Critical Infrastructure Protection Initiative
@ Dalhousie University

Change in the Air: Genetic monitoring of H1N1 during the 2009 outbreak

Robert G. Beiko

Working Paper 01

October 28, 2012

Working papers are in draft form. This working paper is distributed for purposes of comment and discussion only. It may not be reproduced without permission of the corresponding author.

DALHOUSIE UNIVERSITY
Inspiring Minds

University of Strathclyde Glasgow

Author institutions and contact details

Robert G. Beiko[1]
Associate Professor and Canada Research Chair in Bioinformatics
Faculty of Computer Science
Dalhousie University
Halifax
Canada

This paper is a draft. It was prepared for discussion at the CIP Initiative / Pandemics Workshop on October 31, 2012. The event will occur in the immersive world software *Second Life* and is hosted jointly by the School of Public Administration at Dalhousie University, Halifax, Canada, and the Department of Management Science at the University of Strathclyde, Glasgow, Scotland. Comments on the paper are welcome at the workshop or by email to one of the authors. As it is a draft, please do not cite the paper without the permission of the corresponding author.

[1] Corresponding author: Robert G. Beiko. Beiko@cs.dal.ca

Abstract
The nature of Influenza A – its toxicity, its defense mechanisms, and its ability to make myriad copies of itself in a short timespan – is all encoded in its genome. The chemical constitution of its genome, which is made of RNA rather than the more familiar DNA, makes it one of the fastest-changing biological entities known. During the course of a single year, new variants emerge throughout the globe, allowing it to rapidly adapt to new environmental challenges and explore new hosts. Influenza A also packages its genome in a unique way, with no more than one or two genes on any given segment of RNA: this structure allows genes to be swapped between strains, generating novel and potentially dangerous new hybrids.

The 2009 H1N1 pandemic (or H1N1pdm) strain of Influenza A came under unprecedented scrutiny, thanks to the availability of fast and cheap genome sequencing platforms. The combination of technology and need meant that within three weeks of the strain coming into public view, major articles analyzing several H1N1pdm genomes from different geographic locations had been published, offering early insights into the key traits of this new variant. Genetic monitoring continued, with sequences ultimately contributed from over 70 countries. When combined with new analysis tools to infer the evolutionary history of these viruses, associate variation in their genomes with variation in their traits, and visualize their geographic distribution, important insights were gained about the nature of the H1N1pdm strain.

Genetic monitoring of emerging pathogens will soon be routine, as new outbreaks of viruses and bacteria including *E. coli*, *Listeria*, and *Clostridium difficile* are being subjected to sequencing immediately and used to identify probable sources. These rapid diagnostic techniques, which yield rich information, create new opportunities to understand the workings of emerging pathogens and limit their impact.

Background: The dynamic genome of Influenza

The Influenza family of viruses, most notably the Influenza A genus, are a set of deadly pathogens that incur millions of hospital days and over $10 billion in direct medical costs every year in the United States alone (Molinari et al., 2007). While mortality rates from the common subtypes of Influenza A are typically well below 0.1%, epidemic and pandemic strains can spread rapidly through the population with much higher mortality rates. For example, the infamous 1918-1920 "Spanish flu" pandemic strain killed approximately 50 million people worldwide (although this figure may be an understatement: see for instance Johnson and Mueller, 2002), while mortality rates, although difficult to estimate from historical data, were in excess of 2% and in some countries possibly over 10% (Johnson and Mueller, 2002). The variation in virulence and mortality is mirrored by other observable differences among seasonal and pandemic strains: for example, strains differ widely in their host preference, transmissibility, and reproductive rate (Jhung et al., 2011). Particularly worrisome is the increase in resistance to antiviral drugs such as oseltamivir and adamantane (Bright et al., 2005) in seasonal strains over time, presumably in response to the increased use of these drugs in prophylaxis and treatment.

What underpins the substantial differences we see in different variants of the same virus? The key to Influenza is in its lifecycle, which roughly consists of three steps (Figure 1, see also Bouvier and Palese, 2008): (i) adhesion to and infection of a host cell, (ii) replication, where the virus makes many copies of itself, and (iii) destruction of the host cell and release of the newly created viruses. This process allows a single virus to multiply rapidly during its attack on a host, and the disease pathology results from a combination of the direct destruction of cells, and the host's immune response to the viral infection. The effectiveness of host infection is dependent on the progression of the lifecycle, and different variants behave differently during these three steps. For example, a parallel is frequently drawn between seasonal H1N1, which can spread widely but is rarely lethal, and the 'bird flu' H5N1, which is much more dangerous (case fatality ratio as high as 50%: Toner and Adalja, 2012) but spreads only rarely from one human host to another. One important reason for this difference lies in the targeting of specific cells by the different strains: H1N1 attacks our upper respiratory system, while H5N1 attacks our lower respiratory system due to its preference for a different type of lung epithelial cell (Korteweg and Gu, 2008). Upper respiratory tract infections are more transmissible but less severe, whereas the opposite is true in the lower tract. The differences between the 1918-1920 pandemic strain and more recent seasonal variants were explored in a controversial study that rejuvenated and studied the pandemic strain under laboratory conditions (Tumpey et al., 2005). Many differences were found: most notably, the pandemic strain was more effective at replication in bronchial cells, was far more lethal in mice, and induced a much stronger immune response (including the so-called "cytokine storm") which was actually responsible for many deaths in 1919.

These critical differences in transmission and lethality arise from differences at the genetic level in the virus. All living organisms have a genome, made of DNA and comprising many genes that encode the proteins they need to carry out the processes that lead to growth and reproduction. Viruses are considered by many to be non-living because their life cycle depends completely on reproduction within a host cell. They nonetheless also have genomes that contain genes which encode proteins; however, since they make such heavy use of the host's genetic machinery to copy themselves, their genomes can be extremely small by comparison. A typical mammalian genome has just over 3 billion nucleotides ("letters": A, G, C and T) of DNA and 20,000 – 30,000 genes, while most bacterial genomes have several million nucleotides and 2000-5000 genes. The genome of Influenza A is minuscule by contrast, approximately 14,000 nucleotides in length and containing only eleven genes. Each of these genes plays a role in viral attachment, infection, copying, and escape from the host cell. Two genes of particular importance are hemagglutinin (HA), which is responsible for initial binding to the host cell, and neuraminidase (NA), which is responsible for release after the virus has copied itself (Palese and Schulman, 1976). Variation in the HA and NA genes confers the "type" of a particular strain of Influenza: for example, H1N1 and H5N1 have dissimilar HA genes, but similar (although not identical) NA genes. The dissimilarity in HA genes is the primary reason why H1N1 but not H5N1 can readily infect human hosts (Korteweg and Gu, 2008).

