The evolution of epidemic influenza

Martha I. Nelson and Edward C. Holmes*‡*

Abstract | Recent developments in complete-genome sequencing, antigenic mapping and epidemiological modelling are greatly improving our knowledge of the evolution of human influenza virus at the epidemiological scale. In particular, recent studies have revealed a more complex relationship between antigenic evolution, natural selection and reassortment than previously realized. Despite these advances, there is much that remains to be understood about the epidemiology of influenza virus, particularly the processes that determine the virus's strong seasonality. We argue that a complete understanding of the evolutionary biology of this important human pathogen will require a genomic view of genetic diversity, including the acquisition of polymorphism data from within individual hosts and from geographical regions, particularly the tropics, which have been poorly surveyed to date.

Pandemic
An epidemic that occurs over a large geographical area, including multiple countries.

Epidemic
The occurrence of more cases than expected of an infectious disease, in a defined geographical area over a defined time period.

Reassortment
A form of recombination in which two (or more) influenza viruses, of the same or different subtypes, co-infect a single cell and exchange RNA segments to form genetically novel viruses.

**Center for Infectious Disease Dynamics, Department of Biology, The Pennsylvania State University, University Park, Pennsylvania 16802, USA.*
‡Fogarty International Center, National Institutes of Health, Bethesda, Maryland 20892, USA.
Correspondence to E.C.H.
e-mail: ech15@psu.edu
doi:10.1038/nrg2053
Published online
30 January 2007

Influenza is one of the most important infectious diseases of humans. The annual mortality that is caused by influenza in the United States alone is estimated at over 36,000 (REF. 1) (FIG. 1), whereas occasional global pandemics can infect 20–40% of the population in a single year[2]. As a notorious case in point, the pandemic of 1918–1919 caused possibly 20–50 million deaths on a global scale, making it the single most devastating disease outbreak in human history[3]. The recent uncertainty over whether H5N1 avian influenza virus will adapt to human transmission, and how its spread might be controlled[4–7], highlight the threat that is posed by influenza and the need to understand its evolutionary dynamics.

Influenza viruses are single-stranded, negative-sense RNA viruses of the family *Orthomyxoviridae* that cause regular seasonal epidemics in humans, other mammalian species and birds. Three phylogenetically and antigenically distinct viral types — A, B and C — circulate globally in human populations, although type A viruses exhibit the greatest genetic diversity, infect the widest range of host species and cause the vast majority of severe disease in humans, including the great pandemics. The genome of influenza A virus (total length ~13 kb) is composed of eight segments that can be exchanged through reassortment (FIG. 2). Wild waterfowl are the reservoir hosts for type A influenza viruses, harbouring numerous antigenically distinct subtypes (serotypes) of the two main viral antigens, the haemagglutinin (HA) and neuraminidase (NA) surface glycoproteins (16 HA and 9 NA subtypes)[8,9]. These avian viruses occasionally transmit to other species, in which they either cause isolated outbreaks with little or no onward transmission, as is currently the case with avian H5N1 influenza in humans; less frequently, they become established in new hosts, resulting in (irregular) major human pandemics.

Here we review our current understanding of the evolutionary biology of human influenza A virus, showing how recent advances, particularly in comparative genomics and epidemiology, have shed new light on this important pathogen. We focus on the patterns and processes of influenza virus evolution at the level of recurrent human epidemics, highlighting areas in which future research might prove to be particularly profitable. For details of the biology of avian influenza virus and how it manifests as large-scale human outbreaks, see REFS 10–12.

The determinants of influenza virus evolution

The phylodynamics of antigenic drift. Owing to the large amount of available sequence data, particularly from the HA1 domain of the HA protein, many studies have explored the evolutionary processes that shape the genetic diversity of influenza A virus. Investigating the complex interplay between natural selection, phylogeny and epidemiology is key to understanding influenza A virus evolution[13,14]. Because the human immune response to viral infection is not completely cross-protective, natural selection favours amino-acid variants of the HA and NA proteins that allow the virus to evade immunity, infect more hosts and proliferate[15]. This continual change in antigenic structure through time is called antigenic drift[16]. Although both the HA and NA proteins contain antigenic sites in which immune-driven

natural selection can occur, the HA1 domain of the HA protein contains the highest concentration of epitopes and, correspondingly, experiences the most intense positive selection pressure[15,17–22].

At the phylogenetic scale, the continual selective turnover of amino-acid variants is thought to produce the distinctive 'cactus-like' phylogenetic tree of the HA1 domain from A/H3N2 subtype viruses[13,15,17]. A single main trunk lineage depicts the pathway of advantageous mutations that have been fixed by natural selection through time, from past to present, whereas short side branches that stem from this trunk represent those isolates that die out because they were insufficiently antigenically distinct to evade immunity. The apparent regularity of this phylogenetic pattern has generated much interest, because of the potential to predict the future course of viral evolution and, in doing so, aid vaccine strain selection[23]. Likewise, there is still considerable debate over what aspects of influenza epidemiology so strongly favour the survival of a single HA1 trunk lineage in human A/H3N2 viruses, whereas multiple lineages seem to co-circulate more frequently within populations of equine H3N8 (REF. 24), human H1N1[13] and influenza viruses types B[25,26] and C[27] (in which the equivalent haemagglutininesterase protein is termed HEF).

Although antigenic changes in the haemagglutinin protein are clearly important determinants of viral fitness, the 'progressive' model of influenza A evolution, as typified by the cactus-like phylogeny, was formed on the basis of studies that largely focused on HA1 in isolation, considered relatively few sequences from individual time points and geographical locations, and often targeted strains with unusual antigenic properties in the interests of vaccine design. Indeed, the antigenic evolution of HA1 seems to be more clustered than continuous[28]. Moreover, the recent explosion of large-scale genome sequence data from H3N2 viruses has shown that the evolutionary pattern that is observed in the HA1 domain does not always apply to the rest of the viral genome[29,30]. In contrast to the restricted number of lineages that can be observed at any time point in HA1, whole-genome phylogenies show the coexistence of multiple viral lineages, particularly on a limited spatial and temporal scale (FIG. 3). This indicates that the transition among antigenic types does not always proceed in a simple linear manner, that reassortment among coexisting lineages is relatively frequent (see below), and that, for these reasons, predicting the path of influenza virus evolution from sequence data alone will be inherently difficult.

Recent analyses also indicate that positive selection on the HA1 domain occurs in a punctuated manner[14,30,31]. Indeed, the cactus-like structure of the A/H3N2 phylogenetic tree is not in itself conclusive evidence for the action of adaptive evolution, as similar phylogenetic patterns can be generated through a combination of serially sampled (that is, time-structured) data and sequential random population bottlenecks, without strong positive selection. Therefore, the definitive signature of positive selection in the influenza A virus HA protein is not merely the presence of a single trunk lineage, but rather that this trunk is defined by an

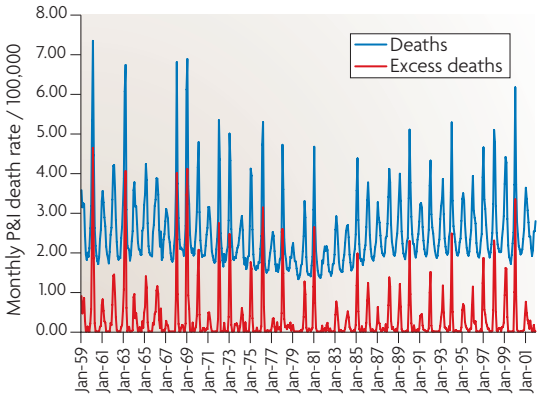

Figure 1 | **The periodicity of pneumonia and influenza mortality and excess mortality rates.** Monthly pneumonia and influenza (P&I) death rates and excess death rates (above the baseline mortality due to other respiratory pathogens) in the United States from 1959 to 2001 are shown (see the web site for the US National Center for Health Statistics). Peaks occur during the winter in northern latitudes at ~2–5 year intervals, usually during H3N2-dominant seasons, since the 1968 pandemic. See REF. 81 for more details.

