

## The future of zoonotic risk prediction

Colin J. Carlson<sup>1,2,\*</sup>, Maxwell J. Farrell<sup>3</sup>, Zoe Grange<sup>4</sup>, Barbara A. Han<sup>5</sup>, Nardus Mollentze<sup>6,7</sup>, Alexandra L. Phelan,<sup>1,8</sup> Angela L. Rasmussen<sup>1</sup>, Gregory F. Albery<sup>9</sup>, Bernard Bett<sup>10</sup>, David M. Brett-Major<sup>11</sup>, Lily E. Cohen<sup>12</sup>, Tad Dallas<sup>13</sup>, Evan A. Eskew<sup>14</sup>, Anna C. Fagre<sup>15</sup>, Kristian M. Forbes<sup>16</sup>, Rory Gibb<sup>17,18</sup>, Sam Halabi<sup>8</sup>, Charlotte C. Hammer<sup>19</sup>, Rebecca Katz<sup>1</sup>, Jason Kindrachuk<sup>20</sup>, Renata L. Muylaert<sup>21</sup>, Felicia B. Nutter<sup>22,23</sup>, Joseph Ogola<sup>24</sup>, Kevin J. Olival<sup>25</sup>, Michelle Rourke<sup>26</sup>, Sadie J. Ryan<sup>27,28</sup>, Noam Ross<sup>25</sup>, Stephanie N. Seifert<sup>29</sup>, Tarja Sironen<sup>30</sup>, Claire J. Standley<sup>1,2</sup>, Kishana Taylor<sup>31</sup>, Marietjie Venter<sup>32</sup>, Paul W. Webala<sup>33</sup>

1. Center for Global Health Science and Security, Georgetown University Medical Center, Washington, D.C., 20007, U.S.A.
2. Department of Microbiology and Immunology, Georgetown University Medical Center, Washington, D.C., 20007, U.S.A.
3. Department of Ecology & Evolutionary Biology, University of Toronto, Toronto, ON, Canada
4. Public Health Scotland, Glasgow, G2 6QE, United Kingdom.
5. Cary Institute of Ecosystem Studies, Millbrook, NY 12545 U.S.A.
6. Medical Research Council - University of Glasgow Centre for Virus Research, Glasgow, G61 1QH, United Kingdom.
7. Institute of Biodiversity, Animal Health & Comparative Medicine, College of Medical, Veterinary & Life Sciences, University of Glasgow, Glasgow, G12 8QQ, United Kingdom
8. O'Neill Institute for National and Global Health Law, Georgetown University Law Center, Washington, D.C., 20001, U.S.A.
9. Department of Biology, Georgetown University, Washington, D.C., 20007, U.S.A.
10. Animal and Human Health Program, International Livestock Research Institute, P. O. Box 30709-00100, Nairobi, Kenya.
11. Department of Epidemiology, College of Public Health, University of Nebraska Medical Center, Omaha, NE, U.S.A.

12. Icahn School of Medicine at Mount Sinai, New York, NY, U.S.A.
13. Department of Biological Sciences, Louisiana State University, Baton Rouge, LA 70806, U.S.A.
14. Department of Biology, Pacific Lutheran University, Tacoma, WA, U.S.A.
15. Department of Microbiology, Immunology, and Pathology, College of Veterinary Medicine and Biomedical Sciences, Colorado State University, Fort Collins, CO, U.S.A.
16. Department of Biological Sciences, University of Arkansas, Fayetteville, AR, 72701, U.S.A.
17. Centre on Climate Change and Planetary Health, London School of Hygiene and Tropical Medicine, London, UK.
18. Centre for Mathematical Modelling of Infectious Diseases, London School of Hygiene and Tropical Medicine, London, UK.
19. Centre for the Study of Existential Risk, University of Cambridge, Cambridge, UK.
20. Department of Medical Microbiology & Infectious Diseases, University of Manitoba, Winnipeg, MB, R3E 0J9, Canada.
21. Molecular Epidemiology and Public Health Laboratory, Hopkirk Research Institute, Massey University, Palmerston North, New Zealand.
22. Department of Infectious Disease and Global Health, Cummings School of Veterinary Medicine, Tufts University, North Grafton, MA, 01536 U.S.A.
23. Department of Public Health and Community Medicine, School of Medicine, Tufts University, Boston, MA 02111 U.S.A.
24. University of Nairobi, Nairobi, Kenya.
25. EcoHealth Alliance, New York, NY, 10018, U.S.A.
26. Law Futures Centre, Griffith Law School, Griffith University, Nathan, QLD 4111, Australia.
27. Department of Geography & Emerging Pathogens Institute, University of Florida, Gainesville, FL, U.S.A.
28. School of Life Sciences, University of KwaZulu-Natal, Durban, South Africa.
29. Paul G. Allen School for Global Health, Washington State University, Pullman, WA, U.S.A.

30. Department of Virology, University of Helsinki, Helsinki, Finland; Department of Veterinary Biosciences, University of Helsinki, Helsinki, Finland.

31. Department of Chemical Engineering, Carnegie Mellon University, Pittsburgh, PA, 15213, U.S.A.

32. Zoonotic arbo and respiratory virus program, Centre for Viral Zoonoses, Department Medical Virology, University of Pretoria, South Africa.

33. Department of Forestry and Wildlife Management, Maasai Mara University, Narok 20500, Kenya.

\* Correspondence should be directed to [colin.carlson@georgetown.edu](mailto:colin.carlson@georgetown.edu)

## Abstract

In light of the urgency raised by the COVID-19 pandemic, global investment in wildlife virology is likely to increase, and new surveillance programs will identify hundreds of novel viruses that might someday pose a threat to humans. In the absence of laboratory characterization, scientists may increasingly rely on data-driven rubrics or machine learning models that learn from known zoonoses to identify which animal pathogens could someday pose a threat to global health. We synthesize the findings of an interdisciplinary workshop on zoonotic risk technologies to answer the following questions: What are the prerequisites, in terms of open data, equity, and interdisciplinary collaboration, to the development and application of those tools? What effect could the technology have on global health? Who would control that technology, who would have access to it, and who would benefit from it? Would it improve pandemic prevention? Could it create new challenges?

## Introduction

After the COVID-19 pandemic ends - or even before [1] - the world will face another emergence of a heretofore-unknown epidemic or pandemic threat, which will most likely be a novel zoonotic virus. This is less a testament to the state of global health, and more a basic consequence of arithmetic: as many as one in every five known mammalian viruses has the ability to make the jump into human populations, and only an estimated one percent of mammal viruses are currently known to science [2,3]. For example, a whole constellation of distinct SARS-related coronaviruses circulate in bats and in China and Southeast Asia [4,5], and at least two-thirds of reservoirs might still be unidentified [6]. But even the most intensively studied viruses and well-sampled hosts can harbor undiscovered diversity: influenza A viruses are perhaps the most widely agreed upon future pandemic threat [7–9], but novel strains emerging through reassortment in wildlife and livestock are often only noticed once they reach or cross the animal-human interface. Despite the urgency of research on zoonotic emergence, the diversity and rapid evolution of viruses poses a problem of scale for actionable science.

The next zoonotic threat might be unfamiliar to virologists, but more likely than not, it will bear at least some similarity to previous counterparts. A handful of viral clades make the zoonotic jump most often, and are more likely to continue spreading within human populations [10–12]. As a result, novel zoonotic epidemics often harken back to previous outbreaks: severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) shares 76% of its genome with SARS-CoV, and much of its pathology [13]; the emergence of HIV-1 group M in the 1920s was followed by a dozen more spillovers of similar primate viruses, including the progenitor of HIV-2 [14–16]; and the emergence of filoviruses with Marburg virus in 1967 was followed almost a decade later by the first *Sudan ebolavirus* and *Zaire ebolavirus* outbreaks, both in 1976 [17].