What makes the Influenza virus family particularly adaptable and dangerous is its propensity for rapid change. Two properties of the genome in particular are responsible for this. First, the Influenza genome is comprised of ribonucleic acid (RNA) rather than DNA. RNA is copied less faithfully than DNA, and the mutation rate in Influenza is consequently about 1000-fold higher than in the human host (Nobusawa and Sato, 2006). Second, although it has only eleven genes, the Influenza A genome is split up into eight distinct segments, each of which contains either one or two genes (Figure 2). If two distinct strains of Influenza infect the same host, their genetic material can be reshuffled through the process of *reassortment*, producing offspring that inherit some of their genes from each of the original strains. The high mutation rate means that new gene variants are constantly being generated and tested: some of these variants have become widespread as they confer resistance to new antiviral drugs such as oseltamivir (Tamiflu) and adamantine. Reassortments have been responsible for several pandemics, by generating new gene combinations that are particularly dangerous in humans (Holmes et al., 2005). For example, the 2009 H1N1 pandemic strain (H1N1pdm) is usually described as a triple or quadruple reassortant, since various components of its tiny genome originated from strains adapted to pigs, birds, and humans (Smith et al., 2009).

The tiny genome of Influenza makes it an ideal target for complete sequencing: indeed, while the first cellular organism genome was not completed until 1995, the eight segments of isolate A/Puerto Rico/8/1934 were all sequenced by 1982 (Winter et al., 1982). Given the extremely high rate of change in the Influenza genome, isolating strains from patients and determining the complete sequence of the corresponding genome is valuable for several reasons. Mutations that

arise in an Influenza genome are passed on to its offspring, so finding similar mutations in a set of isolates can indicate that they all came from the same recent patient. Mutations can therefore be used to track the spread of seasonal or pandemic Influenza. Also, since some mutations can influence key properties such as host specificity, reproduction rate and resistance to antivirals, sequencing a large number of isolates can be used to discover new associations between genotype and properties of the virus. Conversely, if such a relationship is already known, then sequencing many isolates can reveal the incidence of, for instance, resistance to antiviral drugs. Sequence-based monitoring also has the potential to reveal new variants shortly after they emerge; this strategy allowed the rapid identification of the pandemic 2009 strain and estimates of its potential to cause harm.

Early stages of the 2009 outbreak

Until 2009, the three dominant strains of seasonal Influenza infecting humans were an H1N1 strain that originated in 1977 (Webster et al., 1992), an H3N2 strain, and a strain of the related Influenza B virus. In 2009, an outbreak of Influenza occurred in La Gloria, Mexico, starting in the middle of February. One sample from La Gloria, from a five-year-old girl, was not initially typeable as either H1 or H3 (López-Cervantes et al., 2009). The first cases reported in the USA, both from children in southern California, were initially negative for H1N1, H3N2 and H5N1, but confirmed as novel, swine-origin H1N1 through DNA sequencing (Centers for Disease Control, 2009), after which the strain from Mexico was also confirmed to be novel H1N1 (López-Cervantes et al., 2009).

The remarkable convergence of surveillance, DNA sequencing, and novel techniques for analysis of the virus and its genome allowed rapid analyses to be performed in parallel, even before the new H1N1 strain achieved pandemic status. The first 2009 H1N1 sequences were uploaded to the National Center for Biotechnology Information Influenza Virus Resource (NCBI IVR: Bao et al., 2008) in late April, allowing immediate access to all researchers. The Canadian National Microbiology Laboratory played a significant role in sequencing the 2009 strain, contributing the first full genome sequence (Public Health Agency of Canada, 2010). Initial reports on the likely evolutionary origins (Smith et al., 2009); reproductive rate, transmissibility, and lethality (Fraser et al., 2009); antigenic properties (Garten et al., 2009); and case summaries (Dawood et al., 2009) were all published by the month of June, a remarkable occurrence given that stepping up of surveillance only took place in mid-April. This rapid generation of information provided estimates of the likely spread and impact of the virus, as well as highlighting the strength of sequence-based surveillance and identifying gaps in our knowledge.

The evolutionary origin of H1N1pdm was assessed using comparative techniques that infer the relationships between the new sequences and those from earlier pandemic and seasonal strains, and animal-associated strains. These relationships are represented using separate phylogenetic trees for each of the eight genome segments, since each can potentially have a separate origin. In such trees, the closest relatives of the new 2009 sequence were assumed to be the source for the

2009 strain. The analysis showed a tangle of reassortment events: the majority of gene segments including HA came from a swine-associated virus that was itself derived from multiple reassortments, while two segments including NA were derived from a different swine-associated virus that was "avian-like" in nature (Smith et al., 2009; see Figure 3). A comparison of 2009 outbreak gene segments against known sequences in the reference database in each case identified differences of 3-6%, a significant difference corresponding to roughly ten years of evolutionary change (Smith et al., 2009; Garten et al., 2009; Fraser et al., 2009). Given that all segments are derived from swine-associated viruses, it appears that H1N1pdm may have been circulating undetected in swine populations for many years, with international shipping of pigs creating new opportunities for viral reassortment.

Similar analyses, using epidemiological models and permitted the early investigation of important characteristics such as the reproductive rate and lethality of the virus. The case fatality ratio was estimated to be between 0.03% and 1.8% depending on the model used and the inclusion or exclusion of suspected (but not confirmed) cases from the calculation (Fraser et al., 2009). By modeling the genetic diversity of the already-sequenced viral isolates, the origin of the 2009 outbreak was traced to January of that year, albeit with fairly high uncertainty. The basic reproductive number R_0 expresses the number of additional cases generated by an infected individual in a completely susceptible population. The R_0 of H1N1pdm was estimated to be 1.5, with this figure again dependent on model choice and subject to high uncertainty, with the additional confounding factor of age-dependent infection rates; nonetheless, this rate was approximately half that of the estimated number for first round of the 1918 H1N1 pandemic (Mills et al., 2004). R_0 was combined with estimates of current (as of late April 2009) infected population size to calculate the number of generations that had elapsed since the origin of the outbreak. The estimated number of generations had a high associated degree of uncertainty; even with a fixed spreading model the range was 14 to 73 generations. Although the specific number of generations could not be pinned down, from this analysis it was clear that many rounds of human-to-human transmission had taken place.

