Host community structure can shape pathogen outbreak dynamics through a phylogenetic dilution effect.

Marjolein E.M. Toorians[1,*], Isabel M. Smallegange[2], and T. Jonathan Davies[1,3,4]

[1]Department of Botany, Biodiversity Research Centre, University of British Columbia, 2212 Main Mall, Vancouver, BC V6T 1Z4, Canada

[2]School of Natural Environmental Sciences, Newcastle University, Newcastle upon Tyne, NE1 7RU, UK

[3]African Centre for DNA Barcoding, University of Johannesburg, Johannesburg 2092, South Africa

[4]Department Forest & Conservation Sciences, University of British Columbia, 2212 Main Mall, Vancouver, BC V6T 1Z4, Canada

[*]Corresponding author: marjolein.toorians@botany.ubc.ca

April 2, 2024

Acknowledgements: We would like to thank S. Otto for reviewing an earlier version of this manuscript and her advice on modelling approaches. We also thank Ailene MacPherson for her help with the modelling. This research was supported by an NSERC Discovery grant awarded to TJD.

Author contributions: Manuscripts was conceptualized by T.J. Davies and M.E.M. Toorians. MEMT designed and analysed the model with feedback from I.M. Smallegange and TJD. MEMT wrote the paper with TJD. IMS contributed to manuscript edits.

Data accessibility: Data has been generated in simulations in R scripts that have been deposited on https://github.com/Marjoleintoorians/PhyloSIModel

Abstract

Biodiversity loss and anthropogenic modifications to species communities are impacting the frequency and magnitude of disease emergence events. These changes may be related through mechanisms in which biodiversity either increases (amplifies) or decreases (dilutes) disease prevalence. Biodiversity effects can be direct, when contacts among competent hosts are substituted by contacts with sink hosts, or indirect through regulation of host abundances. Whether a disease is diluted or amplified depends on the competence of the host species in the community. Here, we introduce a multi-host compartmental disease model in which we assume that host competence is determined by evolutionary relatedness. By simulating communities of varying phylogenetic structure, and estimating community disease outbreak potential (R_0), we show how differences in host phylogenetic relatedness can switch host communities from diluting to amplifying a disease, even when species richness is unchanged. We additionally show that phylogenetic dilution can occur simultaneously with amplification through species richness, highlighting the multidimensionality of the disease-diversity relationship. We illustrate our model using empirical data describing the relationship between phylogenetic distances separating hosts and their likelihood of disease sharing. Our study demonstrates how a phylogenetic dilution effect can emerge when we allow transmission to depend on host evolutionary histories.

Key words: Amplification, Dilution, Disease-ecology, Epidemiology, Multi-host model, Pathogens, Phylogeny

Word count: 5233

Number of references: 93

Number of figures: 4

Number of tables: 4

1 Introduction

Over the past century, anthropogenic pressures on the environment have accelerated (Steffen et al., 2015), leaving many species threatened with extinction (IPBES, 2020; Barrett et al., 2018; IUCN, 2021). Coinciding with increasing in biodiversity loss has been an increase in the frequency of disease outbreaks (Daszak et al., 2001; UN, 2020) and it has been suggested that these two trends may be linked (Halliday and Rohr, 2019; Rohr et al., 2020). Similar patterns are seen for plants and animals (Daszak et al., 2000; Woolhouse et al., 2005; Jones et al., 2008), suggesting a common process. However, there remains debate about how shifts in the composition and richness of host communities may translate to changes in disease prevalence and outbreak potential (LoGiudice et al., 2003; Joseph et al., 2013). While much attention has been given to the relationship between host species diversity and disease (e.g. Halliday and Rohr, 2019; Rohr et al., 2020; Keesing et al., 2006, 2010; Han et al., 2020), predicting how changes in host species composition may alter disease prevalence requires a mechanistic understanding of multi-host disease dynamics (Woolhouse et al., 2001; Dobson, 2004; Luis et al., 2018).

Theoretical models suggest that an increase in host species richness can either decrease disease prevalence (dilution effect: Table 1) or increase disease prevalence (amplification effect: Table 1) (Keesing et al., 2006, 2010; Ostfeld and Keesing, 2012). Both dilution and amplification can occur via changes in transmission, either directly through modifying encounter rates between competent and less-competent hosts or indirectly through changing densities of competent hosts (Rudolf and Antonovics, 2005; Ostfeld and Keesing, 2000). In practice, the distinction between the processes of amplification and dilution is not straightforward. Some diseases appear to be diluted by diversity in some instances, but amplified by diversity in other instances, depending on the host community species composition, and host shedding rates or relative abundances (Barasona et al., 2019; Streicker et al., 2013; Stewart Merrill et al., 2022). It is also possible for both dilution and amplification to occur simultaneously in the same system, for example, depending on the density and composition of the host community, as has been suggested in rodents infected by hantavirus (Luis et al., 2018) and for RNA viruses in honey and bumble bees, where species abundance and richness separately had amplifying and diluting effects, respectively (Cohen et al., 2022). Keesing and Ostfeld (2021), who first coined the 'dilution effect', acknowledge that the general effect of biodiversity on disease prevalence remains controversial.

Ultimately, it is the composition and relative competence of hosts within a community that drive dilution/amplification dynamics. Species that are able to maintain endemic prevalence—reservoir hosts (Table 1, Haydon et al., 2002)—have

high competence, and can dominate transmission in a multi-host community. Less competent hosts, however, may function as dead-end hosts if they either do not maintain the disease, or do not contribute to onward transmission, thereby reducing disease spread. Data on host competence would allow us to make more accurate predictions about how changes in host diversity would impact disease prevalence; however, such data are frequently lacking, especially for low-competent hosts and hosts with no prior exposure. One approach to bridge this data gap is to approximate host competence from life history data and species' physiology, such as body size or other life-history traits (Dobson, 2004; Joseph et al., 2013; Cohen et al., 2022). Here, we suggest that if relevant traits are evolutionarily conserved, species phylogeny might provide another useful and more general proxy.

There is a growing weight of evidence showing that diseases are more frequently shared between closely related hosts (Gilbert and Webb, 2007; Davies and Pedersen, 2008; Streicker et al., 2010; Parker et al., 2015; Farrell and Davies, 2019a). This phylogenetic structure in the distribution of pathogens among hosts likely reflects phylogenetic conservatism in evolved defense mechanisms (Streicker et al., 2010; Mollentze and Streicker, 2020). We might therefore expect the evolutionary distance between hosts to be correlated with both the probability of sharing a disease and within-host pathogen load following infection. For example, in grassland communities, disease prevalence was higher in species with many close relatives in the community (Parker et al., 2015), whereas hosts that were phylogenetically isolated (Table 1) in their communities obtained a *rare-species advantage* and were less likely to share pathogens with co-occurring species. Phylogenetic isolation—Evolutionary Distinctness (ED, Table 1)—from co-occurring hosts might, therefore, be an indicator of low competence for diseases most prevalent within a given host community (although they may be highly competent hosts for other diseases). We suggest that the phylogenetic composition of hosts may thus be an additional component in the establishment probability and disease dynamics of a pathogen in the community.

Here, using simulations and a simple compartmental (Susceptible-Infected) disease model, we explore the importance of the phylogenetic structure of host communities on outbreak potential (R_0) for a multi-host disease, assuming density dependent (DD) dynamics, where the contact rate is constant. In Appendix Section 6 we show results for a frequency-dependent (FD) model. For simplicity, our model assumes no recovery, such as is the case for *Mycobacterium bovis* infection, the causative agent of bovine Tuberculosis (bTB) (Jolles et al., 2005; Ezenwa et al., 2010), but it could easily be extended. Scaling transmission probability in proportion to the inverse of the phylogenetic distance between hosts, such that transmission is more likely between closely related hosts, we consider two alternative

scenarios: 1) transmission is proportional to the evolutionary distance between infected and susceptible hosts; and 2) transmission is proportional to the distance between the receiving (secondary) and primary (co-evolved) host, irrespective of the identity of the donating host (Figure 1). We provide examples of our model predictions parameterised using empirical data on plant pests (Gilbert et al., 2012). We suggest that our model may help in our understanding of the conflicting evidence for dilution and amplification effects of biodiversity across different host communities and disease systems.

2 Methods

The outbreak potential of a disease may be quantified by its reproductive number (R_0), which represents the average number of new infections that can arise from a single infected individual over the course of their lifetime, considering a population of susceptible individuals. When this value is greater than one, an epidemic can occur. The R_0 can be calculated from the ordinary differential equations describing the susceptible (S) and infected (I) individuals (Kermack and McKendrick, 1927; May and Anderson, 1984; Dobson, 2004):

$$\frac{dS_i}{dt} = b_i S_i - \sum_{j=1}^{n} \beta_{ij}(N_j) I_j p_{ij} - \nu_i S_i \tag{1a}$$

$$\frac{dI_i}{dt} = \sum_{j=1}^{n} \beta_{ij}(N_j) I_j p_{ij} - (\nu_i + \mu_i) I_i \tag{1b}$$

where $\beta(N_j)$ describes the transition from S to I compartments per time unit, ν_i and μ_i are the species-specific disease-induced and natural mortality rates, respectively, b_i is the per-capita host birth rate, and p_{ij} represents the mode of transmission, such that $p_{ij} = \frac{S_i}{N}$ for frequency dependence, or $p_{ij} = \frac{S_i}{N_j}$ for density-dependence. We focus primarily on density-dependent transmission, as this is the most commonly assumed mode of transmission for wildlife diseases White et al. (2017), but provide matching results for frequency dependent dynamics in the supplement. Details of the model and its derivation can be found in Appendix Section 1.