Though these instances are anecdotal and few in number, similarities between emergence events across time and space point to the widely-accepted idea that while individual outbreaks are idiosyncratic and (as standalone stochastic events) unpredictable, they often follow predictable patterns, which might constitute the raw materials for a zoonotic risk assessment procedure - defining, for example, the virus species, conditions, or locations with a greater risk of causing or experiencing these events [18,19]. For example, a 2017 study, which aimed to predict risk factors for future coronavirus emergence, used a regression model to show that wildlife markets predicted higher coronavirus positivity rates in bats; used another model to estimate hundreds of coronaviruses might still be undiscovered; and used a mapping approach to predict most

undiscovered sarbecoviruses would be found in southern China and southeast Asia [20]. These predictions have been reasonably prescient, given the likely origins of SARS-CoV-2 in the southeast Asian peninsula through wildlife farming or trade three years later. Another study used simple models of viral sharing networks to predict that ferret badgers might have been a likely bridge host in the emergence of SARS-CoV-2 a year before the species was flagged by the World Health Organization origins investigation [6]. Though these approaches might not allow scientists to predict exactly where and when the next outbreak will begin, they allow a different kind of prediction, one focused on exploring and explaining biological possibility and socioecological risk factors with an eye towards future threats.

As zoonotic viruses and their non-human animal (hereafter animal) hosts become better characterized, a growing library of virological data is becoming increasingly available and accessible to the scientific community, putting this risk assessment procedure within reach for the first time. This is increasingly possible with the growing application of machine learning to risk assessment problems. Zoonotic origins are often described as a sequential process, in which pathogens must pass through a series of biological, ecological, and social filters that would otherwise prevent their emergence [21,22]. At each of these steps, machine learning has been successfully and reliably applied to predict the animal origins of a novel zoonosis [23,24], the potential hosts of undiscovered zoonoses [6,25], the ecological and anthropogenic risk factors for zoonotic spillover [26,27], the ability of novel viruses to infect humans [28] and their ability to transmit onwards in human populations [10,29]. Such models have also been used to predict the severity of disease [30], and may be extended to predict mortality in the future [29,31]. These methods have been particularly useful when they can harness the genomic signatures of host adaptation and compatibility [23,32], as for many viruses, this may be the only available data [33].

Here we focus on a subset of this emerging set of methods, which we term **zoonotic risk technology** and define as an informatic system, statistical model, or artificial intelligence that identifies at least one of two viral traits we term **zoonotic potential** (which we define as the ability of an animal virus to infect a human host) and **epidemic potential** (the ability of a zoonotic infection to cause disease and transmit onwards in human populations). In calling these tools “zoonotic risk technology,” we aim to encompass a wider range of systems than just predictive models (e.g., informatic systems like databases that report risk factors for emergence based on expert opinion) and that could extend beyond them (e.g., the physical machines that can store these tools or would allow their deployment in field settings). More importantly, by considering these tools as a kind of emerging technology, we aim not to imply anything about the scientific validity or value

of the tools, but instead to focus attention on their user base, implementation, and effects on broader human systems. This set of approaches has a necessarily narrowly-defined scope, which does not encompass every component of “risk.” For example, machine learning algorithms can also be applied to predict zoonotic isolates of bacteria [34,35], and similar models can be applied to identify potential wildlife reservoirs or arthropod vectors of zoonoses [25,36,37]. Similarly, spatiotemporal patterns of viral dynamics in livestock and wildlife reservoirs are a critical missing piece in many spillover risk assessments [38–40]. After a transmissible pathogen reaches human hosts, yet another set of virological, social, economic, and political factors determine whether a spillover event becomes an epidemic or pandemic [41–44]. However, we focus on the narrowly-defined idea of zoonotic risk technology as a way to operationalize a specific set of existing approaches to facilitate the identification of viruses with zoonotic potential, and to interrogate the potential value of these technologies to global health.

To facilitate discussion on these topics, we held a one-day digital workshop (the “Verena Forum on Zoonotic Risk Technology”) at the Georgetown University Center for Global Health Science and Security in January, 2021. This setting allowed scientists to present cutting-edge computational and laboratory approaches, and to discuss potential applications or challenges with global health practitioners, with a focus on equity concerns in data sharing and technology deployment. Here, we report a brief synthesis of our findings.

Zoonotic risk technologies are no longer hypothetical, and are rapidly emerging as practical, concrete applications of scientific knowledge. These tools are only one item in the broader toolkit of prediction in viral ecology, and like other predictive models, are imperfect. Here, we identify three major barriers to actionable science that researchers must consider further:

1. Technologies will have the most value to global health if they are treated as part of the process of characterizing risk, rather than the singular endpoint. Additional work is required to validate predictions, such as laboratory investigations, but may be expensive at-scale and potentially politically sensitive.
2. Academic publishing alone is insufficient to enable the deployment of tools in surveillance programs or rapid outbreak response scenarios; user-friendly, open-source tools must be coupled with global capacity building in risk analyses and mitigation.
3. The development and application of zoonotic risk technology, and the sharing of data to enable these processes, are likely to engage critical issues such as ownership, equity, and

governance; these issues are considered central in global health, but relevant scholarship may not currently interface with existing research on zoonotic risk prediction.

We explore each of these issues in depth here, and discuss possible avenues for interdisciplinary work that might help overcome these barriers, identify conditions precedent to their use, and flag potential limitations.

## How zoonotic risk technology works

At its core, zoonotic risk technology exploits the assumption that viruses with undetected zoonotic potential are more similar to known zoonoses than to non-zoonotic viruses. Early efforts have focused on identifying coarse traits that are common among known zoonoses, such as origins in particular host clades [11,45,46] or a broad host range [47–49]. These approaches are useful for identifying common profiles of what a zoonosis “looks like” - e.g., a vector-borne single-stranded RNA virus with a broad host range including primates - that can be generalized across animal viruses. One of the only examples of zoonotic risk technology available for public use, the SpillOver viral risk ranking platform, uses this approach to rank 887 viruses based on 31 risk factors [50]. These approaches benefit from generality and interpretability, but can be limited by data availability; for example, host range is rarely characterized in wildlife viruses until they are a known threat to human health, and may suffer from biases in study effort or surveillance [51]. Moreover, trait-based assessments may be limited by apparent contradictions in simple patterns. For example, genome size correlates positively with zoonotic risk [52] but has been reported as having contradictory effects on transmissibility [10,12]; replication in the cytoplasm similarly predicts zoonotic potential [11,53,54], but also predicts reduced transmissibility [10]. Each of these has an idiosyncratic effect on risk assessment, but approaches that gather as many lines of evidence as possible can aim to minimize the influence of any given feature to an overall picture of risk.

Genomic data increasingly offer an alternate avenue for predictive work. Genomes are inherently high-dimensional data, encoding meaningful information about microbiology and immunology, and are often the first aspect of a novel virus to be characterized, months or years before its ecology. A simple model might be trained on the nucleotide similarity of viruses compared to known zoonotic threats, while a more advanced one might also include similarity in genomic composition biases [23,28]. This approach has worked well for the parallel problem of inferring viral origins using genomic features that encode coevolutionary signals of host adaptation. For example, CpG

dinucleotide composition can be used to identify vertebrate viruses [55,56], exploiting a viral adaptation that matches genomic composition to the vertebrate genome in order to evade innate immune responses searching for non-self genetic material [57,58]. These patterns are rare and poorly understood today [59], but the subject of significant interest. For zoonotic risk, a model is likely to identify some combination of broadly-transferrable coevolutionary adaptations that allow a virus to cross species barriers more readily within a broad group (e.g., primates, or vertebrates), and random genomic patterns that happen to increase their odds of successful infection of human hosts (which they may never have encountered in their evolutionary history). For example, including similarity of viral genomes to human housekeeping genes and interferon-stimulated genes appears to measurably improve prediction of zoonotic potential [28].

