Identifying genetic markers of adaptation for surveillance of viral host jumps

Kim M. Pepin*, Sandra Lass‡#, Juliet R. C. Pulliam§||#, Andrew F. Read§¶ and James O. Lloyd-Smith||§

Abstract | Adaptation is often thought to affect the likelihood that a virus will be able to successfully emerge in a new host species. If so, surveillance for genetic markers of adaptation could help to predict the risk of disease emergence. However, adaptation is difficult to distinguish conclusively from the other processes that generate genetic change. In this Review we survey the research on the host jumps of influenza A, severe acute respiratory syndrome-coronavirus, canine parvovirus and Venezuelan equine encephalitis virus to illustrate the insights that can arise from combining genetic surveillance with microbiological experimentation in the context of epidemiological data. We argue that using a multidisciplinary approach for surveillance will provide a better understanding of when adaptations are required for host jumps and thus when predictive genetic markers may be present.

*Department of Physics, Pennsylvania State University, University Park, Pennsylvania 16802, USA.
‡Kirchenhölzle 18, 79104 Freiburg, Germany.
§Fogarty International Center, National Institutes of Health, Bethesda, Maryland 20892, USA.
||Department of Ecology and Evolutionary Biology, University of California Los Angeles, California 90095, USA.
¶Centre for Infectious Disease Dynamics, Departments of Biology and Entomology, Pennsylvania State University, University Park, Pennsylvania 16827, USA.
#These authors contributed equally to this work.
Correspondence to K.M.P
e-mail: kmp29@psu.edu

Major public health threats can occur when a virus that is endemic in a reservoir species is transmitted in a new host population, as occurred with severe acute respiratory syndrome-coronavirus (SARS-CoV). The need to assess the risk of such outbreaks in the future has led to calls for genetic surveillance of high-risk viruses (for example, lyssaviruses and avian influenza viruses) in their reservoir hosts (for example, bats, poultry and wild birds)[1–3]. Such approaches assume that certain genetic variants, or combinations of them, are more likely to emerge, and that these can be recognized before a host jump. For example, genomic analysis of the influenza A virus that was linked to the 1918 Spanish flu pandemic, may reveal the genetic factors that made Spanish flu so transmissible and lethal, and allow real-time assessment of the risk for a pandemic and potentially prevention of avian influenza in poultry and wild birds through genetic surveillance. If ten amino acid changes were required to jump from avian hosts to humans in 1918, as proposed[4], then the risk of emergence would be higher when nine of these mutations are present in viruses isolated from poultry than when none is (although note that the nature and origin of genetic changes in the 1918 pandemic remain under debate[5,6]).

To predict host jumps it is necessary to identify the causal genetic markers, which is not a trivial task. As host jumps can be entirely due to ecological change, a causal genetic change in the virus is not actually necessary[7–11]. Even if viral adaptation is responsible for the host jump, genomic information can be used to predict future risk only if genetic markers of adaptation to new hosts can be identified and their influence on viral spread can be characterized. As G. C. Williams first pointed out nearly 50 years ago[12], analyses of adaptation are extremely difficult because of the risk of 'adaptive storytelling' (REFS 13,14): it is extremely easy to infer adaptation when it does not exist. In this Review, we discuss what is required to show that particular viral genetic changes are responsible for host jumps. Microbiological experiments are essential to distinguish between the host jump-associated changes in the viral genome that are caused by adaptation and those that are not, particularly because quantifying viral fitness is the gold standard for rigorously demonstrating adaptation[15].

Host-jump mechanisms

There are different ecological and evolutionary processes involved in viral host jumps (FIG. 1). If the primary factor causing emergence is ecological, and adaptation is not required for the jump to occur, the cause of the host jump is known as an ecological driver (FIG. 1a,b). However, if genetic change in the virus is required for emergence in a new host, the cause is termed an adaptive driver (FIG. 1c,d), although an ecological driver is likely to be

Ecological driver (no selection required)

a

Adaptive driver (selection required)
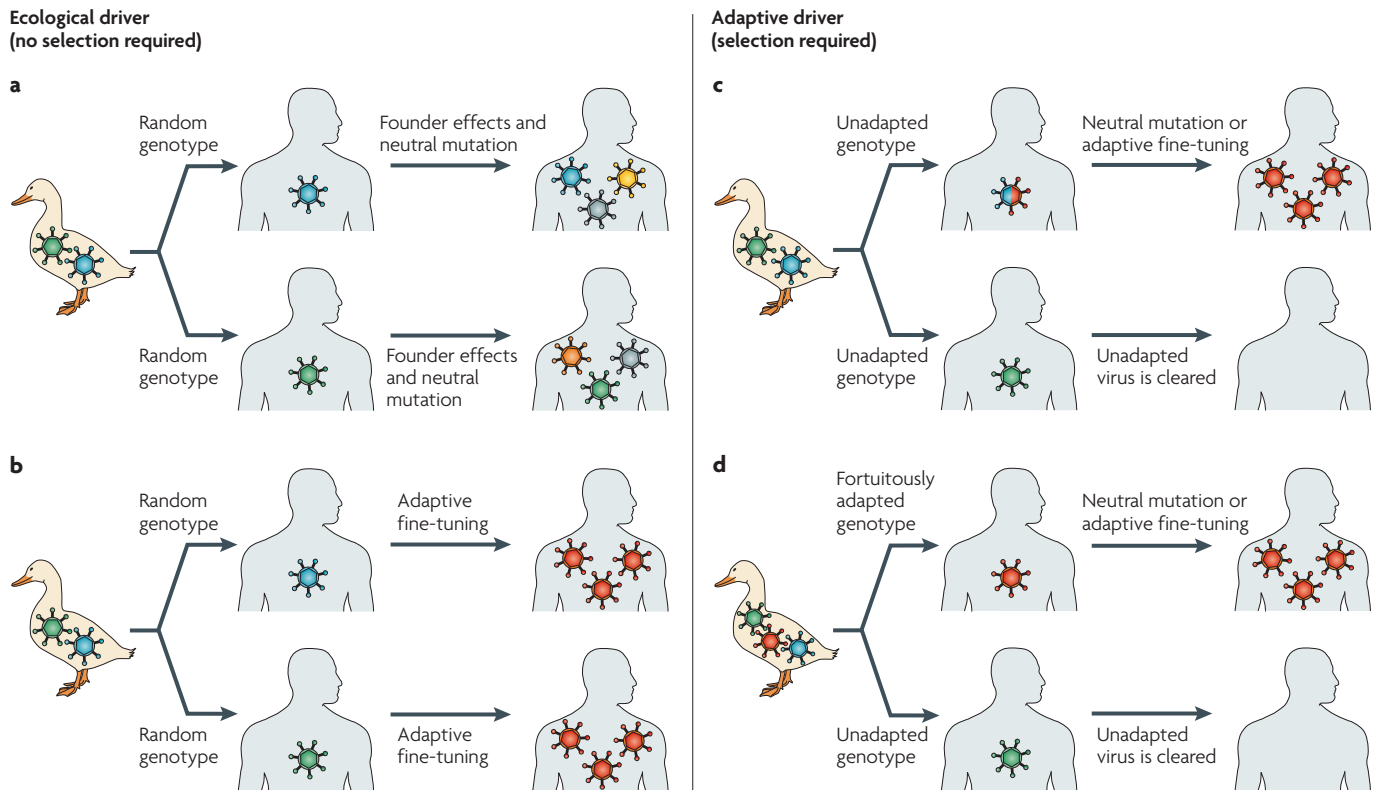
c

b

d

Figure 1 | **Mechanisms of viral emergence in new hosts.** Host jumps and associated genetic diversity can arise through a range of ecological and evolutionary mechanisms. **a** | Ecological driver with founder effects. Any of the viral genotypes circulating in the reservoir are already competent for transmission in the new host; the basic reproductive number (R_0) of the virus strains in the reservoir and the new host are >1. Following host jump, neutral mutations occur during replication and transmission in the new host population. The combination of founder effects and neutral mutation result in a shifted distribution of viral genotypes in the new host, which each have R_0 >1. **b** | Ecological driver with adaptive fine-tuning. The circulating viral genotypes that spill over are already adapted for transmission to the new host (R_0 >1), but adaptive substitutions occur owing to long-term selection in the new host even though they were not required for the initial emergence, such that the R_0 of the adapted virus is greater than the R_0 of the virus that initially spilled over. **c** | Adaptive mutation in an unadapted genotype. The reservoir strain that causes emergence has R_0 <1 in the new host and requires additional genetic change to reach R_0 >1. The change could occur in the initial host individual in whom spillover occurred or during stuttering transmission. Some genotypes will possess a genetic or phenotypic predisposition to acquire the necessary adaptive mutations and achieve R_0 >1, which makes them more likely to emerge. These strains are predictors of emergence risk. **d** | Spillover of a fortuitously adapted genotype. The genotype that spills over and causes emergence already has R_0 >1 in the new host, unlike other viral genotypes circulating in the reservoir population. Scenarios **b** and **d** differ because in **d** some genotypes in the reservoir host can establish in the new host, whereas others cannot; in **b**, the genotype composition in the reservoir host is not a predictor of emergence.

Reservoir
Host population in which a virus is maintained long term.

Host jump
The complete process of a virus transmitting between a reservoir host and a new host followed by sustained transmission among individuals of the new host population.

Emergence
The appearance of new viruses in a population, particularly the appearance and sustained transmission of viruses in a new host species (as opposed to new strains in host populations in which the virus is endemic).

Ecological change
A shift in frequency, nature or outcome of host species contact caused by factors such as demography, migration, invasion or environmental change.

Adaptation
Genetic change driven by natural selection.

present in this situation as well. An adaptive driver is required when a selective 'sieve' is present in the new host population immediately after cross-species transmission, so that at least some of the virus genotypes are excluded from emergence by selection rather than by chance. More precisely, the term adaptive driver is used to describe situations in which specific viral genotypes capable of sustained spread in the new host species are selected over other genotypes that are doomed to fail. A central question for the field is how often adaptation is required for a host jump to occur.

