CellPress

Trends in Parasitology

Forum

Building a comprehensive phylogenetic framework in disease ecology

Antoine Filion [1,*], Jean-François Doherty [1], Robert Poulin [1], and Stephanie S. Godfrey [1]

Disease spillover can have dramatic consequences in multispecies systems, potentially leading to the emergence of zoonoses. To better understand disease emergence patterns, an approach encompassing species relatedness metrics is needed. We show that integrating phylogenetic information in disease ecology is still lagging, and we highlight potential solutions to solve this problem.

Diseases in multihost systems

Diseases act as a major regulating force in community ecology. By regulating interactions between individuals and their environment at multiple scales, and influencing density-dependent population processes, diseases help in preserving ecosystem health [1]. However, the same diseases that promote healthy ecosystems can, in some conditions, become fatal to naïve populations that have no evolutionary defenses against them [1]. With anthropogenic changes constantly modifying the landscape and altering species niche breadth, many species that were once restricted to specific areas are now expanding their range, rapidly shifting community composition [2]. Moreover, accidental or planned species translocation from one system to another can further disturb local species pools, enhancing the risk of pathogen spillover toward naïve populations [3]. Here, we define disease spillover as a pathogen being transmitted from one reservoir host population to another population. This phenomenon can have drastic impacts on newly exposed populations, especially for naïve species that did not coevolve with the pathogen. With no coevolutionary history in new host–pathogen associations, high virulence is likely in new hosts, potentially leading to local extinction of native wildlife [4]. In addition, infectious disease spillover can have dramatic impacts in a multihost system. Indeed, generalist pathogens may follow different evolutionary paths, shaped by phylogenetic relatedness among hosts and ultimately determining the risk of pathogen spillover to other species, in some cases humans [5]. Therefore, integrating metrics of species relatedness, defined here as phylogenetic information, in studies of disease emergence patterns should be of prime importance to disease ecologists. Here, we highlight this crucial gap in study design with a showcase example in avian diseases, discuss potential adverse effects of the omission of phylogenetic information in disease ecology, and offer solutions to establish an integrated phylogenetic framework in this field of research.

Including phylogenetic relatedness in study design

Because closely related species tend to have more similar niches through phylogenetic conservatism than distantly related species [6], disease spillover is not equiprobable among all species in a community. For instance, all else being equal, closely related host species have a higher probability of sharing the same pathogens [7]. This is due to two main processes. First, closely related host species might inherit their pathogens from a common ancestor, in turn shaping their own pathogen communities [8]. Second, similar pathogen communities might be the result of closely related host species sharing the same life history traits and immunological response, allowing their use by the same pathogens [9]. In the case of nonrelated host species, regular host shifts might also be possible due to geographic closeness, enhancing contacts between species [10]. Accordingly, integrating metrics of species relatedness to predict host shifts should be one of the key aspects considered when studying disease emergence in wildlife.

One of the most widely used metrics of species relatedness is the phylogenetic signal, that is, the likelihood that related species resemble each other for a particular trait more than a random subset of species in a phylogenetic tree [7]. In recent decades, many researchers have raised red flags about the omission of phylogeny in conceptual or methodological frameworks, loudly calling for phylogeny to be considered in studies of multispecies systems [7,11]. Indeed, phylogenetic constraints might explain pathogen niche breath, and failure to include phylogeny in analysis can lead to type I error, that is, observing an effect when there actually is none. For example, one of the best-described relationships in host–parasite ecology, that between host body mass and parasite size, often disappears when accounting for phylogenetic relatedness among species in statistical analyses [12]. Moreover, shared traits between similar pathogen species, such as virulence or host range, can be crucial in understanding large-scale patterns of diseases [7]. Ultimately, moving toward a complete integration of phylogeny in disease ecology should be a seamless process as many research questions should not be answered without the inclusion of phylogenetic information.