Haemagglutinin
An influenza virus surface glycoprotein, denoted HA, which is responsible for viral binding and entry into host epithelial cells. Sixteen HA serotypes are present in animal species.

Neuraminidase
An influenza virus surface glycoprotein, denoted NA, which is involved in the budding (release) of new virions from infected cells. Nine NA serotypes are present in animal species.

Antigenic drift
The continual evasion of host immunity by the gradual accumulation of mutations in the haemagglutinin and neuraminidase surface glycoproteins of influenza A virus, changing its antigenic structure.

Epitope
A small sequence of a viral protein that is recognized by either the cellular or humoral arms of the immune system, and therefore frequently undergoes the strongest adaptive selection to rapidly evolve immune-escape mutants.

Population bottleneck
A marked reduction in population size followed by the survival and expansion of a small sample of the original population.

increased frequency of non-synonymous substitutions, which reflects the continual fixation of (advantageous) amino-acid replacements (BOX 1). Similarly, many of the mutations that fall on the side branches of the HA1 tree are likely to be deleterious, and will not achieve fixation even in the absence of immune selection. However, it is important to note that because the computational tools that measure the extent of positive selection are inherently conservative, and quantify the successive fixation of non-synonymous mutations at specific amino-acid sites, adaptive evolution is likely to occur more frequently than is usually detected (BOX 1). In the future, methods that account for the rate of amino-acid fixation[32] (as opposed to simply considering the total number of fixation events) might offer more analytical power.

Although antigenic drift is undoubtedly an important aspect of influenza A virus evolution, as reflected in the changing antigenic profiles[33] and the need for continually updated vaccines, recent data indicate that this process does not occur within the time frame of a single epidemic season in a single locality; few amino-acid changes are fixed in HA1 within populations at the seasonal scale[34]. Consequently, key questions for future research will address the evolutionary and epidemiological processes that drive antigenic drift and the timescale on which this process occurs. To answer these questions will clearly require a far larger sample of influenza virus genomes, with greater resolution in both time and space.

Antigenic maps and cluster jumps. The episodic nature of the antigenic evolution of HA1 has been vividly documented in antigenic maps, one of the most important innovations in studies of viral evolution[33]. Antigenic mapping involves constructing a matrix of haemagglutinin inhibition assay distances among viral isolates (see BOX 2 for more information)

and then plotting these to produce a cartographic surface, analogous to a standard geographical map. This approach provides an important insight into evolutionary dynamics because it allows a direct comparison between changes in viral genotype (reflected by the HA amino-acid change) and an inferred phenotype (reflected by the haemagglutinin inhibition distance), although its relationship to measures of overall viral fitness is less clear. Maps of HA from A/H3N2 isolates, which have been sampled since the first appearance of this subtype in 1968, show that major jumps between antigenically distinct clusters of viral sequences occur with a periodicity of roughly 3 years[33]. Although these antigenic 'cluster jumps' are usually also apparent as long branches on HA1 phylogenetic trees, small genetic changes sometimes have a strong effect on antigenicity. Furthermore, as the cluster jumps tend to correspond to occurrences of vaccine failure[35], they evidently represent a better predictor of antigenic novelty than do data from studies of genotypic evolution alone. Although it is clear that our understanding of influenza A virus evolution will greatly benefit from a better understanding of the rules that govern antigenic evolution, as manifest in the path the virus takes across the cartographic surface, determining the epidemiological processes that underlie the periodicity of antigenic evolution will undoubtedly be more complex[36].

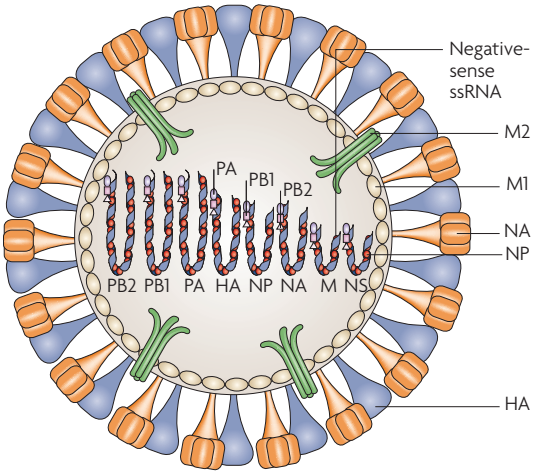

Figure 2 | **The structure of influenza A virus.** The genome of influenza A virus is composed of eight genomic segments that, by convention, are listed from largest to smallest, although their true arrangement within the spherical virion is unknown. Each segment contains a coding region that encodes one or two proteins, as well as short 5′ and 3′ flanking sequences. Three segments encode proteins that form the virus polymerase complex: basic polymerase 2 (PB2; 2,277 nucleotides in the protein-coding region, in segment 1), which controls the recognition of host-cell RNA; basic polymerase 1 (PB1; 2,271 nucleotides, segment 2), which catalyses nucleotide addition (and which also encodes a small pro-apoptotic mitochondrial protein that is translated in a different reading frame — PB1-F2); and the acidic protein (PA; 2,148 nucleotides, segment 3), which might possess a transcriptase protease activity. Two segments encode surface envelope glycoproteins that function as viral antigens: haemagglutinin (HA; 1,698 nucleotides, segment 4), which is responsible for binding to sialic-acid receptors and entry into host cells, and which is divided into two domains (or subunits) — HA1 and HA2 — and neuraminidase (NA; 1,407 nucleotides, segment 6), which is involved in budding of new virions from infected cells. A single segment encodes a nucleoprotein (NP; 1,494 nucleotides, segment 5), which binds to the viral RNA. The seventh segment encodes two proteins that share a short overlapping region: the matrix protein M1 (756 nucleotides) encodes the main component of the viral capsid, and M2 (291 nucleotides), which is an integral membrane protein, functions as an ion channel. Segment 8, the smallest segment of the viral genome, encodes a non-structural protein, NS1 (690 nucleotides), which affects cellular RNA transport, splicing and translation. Also encoded on segment 8 in an overlapping reading frame is the NS2 protein (363 nucleotides), a minor component of the virion, the function of which is currently unknown. A mature virion of influenza A virus is composed of the nucleocapsid, a surrounding layer of M1, and the membrane envelope, which contains the HA, NA and M2 proteins. For more details on the life cycle and replication of influenza virus see REFS 88,89. Figure reproduced with permission from *Nature Reviews Microbiology* REF. 11 © (2005) Macmillan Publishers Ltd.

Reassortment in influenza virus evolution. Severe influenza pandemics can occur following a sudden antigenic shift — when a reassortment event generates a novel combination of HA and NA antigens to which the population is immunologically naive. The segmented genome of the influenza virus facilitates reassortment between isolates that co-infect the same host cell. Reassortment among HA and NA subtypes was fundamental in the human pandemics of 1957 (H2N2 subtype) and 1968 (H3N2 subtype), which also acquired a new basic polymerase 1 (PB1) segment[37]. The origin of the H1N1 strain that caused the severe pandemic of 1918, and whether it jumped to humans directly from an avian reservoir population or first circulated in another mammalian host such as swine, is less clear and the source of much debate[38–40].

The evolutionary importance of reassortment in recurrent influenza epidemics is also uncertain. Reassortment events can be detected when sequences of different segments from the same isolate occupy incongruent positions on phylogenetic trees. Until recently, it was usually only reassortment involving HA and NA that could be detected in this manner, because the vast majority of publicly available influenza virus sequences comprised just these two proteins. However, the expansion of genome sequence data sets has shown that reassortment can also occur among internal segments, and among human strains of the same subtype[29,30,41–44]. As a case in point, a detailed phylogenetic analysis of 413 complete viral genomes from New York State, USA, sampled over a 7-year period, revealed 14 reassortment events that were identified on the basis of incongruent phylogenetic trees of HA, NA and concatenated internal proteins[30]. Even this result is likely to represent a significant underestimate of the true frequency of reassortment. Some reassortment events are undetectable by phylogenetic analysis because they do not lead to major differences in tree topology, they

Antigenic shift
The formation of a new influenza virus subtype with a novel combination of haemagglutinin and neuraminidase segments, which are derived from two different parental influenza strains, that combined through genomic reassortment.

