

UCLA

UCLA Previously Published Works

Title

Improving risk assessment of the emergence of novel influenza A viruses by incorporating environmental surveillance.

Permalink

<https://escholarship.org/uc/item/01q7w1c2>

Journal

Philosophical Transactions of the Royal Society B: Biological Sciences, 374(1782)

Authors

Pepin, Kim
Hopken, Matthew
Shriner, Susan
et al.

Publication Date

2019-09-30

DOI

10.1098/rstb.2018.0346

Peer reviewed

Opinion piece



Cite this article: Pepin KM, Hopken MW, Shriner SA, Spackman E, Abdo Z, Parrish C, Riley S, Lloyd-Smith JO, Piaggio AJ. 2019 Improving risk assessment of the emergence of novel influenza A viruses by incorporating environmental surveillance. *Phil. Trans. R. Soc. B* **374**: 20180346.
<http://dx.doi.org/10.1098/rstb.2018.0346>

Accepted: 12 March 2019

One contribution of 20 to a theme issue 'Dynamic and integrative approaches to understanding pathogen spillover'.

Subject Areas:

computational biology, ecology, genomics

Keywords:

influenza A, surveillance, reassortment, co-infection, environmental transmission, single-virus sequencing

Author for correspondence:

Kim M. Pepin
e-mail: kim.m.pepin@usda.gov

Improving risk assessment of the emergence of novel influenza A viruses by incorporating environmental surveillance

Kim M. Pepin¹, Matthew W. Hopken^{1,2}, Susan A. Shriner¹, Erica Spackman³, Zaid Abdo², Colin Parrish⁴, Steven Riley⁵, James O. Lloyd-Smith^{6,7} and Antoinette J. Piaggio¹

¹National Wildlife Research Center, USDA-APHIS, Fort Collins, CO 80521, USA

²Colorado State University, Fort Collins, CO 80523, USA

³Exotic and Emerging Avian Viral Diseases Research, USDA-ARS, Athens, GA 30605, USA

⁴Baker Institute for Animal Health, Department of Microbiology and Immunology, Cornell University, Ithaca, NY 14853, USA

⁵MRC Centre for Global Infectious Disease Analysis, Imperial College, London, SW7 2AZ, UK

⁶UCLA, Los Angeles, CA 90095, USA

⁷Department of Ecology and Evolutionary Biology, Fogarty International Center, National Institutes of Health, Bethesda MD 20892, USA

KMP, 0000-0002-9931-8312; SAS, 0000-0003-0349-7182; CP, 0000-0002-1836-6655

Reassortment is an evolutionary mechanism by which influenza A viruses (IAV) generate genetic novelty. Reassortment is an important driver of host jumps and is widespread according to retrospective surveillance studies. However, predicting the epidemiological risk of reassortant emergence in novel hosts from surveillance data remains challenging. IAV strains persist and co-occur in the environment, promoting co-infection during environmental transmission. These conditions offer opportunity to understand reassortant emergence in reservoir and spillover hosts. Specifically, environmental RNA could provide rich information for understanding the evolutionary ecology of segmented viruses, and transform our ability to quantify epidemiological risk to spillover hosts. However, significant challenges with recovering and interpreting genomic RNA from the environment have impeded progress towards predicting reassortant emergence from environmental surveillance data. We discuss how the fields of genomics, experimental ecology and epidemiological modelling are well positioned to address these challenges. Coupling quantitative disease models and natural transmission studies with new molecular technologies, such as deep-mutational scanning and single-virus sequencing of environmental samples, should dramatically improve our understanding of viral co-occurrence and reassortment. We define observable risk metrics for emerging molecular technologies and propose a conceptual research framework for improving accuracy and efficiency of risk prediction.

This article is part of the theme issue 'Dynamic and integrative approaches to understanding pathogen spillover'.

1. Introduction

Reassortment is a prominent mechanism by which segmented viruses produce genetic variation. Reassortment occurs when genetic segments from different co-infecting virions within the same cell are packaged together, generating a novel strain. Some of the most devastating outbreaks caused by influenza A virus (IAV; a segmented RNA virus) are believed to have been driven by reassortants [1,2], yet predicting reassortment remains elusive.

For avian IAVs, new reassortants can be generated in natural reservoir hosts (wild water birds), where multiple strains of virus circulate naturally, or following spillover from a reservoir host to an infected spillover host (e.g. poultry, pigs or humans). Although reassortants with increased fitness may be

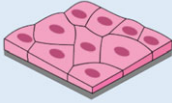


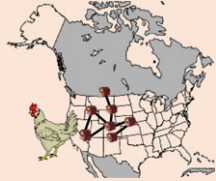
A	B	C	D
<p>1) scales of reassortment</p> <p>within and among cells</p> 	<p>within an environmental reservoir-host system</p> 	<p>among water-host systems (along flyways)</p> 	<p>spillover and emergence in poultry</p> 
<p>2) processes</p> <ul style="list-style-type: none"> • multiplicity of infection • timing that strains co-infect • spatial structure of cells • immune response • viability and fitness of reassortants 	<ul style="list-style-type: none"> • A processes and... <ul style="list-style-type: none"> ◦ environmental transmission <ul style="list-style-type: none"> ◦ virion concentration (host density, viral shedding, water body characteristics) ◦ viral persistence from abiotic factors ◦ host susceptibility ◦ viral uptake rate 	<ul style="list-style-type: none"> • A and B processes and... <ul style="list-style-type: none"> • migration pathways, connectivity and patch residence-time • host density in space and time • host species composition • immunity profile/exposure history 	<ul style="list-style-type: none"> • A, B and C processes and... <ul style="list-style-type: none"> • poultry density in space and time • contact with reservoir populations • biosecurity • spillover force of infection • connectivity between operations • susceptibility/adaptability to reassortants
<p>3) risk metrics</p> <p><i>de novo</i> reassortment rate: change in frequency of new reassortants between the exposure dose and host-level samples</p>	<p>diversity dynamics: number, frequency and relatedness of unique genotypes in the dominant fraction in host-specific samples and water samples <i>over time</i></p> <p>pathogen pressure: concentration (water samples) or prevalence (host-specific samples) of all viable viruses <i>over time</i></p> <p>dominant-strain phenotypes: an absolute score for replication competence for each dominant strain in the sample (within-host fitness)</p>	<p>diversity dynamics, pathogen pressure and dominant-strain phenotypes <i>in reservoir hosts at multiple sites over time along flyways</i></p>	<p>diversity dynamics, pathogen pressure and dominant-strain phenotypes <i>in sentinel poultry at multiple sites over time along flyways</i></p>

Figure 1. Scales at which reassortment occurs (1). In order to predict the emergence of reassortants in poultry (D), reassortment and transmission processes (2) need to be understood at each scale. We propose risk metrics that can be measured at each scale (3) for development of predictive models of epidemiological risk (figure 2).

produced less commonly than deleterious ones [3], reassortants are observed frequently at the population scale among natural virus isolates [4,5] suggesting that reassortment is an important source of adaptation. Both co-infections (a necessary precursor to reassortment) and reassortants are often detected retrospectively during surveys of wild waterfowl [6–11]. A comprehensive analysis of publicly available genomic data from avian and mammalian host species found that a substantial proportion of genomes (3%–646 total) were first time reports of novel reassortants [4]. Yet, emergence of novel IAV reassortants remains unpredictable and depends on host and virus populations [4]. One model estimated a 3-year cycle of emergence and replacement of avian IAV reassortants [9], but the highly pathogenic goose/Guangdong/96 clade 2.3.4.4 H5 appears to be persisting and contributing to the ongoing generation of novel reassortants [12].