Integration of phylogenetic information in disease ecology

To assess the rate of integration of phylogenetic information in disease ecology, we examined the use of phylogenetic information in multihost, multiparasite systems over the past 30 years. We searched for

relevant articles in Web of Science (all databases available) to quantify temporal trends in the integration of host phylogenetic information within all studies of avian diseases in the wild. We focused on avian diseases as an example since the phylogenetic tree for birds is currently one of the best resolved and readily accessible[i], thus making it easy to integrate in avian studies. A total of 876 publications matched our search criteria. We then excluded years for which we had fewer than five publications as they could bias proportion estimates. This effectively reduced our sampling timeframe from January 2011 to September 2021 inclusively, with a total of 144 studies relevant to our criteria. For detailed methods see Box S1 in the supplemental information online. Both the number of articles published per year and the proportion of articles that include phylogenetic information were plotted as a function of year of publication (Figure 1). Overall, we found a positive temporal trend in the proportion of articles including phylogenetic information in their analysis, with more frequent inclusion of phylogeny in data analysis over time. Nevertheless, the proportion of publications taking into account phylogeny in their analysis remains relatively low, especially when considering the ease of access of well-resolved phylogenies including most avian species, demonstrating that there is still much work to do to incorporate phylogenetic information in disease ecology. In the next section we present potential roadmaps to integrate phylogeny into ecological analyses.

Potential solutions

Many online tools can help disease ecologists to access a vast array of phylogenetic information. For instance, fully resolved phylogenies of entire clades, especially birds, can now be downloaded and pruned at will with barely more effort than a mouse click[i,ii,iii].

When online resources are unavailable, useful programs exist to build consensus trees based on species/lineages data, most of them being part of an expanding community, allowing for constant support and improvement for end users. For instance, the 'BEAST' program allows us to build rooted phylogenetic trees with great accuracy, which come with many adequate tutorials to 'tame the BEAST'[iv]; end users can now build a phylogenetic tree based on their own sequences, thus improving the integration of phylogenetic information in disease ecology. In addition, useful R packages such as *pastis*[v] also allow building rooted phylogenetic trees while other R packages such as *phytools*[vi], *ape*[vii] and *ggtree*[viii] are extremely useful to rapidly obtain phylogenetic metrics, such as Pagel's lambda, or to quickly prune and visualize life history traits spread over any given phylogeny, providing users with a solid framework to work with their own phylogeny.

One of the biggest improvements toward the integration of phylogeny in disease ecology lies with the *brms* package[ix]. By allowing users to directly incorporate a phylogenetic covariance matrix with corresponding host/pathogen species as a group-level effect, researchers who are not familiar with more complex Bayesian analyses can now integrate phylogeny in their designs with the help of friendly tutorials[x]. In addition, no prior programing experience is needed, facilitating a smooth transition toward an integrated phylogenetic framework.

Working with multispecies systems can be challenging. In this light, eco-phylogenetic methods provide a suitable framework to integrate phylogenetic information in community disease ecology. For instance, common α-diversity metrics such as Faith's Phylo-Diversity (Faith's PD), Mean Phylogenetic Distance (MPD), and Mean Nearest Taxonomic Distance (MNTD) all provide useful phylogenetic information about communities (see [13] for details).

Thus, with a vast array of solutions suitable to most systems, researchers can move forward to integrate phylogeny in disease ecology.

