PROCEEDINGS B

royalsocietypublishing.org/journal/rspb

Research

Cite this article: Kruger A, Gougherty A, Legault G, Davies TJ. 2025 Pest host expansion as a scale-free stepwise process across the host phylogeny. *Proc. R. Soc. B* **292**: 20251793.
https://doi.org/10.1098/rspb.2025.1793

Received: 11 July 2025
Accepted: 6 October 2025

Subject Category:
Global change and conservation

Subject Areas:
ecology, plant science

Keywords:
host range expansion, phylogenetic ecology, pathogens, insect pests, novel species interactions, spillover, plant disease ecology

Author for correspondence:
Avery Kruger
e-mails: avery.kruger@botany.ubc.ca; avery.m.kruger@gmail.com

Electronic supplementary material is available online at https://doi.org/10.6084/m9.figshare.c.8126260.

THE ROYAL SOCIETY PUBLISHING

Pest host expansion as a scale-free stepwise process across the host phylogeny

Avery Kruger[1], Andrew Gougherty[2], Geoffrey Legault[3] and T. Jonathan Davies[4]

[1]Department of Botany, The University of British Columbia, Vancouver, British Columbia, Canada
[2]USDA Forest Service Northern Research Station, Madison, WI, USA
[3]Unaffiliated
[4]University of British Columbia Department of Forest and Conservation Sciences, Vancouver, British Columbia, Canada

AK, 0000-0003-0410-7940; TJD, 0000-0003-3318-5948

The present and future host ranges of pests have important implications for ecology, economics and health. Most multi-host pests have phylogenetically conserved host symbioses, but jumps to phylogenetically distant hosts are surprisingly common and can have large fitness consequences. We introduce a model representing host distributions as outcomes of random jumps across a host phylogeny, with jump probabilities proportional to a power function of phylogenetic distance, parallel to Lévy flights in foraging behaviour. Using machine learning, we fit our model to empirical data on insect and pathogen pests of North American trees and estimate the parameter controlling the distribution of jump distance. We find that non-native insect pests are more likely to make jumps to phylogenetically distant hosts than native insect pests, although we did not observe a similar distinction between non-native and native pathogens. Finally, we use estimated jump parameters to describe the risk of future host expansion across the phylogeny of North American trees.

1. Introduction

Antagonistic interactions between pests or parasites and their hosts drive fundamental processes in evolution and ecology (e.g. Red Queen dynamics and Janzen–Connell effects [1,2]) and impact human health (e.g. disease, including novel zoonotic spillover [3,4]) and the global economy (e.g. livestock and agricultural losses, reduction in international trade and labour shortages [5]). Intervention can reduce the prevalence of endemic disease in human-managed crops but is often costly [3]. For example, the removal and replacement of ash trees (*Fraxinus* sp.) in the United States following the introduction and spread of the emerald ash borer (*Agrilus planipennis*) has been estimated at billions of dollars due to failure to detect its invasion prior to spread [6,7]. However, action to prevent the introduction and spread of pests can be successful if well-targeted [8,9]. It is therefore important to develop tools to identify candidate hosts that may be susceptible to novel pest invasions.

To successfully emerge on a new host, a pest must recognize it as a potential host, avoid its defences, extract resources from it and successfully reproduce. First, pests with behavioural agency (e.g. insects) must respond to novel chemical or visual stimuli or develop new behaviour in response to old signals [10]. Next, after encountering the novel host, the pest must evade or withstand physical or chemical defences. Fungal, microbial or viral pests, for example, may lose or alter molecular signals to avoid recognition by host receptors and consequent immune response before proliferating [11,12].

Finally, if the pest successfully extracts enough resources to reproduce, it may produce a new generation adapted for exploitation of the new host.

Phylogenetic conservation of both host traits involved in defence and pest traits involved in host recognition and exploitation produces patterns in which closely related hosts tend to share more pests than do more distantly related hosts [13,14]. Phylogeny, thus provides a useful tool for predicting spillover risk (see [15,16]), and it is possible to make risk predictions based on phylogenetic distance between hosts and novel targets [15,17–20]. The phylogenetic distributions of pests have been well-studied (e.g. [14,21,22]). Descriptions of host distributions have typically been characterized by transformations of phylogenetic metrics such as the mean pairwise distance or phylogenetic distance among or between host taxa [13,23]. However, these pattern-based descriptions do not explain how the distributions have come to be, and they may also poorly describe the distributions of pests across a host phylogeny, which are often irregular, featuring multiple disparate clusters of closely related hosts. Pests may often expand their host breadth to encompass closely related sister taxa of established hosts, but they can also make large leaps to distantly related taxa, producing patterns of distinct clusters across the host phylogeny [23,24]. Large phylogenetic jumps can result in highly virulent infections [25], yet we currently lack a process-based model that effectively describes this uneven distribution of pests on a host phylogeny.

Methods to predict unobserved interactions between pests and hosts (link predictions) have developed recently, generally assuming phylogenetic patterns rather than generative processes and calculating probability as an exponential function of phylogenetic distance, which itself may be transformed by a single-parameter evolutionary model like early-burst evolution [26–28]. Though these models are primarily designed to predict unobserved interactions rather than novel interactions, they can in principle be used to do so. However, as currently constructed, these models are scale-sensitive—results depend on the units of phylogenetic distance—and unable to produce patterns expected from high phylogenetic constraint due to the somewhat even relative probabilities they ascribe across the phylogeny of potential hosts.

Here, drawing upon movement patterns described as Lévy flights in the foraging literature, we propose a new scale-free model in which host switches are characterized as a generative, stepwise jumping process across the tips of a host phylogeny, paralleling the actual process of host range expansion. Steps occur with probability proportional to their phylogenetic distance taken to a negative exponent we call the jump parameter (figure 1). One step of this process from a single host is equivalent to a biased random walk on a graph [29,30], which has direct parallels to a Lévy flight model [31], although the parallels diverge as pests associate with increasing numbers of hosts. Under this jump process, when the jump parameter is 0, the random walk is unbiased, meaning pests are phylogenetically unconstrained and can infect all potential hosts with equal probability.

At all positive parameters, pests exhibit some degree of phylogenetic constraint; even when the jump parameter is small, pests are more likely to infect taxa closely related to extant hosts, but they will occasionally make larger jumps to distant clades, nucleating new clusters of closely related hosts. As the jump parameter becomes larger, the bias against large jumps becomes stronger, and pests instead tend to jump only to the closest relatives of their current hosts, becoming more likely to exhibit strong phylogenetic clustering. Interpreted as a diffusion process, jump parameters between 0 and 1 produce superdiffusion of pests across host phylogenetic distance, while values >1 produce subdiffusion, reflecting strong phylogenetic constraints (and values of one produce diffusion, also known as Brownian motion, over phylogenetic distance) [32]. In theory, a negative parameter would be more likely to produce over-dispersion than clustering, but such parameters are not biologically realistic because probabilities entirely depend on the scope of the host phylogeny considered.