The pandemic response varied enormously from country to country, but at first was strongly influenced by the WHO pandemic alert status. Different jurisdictions implemented pharmaceutical (antivirals, to which H1N1pdm was largely sensitive) and non-pharmaceutical (e.g., school closures, border screening and other 'social distancing') interventions to varying degrees of effect (Leung and Nicoll, 2010). In the United States, the case fatality ratio guides estimates of the severity of impact (http://www.flu.gov/planning-preparedness/community/community_mitigation.pdf), and initial sequence analyses showing a likely preference for the upper respiratory tract and a low likelihood of inducing a "cytokine storm" (Nazinitsky and Rosenthal, 2010) gave useful clues as to expected fatalities. However, given that a novel strain is likely to have properties never seen before, it is impossible to predict all epidemiological traits based on sequence similarity to known strains. A deeper understanding

of key attributes of H1N1pdm would depend on observations of the course of the outbreak, at both the genetic and population levels.

Geographic analysis of 2009 sequence data

As the outbreak progressed, more variants were identified and many new sequences were deposited in the public databases. Figure 4 summarizes the distribution by country of 746 HA gene sequences that had been deposited in the NCBI IVR by the end of July 2009.

While the 2009 pandemic strain was distinguishable from other pandemic and seasonal strains based on serological, epidemiological and many other properties, fine-grained distinctions within the 2009 pandemic isolates are more readily explored using sequence data. While not all changes in the RNA sequence of the virus lead to meaningful change in its behavior, the spread of the virus can be tracked using sequence changes whether or not they affect the phenotype. With this sequence data as a basis, our research group explored the evolution and geographic spread of the 2009 pandemic strain using a pair of applications: GenGIS (Parks et al., 2009a) and SeqMonitor (MacDonald et al., 2009). Both of our initial analyses were published in *PLoS Currents: Influenza*, a medium for rapid dissemination of research results that published over 50 articles during the initial pandemic period.

Our initial motivation for developing GenGIS was to study microbial diversity as it relates to physical geography and habitat: GenGIS is a 3D geographic information system that allows a researcher to merge map data with other habitat information such as temperature, alongside biodiversity information. We analyzed several Influenza datasets in GenGIS (Parks et al., 2009b). Rhiza Labs' FluTracker (http://flutracker.rhizalabs.com/) aggregated published reports of flu cases globally from the media with information about geographic location; we formatted these data for use in GenGIS in order to show the spread of the virus in terms of the number of confirmed cases. Figure 5 shows the contrast in confirmed cases at the beginning and end of a ten-day interval in late April that bookended the initial media storm. A handful of confirmed cases in the southwest United States and New York on April 22 became a set of cases covering five continents by May 1, with many locations reporting > 100 cases.

Using sequence data allowed us to profile the spread of particular variants within the 2009 pandemic strain. The timeline of change in the H1N1 genome is days or weeks, and variants that emerge during an outbreak can give clues about patterns of geographic spread. One amino acid in the NA protein showed interesting patterns of variation: although close to the position of a well-known mutation that can confer resistance to the antiviral drug oseltamivir, there is no evidence to suggest the amino acid we identified (at position 248 in the protein) has any relationship with the resistance phenotype. Indeed, nearly all 2009 H1N1pdm isolates were oseltamivir-sensitive. Nonetheless, the global distribution of site 248 amino acids could be informative, and we tracked its progression through time (Figure 6). It was evident from very early on that both sequence types were present, and they both spread globally, with some regions

having only one or the other variant, but other regions (notably China, Japan, Mexico and the US) showing both. This suggests that both variants were present before monitoring of the outbreak began, and both variants spread rapidly around the globe. Although this variant was not reported as being epidemiologically important in the literature, other sites were found to affect the severity of clinical outcome such as position 222 in the HA protein (Kilander et al., 2010).

In addition to tracking individual variant sites in the set of isolates, we used the full set of genetic information from 203 sequenced genomes to construct "family trees" or phylogenies of the 2009 isolates. As Figure 7 shows, while there is some geographic correlation with the relatedness of different isolates, there are also examples where the closest relative to a North American isolate is an isolate from Asia, rather than the other North American isolates. New York is both a global travel hub and a location where extensive sequencing was carried out; its status was reflected by a wide diversity of H1N1 lineages. Influenza A, which spreads rapidly through the air, is known to show little global geographic structure, in sharp contrast with other pathogens which travel more slowly such as the dysentery bacterium *Shigella sonnei* (Holt et al., *Nat Genet* 2012). A study at the University of California, San Diego provided a remarkable demonstration of the "geographic fluidity" of H1N1pdm: sampling carried out over a month in late 2009 showed evidence for at least 24 distinct introductions of the virus onto the campus (Holmes et al., 2011).

We developed the SeqMonitor Web application to combine and harness the wealth of data that appeared online during the 2009 outbreak, including not only sequence information but also information about geographic origin, patient data where available, and information about antiviral resistance that was assessed for some but not all isolates. The key objective of SeqMonitor is to allow the "mining" of sequence and associated data, in order to identify new features that may be important in tracking H1N1 and anticipating its impact. For example, relatively few zanamivir-resistant isolates were identified early on the outbreak, in marked contrast with the resistance profile of seasonal flu strains from previous years (Cheng et al., 2009). SeqMonitor provides a Google Maps-based interface to view the location, timeline, and identifiers of strains matching particular search criteria; Figure 8 shows the distribution of isolates showing the zanamivir resistance trait. These isolates are widespread and suggest a substantial number of unobserved resistant viruses circulating globally.

Since the emergence of new sequence variants can indicate shifts in the properties of the virus, SeqMonitor also allows a user to choose or upload an Influenza sequence and determine whether any of its amino acids are unique or rare relative to the full set of known sequences. In Figure 9 we show an NA sequence that bears the critical mutation at position 275 that confers zanamivir resistance, one of the six shown on the map in Figure 8. The sequence visualization is again based on Google Maps, which allows the user to inspect the sequence and highlight specific amino acids of interest. Coloured bars on top of each amino acid show cases where that amino acid is universally present in all seasonal and pandemic isolates (no bar shown), very common in a set (green bar), relatively rare (yellow bar) or unique (red bar). The top line represents all pandemic and seasonal sequences, and the abundance of green and yellow bars shows that the

majority of sites in the NA protein are modified somewhere in the database, a testament to the rapid changes that take place in the protein. Although this protein is a rarity in the pandemic set because it has a Y at position 275, it has no unique features when compared against the entire sequence database.

Long-term outcomes and prospects for genetic surveillance of pathogens

The 2009 Influenza A pandemic came about in an age of increased pandemic preparedness, and increased availability of antivirals, vaccines and intensive care facilities (Leung and Nicoll, 2010). It was also the first pandemic in which sequencing technology and information technology combined to give a detailed early picture of the virus's origins and likely properties. Analysis of 2009 data has highlighted gaps in surveillance, most notably in the surveillance of swine populations which appear to be not only reservoirs of Influenza but also genetic "melting pots" in which strains can evolve unseen for many years, and reassortment can generate dangerous new combinations (Smith et al., 2009). These efforts have improved markedly since 2009, especially in regions such as Southeast Asia in which many different subtypes are known to circulate (Trevennec et al., 2011). Better surveillance of swine populations will allow us to keep track of emerging variants, and to better anticipate the risk to humans associated with these variants. Recently, it was shown that another swine variant, H3N2, has reasserted with the 2009 pandemic strain to generate a new variant that has become established in Guangxi and Hong Kong (Fan et al., 2012), and in the United States (Liu et al., 2012).