When adaptation is required for the viral host jump (FIG. 1c,d), the adaptive genetic changes can originate either in the new host (known as 'tailor-made'; FIG.1c) or in the reservoir host (known as 'off-the-shelf'; FIG.1d)[16]. If the crucial mutations occur in the new host, a continuum

of scenarios can occur; at one end, genotypes circulating in the reservoir population may be genetically close to genotypes that could emerge successfully in the new host, whereas at the other end the host genotypes could be distant. The definition of close and how many close genotypes may exist, and at what frequencies, are key issues to be resolved for predicting the risk of emergence. Alternatively, all the necessary genetic material required for the viral host jump may arise in the reservoir host, in which case risk prediction depends on understanding which conditions produce the genetic variants that are ready to jump, how often they may be produced and whether they can be maintained in the reservoir host by selection (which could give rise to a particularly high risk of emergence if ecological conditions allow). Regardless of whether the final adaptive steps are taken

REVIEWS

in the reservoir or new host, at least some information about the risk of emergence can in principle be inferred from viral genetic data sampled from the reservoir population (in contrast to the processes depicted in FIG. 1a,b, in which genetic surveillance of reservoir populations is not informative about risk) — the hard part is to deduce which markers predict risk.

Identifying adaptation

Given the range of processes that can drive a host jump, the question is how to distinguish adaptive genetic changes (which are potentially predictive of emergence risk) from those derived from other evolutionary processes. This requires the demonstration that a particular genetic change affects fitness, which is challenging from the standpoint of fundamental evolutionary biology[12–14] and applied eco-epidemiology[17]. Convergent evolution can be used to distinguish adaptation from neutral genetic variation on the basis of sequence data[18,19] (although the absence of convergence, as has been observed for avian influenza viruses that circulate in swine[20], does not rule out adaptation). The logic for this is that given the number of different possibilities for mutations and of opportunities for stochastic changes in any genotype, only a strong selective force is likely to cause multiple occurrences of the same genetic makeup originating from different starting points. In the context of host jumps, evolutionary convergence would mean the repeated evolution of the same or functionally related genetic changes associated with independent jumps of a virus into the same new host species.

One remarkable example of convergent evolution occurred during the early spread of SARS-CoV in humans in 2003. Viral sequences from patients in unconnected outbreaks in Beijing and Guangdong province, China, showed that a five-step mutational pathway had occurred at least twice independently following spillover from the reservoir host[21,22]. Although this is consistent with adaptation, these data alone do not imply that the changes were required for the host jump, as the observed evolutionary convergence occurred after cross-species transmission and could therefore indicate adaptive fine-tuning. However, evidence from an impressive series of studies provides further hints. Although numerous independent cases of SARS-CoV transmission from the reservoir host occurred, most died out after a few human cases, indicating that the introduced strain was not fit for human–human transmission[23]. The putative adaptive mutations, including non-synonymous changes in the spike protein responsible for host cell receptor binding, were found only in viruses transmitted between humans and not in those circulating in palm civets (the reservoir host) or in viruses that were transmitted from palm civets to humans in later spillover events that did not result in human–human transmission[24,25]. A representative viral strain isolated from palm civets could not replicate in cultured human airway epithelial cells, but the introduction of a single amino acid change (Lys479Asn), which is observed in all human isolates, substantially boosted this measure of fitness[24]. Structural analysis of this and another mutation nearby (Ser487Thr) revealed

the molecular mechanism underlying this effect[26]. These studies illustrate the power of integrating surveillance, molecular epidemiology, bioinformatics and microbiology in making a compelling argument that the SARS-CoV host jump required viral adaptation.

Unfortunately, the type of data needed for testing for genetic convergence (for example, replicate host jumps linked to contact-tracing data) can be difficult to obtain. Thus, searches for genetic markers of host jumps often begin with large-scale genetic analyses of viral samples from many host species[27–30]. This identifies genetic differences of viruses in different host species but does not give information about which markers may be linked to adaptation: genetic differences between viral strains can arise in the new host as a consequence of founder effects and neutral evolution (FIG. 1a). Thus, the next step is to test whether amino acids at marker sites are the product of positive selection in the new host species[31–35] to predict which markers are linked to adaptation (FIG. 1b–d). However, these predictions require experimental validation because most sequence-based methods of identifying positive selection rely on the assumption that non-synonymous mutations have much larger fitness effects than synonymous mutations, which can be false, especially in viruses. Indeed, viruses are known to adapt to selective pressures such as tRNA levels, polymerase efficiency controlled by secondary structuring in viral genomic RNA, RNA folding energy and changes in promoter sequences, through positive selection on synonymous mutations[36–41]. Thus, although a ratio of non-synonymous to synonymous mutations that is greater than 1 may indeed be indicative of positive selection, the opposite ratio is not necessarily indicative of negative selection, but could be the result of positive selection of a synonymous mutation (depending on the type of selective pressure). Screens for genetic markers of host adaptation should therefore include investigation of the primary sequence. Protein-based methods are a promising alternative to avoid reliance on this assumption[35] but they are less well developed, and some require detailed organism-specific knowledge of protein evolutionary patterns, making them less accessible. Furthermore, a recent analysis showed that bioinformatic methods for identifying adaptive substitutions can have high rates of false positives or miss important adaptive changes[42]. Thus, experiments that directly measure the viral fitness effects of mutations are important to show adaptation, and, as discussed below, even more data are needed to confirm whether particular genetic markers of adaptation are drivers of host jumps (and therefore that the host jumps are not due to adaptive fine-tuning, disconfirming the mechanism in FIG. 1b in favour of those in FIG. 1c,d).

In addition to confirming whether host species-associated genetic markers are linked to selection, it is crucial to experimentally determine adaptive evolutionary pathways and constraints. In host jumps such as the emergence of feline panleukopenia virus in dogs (now considered a separate viral species, canine parvovirus), cell culture experiments indicate that multiple genetic changes need to occur together in the receptor binding

site of the viral capsid to yield viable viruses in the new host species[43]. In cases in which intermediate viral genotypes have low fitness in the new host, as suspected in the emergence of canine parvovirus, important genetic markers may be undetectable by sequence-based screens of adaptive sites and therefore not available for risk assessment. Identifying how mutations interact in different genetic backgrounds is thus important for predicting the risk of emergence from genetic data. Site-directed mutagenesis, a process in which specific mutations are engineered into different genetic backgrounds, can serve to obtain fitness measures of the unobserved genotypes[44]. By reconstructing possible evolutionary pathways, this type of reverse genetic technique has been used to reveal the number of potential adaptive trajectories that can occur by a seven-step mutational pathway during the adaptation of HIV to alternative co-receptors[45], an adaptive challenge analogous to the host jumps considered here. To effectively assess emergence risk in these cases, in which intermediate genotypes have low fitness, even more emphasis must be placed on identifying how such combinations can overcome the adaptive constraints; for example, by high mutation rates (such that double mutants are likely), by recombination or reassortment (that is, mixing of genetic material)[46], by the use of intermediate hosts[47] in which the mutation does not have low fitness or by increased contact of the reservoir host with the new host species[48].

Measuring fitness

Although central to understanding natural selection and adaptation, fitness is frustratingly difficult to measure in natural settings. A common approach is to use surrogate measures of fitness (known as fitness components). Demonstrations of improved trait performance in the new host, or that an observed mutation improves trait performance or within-host fitness, provide compelling support that a particular substitution is adaptive. However, the relationship between individual traits, within-host fitness and virus transmission are complex and poorly understood[49]. For example, receptor binding affinity does not necessarily predict within-host fitness, and viral titres may not be directly proportional to transmission probability. Such measures provide valuable information, but they should be treated as surrogate measures of viral fitness. Successful emergence at the host population scale requires sustained transmission, and theoretical work has confirmed the importance of integrating within-host and between-host processes for explaining viral evolution[50–54]. Some theoretical advancements have been made towards integrating within-host microbiological insights with their population-scale genetic consequences for the virus[53,54]; however, these insights have yet to be applied in an empirical context. Key questions remain that must be addressed by experiments that directly measure viral transmission.

At the epidemiological scale, the basic reproductive number (R_0) of a virus[55] is probably a good approximation of viral fitness in the endemic case in a reservoir host[56]. However, during an outbreak, the number of secondary cases per case per unit time is a more appropriate

measure of viral fitness than simply the total number of secondary cases per case. Strains with a greater ability to infect new hosts can be favoured as an epidemic builds even if the number of successful infections during their lifetime is lower[57] (a parallel argument holds for viral replication within hosts[58]). On the basis of this, R_0 can be used to measure viral fitness, as an $R_0 < 1$ indicates that a strain is poorly adapted to its host and will die out[59]. Unfortunately, distinguishing whether R_0 is above the threshold value of 1 is difficult, as viral strains with $R_0 < 1$ can occasionally cause small outbreaks before dying out (owing to stuttering transmission[60]), which gives the appearance, at least temporarily, of $R_0 > 1$ (REFS 61,62.) Moreover, if a host jump arises because the virus adapts to the new host during this stuttering transmission (FIG. 1c), which may have occurred during the 2002–2003 SARS epidemic[23], then evidence that the virus was initially not adapted to the new host and had $R_0 < 1$ is essentially erased. To show that $R_0 < 1$ for the original virus, multiple cross-species transmission events of the same viral strain, including multiple failed spillovers, have to be analysed[63]. Ideally, data would include contact-tracing information and host immune history to estimate the effect of host heterogeneities in host susceptibility to the virus or the virus infectivity on R_0 (REF. 64) (for example, the number of individuals that have been in contact with the host or the host immune status). This is obviously difficult but not impossible for diseases in which active surveillance efforts have identified hundreds of failed introductions, such as monkeypox virus[65], H5N1 influenza virus[66] and Nipah virus[67]. For example, H5N1 influenza transmission from avian species to humans has caused 498 cases (294 deaths) in 15 countries worldwide since 2003, but no sustained chains of human–human transmission have been observed[66].