Over time, genomic approaches are likely to move beyond similarity, and start identifying *de novo* predictors of viral compatibility with human cells. A mechanism-agnostic model may simply collapse genomes into hundreds of computational features, identify a small handful of significant predictors, and generalize these patterns - but can only do so successfully with sufficient data. For example, massive clearinghouses of genomic sequences such as GISAID and the NIAID Influenza Research Database have enabled a number of models that accurately classify the zoonotic potential of influenza strains down to the protein level [60,61]. Decomposing viral genomes, and identifying the regions most relevant to zoonotic emergence, can open new avenues for advanced modeling that goes beyond pattern recognition. For example, researchers have developed a number of structural simulations to explore binding affinity between the spike protein of SARS-CoV-2 and ACE2 receptors in animal and human cells [62,63], and structural modeling can be paired with other trait data to better predict the capacity for various mammal species to transmit SARS-CoV-2 [64]. Similar approaches could be used to identify the zoonotic potential of other viruses for which surface protein structure and receptor use has been characterized [65]. These kinds of approaches are ultimately limited by the comparability of different structures in both host and pathogen genomes, and may be most predictive when comparing hosts or pathogens at lower taxonomic levels (i.e. viral strain up to viral genus or family).

No matter how sophisticated these approaches become, they all face the fundamental task of overcoming data limitations [51]. Only a few hundred zoonotic viruses are known—a sample size that is fundamentally limiting for inference, even when treated as the positives in a sample of a few thousand wildlife viruses. Many groups are chronically understudied, even after zoonotic viruses are discovered (e.g., the genus *Tibrovirus* remains poorly characterized despite the 2009 discovery of Bas-Congo virus and the 2015 discoveries of Ekpoma virus 1 and 2), creating bias in training datasets

that will usually lead models to underestimate the risk posed by unfamiliar pathogens. Models that rely on similarity alone will particularly struggle with these data sources, while those that correctly identify underlying mechanisms (e.g., viral matching to host genomic composition) will perform better out-of-sample, though distinguishing the two can be challenging. Similarly, the genomes of many viruses are only partially characterized; models can train on specific regions of the genome, even though targets available from sequencing data may not be those with the greatest biological relevance (e.g., a recent model trained on betacoronavirus RdRp sequences could identify zoonotic viruses with reasonable accuracy [66]). Genome-based models are also inherently constrained by the constant evolution of new viral lineages [19,67]; while influenza sequence data is abundant enough to make protein-based models that are somewhat insensitive to this problem [61], other situations – such as the emergence of a novel recombinant canine coronavirus (alphacoronavirus 1) that can infect humans – are harder to anticipate [68]. Beyond genomic features, ecological data on viruses is often even more limited; for example, metadata on host range – a key predictor in many previous zoonotic risk studies – is even more underdeveloped, with 20-40% of associations missing in even the smallest, best-sampled networks [69]. A recent study showed that graph embeddings of the host-virus network could improve the performance of a genomic model of viral zoonotic potential—adding high-dimensional data to a model otherwise limited to a few hundred points—but using an imputed host-virus network even further improved predictions by overcoming gaps in viral host range data [70]. As these examples highlight, data limitations are likely to change as viruses are discovered at an increasing rate. Still, in the interim, experts will continue developing unique solutions to facing data sparsity that will advance both the basic biology and the computational tools of machine learning.

It is difficult to define what zoonotic risk technology might look like within the next ten years. As models predicting zoonotic potential become more advanced, it may be increasingly possible to also address epidemic potential—one recent example was able to identify transmissible human viruses with > 80% accuracy [10], while others have nodded towards the possibility that mortality might also be predictable [29,31]—and to leverage machine learning more fully for comprehensive risk assessment. This nascent field of research is likely to grow exponentially as post-pandemic investments transform both the available data describing the global virome, and the institutional support for modeling research and development (and associated training in higher education). Especially if this work focuses on improvement through validation and cross-talk between experimentalists and modellers, we anticipate that the predictive accuracy and reliability of these technologies will continue to grow. Previous work, especially from virologists, has been skeptical that these approaches might become a reliable source of inference [51,67]; however, prior work

anticipating the level of predictive resolution that exists today has also historically been subjected to similar skepticism. Moving past these concerns may require a transition from primarily-computational (sometimes mechanism-agnostic) models—which can perform well at any given task, but may be harder to interpret through different disciplinary lenses—and towards a deeper conceptual synthesis of virology and computational biology, focused on identifying the rules of life that underpin host-virus interactions through a computational lens. As models become powered by growing datasets cataloguing the global virome [2,71–75], and more complex microbiological predictors that capture more granular host-virus interactions, it is difficult to imagine today how accurate and valuable they might become. If their potential for global health manifests, we should prepare now to guard against potential misuse, including monopolization in high income countries, and to anticipate important matters of equity, including the equitable sharing of the benefits arising from their use.

### **Connecting computational and empirical work**

Zoonotic risk technology can suggest which viruses may have zoonotic potential, with a nontrivial degree of uncertainty, but further confirming that risk requires laboratory characterization. For example, successful viral replication in humans requires tens to hundreds of protein-protein interactions, which may not be predicted from viral sequence data alone and require laboratory characterization [76–78]. Conversely, one of the greatest strengths of these tools is their ability to narrow down the list of (potentially millions of) viruses for risk assessment procedures that require complicated, sometimes-expensive experiments. For example, experimental evaluation of host competency may require establishment of cell lines from new species [79] or non-model organism systems in the laboratory with a suite of associated challenges including unique housing requirements, low fecundity, a lack of commercial availability, few species-specific laboratory reagents, and often scant baseline data upon which to support health evaluations. Focusing on establishing these systems for the wildlife viruses with the highest predicted risk could be a way to direct effort and minimize costs, particularly if model-to-validation steps are built in that can validate the underlying biological reasoning (e.g., predicting cell entry based on receptor sequences followed by high-throughput functional testing [80]). Conversely, experimental work will point to new needs in model development (e.g., when cell line experiments identify previously-uncharacterized receptors, these can be incorporated back into the modelling process).

This aspect of zoonotic risk assessment can be complicated by concerns that this might require **gain-of-function experiments**, which use genetic editing or forced adaptation experiments to induce new phenotypes, potentially expanding the host range, pathogenesis, or mode of transmission of a pathogen [81]. While these experiments have been critical to previous work — for example, by demonstrating the epidemic potential of highly pathogenic avian influenza through directed mutagenesis and serial passage to recover a virus capable of airborne transmission [82,83], or by demonstrating the potential of SARS-like viruses to jump from bat reservoirs into human populations [84] — they also face tremendous scrutiny given perceived or potential biosafety and biosecurity risks, including those potentially arising from dual-use research of concern [85]. Importantly, most host-virus interaction research (including *in vitro* and *in vivo* experimental infections) are not actually “gain-of-function” experiments, but are mislabeled or misidentified as such by the public or media. These concerns are likely to face even greater scrutiny given public conversations about SARS-CoV-2’s as-yet-unknown origins and the emergence of unsubstantiated origin theories centered around biosecurity lapses [86].

There are a tremendous diversity of experimental approaches stopping short of gain-of-function experiments that may be used to validate predictive models and offer a more operationalized view of the problem. Among these are experimental infections to test the ability for cell entry and receptor usage, replication, pathogenesis, evasion of host immune responses, assembly and egress, and onward transmission [80,87]. The complexity of these experimental systems may begin with host cell lines and surrogate viruses, pseudoviruses, or replicon systems, and expand to include experiments with live virus and organoids [88] or live animal models [89]. Each of these laboratory approaches can offer targeted methods to validate predictions from machine learning models, such as virus-human compatibility, ability for viral replication and productive infection, and disease pathology, tissue tropism, or courses of infection. Many of these methods can be used without the requirement for high-containment laboratories, which is particularly important to ensure that a wide variety of different viral groups can be studied safely yet at scale and across country contexts.