We fit our model to 221 tree pests with comprehensive host records in North America and describe the phylogenetic constraint of the jump process of each pest. These data do not contain any chronological information about host range expansion, so our model is temporally implicit in that steps occur in chronological order, but time elapsed between steps is unknown. Our model assumes that the distribution of pests across the host phylogeny reflects a process of jumps and host breadth expansion; however, it is also possible that tree pests have diversified with their hosts and that their phylogenetic distribution has been additionally shaped by co-evolutionary history and vertical transmission from ancestral to descendent hosts [13]. To examine whether there was evidence to suggest stronger support for one process versus the other, we compared the estimated jump parameter between pests native and non-native to North America. The distribution of hosts of non-native pests must reflect a history of pest jumps and host breadth expansion, whereas the distribution of native pests might reflect a combination of both co-diversification and host switching. Finally, using our jump parameter estimates, we describe the relative risk of novel pest spillover should another pest jump event take place.

2. Methods

We combined a dataset of tree pests and pathogens representing 181 native species and 40 non-native species in North America, including insects, and bacterial and fungal pathogens, observed as pests across 262 genera within 88 families of native North American trees, with information on host phylogeny. Pest host ranges were adapted from Wang *et al*. [22], who expanded on the hosts listed for important North American tree pests from Potter *et al*. [33]. We excluded pests with fewer than four observed hosts, as these did not have sufficient data for modelling the distribution of jump distances, resulting in a reduced dataset of 130 native and 31 non-native pest species. Phylogenetic relationships among host genera were quantified using a phylogeny created with the V.PhyloMaker package [34] in R. Briefly, this package prunes a mega-tree of extant vascular plants to our host genera of interest and appends any genera absent from the mega-tree to the basal node of the respective family.

We parameterize our jump model using a machine learning technique known as convolutional random features [35], here referred to as convolutional kitchen sinks (CKS) [36], to learn jump parameters from simulated data for each pest and then use this information to estimate the parameter best describing the empirically observed distribution of each pest on the host phylogeny.

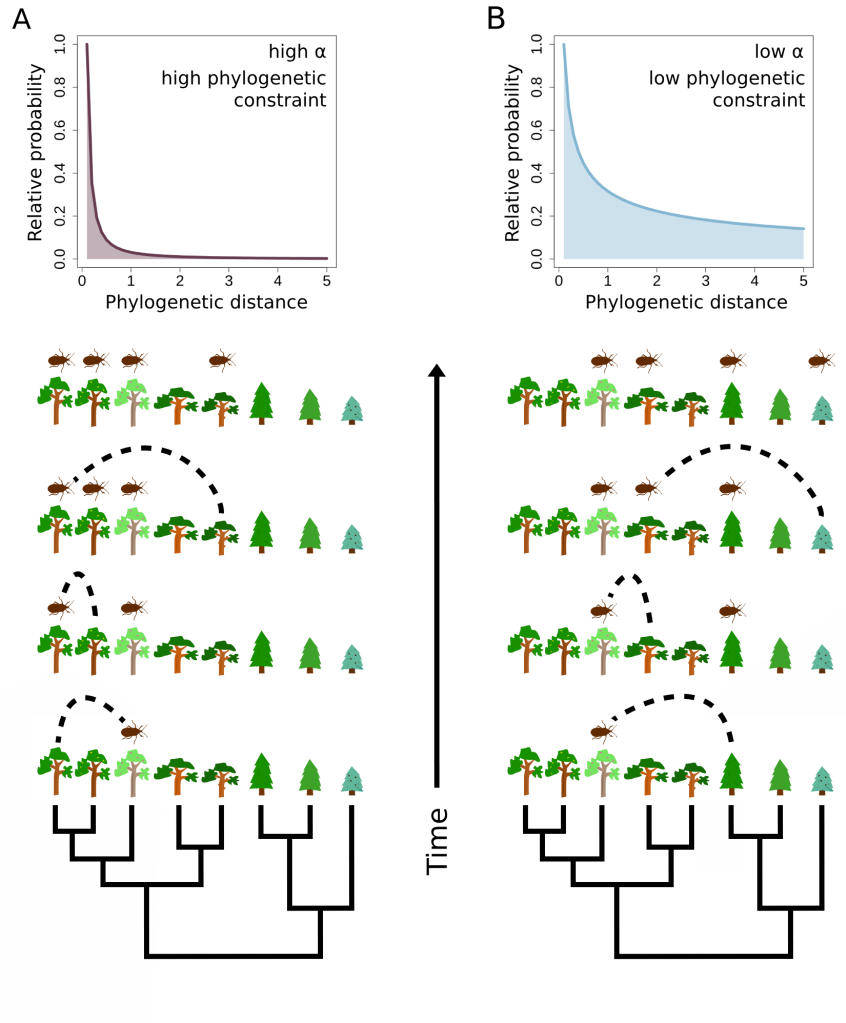

Figure 1. An illustration of host range expansion of a pest over a host phylogeny produced by a stepwise jumping process, contrasting high and low phylogenetic constraint. The probability of jumping to a new host from a given current host decreases with the power of the phylogenetic distance between them. (A) A pest with a large jump parameter (high phylogenetic constraint) is likely to jump to a close relative of its current host(s). If it makes an unlikely jump to a distant host, the frontier of likely jumps expands greatly. When phylogenetic constraints are moderate, this process may produce multiple clusters of closely related hosts. (B) To a pest with a small jump parameter (low phylogenetic constraint), even distant relatives of current hosts are liable to be the recipient of host range expansion. Such pests may be more evenly distributed across their host phylogeny.

(a) Simulating training data

We first constructed for the host plants a symmetric phylogenetic distance matrix D of dimensions $n \times n$, where D_{ij} is the phylogenetic distance (in millions of years, though results are insensitive to units) separating genera i and j (twice the time since they diverged from their most recent common ancestor). This matrix provides the underlying distances used for calculating jump probabilities.

To initiate each simulation, we selected a genus at random as the initial host. We also explored the performance of simulations instead seeded with a random host selected from pests' observed host ranges. We then performed stepwise pest jumps until the observed number of hosts for the pest was reached. At each step, to generate a new pest jump, for each potential host $i \in H$ (where H is the set of uninfected hosts), we calculated the probability of a transition, T_{ij}, of a pest jumping from host j to host i as:

$$T_{ij} = \frac{D_{ij}^{-\alpha} A_{ij}}{\sum_{x \in H} D_{xj}^{-\alpha} A_{xj}}, \tag{2.1}$$

where D_{ij} is the phylogenetic distance between genera i and j, α is the jump parameter and A_{ij} is an element of an adjacency matrix representing the connectedness (0 or 1) of genus i to j. In our model, we assume all hosts and potential hosts are connected, meaning $A_{ij} = 1$, but this can be modified, as we discussed as follows. Equation (2.1) is equivalent to the transition probability for a biased random walk on a graph [29,37] with $D_{ij}^{-\alpha}$ representing the bias. Whenever there were multiple hosts (i.e. all times after the first host jump), the total transition probability for each potential host was the sum of equation (2.1) evaluated for each $j \in N$ (where N is the set of infected hosts), normalized by the size of N. Simply put, this was the relative probability of jumping to a given potential host from any current hosts, normalized by the sum of all the probabilities of possible jumps to any potential hosts from any current hosts.

$$T_{i,\,j\in N} = \frac{1}{n(N)} \sum_{j\in N} \frac{D_{ij}^{-\alpha} A_{ij}}{\sum_{x\in H} D_{xj}^{-\alpha} A_{xj}} \tag{2.2}$$

We calculated the total transition rate for each potential host, creating a vector of probabilities of each potential host being infected upon the next jump. One host was then selected at random based on those probabilities and was added to the set of infected hosts (N) and removed from the set of potential hosts (H), and the process was repeated. For each pest, we simulated a pest jump process 5000 times, which was sufficient to reliably estimate α from the simulated data according to test simulations and analysis.