Distinguishing adaptive drivers of host jumps

Different mechanisms have been detected for several host jumps (TABLE 1). To provide sufficient evidence that adaptation caused a host jump (distinguishing the mechanisms in FIG. 1a,b from those in FIG. 1c,d), adaptation must be shown together with evidence that some genotypes circulating in the reservoir host have $R_0 < 1$ in the new host and that the adaptive mutations to some of the circulating reservoir-derived strains lead to $R_0 > 1$ in the new host. The most conceptually straightforward approach to identify the origin of the adaptive substitutions is to carry out fine-scale longitudinal sampling just before and after the host jump, followed by genome sequencing to identify new mutations and fitness assays of particular genotypes in both the reservoir and new hosts. Such data would be extremely valuable and, indeed, unique. However, there are still caveats. Although the identification of the emerged strain in the reservoir host shows that it originated there, it is challenging to draw firm conclusions if it is absent from the putative reservoir host because the identification of rare viral genotypes require extensive sampling. In the case of segmented viruses, which frequently undergo reassortment with strains from other host species, such as

Table 1 | **The role of adaptation in host jumps to humans for selected zoonotic viruses**

Zoonotic pathogen (zoonotic stage*)	Reservoir hosts	Sources of transmission to humans	Emergence mechanism[†]	Evidence for or against
SARS-CoV (IV)	Bats (putative)	Palm civets	A	Against: convergent sequence evolution suggests adaptation did occur
			B	Against: adaptation was necessary • Repeated independent cross-species transmission to humans followed by limited on-going transmission • All human isolates had numerous non-synonymous mutations not observed in palm civet isolates • Some of these mutations improve the ability of palm civet virus isolates to reproduce in cultured cells bearing human receptors
			C and D	Cannot distinguish between C and D; more intensive sampling from palm civets and humans around the time of initial spillover is required
Ebola virus Zaire (IV)	Fruit bats (putative)	Wild apes, duiker (speculated) and unknown host	A–D	Insufficient sampling of virus in wildlife and along chains of human–human transmission
H5N1 influenza A virus (III)	Wild birds and domestic poultry	Domestic poultry	A and B	Against: repeated independent cross-species transmission to humans followed by little or no ongoing transmission suggest that genetic types predominant in the reservoir host have $R_0 < 1$ in humans and that an adaptive driver is required for a host jump
			C and D	Cannot distinguish between C and D; genetic surveillance in the reservoir species and along chains of human–human transmission is necessary
Nipah virus Bangladesh (III)	Fruit bats	Fruit bats (through contaminated date palm sap) and domestic animals (speculated)	A and B	Against: repeated independent cross-species transmission to humans followed by limited ongoing transmission suggests that genetic types predominant in the reservoir host have $R_0 < 1$ in humans and that an adaptive driver is required for a host jump
			C and D	Cannot distinguish between C and D; greater understanding of potential genetic markers of adaptation is necessary
Monkeypox virus (III)	Unknown	Prairie dogs (USA), rope squirrels (speculated; Africa) and monkeys (Africa)	A and B	Against: repeated independent cross-species transmission to humans followed by limited ongoing transmission suggests that genetic types predominant in the reservoir host have $R_0 < 1$ in humans and that an adaptive driver is required for a host jump
				For: transmission of monkeypox among humans may be severely curtailed because of protective antibodies from past smallpox vaccination; increasing incidence of monkeypox in humans has been observed as population-level antibody prevalence has declined, possibly allowing sustained human–human transmission without viral genetic changes
			C and D	Cannot distinguish between C and D; greater understanding of potential genetic markers of adaptation is necessary
VEEV (II)	Rodents	Horses (by mosquitoes)	A and B	Against: • Repeated independent cross-species transmission to humans with no evidence of ongoing transmission suggests that genetic types predominant in the source host have $R_0 \approx 1$ in humans and that an adaptive driver is required for a host jump • Adaptation seems to be required for successful emergence in horses
			C and D	Cannot distinguish between C and D; greater understanding of potential genetic markers[§] of adaptation is necessary

*Stage I: animal pathogen with no evidence of transmission to humans (not included here). Stage II: transmission to humans occurs with no evidence for subsequent chains of human–human transmission. Stage III: cross-species transmission to humans is followed by human–human transmission with $0 < R_0 < 1$. Stage IV: cross-species transmission to humans is followed by epidemic or pandemic human–human transmission with $R_0 > 1$. Stage V: cross-species transmission in the past produced a new endemic human pathogen (not included here)[60]. [†]In mechanisms A and B the primary factor causing emergence is ecological; although in both cases the virus is competent for transmission to the new host, in mechanism A neutral mutations may occur whereas in mechanism B adaptive fine-tuning is observed. By contrast, in mechanisms C and D adaptation is required; in mechanism C the adaptive change originates in the new host whereas in mechanism D it originates in the reservoir host. See FIG. 1 for details. [§]Although genetic markers of adaptation to horses have been identified, these markers fail to predict transmission in laboratory models; therefore, markers of adaptation to horses should not be considered generally applicable and are unlikely to predict human emergence risk. SARS-CoV, severe acute respiratory syndrome-coronavirus; VEEV, Venezuelan equine encephalitis virus.

influenza A[5], sampling from alternative host species may be required to determine the origin of the adaptive genetic material and the mechanism by which it arises.

Experiments can also ascertain whether the strain is likely to have been derived from the reservoir or new host (distinguishing scenarios in FIG. 1c and FIG. 1d). For example, if the fitness of the emerged strain is low in the reservoir species, the adaptive change probably arose in the new host (supporting the scenario in FIG. 1c) as such a strain would be present at low frequencies in the reservoir host owing to mutation–selection balance (although note that frequent co-infection and relaxed selection could allow low-fitness variants to persist by genetic complementation[68]). However, if strains that are closely related to the emerged strain have high fitness in the reservoir host or are observed at high frequencies in reservoir populations, it is more likely that the emerged strain arose in the reservoir (supporting the scenario in FIG. 1d). This hypothesis could be corroborated by the demonstration that the adaptive genetic material can be produced in the reservoir host, either by gene flow, reassortment or recurrent mutation, and fitness assays would give insight into whether they could be sustained at significant frequencies. Thus, distinguishing adaptive drivers of host jumps relies on microbiolgal techniques, including reverse genetics, experimental measures of fitness and experimental evolution. Site-directed mutagenesis as well as recombination and reassortment studies in reservoir host models are particularly useful approaches to determine whether specific genotypes can be produced and selected in the reservoir, and replicate experimental evolution lines can assess how likely these events are and identify adaptive constraints and sources of selection (TABLE 2).

Empirical foundation of the field

We have carried out a systematic review of the published literature of four viruses for which host-jump mechanisms are well studied: influenza A virus, SARS-CoV, canine parvovirus and Venezuelan equine encephalitis virus (VEEV) (184 publications) (see Supplementary information S1 (table), Supplementary information S2 (box) and Supplementary information S3–S7 (figures)). The survey revealed that research on viral host jumps typically unfolds in four steps (sampling of the virus, *in vitro* culture of the virus, studies in animal models and experiments in reservoir and new hosts (TABLE 2)) and that increased efforts on integrating surveillance practices and bioinformatics with microbiolgal experiments accelerate our ability to confirm genetic markers of viral adaptation.

Sampling and genetic analyses. To date, disease surveillance practices typically are limited to opportunistic sampling, with more systematic protocols coming after host jumps have occurred. There are three scales at which the collection of multiple samples is important: within host individuals, within host species and between host species (including both the reservoir and new host species). In existing studies, strains were most often sampled from many individuals of the host species and/or from individuals of multiple host species (FIG. 2), although many of these studies were limited to bioinformatic analyses of viral genetic data (Supplementary information S1 (table)). By contrast, a remarkably high number of experimental studies focused on detailed analyses of a single strain (and recombinants of that strain), which in practice is often a necessary trade-off to obtain the data needed to validate adaptation. Studies of strains that had previously not been investigated should be prioritized because it is crucial to examine many strains and to develop high-throughput methods for screening putative genetic markers for fitness effects. Multiple samples from a single host individual were almost never studied by any approach (FIG. 2), indicating a data gap. Cross-sectional and longitudinal sampling across both reservoir and new host species is essential for mapping where and when genetic changes take place. Not only do these samples provide the material for phenotypic assays, they are important for decreasing errors that are associated with estimates of evolutionary rates and for linking experimental results to disease characteristics and viral fitness. Experiments could improve our understanding of the propensity for viral adaptation by examining genetic variation throughout experimental infections and its effects on adaptation rates and epidemiological parameters such as infectious period.

Integrating sampling efforts in the field with the design of host-jump studies is important because the ecological and epidemiological context inherently determines changes in viral genetic variation and fitness. More than half of the studies we surveyed focused only on genetic sequence analyses and did not carry out experiments (Supplementary information S5 (figure)), and most did not integrate ecological, epidemiological and host response data (FIG. 3). This is a lost opportunity, because without data on ecological context it is extremely difficult to identify the signature of specific evolutionary drivers from sequence data and to separate changes in transmission that are due to ecology from those that are due to evolutionary processes. Similarly, a lack of epidemiological information, such as direct measures of virus incidence or data on host population serology, hampers the correlation between genetic patterns and factors such as disease virulence or selection by host immune status. For example, recent experiments on a phage–bacterium model system showed that adaptation to new hosts is extremely sensitive to contact rates between the original and new host species[69]. Thus, to obtain the greatest insight into viral evolutionary patterns from sequence data, information about host demographics and contact patterns should be a component of strain sampling protocols (see REF. 70 for a detailed review on current gaps in strain sampling and sequencing strategies).

Reverse genetics. The three fundamental components of adaptive evolution are genetic inputs, phenotypic variation and selection by the environment. To identify and explain the forces that underlie evolutionary change in newly emerged viral populations, the three components need to be measured concurrently (Supplementary information S5 (figure)). For influenza A, site-directed

---

Mutation–selection balance
Steady-state frequency of deleterious genotypes determined by the balance between their continual creation by mutation and their exclusion by selection.

Experimental evolution
Measuring evolutionary change in real-time by applying evolutionary forces experimentally and observing the outcome.

Evolutionary process
A factor that drives genetic change, including genetic drift, mutation, gene flow and natural selection.

