Background and Objectives

In recent years, a variety of different provisional names have been used to refer to emerging lineages of the currently circulating highly pathogenic avian influenza (HPAI) H5N1 viruses. As a result, discussion, comparison, and analysis of the various lineages proved difficult.

Avian H5N1 viruses continue to spread, continue to infect animals and humans, and continue to evolve and diversify. While most viral genes have been replaced through reassortment yielding many different genotypes, the specific H5 hemagglutinin (HA) gene identified in 1996, remarkably, has remained present in all isolates. Therefore, this H5 HA provides a constant to which the evolving strains may be effectively compared, and it was proposed to develop a standard clade nomenclature system based upon the evolution of this H5 HA. This nomenclature system would enable:
1. a unified system to be developed to facilitate the interpretation of sequence/surveillance data from different laboratories;
2. the labelling of clades by geographical reference to be replaced by a more representative system;
3. the phylogenetic tree to be expanded in the future; and
4. a starting point to be established to develop a more extensive system in the near future that takes into consideration antigenic variation and reassortment into multiple genotypes.

The Unified Nomenclature System and Virus Clades

An international consultative group of scientists was convened by the World Health Organization (WHO), the World Organisation for Animal Health (OIE) and the Food and Agriculture Organization (FAO). Phylogenetic analysis was performed using a variety of approaches on all of the publicly available H5 HA sequences that have evolved from the A/goose/Guangdong/1996 H5N1 virus. The initial results showed that the currently circulating H5N1 viruses could be effectively grouped into numerous virus “clades” based on the phylogenetic characterization and sequence homology of the HA gene. Based on criteria used to distinguish various groups of the H5 hemagglutinin (HA) gene, the system has formally identified 20 distinct clades of the virus since its inception in early 2008 [1-2]. These clades are defined as meeting the following three specific clade definition criteria developed by the WHO/OIE/FAO H5N1 Evolution Working Group:
- sharing of a common (clade-defining) node in the phylogenetic tree;
- monophyletic grouping with a bootstrap value of ≥60 at the clade-defining node (after 1000 neighbor-joining bootstrap replicates); and
- average percentage pairwise nucleotide distances between and within clades of >1.5% and <1.5%, respectively.

As the viruses within these clades continue to evolve, new sublineages (potential H5N1 clades) periodically emerge. Once these sublineages meet the same three specific clade definition criteria as the initial clades (listed above), they are designated as separate clades.

See Conceptual Diagram of ongoing H5N1 evolution
pdf, 109kb
These new clades are defined as second, third or fourth-order clades and assigned a numerical 'address' which links them to their original clade using a hierarchical decimal numbering system. For example, within the distinct clade 2.3, third order clades meeting the clade definition were designated as clades 2.3.1 and 2.3.2 and so on. More recently, a new monophyletic clade was identified within clade 2.3.2 and assigned a fourth order designation as clade 2.3.2.1. This logical hierarchical numbering system is objectively related to HA phylogeny, and thus removes geographic designations (e.g., the lineage previously referred to as 'Fujian-like' became third order clade 2.3.4, while the 'Qinghai lineage' became second order clade 2.2). The H5N1 Evolution Working Group recommends that emerging sublineages of H5N1 not be classified using the hierarchical decimal numbering system until that sublineage has met all three of the above criteria for designation of a clade. However, the working group fully encourages the designation of newly emerging distinct sublineages of H5N1 virus by independent investigators using alternative, interim, naming conventions (e.g., using letters).

**Recent divergence of the highly pathogenic H5N1 hemagglutinin gene**

Surveillance for the continuing circulation of highly pathogenic avian influenza A(H5N1) in poultry and wild birds in Asia, the Middle East, Europe, and Africa has led to an increase in the number of sequences available for analysis. The current analysis consisted of 2,947 HA gene sequences consisting of at least 1,600 nucleotides, of which 1,637 sequences (primarily from 2008-2010) were added since the previous nomenclature update. Phylogenetic analysis of the neighbor-joining tree generated from this dataset revealed that the addition of this new sequence data resulted in the generation of one or more monophyletic groups with high bootstrap support within each of the currently circulating clades of H5N1. In addition, many of these groups had long branch lengths separating them from the nearest node in the tree indicating further nucleotide divergence.

![See full tree figure](pdf, 305kb)

Each of the currently circulating clades (1, 2.1.3, 2.2, 2.2.1, 2.3.2, 2.3.4 and 7) were then examined independently to measure the average within-group pairwise nucleotide distances and all were found to have >1.5% within-group divergence indicating the need to split these groups into new order clades. After generation of clade-specific trees, monophyletic groups with bootstrap values >60 were selected and tested to determine the within- and between-group p-distances. Based on the previously defined nomenclature, 12 new second, third, and fourth-order clades were identified.

To summarize these findings and to confirm the consistency of clade topology using a smaller dataset, a small tree containing 196 HA sequences was generated using both NJ and Bayesian methods. Both NJ bootstrap support and Bayesian posterior probabilities (shown on top and bottom of each clade-defining node, respectively) were calculated to demonstrate confidence in clade assignments. This tree is annotated with the 12 new clades presented here and displays older, non-circulating and putatively non-circulating clades as collapsed nodes.

![See small tree figure](pdf, 776kb)

The designation of 12 new H5N1 clades (in contrast with the previous nomenclature update in which only one new clade was identified [2]) is not surprising considering the length of time since the last nomenclature update and improved surveillance/reporting of viruses in recent years, leading to an
increasing number of sequences available for analysis. Interestingly, the clades that had already diverged into third-order groups in previous analyses required the largest number of new clade designations. Previous assignment to third-order clades indicates that the viruses had already evolved considerably, usually because of short generation times and large population sizes in enzootic areas. In addition to the emergence of novel clades of H5N1, a number of previously circulating clades of H5N1 have not been detected for several years (since at least 2008). Although we cannot rule out the possibility that the lack of detection of viruses in these clades may in some cases be related to gaps in surveillance, it is more likely that many of these clades have been supplanted by new clades and become inactive. 13 clades (shown as collapsed nodes in the small tree with the last year of detection shown in parentheses) have not been detected since at least 2008 (0, 2.1.1, 2.1.2, 2.3.1, 2.3.3, 2.4, 2.6, 3, 4, 5, 6, 8, and 9).

As a consequence of HA gene diversification, the classification of H5N1 viruses requires periodic updating, making classification dynamic as the virus has expanded within several disparate ecosystems and along distinct evolutionary trajectories. Although this report focuses strictly on genetic divergence as a measure of H5N1 classification, clearly additional studies related to H5N1 antigenicity and cross-clade protective immunity are needed to better inform pre-pandemic vaccine strain selection as well as risk management policies for both veterinary and human health. As always, continued surveillance for H5N1 in animals and humans is crucial for this process to be most effective.

Updated trees and nucleotide alignments (as well as archived trees/alignments for reference) are available on this website.

References


For more information

Email: gisrs-whohq@who.int