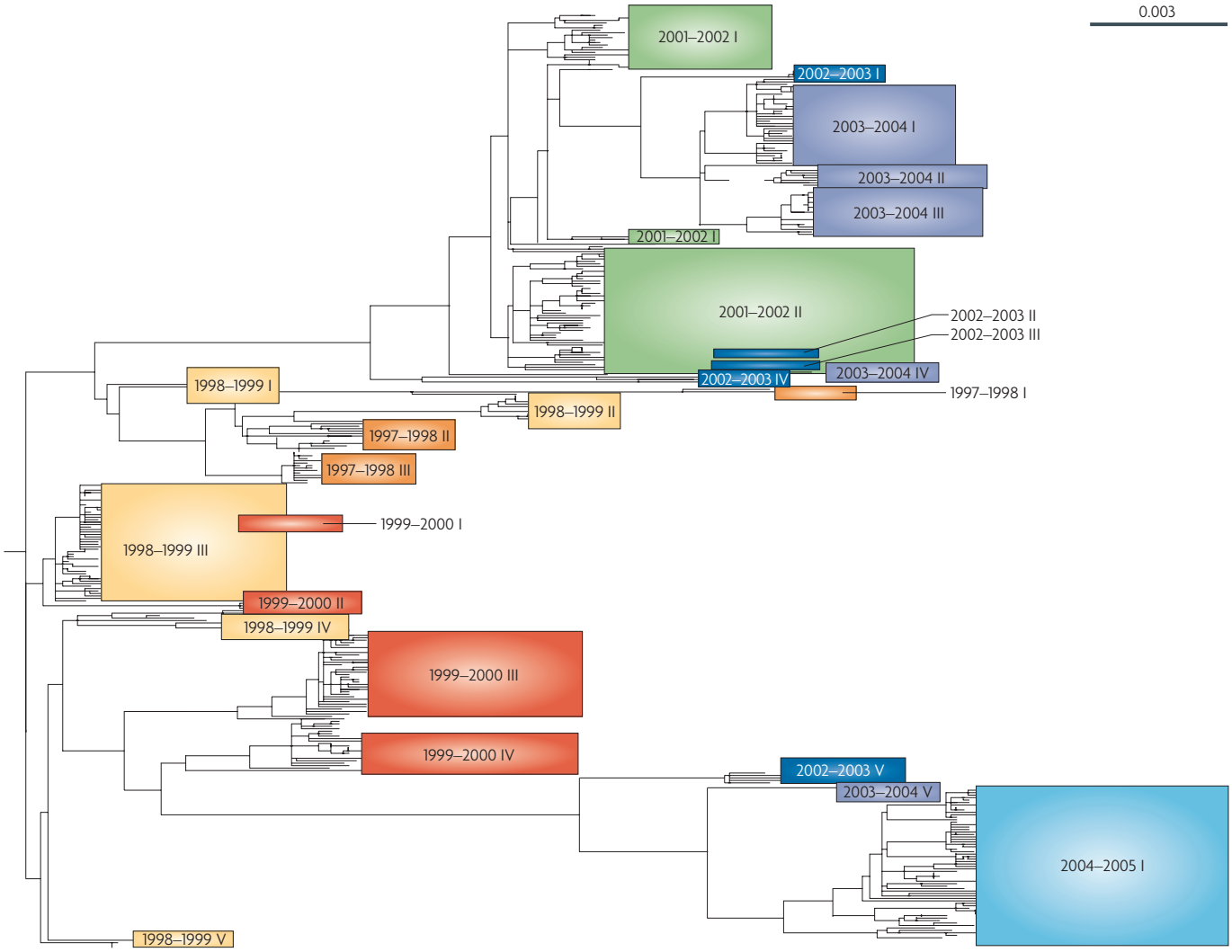

Figure 3 | **Phylogenetic relationships of concatenated internal proteins.** All segments excluding haemagglutinin (HA) and neuraminidase (NA) of A/H3N2 viruses were sampled from New York State, USA, from 1997 to 2005. Rectangles represent distinct clusters of isolates; the size of the rectangle reflects the number of isolates in the lineage. Roman numerals denote distinct viral lineages that were circulating within each influenza season, with seasons coloured individually. The tree is mid-point rooted for purposes of clarity only, and all horizontal branch lengths are drawn to a scale of substitutions per site (as shown by the scale bar). The phylogeny shows the genetic diversity of influenza A virus in a single locality, including the co-circulation of multiple viral lineages. Adapted from REF. 30.

involve 'parental' isolates that have not been sampled, or they result in unfit progeny. It is therefore imperative that more sophisticated methods are developed to estimate both the rate of reassortment, particularly relative to the rate of mutation for each nucleotide, and the background phylogenetic history of the viral genome in the face of such frequent reassortment[45]. Such methods will evidently require the ability to more precisely determine which base changes are the result of mutation and which are due to reassortment[46].

Reassortment might also have an important role in generating evolutionary novelty. For example, reassortment events occurred concurrently with two recent antigenic cluster jumps: the WU95 (Wuhan 1995 strain of H3N2 influenza A virus) to SY97 (Sydney 1997 strain of H3N2 influenza A virus) cluster jump and the SY97 to FU02 (Fujian 2002 strain of H3N2 influenza A virus) cluster jump[29,31,41]. However, determining cause and effect between a reassortment and a cluster jump is inherently difficult. In the case of the SY97 to FU02 antigenic cluster jump, for which a greater number of samples has allowed better resolution, a reassortment event that involved the HA segment coincided with the emergence of a single dominant lineage with a new HI type from an array of co-circulating lineages[29,41]. The evolutionary puzzle is why a clearly fit virus — that is, the virus belonging to the FU02 antigenic cluster — did not immediately rise to fixation, but instead circulated at low frequency for several years. One hypothesis is that the intrinsic fitness of the HA segment could not be

Box 1 | **Measuring selection pressures in influenza virus**

As with many studies of viral evolution, selection pressures in influenza A virus are often quantified as the ratio of non-synonymous (d_N) to synonymous (d_S) substitutions per nucleotide site, with $d_N > d_S$ being indicative of positive selection and $d_N < d_S$ being indicative of purifying selection[19]. In the case of the haemagglutinin HA1 domain, both pairwise and more sophisticated site-specific analyses of d_N/d_S have found compelling evidence for positive selection, particularly in epitope regions. However, analyses of d_N/d_S are inherently conservative in that they require recurrent mutations at the same codon to demonstrate positive selection; any advantageous mutations that occur only once on a single lineage will not be detected. Analysing the phylogenetic distribution of mutations can therefore assist in the detection of positive selection. The higher the fitness of a mutation, the deeper it will fall on a phylogenetic tree, with fixed mutations defining the main clades. Similarly, the shorter the time taken to achieve fixation, the more likely it is that this process was driven by natural selection rather than by genetic drift. By contrast, mutations that fall on terminal branches are more likely to represent transient deleterious mutations, although the possible adaptation to cell culture should also be considered, as this will also generate an excess of non-synonymous mutations on terminal branches[17].

realized until it was placed in a compatible genetic background, a process that was achieved by reassortment. If subsequent compensatory changes are required to increase compatibility within[47] and among segments[48], such reassortment events might also entail a burst of adaptive changes across the viral genome (FIG. 4). If correct, this means that the evolution of influenza A virus is more complex than previously realized, and that the evolutionary dynamics of the HA segment must be considered within the context of evolution at the genomic level. Revealing the evolutionary dynamics and fitness contributions of the other viral proteins during epidemic evolution[49], and their epistatic interactions[14], therefore represents an important direction for future research.