Table 2 | **Steps of investigation of examining viral host jumps***

Sample viruses	*In vitro* and cell culture	Animal models	Reservoir and new hosts
Frequent (included in many of the publications)			
• Sampling of opportunistically viruses from reservoir and/or new host (including important host data such as species and location) • Consensus sequencing of viral genes and genomes • Use of nucleotide-based bioinformatics to identify genetic markers of host specificity and adaptation[28,114]	• Antigenic characterization (for example, haemagglutination-inhibition assays) • Introduction of putative adaptive mutations in wild-type strain and removal of adaptive mutation from adapted virus[110] • Receptor binding and infectivity assays in cells expressing the viral receptor of the new host[115] • Polymerase activity or replication rate assays[71]	• Infections with a single strain • Determination of viral fitness by measuring within-host fitness, tissue tropism, pathogenicity and virulence	• Measurement of viral traits in either the reservoir or the new host using viral genotypes (and/or recombinants) from the reservoir or new host population from one host-jump event[99]
Infrequent (possible to collect but included more rarely)			
• Systematic sampling of viral populations cross-sectionally and longitudinally from many individuals and new host populations • Sequencing of many clones from each viral population • Integration of protein structure prediction, phylogenetic analyses and/or protein-based bioinformatics to identify host-jump markers that are under selection[35] • Collection of additional host data concurrently, including ecology (host population size and density), serology, incidence and clinical data	• Characterization of capsid stability in the environment to evaluate changes in transmission traits • Synthesis of recombinants carrying each potentially adaptive mutation individually and combinations of mutations to measure epistatic effects and determine mutational pathways[43] • Examination of factors affecting adaptation rates by conducting experimental evolution in cell culture	• Examination of multiple isolates sampled longitudinally from reservoir and new host species in the same study • Determination of viral fitness and its link to virulence and the host-jump fitness components by measuring tissue tropism, pathogenicity and virulence[110] • Measurement of transmission traits, such as the rate of dissemination to transmission routes and average number of virions available for transmission • Identification of genetic basis of transmission • Measurment of initial viral titre, consensus sequence of genes for many time points and fitness measures of evolved viral populations relative to ancestor during passage (experimental evolution)[108]	• Measurement of infectivity–dose relationship for natural routes of infection, of host immune response[78] and its effects on viral within-host dynamics and on protection against re-infection, and of transmission, including more than one contact per infected host[116] and more than one transmission • Infection–transmission experiments with viral genotypes from the reservoir and the new host populations obtained from multiple host jump events, for each genotype in each host and cross-species transmissions (transplant experiments)[117]
Ideal (difficult to obtain and rarely included)			
• Characterization of viral genotypes that circulate early in an outbreak and their relation to genotypes in the reservoir host by repeated sampling of infected individuals • Intensive sampling of reservoir and new host populations immediately following host jump[22] • Sequencing of viral genomes from many reservoir hosts to identify rare fortuitously adapted genotypes[32] • Use of genetic data and evolutionary theory to target regions that are hotspots of viral adaptation • Use of next-generation sequencing technology to increase sampling within and between hosts • Collection of clinical data on pathology, infectious period, shedding rates and immune response concurrently with repeated sampling during infection • Integration of genetic and incidence data to determine transmission networks • Sequencing of clones from many time points to link disease characteristics with patterns of genetic variation and transmission networks • Analysis of genetic data from replicate transmission chains to identify sites with convergent evolution[23]	• Examination of multiple isolates from both the reservoir and new host species (if possible strains with an estimated transmission chain) for site-directed mutagenesis and phenotypic assays	• Measurement of viral fitness in multiple host individuals to quantify variability in viral fitness and determine sources of the fitness variation • Sequencing of genomes of strains from the evolved populations at many time points during the infection time course and from multiple replicate experimental evolution lines to assess adaptive constraints and repeatability • Measurement of fitness of isolated genotypes from evolved populations to quantify adaptive constraints • Comparison of evolved differences in the presence and absence of selection on transmission using artificial and experimental transmission	• Measurement of viral fitness and host immune response in many host individuals to quantify effects of host heterogeneities (due to innate or adaptive immunity status or host genetics) on viral fitness • Experimental evolution as in the animal experiments and include three types of experimental transmission lineages: reservoir–reservoir, new–new host and cross-species transmission

*Steps used to address mechanistic hypotheses in the host jump of influenza A, severe acute respiratory syndrome-coronavirus, canine parvovirus and Venezuelan equine encephalitis virus. Note that the time and financial investment increase from step 1 (sample virus) to step 4 (test in reservoir and new host) negatively correlates with research efforts to date (see Supplementary information S1–S7 (figures)). Earlier steps provide useful groundwork for later steps, and although many techniques are important at multiple steps, they are stated only in the step in which they first become relevant.
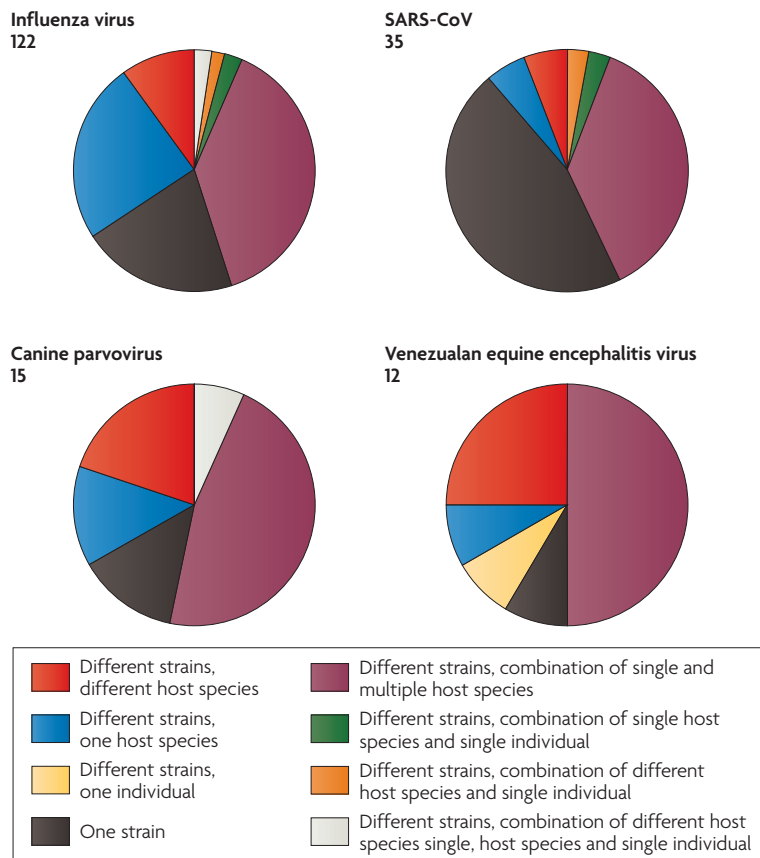
Figure 2 | Origin of strains studied in the surveyed literature. Articles describing evolutionary data of transmission of influenza virus, severe acute respiratory syndrome-coronavirus (SARS-CoV), canine parvovirus and Venezuelan equine encephalitis virus were grouped into one of the eight categories on the basis of the host origin of the viral strains used for experimentation and/or analyses. Inclusion of multiple strains from multiple hosts is important for a comprehensive experimental design. Numbers indicate the total number of published papers surveyed.

Fitness landscape
The relationship of genotype and reproductive success. Often depicted in three dimensional space, in which the X and Y axes are the coordinates that describe all possible genetic combinations in the genome and the height of the Z axis gives the reproductive success for a given set of genetic coordinates.

Experimental transmission
Transmission of a pathogen by placing infected and uninfected animals in close proximity; this is quasi-natural transmission as it excludes the role of contact probability in transmission.

mutagenesis and plasmid-mediated transfection of all eight genetic segments have been successfully applied to identify virulence-enhancing mutations that are selected in new hosts[71,72]. More recent studies have emphasized the importance of repeating these reverse genetic assays in other epidemic strains of the virus. For example, although a Glu627Lys substitution in influenza polymerase basic protein 2 is a virulence-enhancing marker in H5N1 (REF. 71) and H7N7 (REF. 72), it has no effect in the recent H1N1 pandemic strain[73,74]. Other challenges for future research are to expand these types of study so that multiple strains from the reservoir host and the new host are examined in both hosts, and to decrease the use of virus strains that have been heavily passaged in the laboratory (FIG. 3) (Supplementary information S4 (figure)). By testing the phenotypic effects of putative host-specificity mutations in more than one strain, other mutagenesis studies have shown that the genetic background of the virus influences interactions between mutations (known as epistasis[75]), the number of traits affected by a mutation (known as pleiotropy[49]) and the adaptability of the virus[76], even when the genetic distance between strains is minimal.

Measurement of mutational effects in multiple genetic backgrounds and many host models is important to identify molecular mechanisms of viral host jumps and draw general conclusions.

Beyond receptors and antigens. A further priority is to expand the scope of phenotypic analyses. Decades of research have revealed that receptor binding can be an important barrier for a host jump to overcome, and phenotypic analyses of some viruses in this Review were biased towards measuring this and other simple traits such as antigenicity (which is a logical and practical place to start) (Supplementary information S5 (figure)). In the more frequently studied viruses, other traits encoded in viral non-structural proteins (for example, replication rate, modulation of host cell factors that regulate the host innate immunity and RNA folding free energy) have been measured and found to be crucial determinants of within-host viral fitness in new hosts[41,46,71,77,78] (Supplementary information S5 (figure)). Future research should be broadened to include other traits that influence within-host fitness, such as those involved in cell egress or nuclear entry[79], as well as those that determine transmission directly (as done recently for influenza A[80]) and indirectly, including transmission-related traits such as stability outside the host, alternative routes of transmission and dissemination rates to routes of transmission; all are major determinants of viral fitness landscapes. Moreover, as experimental transmission is used to assess transmissibility of newly emerged strains in their reservoirs[81], it is becoming apparent that within-host replication and between-host infectivity do not necessarily correlate[82], which again raises concerns about the use of within-host fitness to measure viral population fitness.

