ECOLOGY LETTERS

Ecology Letters, (2020) **23**: 1178–1188 doi: 10.1111/ele.13520

IDEAS AND PERSPECTIVES

Integrating data mining and transmission theory in the ecology of infectious diseases

Barbara A. Han,[1]* Suzanne M. O'Regan,[2] John Paul Schmidt[3,4] and John M. Drake[3,4]

The peer review history for this article is available at https://publons.com/publon/10.1111/ele.13520

Abstract

Our understanding of ecological processes is built on patterns inferred from data. Applying modern analytical tools such as machine learning to increasingly high dimensional data offers the potential to expand our perspectives on these processes, shedding new light on complex ecological phenomena such as pathogen transmission in wild populations. Here, we propose a novel approach that combines data mining with theoretical models of disease dynamics. Using rodents as an example, we incorporate statistical differences in the life history features of zoonotic reservoir hosts into pathogen transmission models, enabling us to bound the range of dynamical phenomena associated with hosts, based on their traits. We then test for associations between equilibrium prevalence, a key epidemiological metric and data on human outbreaks of rodent-borne zoonoses, identifying matches between empirical evidence and theoretical predictions of transmission dynamics. We show how this framework can be generalized to other systems through a rubric of disease models and parameters that can be derived from empirical data. By linking life history components directly to their effects on disease dynamics, our mining-modelling approach integrates machine learning and theoretical models to explore mechanisms in the macroecology of pathogen transmission and their consequences for spillover infection to humans.

Keywords

Boosted regression, disease dynamics, disease macroecology, pathogen transmission, random forest, statistical learning, zoonosis, zoonotic spillover.

Ecology Letters (2020) **23**: 1178–1188

INTRODUCTION

Naturalists have long observed myriad features distinguishing species from one another. While some of these features capture minor phenotypic differences within a population, others reflect more intrinsic differences among species. Suites of multiple correlated features capture variation in life history across species (Montiglio *et al*. 2018; Polverino *et al*. 2018). These traits have evolved to balance fitness given manifold selection pressures in complex environments where individuals contend with fitness costs and other selection pressures, such as those imposed by parasites and pathogens. As a result, functional traits serve as reliable proxies for complex organismal variables that are more difficult to measure, such as immune strategies, which may strongly influence host responses to pathogens and parasites, and impact disease dynamics in hosts (Lochmiller & Deerenberg 2000). The idea that differences in disease equilibria and dynamics (e.g., prevalence, invasibility, outbreak size) reflect observable differences in host traits has culminated in a number of paradigm hypotheses: (1) The *pace-of-life hypothesis*: Faster living hosts should be more competent (and therefore have higher intensity and prevalence) compared to slow-living species (Johnson *et al*. 2012; Previtali *et al*. 2012; Huang *et al*. 2013; Ostfeld *et al*. 2014); (2) The *invasibility hypothesis*: The invasibility of some host populations by infectious pathogens scales with host behaviors linked to transmission (i.e., R_0 is higher, (Han *et al*. 2015); and (3) The *density-outbreak hypothesis*: Outbreak size should vary by host species as a function of longevity and population density (Dobson 2004; Hily *et al*. 2014).

Other trait-based hypotheses readily follow. The (4) *social transmissibility hypothesis* posits increased opportunities for contact in social species leading to higher rates of transmission by directly transmitted microparasites (Han *et al*. 2015) but reduced transmission of macroparasites (Bordes *et al*. 2007) relative to solitary species. Similarly, species that forage widely may encounter and harbour a higher diversity of parasites than species with smaller home ranges (Ezenwa 2004; Bicca-Marques & Calegaro-Marques 2016). A corollary of this (5) *foraging-diversity hypothesis* is that species with broad diets may exhibit greater tolerance for trophically transmitted parasites (Gutiérrez *et al*. 2019). These hypotheses propose mechanisms by which host traits impact infection dynamics, and identify numerous host traits that may provide insight to the

[1]*Cary Institute of Ecosystem Studies, Box AB Millbrook, NY 12571, USA*
[2]*Department of Mathematics and Statistics, North Carolina A&T State University, 1601 E. Market St., Greensboro, NC 27411, USA*
[3]*Odum School of Ecology, University of Georgia, 140 E. Green St., Athens, GA 30602, USA*
[4]*Center for the Ecology of Infectious Diseases, University of Georgia, 203 D.W. Brooks Drive, Athens, GA 30602, USA*
**Correspondence: E-mail: hanb@caryinstitute.org*

ecological processes underpinning pathogen transmission, which is comparatively difficult to measure across host species.

Previous analyses have demonstrated the utility of species-level traits for predicting transmission risk. Machine learning applied to species-level traits has accurately predicted the zoonotic reservoir status of mammal hosts (Han *et al*. 2015b; Han *et al*. 2019; Plowright *et al*. 2019), the zoonotic vector status of mosquitoes and ticks (Evans *et al*. 2017; Yang & Han 2018) and potential human-to-human transmissibility of zoonotic viruses (Evans *et al*. 2017; Walker *et al*. 2018; Yang & Han 2018; Han *et al*. 2019; Plowright *et al*. 2019). Beyond predictive analytics of zoonotic diseases, machine learning methods have been increasingly applied to modern ecological questions as the volume of available digital data has grown (Peters *et al*. 2014; LaDeau *et al*. 2017). This is in part because machine learning algorithms are adept at learning associations among multiple variables in high dimensional datasets where collinearities and complex hidden interactions are likely to be present but difficult to anticipate *a priori*.

As purely statistical models, machine learning algorithms identify patterns in host traits directly from the trait data themselves without assumptions about underlying ecological processes (Breiman 2001; Hochachka *et al*. 2007). In contrast, mean field models of infectious disease dynamics are theoretical representations of the transmission processes that drive epidemiologic outcomes. These outcomes are themselves strongly influenced by traits. For instance, host traits corresponding to a fast life history strategy (e.g., high population density, short lifespan) are predicted to influence outbreak size, equilibrium prevalence and infectious period. Such traits accurately distinguish among host species in ways that may be consequential for how infection varies through time in their populations. Thus, by linking models of disease dynamics to traits, we can generate predictions about disease dynamics and the risk of disease spillover based on the kinds of hosts present in a system. Moreover, parameterizing dynamical systems models in terms of traits may provide unifying explanations for disease transmission phenomena across multiple host species, such as in the paradigm hypotheses enumerated above (see Box 1), and may prove particularly useful for emerging disease for which prior knowledge of pathogen transmission is often lacking, but for which trait data may be more available.

Here we demonstrate how data mining via machine learning can be combined with mathematical models to generate qualitative predictions of disease dynamics across species. Using rodents and their zoonotic pathogens and parasites (i.e., those transmissible from rodents to humans) as an example, we show how this approach can be applied to trait data to bound the range of dynamical phenomena associated with particular host groups. We then examine associations between estimates of a key epidemiological quantity (equilibrium prevalence) and empirical data on rodent-borne zoonoses in humans to generate testable predictions related to zoonotic spillover risk. To illustrate the generalizability of this data mining-modelling framework, we present a rubric of transmission models and parameters that may be empirically informed. These models cover a range of host-pathogen systems and can be used to explore mechanisms and paradigm hypotheses about disease macroecology and spillover epidemiology (Box 1).

Traits of host species

Due to their high species richness and importance as zoonotic reservoirs (Han *et al*. 2016a), rodents are remarkably well studied, with trait data available for a large proportion of species (Jones *et al*. 2009). We therefore use rodents as an example to investigate features that reliably distinguish zoonotic host from non-host species. By parameterizing transmission models according to traits associated with the propensity to harbour zoonoses, we examine whether such models generate patterns of disease dynamics that give comparative estimates of the risk of human exposure and infection. Specifically, trait values inform SIR models that return a theoretical prediction of equilibrium prevalence, a metric which may indicate exposure risk in humans. Outbreak size and R_0, closely related metrics that we do not calculate here, can also be investigated in the same way. Solving for equilibrium prevalence of a directly transmitted pathogen allows us to determine whether differences in host traits lead to higher mean equilibrium prevalence in some rodent species compared to others. If so, these species may be expected to pose a greater risk of spillover transmission to humans.