(b) Training the machine learning model

We used the simulated data to estimate the jump parameter, α, best fitting the observed host distribution for each pest using CKS. This machine learning method involves transforming an input vector (here, a simulated or observed host distribution ordered by the position of host genera on the host phylogeny) into a feature map (by approximating a convolutional kernel feature map) and then applying ridge regression (a linear model regularized to prevent overfitting using a quadratic loss function) with jump parameter α as the dependent variable and elements of the feature map (i.e. abstract features of the host distribution) as the independent variables (see [36] for detailed CKS methodology). The CKS method is advantageous over other methods because of its computational efficiency and because generation of the feature map is 'automatic', depending only on a small number of hyperparameters (for a full description of how this tool was implemented, see [36]).

Model input data were $1 \times n$ vectors describing whether tree genera were hosts (one) or not (zero) in each simulation, ordered in sequence of the tips on the tree phylogeny. CKS transformed each vector into a feature map that was then used to train a ridge regression model given the known values of α. This produced coefficients for each element of the feature map that could then be used to predict α given new data.

We trained models independently on each pest simulation data set. For each pest, we separated simulated data into 4000 training data and 1000 test data. We trained CKS models on the training data and validated models on the test data by correlating predicted values on known values and calculating the R^2 and slope of the correlation to determine the precision and accuracy of the predictions.

(c) Predicting with machine learning model

For each pest, we used empirical host data to produce a vector of ones and zeroes representing host and non-host, respectively, corresponding to the vectors used in the training simulations. We then applied the trained model to the empirical data to predict the α jump parameter consistent with the empirical observations. Specifically, the input vector was transformed into a feature map, and we then applied the trained coefficients/weights to these features to predict the associated α value. We generated a distribution of α by predicting with CKS performed on 100 bootstraps of the training data; from this distribution, we calculated a mean estimate of α along with the standard deviation.

3. Phylogenetic constraints to jump distances

Using our estimates of jump distances, we compared how phylogenetic constraints to host distribution, as described by jump parameter α, varied between pest type. As the distribution of jump distances was non-normal, we performed non-parametric Wilcoxon signed-rank tests on the mean α estimated for each pest species to examine whether insect and pathogenic (bacterial or fungal) pests differed in their α values, as well as to test for differences between native and non-native pests. We additionally bootstrapped tests on estimates sampled from the posterior distribution of jump parameters to account for uncertainty around estimated α values.

Finally, we used the observed host distribution of pests to estimate the probability of a jump to a new host, again taking the phylogenetic distance to observed hosts and calculating the negative exponent with the estimated jump parameter, α. We repeated these calculations with the upper and lower 95% confidence interval (percentile method) of the estimated parameters to calculate upper and lower 95% confidence intervals of the probability a potential host is the next to be colonized [38]. This produced for each tree genus the relative risk of being the target of a given pest upon its next jump, assuming all observed host genera of a given pest are equally likely to be the source of the jump. Summing these values across all pests provides a weighted probability of a given tree genus becoming host to a novel pest or pathogen.

4. Results

We estimated jump parameters for 161 pest species, with α varying from very low phylogenetic constraint ($\alpha = 0.14$, *Phytophthora cinnamomi*; a 50 million year jump is only 10% more likely than a 100 million year jump, figure 2) to high phylogenetic constraint ($\alpha = 4.63$, the California oak moth *Phryganidia californica*; a 50 million year jump is nearly 25 times more likely than a 100 million year jump, figure 2).

There was a positive relationship between the number of known hosts and the power of models to correctly predict parameters on the test data, revealing that none of our pests have reached a saturated host count at which information about

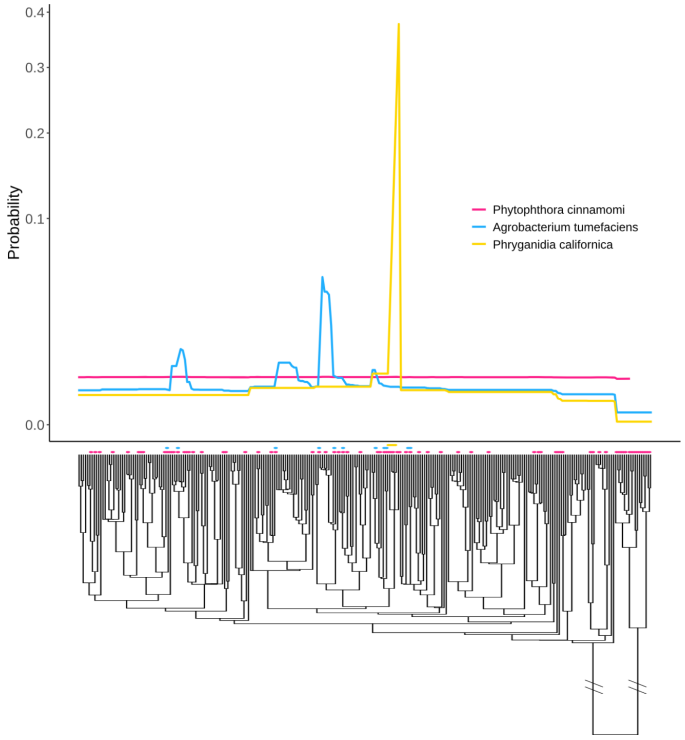

Figure 2. The distribution of jump risk across a phylogeny of candidate host tree genera (below) for pests with the minimum, median and maximum estimated α. Dots shown on the tips of the phylogeny represent observed hosts of the three species. *Phytophthora cinnamomi* (α = 0.14) has observed hosts across the plant kingdom and is estimated to have very low phylogenetic constraints, producing even risk across plant clades. *Agrobacterium tumefaciens* (α = 2.01) is observed on a handful of disparate angiosperms and is estimated to have intermediate phylogenetic constraint, increasing the risk for clades of plants related to its hosts. *Phryganidia californica* (α = 4.63) is observed on only a few closely related hosts and has the maximum estimated phylogenetic constraint in our dataset; it is unlikely to jump to any genera other than one closely related to its observed hosts. The branches separating angiosperms and gymnosperms have been truncated.

assembly process would begin to be lost (electronic supplementary material, figure S1). Correlating predicted values with known values, we achieved an overall model accuracy of $R^2 > 0.49$ for 94 pest species of which 20 were non-native, and $R^2 > 0.50$ for all cases in which the number of hosts was greater than or equal to 18. Model accuracy of $R^2 > 0.70$ (maximum $R^2 = 0.89$) was obtained for 36 species, of which 24 were pathogens, 12 were insects and seven were non-native, and in all cases for which pests had 72 or more hosts. Correlations were close to 1 (mean = 1.02, s.d. = 0.076), suggesting predictions were on average accurate but with variable precision (R^2 values). Models trained on simulations initiated with a random observed host produced results highly correlated with the base models (Pearson's $r = 0.80$), returned higher R^2 (average $R^2 = 0.77$) and tended to predict higher α values. In the main text, we report results from the base models in which the initial host in simulations was unconstrained, as we believe they are more conservative.