Host models. The impact of host genotype (and more broadly, host species) on measures of viral phenotype must not be underestimated. Single viral mutations can affect viral fitness differently depending on host genotype[49,83–85], and amino acid substitution rates and evolutionary patterns depend on both host genotype and the patterns of contact that determine how the virus encounters different host types[86–89]. Thus, when it is necessary to use model hosts, it is important to measure differences in viral transmission in both the reservoir and new hosts using models that mimic the natural hosts as closely as possible. In the case of influenza A, it is feasible to carry out experiments in poultry, waterfowl, passerine birds and swine, and these systems can be used for understanding strain-specific infection characteristics of new epidemic strains[90,91] and the within-host evolutionary processes that generate them[92]. The combined results of three recent studies of H1N1 highlight the importance of using natural host systems whenever possible to assess viral characteristics[93–95] (see ProMED mail post on influenza A H1N1 pandemic). Studies of influenza A viruses from human cases during the recent H1N1 pandemic of (from May 2009 to January 2010) found that the Asp222Gly mutation in haemagglutinin caused significantly more severe disease. Earlier studies

Types of relevant data
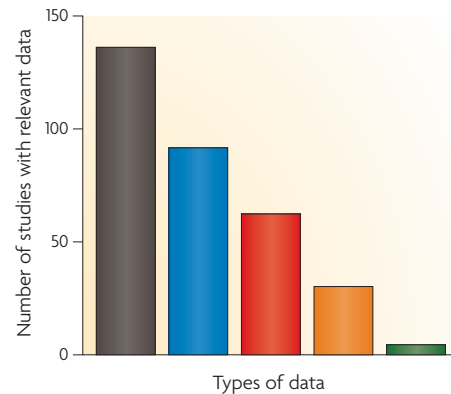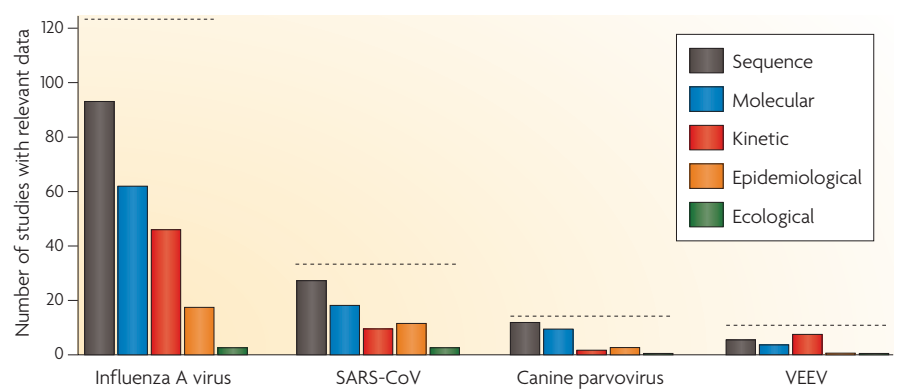
Types of relevant data by virus

Figure 3 | **Types of relevant data collected by surveyed articles.** The Y axis indicates the number of published papers. Data for all viruses are shown on the left. Dashed lines in right plot indicate the total number of papers for each virus shown. Sequence data refer to sequence analysis of genes, genome or proteins; molecular data refer to viral phenotypes, including antigenicity, receptor binding, genome replication, virion packaging and polymerase binding; kinetic data measure the time course of virion production, reflecting within-host fitness; epidemiological data measure incidence of the virus in the host population; and ecological data refer to environmental conditions, including host movement or contact patterns. Note that articles can fall into more than one category. SARS-CoV, severe acure respiratory syndrome-coronavirus; VEEV, Venezuelan equine encephalitis virus.

had used a glycan array to show that this substitution expanded tissue tropism in humans, which could potentially increase disease severity by allowing it to infect more cell types. However, the association between disease severity and the Asp222Gly substitution was not found in ferrets, a commonly used model species for mammalian influenza A[93]. Furthermore, infection of different bird species with H5N1 results in a wide range of morbidity, mortality and virus shedding patterns both within and between taxonomic groups[91,96,97].

Replication. Distinguishing processes that drive host jumps from patterns that occur owing to chance replication of host-jump events is important, and two parameters need investigating: between host-jump events (examining multiple host-jump events for a single virus strain) and within an event (examining multiple strains from the same host-jump event). Replication between events identifies potential generalities of host-jump mechanisms, whereas replication within events helps to quantify the likelihood of re-occurrence, both of which can be used to quantify risk. Although desirable, replication at both levels is a challenge as host jumps are usually infrequent and unpredictable. Despite the logistical challenges involved, more than half of the published studies include some level of replication (see Supplementary information S6 (figure)). Greater investment in experiments that compare strains from independent emergence events under the same experimental conditions will help to elucidate host-jump traits that can be used as markers of emergence risk. This approach has been used to identify genetic markers of host-jump potential of VEEV. First, phylogenetic analyses of VEEV that included multiple strains from independent, successful jumps from rodents to horses (epizootic strains)[98] were used to identify mutations that

were unique to the epizootic strains (that is, potential genetic markers of epizootic risk). One of these mutations, Glu117Lys, alters the charge of the surface-exposed E2 glycoprotein, suggesting that it could also have fitness effects in the new host (horses). Subsequent infection experiments in horses with chimaeric viruses showed that this and other charge-altering mutations in the virus capsid are both necessary and sufficient to produce high viral titres in horses[99], thus allowing for transmission to the mosquito vector and increased epizootic potential. Two charge-altering mutations (Thr213Arg and Thr213Lys) have even been repeatedly associated with VEEV host jumps[98–100] (convergent evolution); however, whether this substitution alone can produce the epizootic phenotype seems to depend on the genetic background in which the substitution occurs[98–100], and host jumps have been observed without this particular change[98].

Experimental evolution. Experimental evolution is a powerful technique to determine the role of different evolutionary processes in defining genetic patterns[101–103]. The procedure typically involves the propagation of an isolated genotype under controlled conditions followed by genetic sequencing and, ideally, fitness measures of evolved genotypes. Viruses are ideal for this because their genomes are small enough that they can be fully sequenced many times, and they replicate fast enough that high rates of evolution are observable on short timescales (days to weeks). In phages, replication of experimental evolution (testing independent lineages of the same strain) has revealed high levels of parallel evolution[104–106], which is not only strong evidence of adaptation but also allows quantification of the role of random variation in virus evolution — a necessary but elusive ingredient for predicting evolutionary outcome and understanding adaptive constraints. Similarly,

experimental evolution of avian influenza strains in mice has revealed convergent evolution with strains isolated from humans, providing a useful way to identify host-jump markers of adaptation from large pools of genetic data[107–109]. Such experiments can also address important knowledge gaps, such as how often adaptive mutations can be produced in the reservoir host or in new hosts, which selective pressures promote or prevent fixation of adaptive mutations, the frequency of parallel and convergent evolution, and the trajectory of adaptive evolution. Intensive sequencing of replicate viral strains and evolutionary lines can reveal the extent of adaptive constraints and the impact of genetic diversity and population bottlenecks on evolutionary trajectories.

The design of experimental evolution experiments is challenging. In a recent experimental evolution study of influenza H3N2, the combination of viral gene sequence data of evolved viral populations with measures of virulence, tissue tropism and pathology revealed adaptive mutations that enabled the evolved population to replicate faster and to higher titres in mouse lungs and other tissues and cell types[110]. This evolution experiment involved serial lung–lung passaging in mice, starting with intranasal infection, harvesting the virus from the lungs of infected animals and subsequent infection of naive mice with homogenized lung tissue. This method was appropriate for the goals in this study, but the homogenization procedure reduces the evolution experiment to a single lineage, which does not allow drawing any conclusions about the adaptation or adaptive constraints of other lineages. It is also noteworthy that the evolved viral population did not transmit well in experimental transmission assays, illustrating that the selection procedure (extracting the virus from lungs and administering it directly to naive mice) did not impose selection on transmission traits and, again, that within-host fitness may not predict between-host fitness. Although this type of experimental design is useful for identifying adaptive mutations and linking them to specific functions, it does not impose selective pressures that are likely to be present in nature and thus does not facilitate the identification of genetic markers that would confer adaption in the wild.

Conclusion

The increasing ease of large-scale genomic sequencing, together with advances in bioinformatics, molecular evolutionary theory and new statistical tools for linking viral genetic variation with epidemiology and phylogeography[111,112], is providing valuable means to visualize viral emergence and generating hypotheses about evolutionary mechanisms. However, full analysis of the resulting hypotheses must also involve biological measurements. Accessible databases have become available to support the growing pool of genetic data that have resulted from increased surveillance. Alongside the genetic data, these databases should include ecological, epidemiological and phenotypic data (an effort that has been piloted for influenza A[113]). This would also help to establish research design standards for studying pathogen emergence so that key public health gaps can be addressed. One approach that would provide balance in data collection and integrative analyses is to develop disease-emergence funding programmes that require interdisciplinary teams (including field ecologists, microbiologists, immunologists, epidemiologists, bioinformaticians and evolutionary biologists) using multiple approaches (field sampling, laboratory experiments, data analysis and theoretical modelling). Such programmes would encourage greater balance in the types of data collected, would help to ensure that data collection is structured in a way that is conducive to analytical goals and would promote the broad collaborations needed to address overarching questions about disease emergence. Detecting adaptation is a great challenge in any context, and the case of viral host jumps is no exception. But understanding the adaptive genetic change involved in host jumps could yield large gains. Not only would we have a more complete account of the role of natural selection in host jumps, we could also generate genetic markers for future risk of epidemic or pandemic disease.