In wild birds, IAV transmission via the environment is a major mechanism facilitating co-infection and host contact with pathogens [13]. Environmental reservoirs, such as ponds, can accumulate virus particles shed in faeces by many individuals, increasing the co-occurrence and exposure to different viral strains [13,14], and hence the likelihood of co-infection. By extension, reassortment events that have occurred in hosts using the same water body could be revealed by sequencing individual whole viruses from the environment over time and conducting genomic analyses, thus increasing detection capabilities for the reassortant combinations present. Environmental reservoirs also increase contact between species—i.e. species that may never contact each other directly drink or feed from the same water source. They additionally increase the infectious period (time frame during which infectious particles can be transmitted following infection) and expand the spatial

availability of viruses (locations where infectious particles can be acquired), thus increasing the likelihood that novel host species will make contact with a variety of different viruses. Therefore, environmental transmission can overcome barriers of spillover by increasing interspecies contact with viable pathogens, and facilitating large jumps in viral diversity and fitness by allowing co-infection and reassortment.

One of the most important public/livestock health challenges is to predict the emergence of novel strains from surveillance systems. Predicting this epidemiological risk for viruses that reassort frequently, such as IAV, requires an understanding of the viral evolutionary ecology in reservoir and spillover hosts and its significance in driving viral fitness across scales [15]. Factors involved include predicting the rate with which new reassortants are generated in cell tissues and individual hosts (*de novo* reassortment), estimating the probability that a new reassortant will become predominant in reservoir-host populations locally and regionally, and the probability of spillover and transmission in spillover-host populations (figure 1–1, 1-2).

We first review what is known from experimental studies of within-host reassortment rates, environmental transmission of IAV and assaying strain-specific fitness. We then outline emerging technologies in environmental sampling, phenotyping, genomics and bioinformatics that could improve detection and assessment of reassortants in environmental surveillance samples. We describe how emerging technology could allow efficient observation of reassortant dynamics in host populations, by incorporating environmental surveillance. Lastly, we define risk metrics (e.g. figure 1-3) that can be measured using these molecular technologies and describe conceptual strategies for using those metrics to inform quantitative frameworks designed for prediction of epidemiological risk from environmental surveillance samples.

2. Knowledge from experimentation

Experiments have provided foundational knowledge on de novo reassortment rates and combinatorial constraints within hosts and cell tissues, yet predicting within-host rates of reassortment from specific factors has just begun [16,17]. Similarly, quantifying reassortment rates within an environmental reservoir-host system (e.g. figure 1b) has only rarely been addressed [11]. Below, we describe experimental approaches that could be used to measure de novo reassortment rates within hosts, and transmission of reassortants in an environmental reservoir-host system, allowing improved prediction of epidemiological risk from surveillance data.

(a) Co-infection and reassortment processes

Significant progress has been made in understanding within-host reassortment processes in spillover host models (i.e. poultry and mammals) using experimental inoculation techniques [16,17]. Rates of de novo reassortment within hosts can be determined by co-infecting single hosts with known parental genotypes and measuring the proportion of shed virions that are reassortants. Such experiments have helped to develop assays for studying reassortment rates within hosts, and demonstrated substantial and highly variable rates of reassortment within-host individuals ([18–22]; e.g. 8.7% of viruses recovered from ferrets co-infected with human and avian strains were reassortants [18], to approximately 50% of viruses recovered from co-infections of chickens with other chicken-derived subtypes [19], to 86% in swine and guinea pigs co-infected with swine or human strains [20,21]). The differences are thought to be owing to functional incompatibilities between viral genome segments or differing receptor binding specificities or other host restrictions, leading to low levels of co-infection in the same cells.

Within-host studies have determined that transmission route, timing of exposures, dose, strains and strain competition all can influence co-infection and reassortment rates [16,17,21]. Higher doses lead to higher rates of co-infected cells [21]. When exposure with two different strains is offset by a short period of time (12 h), co-infection and reassortment rates are higher than when the host is exposed to the two strains simultaneously [21]. Delayed introduction of the second strain likely increases co-infection rates because it allows the first strain time to infect many cells, thus maximizing co-infection potential by the second strain [21]. By contrast, exposures with longer time lags can reduce co-infection because of super-infection exclusion mechanisms that occur later post-infection (i.e. host innate immune function, and viral processes such as receptor interference) [17] and because death of infected cells occurs. As expected, effects of co-infection time lags on reassortment rates can differ depending on host-strain combinations and experimental design [19]. Importantly, these advances in our understanding of within-host reassortment provide a foundation for predictive models to account for co-infection and reassortment processes mechanistically in epidemiological models. An important gap for experimentation is that little work has focused on understanding reassortment in reservoir hosts under natural transmission conditions (figure 1A); how the lessons learned from animal models apply to wild birds infected by environmental transmission remains unknown.

(b) Environmental transmission

Models suggest that environmental reservoirs play an important role in the ecology of IAVs because they allow infection between hosts that infrequently come into direct contact, such as migratory wild waterfowl on breeding grounds [13,14,23,24]. Experimental systems have shown the occurrence of environmental transmission of IAVs among reservoir hosts and from reservoir hosts to spillover hosts [25–28]. However, the role of indirect transmission through the environment relative to other transmission mechanisms is not well understood empirically, which is important for informing models aimed at predicting epidemiological risk. One study identified a correlation between host population density and environmental levels of IAV [29]. Environmental transmission experiments combined with molecular characterization and disease-dynamic models would provide powerful tools for improving our quantitative knowledge of environmental transmission, co-infection and reassortment processes.

One key driver of environmental transmission is persistence of viral viability after shedding into the environment. There is substantial variation in environmental durability of IAVs (between 1 and 600 days) depending on abiotic factors, especially temperature, salinity, pH and mineral content of the medium [30–35] and viral genotype [33,35], including under natural conditions [36]. While some information is available for water and bird faeces, studies with soil [37], sediments [29,33,38] and plant material are few. Variation in viral environmental persistence could be an important component of fitness [39] and should be included in models predicting the ecological and evolutionary dynamics of IAVs. A final important knowledge gap is understanding the relationship between environmental virus concentrations and infectivity.

(c) Strain-specific fitness

Experimental studies are also helpful for evaluating phenotypes (e.g. measures of within-host replication fitness) of parental and reassortant viruses, a predictor of co-infection and reassortment rates within hosts (figure 2a). Fitness of reassortant strains is a crucial filter on emergence risk. Fitness studies on human IAV strains provide some insights where deleterious reassortment events were characterized by fewer reassortant progeny and more limited replication efficiency in spillover host cells [3,40]. Also poorly understood is the connection between within-host fitness and transmissibility (e.g. [41]).

Comparative genomics of IAV genomes from wild bird and poultry samples identify putative genetic constellations that can readily produce reassortants with high fitness in reservoir and spillover host species ('high risk' markers). These could be tested for potential fitness effects *in vitro* with deep-mutational scanning approaches [42] to efficiently phenotype a large number of strains for which the genotype is known. Then putative 'high risk' markers can be identified by statistically analysing the relationship between genotype and phenotype. Using synthetically engineered or naturally isolated genotypes in *in vivo* experimental infection and transmission experiments could produce estimates of fitness for particular genotypes or molecular markers in a given host system (e.g. see techniques in [36]). At present, deep-mutational scanning and *in vivo* techniques are too time-intensive and costly to incorporate in large-scale surveillance systems, but they are feasible for experimental settings and