This broad template can be used to explore the range of possible dynamics that may emerge when multiple species are affected by the same or similar pathogens. For example, we can describe the different classes of transmission or disease dynamics that are possible within a set of constraints set by species' intrinsic features (Table 2). We can directly compare dynamical possibilities among multiple species infected by a common pathogen (even between species not in the same clade). We can ascertain invasibility of a population by a pathogen, or quantify vulnerability of different species susceptible to fatal infection, prior to an empirical challenge. We can also draw testable/verifiable predictions to human spillover infection, for instance, which species are responsible for most of the human transmission; or conversely, when some species have high equilibrium prevalence but are clearly not responsible for the majority of spillover transmission due to extrinsic factors (for example, mismatched phenology, or diluting effects of host community composition (Civitello *et al*. 2015; Keesing & Ostfeld 2015)). In the analyses that follow, we test model predictions directly by mapping the geographic ranges of rodents in the top 10 percent of equilibrium prevalence. We then compared their global distributions with human emerging zoonoses fitting our criteria.

Parameterizing models with host traits

It is well known that certain intrinsic features such as life history traits influence infection dynamics in host populations, as represented by their effects in mechanistic models of disease dynamics. For example, in mammals, birth rates change predictably with litter size and number of litters per year; and natural mortality is the reciprocal of lifespan. Transmission rates depend on contact, which is likely to be affected by traits such as social group size and home range size (Carslake

Box 1. Exploring paradigm hypotheses in disease ecology

In addition to identifying likely reservoirs and regions of human spillover through trait-based dynamics, model parameters can be recast to explore paradigm hypotheses in disease ecology. For example, the *pace-of-life hypothesis* recognizes tradeoffs in species along a life history continuum ranging from fast- to slow-living species (Montiglio *et al*. 2018). Host traits capturing 'pace of life' include those controlling host demography, which reflect fitness tradeoffs between reproduction (e.g., birth rates) and longevity and immune defense strategies. Trait patterns are postulated to influence the likelihood that a parasite successfully invades a susceptible host population (Johnson *et al*. 2012; Han *et al*. 2015, 2016b), the rate of pathogenesis in individual hosts (Cable *et al*. 2007) and the likelihood of population recovery from an outbreak (Cross *et al*. 2005). SIR models parameterized by life history traits suggest that species that live fast (ie, produce more litters per year than others) and die young (ie, are shorter lived than others) have the highest equilibrium prevalence.

In more tailored models, we may also test for differences among host species in their response to warming climates. Recent debate centres on the role warming is playing in the dynamics of human and wildlife diseases whose ranges are postulated to either expand or contract with climate change. Climate effects on disease dynamics may be mediated by many factors. As examples, warming may alter parasite abundance (Gehman *et al*. 2018), shorten parasite developmental periods (Kutz *et al*. 2005), lead to changes in the timing of host and parasite overlap (Altizer *et al*. 2011), alter contact between human hosts and vectors (Pascual & Bouma 2009; Ryan *et al*. 2019) and affect host ability to cope physiologically with infection (Harvell *et al*. 1999). Mining zoonotic pathogen and parasite trait data combined with host traits and mechanistic models, provides a comparative means for testing climate-related hypotheses about infectious diseases (Lafferty & Mordecai 2016), such as the *warmer-means-sicker hypothesis*, across hundreds of host-pathogen pairs.

The rubric in Table 2 identifies a large space of possible models to be parameterized to investigate these and other specific hypotheses. The detailed mechanics of these theoretical models will necessarily depend on which features are most important statistically. Each row corresponds to a transmission mode; quantitative dynamics for each mode are expressed by compartmental models of different classes, e.g., SIR, Ross-MacDonald and 'contaminated environment' transmission systems (column 1). Terms in these equations, such as the birth rate (b(N)), force-of-infection (λ) and infectious period ($1/\gamma$), among others, are hypothesized to be mechanistically influenced by traits of hosts (column 2), parasite traits (column 3) or environmental factors (column 4). To generate testable predictions, these terms can be re-written as functions of measurable traits. For instance, birth rate may be written as a function of body size (in keeping with known allometric relations (Peters 1986)) or virulence may be written as a function of case fatality rates (e.g., (Li *et al*. 2008)). These highly parameterized models can then be used to study how R_0, epidemic period, critical community size and related epidemiological properties are affected by underlying traits.

et al. 2005; Han *et al*. 2015). Mechanistic models also predict that infectious disease prevalence and the basic reproduction number are functions of contact rate and life history parameters such as per capita birth rate, per capita mortality and population density. Thus, statistically identified differences in species traits have dynamical consequences for disease transmission in population models. Directly incorporating features identified from host data into compartmental disease models accounts for the combined effects of these traits on comparative disease dynamics among species (Fig. 1).

We considered the diversity of pathogens found among rodents ($N = 2276$) and identified the transmission modes among rodent-borne zoonoses (Fig. 2, Table S1). In addition to zoonoses transmitted by arthropod vectors, many rodent-borne zoonoses are caused by pathogens that are transmitted environmentally (e.g., via contact with infectious fomites such as aerosolized excreta), or directly via close contact with an infected individual. Since the dynamics of many environmentally transmitted pathogens are well-approximated by models assuming direct transmission ((May & Anderson 1979); Appendix), we consider a system of ordinary differential equations that assumes a density-dependent direct transmission process.

Our model is a mechanistic SIR model for microparasite infections with density-dependent transmission,

$$\begin{aligned} \tfrac{dS}{dt} &= b_0(S+I+R) - b_1(S+I+R)^2 - \beta SI - \mu S, \\ \tfrac{dI}{dt} &= \beta SI - (\gamma+\mu)I, \\ \tfrac{dR}{dt} &= \gamma I - \mu R, \end{aligned} \tag{1}$$

where $N = S + I + R$, b_0 is the per -individual birth rate, μ is the per-individual natural mortality rate,

$r = b_0 - \mu$, K is the carrying capacity, $b_1 = r/K$ is the strength of density dependence, β is the per-individual transmission rate and γ is the per-individual recovery rate. In reservoir hosts, the disease-induced mortality rate is assumed to equal zero, and therefore, at equilibrium, the population size N is equal to the carrying capacity, K. When solving $\frac{dS}{dt} = 0$ for the number of infectious individuals at the endemic steady state I^*, we therefore assume $S + I + R = K$ and then obtain equilibrium prevalence by dividing I^* by K, provided

$$\begin{aligned} R_0 &= \frac{\beta K}{(\gamma+\mu)} > 1, \\ I^*/K &= \mu\left(\frac{1}{\gamma+\mu} - \frac{1}{\beta K}\right). \end{aligned} \tag{2}$$

We next examined how equilibrium prevalence (eqn (2)) varies over trait-dependent parameter values that are bounded according to real data. In what follows, we distinguish between the mechanistic parameters defined above, and m measurable quantities represented by $\theta_1, \theta_2, \ldots \theta_m$. For example, the per-