Following Keeling and Rohani (2011), we can decompose the transmission rate parameter, β as:

$$\beta_{ij}(N_j) = \kappa_{ij}(N_j) \cdot c_{ij} \tag{2}$$

In this equation, the transmission rate, β, is a function of the probability of transmission given a contact, c_{ij}, multiplied by the contact rate, κ. We can define κ as a function of the individuals in the community, N_j, allowing this model to describe both frequency-dependent and density-dependent disease transmission, such that κ is either a constant or varies with N_j, respectively (see Appendix Section 1). This dynamic for κ is captured in the Kronecker function (see Appendix section 4).

In a multi-host community, we can calculate $R_{0,\text{community}}$ by using the Next-Generation-Matrix (Diekmann and Kretzschmar, 1991; Diekmann et al., 2010). A detailed derivation can be found in Appendix Section 1. In a 3-host system, matrix **K** can be represented as:

$$K = \begin{bmatrix} \frac{\beta_{11}S_1}{\mu_1+\nu_1} & \frac{\beta_{12}S_1}{\mu_2+\nu_2} & \frac{\beta_{13}S_1}{\mu_3+\nu_3} \\ \frac{\beta_{21}S_2}{\mu_1+\nu_1} & \frac{\beta_{22}S_2}{\mu_2+\nu_2} & \frac{\beta_{23}S_2}{\mu_3+\nu_3} \\ \frac{\beta_{31}S_3}{\mu_1+\nu_1} & \frac{\beta_{32}S_3}{\mu_2+\nu_2} & \frac{\beta_{33}S_3}{\mu_3+\nu_3} \end{bmatrix} \tag{3}$$

The maximum real eigenvalue of matrix **K** is the *community* R_0 and describes the reproductive success of a generalist pathogen in the fully susceptible host community. On the diagonal of matrix **K** we can find the individual R_0 of species i in isolation, $R_{0,i}$. When $R_{0,i} > 1$, host species i is a reservoir host; when $R_{0,i} < 1$, host species i cannot sustain infection on its own, resulting in obligate multi-host parasitism (Fenton et al., 2015). The sum of a column of matrix **K** shows the total received infection of the nth species in the community (including intraspecific transmission). The sum of a row shows the total contribution of that species to the community infected class.

We can easily rescale the interspecific values for c_{ij} in the β transmission rate term; here, we assume that probability of successful infection increases with decreasing evolutionary distance (Lively, 2010; Gilbert et al., 2012; Parker et al., 2015; Farrell and Davies, 2019a). Now, the probability of successful transmission from host j to i is determined by the evolutionary distance separating the two species (Pairwise Phylogenetic distance, PPd, see Table 1). While we model the effect of phylogenetic distance on the probability of transmission given a contact, we might expect contact rates to also co-vary with phylogeny. However, the weighting by phylogeny can be equally placed on either κ or c, and thus a single weighting term can capture phylogenetic structure in transmission probability, contact rate or both. We assume all species have equal abundances, thus allowing for encounter reduction with the addition of a species to the host community, putatively the main mechanism behind dilution (Keesing et al., 2006). We do not, therefore, explore potential effects of host evenness or other diversity indices combining richness with abundances,

but it would be possible to further expand our model to consider more complex host population dynamics if there was evidence to support them in a given community. We use simulations to explore two alternative scenarios describing primary-secondary host dynamics.

Scenario 1: No primary host

In this scenario, the probability of successful transmission from species j after contact with species i, is modelled as:

$$c_{ij} = \psi \cdot \frac{1}{1 + \alpha PPd_{ij}} \tag{4}$$

where PPd_{ij} is the pairwise phylogenetic distance between hosts i and j (typically measured in millions of years), α is a scalar for PPd, and ψ scales the probability of infection. The probability of establishing an infection in a novel host simply depends on the distance between the transmitting, j, and receiving, i, host. For our simulations, we assume $\alpha = \frac{1}{2} \cdot \alpha_0$ as PPd represents 2 * the time since divergence between hosts; however, alternative evolutionary models can be assumed that allow for non-linear relationships, or α can be estimated directly from empirical observations, as we describe below. This scenario might apply to diseases such as Rabies virus (*Rabies lyssavirus*), a multi-host pathogen affecting a large range of mammal species, where isolation within a species and high mutation rates cause rapid differentiation and local adaptation of the virus to its host (Bourhy et al., 1999; Streicker et al., 2010; Mollentze et al., 2014).

Scenario 2: Single primary host

Here, we scale the interspecific probability of successful transmission by the distance between the secondary, recipient host (i) and the primary host, regardless of the identity of the proximate transmitting host. In this scenario, we assume the pathogen is best adapted to the primary host P, and so the probability of disease transmission will decreases with the evolutionary distance between the recipient host and the primary host, such that:

$$c_{ij} = \psi \cdot \frac{1}{1 + \alpha PPd_{iP}} \tag{5}$$

In a three-species system, there are two unique phylogenetic placements of the primary host, nested within the ingroup sister pair, or as outgroup to the sister pair, we refer to these alternative placements as ingroup and outgroup, respectively. Both scenarios are illustrated in Figure 1.

One well-described empirical example of a disease systems with primary-secondary host dynamics is bovine tuberculosis (bTB). In sub-Saharan Africa, bTB is thought to be endemic in African Buffalo (*Syncerus caffer*), which function as a reservoir, and primary host (Bengis and Erasmus, 1988; Jolles et al., 2005; De Garine-Wichatitsky et al., 2013; Ezenwa and Jolles, 2015). However, bTB frequently spills over to secondary hosts, including other ungulates, where it may establish persistent infections, and, more rarely, to phylogenetically distant species, which act as dead-end hosts with no forward transmission (Michel et al., 2006).

Simulating communities

We used simulations to describe the modelled relationship between host community phylogenetic structure and disease outbreak potential (R_0). We first simulate host phylogenetic topologies under a pure-birth Yule process, using the R-package *phytools* (Revell, 2012), with birth rate, b, drawn from an exponential distribution to simulate *waiting times* (the branching times between speciation events) (See Appendix Section 2 for more information). Communities simulated with low b have an increased distance between branching events, and are thus composed of more distantly related species than communities simulated with high b (Appendix Figure 2).

To study dilution and amplification effects, we explored two alternative effects on host densities, as described by (Ostfeld and Keesing, 2012): (1) Substitutive, adding a species to a community reduces the abundance of other species, such that the total number of individuals in the community, N_T, is constant (where $N_T = S_T + I_T = 100$, and T represents the total n species); (2) Additive, the total number of of individuals, N, increases when a novel species is introduced. Each species has the same number of individuals ($N_i = N_j = ... = N_n$ =100). In all simulations $I \approx 0$ and $N \approx S$ for the calculation of R_0.

We generated 100 host communities for different combinations of richness (n=3,4,5,7,10) and speciation rates ($0 < b < 1$). For each simulated community phylogeny, we calculated i) Phylogenetic Diversity (PD) to capture host tree size; ii) Mean Pairwise Distance between hosts (MPD) as a measure of phylogenetic community structure; and iii) Tree *Imbalance*, Ic, to capture the effect of tree-topology, (Table 1). To explore primary-secondary host dynamics in Scenario 2, for communities with > 3 hosts, we additionally recorded primary host Evolutionary Distinctiveness (ED), a metric describing the phylogenetic isolation of a tip on the phylogeny (Isaac et al., 2007). In simulations, the phylogenetic placement of the primary host (Scenario 2) was assigned randomly. We then calculated the community R_0, with a natural and disease-induced mortality of $\nu = 0.1$ and $\mu = 0.2$, respectively and assuming $\psi = 0.01$ and

$\alpha = 1$ to return realistic values of R_0 ($0 < R_0 < 3$ for option 1 and $0 < R_0 < 15$ for option 2) (Roche et al., 2012). For results reported in the main text, we set contact rates, $\kappa_{ij} = 1$, such that interspecific and intraspecific contacts are assumed equally likely. We explore alternative contact scenarios in Appendix section 4. We use regression models to quantify how different aspects of host phylogeny mediate disease outbreak potential. Both MPD and ED were log transformed, and all variables were scaled using a z-transform before model fitting. Colinearity among predictors was assessed using Variance Inflation Factors (VIF) (Zuur et al., 2010).