Among the 36 pests for which our model had good prediction accuracy ($R^2 > 0.70$), estimates for jump parameters had tight posterior distributions (s.d. < 0.40). Within this group, insects were more phylogenetically constrained than pathogens on average, but the difference was not statistically significant (mean insect $\alpha = 2.07$, mean pathogen $\alpha = 1.54$, Wilcoxon test $p = 0.072$, figure 3). Comparisons across all included pests, irrespective of prediction accuracy, produced qualitatively similar differences in means (mean insect $\alpha = 2.23$, mean pathogen $\alpha = 2.04$, Wilcoxon test $p = 0.34$; mean native $\alpha = 2.13$, mean non-native $\alpha = 2.05$, Wilcoxon test $p = 0.81$). Models trained on observed host-seeded simulations produced qualitatively similar patterns (electronic supplementary material, figure S2).

In within-pest type comparisons across all data, native insects were significantly more phylogenetically constrained than non-native insects (mean native $\alpha = 2.42$, mean non-native $\alpha = 1.84$, Wilcoxon test $p = 0.028$, figure 4A). We did not detect a significant difference between native and non-native pathogens (mean native $\alpha = 1.84$, mean non-native $\alpha = 1.98$, Wilcoxon test $p = 0.11$, figure 4B), although data on non-native pathogens were sparse. Bootstrapped tests on estimates incorporating uncertainty produced qualitatively similar results. Models trained on observed host-seeded simulations again produced qualitatively similar differences in mean estimates, but insect nativity was marginally non-significant while pest nativity became marginally significant (electronic supplementary material, figure S3).

Among all 161 pests, 16 (10%) had 95% confidence intervals overlapping zero and indistinguishable from no phylogenetic constraint, and 32 (20%) had confidence intervals overlapping one, consistent with diffusion across phylogenetic distance. One species (<1%) had confidence intervals between zero and one, while the remaining 112 (70%) had confidence intervals lying entirely above one, consistent with high phylogenetic constraint. These proportions were similar when we subset to the 36 pests with high prediction accuracy ($\alpha = 0$, $n = 4$ (11%), $\alpha = 1$, $n = 7$ (19%), $\alpha > 1$, $n = 25$ (70%)).

We identified several tree taxa with high risk of novel pest or pathogen infestation, i.e. likely target hosts of future pest jumps (figure 5, table 1), as measured by the product of the risks subtracted from one (the likelihood of not being targeted in any of the next jumps). The two plant genera with the highest summed risk were *Chrysolepis* (chinquapin) and *Heteromeles* (toyon). Many of the most probable specific jumps were from native pests to members of Pinaceae (table 1). Two of the most likely

Figure 3. Estimated α values for insect and pathogenic pests for which α was predictable with high confidence. Insect pest $n = 12$, pathogen pest $n = 24$. Coloured violin plots show the density of 100 bootstrap predictions of high confidence estimates. Grey violin plots show the distribution of all estimates, regardless of confidence.

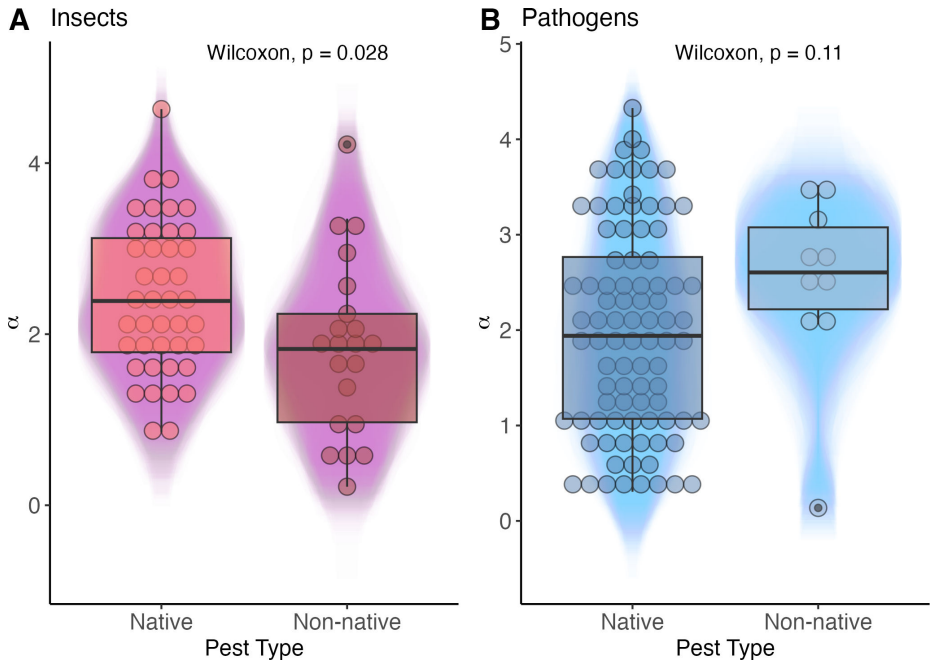

Figure 4. (A) Estimated α values for native and non-native insect pests. Native insect $n = 42$, non-native insect $n = 21$. Background violin plots show the density of 100 samples of the posterior distribution of α estimates. (B) Estimated α values for native and non-native pathogen pests. Native pathogen $n = 88$, non-native pathogen $n = 10$. Background violin plots show the density of 100 samples of the posterior distribution of α estimates.

jumps were to *Larix*, larch, by *Leptographium wageneri*, a fungal pathogen known to cause significant damage to pine forests, and *Arceuthobium tsugense*, a mistletoe species, both specialized on Pinaceae. The most probable jumps from non-native pests were from *Gremmeniella abietina* (a canker fungus) to *Tsuga*, *Tomicus piniperda* (the pine shoot beetle) to *Pseudotsuga*, *Fiorinia externa* (a scale insect) and *Phytophthora lateralis* (cedar root disease) to *Cupressus*. Host taxa of low risk of future jumps include *Quercus*, *Acer* and *Prunus*. We present the full list of probabilities in electronic supplementary material, table S1.

5. Discussion

We modelled host range expansion of tree pests as a stepwise jumping process with power function-distributed jump distances reflecting the strength of phylogenetic constraint describing pest host breadth. Using CKS, a light-weight machine learning tool [36], we were able to accurately recover the jump parameter that describes the distribution of a given pest across the host