1. Calisher, C. H., Childs, J. E., Field, H. E., Holmes, K. V. & Schountz, T. Bats: important reservoir hosts of emerging viruses. *Clin. Microbiol. Rev.* **19**, 531–545 (2006).
2. Guberti, V. & Newman, S. C. Guidelines on wild bird surveillance for highly pathogenic avian influenza H5N1 virus. *J. Wildl. Dis.* **43**, 29–34 (2007).
3. Taubenberger, J. K. & Morens, D. M. 1918 influenza: the mother of all pandemics. *Emerg. Infect. Dis.* **12**, 15–22 (2006).
   Review of the 1918 influenza pandemic claiming that an understanding of the historical, epidemiological and biological aspects of the 1918 influenza as well as extensive sampling and sequencing of influenza A strains from animals are necessary to understand the nature of influenza pandemics.
4. Taubenberger, J. K. *et al.* Characterization of the 1918 influenza virus polymerase genes. *Nature* **437**, 889–893 (2005).
5. Smith, G. J. *et al.* Dating the emergence of pandemic influenza viruses. *Proc. Natl Acad. Sci. USA* **106**, 11709–11712 (2009).
6. Gibbs, M. J. & Gibbs, A. J. Was the 1918 pandemic caused by a bird flu? *Nature* **440**, E8 (2006).
7. Daszak, P., Cunningham, A. A. & Hyatt, A. D. Wildlife ecology - emerging infectious diseases of wildlife - threats to biodiversity and human health. *Science* **287**, 443–449 (2000).
8. Dobson, A. & Foufopoulos, J. Emerging infectious pathogens of wildlife. *Philos. Trans. R. Soc. Lond., B Biol. Sci.* **356**, 1001–1012 (2001).
9. Jones, K. E. *et al.* Global trends in emerging infectious diseases. *Nature* **451**, 990–993 (2008).
10. Plowright, R. K., Sokolow, S. H., Gorman, M. E., Daszak, P. & Foley, J. E. Causal inference in disease ecology: investigating ecological drivers of disease emergence. *Front. Ecol. Environ.* **6**, 420–429 (2008).
11. Woolhouse, M. E. J. & Gowtage-Sequeria, S. Host range and emerging and reemerging pathogens. *Emerging Infect. Dis.* **11**, 1842–1847 (2005).
12. Williams, G. C. in *A Critique of Some Current Evolutionary Thought* (Princeton University Press, Princeton, 1966).
13. Gould, S. J. & Lewontin, R. C. The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. *Proc. R. Soc. Lond. B Biol. Sci.* **205**, 581–598 (1979).
14. Nielsen, R. Adaptionism-30 years after Gould and Lewontin. *Evolution* **63**, 2487–2490 (2009).
    Comment on the difficulties that evolutionary biologists confront when measuring adaptation, emphasizing how these challenges have changed with the increased availability of vast amounts of genetic data and how the data can improve our understanding of evolutionary processes.
15. Orr, H. A. Fitness and its role in evolutionary genetics. *Nature Rev. Genet.* **10**, 531–539 (2009).
16. Woolhouse, M. & Antia, R. in *Evolution in Health and Disease*, 2nd Edition (eds Stearns, S. C. & Koella, J. C.) (Oxford University Press, Oxford, 2007).
17. Matthews, L. & Woolhouse, M. New approaches to quantifying the spread of infection. *Nature Rev. Microbiol.* **3**, 529–536 (2005).
    Review of advances in epidemiological modelling that address disease emergence conditions in the early stages and in small populations.
18. Messier, W. & Stewart, C. B. Episodic adaptive evolution of primate lysozymes. *Nature* **385**, 151–154 (1997).
19. Wood, T. E., Burke, J. M. & Rieseberg, L. H. Parallel genotypic adaptation: when evolution repeats itself. *Genetica* **123**, 157–170 (2005).
    Review of parallel and convergent evolution as signatures of adaptation, drawing from experimentally evolved cases.
20. Dunham, E. J. *et al.* Different evolutionary trajectories of European avian-like and classical swine H1N1 influenza A viruses. *J. Virol.* **83**, 5485–5494 (2009).

21. Liu, W. *et al.* Molecular epidemiology of SARS-associated coronavirus, Beijing. *Emerg. Infect. Dis.* **11**, 1420–1424 (2005).
Report of SARS-CoV genetic sequences isolated from patients in Beijing, showing convergent patterns of evolution with viruses isolated in Guangdong province.

22. Chinese, S. M. E. C. Molecular evolution of the SARS coronavirus during the course of the SARS epidemic in China. *Science* **303**, 1666–1669 (2004).

23. Zhao, G. P. SARS molecular epidemiology: a Chinese fairy tale of controlling an emerging zoonotic disease in the genomics era. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **362**, 1063–1081 (2007).
Summarizes the evidence that SARS-CoV emergence was driven by viral adaptation and illustrates how the globally coordinated response to the SARS outbreak was so successful and what methodological insight was gained for future public health emergencies.

24. Sheahan, T., Rockx, B., Donaldson, E., Corti, D. & Baric, R. Pathways of cross-species transmission of synthetically reconstructed zoonotic severe acute respiratory syndrome coronavirus. *J. Virol.* **82**, 8721–8732 (2008).
Experimental studies of SARS-CoV emergence in which genes from reservoir and new hosts, and from viruses isolated from both hosts, are mixed in chimaeric constructs to test the specificity of particular interactions and the evolutionary potential of the virus.

25. Song, H. D. *et al.* Cross-host evolution of severe acute respiratory syndrome coronavirus in palm civet and human. *Proc. Natl Acad. Sci. USA* **102**, 2430–2435 (2005).

26. Li, F. Structural analysis of major species barriers between humans and palm civets for severe acute respiratory syndrome coronavirus infections. *J. Virol.* **82**, 6984–6991 (2008).

27. Allen, J. E., Gardner, S. N., Vitalis, E. A. & Slezak, T. R. Conserved amino acid markers from past influenza pandemic strains. *BMC Microbiol.* **9**, 77 (2009).

28. Chen, G. W. *et al.* Genomic signatures of human versus avian influenza A viruses. *Emerg. Infect. Dis.* **12**, 1353–1360 (2006).

29. Finkelstein, D. B. *et al.* Persistent host markers in pandemic and H5N1 influenza viruses. *J. Virol.* **81**, 10292–10299 (2007).

30. Miotto, O., Heiny, A. T., Tan, T. W., August, J. T. & Brusic, V. Identification of human-to-human transmissibility factors in PB2 proteins of influenza A by large-scale mutual information analysis. *BMC Bioinformatics* **9**, S18 (2008).

31. Furuse, Y., Suzuki, A., Kamigaki, T. & Oshitani, H. Evolution of the *M* gene of the influenza A virus in different host species: large-scale sequence analysis. *Virol. J.* **6**, 67 (2009).

32. Hoelzer, K., Shackelton, L. A., Parrish, C. R. & Holmes, E. C. Phylogenetic analysis reveals the emergence, evolution and dispersal of carnivore parvoviruses. *J. Gen. Virol.* **89**, 2280–2289 (2008).

33. Tamuri, A. U., Dos Reis, M., Hay, A. J. & Goldstein, R. A. Identifying changes in selective constraints: host shifts in influenza. *PLoS Comput. Biol.* **5**, e1000564 (2009).

34. Tang, X. *et al.* Differential stepwise evolution of SARS coronavirus functional proteins in different host species. *BMC Evol. Biol.* **9**, 52 (2009).

35. Kosakovsky Pond, S. L., Poon, A. F., Leigh Brown, A. J. & Frost, S. D. A maximum likelihood method for detecting directional evolution in protein sequences and its application to influenza A virus. *Mol. Biol. Evol.* **25**, 1809–1824 (2008).
Summarizes current nucleotide-based bioinformatic methods for identifying adaptation in genetic sequences; develops a protein-based method that circumvents these issues and applies it to identify sites under directional selection in influenza A viruses.

36. Novella, I. S., Zarate, S., Metzgar, D. & Ebendick-Corpus, B. E. Positive selection of synonymous mutations in vesicular stomatitis virus. *J. Mol. Biol.* **342**, 1415–1421 (2004).

37. Pepin, K. M., Domsic, J. & McKenna, R. Genomic evolution in a virus under specific selection for host recognition. *Infect. Genet. Evol.* **8**, 825–834 (2008).

38. Aragones, L., Guix, S., Ribes, E., Bosch, A. & Pinto, R. M. Fine-tuning translation kinetics selection as the driving force of codon usage bias in the hepatitis A virus capsid. *PLoS Pathog.* **6**, e1000797 (2010).

39. Bahir, I., Fromer, M., Prat, Y. & Linial, M. Viral adaptation to host: a proteome-based analysis of codon usage and amino acid preferences. *Mol. Syst. Biol.* **5**, 311 (2009).

40. van Hemert, F. J., Berkhout, B. & Lukashov, V. V. Host-related nucleotide composition and codon usage as driving forces in the recent evolution of the Astroviridae. *Virology* **361**, 447–454 (2007).

41. Brower-Sinning, R. *et al.* The role of RNA folding free energy in the evolution of the polymerase genes of the influenza A virus. *Genome Biol.* **10**, R18 (2009).

42. Nozawa, M., Suzuki, Y. & Nei, M. Reliabilities of identifying positive selection by the branch-site and the site-prediction methods. *Proc. Natl Acad. Sci. USA* **106**, 6700–6705 (2009).

43. Hueffer, K., Govindasamy, L., Agbandje-McKenna, M. & Parrish, C. R. Combinations of two capsid regions controlling canine host range determine canine transferrin receptor binding by canine and feline parvoviruses. *J. Virol.* **77**, 10099–10105 (2003).

44. Sanjuan, R. Mutational fitness effects in RNA and single-stranded DNA viruses: common patterns revealed by site-directed mutagenesis studies. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **365**, 1975–1982 (2010).

45. da Silva, J., Coetzer, M., Nedellec, R., Pastore, C. & Mosier, D. E. Fitness epistasis and constraints on adaptation in a human immunodeficiency virus type 1 protein region. *Genetics* **185**, 293–303 (2010).
Examines genetic interactions and potential evolutionary pathways for HIV adaptation to CXC-chemokine receptor 4 by engineering amino acid mutations alone and in combination and measuring their effects on viral fitness.

46. Parrish, C. R. *et al.* Cross-species virus transmission and the emergence of new epidemic diseases. *Microbiol. Mol. Biol. Rev.* **72**, 457–470 (2008).
Review of ecological factors and viral adaptations that can lead to cross-species transmission or host jumps.

47. Wang, L. F. & Eaton, B. T. in *Wildlife and Emerging Zoonotic Diseases: The Biology, Circumstances and Consequences of Cross-Species Transmission*. 325–344 (Springer-Verlag Berlin, Berlin, 2007).

48. Reluga, T., Meza, R., Walton, D. B. & Galvani, A. P. Reservoir interactions and disease emergence. *Theoretical Population Biology* **72**, 400–408 (2007).