The rate of reassortment in influenza A virus also provides insights into the extent of immunological cross-protection, which in turn might have implications for vaccine design. Reassortment among isolates that are assigned to different antigenic types necessarily means that they must co-infect a single cell, implying that protection is not complete at this level of antigenic difference. However, as studies of the intra-host genetic diversity in influenza virus have not been widely undertaken (see below), it remains unknown whether multiple genetically or antigenically distinct lineages co-circulate within individual hosts. Finally, the study of reassortment patterns might also provide important clues to the linkage of genomic segments, as it is expected that closely linked segments will be subject to less frequent reassortment. Analyses of this sort represent a key task for future evolutionary genomics in influenza virus.

A related, although far more controversial, issue is that of intrasegment RNA recombination. Although there is ample evidence that influenza viruses undergo various forms of non-homologous recombination, albeit rarely[50,51], the occurrence of homologous recombination within segments is far from proven. Some comparative studies indicate that complex patterns of genetic diversity might be a footprint of past recombination, although the evidence is not conclusive[52]. The most compelling evolutionary evidence for recombination — the occurrence of incongruent phylogenetic trees — is generally lacking, and previous suggestions of incongruence, for example, in the emergence of 1918 influenza A virus[53], are more likely to be due to differences in substitution rates between the HA1 and HA2 domains[54]. Indeed, low rates of RNA recombination seem to characterize negative-sense RNA viruses in general[52].

Rates of evolutionary change in influenza virus

Accurate estimations of evolutionary rates at both nucleotide and amino-acid levels are central to resolving many long-standing questions about the evolution of influenza virus, including the relative roles of natural selection versus genetic drift, the origins of the 1918 H1N1 pandemic virus, and the ecology of the virus in its avian reservoir. The available literature provides inconsistent reports of evolutionary rates, largely because of differences in methodology, as well as the number and the epidemiological significance of the virus samples that have been used for analysis. Improved analytical methods and greater availability of whole-genome sequence data should enable large-scale systematic comparisons of evolutionary rates of entire genomes from multiple subtypes in numerous host species.

These limitations notwithstanding, the main determinant of variation in substitution rates among influenza viruses seems to be the strength of immune selection pressure; background mutation rates are generally similar among RNA viruses, at approximately one mutation at each genome replication[55], which translates into long-term substitution rates of 10^{-3}–10^{-4} nucleotide

Box 2 | **Antigenic mapping and antigenic cluster jumps**

One of the notable features of influenza A virus is that it is possible to compare patterns of genetic and antigenic change through time, thereby providing a tentative association between genotype and phenotype. Ongoing antigenic change can be measured using data obtained from a haemagglutinin inhibition assay, which assesses the ability of influenza viruses to agglutinate red blood cells, and the corresponding ability of ferret antisera that has been raised against a set of viral isolates to inhibit agglutination. To produce an antigenic map, a point is assigned to each antigen and antiserum reaction and subjected to a form of multidimensional scaling. Application of antigenic mapping to longitudinally sampled isolates of A/H3N2 showed that viral isolates tended to fall into discrete 'clusters' that were defined by amino-acid differences at known (and often positively selected) antigenic sites, and that each cluster remained dominant for approximately 3 years[33]. Antigenic evolution was also found to be more punctuated than genetic evolution (at both nucleotide and amino-acid levels), as measured by phylogenetic analysis, although the fixed amino-acid changes that correspond to antigenic cluster jumps also seem to occur at punctuated time intervals[14,31].

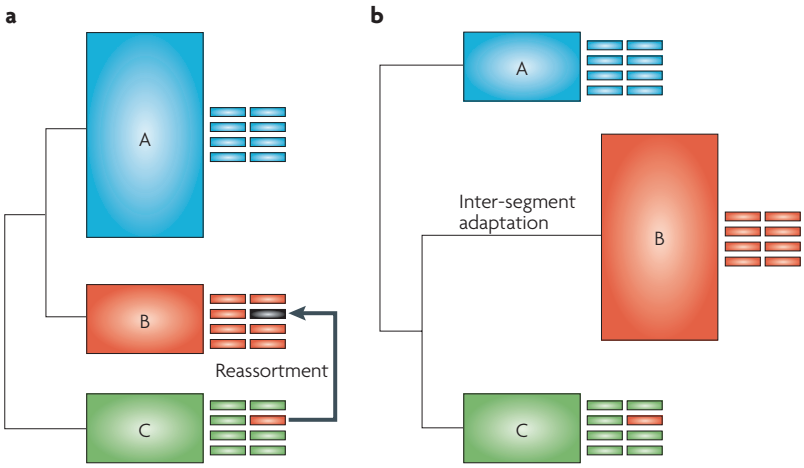

Figure 4 | **A model for the genome-wide evolution of human influenza A/H3N2 virus. a** | Three lineages of influenza A virus — A, B and C — co-circulate in a population. Lineage A has the highest fitness and is therefore dominant. Each lineage has a unique configuration of genomic segments (represented by small rectangles). Whereas the segments in lineage A are internally compatible, lineage B contains a haemagglutinin (HA) segment of low fitness (shown as a black rectangle), which reduces the overall fitness of this lineage. By contrast, lineage C has a high fitness HA (shown as a red rectangle), which is contained within a lower-fitness genomic background. A reassortment event results in the transfer of the high fitness HA from lineage C into the more compatible genetic background of lineage B. **b** | Following the reassortment event, lineage B undergoes a burst of (compensatory) adaptive evolution across its genome, increasing its rate of divergence and fitness so that it becomes the new dominant lineage in the population.

substitutions per site per year[56,57]. Such selection pressures generally reflect the length of time that an influenza virus subtype has been associated with a particular host species. So, 'older' influenza A viruses evolve more slowly (at non-synonymous sites) in the reservoir avian species with which they might have co-adapted, whereas newly emergent viruses, such as A/H3N2 and A/H5N1 in humans and domestic poultry, evolve more rapidly (through positive selection) to evade host immunity and achieve efficient transmission in new host species.

Studies of evolutionary rates of several influenza virus proteins generally support this model: the lowest rates of non-synonymous substitution have been reported in influenza viruses that were sampled from wild aquatic bird species[8,58], the highest rates in viruses that caused human H3N2 epidemics and outbreaks in poultry and swine[59,60], and intermediate rates in older human subtypes (such as H1N1), in type B viruses[61,62] and in internal proteins[63]. However, the hypothesis that avian influenza A viruses have reached an adaptive equilibrium ('evolutionary stasis') following a long-term co-adaptation with wild aquatic bird species[8], could be misleading. Although rates of amino-acid change in influenza A viruses are undoubtedly lower in wild aquatic birds compared with humans, overall rates of nucleotide substitution are not significantly lower than those of most RNA viruses[58,64]. As waterfowl and shorebirds seem to mount only a weak immune response to influenza A virus, it is likely that most amino-acid changes are deleterious and are purged through purifying selection, as shown by the tendency for non-synonymous mutations to fall on terminal branches of phylogenetic trees[65]. By contrast, higher rates of non-synonymous substitution in the HA segment have been observed in some domestic poultry species[58], despite the fact that unvaccinated poultry do not typically mount a strong immune response[59]. These observations indicate that the selection pressures to adapt to a new host can lead to rapid evolution, even in the absence of immune selection, and confirm the importance of whole-genome analysis.

To fully understand the factors that generate variation in the rate of evolutionary change in influenza virus and how these relate to disease emergence and severity, a comprehensive survey of evolutionary rates for all eight genome segments of the main viral subtypes in different host species is needed, using updated methods. Previous linear regression analyses that compared genetic distance against the year of isolation were inherently biased, as data points were non-independent, leading to over-sampling of certain phylogenetic branches. By contrast, maximum likelihood[66] and more recent Bayesian Markov chain Monte Carlo (MCMC)[67,68] approaches explicitly account for phylogenetic structure, time of sampling and rate variation among lineages. Bayesian MCMC methods also provide an indication of statistical uncertainty, because estimates are made on large numbers of sampled trees.