Finally, we parameterised our model using empirically derived transformations of the pairwise phylogenetic distances between hosts for viruses and bacteria from Gilbert et al. (2012). This study estimated host susceptibility in relation to the phylogenetic distance separating donating and recipient hosts for various pathogen groups using the Global Pest and Disease Database from the US Department of Agriculture Animal and Plant Health Inspection Services, a database containing 137,000 pest-host records for 18,500 plant species worldwide. Gilbert et al. (2012) assume the following functional relationship: $logit(c_{ij}) = b_0 + b_1 * log_{10}(PPd_{ij} + 1)$. Because the steepness of the slope describing the relationship between susceptibility and phylogenetic distance differed among pathogen types, we fit two separate models: b_0 = 8.44 and b_1 = -5.2, for viruses, and b_0 = 3.26 and b_1 = -2.97 for bacteria.

Glossery			
Term	**Abbr**	**Definition**	**Source**
Reservoir	-	A single or group of host species able to maintain the prevalence of a disease and persistence of disease within population	Haydon et al. (2002)
Primary host	-	A single host species with which the pathogen has undergone substantial co-evolution with a specific pathogen	
Disease competence	-	Individual or population's susceptibility to and ability to effectively transmit onward a disease	Ostfeld and Keesing (2000)
Dilution effect	-	Reduction in disease prevalence with increasing biodiversity	Ostfeld and Keesing (2000)
Amplification effect	-	Increased disease prevalence with increasing biodiversity	Ostfeld and Keesing (2000)
Phylogenetic diversity	PD	Summed evolutionary history of species set	Faith (1992)
Pairwise Phylogenetic distance	PPd	Evolutionary history separating species	Cadotte and Davies (2016)
Evolutionary Distinctiveness	ED	Measure of phylogenetic isolation of a species	Isaac et al. (2007)
Basic reproduction number	R_0	Expected number of infections arising from a single infected individual in a naive population	Kermack & McKendrick (1927)
Metrics			
Edge length	θ	Length from tip to node (in MY)	-
Pairwise Phylogenetic Distance	PPd_{ij}	Branch length $i \rightarrow j$	Cadotte and Davies (2016)
Phylogenetic Diversity	PD	$\sum_{e \subset z(T)} \theta_e$	Faith (1992)
Mean Phylogenetic Distance	MPD	$\frac{1}{D} \sum_i \sum_{i \neq j} PPd_{ij}$	Webb et al. (2002)
Imbalance (Colless's I)	Ic	$\sum_{u \subset V_{in}(T)} \lvert n_{u_a} - n_{u_b} \rvert$	Colless (1982)
Evolutionary Distinctiveness	ED	$ED(T,i) = \sum_{e \subset q(T,i,r)} (\theta_e \frac{1}{n_e})$	Isaac et al. (2007)

Table 1: Glossary of terms and biodiversity metrics. For phylogenetic metrics, θ_i is the branch length from species i to the root of the tree and n is the number of species in the tree. $e \subset z(T)$ denotes the edges, e, in a set z for tree T, D is the total number of species combinations, and the set $q(T,i,r)$ connecting species i to the root r, and n_e is the number of species that descend from edge e. $V_{in}(T)$ is the set of all inner vertices of the tree, and n_{u_x} are the leaves in the subtree x.

3 Results

We used simulations to explore the relationship between host community phylogenetic structure and disease outbreak potential assuming transmission success is a function of the phylogenetic distance between hosts. We show that both the phylogenetic structure of the host community and the phylogenetic placement of the primary host can have large influence on community R_0.

3.1 Option 1: Substitutive

Adding a species to a community reduces the abundance of other species, such that the total number of individuals in the community remains constant.

Scenario 1: No primary host

In the absence of a single primary host, the establishment probability of the disease (c) simply depends on the phylogenetic distance between the receiving and donating hosts, as described in Eqn (4) and as illustrated in Figure 1A for the 3-taxon case. We modelled R_0 as a function of MPD, species richness (n) and Imbalance (Ic), capturing the key axes of variation in our PCAs (Appendix Figures 3 and 4). There was strong covariation between MPD and PD (VIFs of 111 and 140, respectively), we report multivariate models including MPD here as it is independent of species richness, n; however, models exchanging MPD with PD were qualitatively similar.

In the single-predictor models (Table 2), PD explained the most variation in R_0; more phylogenetically diverse communities generally had lower disease outbreak potential. MPD and host richness, n, were also strong predictors of R_0 (Table 2), with communities composed of more, but less closely related species having a lower outbreak potential (Figure 2A). While Imbalance (Ic) was positively correlated with R_0, more balanced trees had a higher outbreak potential, the relationship was weak (Figure 2C). In our multivariate model, MPD was the strongest predictor of R_0 (Table 2), but both n and Ic explained some additional variation.

Our full model was able to explain a high proportion of the variation ($R^2 = 0.91$) in R_0 across simulations, indicating that these relatively common phylogenetic indices are able to reliably capture the underlying disease transmission probabilities in our simulations.

Scenario 2: Single primary host

When we specify a primary host, the establishment probability of the disease (c) depends on the phylogenetic distance between the receiving and the primary host. Here, we quantify the phylogenetic isolation of the primary host using a metric of Evolutionary Distinctiveness (ED) with the fair proportions method (Isaac et al., 2007).

In the 3-taxon case (Figure 1), when the primary host is nested within the ingroup (Scenario 2a), the distance between the primary host (P) and the most proximate secondary host (S2), is the main driver of disease dynamics (Figures 1B and 1C), whereas the more distant secondary host (S1) contributes little to the outbreak potential because its further distance from the primary host reduces transmission probability. As the phylogenetic distance between the primary and proximate secondary host decreases ($PPd_{P,S2} \rightarrow 0$), the model converges to a two-host system (P+S2 and S1). Similarly, when the primary host is sister to both secondary hosts (Scenario 2b), both secondary hosts are

equidistant from the primary host (Figure 1D), and thus epidemiologically indistinguishable (here the system converges to a two-host system with P and S2+S1). However, as the distance between the primary host and secondary hosts ($PPd_{P,S}$) increases, the system effectively converges on a one-host system, as successful transmission from the donor (P) to either secondary hosts becomes increasingly improbable.

In our single predictor regressions, effect sizes and model R^2 are similar to those described for Scenario 1; however, ED is also strongly correlated with R_0 ($R^2 = 0.70$; Table 2), with R_0 decreasing with increasing host ED (Figure 2E). Fitting the matching multiple regression model from Scenario 1 to data from Scenario 2, returns a notably worse fit ($R^2 = 0.85$). However, when we substitute MPD for ED, we increased our explained variance to 93%, the highest across either scenario, and ED has the largest standardized effect size (Table 2). The model assuming FD transmission returned similar R^2 (Appendix Section 6).

3.2 Option 2: Additive

Adding a species to a community does not effect the abundance of other species, such that the total number of individuals in the community increases.

In contrast to the models above assuming replacement, we here see a classic amplification effect (Figure 3) with species richness, n, positively correlated with R_0. However, MPD and ED still show a diluting effect (Figure 3A, B and E), with absolute effect sizes comparable across all three variables in the multiple regression models. In the single predictor models, it is notable that the effect size of MPD is double that for PD in both scenarios, and PD alone explains little of the variation in R_0 (Scenario 1: R^2= 0.07 and , Scenario 2: R^2=0.08). We suggest that the low explanatory power of PD is because it includes information on both PPd between species as well as species richness, which in this model option have an opposing relationship (classic amplification through n, phylogenetic dilution through MPD). The explanatory power of imbalance is larger than the estimated effect in option 1, and the sign of the relationship is switched—more balanced trees had a higher outbreak potential in a model of classic amplification. However, the contribution of tree balance to R_0 is marginal in the multivariate model (Table 3).

In scenario 2, we again find that the multivariate model with ED is favoured over that with MPD (R^2=0.72 and 0.82 for the models with MPD and ED, respectively). Results for simulations assuming FD dynamics are the same for both additive and substitutive options (see Appendix section 6). Overall explanatory power for models with

additive abundance (option 2) was lower than for models with substitutive abundance (option 1). Regression results are summarized in Table 3.

	Scenario 1			**Scenario 2**		
Variable	Estimate effect on R_0	R^2	AIC	Estimate effect on R_0	R^2	AIC
log(PD)	-0.44	0.89	-3891	-0.45	0.84	-1206
log(MPD)	-0.46	0.77	24	-0.47	0.73	1452
Ic	-0.05	0.22	6108	-0.05	0.21	6765
n	-0.13	0.40	4749	-0.13	0.37	5647
log(ED)	-	-	-	-0.47	0.70	1957
Multivar (a)	1.49-0.38log(MPD)-0.23n+0.04Ic	0.91	-4634	1.46-0.39log(MPD)-0.22n+0.03Ic	0.85	-1539
Multivar (b)	-	-	-	1.46-0.41log(ED)-0.30n+0.05Ic	0.93	-5428

Table 2: Models describing the variation in R_0 with various phylogenetic metrics for simulations with model Option 1 (replacing abundance) and $\nu = 0.1$, $\mu = 0.2$, $\psi = 0.01$, and host species richness of 3, 4, 5, 7 and 10 species. Table shows the standardized effect sizes of each metric. All parameter estimations were significant.