Another important discovery from the 2009 pandemic was the extent of mixed infections and the potential for new variants to emerge in single patients. A study of college students in Beijing found that six of forty Influenza-infected patients were co-infected with both H1N1pdm and seasonal H3N2 (Liu et al., 2010). Ghedin et al. (2011) intensively sequenced Influenza A viruses associated with a 15-year-old immunocompromised patient, and found high levels of genetic diversity. Most strikingly, they discovered that resistance to oseltamivir, a widely used antiviral, emerged in the patient over the course of their infection. This evidence suggests that the prolonged period of infection in immunocomproised individuals may give rise to greater opportunities for viruses to adapt to the host and evolve antiviral resistance.

The trivalent vaccine for 2009 included seasonal H1N1 and H3N2 Influenza A and an Influenza B strain. It appears that the pre-existing seasonal H1N1 strain has been eliminated since the emergence of H1N1pdm, possibly due to the more-virulent H1N1pdm causing immunity to the seasonal variant (Palese and Wang, 2011; Pica et al., 2012). This pattern has been observed before, with previous pandemics leading to replacement of previous seasonal strains in 1957 and 1968, but not in 1977, as the two strains were too dissimilar to produce any cross-strain-derived immunity. Thus, sequencing and other characterizations of new strains are vital to determine the extent to which current vaccines may need to be modified in subsequent years.

Genome sequencing has tremendous potential to transform the monitoring and diagnosis of Influenza and pathogens in general. Sequence-based monitoring programs are in place or planned for pathogens such as *Listeria* and enteric bacteria such as *E. coli*. As shown here, sequence analysis can be used for source tracking, prediction of pathogen traits, modeling of transmission, and estimation of true infection rates. Antigen testing was widely used by clinicians, but the sensitivity of the test was low, potentially leading to increased transmission as individuals falsely tested negative for the virus (Jernigan et al., 2011). By contrast, sequencing efforts by the CDC and other major institutions led to the rapid development of molecular diagnostics that were far more sensitive and could easily be retargeted between the seasonal and pandemic variants. However, these diagnostic tools were primarily used in public health laboratories and research hospitals, and widespread adoption of these techniques will raise new challenges for sample processing and data analysis. The applications we developed offer new ways to access, analyze and visualize critically important genomic variation in emerging pathogens, which can assist in the monitoring of new threats to public health.

Figures

Figure 1. Influenza A lifecycle, from Pinto and Lamb (2006). Critical steps are (1) attachment which is dependent on the HA protein, (5) replication (copying) which requires several other viral proteins, and (7) release, which is dependent on the NA protein. © the American Society for Biochemistry and Molecular Biology.

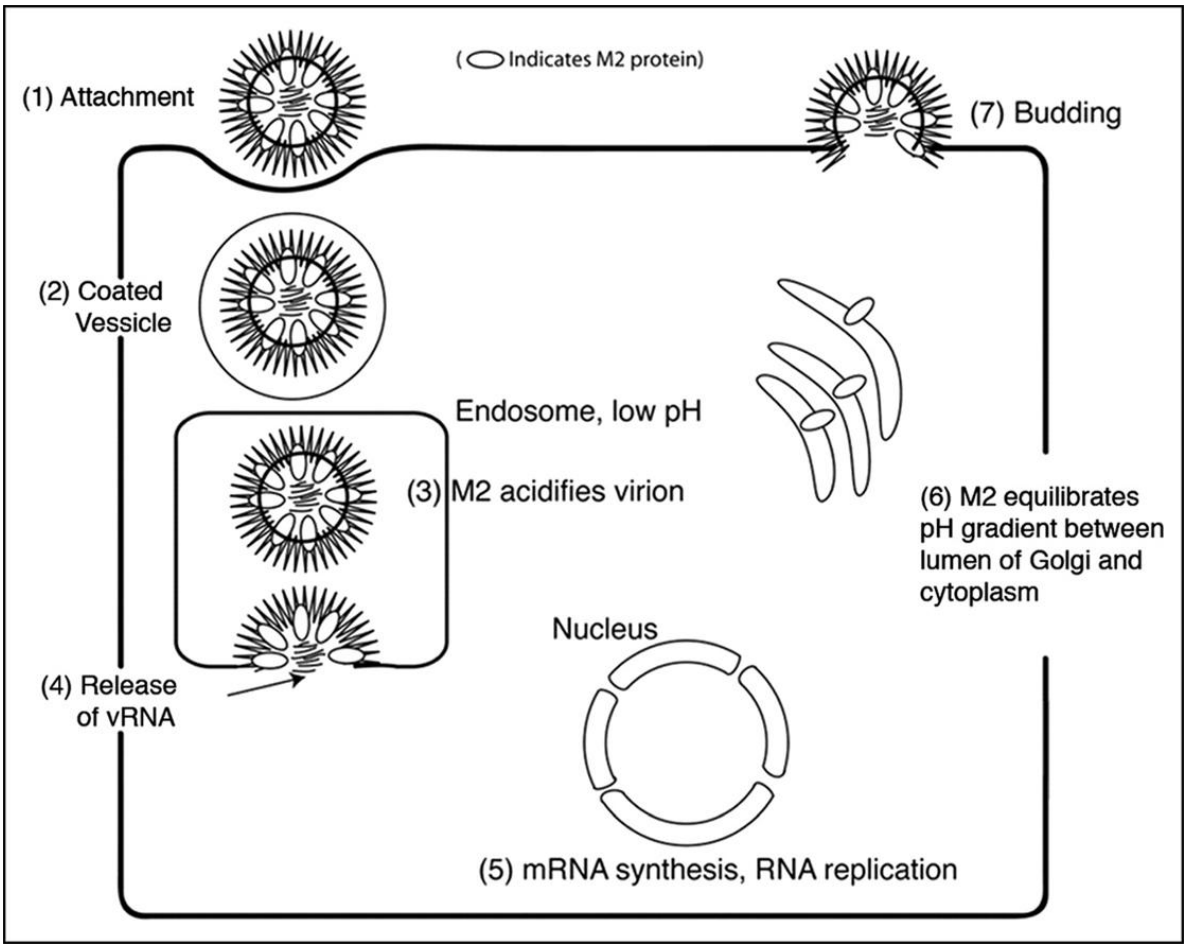

Figure 2. Structure of the Influenza A virus and genome. The eight segments are shown on the bottom of the figure, viral structure is upper right and a guide to the encoded proteins is shown on the left-hand side. From McHardy and Adams (2009), reprinted under Creative Commons attribution license.