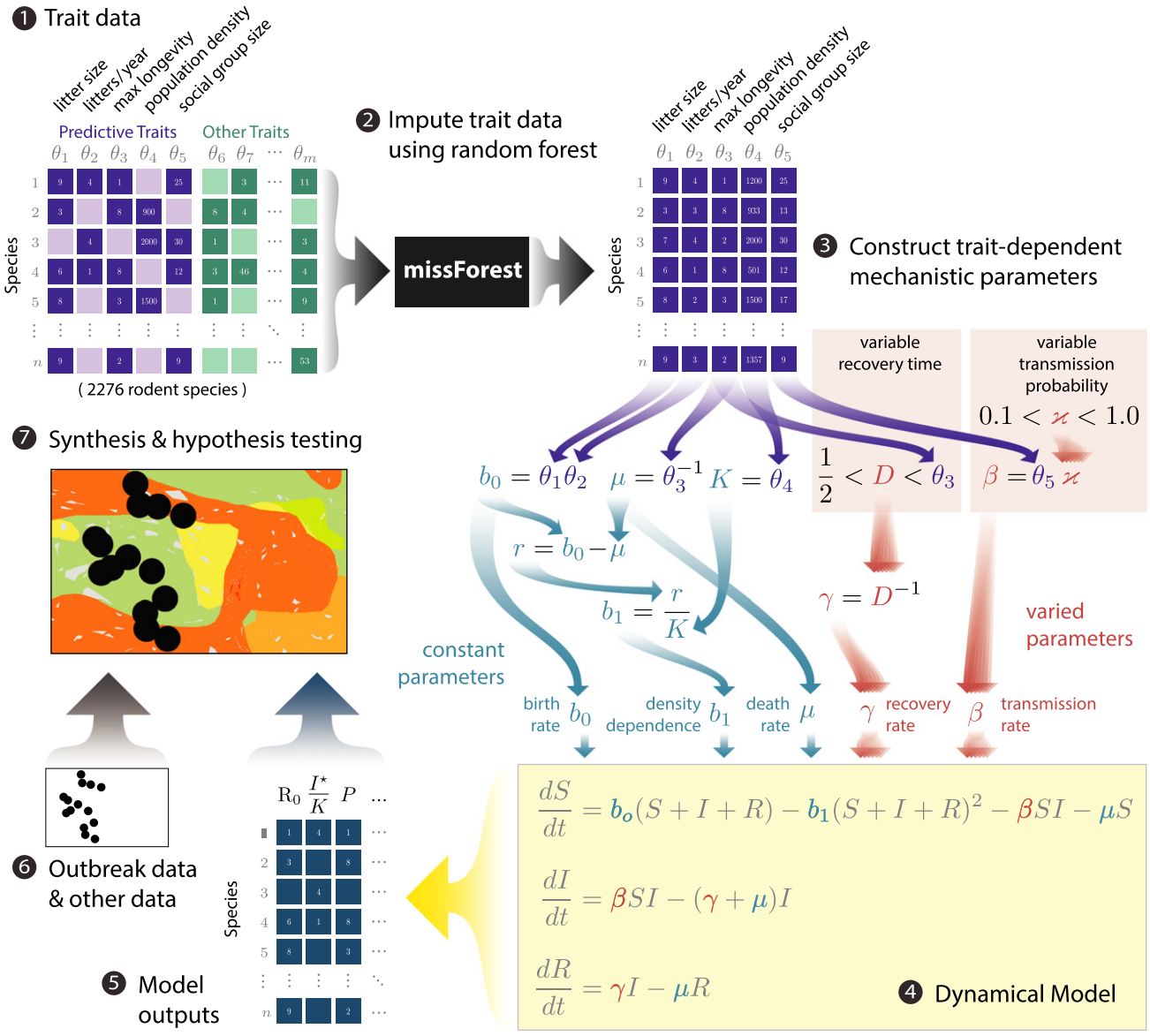

Figure 1 Model parametrization from observable trait data. An incomplete data matrix of species traits (1) is imputed using machine learning to predict the most probable values (2). Species traits determine key parameters of mechanistic models (3) that then inform dynamical models (4), in this case, an epidemiological SIR model is parameterized for each species. Model outputs for all species (5) are compared against relevant empirical data (e.g., human outbreak data, approximating zoonotic spillover events from reservoir species) (6). Model outputs can be visualized by generating a heatmap of values (e.g., R_0, equilibrium prevalence) for overlapping species range polygons and overlaying known human zoonotic outbreak locations (7) to generate hypotheses about the macroecology of zoonotic disease and the geographic risk of human spillover transmission.

individual birth rate b_0 for each species is the product of litter size θ_1 and number of litters per year θ_2. The natural mortality rate μ is given by the reciprocal of longevity θ_3 (in years) and the carrying capacity, K, is defined as the population density per sq. km, θ_4. The strength of density dependence b_1 is a function of life history traits $\theta_1, \theta_2, \theta_3, \theta_4$ (Table 1). For the purposes of this analysis, the per-individual transmission rate β and recovery rate γ are assumed to be parasite-dependent tuning parameters. To account for the variability inherent in transmission and recovery, we computed equilibrium prevalence for each species j over a species-specific grid of 100 (β, γ) values. The per-individual transmission rate β is defined as $\kappa\theta_5$, where social group size θ_5 is a proxy for the number of contacts made with conspecifics, and κ is the probability of transmission given contact. Since κ is likely to depend on the parasite, we varied κ between 0.01 and 1 in increments of 0.1, yielding ten notional values for each species. Recovery time γ was defined as the reciprocal of recovery rate. To capture infections across a spectrum of recovery times, we assumed recovery times varied between a minimum of ½ year and the maximum longevity, θ_3, of each host species (i.e., lifelong infection). Ten equally spaced values between 2 and $1/\theta_3$ were used for γ for each species j. Trait-dependent model parameters are summarized in Table 1.

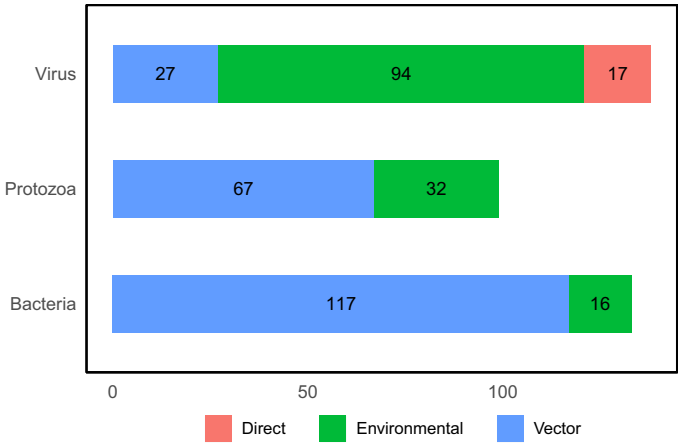

Figure 2 Transmission modes of pathogens and parasites causing 370 rodent-borne zoonoses in 202 host species (of 2276 total rodent species).

To calculate mean SIR equilibrium prevalence P_j using the trait profile of each species j, we put each of the 100 (β, γ) pairs together with the fixed species-specific traits $\theta_i, i = 1, .., 5$ into the SIR equilibrium prevalence formula (2), producing an array of 100 prevalence values I_i/K_j. We then calculated the mean equilibrium prevalence over each species array,

$$P_j = \frac{\sum_{i=1}^{100} \frac{I_i}{K_j}}{100}, j = 1, ..., 2276.$$

Missing data for the five key traits (litter size, litters per year, maximum longevity, population density, social group size) were imputed using the function missForest (Stekhoven & Bühlmann 2012) in the R package of the same name (Stekhoven & Bühlmann 2012). The missForest algorithm uses a random forest trained on the observed values of a data matrix to predict the missing values. The algorithm incorporates complex interactions and nonlinear relations, and yields an out-of-bag imputation error estimate by comparing, at each iteration, model predictions to empirical values for data not included to train the model. Here, error values were litter size ($RMSE$ = 0.04), litters/yr ($RMSE$ = 0.069), max longevity ($RMSE$ = 2.63 months), population density ($RMSE$ = 0.07 individuals/km^2) and social group size ($RMSE$ = 0.12 individuals). For all 2,276 species, life history traits $\theta_i, i = 1, .., 5$ were fixed at the values obtained from the data or from imputation.