	Scenario 1			**Scenario 2**		
Variable	Estimate effect on R_0	R^2	AIC	Estimate effect on R_0	R^2	AIC
log(PD)	-0.64	0.07	23509	-0.72	0.08	23852
log(MPD)	-1.28	0.22	22651	-1.34	0.22	23017
Ic	0.27	0.28	22207	0.23	0.21	23100
n	0.61	0.33	21840	0.57	0.26	22757
log(ED)	-	-	-	-1.87	0.43	21473
Multivar (a)	7.84-1.90log(MPD)+1.95n+0.22Ic	0.81	15561	7.65-1.96log(MPD)+1.98n+0.08Ic	0.72	17893
Multivar (b)	-	-	-	7.65-2.08log(ED)+1.59n+0.19Ic	0.82	15579

Table 3: Models describing the variation in R_0 with various phylogenetic metrics for simulations with model Option 2 (additive abundance) and $\nu = 0.1$, $\mu = 0.2$, $\psi = 0.01$, and host species richness of 3, 4, 5, 7 and 10 species. Table shows the standardized effects size of each metric. All parameter estimations were significant.

3.3 Application to Viruses and Bacteria of plants

We used the phylogenetic branch length transformations from Gilbert *et al.* (2012) for bacteria and viruses to weight probability of transmission success. We explored Scenario 1 only, as we did not have information on primary hosts. Assuming these empirically derived relationships, we show that host phylogeny is less important in determining community R_0 for viruses ($R^2 = 0.49$ for viruses vs $R^2 = 0.87$ for bacteria; Table 4), which demonstrate a threshold-like relationship between MPD and R_0 (Figure 4). Nonetheless, the multivariate models generally had high explanatory power across pathogen types (results for other pathogens are shown in the Appendix Table 2), with MPD the most important predictor variable.

4 Discussion

We are witnessing a major rearrangement of biological diversity. Many species are threatened with extinction, while others are expanding their global distributions, sometimes invasively (IPBES, 2020; Barrett et al., 2018; IUCN, 2021). There is an urgent need to understand how these biotic changes might restructure host communities through altering environmental carrying capacities and species range-overlaps, and the cascading consequences of these changes on the maintenance and potential spillover of pathogens to novel hosts and to altered host communities (Daszak et al., 2001; Webb et al., 2002; Williams and Jackson, 2007; Faust et al., 2018). For example, host range expansions due to climate change may increase risk of pathogen sharing and spillover (Morales-Castilla et al., 2021). In addition, many host communities are also likely changing in diversity, with both decreases and increases in species richness common (Dornelas et al., 2014); yet the effects of biodiversity on disease prevalence remain controversial (Ostfeld and Keesing, 2012; Halliday and Rohr, 2019). Empirical studies provide evidence for both amplification and dilution effects, for example, increasing amphibian species richness has been linked to greater fungal pathogen prevalence (Becker and Zamudio, 2011), but reduced flatworm (*Ribeiroia*) prevalence (Johnson et al., 2012), and the mechanisms underlying

	Bacteria			**Viruses**		
Variable	Estimate effect on R_0	R^2	AIC	Estimate effect on R_0	R^2	AIC
log(MPD)	-0.39	0.86	-5004	-0.13	0.49	-6263
Ic	-0.02	0.07	4644	-0.05	0.02	-3022
n	-0.06	0.15	4218	-0.02	0.04	-3152
Multivar	2.37-0.36log(MPD)+0.01Ic-0.05n	0.87	-5299	2.92-0.13log(MPD)+6.1e^{-5}n+4.3e^{-3}Ic	0.49	-6265

Table 4: Regression coefficients as in Table 2, but assuming the phylogenetic transformation from Gilbert *et al.* (2012) for bacteria and viruses. Results are reported for Scenario 1 only, and are for simulations with $\nu = 0.1$, $\mu = 0.2$ and $\psi = 0.01$ and host species richness of 3, 4, 5, 7 and 10 species. The first three rows show estimates from the univariate regressions.

these contrasting dynamics are often unclear.

Part of the explanation for the mixed support for dilution versus amplification effects likely reflects differences in disease competence among host species. The addition of highly competent hosts to the community will generally promote amplification, while the addition of lower-competent hosts or strong interspecific competitors that reduces the abundance of highly-competent hosts will tend to promote dilution (Cortez and Duffy, 2021). Thus, generating accurate predictions of how changes in host diversity will impact disease prevalence requires data on host competence, which we lack for most species. However, host phylogeny might provide a useful proxy (Gilbert and Parker, 2016; Gougherty and Davies, 2021).

Here, we present a novel framework that builds upon recent observations that disease transmission in multi-host systems is often strongly phylogenetically structured. Using simulations of a compartmental SI disease model with the assumption that the probability of successful transmission is inversely related to the pairwise phylogenetic distance between species, we show that simply by altering the phylogenetic relationships among hosts, a community's disease outbreak potential (R_0) can switch between amplifying ($R_0 > 1$) to diluting ($R_0 < 1$) pathogen invasion risk, while species richness remains unchanged. In our simulations, adding a closely related host to a community will more likely amplify a disease, while adding a distant relative will more likely dilute it. When we include primary-secondary host dynamics, we additionally show that it is not only the phylogenetic structure of the host community that determines the outbreak potential of the pathogen, but also the phylogenetic placement of the primary host. We were able to explain up to 93% of the variation in simulated community R_0 using commonly-used phylogenetic diversity metrics that captured axes of tree topology and phylogenetic clustering, suggesting that these dimensions of host community structure might help in predicting disease spread. While the contribution of host phylogenetic structure to ecosystem functioning has been well studied (e.g. Cadotte et al., 2008) and there is increasing evidence for phylogenetic signal in disease transmission (Parker et al., 2015), these relationships have largely been explored independently; here, we provide a mechanistic model that unites them.

In our model, we make the well-supported assumption that the transmission of diseases among host species co-varies with the phylogenetic distance that separates them (Davies and Pedersen, 2008; Parker et al., 2015; Farrell and Davies, 2019b; Stephens et al., 2019; Wang et al., 2019; Gougherty and Davies, 2021), likely reflecting the tendency for closely related species to share similar immune systems, inherited, with modification, from a common ancestor.

By extension, we assume the likelihood of establishment success of a disease in a novel host is correlated with the immuno-compatibility and thus phylogenetic identity of (donor and recipient) hosts (Davies and Pedersen, 2008; Kuiken et al., 2006; Streicker et al., 2010; Gilbert et al., 2012; Huang et al., 2014). It has already been shown that the genetic diversity of host populations can affect disease outbreak potential (Altizer et al., 2003) and that R_0 may be inversely related to the number of host genotypes (Lively, 2010). We extend this reasoning to encompass host phylogenetic diversity and show that under assumptions of phylogenetic conservatism in host competence, community R_0 is lower in simulated communities with greater host phylogenetic diversity, indexed by the phylogenetic pairwise distances separating hosts in the community.

We present a model that allows for a *phylogenetic dilution* effect, by assuming an inverse relationship between phylogenetic distance between hosts and risk of pathogen sharing. In our model, we include a phylogenetic scaling parameter, α, that allows us to weight the relative importance of phylogenetic distance. We suggest α should be considered as a free parameter, best estimated from the data, and we discuss an example below using a scaling transformation from Gilbert et al. (2012). Importantly, however, any shape can be fit, depending on the host-pathogen system (see Appendix Section 5). It would be straightforward to transform branch lengths to match to alternative evolutionary models, such as the Ornstein–Uhlenbeck (O-U) process (Martins and Hansen, 1997), or more complex, rate variable models.

We explored dynamics across two alternative scenarios. First, we assume transmission success is simply weighted by the phylogenetic distance between donor and recipient hosts (Scenario 1), which we suggest might resemble disease dynamics in pathogen systems such as Rabies virus, characterised by frequent host jumps and fast mutation rates, allowing it to co-evolve with its host (Woolhouse et al., 2001; Mollentze et al., 2014). However, the majority of disease systems likely have more complex primary-secondary host dynamics, in which one or a few host species are the primary (or reservoir) hosts, and host competence co-varies with phylogenetic distance between the secondary (recipient) and primary host (Scenario 2). We highlight bovine tuberculosis (bTB) in buffalo as one exemplar disease system that fits this latter scenario. In Africa, bTB, which originated from cattle imported from Europe, is now endemic in African buffalo, but frequently spills over to other mammal hosts (Michel et al., 2006).

In both modelled scenarios with substitutive abundance (option 1), an increase in species richness (n) decreases R_0. The inclusion of a low competent host in a population with a constant contact rate per time unit (frequency

dependent transmission) can lead to a reduction in contacts among high competent hosts, resulting in a lower overall disease prevalence (Ostfeld and Keesing, 2000; Dobson, 2004; Rudolf and Antonovics, 2005; Keesing et al., 2006). Disease dilution is also possible with density dependent transmission if we assume the total number of individuals in the community remains constant (i.e. substitutive abundance), such that added hosts replace contacts among more competent hosts. Phylogenetic tree shape was weakly correlated with disease outbreak potential in both scenarios: more balanced trees reduce community R_0. A highly imbalanced tree is characterized by one or a few clades of very closely related species, which are more likely to have similar immune-competence, and thus easily share a disease, we suggest this parallels the *phylogenetic clade effect* discussed by Wang et al. (2019). However, in the absence of an assigned primary host, the best predictor of community R_0 (in combination with S and Ic) was the mean of the pairwise phylogenetic distance between hosts in the community (MPD). Communities composed of more closely related hosts had higher outbreak potentials than communities composed of more distantly related hosts because the average transmission success between recipient and donor hosts will be relatively high in the former and low in the latter. When we include a primary host in the model (Scenario 2), we are again able to explain most of the variation in community R_0 using just species richness, tree imbalance, and MPD. However, the model substituting MPD for the evolutionary isolation (measured as Evolutionary Distinctiveness; ED) of the primary host was a better fit to the data. In this modelled scenario, community R_0 is lower when the primary host is phylogenetically isolated from secondary hosts, and primary host ED is more important in mediating disease dynamics than overall host phylogenetic community structure.

