Analyzing a Sequence-Based Phylogenetic Tree to Understand Viral Evolution

In 2009, an influenza A H1N1 virus caused a pandemic and the virus has continued to resurface in outbreaks across the world. Researchers in Taiwan were curious about why the virus kept appearing despite widespread flu vaccine initiatives. They hypothesized that newly evolved variant strains of the H1N1 virus were able to evade human immune system defenses. To test the hypothesis, they needed to determine if each wave of flu infection was caused by a different H1N1 variant strain.

Scientists obtained the genome sequences for 4,703 virus isolates collected from patients with H1N1 flu in Taiwan. They compared the sequences in different strains for the viral hemagglutinin (HA) gene, and based on mutations that had occurred, arranged the isolates into a phylogenetic tree (Figure 1). In the phylogenetic tree, each branch tip is one variant strain of the H1N1 virus with a unique HA gene sequence.

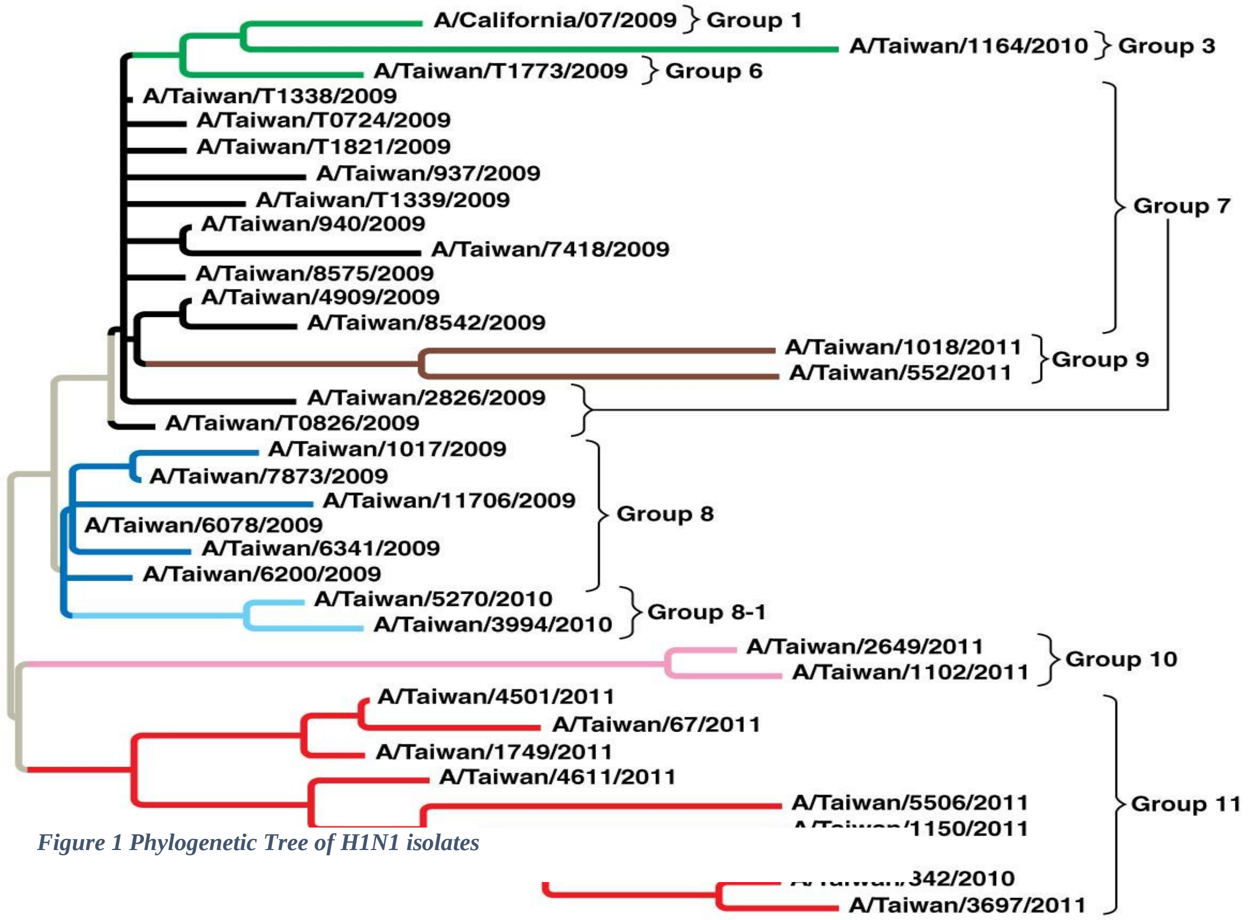


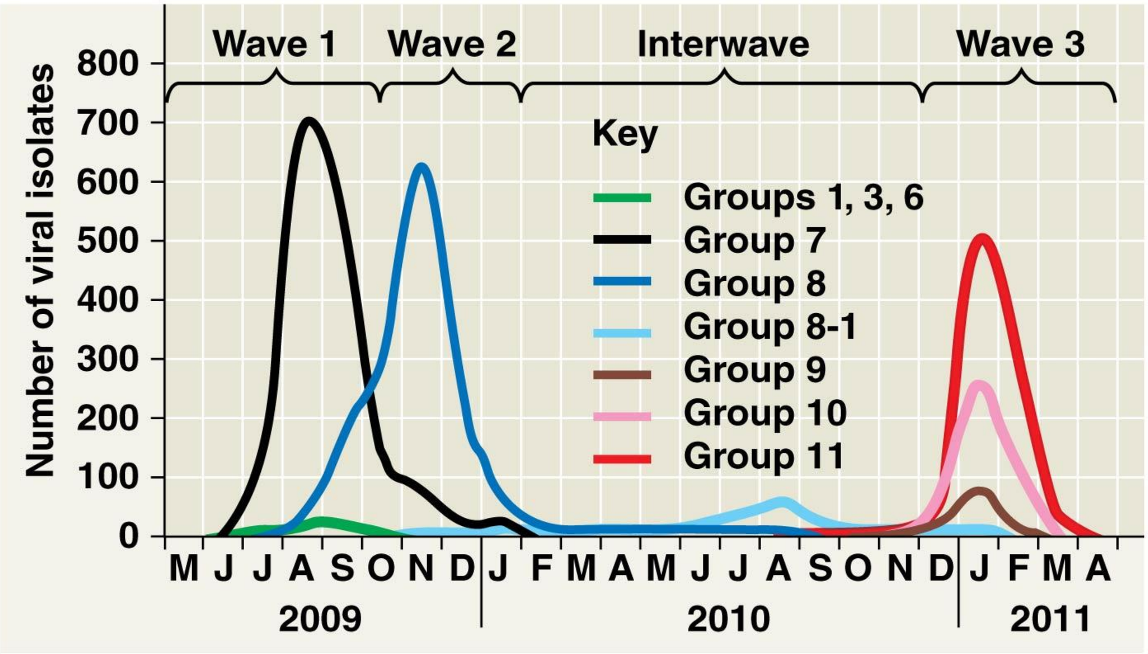
Figure 1 Phylogenetic Tree of H1N1 isolates

Data from J-R. Yang et al., New variants and age shift to high fatality groups contribute to severe successive waves in the 2009 influenza pandemic in Taiwan, *PLoS ONE* 6(11): e28288 (2011). doi:10.1371/journal.pone.0028288.

The phylogenetic tree shows the hypothesized evolutionary relationship between the variant strains of H1N1 virus. Each branch tip is one variant of the H1N1 virus with a unique HA gene sequence. The more closely connected two variants are, the more alike they are in terms of HA gene sequence. Each fork in a branch, called a node, shows where two lineages separate due to different accumulated mutations. The number of nodes separating two variants is an indication of how many DNA sequence differences there are between the variants, thus how distantly related they are.