49. Pepin, K. M., Samuel, M. A. & Wichman, H. A. Variable pleiotropic effects from mutations at the same locus hamper prediction of fitness from a fitness component. *Genetics* **172**, 2047–2056 (2006).

50. Gilchrist, M. A. & Sasaki, A. Modeling host-parasite coevolution: a nested approach based on mechanistic models. *J. Theor. Biol.* **218**, 289–308 (2002).

51. Grenfell, B. T. *et al.* Unifying the epidemiological and evolutionary dynamics of pathogens. *Science* **303**, 327–332 (2004).

52. Volkov, I., Pepin, K. M., Lloyd-Smith, J. O., Banavar, J. R. & Grenfell, B. T. Synthesizing within-host and population-level selective pressures on viral populations: the impact of adaptive immunity on viral immune escape. *J. R. Soc. Interface* **6**, 1311–1318 (2010).

53. Coombs, D., Gilchrist, M. A. & Ball, C. L. Evaluating the importance of within- and between-host selection pressures on the evolution of chronic pathogens. *Theor. Popul. Biol.* **72**, 576–591 (2007).

54. Gilchrist, M. A. & Coombs, D. Evolution of virulence: interdependence, constraints, and selection using nested models. *Theor. Popul. Biol.* **69**, 145–153 (2006).

55. Diekmann, O., Heesterbeek, J. A. & Metz, J. A. On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations. *J. Math. Biol.* **28**, 365–382 (1990).

56. Roberts, M. G. The pluses and minuses of R_0. *J. R. Soc. Interface* **4**, 949–961 (2007).

57. Day, T. Virulence evolution and the timing of disease life-history events. *Trends Ecol. Evol.* **18**, 113–118 (2003).

58. Gilchrist, M. A., Coombs, D. & Perelson, A. S. Optimizing within-host viral fitness: infected cell lifespan and virion production rate. *J. Theor. Biol.* **229**, 281–288 (2004).

59. Anderson, R. M. & May, R. M. *Infectious Diseases of Humans: Dynamics and Control* (Oxford University Press, Oxford, 1991).

60. Lloyd-Smith, J. O. *et al.* Epidemic dynamics at the human-animal interface. *Science* **326**, 1362–1367 (2009).
A comprehensive review of the population dynamics of zoonotic infections.

61. Antia, R., Regoes, R. R., Koella, J. C. & Bergstrom, C. T. The role of evolution in the emergence of infectious diseases. *Nature* **426**, 658–661 (2003).
Modelling study examining whether host jumps occur by adaptive mutation in the new host.

62. Arinaminpathy, N. & McLean, A. R. Evolution and emergence of novel human infections. *Proc. Biol. Sci.* **276**, 3937–3943 (2009).
Discusses epidemiological signs of pathogen adaptation to a new host. Uses a mathematical model that includes within-host evolution and transmission to explore how epidemiological data can be used to monitor the risk of emergence.

63. Farrington, C. P., Kanaan, M. N. & Gay, N. J. Branching process models for surveillance of infectious diseases controlled by mass vaccination. *Biostatistics* **4**, 279–295 (2003).

64. Lloyd-Smith, J. O., Schreiber, S. J., Kopp, P. E. & Getz, W. M. Superspreading and the effect of individual variation on disease emergence. *Nature* **438**, 355–359 (2005).

65. World Health Organization. Communicable disease profile for Democratic Republic of the Congo (WHO, Kinshasa, 2005).

66. [No authors listed]. Cumulative number of confirmed human cases of avian influenza A/(H5N1) reported to WHO. *WHO* [online], http://www.who.int/csr/disease/avian_influenza/country/cases_table_2010_05_06/en/index.html (2010).

67. Dufour, A. in *Waterborne Zoonoses: Identification, Causes and Control*. (eds Cotruvo, J. A. *et al.*) (IWA Publishing, London, 2004).

68. 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).

69. Benmayor, R., Hodgson, D. J., Perron, G. G. & Buckling, A. Host mixing and disease emergence. *Curr. Biol.* **19**, 764–767 (2009).

70. Holmes, E. C. & Grenfell, B. T. Discovering the phylodynamics of RNA viruses. *PLoS Comput. Biol.* **5**, e1000505 (2009).
Opinion on the importance of linking ecological, epidemiological and evolutionary data for studying viral evolution.

71. Gabriel, G. *et al.* The viral polymerase mediates adaptation of an avian influenza virus to a mammalian host. *Proc. Natl Acad. Sci. USA* **102**, 18590–18595 (2005).
Compares two viral strains, one adapted in birds and one in mice, and identifies specific mutations in the viral polymerase of influenza A strains that enhance polymerase activity and increase virulence in mammalian hosts.

72. Shinya, K., Watanabe, S., Ito, T., Kasai, N. & Kawaoka, Y. Adaptation of an H7N7 equine influenza A virus in mice. *J. Gen. Virol.* **88**, 547–535 (2007).

73. Herfst, S. *et al.* Introduction of virulence markers in PB2 of pandemic swine-origin influenza virus does not result in enhanced virulence or transmission. *J. Virol.* **84**, 3752–3758 (2010).
Engineers putative markers of influenza A virulence into a prototype strain of swine flu influenza (S-OIV). Measures replication in cell culture, virulence in mice and ferrets and aerosol transmission in ferrets and finds that the markers of influenza A virulence do not increase virulence in the swine flu strain.

74. Zhu, H. C. *et al.* Substitution of lysine at 627 position in PB2 protein does not change virulence of the 2009 pandemic H1N1 virus in mice. *Virology* **401**, 1–5 (2010).

75. Pepin, K. M. & Wichman, H. A. Variable epistatic effects between mutations at host recognition sites in phiX174 bacteriophage. *Evolution* **61**, 1710–1724 (2007).

76. Cuevas, J. M., Moya, A. & Sanjuan, R. A genetic background with low mutational robustness is associated with increased adaptability to a novel host in an RNA virus. *J. Evol. Biol.* **22**, 2041–2048 (2009).

77. Webby, R., Hoffmann, E. & Webster, R. Molecular constraints to interspecies transmission of viral pathogens. *Nature Med.* **10**, S77–S81 (2004).

78. Szretter, K. J. *et al.* Early control of H5N1 influenza virus replication by the type I interferon response in mice. *J. Virol.* **83**, 5825–5834 (2009).

79. Pulliam, J. R. & Dushoff, J. Ability to replicate in the cytoplasm predicts zoonotic transmission of livestock viruses. *J. Infect. Dis.* **199**, 565–568 (2009).
Statistical analysis of molecular viral traits that correlate with zoonotic emergence.

80. Sorrell, E. M., Wan, H., Araya, Y., Song, H. & Perez, D. R. Minimal molecular constraints for respiratory droplet transmission of an avian-human H9N2 influenza A virus. *Proc. Natl Acad. Sci. USA* **106**, 7565–7570 (2009).
   Investigates the genetic basis of transmission from avian to mammalian hosts. They construct a reassorted avian–human influenza strain (H9N2–H3N2) and adapt this virus to ferrets. They find that the virus gained aerosol transmission ability in ferrets by few genetic changes that also altered the antigenic profile .
81. Yassine, H. M., Al-Natour, M. Q., Lee, C. W. & Saif, Y. M. Interspecies and intraspecies transmission of triple reassortant H3N2 influenza A viruses. *Virol. J.* **4**, 129 (2007).
   Study of the cross-species transmission potential of four reassorted strains of influenza A virus originating from different host species (turkey and swine). They identify a putative marker of cross-species transmission and measure replication and transmission among turkeys, chickens and ducks.
82. Kim, M. C. *et al.* Pathogenicity and transmission studies of H7N7 avian influenza virus isolated from feces of magpie origin in chickens and magpie. *Vet. Microbiol.* **141**, 268–274 (2010).
83. Duffy, S., Turner, P. E. & Burch, C. L. Pleiotropic costs of niche expansion in the RNA bacteriophage Phi 6. *Genetics* **172**, 751–757 (2006).
84. Remold, S. K., Rambaut, A. & Turner, P. E. Evolutionary genomics of host adaptation in vesicular stomatitis virus. *Mol. Biol. Evol.* **25**, 1138–1147 (2008).
85. Smith-Tsurkan, S. D., Wilke, C. O. & Novella, I. S. Incongruent fitness landscapes, not trade-offs, dominate the adaptation of VSV to novel host types. *J. Gen. Virol.* **9**, 1484–1493 (2010).
86. Ciota, A. T. *et al.* Experimental passage of St. Louis encephalitis virus *in vivo* in mosquitoes and chickens reveals evolutionarily significant virus characteristics. *PLoS One* **4**, e7876 (2009).
87. Coffey, L. L. *et al.* Arbovirus evolution *in vivo* is constrained by host alternation. *Proc. Natl. Acad. Sci. USA* **105**, 6970–6975 (2008).
   Compares the adaptive potentials of epizootic and endemic strains of VEEV by experimental evolution under three passage conditions: only mice, only mosquitoes and alternating between the two, and finds that viruses passaged on alternating host types do not show increased fitness.
88. Vasilakis, N. *et al.* Mosquitoes put the brake on arbovirus evolution: experimental evolution reveals slower mutation accumulation in mosquito than vertebrate cells. *Plos Pathog.* **5**, e1000467 (2009).
89. Weaver, S. C., Brault, A. C., Kang, W. & Holland, J. J. Genetic and fitness changes accompanying adaptation of an arbovirus to vertebrate and invertebrate cells. *J. Virol.* **73**, 4316–4326 (1999).
90. Itoh, Y. *et al. In vitro* and *in vivo* characterization of new swine-origin H1N1 influenza viruses. *Nature* **460**, 1021–1025 (2009).
91. Perkins, L. E. & Swayne, D. E. Comparative susceptibility of selected avian and mammalian species to a Hong Kong-origin H5N1 high-pathogenicity avian influenza virus. *Avian Dis.* **47**, 956–967 (2003).
92. Manzoor, R. *et al.* PB2 protein of a highly pathogenic avian influenza virus strain A/chicken/