Notably, when we assume an increase in species richness is additive with respect to community abundance (option 2), the effect of species richness (n) on R_0 switched from diluting to amplifying, while the phylogenetic dilution effect of MPD is largely unchanged. This modelled scenario illustrates how both amplification and dilution effects can operate simultaneously on different axes of community structure. Additionally, we observe a drop in explanatory power of ED in scenario 2, suggesting that when encounters are additive, and thus amplifying, the location of the primary host is of less importance in mediating outbreak potential.

The shape of the relationship between host phylogeny and disease outbreak potential, as well as the relative importance of the different axes of phylogenetic structure we explore, reflect both the underlying structure of our model and the parameter values we fit to it. To examine how these relationships might change under alternative functions describing the phylogenetic signal in the probability of pest sharing, we also fit our model using the functions from

Gilbert et al. (2012) estimated from empirical data for viruses and bacteria on plants. As we do not have information on primary hosts, we fit a model for scenario 1 only. For both pathogen types, we can see the relationship between R_0 and MPD is qualitatively different from that shown in our simulations assuming a simpler phylogenetic scaling. Perhaps most notably, the diluting effect of MPD is only apparent at much larger values, reflecting the ease with which many generalist pathogens can jump between closely related species (Schatz and Park, 2021). Phylogenetic constraints to host shifts are only apparent at high MPD, and for viruses there is evidence of a threshold below which host jumps are effectively unconstrained by phylogeny. Viruses are recognised as being able to make relatively large host jumps (Woolhouse et al., 2005; Park et al., 2018), therefore it may not be so surprising that we also find in our model that host community phylogenetic structure is less important in mediating their outbreak potential.

We recognise that our model, like all models, is an over-simplification, and the high R^2 from our regressions reflect some of the many simplifying assumptions we impose, including equal host population sizes and symmetries in contact rates. Importantly, we also modelled the intraspecific and interspecific contact rates as equal, which would have had an influence on the relative magnitude of amplification/dilution effects (Dobson, 2004; Joseph et al., 2013); however, this is easily modified if different assumptions on host contacts were a better fit to our understanding of a particular empirical system. In Appendix section 4 we show how R^2 from regressions increases when we switch to more realistic contact structures where the interspecific contact rates are reduced relative to intraspecific contacts. Our models are also deterministic, and thus ignore potential for stochastic dynamics, which can have significant impacts on disease outbreaks dynamics (Wolfe et al., 2005). We thus see many exciting extensions to the model framework we present here.

Our simulations, based upon the reasonable assumptions that spillover of pathogens is more likely between closely related hosts, highlight how the phylogenetic structure of host communities might additionally mediate risk of disease outbreaks: communities that are more phylogenetically diverse are also more likely to reduce disease outbreak potential (see also Davies and Pedersen, 2008; Parker et al., 2015; Farrell and Davies, 2019b; Wang et al., 2019, 2021). However, there is evidence that more evolutionarily distinct species—those adding the greatest phylogenetic diversity to communities—might also be more at risk of extinction (Daru et al., 2013), suggesting that extinction might shift communities towards increased potential for disease outbreaks. In addition, the non-random distribution of extinction on the tree-of-life suggests that, at least at local scales where disease dynamics play out, we risk losing a disproportionate amount of phylogenetic diversity if currently threatened species are lost to extinction (Huang et al.,

2012). Thus, we may be most at risk of losing the species and community attributes most likely to suppress disease, while the species less at risk of extinction tend to be those associated with faster life-histories and higher disease competence due to lower investment in pathogen defense (Joseph et al., 2013; Gibb et al., 2020; Halliday et al., 2020; Valenzuela-Sánchez et al., 2021).

References

Altizer, S., Harvell, D., and Friedle, E. (2003). Rapid evolutionary dynamics and disease threats to biodiversity. *Trends in Ecology and Evolution*, 18(11):589–596.

Barasona, J. A., Gortázar, C., de la Fuente, J., and Vicente, J. (2019). Host richness increases tuberculosis disease risk in game-managed areas. *Microorganisms*, 7(6):1–7.

Barrett, M., Belward, A., Bladen, S., Breeze, T., Burgess, N., Butchart, S., Clewclow, H., Cornell, S., Cottam, A., Croft, S., et al. (2018). Living planet report 2018: Aiming higher.

Becker, C. G. and Zamudio, K. R. (2011). Tropical amphibian populations experience higher disease risk in natural habitats. *Proceedings of the National Academy of Sciences*, 108(24):9893–9898.

Bengis, R. and Erasmus, J. (1988). Wildlife diseases in South Africa : a review. *Revue Scientifique et Technique de l'OIE*, 7(4):807–821.

Bourhy, H., Kissi, B., Audry, L., Smreczak, M., Sadkowska-Todys, M., Kulonen, K., Tordo, N., Zmudzinski, J. F., and Holmes, E. C. (1999). Ecology and evolution of rabies virus in europe. *Journal of General Virology*, 80(10):2545–2557.

Cadotte, M. W., Cardinale, B. J., and Oakley, T. H. (2008). Evolutionary history and the effect of biodiversity on plant productivity. *Proceedings of the National Academy of Sciences*, 105(44):17012–17017.

Cadotte, M. W. and Davies, T. J. (2016). *Phylogenies in ecology: a guide to concepts and methods*. Princeton University Press.

Cohen, H., Ponisio, L. C., Russell, K. A., Philpott, S. M., and McFrederick, Q. S. (2022). Floral resources shape parasite and pathogen dynamics in bees facing urbanization. *Molecular Ecology*, 31(7):2157–2171.

Colless, D. H. (1982). Phylogenetics: The theory and practice of phylogenetic systematics.

Cortez, M. H. and Duffy, M. A. (2021). The context-dependent effects of host competence, competition, and pathogen transmission mode on disease prevalence. *The American Naturalist*, 198(2):179–194.

Daru, B. H., Yessoufou, K., Mankga, L. T., and Davies, T. J. (2013). A Global Trend towards the Loss of Evolutionarily Unique Species in Mangrove Ecosystems. *PLoS ONE*, 8(6).

Daszak, P., Cunningham, A. A., and Hyatt, A. D. (2000). Emerging infectious diseases of wildlife - Threats to biodiversity and human health. *Science*, 287(5452):443–449.

Daszak, P., Cunningham, A. A., and Hyatt, A. D. (2001). 1. P. Daszak, A. A. Cunningham, A. D. Hyatt, Acta Trop 78, 103-16 (2001). *Acta Tropica*, 78(2):103–116.

Davies, T. J. and Pedersen, A. B. (2008). Phylogeny and geography predict pathogen community similarity in wild primates and humans. *Proceedings of the Royal Society B: Biological Sciences*, 275(1643):1695–1701.

De Garine-Wichatitsky, M., Caron, A., Kock, R., Tschopp, R., Munyeme, M., Hofmeyr, M., and Michel, A. (2013). A review of bovine tuberculosis at the wildlife–livestock–human interface in sub-saharan africa. *Epidemiology & Infection*, 141(7):1342–1356.

Diekmann, O., Heesterbeek, J., and Roberts, M. G. (2010). The construction of next-generation matrices for compartmental epidemic models. *Journal of the royal society interface*, 7(47):873–885.

Diekmann, O. and Kretzschmar, M. (1991). Patterns in the effects of infectious diseases on population growth. *Journal of Mathematical Biology*, 29(6):539–570.

Dobson, A. (2004). Population dynamics of pathogens with multiple host species. *American Naturalist*, 164(5 SUPPL.).

Dornelas, M., Gotelli, N. J., McGill, B., Shimadzu, H., Moyes, F., Sievers, C., and Magurran, A. E. (2014). Assemblage time series reveal biodiversity change but not systematic loss. *Science*, 344(6181):296–299.

Ezenwa, V. O., Etienne, R. S., Luikart, G., Beja-Pereira, A., and Jolles, A. E. (2010). Hidden consequences of living in a wormy world: Nematode-induced immune suppression facilitates tuberculosis invasion in African buffalo. *American Naturalist*, 176(5):613–624.

Ezenwa, V. O. and Jolles, A. E. (2015). Opposite effects of anthelmintic treatment on microbial infection at individual versus population scales. *Science*, 347(6218):175–177.

Faith, D. P. (1992). Conservation evaluation and phylogenetic diversity. *Biological conservation*, 61(1):1–10.