1. Referring to the phylogenetic tree in Figure 1, which variant is more closely related to A/Taiwan1018/2011 in group 9 (brown): A/Taiwan/552/2011 or A/Taiwan/8542/2009?

2. The scientists arranged the branches into groups made up of one ancestral variant and all of its descendant, mutated variants. They are color-coded in the tree. Do all of the nodes have the same number of branches or branch tips? Are all of the branches in the group the same length? What do these results indicate?



Data from J-R. Yang et al., New variants and age shift to high fatality groups contribute to severe successive waves in the 2009 influenza pandemic in Taiwan, *PLoS ONE* 6(11): e28288 (2011). doi:10.1371/journal.pone.0028288.

Figure 2 Number of Isolates by Month and Year of Collection

3. Figure 2 shows the number of isolates collected (each from an infected patient) by the month and year the isolates were collected. Each group of variants is plotted separately with a line color that matches the phylogenetic tree.
a) Which group of variants was the earliest to cause H1N1 flu in over 100 patients in Taiwan?

b) After a group of variants had a peak number of infections, did members of that same group cause another wave of infection?

c) Does this support or refute the researchers’ hypothesis?

4. One variant in group 1 (green) was used to make a vaccine that was distributed very early in the pandemic. Based on the graph, does it look like that vaccine was effective?

5. Groups 9, 10, and 11 all had H1N1 variants that caused a large number of infections at the same time in Taiwan. Does this mean that the scientists' hypothesis, that new variants cause new waves of infection, was incorrect? Explain your answer.