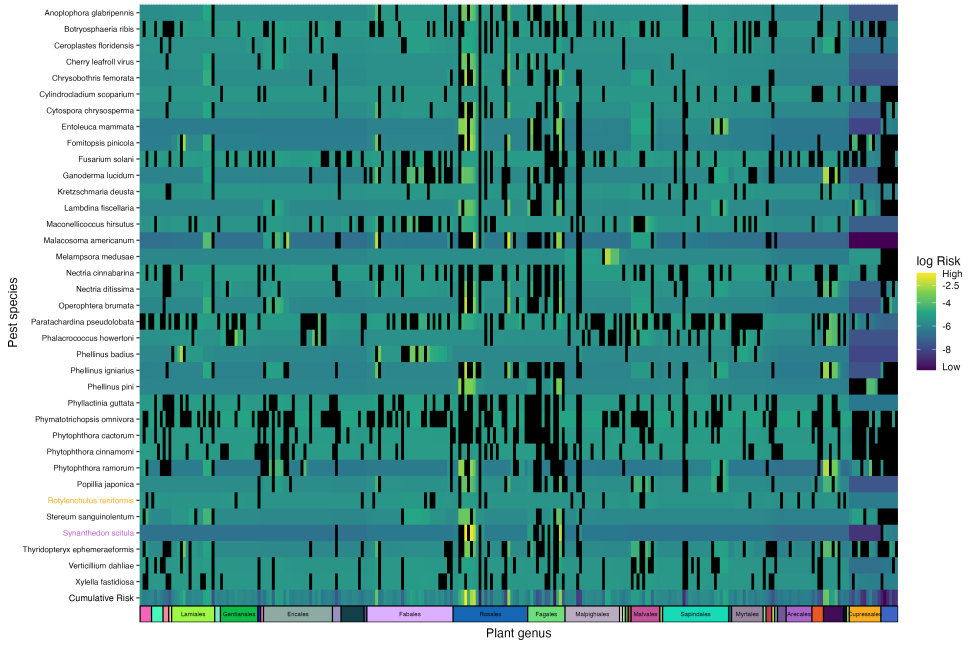

Figure 5. Heat map of the relative risk of tree pests (rows) jumping to North American tree genera (columns). Only pests with prediction accuracy $R^2 > 0.7$ are shown. Black cells represent existing pest–host associations. Relative cumulative risk and plant family are shown in the bottom two rows. Pests with high phylogenetic constraints have risk confined to few taxa, resulting in dark horizontal bands with a few clustered high-risk jumps, e.g. *Synanthedon scitula* (indicated in purple text). In contrast, pests with low phylogenetic constraint have even bands of intermediate risk, e.g. *Rotylenchulus reniformis* (indicated in orange text).

phylogeny. Estimated jump parameters loosely correlate with more traditional indices of phylogenetic conservatism (electronic supplementary material, figure S4), although the assumed underlying processes are entirely different. Evidence of phylogenetic signal in an observed trait is typically interpreted as reflecting evolutionary conservatism in trait change from ancestor to descendent [39]. In contrast, our jump parameter, α, assumes host expansion as a jumping process across the tips of the host phylogeny. Our model is analogous to a Lévy flight process in movement ecology, and we suggest a parallel in how pests may explore potential hosts across phylogenetic space.

Our approach extends previous work assuming some linear transform of phylogenetic distances between hosts (e.g. [13,15,21]), by explicitly estimating the probability of host jumps across varying phylogenetic distances. We show large interspecific variation in the distribution of jump distances across tree pests and pathogens. Our ability to precisely infer phylogenetic constraint improved with increasing observed host breadth, and thus we had more precise estimates for pests with greater host breadth, which are frequently of high concern [40]. We were also able to improve inference by constraining training data to simulations seeded with hosts from observed host ranges rather than random hosts, at the cost of making stronger assumptions about assembly processes. We found that native insects were more constrained than non-native insects, but we did not detect a similar difference in jump distances between native and non-native pathogens. While insect pests tend to be marginally more constrained than pathogens by host phylogeny overall, with a thinner tail of long pest jumps, differences were not significant and more work is necessary to determine if and in what contexts insects may exhibit greater phylogenetic constraint.

The tendency for insect pests to be characterized by shorter jump distances is consistent with observations that insect pests were also on average associated with fewer host genera than pathogens in our dataset. This pattern contrasts with previous suggestions that pathogens tend to be more specialized than insects [15] (but see [22]), but is in keeping with the rapid evolutionary capabilities of bacterial and viral pathogens, which could enable rapid spread across the host phylogeny. Insects may have greater apparent dispersal abilities than many pathogens, but wind- and vector-mediated dispersal distances of pathogens can also be high [41,42], and insect host range expansion may be constrained by behaviour [43,44], mechanical limitations [45,46], and resistance to defensive compounds [47], which might be comparatively slow to evolve due to the longer generation times of insects relative to pathogens. In contrast, plant pathogens are known to evolve rapidly in response to fungicides and antibiotics, analogues to natural plant defences, accessing novel resistance through horizontal gene transfer as well as mutation over large propagule numbers and short generation times [48,49].

(a) Jump parameters and phylogenetic constraint

We show that pests vary in phylogenetic constraint, but that the majority of pests have jump parameters greater than one, indicating significant phylogenetic constraint in host use among pests. Nonetheless, the intermediate jump parameters of many pests still allow for the possibility of jumps to more distant hosts, potentially because hosts have converged in some traits related to susceptibility, such as a loss of immune receptors [50]. Such long jumps may serve as a stepping stone that introduces many potential novel hosts into the window of more frequent short jump distances, facilitating rapid host expansion in a newly colonized clade [51,52]. A small percentage of pests have estimated jump parameters indistinguishable from zero, suggesting

Table 1. Highest specific relative risks of pests jumping to plant genera upon their next pest jump.

pest	type	nativity	plant genus	plant family	risk	lower 95	upper 95
Leptographium wageneri	pathogen	native	*Larix*	Pinaceae	0.967	0.931	0.985
Arceuthobium tsugense	pathogen	native	*Larix*	Pinaceae	0.902	0.732	0.968
Apiosporopsis carpinea	pathogen	native	*Chrysolepis*	Fagaceae	0.856	0.802	0.896
Erwinia cacticida	pathogen	native	*Harrisia*	Cactaceae	0.832	0.422	0.972
Asperisporium sequoiae	pathogen	native	*Calocedrus*	Cupressaceae	0.832	0.771	0.875
Pseudohylesinus sericeus	insect	native	*Larix*	Pinaceae	0.823	0.749	0.873
Venturia inaequalis	pathogen	native	*Vauquelinia*	Rosaceae	0.822	0.584	0.937
Choristoneura pinus	insect	native	*Pseudotsuga*	Pinaceae	0.795	0.758	0.824
Erwinia amylovora	pathogen	native	*Heteromeles*	Rosaceae	0.765	0.613	0.829
Melanophila fulvoguttata	insect	native	*Pseudotsuga*	Pinaceae	0.744	0.669	0.795
Gremmeniella abietina	pathogen	non-native	*Tsuga*	Pinaceae	0.729	0.585	0.835
Candidatus Phytoplasma palmae	pathogen	native	*Serenoa*	Arecaceae	0.717	0.492	0.865
Dendroctonus ponderosae	insect	native	*Tsuga*	Pinaceae	0.698	0.568	0.8
Tomicus piniperda	insect	non-native	*Pseudotsuga*	Pinaceae	0.648	0.554	0.716
Melanophila drummondi	insect	native	*Pinus*	Pinaceae	0.648	0.492	0.772
Dendroctonus pseudotsugae	insect	native	*Abies*	Pinaceae	0.587	0.545	0.622
Fiorinia externa	insect	non-native	*Cupressus*	Cupressaceae	0.561	0.368	0.679
Phytophthora lateralis	pathogen	non-native	*Cupressus*	Cupressaceae	0.509	0.442	0.572
Choristoneura freemani	insect	native	*Tsuga*	Pinaceae	0.457	0.341	0.546
Melampsoridium betulinum	pathogen	native	*Carpinus*	Betulaceae	0.453	0.41	0.492
Phellinus weirii	pathogen	native	*Cupressus*	Cupressaceae	0.43	0.407	0.446
Stigmina palmivora	pathogen	native	*Coccothrinax*	Arecaceae	0.422	0.361	0.477
Dendroctonus rufipennis	insect	native	*Larix*	Pinaceae	0.411	0.362	0.444
Dendroctonus rufipennis	insect	native	*Pseudotsuga*	Pinaceae	0.411	0.362	0.444
Graphiloa phoenicis	pathogen	native	*Thrinax*	Arecaceae	0.388	0.337	0.437
Phryganidia californica	insect	native	*Fagus*	Fagaceae	0.377	0.281	0.484
Apple mosaic virus	pathogen	native	*Heteromeles*	Rosaceae	0.363	0.293	0.413
Oberea myops	insect	native	*Lyonia*	Ericaceae	0.335	0.301	0.361
Dendroctonus pseudotsugae	insect	native	*Picea*	Pinaceae	0.324	0.322	0.319
Seiridium cardinale	pathogen	native	*Taxodium*	Cupressaceae	0.317	0.297	0.333
Apple mosaic virus	pathogen	native	*Amelanchier*	Rosaceae	0.308	0.254	0.344
Oberea myops	insect	native	*Elliottia*	Ericaceae	0.295	0.269	0.315