Farrell, M. J. and Davies, T. J. (2019a). Disease mortality in domesticated animals is predicted by host evolutionary relationships. *Proceedings of the National Academy of Sciences of the United States of America*, 116(16):7911–7915.

Farrell, M. J. and Davies, T. J. (2019b). Disease mortality in domesticated animals is predicted by host evolutionary relationships. *Proceedings of the National Academy of Sciences*, 116(16):7911–7915.

Faust, C. L., McCallum, H. I., Bloomfield, L. S., Gottdenker, N. L., Gillespie, T. R., Torney, C. J., Dobson, A. P., and Plowright, R. K. (2018). Pathogen spillover during land conversion. *Ecology Letters*, 21(4):471–483.

Fenton, A., Streicker, D. G., Petchey, O. L., and Pedersen, A. B. (2015). Are all hosts created equal? Partitioning host species contributions to parasite persistence in multihost communities. *American Naturalist*, 186(5):610–622.

Gibb, R., Redding, D. W., Chin, K. Q., Donnelly, C. A., Blackburn, T. M., Newbold, T., and Jones, K. E. (2020). Zoonotic host diversity increases in human-dominated ecosystems. *Nature*, 584(7821):398–402.

Gilbert, G. S., Magarey, R., Suiter, K., and Webb, C. O. (2012). Evolutionary tools for phytosanitary risk analysis: Phylogenetic signal as a predictor of host range of plant pests and pathogens. *Evolutionary Applications*, 5(8):869–878.

Gilbert, G. S. and Parker, I. M. (2016). The evolutionary ecology of plant disease: a phylogenetic perspective. *Annual Review of Phytopathology*, 54:549–578.

Gilbert, G. S. and Webb, C. O. (2007). Phylogenetic signal in plant pathogen–host range. *Proceedings of the National Academy of Sciences*, 104(12):4979–4983.

Gougherty, A. V. and Davies, T. J. (2021). Towards a phylogenetic ecology of plant pests and pathogens. *Philosophical Transactions of the Royal Society B*, 376(1837):20200359.

Halliday, F. W. and Rohr, J. R. (2019). Measuring the shape of the biodiversity-disease relationship across systems reveals new findings and key gaps. *Nature Communications*, 10(1):1–10.

Halliday, F. W., Rohr, J. R., and Laine, A.-L. (2020). Biodiversity loss underlies the dilution effect of biodiversity. *Ecology letters*, 23(11):1611–1622.

Han, B. A., O'Regan, S. M., Paul Schmidt, J., and Drake, J. M. (2020). Integrating data mining and transmission theory in the ecology of infectious diseases. *Ecology Letters*, 23(8):1178–1188.

Haydon, D. T., Cleaveland, S., Taylor, L. H., and Laurenson, M. K. (2002). Identifying reservoirs of infection: A conceptual and practical challenge. *Emerging Infectious Diseases*, 8(12):1468–1473.

Huang, S., Davies, T. J., and Gittleman, J. L. (2012). How global extinctions impact regional biodiversity in mammals. *Biology Letters*, 8(2):222–225.

Huang, Z. Y., Xu, C., Van Langevelde, F., Prins, H. H., Ben Jebara, K., and De Boer, W. F. (2014). Dilution effect and identity effect by wildlife in the persistence and recurrence of bovine tuberculosis. *Parasitology*, 141(7):981–987.

IPBES (2020). Workshop Report on Biodiversity and Pandemics of the Intergovernmental Platform on Biodiversity and Ecosystem Services. page 108.

Isaac, N. J., Turvey, S. T., Collen, B., Waterman, C., and Baillie, J. E. (2007). Mammals on the edge: conservation priorities based on threat and phylogeny. *PloS one*, 2(3):e296.

IUCN (2021). The iucn red list of threatened species. version 2021-1. https://www.iucnredlist.org. downloaded on 15 june 2021.

Johnson, P. T., Preston, D. L., Hoverman, J. T., Henderson, J. S., Paull, S. H., Richgels, K. L., and Redmond, M. D. (2012). Species diversity reduces parasite infection through crossgenerational effects on host abundance. *Ecology*, 93(1):56–64.

Jolles, A. E., Cooper, D. V., and Levin, S. A. (2005). Hidden effects of chronic tuberculosis in African buffalo. *Ecology*, 86(9):2358–2364.

Jones, K. E., Patel, N. G., Levy, M. A., Storeygard, A., Balk, D., Gittleman, J. L., and Daszak, P. (2008). Global trends in emerging infectious diseases. *Nature*, 451(7181):990–993.

Joseph, M. B., Mihaljevic, J. R., Orlofske, S. A., and Paull, S. H. (2013). Does life history mediate changing disease risk when communities disassemble? *Ecology letters*, 16(11):1405–1412.

Keeling, M. J. and Rohani, P. (2011). *Modeling infectious diseases in humans and animals*. Princeton university press.

Keesing, F., Belden, L. K., Daszak, P., Dobson, A., Harvell, C. D., Holt, R. D., Hudson, P., Jolles, A., Jones, K. E., Mitchell, C. E., Myers, S. S., Bogich, T., and Ostfeld, R. S. (2010). Impacts of biodiversity on the emergence and transmission of infectious diseases. *Nature*, 468(7324):647–652.

Keesing, F., Holt, R. D., and Ostfeld, R. S. (2006). Effects of species diversity on disease risk. *Ecology Letters*, 9(4):485–498.

Keesing, F. and Ostfeld, R. S. (2021). Impacts of biodiversity and biodiversity loss on zoonotic diseases. *Proceedings of the National Academy of Sciences*, 118(17):e2023540118.

Kermack, W. O. and McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character*, 115(772):700–721.

Kuiken, T., Holmes, E. C., McCauley, J., Rimmelzwaan, G. F., Williams, C. S., and Grenfell, B. T. (2006). Host species barriers to influenza virus infections. *Science*, 312(5772):394–397.

Lively, C. M. (2010). The effect of host genetic diversity on disease spread. *American Naturalist*, 175(6):17–20.

LoGiudice, K., Ostfeld, R. S., Schmidt, K. A., and Keesing, F. (2003). The ecology of infectious disease: Effects of host diversity and community composition on lyme disease risk. *Proceedings of the National Academy of Sciences of the United States of America*, 100(2):567–571.

Luis, A. D., Kuenzi, A. J., and Mills, J. N. (2018). Species diversity concurrently dilutes and amplifies transmission in a zoonotic host–pathogen system through competing mechanisms. *Proceedings of the National Academy of Sciences of the United States of America*, 115(31):7979–7984.

Martins, E. P. and Hansen, T. F. (1997). Phylogenies and the comparative method: a general approach to incorporating phylogenetic information into the analysis of interspecific data. *The American Naturalist*, 149(4):646–667.

May, R. M. and Anderson, R. M. (1984). Spatial heterogeneity and the design of immunization programs. *Mathematical Biosciences*, 72(1):83–111.

Michel, A. L., Bengis, R. G., Keet, D., Hofmeyr, M., De Klerk, L., Cross, P., Jolles, A. E., Cooper, D., Whyte, I., Buss, P., et al. (2006). Wildlife tuberculosis in south african conservation areas: implications and challenges. *Veterinary microbiology*, 112(2-4):91–100.

Mollentze, N., Biek, R., and Streicker, D. G. (2014). The role of viral evolution in rabies host shifts and emergence. *Current opinion in virology*, 8:68–72.

Mollentze, N. and Streicker, D. G. (2020). Viral zoonotic risk is homogenous among taxonomic orders of mammalian and avian reservoir hosts. *Proceedings of the National Academy of Sciences of the United States of America*, 117(17):9423–9430.

Morales-Castilla, I., Pappalardo, P., Farrell, M. J., Aguirre, A. A., Huang, S., Gehman, A.-L. M., Dallas, T., Gravel, D., and Davies, T. J. (2021). Forecasting parasite sharing under climate change. *Philosophical Transactions of the Royal Society B*, 376(1837):20200360.

Ostfeld, R. S. and Keesing, F. (2000). The function of biodiversity in the ecology of vector-borne zoonotic diseases. *Canadian Journal of Zoology*, 78(12):2061–2078.

Ostfeld, R. S. and Keesing, F. (2012). Effects of host diversity on infectious disease. *Annual Review of Ecology, Evolution, and Systematics*, 43:157–182.

Park, A. W., Farrell, M. J., Schmidt, J. P., Huang, S., Dallas, T. A., Pappalardo, P., Drake, J. M., Stephens, P. R., Poulin, R., Nunn, C. L., and Davies, T. J. (2018). Characterizing the phylogenetic specialism-generalism spectrum of mammal parasites. *Proceedings of the Royal Society B: Biological Sciences*, 285(1874).

Parker, I. M., Saunders, M., Bontrager, M., Weitz, A. P., Hendricks, R., Magarey, R., Suiter, K., and Gilbert, G. S. (2015). Phylogenetic structure and host abundance drive disease pressure in communities. *Nature*, 520(7548):542–544.

Revell, L. J. (2012). phytools: an r package for phylogenetic comparative biology (and other things). *Methods in ecology and evolution*, (2):217–223.