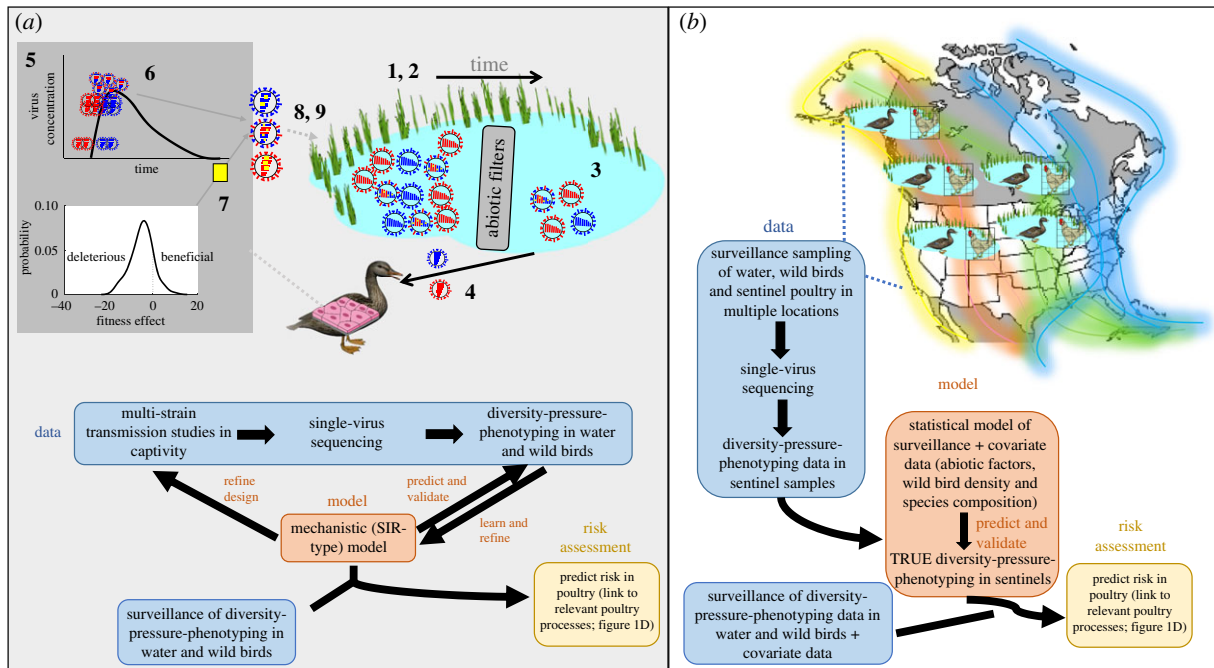


Figure 2. Approaches to predicting epidemiological risk in an environmental reservoir-host system. (a) Bottom-up approach. Develop a mechanistic disease transmission model of the evolutionary and ecological processes (bolded numbers) involved in reassortment. Design controlled environmental transmission studies with multiple strains and hosts. Then, iteratively fit and validate the mechanistic model and refine models and experiments based on learning. Top part of the plot shows processes that need to be considered in the mechanistic model. Note that there are additional within-host processes (not shown here) such as super-infection exclusion, within cell compartmentalization, spatial heterogeneity of target cells and others that can also affect the likelihood of reassortment. Numbers represent the following processes: (1) some viruses degrade owing to environmental conditions, (2) selection for persistence, (3) viruses that persist can be transmitted, (4) more than one virus strain can infect a single host either by co- or super-infection, (5) viral particles propagate within-hosts, as determined by the initial dose and fitness of the inoculating strain(s), (6) co- or super-infection of individual cells and reassortment may occur as a function of multiplicity of infection in target cells, (7) mutations with fitness effects occur during within-host replication, (8) selection for within-host replication influences the bottleneck size at shedding into the environment (9). (b) Top-down approach. Develop a statistical model for predicting diversity-pressure-phenotypes. Collect surveillance data from water, wild birds and sentinel chickens in multiple areas. Fit the model to the data to predict true reassortant diversity-pressure-phenotypes metrics. For both bottom-up and top-down approaches, the first step is to develop predictive tools. Once the predictive tools are validated through research studies, they can be applied to surveillance data (e.g. figure 1D) to predict epidemiological risk in poultry.

thus hold promise for improving our understanding of viral evolutionary dynamics and emergence.

3. Prospects from genomic technology

A fundamental hurdle with studying reassortment using genetic data is knowing how different segments are distributed among virions. In this section, we describe challenges at each step, from acquiring genome sequences to detecting reassortants in a collection of genomes. We further describe cutting-edge advances that may make it possible to detect and understand reassortment from an environmental reservoir-host system (figure 1B). In particular, the rapidly growing field of environmental metagenomics provides a framework for capturing IAV diversity from the environment ([43,44]; figure 1B). When applied in tandem with population genomics analyses, environmental metagenomics could be a useful tool for inferring reassortment events and for understanding the role of the environment in co-infection risk and reassortant emergence (figure 1).

(a) Environmental sampling

A major challenge with environmental surveillance is capturing the strain diversity of IAVs in the environment [38]. Questions include: do you sample water or soils or both? How many samples per site? How deep do you sample? How many technical replicates? A gold standard for capturing,

isolating and identifying true IAV diversity from the environment is still not established [45]. Similarly, the effects of common environmental variables such as water turbidity, rates of dispersion and diffusion and spatial heterogeneity across water or other environments have not been examined, although they affect environmental DNA capture from aquatic organisms [46]. Because knowledge of how to sample the environment for target RNA is still in its infancy, pilot studies structured to optimize sampling protocols are critical (e.g. [47]).

Another major sampling challenge is RNA preservation [48]. Filtration to capture virions/RNA requires hours at a single site (e.g. 800 L) to achieve adequate probability of virus detection [49]. Decontamination of pumping equipment at each site is necessary, further increasing sampling time. While these methods provide a useful means of collecting environmental genetic sequences [46], methodological advancements in RNA preservation/collection techniques are needed to improve the efficiency of collecting high-quality RNA viromes from environmental samples.

(b) Generating sequence data from environmental samples

An initial field-based screen can rapidly prioritize localities for further sampling efforts. Technological advances will soon allow diagnostic PCR assays in the field, with results within hours to minutes. Companies such as Biomeme

(Philadelphia, Pennsylvania, USA) have developed field-based nucleic acid extraction kits (M1 Sample Prep Kits) and a hand held real-time PCR thermocycler that attaches to a smart phone (two3™ Real-Time PCR Thermocycler). Both of these have been deployed in the field to identify mosquito pools positive for RNA viruses in less than 2 h to target samples for more thorough sequencing [50].

New sequencing platforms (e.g. MinION from Oxford Nanopore Technologies) hold promise for field-based surveillance of IAV owing to their rapid library prep and run times, and elimination of PCR amplification. The MinION generates long reads that allow sequencing of entire gene segments in a single read, single-molecule sequencing and direct RNA sequencing without conversion to cDNA and PCR amplification [51,52]. For field-based pathogen surveillance, it is possible to run a sequencer from a smartphone (e.g. SmidgION from Oxford Nanopore Technology). All of these technologies currently lack high-throughput capacities, have cost-prohibitive library prep, low sensitivity and high error rates for application to broad-scale surveillance. Despite these hurdles, there is a continual push to improve mobile diagnostic capabilities [53,54], which will inevitably improve efficiency, accuracy and precision of obtaining IAV genomic data from environmental samples.

All contemporary sequencing platforms require a minimum concentration of input nucleic acids, and water samples tend to fall below this threshold, requiring enrichment of target IAV RNA [55]. Centrifugation following filtration has been effective at concentrating viral particles [56]. The gold standard for IAV enrichment is virus culture in either embryonated chicken eggs or Madin–Darby canine kidney cells. These methods work well for clinical samples, but there can be high costs and challenges with isolating the virus from environmental samples [57–59]. Additionally, using virus isolates to study reassortment can be counter-productive because of subtype and genomic selection bias in media used to propagate isolates. PCR is another method for enriching viral genomes, however, PCR can preferentially amplify specific strains and introduce nucleotide errors that can be difficult to distinguish from true, low-frequency virus variants. The ideal approach is direct sequencing of viral genomes without PCR (see below).