that phylogeny plays little to no role in their patterns of host range expansion. When pests exhibit a lack of phylogenetic constraint, they may have generalizable exploitation strategies that work on a broad range of host phenotypes [53], or they may be exploiting a consistent vulnerability or susceptibility shared across a broad range of hosts [50]. A propensity to long jumps may be important for several reasons. First, pests less constrained by host phylogeny represent a greater risk of emerging on novel hosts when introduced to new regions. Second, novel hosts evolutionarily distant from current hosts may be evolutionarily naive, lacking effective pest defence mechanisms, and thus at greater risk of pest hypervirulence [25,54]. If this increased virulence is not compensated by decreased transmission, a common but not universal trade-off [55], generalists with low phylogenetic constraint may pose significant threats.

(b) Native and non-native pests

We assume a model in which pests and pathogens traverse the host phylogeny in a series of stepwise jumps. However, the distribution of parasites across hosts might also reflect a process of co-diversification and vertical inheritance of parasites from ancestral hosts (see [13]). We explore whether these two alternative scenarios might leave distinct imprints on the pattern of pest distributions on the phylogeny of trees by contrasting phylogenetic constraint exponents between native and non-native pests, the former of which are more likely to have host distributions produced by vertical transmission. We find that native insects are more phylogenetically constrained than non-native insects. It is possible that this difference might capture the longer history of vertical inheritance for native versus non-native insect pests, producing higher levels of phylogenetic constraint in the former.

However, it is not obvious why we do not detect a similar pattern for fungal and bacterial pathogens. For these pest types, either the two processes (vertical and horizontal transmission) may result in similar patterns of phylogenetic distribution (e.g. [56]), or pest jumps predominantly explain the contemporary distribution of both native and non-native tree pathogens on their hosts.

(c) Phylogenetic risk assessment

Using the modelled distribution of pest jumps for both pests and pathogens, we generated a relative risk assessment across the phylogeny of North American tree genera. We estimated some genera to be at much higher risk of novel infestation than others, both cumulatively and from specific pests. For example, *Malus*, the genus containing apples, is an important agricultural crop with well-documented pests. Consequently, its close relatives in Amygdaloideae are the likely recipients of pest jumps from its phylogenetically constrained pests. Some potential novel hosts, such as *Amelanchier*, the saskatoon berry, are cultivated as commercial crops, suggesting important implications for phytosanitary management. It may be desirable, for example, to grow *Malus* apart from other genera in Amygdaloideae or to ensure thorough quarantining procedures before planting these crops together to avoid undesirable spillover events. In contrast, some genera are at consistently low risk. Members of Magnoliidae, especially *Magnolia* and *Liriodendron*, the two genera within Magnoliaceae, are at broadly low risk due to their phylogenetic distance from the diverse Eudicots.

Nonetheless, we caution that phylogenetic risk should not be considered in isolation, as the probability of jumps may be taxonomically biased by host or pest biology, or altered by abundance, geography, anthropogenic activity or other processes. Taxonomic bias may occur when a host clade is predisposed for or against infection. For example, our model predicts larch to be at high spillover risk from *Leptographium wageneri* (black stain fungus), but larch is known to be immune or resistant to this fungus [57], so increased monitoring may not provide much value, barring a shift in pathogenicity or some ecological surprise. When clades within the phylogenetic breadth of host range share such resistances, pest distributions could exhibit whole-clade gaps, though this would not be likely to much influence estimates of jump parameters unless pests have few hosts separated by a large immune clade. On the other hand, predisposition to susceptibility could lead to rapid spread of a pest in a clade; this could bias jump parameters upward as the pest colonizes the clade, but in general, phylogenetic predisposition is a pattern our model is intended to capture.

Pests' probability of spread may also be biased by biological properties linked to their taxonomy, e.g. mode of dispersal (e.g. active, passive or vector-borne) or preadaptations that make them particularly adept at exploiting some host clade. Two pests carried by the same vector, for example, would be likely to reach the same hosts. Pest taxonomic bias would produce phylogenetic signal in the identity of hosts' pests but would not affect estimates of pest jump parameters.

6. Conclusion

We have here presented a novel model of stepwise pest jump over a phylogeny and used it to generate risk predictions of future pest spread, with important applications to agriculture, forestry and natural ecosystems. We suggest that our measure describing the phylogenetic constraints on pest host breadth could also be used to better inform our understanding of the evolution of pest specialism versus generalism and may serve as a useful metric of phylogenetic conservatism that could be applied across the fields of agricultural, conservation or evolution when considering patterns of species interactions shifting over a phylogeny. Variation in phylogenetic constraint may be determined by macroevolutionary differences in host traits across clades or differences in the signature of phylogenetic conservatism among different sets of traits for a given set of hosts; the extent to which each determines pest constraints is worth exploring.

While an understanding of the phylogenetic constraints of pests is critical for modelling host range expansion, it is just one of many factors that shape pest distributions. Improving our understanding of the history of host range expansion (including timing and environmental mediators of pest jumps and severity of novel infestations) would enable us to make better predictions about both where and when jumps will occur. Models could also be improved by including spatial and geographic overlap, which would allow predictions to account for opportunity. Such information would allow us to differentially weight potential source hosts rather than assuming jumps from any host are equally likely; currently, estimates of constraint may be biased upward by the non-random, often clustered phylogenetic distribution of co-occurring trees [58]. Geographically explicit models could be implemented by adjusting weights in the adjacency matrix. For example, adjacency values could simply be set to zero when there is no geographic overlap or scaled to reflect the relative percentage of range or population overlap between host species.

We leveraged a remarkable dataset on tree pests and pathogens in North America. Inevitably, our analysis has some limitations. It is likely that the host–pest association matrix remains highly incomplete—some plant taxa and pests attract more study interest, are easier to detect, etc. [59]. Nonetheless, we believe the patterns we report are generally robust, as indicated by the convergence in jump estimates observed across native and non-native pests and insensitivity to weighting estimates by our level of confidence in them.

Ethics. This work did not require ethical approval from a human subject or animal welfare committee.

Data accessibility. Code for reproducible analysis is archived on Dryad [60]. Data are derived from [33] and [22].

Supplementary material is available online [61].

Declaration of AI use. We have not used AI-assisted technologies in creating this article.

Authors' contributions. A.K.: formal analysis, investigation, methodology, software, validation, visualization, writing—original draft, writing—review and editing; A.G.: data curation, resources, writing—review and editing; G.L.: conceptualization, investigation, methodology, writing—review and editing; J.D.: conceptualization, supervision, writing—review and editing.