Experimental work can further help identify the (model-able) molecular barriers to zoonotic emergence, including rules governing attachment and entry, transcription and genome replication, viral protein expression, innate immune antagonism, viral assembly, and egress. Each stage in the viral life cycle represents an opportunity to improve model performance, but will require the gathering and reconciliation of data across multiple host-virus systems and experimental approaches. For example, laboratory experiments are likely to vastly improve model performance upstream by offering new kinds of predictor data that reflect the various types of host responses to

infection. These may include broad comparative data on host transcriptomic or proteomic responses to infection [90], or host-virus protein-protein interactions [77], which may help identify the mechanisms of infection and pathogenesis in humans even when collected from animal model systems [90]. Through better collaboration among statistical modellers and empiricists, future development of zoonotic risk technologies can iteratively validate or falsify model predictions, helping to improve the accuracy and applicability of predictive models over time. While some modeling publications may therefore recommend further characterization of specific viruses, this will be unlikely to occur without active partnership between modellers and experimentalists, given that the priority is often placed on known and recurrent threats.

### **Theory to technology, technology to toolkit**

Most zoonotic risk technology is developed with the stated intent to contribute positively to human health and reduce the future burden of emerging zoonotic viruses. However, the knowledge that a virus poses a threat to human health often exists for years, even decades, before a catastrophic outbreak [33]. Zoonotic risk technology may therefore have limited benefit to global health without careful, intentional work focused on application and actionability. The pipeline from technology development, to implementation, to risk mitigation is likely to only succeed at first in specific, narrow use cases.

First, and most foundationally, predictive modeling work will always be disconnected from global health if the endpoint is in the academic literature. This is particularly the case if research groups are separated geographically and practically from the direct impacts of potential spillover events, and choose not to pursue collaboration and knowledge exchange with local experts, limiting both the expertise available to properly design and contextualize work, and the channels available for possible dissemination and outreach. Similar challenges have been identified for related modeling problems, like the development and deployment of early warning systems or real-time epidemic forecasting [91,92]. Engaging practitioners, policymakers, and stakeholders in the participatory design, release, and ongoing improvement of infrastructure is likely to increase the value of zoonotic risk technology, as will designing open tools with public interfaces (open-source software or websites, e.g. FluLeap: <https://fluleap.bic.nus.edu.sg/>) and based on open, interpretable data. These tools should be adaptable in order to more accurately reflect scientific advances, as well as changing user needs, over time. They must report appropriate, interpretable, and locally-relevant metrics of risk to inform decision-making, with uncertainty presented as transparently as possible, including where

uncertainty comes from (e.g., data limitations vs. model calibration) and how uncertainty correlates with the outcome variables. Scientists may also be called on to develop new language for conveying context-dependent risk and communicating uncertainty to different audiences, including properly disclaiming results such that public, private, or health sector responses neither over-react to high-risk predictions nor under-react to low-risk predictions. This enduring challenge extends into many other aspects of global health and disease ecology.

At least in the near term, it is unlikely that any specific, coordinated, and effective response will be mobilized based solely on the identification of a novel virus with zoonotic potential. Resource scarcity prevents the development of individual surveillance systems or biomedical research-and-development programs for each of the thousands of wildlife viruses with zoonotic potential; existing programs focused on the narrowest set of expert-assessed high-risk threats (e.g., influenza A viruses, betacoronaviruses, or henipaviruses) are already over-encumbered. These programs may, however, benefit from technology that identifies zoonosis-relevant evolutionary shifts in viruses circulating in wildlife (e.g., the emergence of a Nipah virus lineage with greater estimated transmissibility, a long-standing concern in some biosecurity circles [93]).

More broadly, these tools may find applications in existing One Health surveillance programs focused on high-risk interfaces between wildlife, domestic and captive animals, and humans. Zoonotic risk technology is likely to be most actionable at small scales: a local inventory of wildlife viruses can be ranked according to risk, frequency, and degree of human-animal contact, with the highest-risk viruses incorporated back into local surveillance priorities. For studies reporting the discovery of novel animal viruses, this step can be simple as an additional analysis, with at least one published example of this use case [94]. However, if studies are designed with these kinds of assessments in mind, they may also be able to collect additional data with tremendous value. For example, sequence-based viral discovery often focuses only on viral reads and discards data from the host [95]. The host-derived sequence data contains crucial information about the host response that could provide insight into a given virus' pathogenic potential in an animal or human host, allowing for further surveillance prioritization of both host and virus species. The utility of these studies is also limited by the quality of data shared in the public domain: sharing of standardised and validated data such as host species identification, location, specimen type, and date of collection in a centralised resource, rather than lost in a publication (if at all), is one of the first steps to a truly collaborative approach.

Once high-risk viruses are identified and reported in wildlife monitoring studies, these pathogens may also be identified and flagged earlier in samples collected by programs that passively monitor the health of high-risk human populations (e.g., livestock keepers or wildlife traders) and sentinel hosts like livestock, or actively screen human populations for novel pathogens by investigating undiagnosed febrile illnesses [96–99]. Behavioral change or occupational safety interventions may then be targeted to reduce spillover risk for high-risk human populations, though they may be most feasible or successful if they target risky exposure to specific host species with multiple high-risk viruses and frequent human contact (especially if they already match local priorities), thereby protecting against their entire **zoonotic virome**. For example, while Nipah virus is the highest-priority zoonotic threat hosted by the Indian fruit bat (*Pteropus medius*), interventions that reduce Nipah exposure in humans may also protect against the other 50+ viruses that these bats host [100]. Similarly, the reservoir for Lassa virus carries a number of other bacterial zoonoses, and rodent control can reduce transmission risk across this range of threats [101]. While many of these reservoirs are known today from their role in pathogen spillover, a number of other high-risk species are presumably unknown; building on previous work that characterizes zoonotic risk using ecological traits correlated with “hyperreservoirs” [25], fruitful avenues of future research include characterizing these species’ viromes, and estimate the risk that they pose in aggregate.

### **Failure to equitably share benefits may limit impact**

As the failure to ensure global access to diagnostics, therapeutics, and vaccines during the COVID-19 pandemic has demonstrated [102], the distribution of the benefits from health technologies, particularly novel technologies, is inequitable and a global injustice. Global health must not simply aspire to principles of health equity and social justice, but must also make equitable access to life-saving technologies a condition precedent to their development and use. This must be a priority in the development and use of zoonotic risk technology, which may also pose a unique set of problems for both researchers and practitioners. These technologies depend on open data sharing, both to create sufficient training sets for artificial intelligence, and to actually apply them for risk assessment and subsequent mitigation. Community efforts to share human and animal sequence data at sufficient scales (i.e., to generate feature sets for advanced machine learning) exist for just a handful of high profile viruses, nearly exclusively as part of international coordination on pandemic preparedness and response (e.g. influenza A and SARS-CoV-2 data sharing via the GISAID platform), while all-purpose repositories like GenBank still only capture a fraction of known viruses

(as many are bottlenecked by taxonomic ratification), and lack essential metadata needed for prediction. Both face challenges with regard to contributors receiving credit and attribution for research (especially modeling studies) based on the data they submit (Box 1).

These problems become more complex with regards to the deployment of zoonotic risk technology itself. Initially, there may be resistance to using these tools: scientists who gather novel sequence data may rightfully be hesitant to upload unpublished data to online web tools for zoonotic risk prediction without clear and enforceable protections against data reuse by the curators of such tools, even if these tools are curated by a trusted third-party (though access to this technology may inherently change power dynamics). This is only one concern out of a broader set of issues around **access and benefit sharing** for viral surveillance. Based in countries' sovereign rights to determine the use of resources within their territory, access and benefit-sharing regimes seek to redress and prevent injustices arising from the exploitation of genetic resources, and from the inequitable sharing of the benefits that arise from their use. Some protections and norms around the sharing of physical pathogen samples and the benefits arising from their use are reflected under the Nagoya Protocol on Access to Genetic Resources and the Fair and Equitable Sharing of Benefits Arising from their Utilization (Nagoya Protocol) to the Convention on Biological Diversity. Under the Nagoya Protocol, countries may implement domestic legislation that requires foreign researchers seeking access to pathogen samples, or in some cases even data related to those samples, to obtain the country's prior informed consent and the conclusion of mutually agreed terms that include benefit sharing, such as attribution in publications, capacity building, technology transfer, or intellectual property rights. Depending on the terms agreed, the use of genetic sequence data derived from pathogen samples may be restricted, including both preventing the sharing of sequence data in open access databases, requiring open sharing, or the sharing of any diagnostic tools developed utilizing the sequence data.