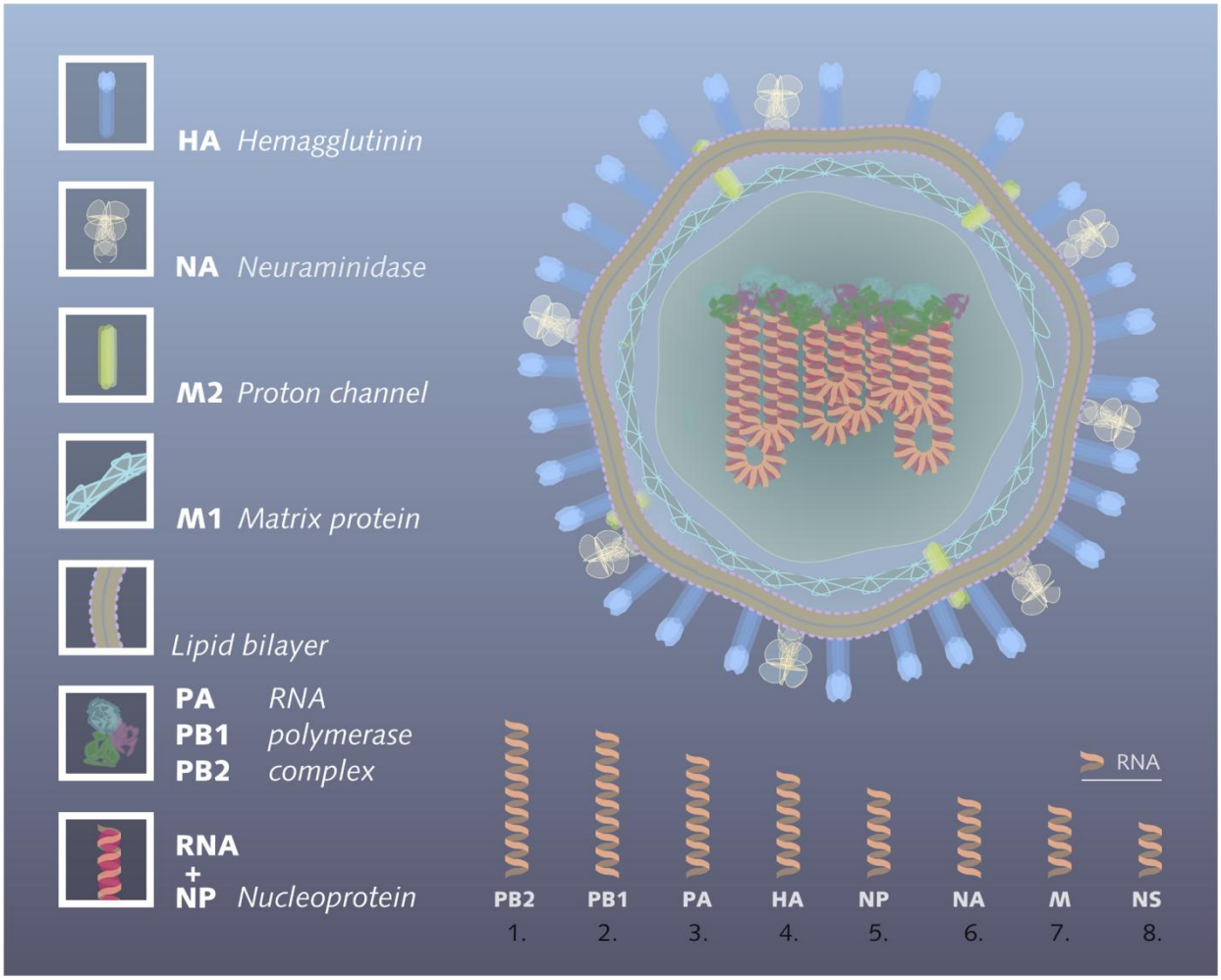

Figure 3. Reassortment history of H1N1pdm as inferred from phylogenetic analysis of genetic sequence data from many strains. Colored horizontal boxes indicate carriage in avian, swine, and human hosts; individual tracks show the movement of different genome segments across hosts through time. From Smith et al. (2009) and http://tree.bio.ed.ac.uk/groups/Influenza/wiki/aea97/Phylogenetic_analysis_and_reassortment.html, reprinted under Creative Commons attribution – no commercial derivatives licence.