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Conflict of interest declaration. We declare we have no competing interests.

Funding. This project was made possible by an NSERC Discovery Grant awarded to T.J.D.

Acknowledgements. Thanks to Maxwell Farrell for participation in early discussions on the project.

References

1. Decaestecker E, Gaba S, Raeymaekers JAM, Stoks R, Van Kerckhoven L, Ebert D, De Meester L. 2007 Host–parasite 'Red Queen' dynamics archived in pond sediment. *Nature* **450**, 870–873. (doi:10.1038/nature06291)
2. Petermann JS, Fergus AJF, Turnbull LA, Schmid B. 2008 Janzen–Connell effects are widespread and strong enough to maintain diversity in grasslands. *Ecology* **89**, 2399–2406. (doi:10.1890/07-2056.1)
3. Bradshaw CJA *et al*. 2016 Massive yet grossly underestimated global costs of invasive insects. *Nat. Commun.* **7**, 12986. (doi:10.1038/ncomms12986)
4. Murray KA, Preston N, Allen T, Zambrana-Torrelio C, Hosseini PR, Daszak P. 2015 Global biogeography of human infectious diseases. *Proc. Natl Acad. Sci. USA* **112**, 12746–12751. (doi:10.1073/pnas.1507442112)
5. Smith KM, Machalaba CC, Seifman R, Feferholtz Y, Karesh WB. 2019 Infectious disease and economics: the case for considering multi-sectoral impacts. *One Health* **7**, 100080. (doi:10.1016/j.onehlt.2018.100080)
6. Kovacs KF, Mercader RJ, Haight RG, Siegert NW, McCullough DG, Liebhold AM. 2011 The influence of satellite populations of emerald ash borer on projected economic costs in U.S. communities, 2010–2020. *J. Environ. Manag.* **92**, 2170–2181. (doi:10.1016/j.jenvman.2011.03.043)
7. Liebhold AM, Kean JM. 2019 Eradication and containment of non-native forest insects: successes and failures. *J. Pest Sci.* **92**, 83–91. (doi:10.1007/s10340-018-1056-z)
8. Brockerhoff EG, Liebhold AM, Richardson B, Suckling DM. 2010 Eradication of invasive forest insects: concepts, methods, costs and beneits. *N. Z. J. For. Sci* **40 (Suppl.)**, S117–S135.
9. Haack RA *et al*. 2014 Effectiveness of the international phytosanitary standard ISPM no. 15 on reducing wood borer infestation rates in wood packaging material entering the United States. *PLoS One* **9**, e96611. (doi:10.1371/journal.pone.0096611)
10. Anton S, Cortesero AM. 2022 Plasticity in chemical host plant recognition in herbivorous insects and its implication for pest control. *Biology* **11**, 1842. (doi:10.3390/biology11121842)
11. Morris CE, Moury B. 2019 Revisiting the concept of host range of plant pathogens. *Annu. Rev. Phytopathol.* **57**, 63–90. (doi:10.1146/annurev-phyto-082718-100034)
12. Thines M. 2019 An evolutionary framework for host shifts—jumping ships for survival. *New Phytol.* **224**, 605–617. (doi:10.1111/nph.16092)
13. Cooper N, Griffin R, Franz M, Omotayo M, Nunn CL. 2012 Phylogenetic host specificity and understanding parasite sharing in primates. *Ecol. Lett.* **15**, 1370–1377. (doi:10.1111/j.1461-0248.2012.01858.x)
14. Gougherty AV, Davies TJ. 2022 Host phylogenetic diversity predicts the global extent and composition of tree pests. *Ecol. Lett.* **25**, 101–112. (doi:10.1111/ele.13908)
15. Gilbert GS, Magarey R, Suiter K, Webb CO. 2012 Evolutionary tools for phytosanitary risk analysis: phylogenetic signal as a predictor of host range of plant pests and pathogens. *Evol. Appl.* **5**, 869–878. (doi:10.1111/j.1752-4571.2012.00265.x)
16. Gougherty AV, Davies TJ. 2021 Towards a phylogenetic ecology of plant pests and pathogens. *Phil. Trans. R. Soc. B* **376**, 20200359. (doi:10.1098/rstb.2020.0359)
17. Gougherty AV, Davies TJ. 2022 A global analysis of tree pests and emerging pest threats. *Proc. Natl Acad. Sci. USA* **119**, e2113298119. (doi:10.1073/pnas.2113298119)
18. Mech AM *et al*. 2019 Evolutionary history predicts high-impact invasions by herbivorous insects. *Ecol. Evol.* **9**, 12216–12230. (doi:10.1002/ece3.5709)
19. Pedersen AB, Davies TJ. 2009 Cross-species pathogen transmission and disease emergence in primates. *Ecohealth* **6**, 496–508. (doi:10.1007/s10393-010-0284-3)
20. Uden DR *et al*. 2023 Phylogenetic risk assessment is robust for forecasting the impact of European insects on North American conifers. *Ecol. Appl.* **33**, e2761. (doi:10.1002/eap.2761)
21. Gilbert GS, Webb CO. 2007 Phylogenetic signal in plant pathogen–host range. *Proc. Natl Acad. Sci. USA* **104**, 4979–4983. (doi:10.1073/pnas.0607968104)
22. Wang SS, Gougherty AV, Davies TJ. 2022 Non-native tree pests have a broader host range than native pests and differentially impact host lineages. *J. Ecol.* **110**, 2898–2910. (doi:10.1111/1365-2745.13995)
23. Park AW *et al*. 2018 Characterizing the phylogenetic specialism–generalism spectrum of mammal parasites. *Proc. R. Soc. B* **285**, 20172613. (doi:10.1098/rspb.2017.2613)
24. Ouvrard D, Chalise P, Percy DM. 2015 Host-plant leaps versus host-plant shuffle: a global survey reveals contrasting patterns in an oligophagous insect group (Hemiptera, Psylloidea). *Syst. Biodivers.* **13**, 434–454. (doi:10.1080/14772000.2015.1046969)
25. Farrell MJ, Davies TJ. 2019 Disease mortality in domesticated animals is predicted by host evolutionary relationships. *Proc. Natl Acad. Sci. USA* **116**, 7911–7915. (doi:10.1073/pnas.1817323116)
26. Elmasri M, Farrell MJ, Davies TJ, Stephens DA. 2020 A hierarchical Bayesian model for predicting ecological interactions using scaled evolutionary relationships. *Ann. Appl. Stat.* **14** 221–240. (doi:10.1214/19-aoas1296)
27. Farrell MJ, Elmasri M, Stephens DA, Davies TJ. 2022 Predicting missing links in global host–parasite networks. *J. Anim. Ecol.* **91**, 715–726. (doi:10.1111/1365-2656.13666)
28. Werner CS, Kasan K, Geyer JK, Elmasri M, Farrell MJ, Nunn CL. 2023 Using phylogeographic link-prediction in primates to prioritize human parasite screening. *Am. J. Biol. Anthropol.* **182**, 583–594. (doi:10.1002/ajpa.24604)
29. Codling EA, Plank MJ, Benhamou S. 2008 Random walk models in biology. *J. R. Soc. Interface* **5**, 813–834. (doi:10.1098/rsif.2008.0014)
30. Lambiotte R, Sinatra R, Delvenne JC, Evans TS, Barahona M, Latora V. 2011 Flow graphs: interweaving dynamics and structure. *Phys. Rev. E* **84**, 017102. (doi:10.1103/PhysRevE.84.017102)
31. Riascos AP, Mateos JL. 2012 Long-range navigation on complex networks using Lévy random walks. *Phys. Rev. E* **86**, 056110. (doi:10.1103/physreve.86.056110)
32. Oliveira FA, Ferreira RMS, Lapas LC, Vainstein MH. 2019 Anomalous diffusion: a basic mechanism for the evolution of inhomogeneous systems. *Front. Phys* **7**, 1–17. (doi:10.3389/fphy.2019.00018)
33. Potter KM, Escanferla ME, Jetton RM, Man G. 2019 Important insect and disease threats to United States tree species and geographic patterns of their potential impacts. *Forests* **10**, 304. (doi:10.3390/f10040304)
34. Jin Y, Qian H. 2019 V.phylomaker: an R package that can generate very large phylogenies for vascular plants. *Ecography* **42**, 1353–1359. (doi:10.1111/ecog.04434)

