

# SOP for New Pandemic (H1N1) Classification

**Author:** Catherine Macken (cmacken@lanl.gov)

**Submission Date:** 05/31/2011

**Introduction:** In 2009, a new H1N1 pandemic was recognized in the human population. The origin of this novel reassortant virus is believed to be swine. Subsequently, this virus transmitted to swine, and now transmits among swine. Reassortment of the new pandemic virus with viruses of other serotypes in swine and/or human hosts is possible. The Influenza Research Database (IRD) contains sequences from human and swine viruses containing segments that may be appropriately classified as “new pandemic,” i.e., having a sequence highly similar to the same segment in a known new pandemic virus. This document serves to describe how sequences are used to classify segments of human or swine viruses as “new pandemic.” From this classification, a classification of the strain to which the segments belong can be derived, using a separate SOP found below.

## Implementation Specifics for classification of segments/proteins:

### 1. Input

- a. All segment sequences in the IRD database will be evaluated

### 2. Evaluation criteria

- a. To prepare for BLAST-based classification steps, generate a classification database from all sequences of all strains in the “New pandemic” data set and in the “Other” dataset (see below). Use only 8 of the 17 available sequences for A/California/04/2009, those with GenBank accession numbers FJ966079 through FJ966086.

Classification Dataset	
Pandemic H1N1 Reference Strains (NPDM)	Non-Pandemic H1N1 Reference Strains (Other)
A/California/04/2009	A/AA/Huston/1945
A/Germany-BY/74/2009	A/AA/Marton/1943
A/Mexico/47N/2009	A/Albany/1618/1951
A/Mexico/InDRE4115/2009	A/Albany/8/1979
A/Rhode Island/08/2009	A/California/VRDL225/2009
A/Russia/190/2009	A/Canterbury/48/2001
A/Tallinn/INS372/2009	A/Fiji/15899/83
A/Thailand/CU-H1222/2010	A/Fort Monmouth/1/1947

A/Thailand/CU-1821/2010	A/Iowa/1943
A/Thailand/H1818/2010	A/Malaysia/54
A/Ulaanbaatar/190/2011	A/Memphis/2/1996
A/Yamaguchi/22/2009	A/Puerto Rico/8/34/Mount Sinai
A/Newark/INS429/2010	A/Roma/1949
	A/South Carolina/1/18
	A/swine/California/T9001707/1991
	A/swine/Chonburi/NIAH589/2005
	A/swine/England/WVL15/1997
	A/swine/France/WVL8/1992
	A/swine/Haseluenne/IDT2617/2003
	A/swine/Hong Kong/9656/2001
	A/swine/Hong Kong/NS1179/2007
	A/swine/Illinois/1/1975
	A/swine/Iowa/1/1986
	A/swine/Iowa/100/1977
	A/swine/Iowa/H02NJ56371/2002
	A/swine/Iowa/H02PW7/2002
	A/swine/Iowa/H03UWF2/2003
	A/swine/North Carolina/02084/2008
	A/swine/Ohio/24366/2007
	A/swine/Ontario/23866/04
	A/swine/Ratchaburi/NIAH1481/2000
	A/swine/Shandong/1123/2008
	A/swine/Tennessee/10/1977
	A/swine/Tennessee/112/1977
	A/swine/Tennessee/17/1976
	A/swine/Tennessee/48/1977
	A/swine/Tennessee/82/1977
	A/swine/Wisconsin/1/1957
	A/swine/Wisconsin/1/1971
	A/swine/Wisconsin/2/1970
	A/Texas/36/1991

- i. Looping through all sequences in the IRD database, BLAST the current sequence against the classification database and evaluate the three best BLAST hits for self-consistency of classification as “NPDM (new pandemic)” or “other (non-pandemic H1N1)”.
  - a. If the top 3 hits are self-consistent, and
    1. The lowest of the three scores is  $> 97\%$ , the classification associated with these 3 hits (NPDM or Other) is assigned.
    2. The lowest of the 3 scores is  $\leq 97\%$ , then the sequence is Other.
  - b. If the top 3 hits are inconsistent, but
    1. The best BLAST hit has a score  $> 98\%$ , assign the classification associated with the best BLAST hit.
    2. The best BLAST hit has a score  $\leq 98\%$ , then the sequence is Other.

### 3. Outputs

Each segment in the IRD database is assigned a classification of Y (pH1N1-like) or N (NOT pH1N1-like) corresponding to an output from the algorithm of NPDM and Other.

### 4. Classification of Strains

Since the new pandemic strain was not isolated before 2009, only influenza strains isolated from January 1, 2009 to the present (July 2011 at writing of this document) are classified as being 2009 Pandemic H1N1-like. For each post-2008 strain in the IRD database (including those isolated from all hosts and geographic locations and having any subtype), each segment is evaluated for a Y or N flag. Each strain is assigned to one of five categories based upon the following criteria:

- a. **Positive (Pandemic H1N1-like):** All 8 segments are sequenced and all 8 are classified as Y
- b. **Probable Positive (Probable Pandemic H1N1-like):** Less than 8 segments are sequenced and all sequenced segments are classified as Y
- c. **Mixed Positive and Negative Segments:** At least one segment is classified as Y and at least one segment is classified as N
- d. **Probable Negative (Probably Not Pandemic H1N1-like):** Less than 8 segments are sequenced and all sequenced segments are classified as N

- e. **Negative (Not Pandemic H1N1-like):** All 8 segments are sequenced and all sequenced segments are classified as N