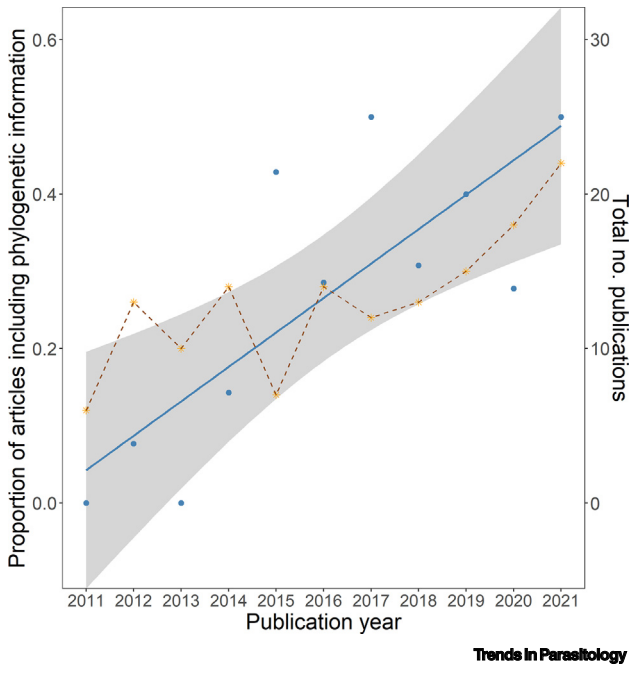

Figure 1. Temporal change in integration of phylogenetic information in community studies of avian diseases. Blue points represent the annual proportions of publications that account for host phylogeny in their analysis; the gray area represents the 95% confidence bands of the trend line. Orange stars and broken line represent the number of publications associated with each datapoint. Data were compiled from January 2011 to September 2021 inclusively. See Box S1 for detailed methods.

Future directions

To successfully establish an integrated phylogenetic framework in disease ecology we strongly believe that a change of perspective in data analysis is needed. As recommended by the American Statistical Association [14], we promote a shift from more conservative analyses based on *P* values toward statistics allowing for the integration of sampling frameworks and complex interactions. Ultimately, complexity in data analysis should not be a deterrent but a motivation to adopt an integrated approach to produce the most accurate representation of natural phenomena.

In conclusion, we show that the integration of phylogeny in disease ecology remains a slow process. With most studies still not incorporating any phylogenetic information in their conceptual framework, one can only ponder about the subsequent cascading effects that biased variable estimates could have on our broader understanding of disease ecology (Box 1). This may even lead to erroneous conclusions being derived from meta-analyses. To better predict disease emergence patterns in the wild, and with the potential for many diseases to become zoonotic, we posit that community-level studies in disease ecology should be at the forefront of the inclusion of phylogenetic information into their conceptual framework. With recent reviews promoting the establishment of an integrated phylogenetic framework when studying disease patterns ([13,15]), and with decades of warning about the potential consequences of ignoring phylogeny ([7,11]), we strongly believe that it is time for a change in the conceptual and analytical philosophy for multispecies systems in disease ecology.

Box 1. Including phylogenetic information in tests of disease drivers

Here, we provide a showcase example of three potential approaches to deal with multispecies datasets in disease ecology. We base our example on previously published data[xi] on avian malaria. We used local temperature seasonality as our main population-level (fixed effect) predictor of malaria (*Plasmodium* spp.) infection in birds sampled across the globe; for each host individual, data were available on infection status (infected or not), the species it belonged to, and where it was sampled. We considered three data-modelling scenarios. Firstly, we made a simple model using only sampling locality as a group-level (random) effect. Second, we used a more conservative approach to control for phylogenetic effects by including bird species as a group-level effect. Finally, we used a full species covariance matrix in a multilevel phylogenetic model, that is, full information on the phylogenetic relatedness among species, based on a global molecular phylogeny of birds, was integrated in the analysis. We show that including phylogeny in data analysis can severely reduce the estimated effect of the main predictor variable. Furthermore, we show that a more simplistic approach to control for phylogeny, that is, including species as a group-level effect, might not be a suitable approach to control for phylogenetic relatedness (Figure I).

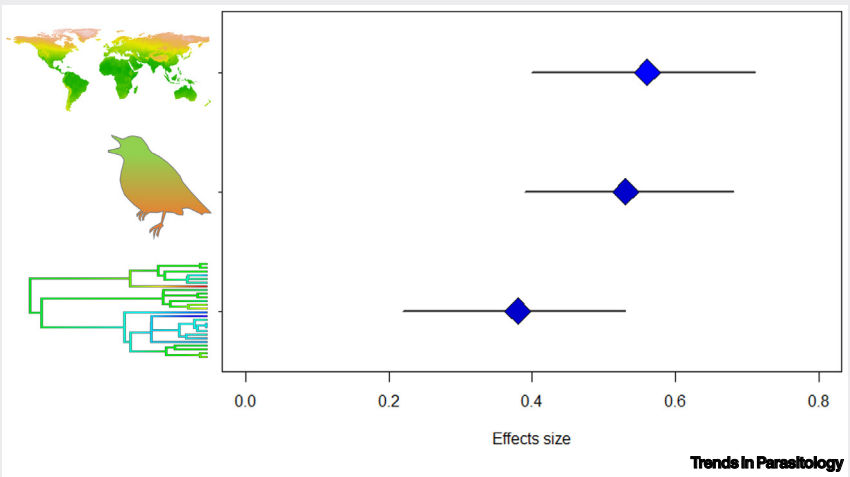

Figure I. Effect size of local temperature seasonality on avian malaria prevalence under three data analysis scenarios. From top to bottom: (i) controlling for geographic location, (ii) controlling for geographic location and species identity, and (iii) controlling for geographic location and species relatedness using a phylogenetic covariance matrix.

