The problem of scale in the prediction and management of pathogen spillover

Daniel J. Becker[1-3*], Alex D. Washburne[1], Christina L. Faust[4], Erin A. Mordecai[5], Raina K. Plowright[1]

[1]Department of Microbiology and Immunology, Montana State University, Bozeman, Montana

[2]Center for the Ecology of Infectious Diseases, University of Georgia, Athens, Georgia

[3]Department of Biology, Indiana University, Bloomington, Indiana

[4]Institute of Biodiversity Animal Health & Comparative Medicine, University of Glasgow, Glasgow, United Kingdom

[5]Department of Biology, Stanford University, Stanford, California

[*]danbeck@iu.edu

Running head: Scales of spillover prediction

Keywords: mechanistic; cross-species transmission; surveillance; macroecology; pathogen discovery; zoonosis

Abstract

Epidemics, pandemics, and disease emergence events all underscore the need to predict zoonotic pathogen spillover. Because cross-species transmission is inherently hierarchical, involving processes that occur at varying levels of biological organization, such predictive efforts can be complicated by the many scales and vastness of data potentially required for forecasting. A wide range of approaches are currently used to forecast spillover risk (e.g., macroecology, pathogen discovery, surveillance of human populations, among others), each of which is bound within particular phylogenetic, spatial, and temporal scales of prediction. Here, we contextualize these diverse approaches within their forecasting goals and resulting scales of prediction to illustrate critical areas of conceptual and pragmatic overlap. Specifically, we focus on an ecological perspective to envision a research pipeline that connects these different scales of data and predictions from the aims of discovery to intervention. Pathogen discovery and predictions focused at the phylogenetic scale can first provide coarse and pattern-based guidance for which reservoirs, vectors, pathogens are likely to be involved in spillover, thereby narrowing surveillance targets and where such efforts should be conducted. Next, these predictions can be followed with ecologically driven spatiotemporal studies of reservoirs and vectors to quantify spatiotemporal fluctuations in infection and to mechanistically understand how pathogens circulate and are transmitted to humans. This approach can also help identify general regions and periods for which spillover is most likely. We illustrate this point by highlighting several case studies where long-term, ecologically focused studies (e.g., Lyme disease in the northeast United States, Hendra virus in eastern Australia, *Plasmodium knowlesi* in Southeast Asia) have facilitated predicting spillover in space and time and facilitated design of possible intervention strategies. Such studies can in turn help narrow human surveillance efforts and help refine and improve future large-scale, phylogenetic predictions. We conclude by discussing how greater integration and exchange between data and predictions generated across these varying scales could ultimately help generate more actionable forecasts and interventions.

Introduction

Scale is a key challenge for developing actionable forecasts of pathogen spillover. The processes that connect reservoir and recipient hosts to facilitate cross-species transmission—which include infection dynamics in reservoirs, release of infectious agents and survival in the environment, and recipient host exposure and susceptibility to infection—are inherently hierarchical and occur over multiple scales of biological organization [1]. Efforts to predict pathogen spillover can thus be complicated by the many scales and vastness of data potentially required for such forecasts [Washburne et al., this issue]. However, recent epidemics (e.g., Ebola virus in West Africa [2]), pandemics (e.g., Zika virus across the Americas [3]), and disease emergence events (e.g., *Plasmodium knowlesi* in Malaysia [4]) underscore the need to improve such predictive efforts.

Any prediction of pathogen spillover will be bound to a specific scale of space and time [5], and these scales are determined by objectives of the particular forecast. We rarely aim to predict an individual that will become infected with a zoonotic pathogen; instead, forecasts often aim to predict the region and period within which interventions are practical. As an analogue, weather forecasts do not aim to predict the specific time and place of a particular thunderstorm but instead predict the probability of rainfall over regional spatial scales (e.g., cities) at hourly, daily, and weekly temporal scales. Forecasts of pathogen spillover are further restricted not only to spatial and temporal scales but also to the scale of organisms for which we forecast risk (e.g., reservoirs, vectors, pathogens). We refer to this as the "phylogenetic scale", following recent developments in ecology and evolution, as it can be defined along a tunable DNA sequence similarity or given phylogenetic depth [6,7]. As the concept of scale is based upon the order of entities within some hierarchy (e.g., for spatial scale, continents, biomes, ecoregions, etc [8]), phylogenetic scale can accordingly encompass several measures, including not only taxonomic ranks of organisms (e.g., family, genus, species) but also clade age, clade size, and node-to-root distance, among others [9]. Though we use the term "phylogenetic scale" to be inclusive, its application here is generally synonymous with taxonomic level. Any spillover prediction will thus correspond to a particular phylogenetic scale or scales (i.e., a particular lineage of pathogens from a particular taxon of reservoirs or vectors) within a particular region of space and time.

Various approaches are currently used to forecast spillover risk within the bounds of these scales of prediction. Many macroecological studies focus on processes located upstream in the spillover pathway and have thus examined trait profiles or cladistic patterns in which species are likely to serve as reservoirs or vectors of zoonotic pathogens. For example, rodents with a particularly fast pace of life are more likely to host zoonotic pathogens [10]. Related approaches have been applied to pathogens [11], and recent work has used viral sequences to predict reservoirs and vectors [12]. By generating predictions at the phylogenetic scale, these approaches can guide surveillance, and mapping distributions of known or predicted reservoirs, vectors, and zoonoses can allow spatial predictions that identify regions where spillover is likely [13,14]. Given phylogenetic and spatial biases in such research [15,16], a related approach has used pathogen discovery to characterize pathogen diversity; examples include but are not limited to the U.S. Agency for International Development PREDICT project and Global Virome Project [17,18]. Many pathogen discovery projects also enhance local capacity and conduct human surveillance (e.g., VIZIONS in Vietnam [19]), which is another (but not mutually exclusive) approach [Das et al., this issue]. Spatiotemporal human surveillance can identify spillovers by screening persons with symptoms that are not easily diagnosable [20] and can facilitate early detection of known pathogens [21]. These approaches focus on upstream (e.g., reservoirs) and

downstream (e.g., humans) processes of the spillover pathway and can generate predictions at different scales, which has prompted discussions of how best to allocate research efforts [22,23].