Table 1 Trait-dependent mechanistic parameters used in SIR model. Mechanistic parameters depend on measurable quantities obtained θ_i, i = 1,. . ., 5, from table of traits for all rodent species

Parameter	Function of Traits θ_i	Source
Per-individual birth rate $b_0(\theta_1, \theta_2)$	Litter size × Litters per year $\theta_1 \times \theta_2$	θ_1 and θ_2 from Data/Imputed data
Per-individual mortality rate $\mu(\theta_3)$	1/Max Longevity (years^{-1}) $1/\theta_3$	θ_3 from Data/Imputed data
Carrying capacity $K(\theta_4)$	Population density per sq. km θ_4	θ_4 from Data/Imputed data
Strength of density dependence $b_1(\theta_1, \theta_2, \theta_3, \theta_4)$	$\frac{b_0-\mu}{K} = \frac{\theta_1 \times \theta_2 - 1/\theta_3}{\theta_4}$ per year density per sq. km	θ_1, θ_2, θ_3 and θ_4 from Data/Imputed data
Per-individual transmission rate $\beta(\theta_5)$	Probability of transmission given contact × social group size $\kappa \times \theta_5$	κ varied between 0.01 and 1. θ_5 from Data/Imputed data
Per-individual recovery rate $\gamma(\theta_3)$	1/(recovery time) (years^{-1}) $1/D$, $1/2 \le D \le \theta_3$	Recovery time D varied between 1/2 year and max longevity θ_3

Rodent-borne zoonoses

To a first approximation, for a pathogen that is density dependent and transmitted via direct contact between animal reservoirs to humans, the risk of human infection is likely to scale with host prevalence (Plowright *et al.* 2017). To investigate this, we cross-examined species with high equilibrium prevalence from our model with zoonotic disease data collected from wild rodent species. Specifically, we identified directly transmitted rodent-borne zoonoses and their rodent host species from the primary literature and from the GIDEON data repository (Berger 2005).

Among 2276 rodents, 156 species did not have geographic ranges that could be assigned to ecozones due to data deficiency or recent extinction, and three species were subsequently found to be reassigned as other species due to changing taxonomic standards (N = 2117). There were many unique species whose features generated high values of mean equilibrium prevalence (537 in the top quartile of prevalence, 212 in the 90th percentile; Table S2). These species are geographically widespread, ranging across all major ecozones except Antarctica (Fig. 3). For example, in the top 10 percent of 'high prevalence' species, 29 species are confirmed hosts associated with 22 unique zoonoses. Of these zoonoses, seven are caused by pathogens that are directly or environmentally transmitted (Table S3). The majority of rodent reservoir species whose features generated high equilibrium prevalence (N = 23/29) carry at least one of these seven zoonoses.

Wild reservoirs whose traits suggest an intrinsic capacity to generate high equilibrium prevalence represent surveillance targets for pathogen discovery, particularly for directly transmitted pathogens including zoonotic viruses and their strain diversity in shared human environments. Among species in the 90th percentile prevalence, there were two species that are currently not known to carry any zoonoses (e.g., *Myopus schistocolor* and *Eligmondontia typus*, Fig. 3). In contrast, other 'high prevalence' species are known to harbour zoonoses, but none that are directly transmitted. For example, traits of *Abrothrix olivaceus*, *Microtus pennsylvanicus* and *Peromyscus gossypinus* generated among the highest equilibrium prevalence values, but only *P. gossypinus* is associated with a directly transmitted zoonosis (Hantavirus pulmonary syndrome) while the other two species are only known to carry zoonoses caused by macroparasites, or pathogens transmitted

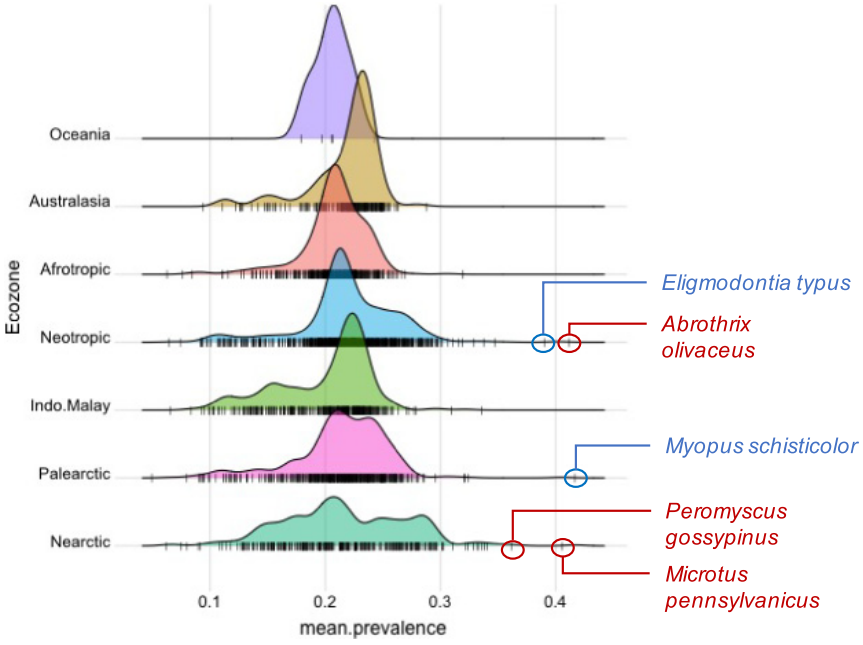

Figure 3 Mean equilibrium prevalence values generated by an SIR density-dependent transmission model parameterized according to intrinsic traits of rodents. The jittered margin rugs represent the distribution of equilibrium prevalence values for all rodent species in each ecozone (jitter width = 0.02). Species whose geographic ranges spanning multiple ecozones are included in the plots for every ecozone in which they occur. Circles identify particular species whose traits generated the highest mean equilibrium prevalence values, with red circles indicating confirmed reservoir species and blue circles indicating species not currently known to carry any zoonoses.

primarily through water (cryptosporidiosis, multilocular echinococcosis, American trypanosomiasis) (Fig. 3).

Global maps of species in the top quartile of equilibrium prevalence reveal regions of co-occurrence (geographical range overlap) that overlap with outbreaks of several directly transmitted rodent-borne zoonoses observed in humans since the 1950s (Fig. 4), with some variation in the range of equilibrium prevalence values exhibited by overlapping species (Fig. 5). There are several rodent-borne zoonoses found in South America, where rodent species richness is particularly high and where we also see the highest numbers of high prevalence (top quartile) species. We also observe places where there are no rodent species in the top quartile of mean prevalence (i.e., the highest observed prevalence values are still relatively low) - South Asia, most of Australia and New Zealand, the Philippines, the Arabian peninsula, Madagascar and much of Saharan and southern Africa. These locations also have few outbreaks. Most of the locations where we observe outbreaks of rodent-borne zoonoses show four or more rodent species in the top quartile of mean equilibrium prevalence. The exception to this is central Africa where monkeypox appears to spillover in the absence of multiple high prevalence species. Comparison of our maps with GBIF records suggests that the IUCN range map for *Lemniscomys striatus* should be extended to include all monkeypox outbreaks and may reflect a more general need for updated species range maps, particularly for the African subcontinent whose rodent biogeography is not as well studied.

Human disease outbreaks caused by rodent-borne pathogens include Argentine, Brazilian, Bolivian, Venezuelan, and Whitewater Arroyo haemorrhagic fevers; hantaviruses (Old World, and hantavirus pulmonary syndrome); and Monkeypox (incidentally, all of these pathogens are viruses). To calculate the degree of concordance between equilibrium prevalence predictions and outbreak occurrence, we conducted nonparametric Wilcoxon rank sum tests comparing the highest value of equilibrium prevalence and the richness of high prevalence species at each georeferenced outbreak location. We compared these values to the same measures at an equal number of background locations (n = 114) that were randomly generated across the global distribution of rodents. The highest value of estimated mean prevalence (P = 0.0094) and the number of species in the top quartile of estimated mean prevalence (P = 0.00029) were both significantly greater at outbreak locations than at background points. To check that these differences are robust to choice of background locations, we repeated tests for each of 1000 random sets of background points. Outbreak locations were significantly greater ($P < 0.05$) in 99% of comparisons of number of top quartile species and in 91% of comparisons of highest prevalence value (Fig. S1).