Another major challenge with viromic data from environmental samples is delineation of individual genomes—i.e. how do we distinguish potential reassortants from multiple parental genotypes or viral mixtures? A common method is to isolate individual genotypes using limiting dilution plaque assays and subsequent whole-genome sequencing. However, plaque assays can cause biased replication levels between multiple strains because of tissue tropism [60], *in vitro* strain mutation [61] and overlapping plaques [62], all which can obscure the delineation of genotypes. Serial dilution can reduce plaque overlap but is very time consuming.

The advent of single-virus sequencing techniques [63,64], similar to single-cell genome sequencing techniques of eukaryotic and bacterial genomes [65], holds promise for circumventing several potential biases from viral isolation techniques while providing sequence data from individual virions. Using flow cytometry to isolate virions has improved sequencing of single-virus genomes [63,66], but RNA input requirements for library prep are still too high to allow the sequencing of a single, non-enriched virus genome from

environmental samples. However, methods for conducting single-virus RNA sequencing on a Nanopore or similar single-molecule sequencing platform will undoubtedly be realized in the near future, providing a valuable method of delineating IAV genotypes that have not been subject to selective media.

(c) Assembling genes and genomes for reassortment analysis

The most widely used sequencing platforms generate short reads, which makes genome assembly challenging because there may not be enough overlap between reads to determine alignments with high enough certainty. A common first step is to remove non-target sequence reads. Environmental samples include nucleic acids from non-target organisms, for which sequenced genomes are not available. The effectiveness of filtering is thus reduced, which increases computing time for genome assembly. The two main approaches for viral genome assembly are reference-based mapping and *de novo* assembly. Mapping reads to a reference genome can simplify assembly, but reference genome choice is not trivial and IAV genomic diversity in wild birds is very high, which can bias results when trying to recover viral population diversity [67]. *De novo* assembly of segmented viral genomes is complicated because of incomplete and uneven coverage of different segments, which can lead to skewed results or failed assemblies [68]. A useful approach could be *de novo* assembly to build reference genomes directly from the sample reads then map all the reads back to this genome to recover the viral population diversity, infer haplotypes and estimate their frequencies [67,69,70]. The ultimate goal is to accurately assemble all genome variants in systems with high levels of diversity and determine dominant and low-frequency strains, which can greatly influence epidemiologic outcomes.

(d) Detecting reassortants

Classically reassortants are identified by sequence alignment, and if necessary by generating phylogenetic trees of each gene segment to look for incongruent clades. An isolate can be considered to be a reassortant if one or more gene segments have a significantly different position in a fully resolved phylogenetic tree relative to other gene segments [71]. However, in wild birds, IAV diversities are frequently very high with multiple strains co-circulating, such that both mixtures of parental sequences and reassortants may be quite prevalent, resulting in phylogenetic trees that do not differ significantly from trees generated at random [6]. To detect such reassortants, we must define a minimum degree of difference in the topologies (e.g. [4]). The chosen thresholds will impact inferred rates of reassortment in surveillance samples.

Another challenge of detecting reassortants is that some assumptions of phylogenetic methods are violated, as ecological and evolutionary processes occur simultaneously in viral populations. For example, phylogenetic trees assume ancestral alleles are extinct, evolutionary patterns are bifurcating and no reticulation events have occurred [72]. Even creative approaches to estimating the frequency of reassortants in phylogenetic trees, such as discrete trait mapping and Bayesian ancestral state reconstruction [73], do not

circumvent assumptions. Furthermore all eight gene segments need to be analysed separately, which is computationally burdensome. To better detect viral reassortants, we need to develop or apply approaches that align with viral evolutionary processes rather than applying theories and concepts developed for organisms with highly disparate biology [74].

Population genetic approaches may be more appropriate for identifying the frequency of reassortants in a set of genomes. Phylogenetic networks are methods that account for reticulate evolution and can identify reassortment events [11,72,75,76]. There is promise in using coalescent theory [77] and maximum-likelihood or Bayesian estimates of migration [78]. An additional approach could be from population genetic clustering algorithms (e.g. [79,80]), using hemagglutinin (HA) and neuraminidase (NA) subtypes as population delimiters and then using those as a framework for estimating the replacement and ‘gene flow’ of the internal segments. One final method could take advantage of recombination metrics by using concatenated IAV gene segments and treating the genomes as chromosomes. The exchange of gene segments will thus lead to estimates of recombination rates that can serve as a proxy for reassortment frequencies [76].

4. Approaches to predicting epidemiological risk: the way forward

(a) Observing reassortment across scales

Within individual hosts (figure 1A), the important risk metric is the *de novo* reassortment rate. Reassortment rates ideally would be measured using the genomic diversity (number of unique genotypes, frequency of each unique genotype, and relatedness between genotypes) in samples from individual hosts over time during an infection, where the first time point is the exposure dose of co-infecting viruses [22]. The frequency of reassortants in each sample relative to parental viruses at the first time point would give an empirical index of the host-level rate of reassortment. From the distribution of frequencies for all samples from an individual host over time, one could calculate the mean and variation of host-level reassortment rates. However, except in controlled field experiments the genomic diversity of the exposure dose will be unknown. As a proxy, genomic diversity of samples from the environment could be used.

Extending these ideas, the next scale is reassortment in an individual environmental reservoir-wild bird system (figure 1B). We propose three complementary risk metrics at this scale: the ‘diversity dynamics’ are captured by measuring the genomic diversity in the water and faecal samples from individual hosts over time; the ‘pathogen pressure’ [81] is the concentration of virions in water and their prevalence in the host population; and the ‘dominant-strain phenotypes’ is a rapid assessment of replicative fitness for strains that are most frequent in the surveillance samples. We refer to these three important risk metrics together as ‘diversity-pressure-phenotyping’: data that can be observed to understand and predict epidemiological risk. An important condition is that these data are collected over time at the same site (longitudinal sampling) in order to monitor the process of reassortants being generated and selected

within hosts, and then shed and selected for persistence in the environment (e.g. figure 2a). Linking the molecular markers that are identified in the ‘diversity’ data to phenotypes will provide a mechanistic foundation for diversity dynamics.

At the next level (figure 1C), diversity-pressure-phenotyping in water and hosts could be measured across space and time to surveil for reassortant emergence along flyways. Lastly, concurrent surveillance of diversity-pressure-phenotyping in sentinel poultry at sites where water-host systems are being surveilled would provide an ideal measure of predicted epidemiological risk for spillover and emergence of reassortants in poultry (figure 1D).

(b) Bottom-up approaches: disease-dynamic modelling

Within-host viral dynamics of IAV have now been reasonably well studied using experimental data [82,83], providing a strong foundation for developing predictive models of reassortment dynamics. Important factors determining within-host viral kinetics for IAV include target-cell availability [84], immune system factors [85], spatial distribution of target cells [86] and initial dose [87]. While these viral kinetic models have been developed using data from human infections [82,83], their insights and methods have begun to be applied to avian hosts [88]. The most important factor for reassortment is co-infection [16,17]. Thus, initial dose, target-cell availability and the distribution of target cells will play a primary role in predicting *de novo* reassortment rates. Once reassortment has occurred within a host, selection within the host and the size of the transmission bottleneck will determine whether a reassortant is shed into the environment [16] (figure 2a). The recent advances in understanding of intracellular and within-host processes that govern reassortment [16,17] provide conceptual and quantitative knowledge for models of within-host viral and reassortment dynamics. In the larger context, the reassortment processes and within-host viral dynamic models can be nested in between-host models that account for selection, transmission bottlenecks and environmental transmission processes (e.g. [39]) to develop a predictive framework that can be validated with experimental data (figure 2a).