The problem of scale has been well described for ecology and evolution [8,24] and could inform discussions of how various research approaches and data streams can contribute to forecasting pathogen spillover risks [22,25,26]; in particular, scales define the interface between data collection, forecasts, and interventions. Here, we argue that these various methods, data streams, and predictions fit along a research pipeline that spans from discovery to intervention (e.g., Plowright et al. 2019), and we accordingly contextualize these approaches within their forecasting goals and resulting scales of prediction (Table 1). On the one hand, many approaches aim to predict spillover of *novel* pathogens (e.g., pathogen discovery in wildlife or in human populations), and these efforts and the data they collect inform the phylogenetic scales of prediction (e.g., identifying likely reservoir, vector, and pathogen lineages involved in spillover) and by extrapolation the spatial scales of prediction (e.g., mapping distributions). On the other hand, a different suite of efforts aims to predict the spillover of *known* pathogens, which often focus explicitly on prediction at the scales of space and time within a well-defined phylogenetic scale (e.g., yellow fever virus [Childs et al., this issue] or Hendra virus [27]). A central concept connecting these different approaches is scale, but no ecological system has a single scale for which complex phenomena such as spillover should be studied or predicted [24]. For pathogen spillover, the scales for which we aim to forecast risk (i.e., phylogenetic, spatial, temporal) vary, and the resulting predictions made across these variable scales can have different applications.

In this paper, we (*i*) highlight what practical information can be gained from data at different phylogenetic and spatiotemporal scales and (*ii*) identify areas of complement between coarse- and fine-scale forecasts through ecological perspective. Specifically, we envision a research pipeline that spans from discovery to intervention and connects the data and predictions provided by macroecology, pathogen discovery, and surveillance in wildlife and humans (Fig. 1). We also highlight statistical tools for defining the appropriate phylogenetic, spatial, and temporal scales at which such predictions can be made, which are necessary to better connect data with proposed interventions. The problem of predicting pathogen spillover is thus best framed as a challenge in identifying over what scales our data permit forecasts, at what benefits, and at what costs. Each of these can be quantified for improved surveillance and management decisions.

Phylogenetic scales of spillover prediction

Do lineages of reservoirs, vectors, and pathogens have common patterns in spillover occurrence? Phylogenetic predictions for which animals are likely to be reservoirs or vectors and which pathogens are likely to be zoonotic range from the phylogenetic scale of species to order or even class. Species-specific predictions, often generated from trait-based analyses and machine learning algorithms, can offer the most operable resolution of spillover forecasting given a sufficiently high degree of cross-validation (e.g., area under the receiver operating characteristic curve [AUC] for out-of-sample prediction) and resulting rank order of likely hosts of zoonotic pathogens. For example, boosted regression trees with high classification accuracy (AUC=87%) predicted 112 bat species to be within the 90th percentile of likely novel filovirus reservoirs [28]. Mapping the geographic distributions of these likely reservoirs further provided spatially explicit predictions, with Southeast Asia being a notable hotspot of filovirus-positive bat species where evidence has otherwise been moderate. Recent filovirus surveys of bats listed in this 90th percentile (e.g., *Eonycteris spelaea*) validate these fine-scale phylogenetic predictions [29].

Similar species-specific predictions have been generated for rodent zoonoses [10], Zika virus in mosquitoes and primates [14,30], and human-to-human transmissibility of zoonotic viruses [31]. Other trait-based approaches generate broader predictions focused on the life history profiles of likely reservoirs or zoonotic pathogens rather than species predictions [13,32,33].

Phylogenetic predictions are also commonly generated at coarser scales, such as clades (Fig. 1a). Such efforts can be especially insightful when trait data are limited. A comprehensive study across mammal viruses found that host Order had greater predictive power for explaining variation in viral diversity compared to traits such as sympatry and body mass [13]. Similarly, trait-based analyses found no host traits to be predictive of how helminth infections in wildlife respond to environmental change, yet Old World primates showed high phylogenetic signal [34]; recent work on helminths in more urban vervet monkeys supports such cladistic predictions [35].

In general, it may not be safe to assume a single phylogenetic scale (e.g., class, order, genus, species) captures any pattern in an ecological dataset. As traits driving spillover may evolve along branches in a phylogeny, there may not be any *a priori* reason to expect the most important driving patterns to be concentrated along branches of a fixed depth (e.g., only genus). A novel machine learning algorithm, phylogenetic factorization, was recently developed to more flexibly identify phylogenetic scales at various depths (e.g., genera and families simultaneously) underlying patterns in ecological data [36]. A recent application of this method to the taxonomy of mammal viruses found that the propensity of viruses to be zoonotic is best partitioned along clades of different phylogenetic scales, such as the order *Nidovirales*, family *Papillomaviridae*, and genera *Alphavirus* and *Deltaretrovirus* [37]. Such consideration of multiple phylogenetic scales in future studies of cladistic patterns in reservoirs and vectors would be informative.

Generating even coarse predictions of reservoir, vector, or pathogen clades involved in spillover (Fig. 1a) depends on pathogen discovery and its data on pathogen diversity. However, such data streams and phylogenetic forecasts provide the first steps that are needed to move from discovery to intervention by narrowing the range of which reservoirs, vectors, and pathogens should be the focus of surveillance and where such efforts should be conducted (Fig. 1b). This initial stage of the pipeline can apply to diverse contexts of spillover (Table 1). Predicting broad clades of reservoirs or vectors could be especially important for guiding surveillance of novel pathogens for which little *a priori* data exist [26]; for example, viral genome sequences were used to forecast artiodactyls as the likely origin of the recently emerged Bas-Congo virus [12]. For known pathogens of public health concern, coarse phylogenetic predictions can still guide field surveillance; for example, Old World fruit bats, primates, and artiodactyls could be targeted for further Ebola virus survey efforts [Schmidt et al., this issue]. To move toward interventions, however, these phylogenetic predictions should be next followed by spatiotemporal studies.