Evolutionary aspects of seasonality

The clock-like consistency of the winter incidence peaks of influenza virus represents one of the strongest examples of seasonality in infectious disease (FIG. 5). However, the reasons that human influenza epidemics arise and then peak at consistent 6-month intervals across temperate regions of the northern and southern hemispheres are unknown. Various theories have been proposed to explain how seasonal change might stimulate influenza activity: transmission rates might increase during school terms and winter crowding, the stability of the virus might be enhanced by cooler temperatures, or host immunity might decline during colder weather (reviewed in REF. 69). All of these hypotheses remain largely untested.

Experiments from half a century ago are still cited as evidence that influenza virus is most stable in cool, dry temperatures[70]. However, later work weakened this correlation by showing an increase in susceptibility in mice during winter, even when temperature and humidity were held constant[71]. Recent research in tropical regions has also shown a significant burden of disease in areas with warm, humid climates[72,73]. Therefore, cooler temperature alone does not explain influenza seasonality. Various aspects of human behaviour, such as time spent indoors or in school, have been implicated in changing transmission rates; the effect of seasonal climate change on immune function and host susceptibility has also been documented[74,75]. For example, antibody responses are believed to fluctuate with melatonin secretion during seasonal light and dark cycles, potentially increasing human susceptibility to influenza infection at certain times of the year[69].

Maximum likelihood
A statistical method that selects the phylogenetic tree with the highest probability of explaining the sequence data, under a specific model of substitution (changes in the nucleotide or amino-acid sequence).

Bayesian Markov chain Monte Carlo
(MCMC); Bayesian statistical inference allows the use of prior knowledge in assessing the probability of model parameters in the presence of new data. The prior distribution can strongly affect the posterior (the results). MCMC is a stochastic algorithm for drawing samples from a posterior distribution, therein providing an estimate of the distribution.

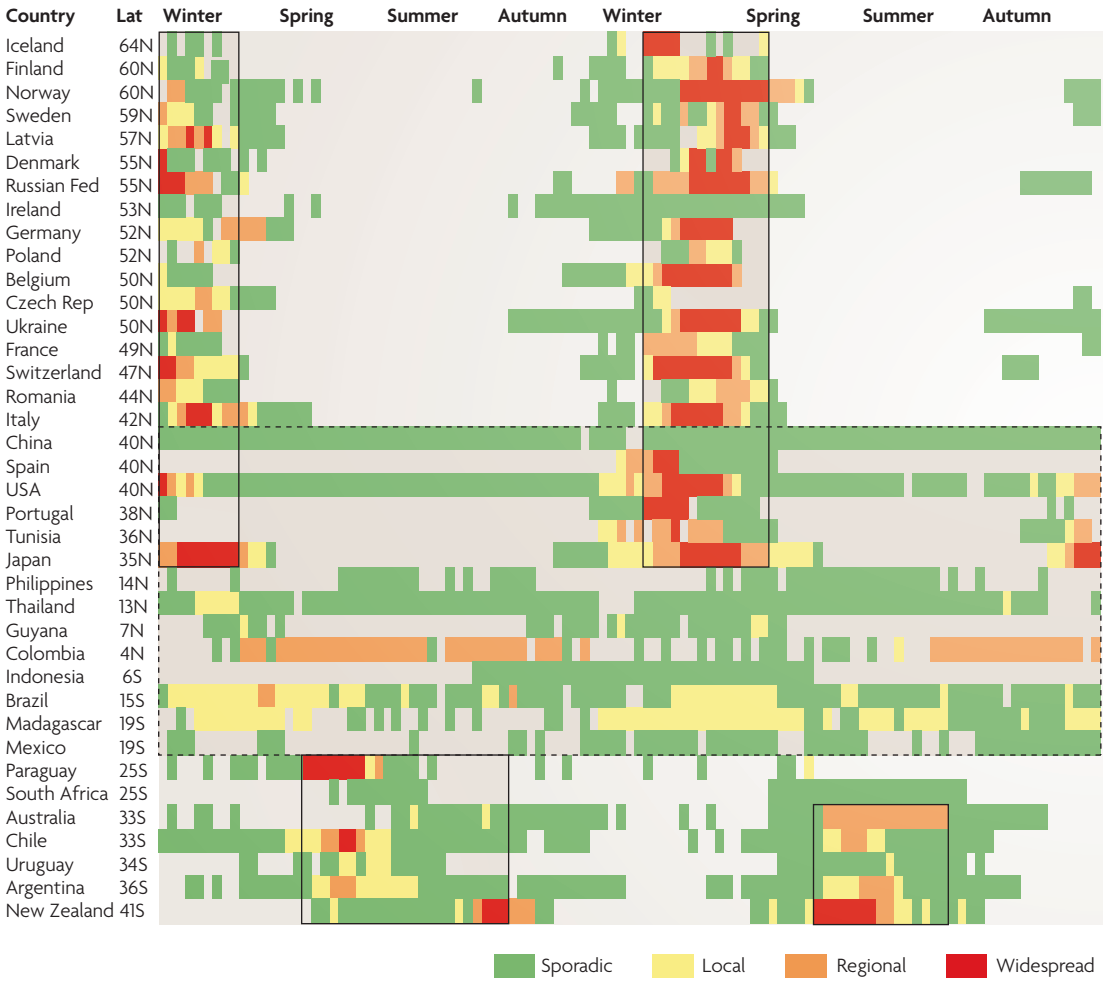

Figure 5 | **The seasonality of influenza virus.** Weekly reports of influenza-like illness (ILI) are classed as 'Sporadic', 'Local Outbreak', 'Regional Outbreak' or 'Widespread Outbreak' by the World Health Organization's (WHO) FluNet surveillance system. Influenza virus activity peaks at similar times in countries at similar latitude, during winter and early spring in the northern hemisphere and during late spring and summer in the southern hemisphere (depicted by the heavily outlined boxes). In countries that are closer to the equator, influenza virus activity is more consistent throughout the year (depicted by the dotted-line box), with dampened epidemic fluctuations, indicating that these areas could potentially serve as year-round reservoirs for the virus.

It has also been proposed that a deficiency of vitamin D during the winter months, which in turn might reduce the effectiveness of the innate immune system, could help to shape influenza seasonality[76].

Explaining this phenomenon has been impeded by gaps in our basic knowledge of the epidemiology of the influenza virus, particularly in the reservoir avian species, which offer a potential avenue for further investigation of seasonal dynamics. Epidemics in wild aquatic birds peak in late summer and early autumn in the northern hemisphere, when fledgling and congregation lead to an increase in population density[8,77]. The virus emerges 6–8 weeks later in domestic turkeys, although the causative factors of this time lag are still unknown[77]. Overall, the apparent importance of pre-migratory congregation and the birth of new susceptible hosts in triggering avian epidemics implies that the indirect effects of colder conditions on human social behaviour might in part drive influenza virus activity.

Another poorly understood aspect of seasonality is the spatial link between epidemics in the northern and southern hemispheres, and how tropical regions fit within this cross-hemispheric dynamic. For example, although epidemics in the northern hemisphere occur on a regular timescale, those that occur approximately 6 months later in the southern hemisphere exhibit weaker correlation with the epidemic timing of the northern hemisphere[78]. Epidemic periodicity is even more variable in tropical regions, where available data indicate that influenza often occurs year-round, although incidence peaks sometimes coincide with rainy seasons[79] (FIG. 5). Tropical regions might therefore serve as year-round reservoirs for influenza virus and clearly need to be surveyed more intensively[80].

One highly informative, albeit indirect, approach to studying seasonality has been to improve documentation of the overall spatio-temporal dynamics of the influenza virus. Analysis of influenza-related mortality data from

Gravity model
A methodology that extends Newtonian gravitational laws to models of behavioural patterns that mimic gravitational interaction, in that the effect of one population (mass) on another is inversely related to the spatial distance between them.

