequences of at least 300 nucleotides of HA1. See [US Swine Classification SOP](#) for more details. Those HA's not belonging to any of the recognized US swine H1 clades are given the classification "Other." Non-segment 4 sequences from a virus with a Global and US swine H1 classification are given the same assignment as the HA.
[Representative tree of swine HA\(H1\) sequences showing named US swine clades](#)
[Description of US clades with name that include "-like"](#)

Results matching your criteria: 3,326

DATA TO RETURN

☒ Segment / Nucleotide

☐ Protein

☐ Strain

CLADE CLASSIFICATION

☒ Global Swine H1 Clade (SOP)

☐ US Swine H1 Clade (SOP)

All
 1A.1
 1A.1-like
 1A.1.1
 1A.1.2
 1A.1.3
 1A.2
 1A.2-3-like
 1A.3
 1A.3.1
 1A.3.2
 1A.3.2-3-like

SELECT SEGMENTS

All
 1 PB2
 2 PB1
 3 PA
 4 HA
 5 NP
 6 NA
 7 M
 8 NS

COMPLETE SEQUENCES

☐ Include Partial Sequences

☒ Complete Segments Only

☐ Include near-complete sequences (IVR) [?](#)

☐ Complete Genomes only

DATE RANGE

From: To:

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

HOST

Panda
 Reassortant
 Sea Mammal
 Swine
 Unknown
 Weasel

GEOGRAPHIC GROUPING

Europe
 North America
 Oceania
 South America

COUNTRY

Mexico
 Nicaragua
 Panama
 Puerto Rico
 Trinidad and Tobago
 USA

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

ADVANCED OPTIONS [Show All](#)

Select Advanced Option

Flu Season

FLU SEASON (SOP) [Remove](#)

16-17
 15-16
 14-15
 13-14
 12-13

Select Advanced Option

Remove Duplicate Sequences

REMOVE DUPLICATE SEQUENCES [Remove](#)

☒ Remove Duplicate Sequences - Only include one of the results if multiple nucleotide sequences are exactly identical.

[Add Another Advanced Option](#)

Clear Search

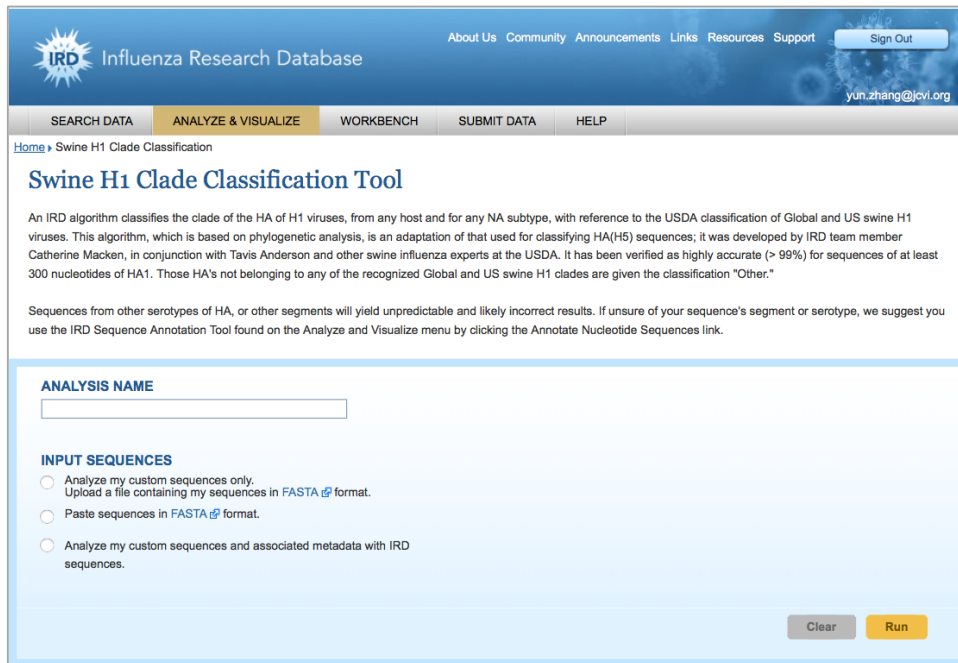
Figure 3. The Swine H1 Clade Sequence Search page