Controlled experiments that allow transmission to occur naturally coupled with data-driven modelling frameworks (as above, figure 2a) together form a valuable stepping-stone toward inferring epidemiological risk from genetic samples. Controlled experiments are important for restricting the range of potential genotypes and environmental factors so that models can be designed to capture epidemiological behaviour based on a limited and known set of factors. In high-dimensional systems such as IAV, model-guided fieldwork [89]—i.e. iterative model development, prediction and experimentation (figure 2a)—can be applied for efficient experimental design and knowledge building. By this approach, models could be used to predict which processes are the most important to assess. Then experiments would test our knowledge of how ill-understood processes such as environmental transmission contribute to the observed outcome. For example, a poor-fitting model of diversity-pressure-phenotyping would suggest that there are ill-defined processes occurring in the system that are not well captured or are absent in the model. Then the model and experimental design could be adjusted to elucidate these ill-understood mechanisms. This concept is not new [89] but it

is surprisingly underused for designing empirical studies of complex ecological and epidemiological systems with ample process uncertainty. The iterative approach of hypothesis generation → model building and prediction → data collection → model evaluation aims to reduce process uncertainty by targeted learning. Thus, it is a powerful approach for identifying important gaps in our ability to predict reassortment in environmental reservoir-host systems (figure 1B), and to design effective experiments that will best inform the identified gaps. Once this type of mechanistic epidemiological model is developed and validated using experimental data from controlled studies with natural transmission, it could be used to predict epidemiological risk from surveillance of diversity-pressure-phenotypes data in environmental reservoir-host systems.

As an illustration of how these cross-scale processes could be explored empirically, we describe a possible experimental study to measure diversity-pressure-phenotypes and resulting infection risks in an environmental reservoir-host system. A potential study would simulate a scenario where multiple hosts infected with different virus strains would be allowed to contaminate a quasi-natural environment. Infected hosts would then be removed and susceptible individuals would be exposed to the contaminated environment to allow for infection. Hosts and the environment would be sampled longitudinally at regular intervals throughout the experiment. All samples would be sequenced to determine *de novo* reassortment rates and diversity-pressure-phenotypes (figure 1; which measures the frequency dynamics of particular genotypes and their associated phenotypes, i.e. reassortant emergence). Testing both reservoir and spillover host species (e.g. mallards and chickens) would allow for determination of potential host differences in susceptibility and the impacts of diversity-pressure-phenotypes on epidemiological dynamics. A similar design could be applied to conditions that only allow direct contact between hosts. Comparing results from both experiments would help quantify transmission mechanisms and determine which mechanisms lead to higher reassortment risk.

(c) Top-down approaches: statistical modelling of sentinel surveillance (functional surveillance)

A second important approach to understanding and predicting epidemiological risk is to apply statistical methods to surveillance data from the environment and from reservoir and sentinel hosts. While bottom-up approaches based on experimental data have the power to improve inference by reducing complexity, the inference is specific to the conditions that are tested. For this reason, approaches that measure diversity-pressure-phenotypes in natural populations are equally important. Phylogenetic or phylodynamic methods have mostly been applied to estimate reassortment occurrence, rather than to quantify its frequency or dynamics (i.e. pattern-based surveillance as opposed to process-based, which is 'functional surveillance'). Additionally, because these methods rely on historical genomic diversity for inference of current genomic diversity, they are retrospective not predictive. We propose that longitudinal surveillance of environmental reservoir-host systems, paired with systematic deployment of sentinel host individuals (e.g. [11]), is an important research direction that could improve the power

of top-down statistical tools for prediction of epidemiological risk from environmental surveillance data.

Sentinel host sampling can allow for direct measurement of spillover rates between reservoir and poultry host populations when surveillance occurs at the same locations over time [90]. An example approach would include enclosures of sentinel poultry adjacent to environmental reservoirs used by waterfowl. This type of sentinel host design has allowed for documentation of seasonal dynamics of spillover [91] and evaluation of vaccination effectiveness [92]. The sentinel host design would be coupled with frequent measures of diversity-pressure-phenotypes in the environmental reservoir-host system, and in the sentinel hosts themselves, by applying single-virus sequencing and phenotyping assays to the samples (figure 2b). Analysis of genotype frequencies, linked to measured phenotypes, could identify molecular markers that are most likely to lead to outbreaks in poultry. Another useful dimension within this design would be to collect longitudinal data on the abiotic conditions of environmental samples, densities of hosts and host-species composition, concurrently with the longitudinal diversity-pressure-phenotypes data. These data would be used as covariate data for predicting emergence risk in sentinel hosts (figure 2b). Once validated with out-of-sample data from multiple sites, this type of statistical framework could be used to predict epidemiological risk in spillover hosts using diversity-pressure-phenotypes and abiotic data collected longitudinally from environmental reservoir-host systems. Although these systems might not capture all of the real-life complexity of contact structure between hosts, they would include natural complexity from strain diversity and abiotic factors, which is important for optimizing environmental surveillance technologies and inference from environmentally sampled data.

5. Conclusion

Understanding the emergence of novel reassortments across ecological scales that include environmental transmission, could transform risk assessment capabilities and add to current approaches that use non-mechanistic or retrospective genetic analyses. For example, shifting our focus from inferring which strains have spilled over to predicting which strains will successfully spill over from wild birds to poultry in the near future. Advancements in molecular techniques have brought the goal of predicting reassortment and spillover risk of IAVs onto the near-term horizon. Coupling risk-metric data, such as diversity-pressure-phenotyping, with bottom-up and top-down quantitative models is possible and can yield frameworks for both understanding and predicting reassortant emergence.

Although it is still not possible to collect diversity-pressure-phenotyping data routinely as part of large-scale surveillance systems, technological refinements that dramatically improve efficiency can occur rapidly. Once more efficient molecular technologies are available, scaling up their application to surveillance programmes will require a strong research foundation that has addressed fundamental gaps in the evolutionary ecology of IAVs; a knowledge base that is inextricably linked to the natural life cycle and complexity of this host-pathogen system. Thus, addressing the basic science knowledge gaps with current technology, as

we have outlined, is an important stepping-stone towards developing more accurate and efficient risk assessment and surveillance programmes. However, even at a fine scale, the research approaches we outlined will require substantial investment because they require wide-ranging interdisciplinary collaboration and specialized facilities (i.e. animal pathobiology, cutting-edge sequencing technology, bioinformatic resources and skills, and expertise in mathematical and statistical modelling). Collaborations of this breadth are expensive but are feasible. The potential payoffs from improvements to risk assessment and outbreak prevention through evidence-based biosecurity could greatly outweigh the costs in the long term.

The improved surveillance system we envision can be characterized as ‘functional surveillance’: collecting and integrating appropriate surveillance data into analytical frameworks that describe emergence processes (as opposed to patterns). Functional surveillance would complement current approaches of using pattern-based surveillance

by improving our understanding of emergence processes (e.g. [93]). The mechanistic underpinnings of functional surveillance will allow for better prediction as conditions change, a current weakness of retrospective risk assessments.

Data accessibility. This article has no additional data.

Authors’ contributions. K.M.P., M.W.H., S.A.S., E.S. and A.J.P. wrote the first draft. K.M.P. designed the figures. All authors contributed to idea development and manuscript/figure editing.

Competing interests. We declare we have no competing interests.

Funding. K.M.P., A.J.P., M.W.H., E.S. and S.A.S. were supported by the U.S. Department of Agriculture, APHIS.

Acknowledgements. The findings and conclusions in this preliminary publication have not been formally disseminated by the U.S. Department of Agriculture and should not be construed to represent any agency determination or policy. The authors thank three anonymous reviewers for helpful comments and the organizers of a workshop on pathogen spillover in MT 20–23 February, which helped with idea development, and DARPA for funding the workshop.