Spatiotemporal scales of spillover prediction

As pathogen circulation in the reservoir or vector over space and time is the first requisite for spillover [1], the next stage of the pipeline from discovery to intervention requires phylogenetic predictions to be supplemented by spatial and temporal surveillance (Fig. 1c). Specifically, such studies in the proposed reservoir(s) or vector(s) are necessary to quantify spatiotemporal fluctuations in infection (i.e., pathogen pressure) [Plowright at al., this issue] and assess their ecological drivers [38–42]. Mechanistic models can help explain whether pathogen pressure is driven by birth seasonality, metapopulation dynamics, environmental synchrony, and within-host processes, among others [Glennon et al., this issue][43,44]. Models that integrate such drivers

can be highly predictive of spillover. For example, an ecologically driven model of yellow fever virus spillover that integrated various spatiotemporal data streams had the strongest predictive accuracy (AUC=0.79) when considering cyclical infection dynamics in wild primate reservoir hosts; critically, models considering the ecology of mosquito vectors and wild primate reservoirs were more predictive than those that also included human population size and immunity [Childs et al., this issue]. As with phylogenetic predictions, such ecologically driven efforts may not generate exact spatial and temporal spillover predictions. However, these coarser predictions at the scale of months, seasons, geographic regions, and habitat types can be actionable for guiding spillover prevention, surveillance, and possible interventions (Fig. 1c).

As with phylogenetic predictions, one may not wish to *a priori* assume a given spatial (e.g., 5 km, 50 km) and temporal scale (e.g., weeks, months) for studying infection dynamics in reservoirs and vectors. Analyses of spatiotemporal autocorrelation of case data in recipient hosts or prevalence in related sylvatic systems can facilitate sampling decisions by determining the spatial and temporal scales at which sampling should occur (e.g., by identifying the range of spatial and temporal dependence with semivariograms) [45]. Where spatially and temporally explicit data exist across many reservoir or vector species, phylogenetic factorization could be applied to identify which spatial and temporal scales may be most appropriate for surveillance or intervention for specific clades [36]. Determining the spatial and temporal scales at which infection is correlated can also help identify ecological correlates of infection dynamics and spillover risk. For example, spatial dependence across large scales can suggest effects of major climatic drivers, while spatial dependence between nearby locations can instead suggest a highly localized infection process [46,47]. Additionally, other time series analyses can provide further epidemiological inference and timescales of spillover risk (e.g., early warning signals) [48].

Spatiotemporal surveillance of human populations, especially those in regular contact with wildlife, is also important for predicting places and times of high spillover risk [22]. For example, focused surveillance for Ebola virus in humans within West and Central Africa could alert health systems to early virus detection and avert costs of containing large outbreaks [21]. However, focusing on spatiotemporal data and predictions in the context of reservoirs and vectors can be particularly informative given the hierarchical nature of spillover, as processes that occur upstream in the pathway to spillover could have a greater influence on human risk [Childs et al., this issue; Washburne et al., this issue]. Identifying regions and times of high pathogen pressure in proposed reservoirs or vectors (e.g., Fig. 1c) could also help prioritize human surveillance. Lastly, interventions focused upstream on reservoirs or vectors could be more cost effective and have sustained influence on minimizing risk [Sokolow et al, this issue].

Below, we highlight three case studies that demonstrate the value of spatiotemporal data and understanding ecological mechanism to generate actionable predictions of spillover risks and to guide long-term interventions (Fig. 2). However, we acknowledge that such efforts can accordingly carry high logistical costs (e.g., years to establish seasonal pulses of infection from reservoir populations), and thus ecologically driven studies should be viewed as complementary to the recipient host surveillance efforts involved in a response to spillover.

Lyme disease

Lyme disease, which is caused by the bacterium *Borrelia burgdorferi* and transmitted to humans by *Ixodes* ticks, is the most common vector-borne disease in the United States [49]. Ticks require three bloodmeals to complete their life cycle. As larvae are born naïve, ticks only obtain *Borrelia*

infection after feeding on a competent small vertebrate host. Infected larvae then molt to become nymphs, which are the most likely life stage to transmit infection to humans. Accordingly, the density of infected nymphal ticks is a useful proxy for human Lyme disease risk [50,51].

Two decades of consistent annual monitoring across replicate sites for the density of infected nymphs and relevant biotic and abiotic covariates (e.g., acorn abundance, small mammal abundance, deer abundance, temperature, precipitation) have shown that Lyme disease risk can be predictable [52–54]. A given year's density of infected nymphs is predicted by the prior year's abundance of white-footed mice and eastern chipmunks and by acorn abundance from two years prior [52]. Years with strong acorn mast generate high rodent abundances in the next year, which subsequently drive high nymphal abundances the following year. Such analyses drawn from years of repeated, ecologically driven surveillance thus show that masting indices can provide relevant spatial and temporal (e.g., two-year lag) predictions for spillover risk (Fig. 2a).

In addition to identifying regions and years for which human risk of exposure to infected nymphs is high, this mechanistic understanding highlights opportunities to interrupt *Borrelia* transmission by targeting small mammal reservoirs and their contribution to hosting and infecting larval ticks. Field experiments have found that direct immunization of white-footed mice can reduce nymphal infection prevalence [55], though effects were strongest when oral bait vaccines were used to immunize the broader small mammal host community [56]. Such ecological approaches were more effective across longer study durations, highlighting how such interventions may be more sustainable [Sokolow et al., this issue]. Ongoing efforts through The Tick Project (https://www.tickproject.org/) are further assessing what combination of acaricide treatments, deployed at which spatial and temporal scales, are most capable of reducing risk [57].

Hendra virus

Hendra virus is a RNA virus that emerged in 1994, causing an outbreak of a lethal respiratory and neurological disease in horses and subsequently to humans that has been followed by over 60 spillover events through 2018 across eastern Australia [58,59]. Flying foxes of the genus *Pteropus* are the natural reservoir hosts, and transmission of virus to horses is assumed to occur through the ingestion of food or water contaminated with urine [39,60]. Surveillance shows pulses of bat viral shedding in winter [61], but these do not follow uniform seasonality [47].

Despite the relatively small number of spillover events compared to Lyme disease cases, a mechanistic understanding of Hendra virus spillover informed by long-term ecological work has likewise provided coarse predictions over space and time [58]. The unprecedented cluster of spillovers in 2011 was preceded by distinct environmental conditions: a rise in the southern oscillation index in 2010 that shifted eucalypt plants, the preferred nectar source of bats, into a growth rather than a flowering phase, in turn restricting food availability [62]. Such food shortages have been exacerbated by agricultural land conversion and cause periods of intense nutritional stress for bats [63,64]. Nutritional stress not only drives flying foxes into urban habitats, where they form sedentary camps near abundant but poor-quality food resources, but also likely amplifies Hendra virus shedding by impairing bat immunity; for example, nutritional stress was associated with greater seroprevalence in little red flying foxes [65,66] (Fig. 2b).