Roche, B., Dobson, A. P., Guégan, J. F., and Rohani, P. (2012). Linking community and disease ecology: The impact of biodiversity on pathogen transmission. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1604):2807–2813.

Rohr, J. R., Civitello, D. J., Halliday, F. W., Hudson, P. J., Lafferty, K. D., Wood, C. L., and Mordecai, E. A. (2020). Towards common ground in the biodiversity–disease debate. *Nature Ecology and Evolution*, 4(1):24–33.

Rudolf, V. H. and Antonovics, J. (2005). Species coexistence and pathogens with frequency-dependent transmission. *American Naturalist*, 166(1):112–118.

Schatz, A. M. and Park, A. W. (2021). Host and parasite traits predict cross- species parasite acquisition by introduced mammals.

Steffen, W., Broadgate, W., Deutsch, L., Gaffney, O., and Ludwig, C. (2015). The trajectory of the anthropocene: the great acceleration. *The Anthropocene Review*, 2(1):81–98.

Stephens, P. R., Altizer, S., Ezenwa, V. O., Gittleman, J. L., Moan, E., Han, B., Huang, S., and Pappalardo, P. (2019). Parasite sharing in wild ungulates and their predators: effects of phylogeny, range overlap, and trophic links. *Journal of Animal Ecology*, 88(7):1017–1028.

Stewart Merrill, T. E., Calhoun, D. M., and Johnson, P. T. (2022). Beyond single host, single parasite interactions: Quantifying competence for complete multi-host, multi-parasite communities. *Functional Ecology*, 36(8):1845–1857.

Streicker, D. G., Fenton, A., and Pedersen, A. B. (2013). Differential sources of host species heterogeneity influence the transmission and control of multihost parasites. *Ecology Letters*, 16(8):975–984.

Streicker, D. G., Turmelle, A. S., Vonhof, M. J., Kuzmin, I. V., McCracken, G. F., and Rupprecht, C. E. (2010). Host phylogeny constrains cross-species emergence and establishment of rabies virus in bats. *Science*, 329(5992):676–679.

UN (2020). *Zoonotic diseases and how to break the chain of transmission A Scientific Assessment with Key Messages for Policy-Makers.*

Valenzuela-Sánchez, A., Wilber, M. Q., Canessa, S., Bacigalupe, L. D., Muths, E., Schmidt, B. R., Cunningham, A. A., Ozgul, A., Johnson, P. T., and Cayuela, H. (2021). Why disease ecology needs life-history theory: a host perspective. *Ecology Letters*, 24(4):876–890.

Wang, Y. X., Matson, K. D., Prins, H. H., Gort, G., Awada, L., Huang, Z. Y., and de Boer, W. F. (2019). Phylogenetic structure of wildlife assemblages shapes patterns of infectious livestock diseases in Africa. *Functional Ecology*, 33(7):1332–1341.

Wang, Y. X., Matson, K. D., Santini, L., Visconti, P., Hilbers, J. P., Huijbregts, M. A., Xu, Y., Prins, H. H., Allen, T., Huang, Z. Y., et al. (2021). Mammal assemblage composition predicts global patterns in emerging infectious disease risk. *Global Change Biology*, 27(20):4995–5007.

Webb, C. O., Ackerly, D. D., McPeek, M. A., and Donoghue, M. J. (2002). Phylogenies and community ecology. *Annual review of ecology and systematics*, 33(1):475–505.

White, L. A., Forester, J. D., and Craft, M. E. (2017). Using contact networks to explore mechanisms of parasite transmission in wildlife. *Biological Reviews*, 92(1):389–409.

Williams, J. W. and Jackson, S. T. (2007). Novel climates, no-analog communities, and ecological surprises. *Frontiers in Ecology and the Environment*, 5(9):475–482.

Wolfe, N. D., Daszak, P., Kilpatrick, A. M., and Burke, D. S. (2005). Bushmeat hunting, deforestation, and prediction of zoonotic disease emergence. *Emerging Infectious Diseases*, 11(12):1822–1827.

Woolhouse, M. E., Haydon, D. T., and Antia, R. (2005). Emerging pathogens: The epidemiology and evolution of species jumps. *Trends in Ecology and Evolution*, 20(5):238–244.

Woolhouse, M. E., Taylor, L. H., and Haydon, D. T. (2001). Population biology of multihost pathogens. *Science*, 292(5519):1109–1112.

Zuur, A. F., Ieno, E. N., and Elphick, C. S. (2010). A protocol for data exploration to avoid common statistical problems. *Methods in ecology and evolution*, 1(1):3–14.

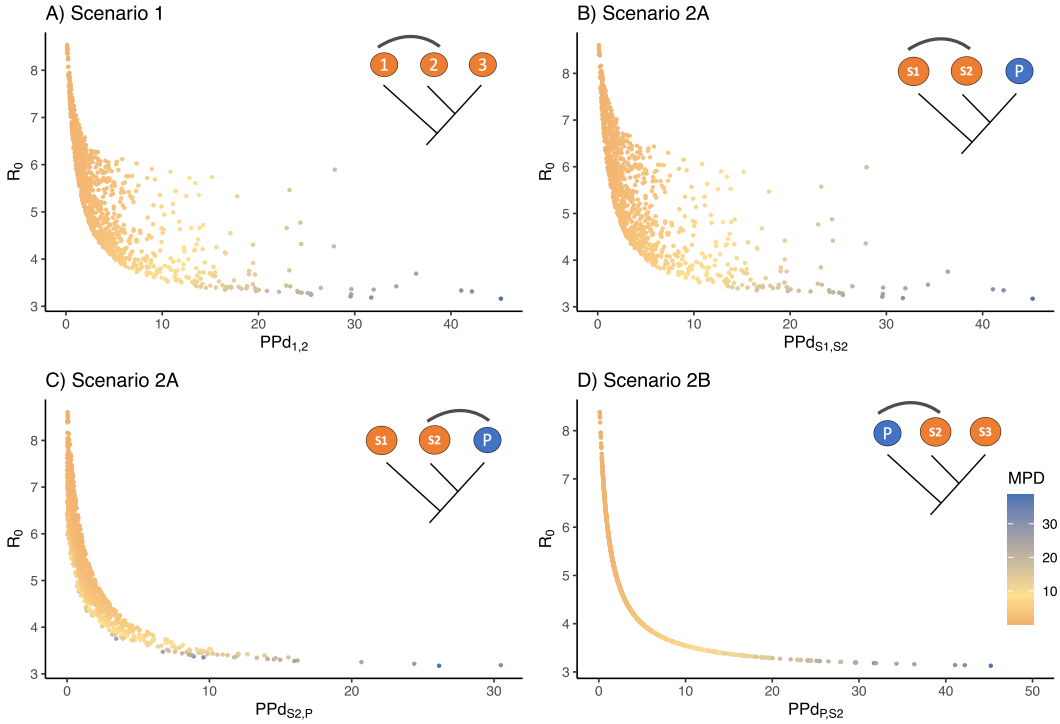

Figure 1: Community disease outbreak potential (R_0) plotted against the phylogenetic distance (PPd) between host pairs indicated on the x-axes. Shading is proportional to MPD. Inset within each graph provides an illustration of the alternative primary-secondary host dynamics for a simple three-taxon host system. The first panel (A) shows scenario 1, with no assigned primary host. The second (B) and third (C) panels show scenario 2a (primary host as ingroup) with blue circles representing the primary host (P), and orange circles the secondary hosts (S1 and S2). (B) shows the pairwise phylogenetic distance (PPd) between the two secondary hosts (S1 and S2) on the x-axis, whereas (C) shows the PPd between the primary host (P) and secondary host 2 (S2). The fourth panel (D) shows Scenario 2b, with the primary host sister to both secondary hosts (primary host as outgroup). Here both secondary hosts are equidistant from the primary host, and thus the system converges on a simple two-host model, and R_0 is effectively determined by tree depth. Grey line in illustration shows the distance between the two hosts on x-axis of each graph. Simulated data from the model with $\alpha = 1$, $\nu = 0.1$, $\mu = 0.2$, $\psi = 0.01$.