Given that a growing number of labs are readily able to synthesize viruses from their genome sequences (e.g., horsepox [103] and SARS-CoV-2 [104]), there are concerns that the bargain underpinning access and benefit sharing regimes provided by physical samples - like those established in the Nagoya Protocol - may be in flux. If zoonotic risk technologies allow researchers to identify high-risk viruses with reasonable certainty before laboratory characterization, this could add an additional layer of complexity. Sequence sharing and synthesis of SARS-CoV-2, in addition to the global failure to equitably distribute associated vaccines, diagnostics, and therapeutics, may motivate attempts to expressly address these gaps in international legal instruments. These could include, for example, potential revision to the International Health Regulations (2005) or the

Nagoya Protocol, or new international law, such as a Pandemic Treaty. Any international governance reform should actively consider the importance, on equal footing, of open data sharing and the equitable sharing of the benefits of novel technologies like zoonotic risk technology.

Even if zoonotic risk technologies are easily applied without challenges around sequence data sharing, there may be gaps between intentions and actionable science. When high-risk viruses are identified, findings may be kept private until they are published much later in peer-reviewed journals, both to protect credit for scientific discoveries, and to accommodate governments' hesitancy to release information that could create fear or stigmatization. This could reinforce the disconnect between viral sampling and actionable science for global public health. At present, announcements about the discovery of notable animal viruses are often made *ad hoc* either by press release or conventional publishing methods. If zoonotic risk technology becomes a widely-adopted part of surveillance, new governance processes will likely need to be developed that protect researchers' careers and credit, but also ensure that announcements are transparent and verifiable, particularly if alarming or unusual results (e.g., the hypothetical discovery of a filovirus with zoonotic potential in bats in the United States) are likely to motivate public or international concern.

Another set of issues could arise around who benefits from zoonotic risk technology. It seems plausible that these technologies might mostly benefit from the research effort and data sharing occurring in tropical countries, where zoonotic viral diversity is believed to be highest [11]. However, their development might mostly further the careers of researchers in high income countries in North America and Europe, particularly if developed by experts who are unattuned to power dynamics in global health. Equally concerning, we identify a possibility that these tools will largely be developed as proprietary "risk assessment algorithms" by corporate "data science for impact" programs, for-profit global health firms, and non-profit organizations, just as they have been for the development of pandemic insurance programs or similar analytics. In these circumstances, and without appropriate governance, the countries with the highest burden of zoonotic emergence might find their own data (repackaged in an analytic format) sold back to them at a premium by scientists and corporations from high-income countries. Open sharing of academic research could help scientists undermine this trend and provide tools directly to end users in public health, or assist them in developing their own tools, but may simply accelerate advances in zoonotic risk technology without changing the existing colonial framework of global health. Involving researchers from low- and middle-income countries – and supporting their leadership in this field (particularly, to a greater extent, in future workshops on this topic, which could advance this issue

further than the present workshop did) – will help limit these shortfalls. This can be particularly facilitated by using virtual workshops (with attention to different accessibility challenges) and by supporting in-person workshops around the world, and by coordinating participation proactively through existing zoonotic disease and laboratory networks (e.g., SEAOHUN, OHCEA, or the CREID center network). Involving researchers from the discipline of science and technology studies (STS) may also lead to more honest and critical appraisal of the ethical issues surrounding the emerging technology, and who it benefits or harms.

Finally, we anticipate that zoonotic risk technology may replicate existing, and potentially create new, ethics and governance problems in **synthetic biology**. Just as convolutional neural networks and other kinds of artificial intelligence can be used to fabricate realistic images entirely through predictive algorithms (e.g., [thispersondoesnotexist.com](http://thispersondoesnotexist.com)), zoonotic risk technology might be used to generate novel viral sequences (and potentially synthetic viruses) with high predicted zoonotic and epidemic potential. Already, researchers have used these approaches to simulate alternate coronavirus spike protein sequences that might be able to infect human cells [105]. These approaches might support biomedical work; for example, synthetic spike proteins could be used to test a candidate universal betacoronavirus vaccine for its value across “unsampled evolutionary space.” However, if biomedical companies attempt to patent these sequences, they could create new problems for future sample sharing, therapeutic and vaccine development, or outbreak response if viruses with the relevant sequence someday emerge - potentially at the expense of some countries more than others. While similar issues have been raised before during zoonotic outbreaks [106], the novelty of **simulated zoonoses** might create new complications for intellectual property law. Moreover, viral ranking algorithms or artificially simulated virus sequences might also be used by a malicious actor, highlighting the need to involve scholarship from the “dual use” field of bioethics.

### Prediction isn't prevention

Zoonotic risk technology may become an asset in the emerging disease toolkit, but overselling this technology or understating uncertainty will lead to preventable divergences between expectations and scientific possibility. Models may ultimately have profound clinical and field applications, but the uncertainty around risk estimates and likelihood of inaccurate predictions must be carefully communicated. As part of that, epistemic differences in disciplinary conceptions of uncertainty may need to be bridged: for example, a model may make “errors” simply because reality is a stochastic observation of underlying risk landscapes, and a technology that correctly infers probabilities or risk

landscapes will still never perfectly represent reality. (These may play into other disciplinary tensions about what “prediction” means: to public health experts and the public, prediction is often synonymous with anticipating future events, but to computational biologists, it may more often be used to describe accurate inference about biological possibility.) Further, there is no substitute for experimental work, and bench virology will play a critical role to generate the necessary data for model development and validation. Zoonotic risk technology is also no substitute for general public health preparedness; even though these tools could be used in the future to estimate the risk posed by newly-discovered viruses as soon as the first genome becomes available, many viruses are still likely to continue to enter human populations before they have been characterized in animals. Whether these outbreaks become epidemics or pandemics is a problem outside the scope of the technologies we discuss.

Therefore, we warn that investments in research and development on topics like machine learning or animal virus genomics must not come at the expense of other essential kinds of modeling work (e.g., work focused on virus transmission and spread, or identifying the most consequential surveillance gaps), or more importantly, at the expense of non-technological investments in health systems strengthening, including attainment of universal health coverage, and similar aspects of pandemic preparedness. Similarly, it is possible that interest in pre-emergence zoonotic viruses might conflict with, redirect, or undermine local priorities like water and food-borne diseases (and sanitation), agricultural, high-burden communicable diseases (e.g., HIV-AIDS, tuberculosis, and malaria) or non-communicable diseases; interventions may even disrupt local interests and norms, potentially weakening outbreak response during emergencies. If the post-pandemic period becomes dominated by this narrow subset of research priorities, researchers will need to be individually careful in order to accurately and fairly present the value and importance of their work (an imperative that will be encouraged by efforts to reduce funding scarcity in this space).

At the same time, it is indisputable that zoonotic risk technology is currently limited in both development and application by data scarcity, and that the only solution to this is continued or greater investment in data collection—particularly in basic science. Post-pandemic investment in coordinated programs for viral discovery, One Health surveillance, bench virology, and other kinds of laboratory capacity are all likely to generate vital data that can improve the performance of these technologies, and remedy critical gaps in our current understanding of the global virome. These will be most effective if investments are maximized in the hotspots of zoonotic emergence, if modellers are engaged in the process to support data collection and processing in reusable formats, and - perhaps most importantly - if these investments are made with the aim of improving outbreak

prevention and preparedness entirely independent of the success or failure of zoonotic risk prediction as a scientific outcome.

Finally, we suggest that ongoing work is required to benchmark the accuracy and value of these technologies, that transparency and uncertainty be key facets of their presentation, and most importantly that the scientific community remains prepared for “surprises.” (In a strikingly timely example, only days before the submission of this manuscript, the first ever report of human infections with H5N8 avian influenza A virus was released—a strain that was, surprisingly enough, able to be correctly identified as human by a previously-published model that had never encountered a zoonotic H5N8 virus in the training data [107].) Models are only as powerful as the data that inform them, and with such a small percentage of the global virome described to date - and new viruses evolving constantly - it seems likely that the next generation of risk prediction systems, and public health infrastructure that may come to rely on them, will face a number of entirely unexpected threats.