Recently, environmental conditions and weather events in eastern Australia mirrored those seen prior to the 2011 Hendra virus spillovers; a severe El Niño in late 2015 and early 2016 was followed by a rise in the southern oscillation index in winter 2016 and a subsequent food shortage in summer 2016 with concomitant nutritional stress observed in flying foxes [58]. A

spillover event in horses was coincident in space and time with nutritional stress in flying foxes [67]. This chain of predictable events enabled researchers to suggest veterinarians increase winter monitoring of horses and urge precautionary actions such as vaccination [58]. As with Lyme disease, a mechanistic ecological understanding facilitated generating such predictions.

Plasmodium knowlesi

A species of zoonotic malaria, *Plasmodium knowlesi* primarily circulates in monkeys across Southeast Asia. While human infections were first recorded in the 1960s, only since 2004 have larger epidemiological clusters been recorded [4]; in some regions (e.g., Sabah, Malaysia), *P. knowlesi* accounts for the majority of all malaria cases [68]. Wild macaques and leaf monkeys are the primary reservoirs, and transmission occurs through an infectious bite of *Anopheles leucosphyrus* mosquitoes [69]. Most if not all human cases are spillover from primates [70].

Unlike Lyme disease and Hendra virus, most research on *P. knowlesi* has focused on humans, including a large-scale case–control study, MONKEYBAR, in the Philippines and Malaysian Borneo [71]. Human risk factors for infection include recent activities within or at the edge of forests (e.g., vegetation clearance, agriculture) where wild macaques occur [72,73], and human cases have been positively associated with high degrees of local forest cover (2 km) and recent forest loss [74]. In regions where *P. knowlesi* cases are less common, human risk has been associated with land clearing activities within 500 meters of an exposed person's home [75].

These findings suggest land conversion may alter interactions between humans, reservoirs, and vectors differently at distinct spatial scales. Recent work thus used machine learning tools to flexibly consider the effects of multiple spatial scales on malaria cases [76]. Risk was high when the proportion of cleared land within 1 km was low, suggesting that isolated households could be more prone to high vector densities in forested habitats. The effect of deforestation was also high at larger spatial scales (4–5 km), suggesting elevated human exposure to mosquitoes during commutes to agricultural work or a change in macaque behavior or demography [76]. Such large-scale deforestation could reduce reservoir host densities and promote a subsequent behavior change in mosquitoes in these cleared habitats [77]. While identifying the spatial scales of human risk cannot identify the processes by which deforestation drives *P. knowlesi* spillover, such work can narrow the space of possible ecological mechanisms to be elucidated by detailed spatiotemporal studies of reservoirs and vectors (Fig. 2c) [78].

Interplay between phylogenetic and spatiotemporal scales of prediction

The above examples highlight that ecologically minded spatiotemporal studies of reservoirs and vectors can provide actionable predictions that can be used to mitigate spillover risks. These predictions are often coarse in scale and cannot identify the precise spatial location and narrow timepoint of observed spillover events. However, predictions need not be generated on extremely fine scales to be useful for prevention or intervention. For Lyme disease, human risk is greatest in regions with high oak abundance and where acorn masting occurred two years prior. People can become more vigilant in spring and summer when nymphs are questing, and ecological interventions can work to break the transmission cycle between rodents and ticks [52,54,57]. For Hendra virus, veterinarians in regions with active flying fox camps can expect to more closely monitor horse health during the seasons that follow notable shifts in the southern oscillation index or observed nutritional stress events in flying foxes [47,58,62]. Although the ecological mechanisms connecting deforestation to *P. knowlesi* spillover remain less resolved than for

Lyme disease or Hendra virus, observed associations between human cases and different spatial scales of land conversion can guide future ecological studies as well as spatially explicit interventions [74,76,78]. Such work collectively highlights that spillover can be predicted, but at varying phylogenetic, spatial, and temporal scales, which carry distinct benefits and costs.

We here argue that generating phylogenetic predictions, conducting ecologically driven spatiotemporal studies, and surveying human populations are complementary approaches to investigating and predicting spillover. Ecological studies at fine spatial and temporal scales can complement large-scale pathogen discovery projects, macroecology of zoonoses in reservoirs and vectors, and human surveillance (Fig. 1). Pathogen discovery can generate the data needed to narrow the wide range of which reservoirs, vectors, and pathogens should be targets of surveillance. Detailed ecological field studies can next generate a mechanistic understanding of pathogen circulation and spillover that coarsely predict when and where cross-species transmission is most likely. This work can also identify ecological interventions that could occur prior to human exposure and that may have pronounced and long-term impacts on limiting risks. Further, spatiotemporal studies of reservoirs and vectors can also narrow human surveillance efforts toward pathogens with the greatest likelihood for being zoonotic and around those regions and periods where spillover is most likely. Indeed, projects such as PREDICT currently adhere to a similar model by integrating wildlife and human data streams to guide surveillance efforts [25]. Moreover, these various data streams on spatial and temporal infection in reservoirs, vectors, and recipient hosts could help refine and improve macroecological analyses and predictions (Fig. 1d). This pipeline connecting macroecology, pathogen discovery, and surveillance could facilitate synergistic hypothesis generation (i.e., prediction) and testing (i.e., surveillance) to continually refine research efforts. This integrated approach also highlights how spillover predictions can catalyze both discovery and intervention. Greater exchange between macroecology, pathogen discovery, and surveillance (e.g., through interdisciplinary working groups [Becker et al., this issue; [79,80]) could ultimately help generate more actionable predictions and public health interventions to limit pathogen spillover risks.

Acknowledgements

This work was supported by the National Science Foundation (DEB-1716698), the Defense Advanced Research Projects Agency (DARPA D16AP00113), and the DARPA PREEMPT program administered though DARPA Cooperative Agreement D18AC00031. The views, opinions, and/or findings expressed are those of the authors and should not be interpreted as representing the official views or policies of the Department of Defense or the U.S. Government. DJB was also supported by an appointment to the Intelligence Community Postdoctoral Research Fellowship Program at Indiana University, administered by Oak Ridge Institute for Science and Education through an interagency agreement between the U.S. Department of Energy and the Office of the Director of National Intelligence. RKP was also supported by the U.S. National Institutes of General Medical Sciences IDeA Program (P20GM103474 and P30GM110732), Strategic Environmental Research and Development Program (RC-2633), and USDA National Institute of Food and Agriculture (Hatch project 1015891). CLF was supported by the Wellcome Trust (204820/Z/16/Z). EAM was supported by NSF DEB-1518681, the Stanford University Woods Institute for the Environment Environmental Ventures Program, a Hellman Faculty Fellowship, and a Terman Fellowship. We thank Barbara Han, Ellen Ketterson, and three reviewers for helpful feedback on earlier versions of this manuscript.