Acknowledgment

The authors would like to thank two anonymous reviewers for their comments and suggestions that improved the manuscript. A. Filion is supported by a University of Otago Doctoral Scholarship grant.

Declaration of interests

The authors declare no competing interests.

Supplemental information

Supplemental information associated with this article can be found online at https://doi.org/10.1016/j.pt.2022.01.008.

Resources

[i]https://birdtree.org/downloads/

[ii]www.timetree.org

[iii]www.nature.com/articles/nature11631

[iv]https://beast.community/first_tutorial

[v]https://cran.r-project.org/web/packages/pastis/vignettes/pastis-vignette.pdf

[vi]https://cran.r-project.org/web/packages/phytools/phytools.pdf

[vii]https://cran.r-project.org/web/packages/ape/ape.pdf

[viii]https://yulab-smu.top/treedata-book/

[ix]https://cran.r-project.org/web/packages/brms/vignettes/brms_overview.pdf

[x]https://cran.r-project.org/web/packages/brms/vignettes/brms_phylogenetics.html

[xi]https://link.springer.com/article/10.1007/s00442-020-04782-x#citeas

[1]Department of Zoology, University of Otago, Dunedin, New Zealand

*Correspondence:
afilion90@gmail.com (A. Filion).

https://doi.org/10.1016/j.pt.2022.01.008

References

1. Tompkins, D.M. *et al.* (2011) Wildlife diseases: from individuals to ecosystems. *J. Anim. Ecol.* 80, 19–38
2. Gottdenker, N.L. *et al.* (2014) Anthropogenic land use change and infectious diseases: a review of the evidence. *EcoHealth* 11, 619–632

3. Daszak, P. *et al.* (2000) Emerging infectious diseases of wildlife – threats to biodiversity and human health. *Science* 287, 443–449
4. Lymbery, A.J. *et al.* (2014) Co-invaders: the effects of alien parasites on native hosts. *Int. J. Parasitol.* 3, 171–177
5. Suzán, G. *et al.* (2015) Metacommunity and phylogenetic structure determine wildlife and zoonotic infectious disease patterns in time and space. *Eco. Evol.* 5, 865–873
6. Peterson, A.T. *et al.* (1999) Conservatism of ecological niches in evolutionary time. *Science* 285, 1265–1267
7. Poulin, R. *et al.* (2011) The comparative ecology and biogeography of parasites. *Philos. Trans. R. Soc. B* 366, 2379–2390
8. Page, R. (2003) *Tangled Trees, Phylogeny, Cospeciation, and Coevolution*, University of Chicago Press
9. Ricklefs, R.E. and Fallon, S.M. (2002) Diversification and host switching in avian malaria parasites. *Proc. R. Soc. B* 269, 885–892
10. Antonovics, J. *et al.* (2002) The ecology and genetics of a host shift: *Microbotryum* as a model system. *Am. Nat.* 160, S40–S53
11. Morand, S. and Poulin, R. (2003) Phylogenies, the comparative method and parasite evolutionary ecology. *Adv. Parasitol.* 54, 282–294
12. Poulin, R. *et al.* (2003) A comparative analysis of adult body size and its correlates in acanthocephalan parasites. *Int. J. Parasitol.* 33, 799–805
13. Fountain-Jones, N.M. *et al.* (2018) Towards an ecophylogenetic framework for infectious disease ecology. *Biol. Rev.* 93, 950–970
14. Wasserstein, R.L. *et al.* (2019) Moving to a world beyond 'p<0.05'. *Am. Stat.* 73, 1–19
15. Davies, J.T. (2021) Ecophylogenetics redux. *Ecology* 24, 1073–1088