Targeting surveillance

As none of these viruses is known to have sustained human-to-human transmission, repeated human outbreaks likely represent spillover events from sylvatic reservoirs (Fine *et al.* 1988; Hutin *et al.* 2001). For example, rodent species exhibiting traits leading to high equilibrium prevalence and that overlap geographically with hantavirus outbreaks in humans (New World) should receive enhanced surveillance for hantavirus. In this context, *P. gossypinus* could be a focal species. In addition to exhibiting high equilibrium prevalence, it is an important reservoir in the nearctic region (primarily in the United States) where seasonal population dynamics, especially in response to heavy rainfall and ephemeral flooding, may temporarily displace *Peromyscus leucopus* as the main reservoir of hantavirus spillover to humans in some areas (Barko & Feldhamer 2002; Tian & Stenseth 2019). In the neotropics, surveillance might centre on *Abrothrix olivaceus* and *Eligmodontia typus* as potential reservoirs for hantavirus given their high equilibrium prevalence for directly transmitted pathogens

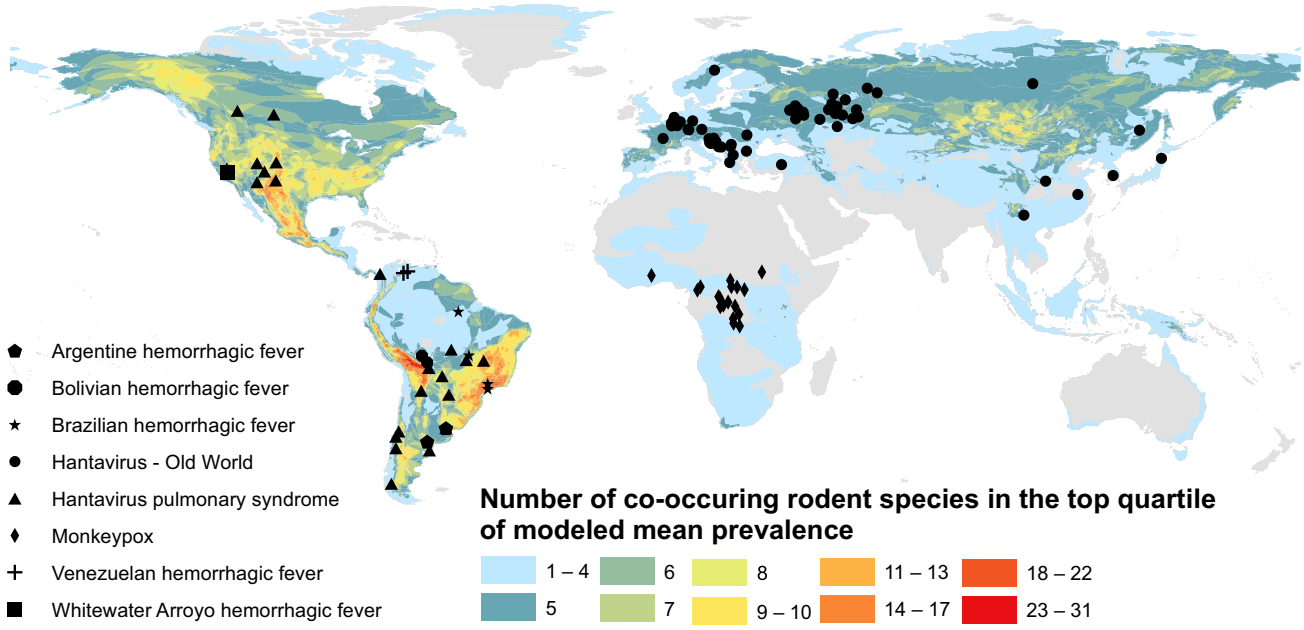

Figure 4 Colour bands show the richness of rodent species (n = 537) whose traits generate high equilibrium prevalence (top quartile) of density dependent, directly transmitted zoonoses. Points represent the locations of human outbreaks of eight zoonoses that have been recently recorded.

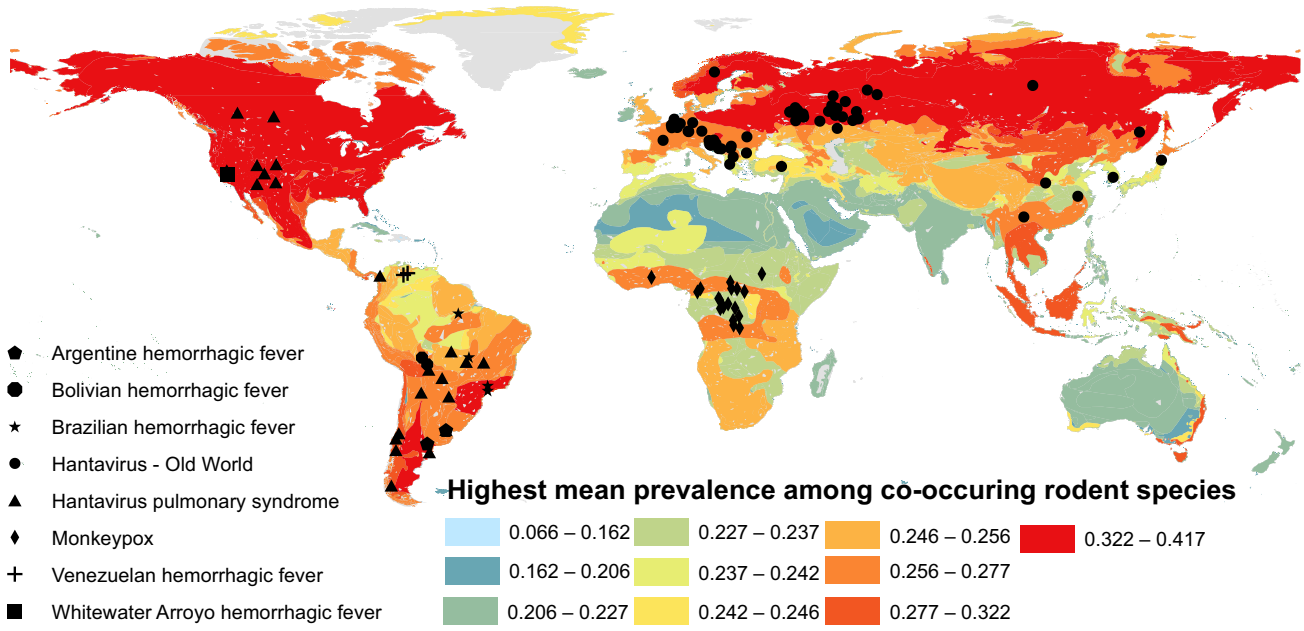

Figure 5 Mean equilibrium prevalence of the rodent species with the highest model-generated value at a given location. Globally, of 2117 species for which IUCN geographic range maps were available, 866 rodent species had the highest value of modelled mean equilibrium prevalence among co-occurring rodent species. Mean equilibrium prevalence values were generated by parameterizing a density dependent, directly transmitted model of rodent zoonoses using empirical and imputed traits of each species

and geographic overlap with human hantavirus outbreaks (particularly in Chile, (Astorga *et al.* 2018). In addition to understanding reservoir diversity, trait-based model results could also support the design of surveys for hantavirus diversity by prioritizing the subset of species with traits predicting high equilibrium prevalence (Luong *et al.* 2011). Similarly, surveillance for novel directly transmitted zoonotic pathogens might focus on *Myopus schistocolor*, *Eligmondontia typus*, *Abrothrix olivaceus* and *Microtus pennsylvanicus* (Fig. 3) in ecozones in which they are synanthropic or exhibit seasonal or increasing contact with humans due to patterns of land use change.

In the Afrotropical ecozone, monkeypox is a major directly transmitted zoonosis. Monkeypox outbreaks are most common in the Democratic Republic of Congo, but are increasing in frequency across the Afrotropical ecozone (Durski *et al.*

2018) (Fig. 4). Although sylvatic reservoirs have not been confirmed for monkeypox, direct contact with infected mammals, especially rodents in peridomestic settings, is suspected as the dominant mode of spillover transmission to humans (Khodakevich *et al.* 1988; Hutin *et al.* 2001). In the Afrotropics, of the three rodent species whose traits generated equilibrium prevalence in the 90th percentile (and the 29 species in the 80th percentile) from trait-based SIR models, none are currently known to be reservoirs of directly transmitted human pathogens. However, among these high prevalence species, those whose geographic ranges overlap recent human outbreaks of monkeypox include *Lemniscomys striatus*, which is geographically widespread and comprised of several subtaxa across sub-Saharan Africa (Nicolas *et al.* 2008). It is also a common source of food in some countries (Assogbadjo *et al.* 2005), raising the possibility of human transmission through direct contact via hunting or the bushmeat trade. Negative publication bias notwithstanding, monkeypox surveillance efforts do not seem to have focused on rodent species, either based on risk or range overlap with human monkeypox cases (Table S4). While there appears to be a gap in the range area of this species across parts of the Congo Basin and Central Africa (Fig. 4), recent occurrence records confirm that *L. striatus* occurs throughout West and Central Africa (https://www.gbif.org/species/2438204), including Equatorial Guinea, Gabon, the Republic of Congo and the Democratic Republic of Congo where monkeypox is emerging (Thomassen *et al.* 2013).