## Acknowledgements

We thank Simon Anthony, Eddie Holmes, Michelle Wille, three anonymous reviewers, and the other participants in the workshop for their time and substantive contributions to the discussion, as well as the Georgetown Center for Global Health Science and Security, the Georgetown Global Health Initiative, and the Georgetown Environment Initiative for hosting the Verena Forum workshop. CJC, MJF, and the Verena Consortium were supported by NSF BII 2021909. MJF was supported by the University of Toronto EEB Fellowship. NM was supported by the Wellcome Trust (217221/Z/19/Z). KJO is supported by the National Institute of Allergy and Infectious Diseases of the National Institutes of Health (U01AI151797) and the Defense Threat Reduction Agency (HDTRA11710064). Our workshop was organized remotely from Georgetown University, and we acknowledge that unceded land was and still is the homeland of the Nacotchtank and their descendants, the Piscataway Conoy people (and we thank the Georgetown Native American Student Council for the language adapted here). Finally, we thank the researchers from around the world who were invited to our workshop but unable to attend as they responded to the COVID-19 pandemic, working tirelessly to mitigate the impacts of the latest zoonotic emergence in their own communities.

### **Box 1. Crediting researchers for reused, open sequence data**

When “big data” becomes available at scales that allow machine learning (or other intensive secondary analysis), the researchers who compiled the data often receive exponentially-diminishing credit through academic incentives. Existing public data repositories, including GISAID and NCBI GenBank, have no indexable source attribution for sequence data. GISAID requires acknowledgement of the source, but such acknowledgement is not a trackable metric contributing to career development; similarly, GenBank accession numbers assist in reproducibility of analyses, but are not indexed, and cannot be easily tracked by contributors as a career metric. This can disincentivize open data sharing if it is seen as a “non-promotable” task for those generating the data, given that other indexed metrics like citations may be used to determine scientific impact when evaluating funding proposals or in hiring and promotion decisions.

Moreover, this system currently benefits users of public data repositories more than those who generate the data. In several instances during the COVID-19 pandemic, laboratories generating SARS-CoV-2 sequence data have been stretched thin with pandemic response and were unable to annotate, analyze, and publish on their data before computational or academic laboratories used the data in their own publications. Similar practices are particularly divisive when researchers use data generated by public health laboratories in developing nations without co-authorship, collaboration, or indexed citation of the source.

One potential solution would be an indexed DOI for sequence data; similar approaches are used for aggregate data in biodiversity research (e.g., by the Global Biodiversity Informatics Facility; gbif.org), and while many studies fail to follow recommendations for proper attribution, these procedures are a reasonable first step for fair credit.

## Literature cited

1. Yong E. 2020 America Should Prepare for a Double Pandemic. *The Atlantic*, 15 July. , <https://www.theatlantic.com/health/archive/2020/07/double-pandemic-covid-flu/614152/>.
2. Carroll D, Daszak P, Wolfe ND, Gao GF, Morel CM, Morzaria S, Pablos-Méndez A, Tomori O, Mazet JAK. 2018 The Global Virome Project. *Science* **359**, 872–874.
3. Carlson CJ, Zipfel CM, Garnier R, Bansal S. 2019 Global estimates of mammalian viral diversity accounting for host sharing. *Nat Ecol Evol* **3**, 1070–1075.
4. Wacharapluesadee S *et al.* 2021 Evidence for SARS-CoV-2 related coronaviruses circulating in bats and pangolins in Southeast Asia. *Nat. Commun.* **12**, 972.
5. Latinne A *et al.* 2020 Origin and cross-species transmission of bat coronaviruses in China. *Nat. Commun.* **11**, 4235.
6. Becker DJ *et al.* 2020 Predicting wildlife hosts of betacoronaviruses for SARS-CoV-2 sampling prioritization: a modeling study. *Cold Spring Harbor Laboratory.* , 2020.05.22.111344. (doi:10.1101/2020.05.22.111344)
7. Eccleston-Turner M, Phelan A, Katz R. 2019 Preparing for the Next Pandemic - The WHO's Global Influenza Strategy. *N. Engl. J. Med.* **381**, 2192–2194.
8. Lipsitch M *et al.* 2016 Viral factors in influenza pandemic risk assessment. *Elife* **5**. (doi:10.7554/eLife.18491)
9. Fan VY, Jamison DT, Summers LH. 2018 Pandemic risk: how large are the expected losses? *Bull. World Health Organ.* **96**, 129–134.
10. Walker JW, Han BA, Ott IM, Drake JM. 2018 Transmissibility of emerging viral zoonoses. *PLoS One* **13**, e0206926.
11. Olival KJ, Hosseini PR, Zambrana-Torrel C, Ross N, Bogich TL, Daszak P. 2017 Host and Viral Traits Predict Zoonotic Spillover from Mammals. (doi:10.5281/zenodo.807517)
12. Geoghegan JL, Senior AM, Di Giallonardo F, Holmes EC. 2016 Virological factors that increase the transmissibility of emerging human viruses. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 4170–4175.
13. Wang H, Li X, Li T, Zhang S, Wang L, Wu X, Liu J. 2020 The genetic sequence, origin, and diagnosis of SARS-CoV-2. *Eur. J. Clin. Microbiol. Infect. Dis.* **39**, 1629–1635.
14. Worobey M *et al.* 2010 Island biogeography reveals the deep history of SIV. *Science* **329**, 1487.

15. Sharp PM, Hahn BH. 2010 The evolution of HIV-1 and the origin of AIDS. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **365**, 2487–2494.
16. Korber B, Muldoon M, Theiler J, Gao F, Gupta R, Lapedes A, Hahn BH, Wolinsky S, Bhattacharya T. 2000 Timing the ancestor of the HIV-1 pandemic strains. *Science* **288**, 1789–1796.
17. Emanuel J, Marzi A, Feldmann H. 2018 Filoviruses: Ecology, Molecular Biology, and Evolution. *Adv. Virus Res.* **100**, 189–221.
18. Morse SS, Mazet JAK, Woolhouse M, Parrish CR, Carroll D, Karesh WB, Zambrana-Torrel C, Lipkin WI, Daszak P. 2012 Prediction and prevention of the next pandemic zoonosis. *Lancet* **380**, 1956–1965.
19. Geoghegan JL, Holmes EC. 2017 Predicting virus emergence amid evolutionary noise. *Open Biol.* **7**. (doi:10.1098/rsob.170189)
20. Anthony SJ *et al.* 2017 Global patterns in coronavirus diversity. *Virus Evol* **3**, vex012.
21. Wolfe ND, Dunavan CP, Diamond J. 2007 Origins of major human infectious diseases. *Nature* **447**, 279–283.
22. 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.
23. Babayan SA, Orton RJ, Streicker DG. 2018 Predicting reservoir hosts and arthropod vectors from evolutionary signatures in RNA virus genomes. *Science* **362**, 577–580.
24. Brierley L, Fowler A. 2020 Predicting the animal hosts of coronaviruses from compositional biases of spike protein and whole genome sequences through machine learning. *Cold Spring Harbor Laboratory.*, 2020.11.02.350439. (doi:10.1101/2020.11.02.350439)
25. Han BA, Schmidt JP, Bowden SE, Drake JM. 2015 Rodent reservoirs of future zoonotic diseases. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 7039–7044.
26. Allen T, Murray KA, Zambrana-Torrel C, Morse SS, Rondinini C, Di Marco M, Breit N, Olival KJ, Daszak P. 2017 Global hotspots and correlates of emerging zoonotic diseases. *Nat. Commun.* **8**, 1124.
27. Carlson CJ, Albery GF, Merow C, Trisos CH, Zipfel CM. 2020 Climate change will drive novel cross-species viral transmission. *bioRxiv*
28. Mollentze N, Babayan SA, Streicker DG. 2020 Identifying and prioritizing potential human-infecting viruses from their genome sequences. *Cold Spring Harbor Laboratory.*,