Yamaguchi/7/2004 (H5N1) determines its replication potential in pigs. *J. Virol.* **83**, 1572–1578 (2009).
93. [No authors listed]. Preliminary review of D222G amino acid substitution in the haemagglutinin of pandemic influenza A (H221N1) 2009 viruses. *Wkly Epidemiol. Rec.* **85**, 21–22 (2010).
94. Kilander, A., Rykkvin, R., Dudman, S. G. & Hungnes, O. Observed association between the HA1 mutation D222G in the 2009 pandemic influenza A(H221N1) virus and severe clinical outcome, Norway 2009–2010. *Euro Surveill.* **15**, 19498 (2010).
95. Shinya, K. *et al.* Avian flu: influenza virus receptors in the human airway. *Nature* **440**, 435–436 (2006).
96. Brown, J. D., Stallknecht, D. E., Beck, J. R., Suarez, D. L. & Swayne, D. E. Susceptibility of North American ducks and gulls to H5N1 highly pathogenic avian influenza viruses. *Emerg. Infect. Dis.* **12**, 1663–1670 (2006).
97. Keawcharoen, J. *et al.* Wild ducks as long-distance vectors of highly pathogenic avian influenza virus. *Emerging Infect. Dis.* **14**, 600–607 (2008).
98. Brault, A. C., Powers, A. M., Holmes, E. C., Woelk, C. H. & Weaver, S. C. Positively charged amino acid substitutions in the e2 envelope glycoprotein are associated with the emergence of Venezuelan equine encephalitis virus. *J. Virol.* **76**, 1718–1730 (2002).
99. Greene, I. P. *et al.* Envelope glycoprotein mutations mediate equine amplification and virulence of epizootic Venezuelan equine encephalitis virus. *J. Virol.* **79**, 9128–9133 (2005).
100. Anishchenko, M. *et al.* Venezuelan encephalitis emergence mediated by a phylogenetically predicted viral mutation. *Proc. Natl Acad. Sci. USA* **103**, 4994–4999 (2006).
101. Bull, J. J. & Molineux, I. J. Predicting evolution from genomics: experimental evolution of bacteriophage T7. *Heredity* **100**, 453–463 (2008).
102. Elena, S. F. & Lenski, R. E. Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. *Nature Rev. Genet.* **4**, 457–469 (2003).
103. Elena, S. F. & Sanjuan, R. Virus evolution: Insights from an experimental approach. *Annu. Rev. Ecol. Evol. Syst.* **38**, 27–52 (2007).
104. Crill, W. D., Wichman, H. A. & Bull., J. J. Evolutionary reversals during viral adaptation to alternating hosts. *Genetics* **154**, 27–37 (2000).
105. Wichman, H. A., Badgett, M. R., Scott, L. A., Boulianne, C. M. & Bull., J. J. Different trajectories of parallel evolution during viral adaptation. *Science* **285**, 422–424 (1999).
106. Wichman, H. A., Scott, L. A., Yarber, C. D. & Bull., J. J. Experimental evolution recapitulates natural evolution. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **355**, 1677–1684 (2000).
107. Brown, E. G., Liu, H., Kit, L. C., Baird, S. & Nesrallah, M. Pattern of mutation in the genome of influenza A virus on adaptation to increased virulence in the mouse lung: identification of functional themes. *Proc. Natl Acad. Sci. USA* **98**, 6883–6888 (2001).
108. Keleta, L., Ibricevic, A., Bovin, N. V., Brody, S. L. & Brown, E. G. Experimental evolution of human influenza virus H3 hemagglutinin in the mouse lung identifies adaptive regions in HA1 and HA2. *J. Virol.* **82**, 11599–11608 (2008).
109. Wu, R. *et al.* Multiple amino acid substitutions are involved in the adaptation of H9N2 avian influenza virus to mice. *Vet. Microbiol.* **138**, 85–91 (2009).

110. Narasaraju, T. *et al.* Adaptation of human influenza H3N2 virus in a mouse pneumonitis model: insights into viral virulence, tissue tropism and host pathogenesis. *Microbes Infect.* **11**, 2–11 (2009).
111. Lemey, P., Rambaut, A., Drummond, A. J. & Suchard, M. A. Bayesian phylogeography finds its roots. *PLoS Comput. Biol.* **5**, e1000520 (2009).
112. Janies, D. A. *et al.* The Supramap project: linking pathogen genomes with geography to fight emergent infectious diseases. *Cladistics* **2010**, 1–6 (2010).
113. Liechti, R. *et al.* OpenFluDB, a database for human and animal influenza virus. *Database* **6**, baq004 (2010).
   Development of an application, Supramap (http://supramap.osu.edu), that incorporates the dynamics of pathogen epidemiology and evolution into a geographical context to test genetic markers for host-jump potential and to estimate their global transmission patterns.
114. Bush, R. M. Influenza as a model system for studying the cross-species transfer and evolution of the SARS coronavirus. *Philos. Trans. R. Soc. Lond., B Biol. Sci.* **359**, 1067–1073 (2004).
115. Qu, X. X. *et al.* Identification of two critical amino acid residues of the severe acute respiratory syndrome coronavirus spike protein for its variation in zoonotic tropism transition via a double substitution strategy. *J. Biol. Chem.* **280**, 29588–29595 (2005).
116. Song, M. S. *et al.* Ecology of H3 avian influenza viruses in Korea and assessment of their pathogenic potentials. *J. Gen. Virol.* **89**, 949–957 (2008).
117. Hoelzer, K., Shackelton, L. A., Holmes, E. C. & Parrish, C. R. Within-host genetic diversity of endemic and emerging parvoviruses of dogs and cats. *J. Virol.* **82**, 11096–11105 (2008).
   Investigates within-host genetic diversity produced by canine parvovirus and feline panleukopenia virus during natural and experimental infections.

FURTHER INFORMATION
Kim Pepin's homepage:
http://www.cidd.psu.edu/people/kmp29
ProMED mail post on influenza A H1N1 pandemic:
www.promedmail.org/pls/apex/f?p=2400:1001:5573072427
14929::NO::F2400_P1001_BACK_PAGE,F2400_P1001_PUB_
MAIL_ID:1000,84555

SUPPLEMENTARY INFORMATION
See online article: S1 (table) | S2 (box) | S3 (figure) | S4 (figure) | S5 (figure) | S6 (figure) | S7 (figure)

ALL LINKS ARE ACTIVE IN THE ONLINE PDF

Author Bios

Kim M. Pepin received a B.Sc. in ecology at the University of British Columbia, Canada, and her Ph.D. in biology at the University of Idaho, USA, where she examined genetic interactions and adaptation in viruses by mutagenesis and experimental evolution. During a short postdoctoral position at New Mexico State University, USA, she studied the role of intracellular competition in dengue virus emergence. She is currently a research associate at Pennsylvania State University, USA, where she develops dynamical models of viral emergence and Evolution, and statistical models of climatic drivers of disease.

Sandra Lass received a diploma in biology from Christian-Albrechts-University in Kiel, Germany, and a Ph.D. in environmental sciences from the Swiss Federal Institute of Technology in Zürich (ETHZ), Switzerland, studying inducible defences and chemical communication in a predator–prey system. Her postdoctoral projects at the Universities of Fribourg and Basel, Switzerland, and at Pennsylvania State University focused on the evolutionary ecology of host–parasite interactions and the epidemiology of different micro- and macroparasites using experimental approaches in the field and the laboratory.

Juliet R.C. Pulliam is a research and policy for infectious disease dynamics (RAPIDD) programme fellow at the Fogarty International Center, National Institutes of Health, Bethesda, USA. She received a Ph.D. in ecology and evolutionary biology from Princeton University, New Jersey, USA, in 2007. Her research focuses on the determinants and dynamics of viral host jumps and on the interactions between human, wildlife and domestic animal health.

Andrew R. Read is Professor of Biology and Entomology at Pennsylvania State University, working on the ecology and evolution of host–parasite interactions, and particularly the adaptation of disease-causing organisms to drugs, vaccines and insecticides. At his age, details of his graduate studies are irrelevant.

James O. Lloyd-Smith is Assistant Professor at University of California Los Angeles, USA, and holds the De Logi Chair in Biological Sciences. His research programme addresses the ecological and evolutionary dynamics of zoonotic diseases, including monkeypox, severe acute respiratory syndrome-coronavirus, influenza and leptospirosis. He received a Ph.D. in biophysics from the University of California, Berkeley, USA, in 2005 for his study of disease transmission dynamics in heterogeneous populations, and carried out postdoctoral studies at Pennsylvania State University.

On-line summary

- Viral host jumps can lead to major public health threats. The most recent pandemics were caused by viruses that were transmitted from animal reservoirs to humans, such as influenza A viruses and severe acute respiratory syndrome-coronavirus. Adaptation of the virus to the new host is often cited as the cause of such emergence.

- Distinguishing the genetic changes that are due to adaptation from those that are due to random events is hard in any biological context; virus host jumps are no exception. We present four different mechanisms by which viruses may emerge in a new host. Although all four mechanisms could produce the same genetic pattern in new hosts, only two are due to adaptation. We illustrate which data need to be collected to distinguish between the four mechanisms.

- Future risk of viral host jumps to humans could be assessed by genetic surveillance of viruses in reservoir hosts, but only when genetic adaptation is required for a host jump and when precursors of this adaptation can be detected.

- Bioinformatic analyses of surveillance data are key stepping stones for identifying putative genetic markers of viral adaptation from enormous pools of genetic data. Confirmation of which of these putative markers are due to adaptation requires experimental validation by using reverse genetics and host models from reservoir and new host species, and corroborating results with epidemiological and ecological data.

- Our review of the current literature on four well-studied viral host jumps shows that research on host-jump processes unfolds in four broad stages: virus sample collection and genetic analysis; experiments *in vitro* or in cell culture; *in vivo* experiments in model hosts; and *in vivo* experiments in natural hosts. We evaluate the issues in using these types of data for validating adaptive hypotheses, and identify opportunities to collect further data that would enable better discrimination among emergence mechanisms.

- A detailed understanding of viral host jumps and the assessment of future risk requires multidisciplinary research efforts with input from field ecologists, microbiologists, immunologists, epidemiologists, bioinformaticians and evolutionary biologists, and the use of use of diverse approaches (field sampling, laboratory experiments, data analysis and mathematical modelling).