316
317 **Author contributions**
318 All authors contributed to the development of ideas and to the writing of this manuscript.

References

1. Plowright RK, Parrish CR, McCallum H, Hudson PJ, Ko AI, Graham AL, Lloyd-Smith JO. 2017 Pathways to zoonotic spillover. *Nature Reviews Microbiology* **15**, 502–510.

2. Spengler JR, Ervin ED, Towner JS, Rollin PE, Nichol ST. 2016 Perspectives on West Africa Ebola Virus Disease Outbreak, 2013–2016. *Emerg Infect Dis* **22**, 956–963. (doi:10.3201/eid2206.160021)

3. Fauci AS, Morens DM. 2016 Zika Virus in the Americas — Yet Another Arbovirus Threat. *New England Journal of Medicine* **374**, 601–604. (doi:10.1056/NEJMp1600297)

4. Singh B, Sung LK, Matusop A, Radhakrishnan A, Shamsul SS, Cox-Singh J, Thomas A, Conway DJ. 2004 A large focus of naturally acquired Plasmodium knowlesi infections in human beings. *The Lancet* **363**, 1017–1024.

5. Han BA, Drake JM. 2016 Future directions in analytics for infectious disease intelligence. *EMBO reports* , e201642534. (doi:10.15252/embr.201642534)

6. Washburne AD, Silverman JD, Morton JT, Becker DJ, Crowley D, Mukherjee S, David LA, Plowright RK. 2019 Phylofactorization: a graph partitioning algorithm to identify phylogenetic scales of ecological data. *Ecological Monographs* **0**, e01353. (doi:10.1002/ecm.1353)

7. Münkemüller T *et al.* 2014 Scale decisions can reverse conclusions on community assembly processes. *Global Ecology and Biogeography* **23**, 620–632.

8. Wiens JA. 1989 Spatial scaling in ecology. *Functional ecology* **3**, 385–397.

9. Graham CH, Storch D, Machac A. 2018 Phylogenetic scale in ecology and evolution. *Global Ecology and Biogeography* **27**, 175–187. (doi:10.1111/geb.12686)

10. Han BA, Schmidt JP, Bowden SE, Drake JM. 2015 Rodent reservoirs of future zoonotic diseases. *Proceedings of the National Academy of Sciences* **112**, 7039–7044.

11. Cleaveland S, Laurenson MK, Taylor LH. 2001 Diseases of humans and their domestic mammals: pathogen characteristics, host range and the risk of emergence. *Phil. Trans. R. Soc. Lond. B* **356**, 991–999. (doi:10.1098/rstb.2001.0889)

12. Babayan SA, Orton RJ, Streicker DG. 2018 Predicting reservoir hosts and arthropod vectors from evolutionary signatures in RNA virus genomes. *Science* **362**, 577–580. (doi:10.1126/science.aap9072)

13. Olival KJ, Hosseini PR, Zambrana-Torrelio C, Ross N, Bogich TL, Daszak P. 2017 Host and viral traits predict zoonotic spillover from mammals. *Nature* **546**, 646–650. (doi:10.1038/nature22975)

14. Evans MV, Dallas TA, Han BA, Murdock CC, Drake JM. 2017 Data-driven identification of potential Zika virus vectors. *elife* **6**, e22053.

15. Jones KE, Patel NG, Levy MA, Storeygard A, Balk D, Gittleman JL, Daszak P. 2008 Global trends in emerging infectious diseases. *Nature* **451**, 990–993. (doi:10.1038/nature06536)

16. Han BA, Kramer AM, Drake JM. 2016 Global Patterns of Zoonotic Disease in Mammals. *Trends in Parasitology* **32**, 565–577.

17. 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. (doi:10.1126/science.aap7463)

18. Anthony SJ *et al.* 2013 A strategy to estimate unknown viral diversity in mammals. *MBio* **4**, e00598–13.

19. Rabaa MA *et al.* 2015 The Vietnam Initiative on Zoonotic Infections (VIZIONS): A Strategic Approach to Studying Emerging Zoonotic Infectious Diseases. *Ecohealth* **12**, 726–735. (doi:10.1007/s10393-015-1061-0)

20. Gardy JL, Loman NJ. 2018 Towards a genomics-informed, real-time, global pathogen surveillance system. *Nature Reviews Genetics* **19**, 9–20. (doi:10.1038/nrg.2017.88)

21. Munster VJ *et al.* 2018 Outbreaks in a Rapidly Changing Central Africa — Lessons from Ebola. *New England Journal of Medicine* **379**, 1198–1201. (doi:10.1056/NEJMp1807691)

22. Holmes EC, Rambaut A, Andersen KG. 2018 Pandemics: spend on surveillance, not prediction. *Nature* **558**, 180. (doi:10.1038/d41586-018-05373-w)

23. Pike J, Bogich T, Elwood S, Finnoff DC, Daszak P. 2014 Economic optimization of a global strategy to address the pandemic threat. *PNAS* **111**, 18519–18523. (doi:10.1073/pnas.1412661112)

24. Levin SA. 1992 The problem of pattern and scale in ecology: the Robert H. MacArthur award lecture. *Ecology* **73**, 1943–1967.

25. Morse SS, Mazet JA, Woolhouse M, Parrish CR, Carroll D, Karesh WB, Zambrana-Torrelio C, Lipkin WI, Daszak P. 2012 Prediction and prevention of the next pandemic zoonosis. *The Lancet* **380**, 1956–1965. (doi:10.1016/S0140-6736(12)61684-5)

26. Woolhouse M. 2018 Sources of human viruses. *Science* **362**, 524–525. (doi:10.1126/science.aav4265)

27. Martin G, Yanez-Arenas C, Plowright RK, Chen C, Roberts B, Skerratt LF. 2018 Hendra Virus Spillover is a Bimodal System Driven by Climatic Factors. *EcoHealth* , 1–17.

28. Han BA, Schmidt JP, Alexander LW, Bowden SE, Hayman DT, Drake JM. 2016 Undiscovered bat hosts of filoviruses. *PLOS Negl Trop Dis* **10**, e0004815.