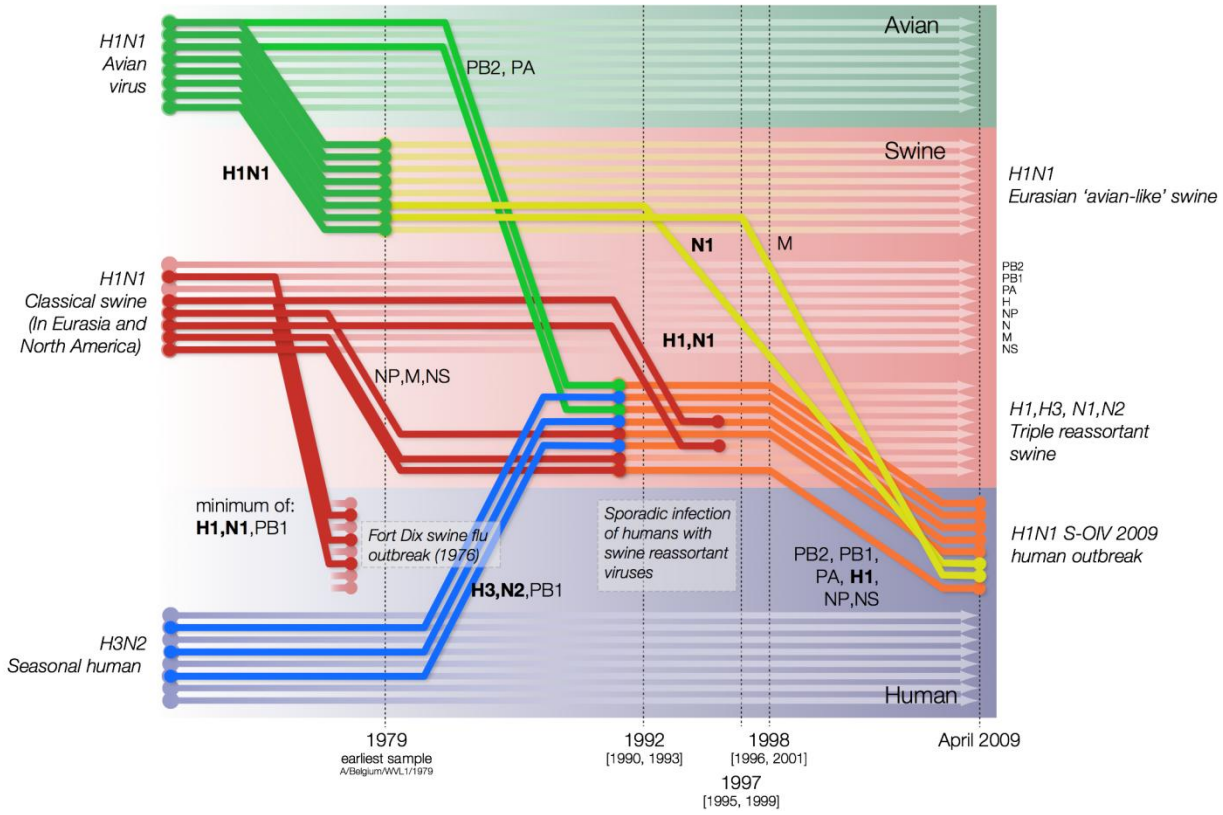

Figure 4. Pandemic strain hemagglutinin (HA) sequences deposited in the NCBI IVR as of July 2009, grouped by country of origin.

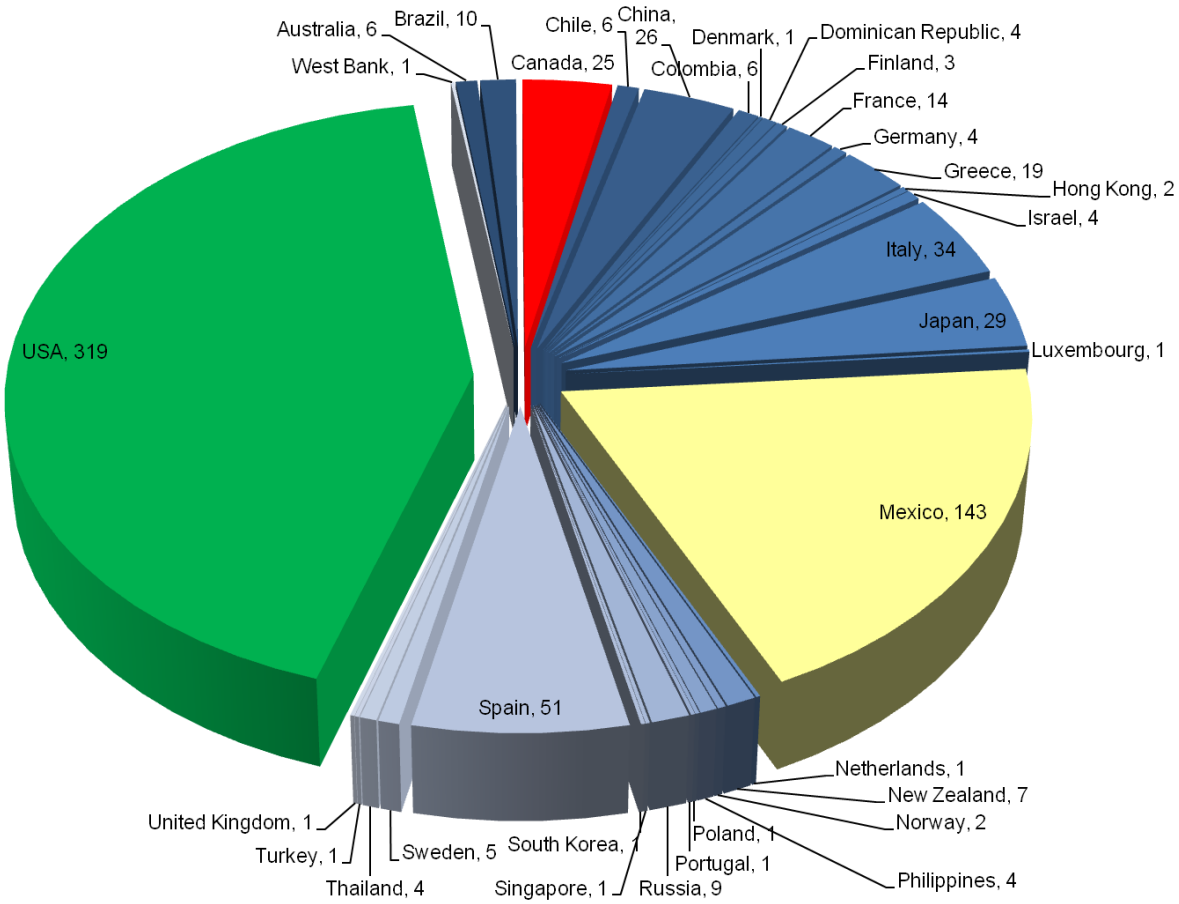

Figure 5. Screenshots from a GenGIS animation showing the number of confirmed 2009 H1N1 Influenza cases at two time points. (A) April 22. (B) May 1.