Box 3 | **Ten questions for future research on influenza virus evolution**

- What are the main fitness determinants of influenza A virus, particularly in proteins other than haemagglutinin?
- What evolutionary and epidemiological processes underpin antigenic drift?
- When and where does most antigenic drift occur?
- What is the rate of reassortment in influenza virus (particularly relative to the rate of mutation)?
- What role does reassortment have in the generation of genetic and antigenic novelty?
- Does homologous RNA recombination occur in influenza virus?
- How frequent is epistasis across the influenza virus genome and what role does it have in viral evolution?
- What is the extent of intra-host genetic and antigenic variation in influenza virus?
- How severe is the population bottleneck at inter-host viral transmission?
- What processes determine the seasonality of influenza virus?

Obtaining the answers to these questions will require a combination of experimental, epidemiological and genomic approaches. In particular, it will be necessary to undertake a much larger and more systematic sampling of influenza virus genome sequences that is comprehensive in both time (within and between epidemic seasons) and space (including geographical regions that are currently poorly surveyed, particularly the tropics, Africa and Latin America). Other key data sets include those that document intra-host viral genetic diversity, influenza transmission chains (such as infected families) and viral genotypes and phenotypes in samples from patients that are asymptomatic for influenza disease.

the United States over the past 30 years, using a newly developed gravity model, showed that the timing of epidemics is most synchronized between the most populous states and during the most severe disease seasons[81]. Furthermore, the long-distance spread of influenza between cities and states was better correlated with adult workflow traffic patterns than with simple geographical distance. Nevertheless, the full epidemiology of the virus remains complex, and children are still believed to drive the spread of influenza at more local levels: within schools, households and communities in general.

In the future, phylogenetic analysis could help to reveal key aspects of influenza virus seasonality. By inferring the evolutionary relationships that exist between viruses that have been sampled from spatially disjunct regions, particularly the tropics, it might be possible to determine the directions of global viral migration and the location of the virus during non-epidemic periods. Indeed, the recent phylogenetic analysis of viruses from single populations has shown that the virus does not 'over-summer', but dies out at the end of each seasonal epidemic, and that subsequent seasonal viral re-emergence is ignited by imported genetic variation[30].

Conclusions

Improvements in methods for bioinformatic[67] and epidemiological analysis[81], as well as a greatly expanded GenBank database of influenza virus genome sequences, provides an unprecedented opportunity to investigate long-standing questions in influenza virus epidemiology and evolution[82,83] (for a discussion of key research questions in influenza virus evolution, see BOX 3).

Recent work indicates that the evolutionary dynamics of influenza virus might be more complex than was previously thought, reflecting an intricate interplay between antigenic variation, natural selection and reassortment. For example, at the local spatial level, migration and reassortment among multiple co-circulating lineages of the same subtype might be more important determinants of the seasonal evolution of influenza virus than antigenic drift[30], with periodic, selection-driven cluster jumps that result in major changes in antigenic phenotype[14,31,33]. Despite these important insights, a comprehensive understanding of influenza virus evolution will require a far broader analysis of whole-genome sequences from a wider range of subtypes, host species and geographical areas, including tropical regions, as well as the development of more realistic epidemiological models.

It is also striking that, despite the huge amount of sequence data that has been generated for influenza A virus, studies of intra-host genetic variation are largely absent. However, the high rates of mutation and replication that are common to most RNA viruses mean that intra-host population diversity is likely to be extensive, even in viruses that cause acute infections[84]. Furthermore, if the population bottleneck at inter-host transmission is not particularly severe, multiple viral lineages, including reassortants, viruses with new antigenic characteristics or even defective viruses[85], are likely to be transmitted among hosts. A crucial task for future studies in influenza virus evolution is therefore to quantify the extent of intra-host genetic variation within single individuals to determine whether this includes isolates that are antigenically distinct, and reveal how much genetic diversity is transmitted among hosts and how this might differ among avian and mammalian influenza viruses.

An important shortcoming in research on influenza virus has been the lack of a unifying framework that integrates genome sequence, phenotypic (including antigenic) and epidemiological data. The recent reconstruction of the 1918 pandemic influenza virus genome sequence[86] demonstrates how a whole-genome analysis can provide crucial insights into long-standing questions about the virulence and aetiology of this catastrophic disease event[87]. Similarly, the ability to simultaneously analyse genetic and phenotypic influenza virus data has had a strong influence on our understanding of the patterns and timing of the antigenic evolution of influenza virus[33]. The Influenza Virus Resource, which is now available on GenBank, exemplifies the most recent attempt to integrate epidemiological and molecular data by making various influenza virus data publicly available, including whole-genome sequences along with the date of isolation, patient characteristics and geographical locations[83]. Notably, antigenic data are still excluded from this resource. Other influenza virus data sets are less conducive for research, as epidemiological data has rarely been collected in conjunction with sequence data, and much data have not been made publicly accessible.

Although new analytical methods and faster sequencing technology offer the opportunity to address crucial questions about influenza virus evolution through phylogenetic analyses, greater surveillance of viral populations and access to data underpin the advancement of this key field of viral research.