29. Yang X-L *et al.* 2017 Genetically Diverse Filoviruses in Rousettus and Eonycteris spp. Bats, China, 2009 and 2015. *Emerging Infect. Dis.* **23**, 482–486. (doi:10.3201/eid2303.161119)

30. Han BA *et al.* 2019 Confronting data sparsity to identify potential sources of Zika virus spillover infection among primates. *Epidemics* (doi:10.1016/j.epidem.2019.01.005)

31. Walker JW, Han BA, Ott IM, Drake JM. 2018 Transmissibility of emerging viral zoonoses. *PLOS ONE* **13**, e0206926. (doi:10.1371/journal.pone.0206926)

32. Plourde BT, Burgess TL, Eskew EA, Roth TM, Stephenson N, Foley JE. 2017 Are disease reservoirs special? Taxonomic and life history characteristics. *PLOS ONE* **12**, e0180716. (doi:10.1371/journal.pone.0180716)

33. Geoghegan JL, Senior AM, Giallonardo FD, Holmes EC. 2016 Virological factors that increase the transmissibility of emerging human viruses. *PNAS* **113**, 4170–4175. (doi:10.1073/pnas.1521582113)

34. Becker DJ, Streicker DG, Altizer S. 2018 Using host species traits to understand the consequences of resource provisioning for host–parasite interactions. *Journal of Animal Ecology* **87**, 511–525. (doi:10.1111/1365-2656.12765)

35. Thatcher HR, Downs CT, Koyama NF. 2018 Using Parasitic Load to Measure the Effect of Anthropogenic Disturbance on Vervet Monkeys. *EcoHealth* **15**, 676–681. (doi:10.1007/s10393-018-1349-y)

36. Washburne AD, Silverman JD, Morton JT, Becker D, Crowley D, Mukherjee S, David LA, Plowright RK. 2017 Phylofactorization - a graph partitioning algorithm to identify phylogenetic scales of ecological data. *bioRxiv* , 235341. (doi:10.1101/235341)

37. 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. (doi:10.7717/peerj.5979)

38. Amman BR *et al.* 2012 Seasonal Pulses of Marburg Virus Circulation in Juvenile Rousettus aegyptiacus Bats Coincide with Periods of Increased Risk of Human Infection. *PLoS Pathog.* **8**. (doi:10.1371/journal.ppat.1002877)

39. Plowright RK *et al.* 2015 Ecological dynamics of emerging bat virus spillover. *Proc. R. Soc. B* **282**, 20142124. (doi:10.1098/rspb.2014.2124)

40. Tersago K, Verhagen R, Vapalahti O, Heyman P, Ducoffre G, Leirs H. 2011 Hantavirus outbreak in Western Europe: reservoir host infection dynamics related to human disease patterns. *Epidemiology & Infection* **139**, 381–390. (doi:10.1017/S0950268810000956)

41. Pepin KM, Eisen RJ, Mead PS, Piesman J, Fish D, Hoen AG, Barbour AG, Hamer S, Diuk-Wasser MA. 2012 Geographic Variation in the Relationship between Human Lyme Disease Incidence and Density of Infected Host-Seeking Ixodes scapularis Nymphs in the Eastern

United States. *The American Journal of Tropical Medicine and Hygiene* **86**, 1062–1071. (doi:10.4269/ajtmh.2012.11-0630)

42. Jones BA *et al.* 2013 Zoonosis emergence linked to agricultural intensification and environmental change. *Proceedings of the National Academy of Sciences* **110**, 8399–8404.

43. Plowright RK, Peel AJ, Streicker DG, Gilbert AT, McCallum H, Wood J, Baker ML, Restif O. 2016 Transmission or Within-Host Dynamics Driving Pulses of Zoonotic Viruses in Reservoir–Host Populations. *PLOS Neglected Tropical Diseases* **10**, e0004796. (doi:10.1371/journal.pntd.0004796)

44. Restif O *et al.* 2012 Model-guided fieldwork: practical guidelines for multidisciplinary research on wildlife ecological and epidemiological dynamics. *Ecology letters* **15**, 1083–1094.

45. Isaaks EH, Srivastava MR. 1989 *Applied geostatistics*.

46. Pullan RL, Sturrock HJ, Magalhaes RJS, Clements AC, Brooker SJ. 2012 Spatial parasite ecology and epidemiology: a review of methods and applications. *Parasitology* **139**, 1870–1887.

47. Páez DJ, Giles J, McCallum H, Field H, Jordan D, Peel AJ, Plowright RK. 2017 Conditions affecting the timing and magnitude of Hendra virus shedding across pteropodid bat populations in Australia. *Epidemiol. Infect.* , 1–11. (doi:10.1017/S0950268817002138)

48. Brett Tobias S., Drake John M., Rohani Pejman. 2017 Anticipating the emergence of infectious diseases. *Journal of The Royal Society Interface* **14**, 20170115. (doi:10.1098/rsif.2017.0115)

49. Schwartz AM. 2017 Surveillance for Lyme Disease — United States, 2008–2015. *MMWR Surveill Summ* **66**. (doi:10.15585/mmwr.ss6622a1)

50. Mather TN, Nicholson MC, Donnelly EF, Matyas BT. 1996 Entomologic index for human risk of Lyme disease. *American Journal of Epidemiology* **144**, 1066–1069.

51. Ostfeld R. 2010 *Lyme disease: the ecology of a complex system*. OUP USA.

52. Ostfeld RS, Canham CD, Oggenfuss K, Winchcombe RJ, Keesing F. 2006 Climate, Deer, Rodents, and Acorns as Determinants of Variation in Lyme-Disease Risk. *PLoS Biology* **4**, e145. (doi:10.1371/journal.pbio.0040145)

53. Ostfeld RS, Levi T, Keesing F, Oggenfuss K, Canham CD. 2018 Tick-borne disease risk in a forest food web. *Ecology* **99**, 1562–1573. (doi:10.1002/ecy.2386)

54. Schauber EM, Ostfeld RS, Evans AS. 2005 What is the best predictor of annual Lyme disease incidence: weather, mice, or acorns? *Ecological Applications* **15**, 575–586.