35. Rahimi A, Recht B. 2008 Weighted sums of random kitchen sinks: replacing minimization with randomization in learning. In *Advances in neural information processing systems* (eds D Koller, D Schuurmans, Y Bengio, L Bottou). Red Hook, NY: Curran Associates, Inc.
36. Kruger A, Shankar V, Jonathan Davies T. 2024 Inferring the evolutionary model of community-structuring traits with convolutional kitchen sinks. *Syst. Biol.* **73**, syae026. (doi:10.1093/sysbio/syae026)
37. Gómez-Gardeñes J, Latora V. 2008 Entropy rate of diffusion processes on complex networks. *Phys. Rev. E* **78**, 065102. (doi:10.1103/physreve.78.065102)
38. Davison AC, Hinkley DV. 1997 *Bootstrap methods and their application*. Cambridge, UK: Cambridge University Press.
39. Blomberg SP, Garland T, Ives AR. 2003 Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution* **57**, 717–745. (doi:10.1111/j.0014-3820.2003.tb00285.x)
40. Olival KJ, Hosseini PR, Zambrana-Torrelio C, Ross N, Bogich TL, Daszak P. 2017 Host and viral traits predict zoonotic spillover from mammals. *Nature* **546**, 646–650. (doi:10.1038/nature22975)
41. Brown JKM, Hovmøller MS. 2002 Aerial dispersal of pathogens on the global and continental scales and its impact on plant disease. *Science* **297**, 537–541. (doi:10.1126/science.1072678)
42. Golan JJ, Pringle A. 2017 Long-distance dispersal of fungi. *Microbiol. Spectr.* **5**, 1–24. (doi:10.1128/microbiolspec.funk-0047-2016)
43. Forister ML, Dyer LA, Singer MS, Stireman JO, Lill JT. 2012 Revisiting the evolution of ecological specialization, with emphasis on insect–plant interactions. *Ecology* **93**, 981–991. (doi:10.1890/11-0650.1)
44. Futuyma DJ, Moreno G. 1988 The evolution of ecological specialization. *Annu. Rev. Ecol. Syst*. **19**, 207–233.
45. Salerno G, Rebora M, Gorb E, Gorb S. 2024 Mechanoecology: biomechanical aspects of insect–plant interactions. *J. Comp. Physiol. A* **210**, 249–265. (doi:10.1007/s00359-024-01698-2)
46. Whitney HM, Federle W. 2013 Biomechanics of plant–insect interactions, growth and development. *Curr. Opin. Plant Biol.* **16**, 105–111. (doi:10.1016/j.pbi.2012.11.008)
47. Harborne JB. 2001 Twenty-five years of chemical ecology. *Nat. Prod. Rep*. **18**, 361–379. (doi:10.1039/b005311m)
48. Lucas JA, Hawkins NJ, Fraaije BA. 2015 Chapter two—the evolution of fungicide resistance. In *Advances in applied microbiology* (eds S Sariaslani, GM Gadd), pp. 29–92, vol. **90**. Cambridge, MA: Academic Press. (doi:10.1016/bs.aambs.2014.09.001)
49. Sundin GW, Wang N. 2018 Antibiotic resistance in plant-pathogenic bacteria. *Annu. Rev. Phytopathol*. **56**, 161–180. (doi:10.1146/annurev-phyto-080417-045946)
50. Li SX, Liu Y, Zhang YM, Chen JQ, Shao ZQ. 2025 Convergent reduction of immune receptor repertoires during plant adaptation to diverse special lifestyles and habitats. *Nat. Plants* **11**, 248–262. (doi:10.1038/s41477-024-01901-x)
51. Braga MP, Razzolini E, Boeger WA. 2015 Drivers of parasite sharing among Neotropical freshwater fishes. *J. Anim. Ecol.* **84**, 487–497. (doi:10.1111/1365-2656.12298)
52. Hoberg E, Boeger W, Brooks D, Trivellone V, Agosta S. 2022 Stepping-stones and mediators of pandemic expansion—a context for humans as ecological super-spreaders. *Manter (Linc)* 1–12. (doi:10.32873/unl.dc.manter18)
53. Joy JB. 2013 Symbiosis catalyses niche expansion and diversification. *Proc. R. Soc. B* **280**, 20122820. (doi:10.1098/rspb.2012.2820)
54. Guth S, Visher E, Boots M, Brook CE. 2019 Host phylogenetic distance drives trends in virus virulence and transmissibility across the animal–human interface. *Phil. Trans. R. Soc. B* **374**, 20190296. (doi:10.1098/rstb.2019.0296)
55. Alizon S, Hurford A, Mideo N, Van baalen M. 2009 Virulence evolution and the trade-off hypothesis: history, current state of affairs and the future. *J. Evol. Biol*. **22**, 245–259. (doi:10.1111/j.1420-9101.2008.01658.x)
56. Mazel F, Davis KM, Loudon A, Kwong WK, Groussin M, Parfrey LW. 2018 Is host filtering the main driver of phylosymbiosis across the tree of life? *mSystems* **3**, e00097-18. (doi:10.1128/mSystems.00097-18)
57. Hessburg PF, Goheen DJ, Bega RV. 1995 *Black stain root disease of conifers (no. 145548 (revised)*. Washington, DC: Forest Insect & Disease Leaflet. USDA Forest Service.
58. Potter KM, Koch FH. 2014 Patterns of forest phylogenetic community structure across the United States and their possible forest health implications. *For. Sci*. **60**, 851–861. (doi:10.5849/forsci.13-115)
59. Gougherty AV, Davies TJ. 2022 Most countries are vulnerable to novel pest invasions and under-report the diversity of tree pests. *Glob. Ecol. Biogeogr.* **31**, 2314–2322. (doi:10.1111/geb.13580)
60. Kruger A, Gougherty A, Legault G, Davies TJ. 2025 Data and code from: Pest host expansion as a scale-free stepwise process across the host phylogeny. Dryad Digital Repository. (doi:10.5061/dryad.m0cfxppdd)
61. Kruger A, Gougherty A, Legault G, Davies J. 2025 Supplementary material from: Pest host expansion as a scale-free stepwise process across the host phylogeny. Figshare. (doi:10.6084/m9.figshare.c.8126260)