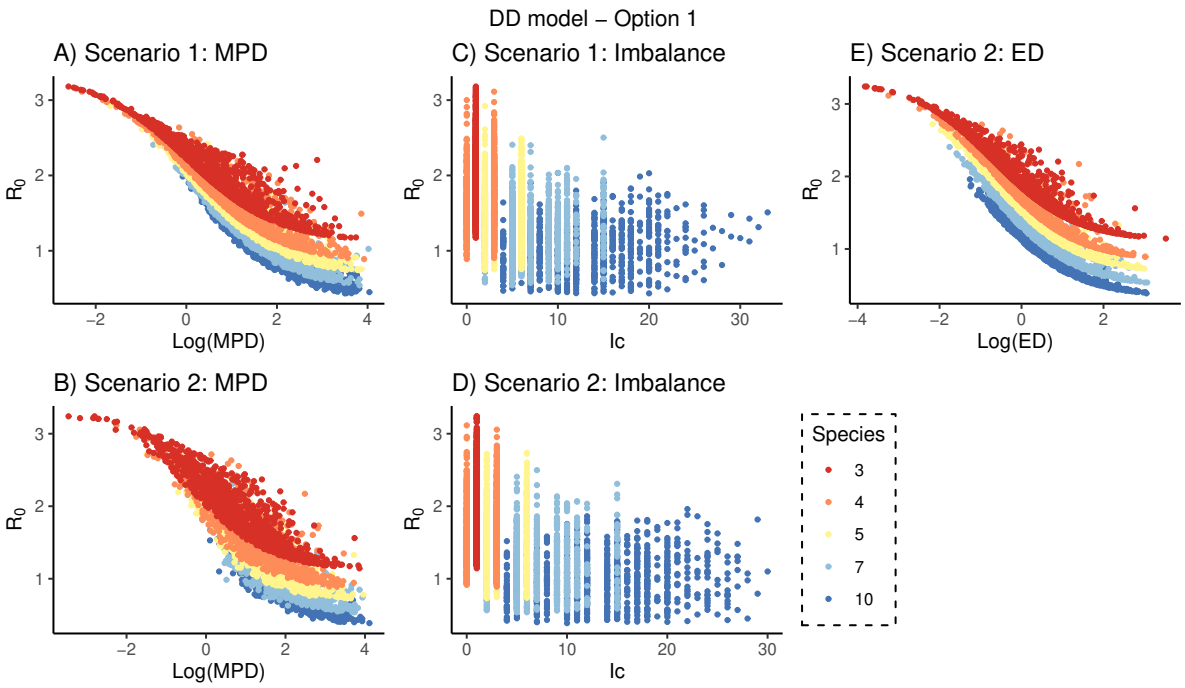

Figure 2: Modelled relationships between phylogenetic community structure and R_0 assuming transmission success is inversely weighted by the phylogenetic distance between hosts and replacing individuals (Option 1, substitutive). Scenario 1 (no primary host): A) Mean Pairwise Distance (MPD) of the community is inversely related to R_0. B) Relationship between R_0 and the Mean Pairwise phylogenetic Distance (MPD) between host species, with outbreak potential decreasing logistically with increasing MPD. C) Imbalance of the phylogenetic tree representing the host species community shows a significant but weak trend for R_0 to increase with increasing tree balance. Scenario 2 (single primary host). D) Relationship between R_0 and phylogenetic tree Imbalance (Ic), shows high variation. E) Relationship between R_0 and Evolutionary distinctiveness (ED) of the primary host, demonstrating similar dynamics to MPD. Data from simulations with $\kappa = 1$, $\alpha = 1$, $\nu = 0.1$, $\mu = 0.2$ and $\psi = 0.01$. Colours indicate simulated communities with different host richness (n=3,4,5,7,10).

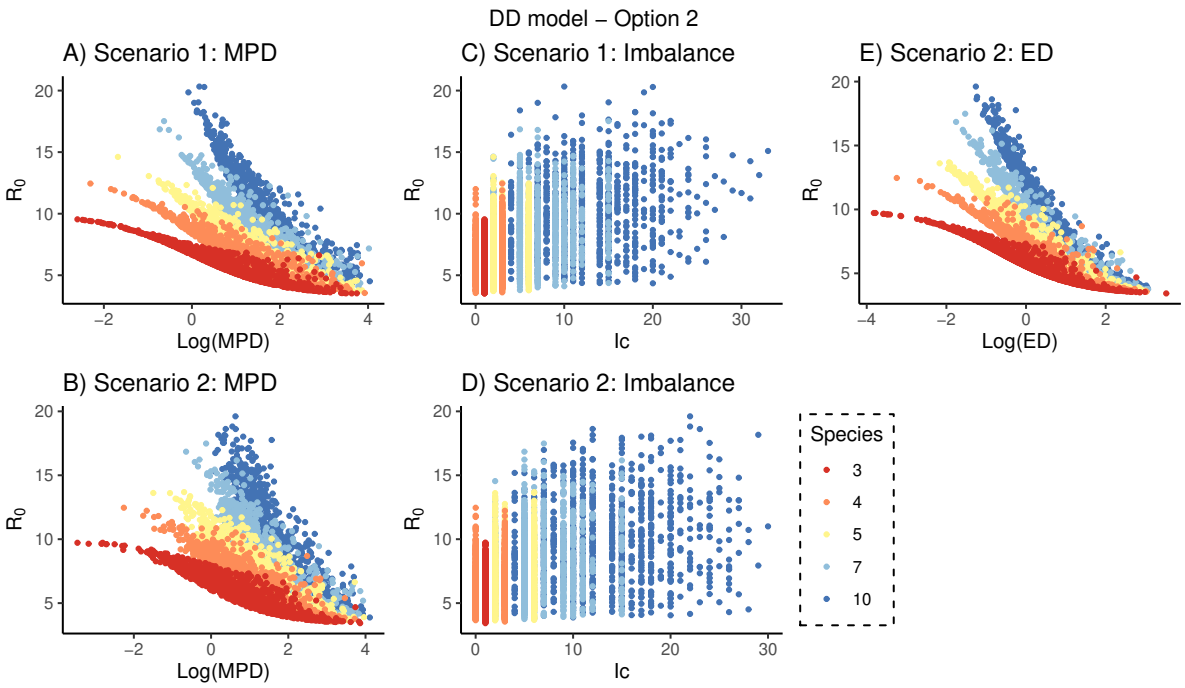

Figure 3: Modelled relationships between phylogenetic community structure and R_0 assuming transmission success is inversely weighted by the phylogenetic distance between hosts and adding species increases total community size (Option 2, additive). Scenario 1 (no primary host): A) Mean Pairwise Distance (MPD) of the community is inversely related to R_0. B) Relationship between R_0 and the Mean Pairwise phylogenetic Distance (MPD) between host species, with outbreak potential decreasing logistically with increasing MPD, but here the relationship becomes less clear once species richness increases. C) Imbalance of the phylogenetic tree representing the host species community shows a significant but weak trend for R_0 to increase with increasing tree balance. Scenario 2 (single primary host). D) Relationship between R_0 and phylogenetic tree Imbalance (Ic), illustrating high variation. E) Relationship between R_0 and Evolutionary distinctiveness (ED) of the primary host, demonstrating similar dynamics to MPD. Data from simulations with $\kappa = 1$, $\alpha = 1$, $\nu = 0.1$, $\mu = 0.2$ and $\psi = 0.01$. Colours indicate simulated communities with different host richness (n=3,4,5,7,10).

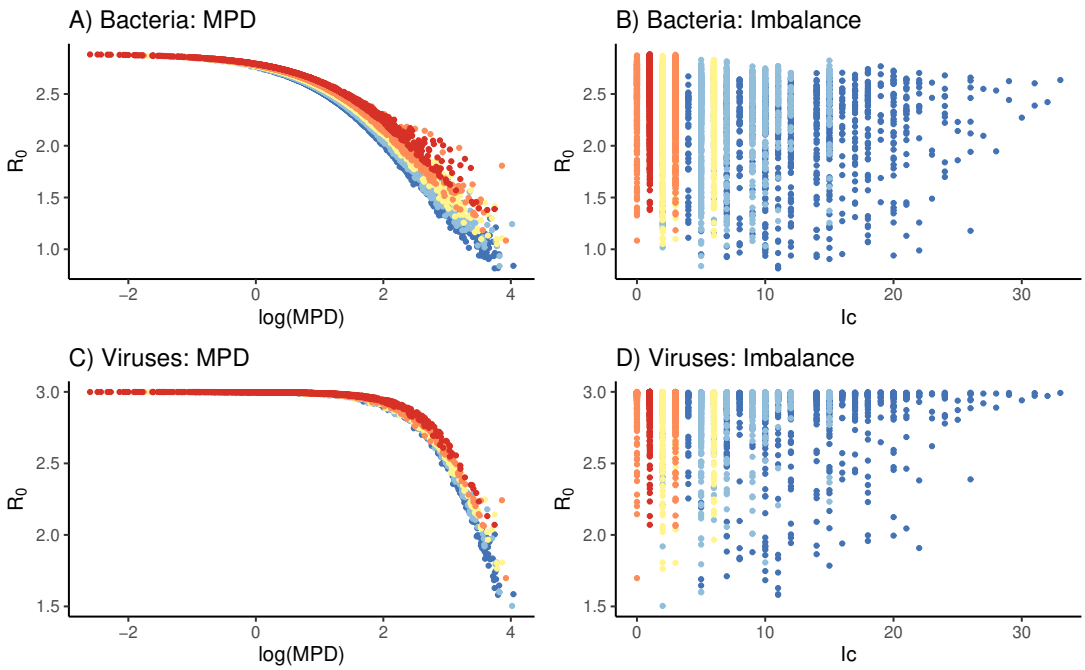

Figure 4: Modelled relationships between phylogenetic community structure and R_0 assuming the phylogenetic transformation from Gilbert *et al.* (2012) for bacteria and viruses, Scenario 1 and model Option 1 (replacing individuals). A and C illustrate the effect of Mean Pairwise distance between hosts, B and D illustrate the effect of tree Imbalance.