55. Gomes-Solecki MJC, Brisson DR, Dattwyler RJ. 2006 Oral vaccine that breaks the transmission cycle of the Lyme disease spirochete can be delivered via bait. *Vaccine* **24**, 4440–4449. (doi:10.1016/j.vaccine.2005.08.089)

56. Tsao JI, Wootton JT, Bunikis J, Luna MG, Fish D, Barbour AG. 2004 An ecological approach to preventing human infection: Vaccinating wild mouse reservoirs intervenes in the Lyme disease cycle. *PNAS* **101**, 18159–18164. (doi:10.1073/pnas.0405763102)

57. Keesing F, Ostfeld RS. 2018 The Tick Project: Testing Environmental Methods of Preventing Tick-borne Diseases. *Trends in Parasitology* **34**, 447–450. (doi:10.1016/j.pt.2018.02.008)

58. Peel A, Eby P, Kessler M, Lunn T, Breed A, Plowright R. 2017 Hendra virus spillover risk in horses: heightened vigilance and precautions being urged this winter. *Australian Veterinary Journal* **95**, N20–N21.

59. Murray K *et al.* 1995 A morbillivirus that caused fatal disease in horses and humans. *Science* **268**, 94–97.

60. Halpin K *et al.* 2011 Pteropid bats are confirmed as the reservoir hosts of henipaviruses: a comprehensive experimental study of virus transmission. *Am. J. Trop. Med. Hyg.* **85**, 946–951. (doi:10.4269/ajtmh.2011.10-0567)

61. Field H *et al.* 2015 Spatiotemporal Aspects of Hendra Virus Infection in Pteropid Bats (Flying-Foxes) in Eastern Australia. *PLOS ONE* **10**, e0144055. (doi:10.1371/journal.pone.0144055)

62. Giles JR, Plowright RK, Eby P, Peel AJ, McCallum H. 2016 Models of Eucalypt phenology predict bat population flux. *Ecology and Evolution* **6**, 7230–7245.

63. Eby P, Richards G, Collins L, Parry-Jones K. 1999 The distribution, abundance and vulnerability to population reduction of a nomadic nectarivore, the Grey-headed Flying-fox Pteropus poliocephalus in New South Wales, during a period of resource concentration. *Australian Zoologist* **31**, 240–253. (doi:10.7882/AZ.1999.024)

64. Eby P, Law B. 2008 Ranking the feeding habitat of grey-headed flying foxes for conservation management. *Canberra: Department of Environment, Heritage, Water and the Arts*

65. Kessler MK *et al.* 2018 Changing resource landscapes and spillover of henipaviruses. *Annals of the New York Academy of Sciences* **1429**, 78–99. (doi:10.1111/nyas.13910)

66. Plowright RK, Field HE, Smith C, Divljan A, Palmer C, Tabor G, Daszak P, Foley JE. 2008 Reproduction and nutritional stress are risk factors for Hendra virus infection in little red flying foxes (Pteropus scapulatus). *Proc. R. Soc. B* **275**, 861–869. (doi:10.1098/rspb.2007.1260)

490 67. Internataional Society for Infection Diseases. 2017 PRO/AH/EDR> Hendra virus, equine -
491 Australia (02): (NS).

492 68. William T *et al.* 2014 Changing epidemiology of malaria in Sabah, Malaysia: increasing
493 incidence of Plasmodium knowlesi. *Malaria journal* **13**, 390.

494 69. Faust C, Dobson AP. 2015 Primate malarias: diversity, distribution and insights for zoonotic
495 Plasmodium. *One Health* **1**, 66–75.

496 70. Lee K-S, Divis PC, Zakaria SK, Matusop A, Julin RA, Conway DJ, Cox-Singh J, Singh B.
497 2011 Plasmodium knowlesi: reservoir hosts and tracking the emergence in humans and
498 macaques. *PLoS pathogens* **7**, e1002015.

499 71. Grigg MJ *et al.* 2014 Factors that are associated with the risk of acquiring Plasmodium
500 knowlesi malaria in Sabah, Malaysia: a case-control study protocol. *BMJ Open* **4**, e006004.
501 (doi:10.1136/bmjopen-2014-006004)

502 72. Cox-Singh J, Singh B. 2008 Knowlesi malaria: newly emergent and of public health
503 importance? *Trends in Parasitology* **24**, 406–410. (doi:10.1016/j.pt.2008.06.001)

504 73. Grigg MJ *et al.* 2017 Individual-level factors associated with the risk of acquiring human
505 Plasmodium knowlesi malaria in Malaysia: a case-control study. *The Lancet Planetary*
506 *Health* **1**, e97–e104.

507 74. Fornace KM *et al.* 2016 Association between landscape factors and spatial patterns of
508 Plasmodium knowlesi infections in Sabah, Malaysia. *Emerging Infectious Diseases* **22**, 201.

509 75. Fornace KM *et al.* 2018 Exposure and infection to Plasmodium knowlesi in case study
510 communities in Northern Sabah, Malaysia and Palawan, The Philippines. *PLOS Neglected*
511 *Tropical Diseases* **12**, e0006432. (doi:10.1371/journal.pntd.0006432)

512 76. Brock PM, Fornace KM, Grigg MJ, Anstey NM, William T, Cox J, Drakeley CJ, Ferguson
513 HM, Kao RR. 2019 Predictive analysis across spatial scales links zoonotic malaria to
514 deforestation. *Proceedings of the Royal Society B* **286**, 20182351.

515 77. Yakob L, Lloyd AL, Kao RR, Ferguson HM, Brock PM, Drakeley C, Bonsall MB. 2018
516 Plasmodium knowlesi invasion following spread by infected mosquitoes, macaques and
517 humans. *Parasitology* **145**, 101–110. (doi:10.1017/S0031182016002456)

518 78. Brock PM, Fornace KM, Parmiter M, Cox J, Drakeley CJ, Ferguson HM, Kao RR. 2016
519 Plasmodium knowlesi transmission: integrating quantitative approaches from epidemiology
520 and ecology to understand malaria as a zoonosis. *Parasitology* **143**, 389–400.

521 79. Wood JLN *et al.* 2012 A framework for the study of zoonotic disease emergence and its
522 drivers: spillover of bat pathogens as a case study. *Phil. Trans. R. Soc. B* **367**, 2881–2892.
523 (doi:10.1098/rstb.2012.0228)

524 80. Daszak P, Epstein JH, Kilpatrick AM, Aguirre AA, Karesh WB, Cunningham AA. 2007
525 Collaborative Research Approaches to the Role of Wildlife in Zoonotic Disease Emergence.
526 In *Wildlife and Emerging Zoonotic Diseases: The Biology, Circumstances and*
527 *Consequences of Cross-Species Transmission* (eds SRSJE Childs, PJS Mackenzie, VMOJA
528 Richt), pp. 463–475. Springer Berlin Heidelberg.