Similarly, focusing surveillance on high prevalence rodent species that overlap geographically with haemorrhagic fever outbreaks (Figs 4 and 5) may confirm suspected rodent reservoirs, identify additional reservoirs or determine surveillance priorities for viral haemorrhagic fevers whose sylvatic reservoirs have yet to be identified (e.g., Bolivian and Brazilian haemorrhagic fevers). In general, when reservoir hosts for zoonoses are unknown, applying a comparative, macroecological trait-based approach may help to distinguish among many potential reservoir species.

Testing model predictions

Predictions about individual reservoir species and their role in human spillover transmission may be directly tested with targeted field data. These data should be collected at frequencies high enough to ensure capture of transient viral shedding often exhibited by wild reservoirs (Peel *et al.* 2019). Beyond validating predictions, if a rodent with low equilibrium prevalence in our model is found to have high prevalence in the field, or if it is otherwise confirmed as a major source of human spillover infection, such findings would suggest that our assumptions about the functional form of transmission, or some other parameters in our model, require updating (for example, longevity or reproductive output). Patterns emerging from empirical trait data thus initiate hypothesis generation through data mining, hypotheses which are then concretely encoded in dynamical models of infectious disease.

Table 2 A rubric for classifying models of disease transmission. Susceptible hosts are denoted by infectious S, hosts by I. In the vector-borne model we distinguish between susceptible/infectious hosts (S_H/ I_H) and susceptible/infectious vectors (S_V/ I_V). Each transmission model can be parameterized using host, parasite and/or environmental features (traits)

Transmission mode	Host traits	Parasite traits	ENV. traits
Direct $\frac{dS}{dt} = b(S+I) - \beta SI - \mu S + \gamma I$ $\frac{dI}{dt} = \beta SI - (\gamma + \mu)I$	• Transmission rate, β • Susceptibility to infection • Probability of successful transmission • Infectious period, $1/\gamma$ • Host longevity, $1/\mu$ • Host reproductive rate, $b(S + I)$ • Host carrying capacity	• Parasite load required for successful transmission • Lethality to host (infection intensity, case-fatality rates) • Ability to evade host immune response	• Temperature • Precipitation • Changes affecting host resource abundance
Vector-borne $\frac{dS_H}{dt} = b_H(S_H + I_H)[S_H + I_H] - kpI_V S_H/(S_H + I_H) - \mu_H S_H + \gamma I_H$ $\frac{dI_H}{dt} = kpI_V S_H/(S_H + I_H) - (\gamma_H + \mu_H)I_H$ $\frac{dS_V}{dt} = b_V(S_V + I_V)[S_V + I_V] - kqS_V I_H/(S_H + I_H) - \mu_V S_V$ $\frac{dI_V}{dt} = kqS_V I_H/(S_H + I_H) - (\gamma_V + \mu_V)I_V$	• Per-individual biting rate per host k • Transmission efficiency from vector to host, and host to vector p, q • Host & vector lifespan, $1/\mu_H$, $1/\mu_v$ • Infectious period of the host and vector, $1/\gamma_H$, $1/\gamma_V$ • Vector & host reproductive rates, $b_H(S_H + I_H)$, $b_V(S_V + I_V)$	• Parasite load required for successful transmission • Lethality to host (infection intensity, case-fatality rates) • Ability to evade host immune response	• Temperature, precipitation changes affecting breeding habitat for the vector
Environmental $\frac{dS}{dt} = b(S+I)[S+I] - \tilde{\beta}SV - \mu S + \gamma I$ $\frac{dI}{dt} = \tilde{\beta}SV - (\gamma + \mu)I$ $\frac{dV}{dt} = \omega I - \delta V$	• Per-individual shedding rate into environment ω • Transmission rate upon contact with free living stage, $\tilde{\beta}$ • Probability that contact with free-living pathogen results in infection	• Durability in the environment, $1/\delta$	• Change to properties of the environment that make it a reservoir

There are limitations to the approach we have explored here. Georeferencing human cases can be difficult and imprecise for diseases that take several days or weeks to present clinically (such as hantavirus; (Astorga *et al*. 2018)). For animal hosts, species with traits predicting higher equilibrium prevalence (or R_0, or outbreak sizes) via transmission models may exhibit seasonal shifts in behaviour or population cycles that reduce contact (and therefore spillover risk) to humans (Davis & Calvet 2005). These caveats notwithstanding, testing model predictions with field data can target the collection of additional data to refine model structure (i.e., our understanding of the transmission process), and ultimately, to improve model predictions of likely host species and transmission dynamics within these species.

Theoretical conditions suggested by these models can be explicitly tested with empirical data. For example, empirical data can assist in identifying when assumptions about the functional form of transmission are too simplistic (e.g., (Borremans *et al*. 2017); or when presumed knowledge of host species is inconsistent with observed dynamics (i.e., when latent (unobserved) variables are at play, or when models do not match reality). Trait-based models can be further refined through parameter estimation methods that select values that best fit observations (e.g., maximum likelihood (Bretó 2018), plausible parameter sets (Drake *et al*. 2015; Kramer *et al*. 2019), Markov chain Monte Carlo (MCMC; Streftaris & Gibson 2004) and others (Beaumont 2010; Ionides *et al*. 2015)). This mining-modelling approach compares possible disease dynamics and equilibria across numerous candidate host species by parameterizing models based on central tendencies of observable traits at the species level. Thus, this approach identifies species that are intrinsically more likely to generate particular disease dynamics. For zoonotic pathogens carried by multiple host species, applying a trait-based approach to transmission modelling may illuminate the network topology of interacting host species (Truitt *et al*. 2019) and may better predict when species diversity leads to dilution or amplification effects (Dobson 2004; Dizney & Dearing 2016; Faust *et al*. 2017).

Generalizing to other ecological processes

Beyond infectious disease dynamics, combining data mining with process modelling offers a comparative dimension to model-guided fieldwork (Restif *et al*. 2012) and may also be useful across a wide range of scenarios where ecological process knowledge is sparse (Hochachka *et al*. 2007; Kelling *et al*. 2009). For example, previous studies suggest that there are predictable differences in the traits of invasive vs. noninvasive plants (Schmidt & Drake 2011; Schmidt *et al*. 2012), threatened vs. non-threatened mammals (Davidson *et al*. 2009) or populations with extreme variations in key behaviours impacting fitness (e.g., (Dingle 2006; Sih *et al*. 2012)). Thus, features that distinguish between organisms may also represent features that drive ecological processes. Updating or building process models to directly incorporate such features constitutes a powerful new means for generating hypotheses to explain and predict ecological dynamics across distinct functional groups in nature.

ACKNOWLEDGEMENTS

This work was supported by a short-term visitor fellowship from NIMBioS (BAH, SO), by the NSF ADVANCE Institutional Transformation program (HRD-1409799 to SO) and by the NSF Ecology and Evolution of Infectious Diseases program (DEB 1717282 to BAH, SO, JPS, JD). The AUTHORS would also like to thank Adrian Castellanos, and Eric Marty for creating Fig. 1.

AUTHORSHIP

BH, SO and JD conceived of the study. BH, SO and JPS collected the data and performed all analyses. BH, SO and JPS wrote the first draft of the manuscript, and all authors contributed substantially to revisions.

DATA ACCESSIBILITY STATEMENT

All data supporting the results are archived in Figshare (https://doi.org/10.25390/caryinstitute.c.4912389).