4. Clade Classification of Custom Sequences

The Swine H1 Clade Classification tool is also made available as a stand-alone tool for annotating user-provided H1 sequences. The tool is accessible from **Analyze & Visualize > Swine H1 Clade Classification Tool**

[\[https://www.fludb.org/brc/h1CladeClassifier.spg?method=ShowCleanInputPage&decorator=influenza\]](https://www.fludb.org/brc/h1CladeClassifier.spg?method=ShowCleanInputPage&decorator=influenza). User sequences can be provided by uploading a file, pasting in FASTA-formatted sequences, or a sequence file uploaded to the Workbench (Figure 4). Both the Global & the US Swine H1 Clade Classification tools will be run against the input sequences. The Clade

Classification Report provides both the clade assignment and the phylogenetic tree used to assign the clade designation to the input sequence (Figure 5).



Swine H1 Clade Classification Tool

An IRD algorithm classifies the clade of the HA of H1 viruses, from any host and for any NA subtype, with reference to the USDA classification of Global and US swine H1 viruses. This algorithm, which is based on phylogenetic analysis, is an adaptation of that used for classifying HA(H5) sequences; it was developed by IRD team member Catherine Macken, in conjunction with Tavis Anderson and other swine influenza experts at the USDA. It has been verified as highly accurate (> 99%) for sequences of at least 300 nucleotides of HA1. Those HA's not belonging to any of the recognized Global and US swine H1 clades are given the classification "Other."

Sequences from other serotypes of HA, or other segments will yield unpredictable and likely incorrect results. If unsure of your sequence's segment or serotype, we suggest you use the IRD Sequence Annotation Tool found on the Analyze and Visualize menu by clicking the Annotate Nucleotide Sequences link.

ANALYSIS NAME

INPUT SEQUENCES

☐ Analyze my custom sequences only.
Upload a file containing my sequences in [FASTA](#) format.

☐ Paste sequences in [FASTA](#) format.

☐ Analyze my custom sequences and associated metadata with IRD sequences.

Figure 4. The Swine H1 Clade Classification Tool landing page



Swine H1 Clade Classification Report

| Sequence Identifier | Global Swine H1 Clade Classification (SOP) | | | US Swine H1 Clade Classification (SOP) | | |
|--|--|--------------------------|---------------------------|--|--------------------------|---------------------------|
| | Clade Assignment | Unified Tree | Tree Viewer | Clade Assignment | Unified Tree | Tree Viewer |
| 2016054341 Influenza_A_Virus -N/A- H1 2016 Swine North_America USA Indiana delta1 | 1B.2.2.1 | Download | View Tree | delta1 | Download | View Tree |
| 400183 Influenza_A_Virus -N/A- H1 2015 Swine North_America USA Iowa delta2 | 1B.2.1 | Download | View Tree | delta2 | Download | View Tree |
| >33325 Influenza_A_Virus -N/A- H4 2015 Swine North_America USA Missouri -N/A- | Not H1 Segment 4 Sequence - Unable to determine classification | | | Not H1 Segment 4 Sequence - Unable to determine classification | | |
| >63706B Influenza_A_Virus -N/A- H3 2015 Swine North_America USA North_Carolina -N/A- | Not H1 Segment 4 Sequence - Unable to determine classification | | | Not H1 Segment 4 Sequence - Unable to determine classification | | |

Figure 5. An example of the Swine H1 Clade Classification Report

References

Anderson TK, et al. A Phylogeny-Based Global Nomenclature System and Automated Annotation Tool for H1 Hemagglutinin Genes from Swine Influenza A Viruses. *mSphere*. 2016, 1(6): e00275-16. PMID: 27981236.

Influenza Research Database. IRD US swine H1 hemagglutinin clade classification tool.
https://www.fludb.org/brcDocs/documents/SOP_SwineH1CladeClassn_062415.pdf

Matsen FA, et al. pplacer: linear time maximumlikelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. BMC Biomathematics. 2010, 11:538. PMID: 21034504.