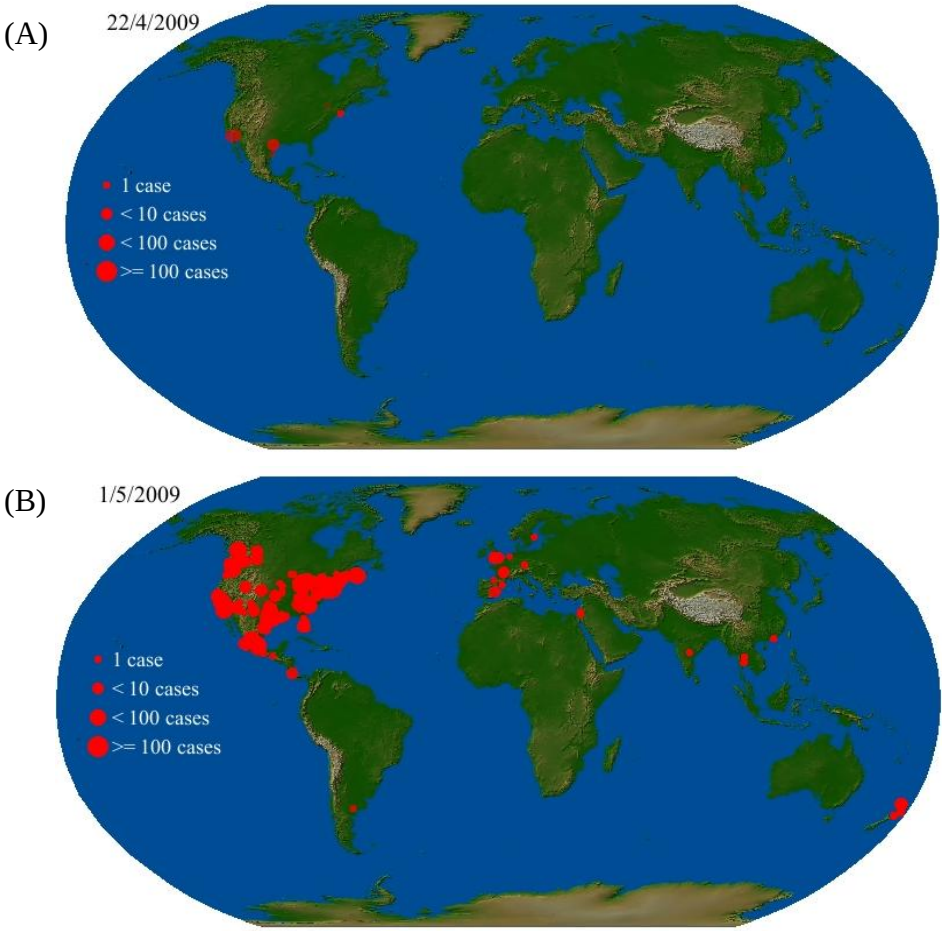

Figure 6. Screenshots from a GenGIS animation showing the spread and distribution of two sequence variants (based on the amino acid present at position 248 of the NA protein) of 2009 pandemic Influenza A. (A) April 18, all sequences are of the N type, indicating the amino acid asparagine. (B) April 19, showing the first occurrence of the variant D type, with aspartic acid at position 248. (C) April 28, more sequences of both types in North America, with the first European sequence of the N type. (D) June 4, widespread distribution of both sequence types.

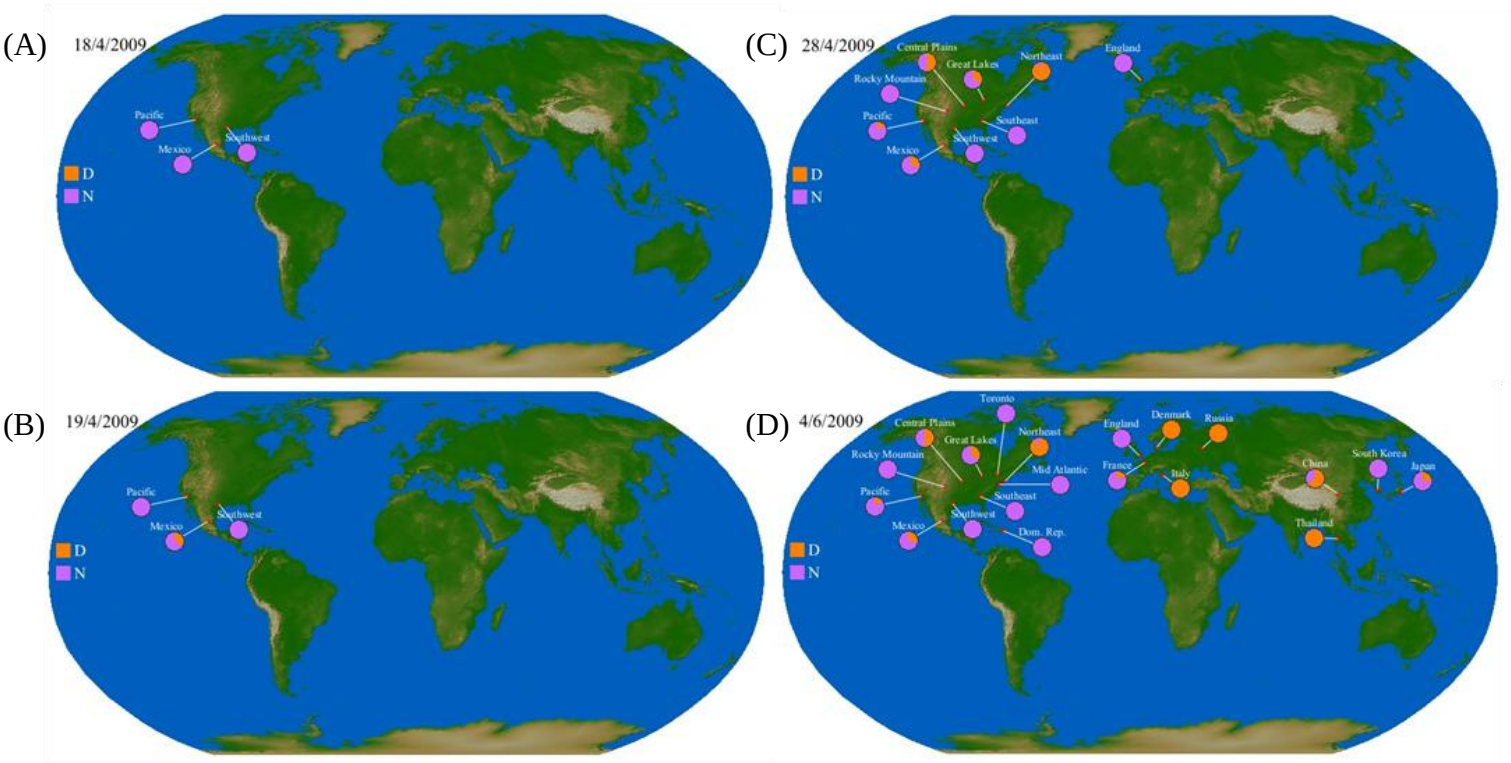

Figure 7. Phylogenetic tree in GenGIS, showing a lack of perfect association between geographic location and relatedness of isolates. Colors of location points and tree branches indicate continent of origin, with orange = North America, purple = Europe, and green = Asia. GenGIS attempts to align the tree at the bottom with the ordering of geographic points along the dashed gray lines; crossings between the two lines indicate disagreement between geography and the evolutionary history of the isolates. The implied common ancestor of the observed isolates is represented by the white dot at the bottom of the tree. White branches indicate are ancestors of isolates present on > 1 continent.