REFERENCES

Altizer, S., Bartel, R. & Han, B.A. (2011). Animal migration and infectious disease risk. *Science*, 331, 296–302.

Assogbadjo, A.E.,Codjia, C.J.T., Sinsin, B., Ekué, M.R.M. & Mensah, G.A. (2005). Importance of rodents as a human food source in Benin. *Belgian Journal of Zoology*, 135, 11–15.

Astorga, F., Escobar, L.E., Poo-Muñoz, D., Escobar-Dodero, J., Rojas-Hucks, S., Alvarado-Rybak, M. *et al*. (2018). Distributional ecology of Andes hantavirus: a macroecological approach. *Int. J. Health Geogr*., 17, 22.

Barko, V.A. & Feldhamer, G.A. (2002). Cotton Mice (Peromyscus gossypinus) in Southern Illinois: Evidence for Hybridization with White-footed Mice (Peromyscus leucopus). *Am. Midl. Nat*., 147, 109–115.

Beaumont, M.A. (2010). Approximate Bayesian Computation in Evolution and Ecology. *Annu. Rev. Ecol. Evol. Syst*., 41, 379–406.

Berger, S.A. (2005). GIDEON: a comprehensive Web-based resource for geographic medicine. *Int. J. Health Geogr*., 4, 10.

Bicca-Marques, J.C. & Calegaro-Marques, C. (2016). Ranging behavior drives parasite richness: A more parsimonious hypothesis. *Am. J. Primatol*., 78, 923–927.

Bordes, F., Blumstein, D.T. & Morand, S. (2007). Rodent sociality and parasite diversity. *Biol. Lett*., 3, 692–694.

Borremans, B., Reijniers, J., Hens, N. & Leirs, H. (2017). The shape of the contact–density function matters when modelling parasite transmission in fluctuating populations. *Royal Society Open Science*, 4, 171308.

Breiman, L. (2001). Statistical modeling: the two cultures. *Stat. Sci*., 16, 199–231.

Bretó, C. (2018). Modeling and inference for infectious disease dynamics: a likelihood-based approach. *Stat. Sci*., 33, 57–69.

Cable, J.M., Enquist, B.J. & Moses, M.E. (2007). The allometry of host-pathogen interactions. *PLoS ONE*, 2, e1130.

Carslake, D., Bennett, M., Bown, K., Hazel, S., Telfer, S. & Begon, M. (2005). Space-time clustering of cowpox virus infection in wild rodent populations. *J. Anim. Ecol*., 74, 647–655.

Civitello, D.J., Cohen, J., Fatima, H., Halstead, N.T., Liriano, J., McMahon, T.A. *et al*. (2015). Biodiversity inhibits parasites: Broad evidence for the dilution effect. *Proc. Natl. Acad. Sci. USA*, 112, 8667–8671.

Cross, P.C., Lloyd-Smith, J.O., Johnson, P.L.F. & Getz, W.M. (2005). Duelling timescales of host movement and disease recovery determine invasion of disease in structured populations. *Ecol. Lett*., 8, 587–595.

Davidson, A.D., Hamilton, M.J., Boyer, A.G., Brown, J.H. & Ceballos, G. (2009). Multiple ecological pathways to extinction in mammals. *Proc. Natl Acad. Sci. USA*, 106, 10702–10705.

Davis, S. & Calvet, E. (2005). Fluctuating rodent populations and risk to humans from rodent-borne zoonoses. *Vector-Borne and Zoonotic Diseases*, 5, 305–314.

Dingle, H. (2006). Animal migration: is there a common migratory syndrome? *J. Ornithol*., 147, 212–220.

Dizney, L. & Dearing, M.D. (2016). Behavioural differences: a link between biodiversity and pathogen transmission. *Anim. Behav*., 111, 341–347.

Dobson, A. (2004). Population dynamics of pathogens with multiple host species. *Am. Nat*., 164(Suppl 5), S64–78.

Drake, J.M., Kaul, R.B., Alexander, L.W., O'Regan, S.M., Kramer, A.M., Pulliam, J.T. *et al*. (2015). Ebola cases and health system demand in liberia. *PLoS Biol*., 13, e1002056.

Durski, K.N., McCollum, A.M., Nakazawa, Y., Petersen, B.W., Reynolds, M.G., Briand, S. *et al*. (2018). Emergence of Monkeypox - West and Central Africa, 1970–2017. *MMWR Morb. Mortal. Wkly. Rep*., 67, 306–310.

Evans, M.V., Dallas, T.A., Han, B.A., Murdock, C.C. & Drake, J.M. (2017). Data-driven identification of potential Zika virus vectors. *Elife*, 6, 077966.

Ezenwa, V.O. (2004). Host social behavior and parasitic infection: a multifactorial approach. *Behav. Ecol*., 15, 446.

Faust, C.L., Dobson, A.P., Gottdenker, N., Bloomfield, L.S.P., McCallum, H.I., Gillespie, T.R. *et al*. (2017). Null expectations for disease dynamics in shrinking habitat: dilution or amplification? *Biol. Sci*., 372, 20160173

Fine, P.E., Jezek, Z., Grab, B. & Dixon, H. (1988). The transmission potential of monkeypox virus in human populations. *Int. J. Epidemiol*., 17, 643–650.

Gehman, A.-L.-M., Hall, R.J. & Byers, J.E. (2018). Host and parasite thermal ecology jointly determine the effect of climate warming on epidemic dynamics. *Proc. Natl Acad. Sci. USA*, 115, 744–749.

Gutiérrez, J.S., Piersma, T. & Thieltges, D.W. (2019). Micro- and macroparasite species richness in birds: The role of host life history and ecology. *Journal of Animal Ecology, Early View*, 88, 1226–1239.

Han, B.A., Kramer, A.M. & Drake, J.M. (2016a). Global patterns of zoonotic disease in mammals. *Trends Parasitol*., 32, 565–577.

Han, B.A., Majumdar, S., Calmon, F.P., Glicksberg, B.S., Horesh, R., Kumar, A. *et al*. (2019). Confronting data sparsity to identify potential sources of Zika virus spillover infection among primates. *Epidemics*, 27, 59–65.

Han, B.A., Park, A.W., Jolles, A.E. & Altizer, S. (2015). Infectious disease transmission and behavioural allometry in wild mammals. *J. Anim. Ecol*., 84, 637–646.

Han, B.A., Schmidt, J.P., Alexander, L.W., Bowden, S.E., Hayman, D.T.S. & Drake, J.M. (2016b). Undiscovered Bat Hosts of Filoviruses. *PLoS Negl. Trop. Dis*., 10, e0004815.

Harvell, C.D., Kim, K., Burkholder, J.M., Colwell, R.R., Epstein, P.R., Grimes, D.J. *et al*. (1999). Emerging marine diseases-climate links and anthropogenic factors. *Science*, 285, 1505–1510.

Hily, J.M., García, A., Moreno, A., Plaza, M., Wilkinson, M.D., Fereres, A. *et al*. (2014). The relationship between host lifespan and pathogen reservoir potential: an analysis in the system Arabidopsis thaliana–cucumber mosaic virus. *PLoS Pathog*., 10, e1004492.

Hochachka, W.M., Caruana, R., Fink, D., Munson, A., Riedewald, M., Sorokina, D. *et al*. (2007). Data-mining discovery of pattern and process in ecological systems. *J. Wildl. Manage*., 71, 2427.

Huang, Z.Y.X., de Boer, W.F., van Langevelde, F., Olson, V., Blackburn, T.M. & Prins, H.H.T. (2013). Species' life-history traits explain interspecific variation in reservoir competence: a possible mechanism underlying the dilution effect. *PLoS ONE*, 8, e54341.

Hutin, Y.J., Williams, R.J., Malfait, P., Pebody, R., Loparev, V.N., Ropp, S.L. *et al*. (2001). Outbreak of human monkeypox, Democratic Republic of Congo, 1996 to 1997. *Emerg. Infect. Dis*., 7, 434–438.

Ionides, E.L., Nguyen, D., Atchadé, Y., Stoev, S. & King, A.A. (2015). Inference for dynamic and latent variable models via iterated, perturbed Bayes maps. *Proc. Natl Acad. Sci. USA*, 112, 719–724.