529 81. Baddeley A, Rubak E, Turner R. 2015 *Spatial Point Patterns: Methodology and Applications*
530 *with R*. CRC Press.

531 Plowright RK, Becker DJ, Crowley DE, Washburne A, Huang T, Nameer PO, Han BA.
532 Prioritizing surveillance of Nipah virus in India. *PLoS Neglected Tropical Diseases*. 2019.
533 10.1371/journal.pntd.0007393
534

Tables

Table 1. Summary of select approaches to predict pathogen spillover, stratified by the goal and the scales of predictions.

Prediction goal	Approach	Scale	Example method	Source
Emergence of *novel* zoonotic pathogens	Reservoir or vector surveillance	Phylogenetic (reservoir, vector, pathogen), space	Quantify pathogen diversity across reservoirs or vectors globally	[17]
	Macroecology	Phylogenetic (reservoir, vector, pathogen), space[1]	Machine learning to identify likely or novel reservoirs or vectors	[10,12]
	Human surveillance	Phylogenetic (pathogen), space	Survey human populations with high wildlife contact	[22]
Where and when *known* pathogens may spillover	Macroecology	Phylogenetic (host), space[1]	Machine learning to identify which reservoirs to sample	[14]
	Human surveillance	Space, time	Improve diagnostic capacity to detect early zoonotic outbreaks	[21]
	Early warning signals	Time	Detect transition from stuttering chains to sustained transmission	[48]
	Reservoir surveillance	Space, time	Identify ecological predictors of pathogen spillover (Fig. 2)	[52,58]

[1]Space is here implicit through mapping the distribution of predicted reservoirs or vectors

Figures

Figure 1. Interplay between scales of pathogen spillover prediction and a proposed pipeline for their integration. Even coarse phylogenetic predictions (*a*) can narrow the scope of what reservoirs and vectors for a given set of pathogens, and in which geographic regions, should be prioritized for surveillance (*b*); examples based on trait-based or cladistic analyses include filoviruses in Neotropical bats [28], zoonotic pathogens in European rodents [10], and helminths in Old world primates [34]. Spatiotemporal studies can next elucidate how zoonotic pathogens circulate in reservoir or vector populations, identify broad spatial and temporal scales at which pathogen pressure (e.g., shedding) is greatest, and uncover the ecological mechanisms leading to spillover (*c*). Based on these regions and times where risk is greatest (circles), managers can design preemptive interventions and prioritize human surveillance. Data from spatiotemporal studies can further address information gaps and refine future macroecological analyses and predictions in an iterative fashion (*d*). The map (*b*) is adapted from Han et al. [10], and perspective plots (*c*) were generated using random realizations of a binomial point process with varying intensities [81]; both are used here simply as heuristic devices.

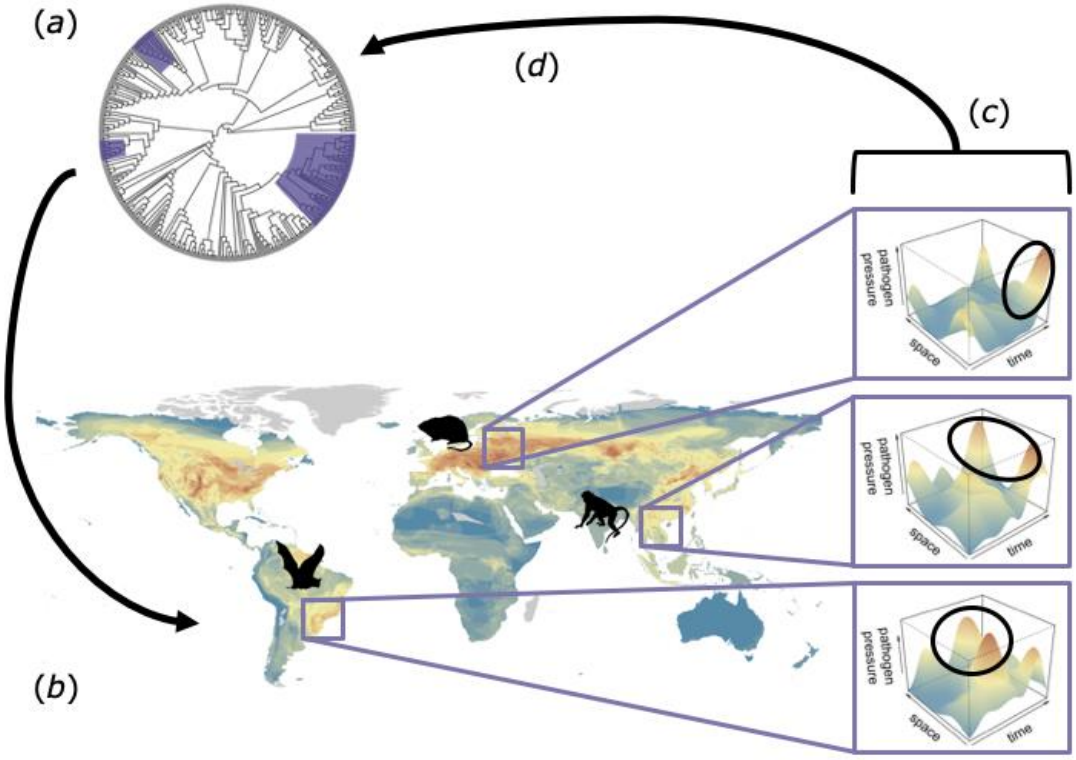

Figure 2. Predictive insights into pathogen spillover risk gained from long-term, spatiotemporal studies of reservoir hosts and vectors. For Lyme disease in North America (*a*) and Hendra virus in eastern Australia (*b*), ecological and mechanistic approaches have identified both spatial and temporal proxies for spillover risk in recipient hosts. Columns indicate ecological correlates of spillover risk at varying time lags, and colors represent differentiate those that occur in autumn and winter (blue) or in spring and summer (yellow) for systems with strong seasonality (*a*–*b*). The ecological mechanisms linking land clearance with *P. knowlesi* spillover in Southeast Asia are less well understood (*c*), but analyses of spatial scale and human cases have generated hypotheses for spatiotemporal studies of reservoirs and vectors.