2020.11.12.379917. (doi:10.1101/2020.11.12.379917)

29. Guth S, Visser E, Boots M, Brook CE. 2019 Host phylogenetic distance drives trends in virus virulence and transmissibility across the animal-human interface. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **374**, 20190296.
30. Brierley L, Pedersen AB, Woolhouse MEJ. 2019 Tissue tropism and transmission ecology predict virulence of human RNA viruses. *PLoS Biol.* **17**, e3000206.
31. Farrell MJ, Davies TJ. 2019 Disease mortality in domesticated animals is predicted by host evolutionary relationships. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 7911–7915.
32. 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**, 802–813.
33. Carlson CJ. 2020 From PREDICT to prevention, one pandemic later. *Lancet Microbe* **1**, e6–e7.
34. Lupolova N, Dallman TJ, Matthews L, Bono JL, Gally DL. 2016 Support vector machine applied to predict the zoonotic potential of E. coli O157 cattle isolates. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 11312–11317.
35. Lupolova N, Dallman TJ, Holden NJ, Gally DL. 2017 Patchy promiscuity: machine learning applied to predict the host specificity of and. *Microb Genom* **3**, e000135.
36. Yang LH, Han BA. 2018 Data-driven predictions and novel hypotheses about zoonotic tick vectors from the genus Ixodes. *BMC Ecol.* **18**, 7.
37. Guy C, Ratcliffe JM, Mideo N. 2020 The influence of bat ecology on viral diversity and reservoir status. *Ecol. Evol.* **10**, 5748–5758.
38. Kessler MK *et al.* 2018 Changing resource landscapes and spillover of henipaviruses. *Ann. N. Y. Acad. Sci.* **1429**, 78–99.
39. Schmidt JP, Park AW, Kramer AM, Han BA, Alexander LW, Drake JM. 2017 Spatiotemporal Fluctuations and Triggers of Ebola Virus Spillover. *Emerg. Infect. Dis.* **23**, 415–422.
40. Epstein JH *et al.* 2020 Nipah virus dynamics in bats and implications for spillover to humans. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 29190–29201.
41. Ali S *et al.* 2017 Environmental and Social Change Drive the Explosive Emergence of Zika Virus in the Americas. *PLoS Negl. Trop. Dis.* **11**, e0005135.
42. Singer M, Bulled N, Ostrach B, Mendenhall E. 2017 Syndemics and the biosocial conception of health. *Lancet* **389**, 941–950.

43. Moon S *et al.* 2015 Will Ebola change the game? Ten essential reforms before the next pandemic. The report of the Harvard-LSHTM Independent Panel on the Global Response to Ebola. *Lancet* **386**, 2204–2221.
44. McCloskey B, Dar O, Zumla A, Heymann DL. 2014 Emerging infectious diseases and pandemic potential: status quo and reducing risk of global spread. *Lancet Infect. Dis.* **14**, 1001–1010.
45. Washburne AD, Crowley DE, Becker DJ, Olival KJ, Taylor M, Munster VJ, Plowright RK. 2018 Taxonomic patterns in the zoonotic potential of mammalian viruses. *PeerJ* **6**, e5979.
46. Luis AD *et al.* 2013 A comparison of bats and rodents as reservoirs of zoonotic viruses: are bats special? *Proc. Biol. Sci.* **280**, 20122753.
47. Cleaveland S, Laurenson MK, Taylor LH. 2001 Diseases of humans and their domestic mammals: pathogen characteristics, host range and the risk of emergence. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **356**, 991–999.
48. Woolhouse MEJ, Gowtage-Sequeria S. 2005 Host range and emerging and reemerging pathogens. *Emerg. Infect. Dis.* **11**, 1842–1847.
49. Kreuder Johnson C *et al.* 2015 Spillover and pandemic properties of zoonotic viruses with high host plasticity. *Sci. Rep.* **5**, 14830.
50. Grange ZL *et al.* 2021 Ranking the risk of animal-to-human spillover for newly discovered viruses. *Proceedings of the National Academy of Sciences*, in press.
51. Wille M, Geoghegan JL, Holmes EC. 2021 How accurately can we assess zoonotic risk? *PLoS Biology*, in press.
52. Grewelle RE. 2020 Larger viral genome size facilitates emergence of zoonotic diseases. *Cold Spring Harbor Laboratory.*, 2020.03.10.986109. (doi:10.1101/2020.03.10.986109)
53. Pulliam JRC, Dushoff J. 2009 Ability to replicate in the cytoplasm predicts zoonotic transmission of livestock viruses. *J. Infect. Dis.* **199**, 565–568.
54. Mollentze N, Streicker DG. 2020 Viral zoonotic risk is homogenous among taxonomic orders of mammalian and avian reservoir hosts. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 9423–9430.
55. Kapoor A, Simmonds P, Lipkin WI, Zaidi S, Delwart E. 2010 Use of nucleotide composition analysis to infer hosts for three novel picorna-like viruses. *J. Virol.* **84**, 10322–10328.
56. Colmant AMG *et al.* 2017 A New Clade of Insect-Specific Flaviviruses from Australian Mosquitoes Displays Species-Specific Host Restriction. *mSphere* **2**. (doi:10.1128/mSphere.00262-17)

57. Ficarella M, Wilson H, Pedro Galão R, Mazzon M, Antzin-Anduetza I, Marsh M, Neil SJ, Swanson CM. 2019 KHNYN is essential for the zinc finger antiviral protein (ZAP) to restrict HIV-1 containing clustered CpG dinucleotides. *Elife* **8**. (doi:10.7554/eLife.46767)
58. Velazquez-Salinas L, Zarate S, Eschbaumer M, Pereira Lobo F, Gladue DP, Arzt J, Novella IS, Rodriguez LL. 2016 Selective Factors Associated with the Evolution of Codon Usage in Natural Populations of Arboviruses. *PLoS One* **11**, e0159943.
59. Di Giallonardo F, Schlub TE, Shi M, Holmes EC. 2017 Dinucleotide Composition in Animal RNA Viruses Is Shaped More by Virus Family than by Host Species. *J. Virol.* **91**. (doi:10.1128/JVI.02381-16)
60. Eng CLP, Tong JC, Tan TW. 2014 Predicting host tropism of influenza A virus proteins using random forest. *BMC Med. Genomics* **7 Suppl 3**, S1.
61. Eng CLP, Tong JC, Tan TW. 2017 Predicting Zoonotic Risk of Influenza A Viruses from Host Tropism Protein Signature Using Random Forest. *Int. J. Mol. Sci.* **18**. (doi:10.3390/ijms18061135)
62. Rodrigues JPGLM, Barrera-Vilarmau S, M C Teixeira J, Sorokina M, Seckel E, Kastitis PL, Levitt M. 2020 Insights on cross-species transmission of SARS-CoV-2 from structural modeling. *PLoS Comput. Biol.* **16**, e1008449.
63. Junwen Luan, Yue Lu, Xiaolu Jin, Leiliang Zhang. 2020 Spike protein recognition of mammalian ACE2 predicts the host range and an optimized ACE2 for SARS-CoV-2 infection. *Biochem. Biophys. Res. Commun.* **526**, 165–169.
64. Fischhoff IR, Castellanos AA, Rodrigues JPGLM, Varsani A, Han BA. 2021 Predicting the zoonotic capacity of mammal species for SARS-CoV-2. *bioRxiv* (doi:10.1101/2021.02.18.431844)
65. Liu-Wei W, Kafkas Ş, Chen J, Dimonaco NJ, Tegnér J, Hoehndorf R. 2021 DeepViral: prediction of novel virus–host interactions from protein sequences and infectious disease phenotypes. *Bioinformatics* (doi:10.1093/bioinformatics/btab147)
66. Wadhawan K, Das P, Han BA, Fischhoff IR, Castellanos AC, Varsani A, Varshney KR. 2021 Towards interpreting zoonotic potential of betacoronavirus sequences with attention.
67. Holmes EC, Rambaut A, Andersen KG. 2018 Pandemics: spend on surveillance, not prediction. *Nature* **558**, 180–182.
68. Vlasova AN, Diaz A, Dantie D, Xiu L, Toh T-H, Lee JS-Y, Saif LJ, Gray GC. 2021 Novel Canine Coronavirus Isolated from a Hospitalized Pneumonia Patient, East Malaysia. *Clin. Infect.*