Johnson, P.T.J., Rohr, J.R., Hoverman, J.T., Kellermanns, E., Bowerman, J. & Lunde, K.B. (2012). Living fast and dying of infection: host life history drives interspecific variation in infection and disease risk. *Ecol. Lett*., 15, 235–242.

Jones, K.E., Bielby, J., Cardillo, M., Fritz, S.A. & O'Dell, J. (2009). PanTHERIA: a species-level database of life history, ecology, and geography of extant and recently extinct mammals. *Ecology*, 90, 2648–2648

Keesing, F. & Ostfeld, R.S. (2015). Ecology. Is biodiversity good for your health? *Science*, 349, 235–236.

Kelling, S., Hochachka, W.M., Fink, D., Riedewald, M., Caruana, R., Ballard, G. *et al*. (2009). Data-intensive Science: A New Paradigm for Biodiversity Studies. *Bioscience*, 59, 613–620

Khodakevich, L., Jezek, Z. & Messinger, D. (1988). Monkeypox virus: ecology and public health significance. *Bull. World Health Organ*., 66, 747–752.

Kramer, A.M., Teitelbaum, C.S., Griffin, A. & Drake, J.M. (2019). Multiscale model of regional population decline in little brown bats due to white-nose syndrome. *Ecol. Evol*., 9, 8639–8651.

Kutz, S.J., Hoberg, E.P., Polley, L. & Jenkins, E.J. (2005). Global warming is changing the dynamics of Arctic host–parasite systems. *Proceedings of the Royal Society B: Biological Sciences*, 272, 2571–2576.

LaDeau, S.L., Han, B.A., Rosi-Marshall, E.J. & Weathers, K.C. (2017). The next decade of big data in ecosystem science. *Ecosystems*, 20, 274–283.

Lafferty, K.D. & Mordecai, E.A. (2016). The rise and fall of infectious disease in a warmer world. *F1000Research*, 5, 2040

Li, F.C.K., Choi, B.C.K., Sly, T. & Pak, A.W.P. (2008). Finding the real case-fatality rate of H5N1 avian influenza. *J. Epidemiol. Community Health*, 62, 555–559.

Lochmiller, R.L. & Deerenberg, C. (2000). Trade-offs in evolutionary immunology: just what is the cost of immunity? *Oikos*, 88, 87–98.

Luong, L.T., Vigliotti, B.A., Campbell, S., Comer, J.A., Mills, J.N. & Hudson, P.J. (2011). Dynamics of hantavirus infection in Peromyscus leucopus of central Pennsylvania. *Vector Borne Zoonotic Dis*., 11, 1459–1464.

May, R.M. & Anderson, R.M. (1979). Population biology of infectious diseases. *Nature*, 280, 455–461.

Montiglio, P.-O., Dammhahn, M., Dubuc Messier, G. & Réale, D. (2018). The pace-of-life syndrome revisited: the role of ecological conditions and natural history on the slow-fast continuum. *Behav. Ecol. Sociobiol*., 72, 116.

Nicolas, V., Mboumba, J.-F., Verheyen, E., Denys, C., Lecompte, E., Olayemi, A. *et al*. (2008). Phylogeographic structure and regional history of Lemniscomys striatus (Rodentia: Muridae) in tropical Africa. *J. Biogeogr*., 35, 2074–2089.

Ostfeld, R.S., Levi, T., Jolles, A.E., Martin, L.B., Hosseini, P.R. & Keesing, F. (2014). Life history and demographic drivers of reservoir competence for three tick-borne zoonotic pathogens. *PLoS ONE*, 9, e107387.

Pascual, M. & Bouma, M.J. (2009). Do rising temperatures matter. *Ecology*, 90, 906–912.

Peel, A.J., Wells, K., Giles, J., Boyd, V., Burroughs, A., Edson, D. *et al*. (2019). Synchronous shedding of multiple bat paramyxoviruses coincides with peak periods of Hendra virus spillover. *Emerg. Microbes Infect*., 8, 1314–1323.

Peters, D.P.C., Havstad, K.M., Cushing, J., Tweedie, C., Fuentes, O. & Villanueva-Rosales, N. (2014). Harnessing the power of big data: infusing the scientific method with machine learning to transform ecology. *Ecosphere*, 5, art67.

Peters, R.H. (1986). *The Ecological Implications of Body Size*. Cambridge Univ Press, New York.

Plowright, R.K., Becker, D.J., Crowley, D.E., Washburne, A.D., Huang, T., Nameer, P.O. *et al.* (2019). Prioritizing surveillance of Nipah virus in India. *PLoS Negl. Trop. Dis.*, 13, e0007393.

Plowright, R.K., Parrish, C.R., McCallum, H., Hudson, P.J., Ko, A.I., Graham, A.L. *et al.* (2017). Pathways to zoonotic spillover. *Nat. Rev. Microbiol.*, 15, 502–510

Polverino, G., Santostefano, F., Díaz-Gil, C. & Mehner, T. (2018). Ecological conditions drive pace-of-life syndromes by shaping relationships between life history, physiology and behaviour in two populations of Eastern mosquitofish. *Sci. Rep.*, 8, 14673.

Previtali, M.A., Ostfeld, R.S., Keesing, F., Jolles, A.E., Hanselmann, R. & Martin, L.B. (2012). Relationship between pace of life and immune responses in wild rodents. *Oikos*, 121, 1483–1492.

Restif, O., Hayman, D.T.S., Pulliam, J.R.C., Plowright, R.K., George, D.B., Luis, A.D. *et al.* (2012). Model-guided fieldwork: practical guidelines for multidisciplinary research on wildlife ecological and epidemiological dynamics. *Ecol. Lett.*, 15(10), 1083–1094

Ryan, S.J., Carlson, C.J., Mordecai, E.A. & Johnson, L.R. (2019). Global expansion and redistribution of Aedes-borne virus transmission risk with climate change. *PLoS Negl. Trop. Dis.*, 13, e0007213.

Schmidt, J.P. & Drake, J.M. (2011). Why are some plant genera more invasive than others? *PLoS ONE*, 6, e18654.

Schmidt, J.P., Springborn, M. & Drake, J.M. (2012). Bioeconomic forecasting of invasive species by ecological syndrome. *Ecosphere*, 3, art46.

Sih, A., Cote, J., Evans, M., Fogarty, S. & Pruitt, J. (2012). Ecological implications of behavioural syndromes. *Ecol. Lett.*, 15, 278–289.

Stekhoven, D.J. & Bühlmann, P. (2012). MissForest–non-parametric missing value imputation for mixed-type data. *Bioinformatics*, 28, 112–118.

Streftaris, G. & Gibson, G.J. (2004). Bayesian analysis of experimental epidemics of foot-and-mouth disease. *Proc. Biol. Sci.*, 271, 1111–1117.

Thomassen, H.A., Fuller, T., Asefi-Najafabady, S., Shiplacoff, J.A.G., Mulembakani, P.M., Blumberg, S. *et al.* (2013). Pathogen-host associations and predicted range shifts of human monkeypox in response to climate change in central Africa. *PLoS ONE*, 8, e66071.

Tian, H. & Stenseth, N.C. (2019). The ecological dynamics of hantavirus diseases: From environmental variability to disease prevention largely based on data from China. *PLoS Negl. Trop. Dis.*, 13, e0006901.

Truitt, L., McArt, S., Vaughn, A. & Ellner, S.P. (2019). Trait-based modeling of multi-host pathogen transmission: Plant-pollinator networks. *Am. Nat.*, 193, E149–E167.

Walker, J.W., Han, B.A., Ott, I.M. & Drake, J.M. (2018). Transmissibility of emerging viral zoonoses. *PLoS ONE*, 13, e0206926.

Yang, L.H. & Han, B.A. (2018). Data-driven predictions and novel hypotheses about zoonotic tick vectors from the genus Ixodes. *BMC Ecol.*, 18, 7.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Editor, Amanda Bates
Manuscript received 6 December 2019
First decision made 21 January 2020
Manuscript accepted 27 March 2020