References

- Chen Y *et al.* 2013 Human infections with the emerging avian influenza A H7N9 virus from wet market poultry: clinical analysis and characterisation of viral genome. *Lancet* **381**, 1916–1925. (doi:10.1016/S0140-6736(13)60903-4)
- Worobey M, Han GZ, Rambaut A. 2014 Genesis and pathogenesis of the 1918 pandemic H1N1 influenza A virus. *Proc. Natl Acad. Sci. USA* **111**, 8107–8112. (doi:10.1073/pnas.1324197111)
- Villa M, Lassig M. 2017 Fitness cost of reassortment in human influenza. *PLoS Pathog.* **13**, e1006685. (doi:10.1371/journal.ppat.1006685)
- Pinsent A, Fraser C, Ferguson NM, Riley S. 2016 A systematic review of reported reassortant viral lineages of influenza A. *BMC Infect. Dis.* **16**, 3. (doi:10.1186/s12879-015-1298-9)
- Xu KM *et al.* 2007 The genesis and evolution of H9N2 influenza viruses in poultry from southern China, 2000 to 2005. *J. Virol.* **81**, 10 389–10 401. (doi:10.1128/JVI.00979-07)
- Dugan VG *et al.* 2008 The evolutionary genetics and emergence of avian influenza viruses in wild birds. *PLoS Pathog.* **4**, e1000076. (doi:10.1371/journal.ppat.1000076)
- Hatchette TF *et al.* 2004 Influenza A viruses in feral Canadian ducks: extensive reassortment in nature. *J. Gen. Virol.* **85**(Pt 8), 2327–2337. (doi:10.1099/vir.0.79878-0)
- Huang Y *et al.* 2014 Perpetuation and reassortment of gull influenza A viruses in Atlantic North America. *Virology* **456–457**, 353–363. (doi:10.1016/j.virol.2014.04.009)
- Macken CA, Webby RJ, Bruno WJ. 2006 Genotype turnover by reassortment of replication complex genes from avian influenza A virus. *J. Gen. Virol.* **87**(Pt 10), 2803–2815. (doi:10.1099/vir.0.81454-0)
- Sharp GB *et al.* 1997 Coinfection of wild ducks by influenza A viruses: distribution patterns and biological significance. *J. Virol.* **71**, 6128–6135.
- Wille M, Tolf C, Avril A, Latorre-Margalef N. 2013 Frequency and patterns of reassortment in natural influenza A virus infection in a reservoir host. *Virology* **443**, 150–160. (doi:10.1016/j.virol.2013.05.004)
- Lee DH, Bertran K, Kwon J-H, Swayne DE. 2017 Evolution, global spread, and pathogenicity of highly pathogenic avian influenza H5Nx clade 2.3.4.4. *J. Vet. Sci.* **18**, 269–280. (doi:10.4142/jvs.2017.18.S1.269)
- Roche B, Drake JM, Brown J, Stallknecht DE, Bedford T, Rohani P. 2014 Adaptive evolution and environmental durability jointly structure phylogenetic patterns in avian influenza viruses. *PLoS Biol.* **12**, e1001931. (doi:10.1371/journal.pbio.1001931)
- Breban R, Drake JM, Stallknecht DE, Rohani P. 2009 The role of environmental transmission in recurrent avian influenza epidemics. *PLoS Comput. Biol.* **5**, 11. (doi:10.1371/journal.pcbi.1000346)
- Pepin KM, Lass S, Pulliam JRC, Read AF, Lloyd-Smith JO. 2010 Identifying genetic markers of adaptation for surveillance of viral host jumps. *Nat. Rev. Microbiol.* **8**, 802. (doi:10.1038/nrmicro2440)
- Lowen AC. 2017 Constraints, drivers, and implications of influenza A virus reassortment. *Annu. Rev. Virol.* **4**, 105–121. (doi:10.1146/annurev-virology-101416-041726)
- Lowen AC. 2018 It’s in the mix: reassortment of segmented viral genomes. *PLoS Pathog.* **14**, e1007200. (doi:10.1371/journal.ppat.1007200)
- Jackson S, Van Hoeven N, Chen L-M, Maines TR, Cox NJ, Katz JM, Donis RO. 2009 Reassortment between Avian H5N1 and human H3N2 influenza viruses in ferrets: a public health risk assessment. *J. Virol.* **86**, 8131–8140. (doi:10.1128/JVI.00534-09)
- Li X, Liu B, Ma S, Cui P, Liu W, Li Y, Guo J, Chen H. 2018 High frequency of reassortment after co-infection of chickens with the H4N6 and H9N2 influenza A viruses and the biological characteristics of the reassortants. *Vet. Microbiol.* **222**, 11–17. (doi:10.1016/j.vetmic.2018.06.011)
- Ma MJ *et al.* 2018 Evidence for cross-species influenza A virus transmission within swine farms, China: a one health, prospective cohort study. *Clin. Infect. Dis.* **66**, 533–540. (doi:10.1093/cid/cix823)
- Marshall N, Priyamvada L, Ende Z, Steel J, Lowen AC. 2013 Influenza virus reassortment occurs with high frequency in the absence of segment mismatch. *PLoS Pathog.* **9**, 11. (doi:10.1371/journal.ppat.1003421)
- Zhang X *et al.* 2018 Tissue tropisms opt for transmissible reassortants during avian and swine influenza A virus co-infection in swine. *PLoS Pathog.* **14**, e1007417. (doi:10.1371/journal.ppat.1007417)
- Rohani P, Breban R, Stallknecht DE, Drake JM. 2009 Environmental transmission of low pathogenicity avian influenza viruses and its implications for pathogen invasion. *Proc. Natl Acad. Sci. USA* **106**, 10 365–10 369. (doi:10.1073/pnas.0809026106)
- Vittecoq M, Gauduin H, Oudart T, Bertrand O, Roche B, Guillemin M, Boutron O. 2017 Modeling the spread of avian influenza viruses in aquatic reservoirs: a novel hydrodynamic approach applied to the Rhône delta (southern France). *Sci. Total Environ.* **595**, 787–800. (doi:10.1016/j.scitotenv.2017.03.165)
- Root JJ, Bentler KT, Shriner SA, Mooers NL, Vandalen KK, Sullivan HJ, Franklin AB. 2014 Ecological routes of avian influenza virus transmission to a common mesopredator: an experimental evaluation of alternatives. *PLoS ONE* **9**, e102964. (doi:10.1371/journal.pone.0102964)
- Root JJ, Shriner SA, Ellis JW, Vandalen KK, Sullivan HJ, Franklin AB. 2015 When fur and feather occur together: interclass transmission of avian influenza A virus from mammals to birds through common