1. Thompson, W. W. *et al.* Mortality associated with influenza and respiratory syncytial virus in the United States. *JAMA* **289**, 179–186 (2003).
2. Taubenberger, J. K., Reid, A. H., Janczewksi, T. A. & Fanning, T. G. Integrating historical, clinical and molecular genetic data in order to explain the origin and virulence of the 1918 Spanish influenza virus. *Phil. Trans. R. Soc. Lond. B Biol. Sci.* **356**, 1857–1859 (2001).
3. Johnson, N. P. A. S. & Mueller, J. Updating the accounts: global mortality of the 1918–20 'Spanish' influenza epidemic. *Bull. Hist. Med.* **76**, 105–115 (2002).
4. Ferguson, N. M., Fraser, C., Donnelly, C. A, Ghani, A. C. & Anderson, R. M. Public health risk from the avian H5N1 influenza epidemic. *Science* **304**, 968–969 (2004).
5. Ferguson, N. M. *et al.* Strategies for containing an emerging influenza pandemic in Southeast Asia. *Nature* **437**, 209–214 (2005).
6. Ferguson, N. M. *et al.* Strategies for mitigating an influenza pandemic. *Nature* **442**, 448–452 (2006).
7. Longini, I. M. Jr *et al.* Containing pandemic influenza at the source. *Science* **309**, 1083–1087 (2005).
8. Webster, R. G., Bean, W. J., Gorman, O. T., Chambers, T. M. & Kawaoka, Y. Evolution and ecology of influenza A viruses. *Microbiol. Rev.* **56**, 152–179 (1992).
9. Fouchier, R. A. *et al.* Characterization of a novel influenza A virus hemagglutinin subtype (H16) obtained from black-headed gulls. *J. Virol.* **79**, 2814–2822 (2005).
10. Baigent, S. & McCauley, J. Influenza type A in humans, mammals and birds: determinants of virus virulence, host-range and interspecies transmission. *BioEssays* **25**, 657–671 (2003).
11. Horimoto, T. & Kawaoka, Y. Influenza: lessons from past pandemics, warnings from current incidents. *Nature Rev. Microbiol.* **3**, 591–600 (2005).
12. Kuiken, T. *et al.* Host species barriers to influenza virus infections. *Science* **312**, 394–397 (2006).
13. Ferguson, N. M., Galvani, A. P. & Bush, R. M. Ecological and immunological determinants of influenza evolution. *Nature* **422**, 428–433 (2003).
The first attempt to link the phylogenetic tree structure of HA1 from influenza A virus with an explicit epidemiological model.
14. Koelle, K., Cobey, S., Grenfell, B. & Pascual, M. Epochal evolution shapes the phylodynamics of interpandemic influenza A (H3N2) in humans. *Science* **314**, 1898–1903 (2006).
An important new phylodynamic model of evolution for the HA protein of influenza A virus, which was based on neutral networks. It links evolutionary change at the genetic, antigenic and epidemiological scales.
15. Fitch, W. M., Bush, R. M., Bender, C. A. & Cox, N. J. Long term trends in the evolution of H(3) HA1 human influenza type A. *Proc. Natl Acad. Sci.USA* **94**, 7712–7718 (1997).
16. Webster, R. G., Laver, W. G., Air, G. M. & Schild, G. C. Molecular mechanisms of variation in influenza viruses. *Nature* **296**, 115–121 (1982).
17. Bush, R. M., Fitch, W. M., Bender, C. A. & Cox, N. J. Positive selection on the H3 hemagglutinin gene of human influenza virus A. *Mol. Biol. Evol.* **16**, 1457–1465 (1999).
18. Ina, Y. & Gojobori, T. Statistical analysis of nucleotide sequences of the hemagglutinin gene of human influenza A viruses. *Proc. Natl Acad. Sci. USA* **91**, 8388–8392 (1994).
19. Suzuki, Y. Natural selection on the influenza virus genome. *Mol. Biol. Evol.* **23**, 1902–1911 (2006).
20. Fitch, W. M., Leiter J. M. E., Li X. & Palese P. Positive Darwinian evolution in human influenza A viruses. *Proc. Natl Acad. Sci. USA* **88**, 4270–4274 (1991).
21. Ratner, V. A., Kolchanov, N. A. & Omel'ianchuk, L. V. Phylogenetic analysis of genes of the influenza virus. Relationship between adaptability and neutrality. *Genetika* **25**, 1499–1507 (1989).
22. Bean, W. J. *et al.* Evolution of the H3 influenza virus hemagglutinin from human and nonhuman hosts. *J. Virol.* **66**, 1129–1138 (1992).
23. Bush, R. M., Bender, C. A., Subbarao, K., Cox, N. J. & Fitch, W. M. Predicting the evolution of human influenza A. *Science* **286**, 1921–1925 (1999).
24. Daly, J. M. *et al.* Antigenic and genetic evolution of equine H3N8 influenza A viruses. *J. Gen. Virol.* **77**, 661–671 (1996).
25. Yamashita, M., Krystal, M., Firch, M. & Palese, P. Influenza B virus evolution: co-circulating lineages and comparison of evolutionary pattern with those of influenza A and C viruses. *Virology* **163**, 112–122 (1988).
26. Kanegae, Y. *et al.* Evolutionary pattern of the hemagglutinin gene of influenza B viruses isolated in Japan: cocirculating lineages in the same epidemic season. *J. Virol.* **64**, 2860–2865 (1990).
27. Buonagurio, D. A. *et al.* Noncumulative sequence changes in the hemagglutinin genes of influenza C virus isolates. *Virology* **146**, 221–232 (1985).
28. Plotkin, J. B., Dushoff, J. & Levin, S. A. Hemagglutinin sequence clusters and the antigenic evolution of influenza A virus. *Proc. Natl Acad. Sci. USA* **99**, 6263–6268 (2002).
29. Holmes, E. C. *et al.* Whole genome analysis of human influenza A virus reveals multiple persistent lineages and reassortment among recent H3N2 viruses. *PLoS Biol.* **3**, e300 (2005).
The first large-scale analysis of the genome-wide evolution of influenza A virus. The authors show that multiple lineages can co-circulate within populations (in this case, the state of New York, USA) and reassort.
30. Nelson, M. I. *et al.* Stochastic processes are key determinants of the short-term evolution of influenza A virus. *PLoS Pathog.* **2**, e125 (2006).
31. Wolf, Y. I., Viboud, C., Holmes, E. C., Koonin, E. V. & Lipman, D. J. Long intervals of stasis punctuated by bursts of positive selection in the seasonal evolution of influenza A virus. *Biol. Direct.* **1**, 34 (2006).
A detailed analysis of the selection pressures that affect influenza A virus at the epidemic scale. The authors show that positive selection is punctuated rather than continuous, and is focused on epitopes in HA1.
32. Williamson, S. Adaptation in the *env* gene of HIV-1 and evolutionary theories of disease progression. *Mol. Biol. Evol.* **20**, 1318–1325 (2003).
33. Smith, D. J. *et al.* Mapping the antigenic and genetic evolution of influenza virus. *Science* **305**, 371–376 (2004).
A landmark paper that describes the antigenic cartography of influenza A virus and how it relates to evolution at the genetic level.
34. Lavenu, A. *et al.* Detailed analysis of the genetic evolution of influenza virus during the course of an epidemic. *Epidemiol. Infect.* **134**, 514–520 (2005).
35. de Jong, J. C., Beyer, W. E. P., Palache, A. M., Rimmelzwaan, G. F. & Osterhaus, A. D. M. E. Mismatch between the 1997/1998 influenza vaccine and the major epidemic A(H3N2) virus strain as the cause of an inadequate vaccine-induced antibody response to this strain in the elderly. *J. Med. Virol.* **61**, 94–99 (2000).
36. Boni, M. F., Gog, J. R., Andreasen, V. & Feldman, M. W. Epidemic dynamics and antigenic evolution in a single season of influenza A. *Proc. Biol. Sci.* **273**, 1307–1316 (2006).
37. Lindstrom, S. E., Cox, N. J. & Klimov, A. Genetic analysis of human H2N2 and early H3N2 influenza viruses, 1957–1972: evidence for genetic divergence and multiple reassortment events. *Virology* **328**, 101–119 (2004).
38. Taubenberger, J. K. *et al.* Characterization of the 1918 influenza virus polymerase genes. *Nature* **437**, 889–893 (2005).
39. Antonovics, J., Hood, M. & Baker, C. H. Molecular virology: was the 1918 flu avian in origin? *Nature* **440**, e9 (2006).
40. Gibbs, M. J. & Gibbs, A. Molecular virology: was the 1918 pandemic caused by a bird flu? *Nature* **440**, e8 (2006).
41. Barr, I. G. *et al.* An influenza A(H3) reassortant was epidemic in Australia and New Zealand in 2003. *J. Med. Virol.* **76**, 391–397 (2005).
42. Lin, Y. P., Bennett, G. M., & Hay, A. Recent changes among human influenza viruses. *Virus Res.* **103**, 47–52 (2004).
43. Lindstrom, S. E. *et al.* Phylogenetic analysis of the entire genome of influenza A (H3N2) viruses from Japan: evidence for genetic reassortment of the six internal genes. *J. Virol.* **72**, 8021–8031 (1998).
44. Lindstrom, S. *et al.* Comparative analysis of evolutionary mechanisms of the hemagglutinin and three internal protein genes of influenza B virus: multiple cocirculating lineages and frequent reassortment of the NP, M, and NS genes. *J. Virol.* **73**, 4413–4426 (1999).
45. Macken, C. A., Webby, R. J. & Bruno, W. J. Genotype turnover by reassortment of replication complex genes from avian Influenza A virus. *J. Gen. Virol.* **87**, 2803–2815 (2006).
The first detailed attempt to use phylogenetic methods to measure the rate of reassortment in influenza virus, using avian isolates as a model system.
46. Didelot, X. & Falush, D. Inference of bacterial microevolution using multilocus sequence data. *Genetics* 6 Dec 2006 (doi: 10.1534/genetics.106.063305).
47. Rimmelzwaan, G. F., Berkhoff, E. G., Nieuwkoop, N. J., Fouchier, R. A. & Osterhaus, A. D. M. E. Functional compensation of a detrimental amino acid substitution in a cytotoxic-T-lymphocyte epitope of influenza A viruses by comutations. *J. Virol.* **78**, 8946–8949 (2004).
48. Mitnaul, L. J. *et al.* Balanced hemagglutinin and neuraminidase activities are critical for efficient replication of influenza A virus. *J. Virol.* **74**, 6015–6020 (2000).
49. Voeten, J. T. M. *et al.* Antigenic drift in the influenza A virus (H3N2) nucleoprotein and escape from recognition by cytotoxic T lymphocytes. *J. Virol.* **74**, 6800–6807 (2000).
50. Khatchikian, D., Orlich, M. & Rott, R. Increased viral pathogenicity after insertion of a 28S ribosomal RNA sequence into the haemagglutinin gene of an influenza virus. *Nature* **340**, 1567 (1989).
51. Orlich, M., Gottwals, H. & Rott, R. Nonhomologous recombination between the hemagglutinin gene and the nucleoprotein gene of an influenza virus. *Virology* **204**, 462–465 (1994).
52. Chare, E. R., Gould, E. A. & Holmes, E. C. Phylogenetic analysis reveals a low rate of homologous recombination in negative-sense RNA viruses. *J. Gen. Virol.* **84**, 2691–2703 (2003).
53. Gibbs, M. J., Armstrong, J. S. & Gibbs, A. J. Recombination in the hemagglutinin gene of the 1918 'Spanish flu'. *Science* **293**, 1842–1845 (2001).
54. Worobey, M., Rambaut, A., Pybus, O. P. & Robertson D. L. Questioning the evidence for genetic recombination in the 1918 'Spanish flu' virus. *Science* **296**, 211 (2002).
55. Drake, J. W. & Holland, J. J. Mutation rates among RNA viruses. *Proc. Natl Acad. Sci. USA* **96**, 13910–13913 (1999).
56. Hanada, K., Suzuki, Y. & Gojobori, T. A large variation in the rates of synonymous substitution for RNA viruses and its relationship to diversity of viral infection and transmission modes. *Mol. Biol. Evol.* **21**, 1074–1080 (2004).
57. Jenkins, G. M., Rambaut, A., Pybus, O. G. & Holmes, E. C. Rates of molecular evolution in RNA viruses: a quantitative phylogenetic approach. *J. Mol. Evol.* **54**, 156–165 (2002).
58. Chen, R. & Holmes, E. C. Avian influenza virus exhibits rapid evolutionary dynamics. *Mol. Biol. Evol.* **23**, 2336–2341 (2006).
59. Suarez, D. L. Evolution of avian influenza viruses. *Vet. Microbiol.* **74**, 15–27 (2000).
60. Ludwig, S. *et al.* Recent influenza A (H1N1) infections of pigs and turkeys in Northern Europe. *Virology* **202**, 281–286 (1994).
61. Gorman, O. T. *et al.* Evolution of influenza A virus nucleoprotein genes: implications for the origins of H1N1 human and classical swine viruses. *J. Virol.* **65**, 3704–3714 (1991).
62. Air, M., Gibbs, A. J., Laver, W. G. & Webster, R. G. Evolutionary changes in influenza B are not primarily governed by antibody selection. *Proc. Natl Acad. Sci. USA* **87**, 3884–3888 (1990).
63. Gorman, O. T., Bean, W. J., Kawaoka, Y. & Webster, R. G. Evolution of nucleoprotein gene of influenza A virus. *J. Virol.* **64**, 1487–1497 (1990).
64. Suzuki, Y. & Nei, M. Origin and evolution of influenza virus hemagglutinin genes. *Mol. Biol. Evol.* **19**, 501–509 (2002).
65. Bean, W. J. *et al.* Evolution of H3 influenza virus hemagglutinin from human and non-human hosts. *J. Virol.* **66**, 1129–1138 (1992).
66. Rambaut, A. Estimating the rate of molecular evolution: incorporating non-contemporaneous sequences into maximum likelihood phylogenies. *Bioinformatics* **16**, 395–399 (2000).