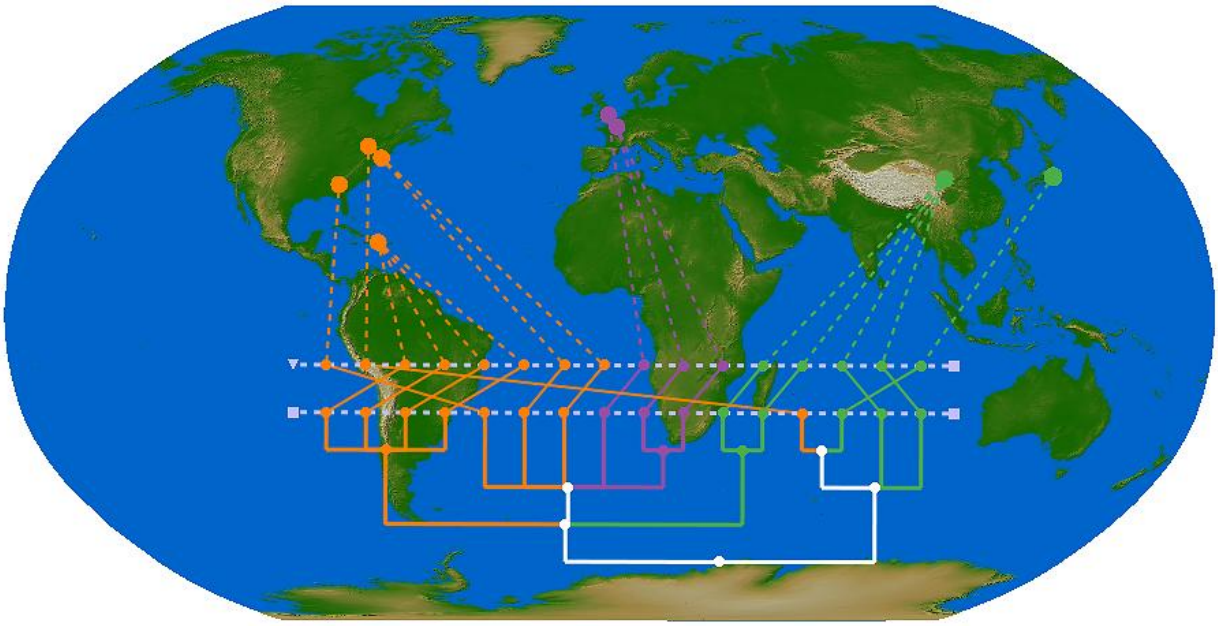

Figure 8. Distribution of 2009 pandemic Influenza A sequences showing laboratory-verified zanamivir resistance. Details of each isolate are given at the bottom.

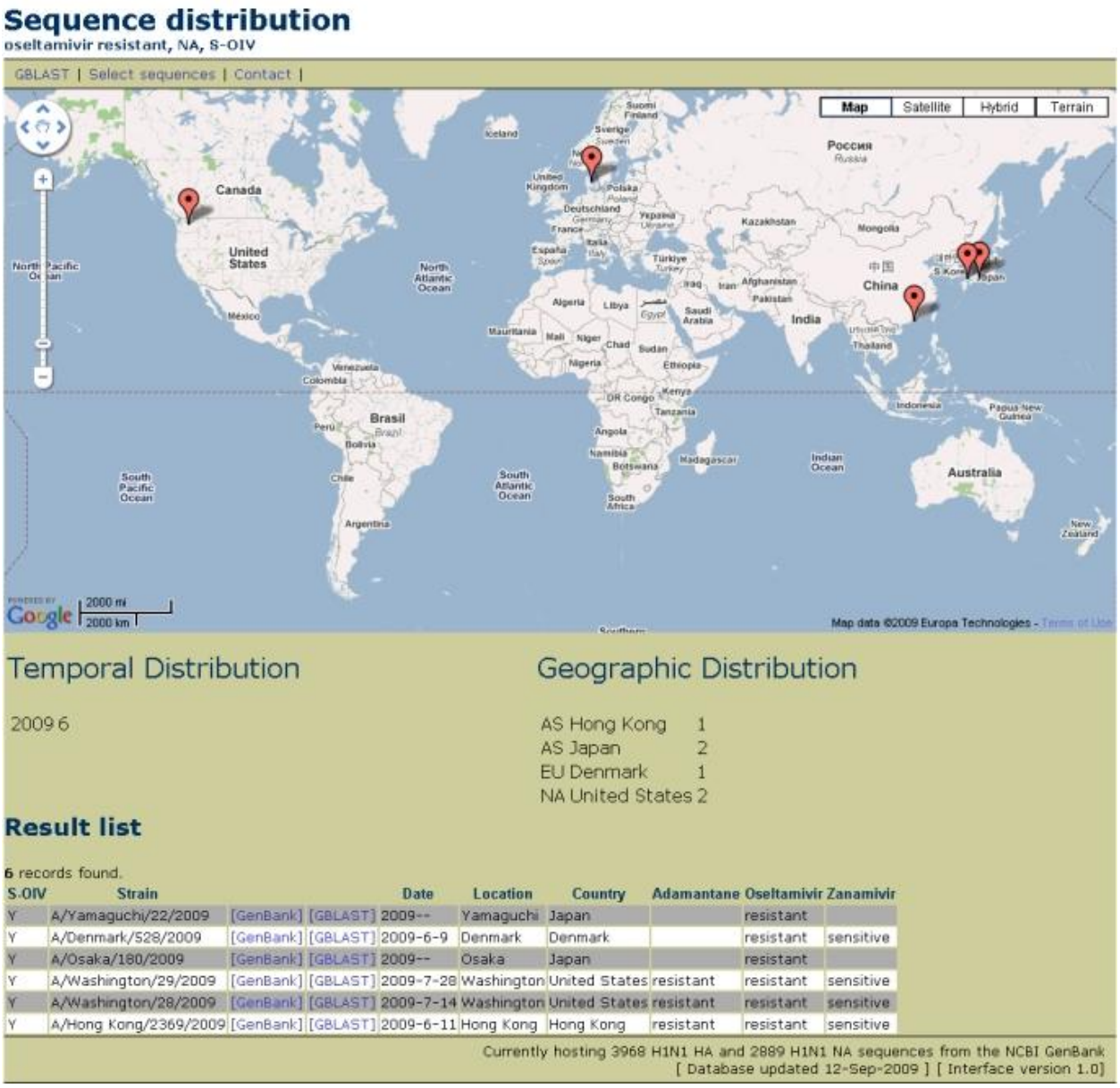

Figure 9. Modified Google Maps interface showing the amino acid sequence of a query protein, compared against two different groups of Influenza A sequences, indicated by the two tracks of coloured lines in the figure. Bottom line = all zanamivir-resistant sequences, top line = all seasonal and pandemic sequences in the reference database. Each letter indicates a single amino acid in the query protein sequence. No colored line above a letter in a given track indicates that 100% of sequences in the corresponding reference group have the same amino acid. A green line indicates that 50% or more of the sequences in the reference set have the same amino acid, while yellow indicates presence in < 50% of reference sequences. A red bar indicates that a given amino acid is not observed at all in the reference set.