*Dis.* (doi:10.1093/cid/ciab456)

69. Dallas T, Park AW, Drake JM. 2017 Predicting cryptic links in host-parasite networks. *PLoS Comput. Biol.* **13**, e1005557.
70. Poisot T, Ouellet M-A, Mollentze N, Farrell MJ, Becker DJ, Albery GF, Gibb RJ, Seifert SN, Carlson CJ. 2021 Imputing the mammalian virome with linear filtering and singular value decomposition. *arXiv [q-bio.QM]*.
71. Edgar RC *et al.* 2020 Petabase-scale sequence alignment catalyses viral discovery. *Cold Spring Harbor Laboratory.* , 2020.08.07.241729. (doi:10.1101/2020.08.07.241729)
72. Gibb R *et al.* 2021 Data proliferation, reconciliation, and synthesis in viral ecology. *Cold Spring Harbor Laboratory.* , 2021.01.14.426572. (doi:10.1101/2021.01.14.426572)
73. Nieuwenhuijse DF, Oude Munnink BB, Phan MVT, Global Sewage Surveillance project consortium, Munk P, Venkatakrishnan S, Aarestrup FM, Cotten M, Koopmans MPG. 2020 Setting a baseline for global urban virome surveillance in sewage. *Sci. Rep.* **10**, 13748.
74. Schmidt C. 2018 The virome hunters. *Nat. Biotechnol.* **36**, 916–919.
75. Mina MJ, Metcalf CJE, McDermott AB, Douek DC, Farrar J, Grenfell BT. 2020 A Global Immunological Observatory to meet a time of pandemics. *Elife* **9**. (doi:10.7554/eLife.58989)
76. Warren CJ, Sawyer SL. 2019 How host genetics dictates successful viral zoonosis. *PLoS Biol.* **17**, e3000217.
77. Brito AF, Pinney JW. 2017 Protein-Protein Interactions in Virus-Host Systems. *Front. Microbiol.* **8**, 1557.
78. Cook HV, Doncheva NT, Szklarczyk D, Von Mering C, Jensen LJ. 2018 Viruses.STRING: A Virus-Host Protein-Protein Interaction Database. *Viruses* **10**, 519.
79. Banerjee A, Misra V, Schountz T, Baker ML. 2018 Tools to study pathogen-host interactions in bats. *Virus Res.* **248**, 5–12.
80. Letko M, Marzi A, Munster V. 2020 Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. *Nat Microbiol* **5**, 562–569.
81. Duprex WP, Fouchier RAM, Imperiale MJ, Lipsitch M, Relman DA. 2015 Gain-of-function experiments: time for a real debate. *Nat. Rev. Microbiol.* **13**, 58–64.
82. Herfst S *et al.* 2012 Airborne transmission of influenza A/H5N1 virus between ferrets. *Science* **336**, 1534–1541.

83. Sutton TC, Finch C, Shao H, Angel M, Chen H, Capua I, Cattoli G, Monne I, Perez DR. 2014 Airborne transmission of highly pathogenic H7N1 influenza virus in ferrets. *J. Virol.* **88**, 6623–6635.
84. Menachery VD *et al.* 2015 A SARS-like cluster of circulating bat coronaviruses shows potential for human emergence. *Nat. Med.* **21**, 1508–1513.
85. Casadevall A, Imperiale MJ. 2014 Risks and benefits of gain-of-function experiments with pathogens of pandemic potential, such as influenza virus: a call for a science-based discussion. *MBio* **5**, e01730–14.
86. Andersen KG, Rambaut A, Lipkin WI, Holmes EC, Garry RF. 2020 The proximal origin of SARS-CoV-2. *Nat. Med.* **26**, 450–452.
87. Bosco-Lauth AM *et al.* 2020 Experimental infection of domestic dogs and cats with SARS-CoV-2: Pathogenesis, transmission, and response to reexposure in cats. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 26382–26388.
88. Kim J, Koo B-K, Knoblich JA. 2020 Human organoids: model systems for human biology and medicine. *Nat. Rev. Mol. Cell Biol.* **21**, 571–584.
89. Park MS, Kim JI, Bae J-Y, Park M-S. 2020 Animal models for the risk assessment of viral pandemic potential. *Lab. Anim. Res.* **36**, 11.
90. Price A *et al.* 2020 Transcriptional Correlates of Tolerance and Lethality in Mice Predict Ebola Virus Disease Patient Outcomes. *Cell Rep.* **30**, 1702–1713.e6.
91. Hussain-Alkhateeb L *et al.* 2018 Early warning and response system (EWARS) for dengue outbreaks: Recent advancements towards widespread applications in critical settings. *PLoS One* **13**, e0196811.
92. George DB *et al.* 2019 Technology to advance infectious disease forecasting for outbreak management. *Nat. Commun.* **10**, 3932.
93. Luby SP. 2013 The pandemic potential of Nipah virus. *Antiviral Res.* **100**, 38–43.
94. Bergner LM, Mollentze N, Orton RJ, Tello C, Broos A, Biek R, Streicker DG. 2021 Characterizing and Evaluating the Zoonotic Potential of Novel Viruses Discovered in Vampire Bats. *Viruses* **13**, 252.
95. Bergner LM, Orton RJ, da Silva Filipe A, Shaw AE, Becker DJ, Tello C, Biek R, Streicker DG. 2019 Using noninvasive metagenomics to characterize viral communities from wildlife. *Mol. Ecol. Resour.* **19**, 128–143.

96. Grace D, Lindahl J, Wanyoike F, Bett B, Randolph T, Rich KM. 2017 Poor livestock keepers: ecosystem–poverty–health interactions. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **372**, 20160166.
97. Forbes KM *et al.* 2020 Towards a coordinated strategy for intercepting human disease emergence in Africa. *The Lancet Microbe* (doi:10.1016/s2666-5247(20)30220-2)
98. Li H *et al.* 2019 Human-animal interactions and bat coronavirus spillover potential among rural residents in Southern China. *Biosaf Health* **1**, 84–90.
99. Wang N *et al.* 2018 Serological Evidence of Bat SARS-Related Coronavirus Infection in Humans, China. *Viol. Sin.* **33**, 104–107.
100. Anthony SJ *et al.* 2013 A strategy to estimate unknown viral diversity in mammals. *MBio* **4**, e00598–13.
101. Mari Saez A, Cherif Haidara M, Camara A, Kourouma F, Sage M, Magassouba N 'faly, Fichet-Calvet E. 2018 Rodent control to fight Lassa fever: Evaluation and lessons learned from a 4-year study in Upper Guinea. *PLoS Negl. Trop. Dis.* **12**, e0006829.
102. Phelan AL, Eccleston-Turner M, Rourke M, Maleche A, Wang C. 2020 Legal agreements: barriers and enablers to global equitable COVID-19 vaccine access. *Lancet* **396**, 800–802.
103. Rourke MF, Phelan A, Lawson C. 2020 Access and benefit-sharing following the synthesis of horsepox virus. *Nat. Biotechnol.* **38**, 537–539.
104. Thi Nhu Thao T *et al.* 2020 Rapid reconstruction of SARS-CoV-2 using a synthetic genomics platform. *Nature* **582**, 561–565.
105. Crossman LC. In press. Leveraging Deep Learning to Simulate Coronavirus Spike proteins has the potential to predict future Zoonotic sequences. (doi:10.1101/2020.04.20.046920)
106. The Lancet. 2013 MERS-CoV: a global challenge. *Lancet* **381**, 1960.
107. Carlson CJ. Evolutionary surprise, artificial intelligence, and H5N8. *The Verena Consortium Blog*. <https://www.viralemergence.org/blog/evolutionary-surprise-artificial-intelligence-and-h5n8>.