67. Drummond, A. J., Ho, S. Y. W., Phillips, M. J. & Rambaut, A. Relaxed phylogenetics and dating with confidence. *PLoS Biol*. **4**, e88 (2006).
An important methodological paper that shows how key evolutionary parameters, particularly substitution rates and times of divergence, can be estimated from temporally sampled data sets such as those from RNA viruses.
68. Drummond, A., Pybus, O. G. & Rambaut, A. Inference of viral evolutionary rates from molecular sequences. *Adv. Parasitol*. **54**, 331–358 (2003).
69. Dowell, S. F. Seasonal variation in host susceptibility and cycles of certain infectious diseases. *Emerg. Infect. Dis.* **7**, 369–374 (2001).
An excellent review of the possible causes of seasonality in influenza virus.
70. Hemmes, J., Winkler, K. & Kool, S. Virus survival as a seasonal factor in influenza and poliomyeletitis. *Nature* **188**, 430–431 (1960).
71. Schulman, J. & Kilbourne, E. Experimental transmission of influenza virus infection in mice. II. Some factors affecting the incidence of transmitted infections. *J. Exp. Med.* **118**, 267–275 (1963).
72. Chiu, S. S. *et al.* Influenza-related hospitalizations among children in Hong Kong. *N. Engl. J. Med.* **347**, 2097–2103 (2002).
73. Wong, C. M. *et al.* Influenza-associated hospitalization in a subtropical city. *PLoS Med.* **3**, e121 (2006).
74. Nelson, R. J., Demas, G. E., Kelin, S. L. & Kriegsfeld, L. J. *Seasonal Patterns Of Stress, Immune Function, And Disease* (Cambridge Univ. Press, London, 2002).
75. Shephard, R. J. & Shek, P. N. Cold exposure and immune function. *Can. J. Physiol. Pharmacol.* **76**, 828–36 (1998).
76. Cannell, J. J. *et al.* Epidemic influenza and vitamin D. *Epidemiol. Infect*. **134**, 1129–1140 (2006).
77. Halvorson, D. A., Kelleher, C. J. & Senne, D. A. Epizootiology of avian influenza: effect of season on incidence in sentinal ducks and domestic turkeys in Minnesota. *Appl. Environ. Micro.* **49**, 914–919 (1985).
78. Viboud, C. *et al.* Influenza epidemics in the United States, France, and Australia, 1972–1997. *Emerg. Infect. Dis*. **10**, 32–39 (2004).
79. Shek, L. P. C. & Lee, B. W. Epidemiology and seasonality of respiratory tract virus infections in the tropics. *Paediatr. Respir. Rev*. **4**, 105–111 (2003).
80. Viboud, C., Alonso, W. J. & Simonsen, L. Influenza in tropical regions. *PLoS Med*. **3**, e89–e90 (2006).
81. Viboud, C. *et al.* Synchrony, waves, and spatial hierarchies in the spread of influenza. *Science* **312**, 447–451 (2006).
An elegant description of the spatial dynamics of influenza virus within the United Sates. The authors show that the local spread of influenza A virus is largely determined by patterns of workflow.
82. Smith, D. J. Predictability and preparedness in influenza control. *Science* **312**, 392–394 (2006).
83. Ghedin, E. *et al.* Large-scale sequencing of human influenza reveals the dynamic nature of viral genome evolution. *Nature* **437**, 1162–1166 (2005).
84. Domingo, E. & Holland, J. J. RNA virus mutations for fitness and survival. *Ann. Rev. Microbiol*. **51**, 151–178 (1997).
85. Aaskov, J., Buzacott, K., Thu, H. M., Lowry, K. & Holmes, E. C. Long-term transmission of defective RNA viruses in humans and *Aedes* mosquitoes. *Science* **311**, 236–238 (2006).
86. Tumpey, T. M. *et al.* Characterization of the reconstructed 1918 Spanish influenza pandemic virus. *Science* **310**, 77–80 (2005).
A reconstruction of the genome of the H1N1 influenza A virus that was associated with the 1918 pandemic, and an investigation of its virulence determinants.
87. Kash, J. C. *et al.* Genomic analysis of increased host immune and cell death responses induced by 1918 influenza virus. *Nature* **443**, 578–581 (2006).
88. Steinhauer, D. A. & Skehel, J. J. Genetics of influenza viruses. *Ann. Rev. Genet*. **36**, 305–332 (2002).
89. Lamb, R. A. & Klug, R. M. in *Fundamental Virology* 3rd edn (eds Fields, B. N., Knipe, D. M. & Howley, P. M. *et al.*) 605–647 (Lippincott–Raven, Philadelphia, 2006).

Competing interests statement
The authors declare no competing financial interests.

FURTHER INFORMATION

Center for Infectious Disease Dynamics:
http://www.cidd.psu.edu
Influenza Virus Resource:
http://www.ncbi.nlm.nih.gov/genomes/FLU/FLU.htm
National Center for Health Statistics:
http://www.cdc.gov/nchs
Access to this links box is available online.