- resources. *Sci. Rep.* **5**, 14354. (doi:10.1038/srep14354)
27. VanDalen KK, Franklin AB, Mooers NL, Sullivan HJ, Shriner SA. 2010 Shedding light on avian influenza H4N6 infection in mallards: modes of transmission and implications for surveillance. *PLoS ONE* **5**, e12851. (doi:10.1371/journal.pone.0012851)
28. Wille M *et al.* 2018 Alternate routes of influenza A virus infection in Mallard (*Anas platyrhynchos*). *Vet. Res.* **49**, 110. (doi:10.1186/s13567-018-0604-0)
29. Densmore CL, Iwanowicz DD, Ottinger CA, Hindman LJ, Bessler AM, Iwanowicz LR, Prosser DJ, Whitbeck M, Driscoll CP. 2017 Molecular detection of avian influenza virus from sediment samples in waterfowl habitats on the Delmarva Peninsula, United States. *Avian Dis.* **61**, 520–525. (doi:10.1637/11687-060917-ResNote.1)
30. Brown JD, Goekjian G, Poulson R, Valeika S, Stallknecht DE. 2009 Avian influenza virus in water: infectivity is dependent on pH, salinity and temperature. *Vet. Microbiol.* **136**, 20–26. (doi:10.1016/j.vetmic.2008.10.027)
31. Chumpolbanchorn K *et al.* 2006 The effect of temperature and UV light on infectivity of avian influenza virus (H5N1, Thai field strain) in chicken fecal manure. *Southeast Asian J. Trop. Med. Public Health* **37**, 102–105.
32. Graiver DA, Topliff CL, Kelling CL, Bartelt-Hunt SL. 2009 Survival of the avian influenza virus (H6N2) after land disposal. *Environ. Sci. Technol.* **43**, 4063–4067. (doi:10.1021/es900370x)
33. Nazir J, Haumacher R, Ike AC, Marschang RE. 2011 Persistence of avian influenza viruses in lake sediment, duck feces, and duck meat. *Appl. Environ. Microbiol.* **77**, 4981–4985. (doi:10.1128/AEM.00415-11)
34. Stallknecht DE, Kearney MT, Shane SM, Zwank PJ. 1990 Effects of pH, temperature, and salinity on persistence of avian influenza viruses in water. *Avian Dis.* **34**, 412–418. (doi:10.2307/1591429)
35. Stallknecht DE, Shane SM, Kearney MT, Zwank PJ. 1990 Persistence of avian influenza viruses in water. *Avian Dis.* **34**, 406–411. (doi:10.2307/1591428)
36. Lebarbenchon C, Choi YW, Chappie DJ, Rogers JV, Kaye JZ. 2012 Reassortant influenza A viruses in wild duck populations: effects on viral shedding and persistence in water. *Proc. R. Soc. B* **279**, 3967–3975. (doi:10.1098/rspb.2012.1271)
37. Wood JP *et al.* 2010 Environmental persistence of a highly pathogenic avian influenza (H5N1) virus. *Environ. Sci. Technol.* **44**, 7515–7520. (doi:10.1021/es1016153)
38. Lang AS, Kelly A, Runstadler JA. 2008 Prevalence and diversity of avian influenza viruses in environmental reservoirs. *J. Gen. Virol.* **89**, 509–519. (doi:10.1099/vir.0.83369-0)
39. Handel A, Lebarbenchon C, Stallknecht D, Rohani P. 2014 Trade-offs between and within scales: environmental persistence and within-host fitness of avian influenza viruses. *Proc. R. Soc. B* **281**, 20133051. (doi:10.1098/rspb.2013.3051)
40. Kreibich A, Stech O, Hundt J, Ziller M, Mettenleiter TC, Stech J. 2013 Avian influenza virus h3 hemagglutinin may enable high fitness of novel human virus reassortants. *PLoS ONE* **8**, e79165. (doi:10.1371/journal.pone.0079165)
41. Handel A, Rohani P. 2015 Crossing the scale from within-host infection dynamics to between-host transmission fitness: a discussion of current assumptions and knowledge. *Phil. Trans. R. Soc. B* **370**, 20140302. (doi:10.1098/rstb.2014.0302)
42. Bloom JD. 2015 Software for the analysis and visualization of deep mutational scanning data. *BMC Bioinf.* **16**, 168. (doi:10.1186/s12859-015-0590-4)
43. Edwards RA, Rohwer F. 2005 Viral metagenomics. *Nat. Rev. Microbiol.* **3**, 504. (doi:10.1038/nrmicro1163)
44. Rosario K, Breitbart M. 2011 Exploring the viral world through metagenomics. *Curr. Opin. Virol.* **1**, 289–297. (doi:10.1016/j.coviro.2011.06.004)
45. Stallknecht DE, Goekjian VH, Wilcox BR, Poulson RL, Brown JD. 2010 Avian influenza virus in aquatic habitats: what do we need to learn? *Avian Dis.* **54**, 461–465. (doi:10.1637/8760-033109-Reg.1)
46. Thomsen PF, Willerslev E. 2015 Environmental DNA – an emerging tool in conservation for monitoring past and present biodiversity. *Biol. Conserv.* **183**, 4–18. (doi:10.1016/j.biocon.2014.11.019)
47. Davis AJ, Williams KE, Snow NP, Pepin KM, Piaggio AJ. 2018 Accounting for observation processes across multiple levels of uncertainty improves inference of species distributions and guides adaptive sampling of environmental DNA. *Ecol. Evol.* **8**, 10 879–10 892. (doi:10.1002/ece3.4552)
48. Forster JL, Harkin VB, Graham DA, McCullough SJ. 2008 The effect of sample type, temperature and RNAlater™ on the stability of avian influenza virus RNA. *J. Virol. Methods* **149**, 190–194. (doi:10.1016/j.jviromet.2007.12.020)
49. Hubbard LE, Kolpin DW, Fields CL, Hladik ML, Iwanowicz LR. 2017 Highlighting the complexities of a groundwater pilot study during an avian influenza outbreak: methods, lessons learned, and select contaminant results. *Environ. Res.* **158**, 212–224. (doi:10.1016/j.envres.2017.06.010)
50. Russell JA, Campos B, Stone J, Blosser EM, Burkett-Cadena N, Jacobs JL. 2018 Unbiased strain-typing of arbovirus directly from mosquitoes using nanopore sequencing: a field-forward biosurveillance protocol. *Sci. Rep.* **8**, 5417. (doi:10.1038/s41598-018-23641-7)
51. Keller MW *et al.* 2018 Direct RNA sequencing of the complete influenza A virus genome. *Sci. Rep.* **8**, 14408. (doi:10.1038/s41598-018-32615-8)
52. Liu KT *et al.* 2018 Genetic and biological characterization of three poultry-origin H5N6 avian influenza viruses with all internal genes from genotype S H9N2 viruses. *Arch. Virol.* **163**, 947–960. (doi:10.1007/s00705-017-3695-4)
53. Farià NR, Sabino EC, Nunes MRT, Alcántara LCJ, Loman NJ, Pybus OG. 2016 Mobile real-time surveillance of Zika virus in Brazil. *Genome Med.* **8**, 97. (doi:10.1186/s13073-016-0356-2)
54. Perkel JM. 2017 Pocket laboratories. *Nature* **545**, 119. (doi:10.1038/545119a)
55. Lebarbenchon C, Yang M, Keeler SP, Ramakrishnan MA, Brown JD, Stallknecht DE, Sreevatsan S. 2011 Viral replication, persistence in water and genetic characterization of two influenza A viruses isolated from surface lake water. *PLoS ONE* **6**, 8. (doi:10.1371/journal.pone.0026566)
56. Millen HT, Gonnering JC, Berg RK, Spencer SK, Jokela WE, Pearce JM, Borchardt JS, Borchardt MA. 2012 Glass wool filters for concentrating waterborne viruses and agricultural zoonotic pathogens. *J. Vis. Exp.* (61), e3930. (doi:10.3791/3930)
57. Khalenkov A, Laver WG, Webster RG. 2008 Detection and isolation of H5N1 influenza virus from large volumes of natural water. *J. Virol. Methods* **149**, 180–183. (doi:10.1016/j.jviromet.2008.01.001)
58. Vong S, Ly S, Mardy S, Holl D, Buchy P. 2008 Environmental contamination during influenza A virus (H5N1) outbreaks, Cambodia, 2006. *Emerg. Infect. Dis.* **14**, 1303–1305. (doi:10.3201/eid1408.070912)
59. Zhang H, Xu B, Chen Q, Chen Z. 2011 Characterization of H9N2 influenza viruses isolated from Dongting Lake wetland in 2007. *Arch. Virol.* **156**, 95–105. (doi:10.1007/s00705-010-0836-4)
60. Ilyushina NA, Ikizler MR, Kawaoka Y, Rudenko LG, Treanor JJ, Subbarao K, Wright PF. 2012 Comparative study of influenza virus replication in MDCK cells and in primary cells derived from adenoids and airway epithelium. *J. Virol.* **86**, 11725–11 734. (doi:10.1128/JVI.01477-12)
61. Yang J *et al.* 2014 Mutations of novel influenza A(H10N8) virus in chicken eggs and MDCK cells. *Emerg. Infect. Dis.* **20**, 1541–1543. (doi:10.3201/eid2009.140257)
62. Drayman N. 2017 Observed high coinfection rates seem to be a result of overlapping plaques. *Mbio* **8**, e01000-17. (doi:10.1128/mBio.01000-17)
63. Allen LZ, Ishoey T, Novotny MA, McLean JS, Lasken RS, Williamson SJ. 2011 Single virus genomics: a new tool for virus discovery. *PLoS ONE* **6**, e17722. (doi:10.1371/journal.pone.0017722)
64. Steuerma Y *et al.* 2018 Dissection of influenza infection *in vivo* by single-cell RNA sequencing. *Cell Syst.* **6**, 679–691. (doi:10.1016/j.cels.2018.05.008)
65. Shapiro E, Biezuner T, Linnarsson S. 2013 Single-cell sequencing-based technologies will revolutionize whole-organism science. *Nat. Rev. Genet.* **14**, 618–630. (doi:10.1038/nrg3542)
66. Zamora JLR, Aguilar HC. 2018 Flow virometry as a tool to study viruses. *Methods* **134**, 87–97. (doi:10.1016/j.ymeth.2017.12.011)
67. Baaijens JA, Aabidine AZE, Rivals E. 2017 De novo assembly of viral quasispecies using overlap graphs. *Genome Res.* **27**, 835–848. (doi:10.1101/gr.215038.116)
68. Ibrahim B, McMahon DP, Hufsky F, Beer M, Deng L, Mercier PL, Palmarini M, Thiel V, Marz M. 2018 A new era of virus bioinformatics. *Virus Res.* **251**, 86–90. (doi:10.1016/j.virusres.2018.05.009)

69. Baaijens J, Van der Roest B, Koester J, Stougie L, Schoenhuth A. 2018 Full-length *de novo* viral quasispecies assembly through variation graph construction. *Bioinformatics* **btz443**. (doi:10.1093/bioinformatics/btz443)
70. Yang X, Charlebois P, Macalalad A, Henn MR, Zody MC. 2013 V-Phaser 2: variant inference for viral populations. *BMC Genomics* **14**, 674. (doi:10.1186/1471-2164-14-674)
71. Lam TT-Y, Zhu H, Guan Y, Holmes EC. 2016 Genomic analysis of the emergence, evolution, and spread of human respiratory RNA viruses. *Annu. Rev. Genomics Hum. Genet.* **17**, 193–218. (doi:10.1146/annurev-genom-083115-022628)
72. Posada D, Crandall KA. 2001 Intraspecific gene genealogies: trees grafting into networks. *Trends Ecol. Evol.* **16**, 37–45. (doi:10.1016/S0169-5347(00)02026-7)
73. Lu L, Lycett SJ, Brown AJL. 2014 Reassortment patterns of avian influenza virus internal segments among different subtypes. *BMC Evol. Biol.* **14**, 16. (doi:10.1186/1471-2148-14-16)
74. Robbins RJ, Krishtalka L, Wooley JC. 2016 Advances in biodiversity: metagenomics and the unveiling of biological dark matter. *Stand. Genomic Sci.* **11**, 69. (doi:10.1186/s40793-016-0180-8)
75. Huson DH. 1998 SplitsTree: analyzing and visualizing evolutionary data. *Bioinformatics* **14**, 68–73. (doi:10.1093/bioinformatics/14.1.68)
76. Varsani A, Lefevre P, Roumagnac P, Martin D. 2018 Notes on recombination and reassortment in multipartite/segmented viruses. *Curr. Opin. Virol.* **33**, 156–166. (doi:10.1016/j.coviro.2018.08.013)
77. Pybus OG, Rambaut A. 2009 Evolutionary analysis of the dynamics of viral infectious disease. *Nat. Rev. Genet.* **10**, 540–550. (doi:10.1038/nrg2583)
78. Beerli P, Felsenstein J. 2001 Maximum likelihood estimation of a migration matrix and effective population sizes in *n* subpopulations by using a coalescent approach. *Proc. Natl Acad. Sci. USA* **98**, 4563–4568. (doi:10.1073/pnas.081068098)
79. Corander J, Marttinen P, Sirén J, Tang J. 2008 Enhanced Bayesian modelling in BAPS software for learning genetic structures of populations. *BMC Bioinf.* **9**, 539. (doi:10.1186/1471-2105-9-539)
80. Pritchard JK, Stephens M, Donnelly P. 2000 Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959.
81. Plowright RK, Parrish CR, McCallum H, Hudson PJ, Ko AI, Graham AL, Lloyd-Smith JO. 2017 Pathways to zoonotic spillover. *Nat. Rev. Microbiol.* **15**, 502–510. (doi:10.1038/nrmicro.2017.45)
82. Beauchemin CA, Handel A. 2011 A review of mathematical models of influenza A infections within a host or cell culture: lessons learned and challenges ahead. *BMC Public Health* **11**(Suppl. 1), S7. (doi:10.1186/1471-2458-11-S1-S7)
83. Smith AM, Ribeiro RM. 2010 Modeling the viral dynamics of influenza A virus infection. *Crit. Rev. Immunol.* **30**, 291–298. (doi:10.1615/CritRevImmunol.v30.i3.60)
84. Baccam P, Beauchemin C, Macken CA, Hayden FG, Perelson AS. 2006 Kinetics of influenza A virus infection in humans. *J. Virol.* **80**, 7590–7599. (doi:10.1128/JVI.01623-05)
85. Miao H, Hollenbaugh JA, Zand MS, Holden-Wiltse J, Mosmann TR, Perelson AS, Wu H, Topham DJ. 2010 Quantifying the early immune response and adaptive immune response kinetics in mice infected with influenza A virus. *J. Virol.* **84**, 6687–6698. (doi:10.1128/JVI.00266-10)
86. Beauchemin C. 2006 Probing the effects of the well-mixed assumption on viral infection dynamics. *J. Theor. Biol.* **242**, 464–477. (doi:10.1016/j.jtbi.2006.03.014)
87. Hancioglu B, Swigon D, Clermont G. 2007 A dynamical model of human immune response to influenza A virus infection. *J. Theor. Biol.* **246**, 70–86. (doi:10.1016/j.jtbi.2006.12.015)
88. Handel A, Brown J, Stallknecht D, Rohani P. 2013 A multi-scale analysis of influenza A virus fitness tradeoffs due to temperature-dependent virus persistence. *PLoS Comput. Biol.* **9**, e1002989. (doi:10.1371/journal.pcbi.1002989)
89. Restif O *et al.* 2012 Model-guided fieldwork: practical guidelines for multidisciplinary research on wildlife ecological and epidemiological dynamics. *Ecol. Lett.* **15**, 1083–1094. (doi:10.1111/j.1461-0248.2012.01836.x)
90. Halvorson D *et al.* 1983 Epizootiology of avian influenza—simultaneous monitoring of sentinel ducks and turkeys in Minnesota. *Avian Dis.* **27**, 77–85. (doi:10.2307/1590374)
91. Halvorson DA, Kelleher CJ, Senne DA. 1985 Epizootiology of avian influenza: effect of season on incidence in sentinel ducks and domestic turkeys in Minnesota. *Appl. Environ. Microbiol.* **49**, 914–919. (doi:10.2307/1590374)
92. Henning J, Henning KA, Morton JM, Long NT, Ha NT, Vu LT, Vu PP, Hoa DM, Meers J. 2011 Highly pathogenic avian influenza (H5N1) in ducks and in-contact chickens in backyard and smallholder commercial duck farms in Viet Nam. *Prev. Vet. Med.* **101**, 229–240. (doi:10.1016/j.prevetmed.2010.05.016)
93. Becker DJ, Washburne AD, Faust CL, Pulliam JRC, Mordecai EA, Lloyd-Smith JO, Plowright RK. 2019 Dynamic and integrative approaches to understanding pathogen spillover. *Phil. Trans. R. Soc. B* **374**, 20190014. (doi:10.1098/rstb.2019.0014)