Title: On the evolution of host specificity: a case study of helminths

Authors: Alaina C. Pfenning-Butterworth^{1*}, Sebastian Botero-Cañola¹ and Clayton E. Cressler¹

¹*School of Biological Sciences, University of Nebraska-Lincoln, NE 68588*

*Corresponding Author: A. C. Pfenning-Butterworth, apfenning@huskers.unl.edu

Competing Interests: The authors have no competing interests to declare.

Keywords: host specificity, phylogeny, helminth, parasite evolution, zoonoses

ABSTRACT

The significant variation in host specificity exhibited by parasites has been separately linked to evolutionary history and ecological factors in specific host-parasite associations. Yet, whether there are any general patterns in the factors that shape host specificity across parasites more broadly is unknown. Here we constructed a molecular phylogeny for 249 helminth species infecting free-range mammals and find that the influence of ecological factors and evolutionary history varies across different measures of host specificity. Whereas the phylogenetic range of hosts a parasite can infect shows a strong signal of evolutionary constraint, the number of hosts a parasite infects does not. Our results shed new light on the evolution of host specificity in parasites, suggesting that phylogenetic breadth may capture the evolutionary potential of a parasite to jump between hosts, whereas the number of hosts may reflect ecological opportunity. Finally, we show parasite phylogenies can also provide an alternative perspective on zoonosis by identifying which hosts are infected by a broad phylogenetic range of parasites.

INTRODUCTION

Parasites vary considerably in their host specificity, or the range of hosts they can infect, and previous work has identified a large number of ecological and evolutionary factors that may lead to variation in host specificity (Bernays and Graham 1988; Poulin 1992; Desdevises et al. 2002; Mouillot et al. 2006; Clark and Clegg 2017). In some systems, host specificity appears to be more strongly shaped by local environment (e.g., host diversity/abundance and abiotic factors, Krasnov et al. 2005; Loiseau et al. 2012; Dallas and Presley 2014), whereas in other systems, host specificity appears to be constrained by evolutionary history (e.g., outcomes of adaptive evolution, Desdevises et al. 2002; Clark and Clegg 2017). These contrasting results are perhaps not surprising, given the phylogenetic, phenotypic, and ecological diversity of parasites – perhaps there are no general patterns that underlie the evolution of specificity across parasite lineages. However, there is value in exploring that question, as a broader understanding of the ecological and evolutionary factors that shape host specificity may help to identify hosts and parasites that are the most likely reservoirs of novel zoonoses.

Studies like those referenced above have assessed the determinants of variation in host specificity using taxonomically restricted datasets that focus on a single group of parasites (e.g., fleas, avian malaria, etc.). This taxonomic restriction can allow researchers to test specific hypotheses (e.g., whether host specificity varies with geography; Krasnov et al. 2005; Krasnov et al. 2008; reviewed in Poulin et al. 2011) and to more readily identify ecological and physiological factors likely to be driving differences in specificity among parasite species (e.g., whether host specificity is shaped by host phylogeny; McCoy et al. 2001; Fallon et al. 2005; Clark and Clegg 2017). However, the widespread variation in host specificity among groups of parasites suggests that inferences that apply to one group will likely be of limited value in other

groups. For instance, monogeneans, ecto-parasitic flat worms, have high host-specificity (~74% infect a single host, Bikhovski 1957) and are thought to have diversified during the Cretaceous period (~100 mya, Kearn 1994). Monogeneans diverged from Platyhelminthes (which emerged ~270 mya, Dentzien-Dias et al. 2013), and the oldest known helminth lineages emerged ~550 mya (Zhang et al 2020). Given such ancient divergences, it would be unwise to conclude that all helminths are highly host-specific, or that the factors that drive variation in specificity among monogeneans will also be important to Platyhelminthes. Nevertheless, helminths like these are perhaps the best group of parasites to study to determine whether local environment or evolutionary history are the primary determinants of host specificity across the broadest phylogenetic scale studied to date.

Helminths are perhaps the most successful parasites in the world, in terms of global prevalence (Dallas et al. 2018). They are a diverse group consisting of four major groups of parasitic worms—acanthocephalans, cestodes, nematodes, and trematodes—that differ in their morphology, transmission, and host specificity (Mackiewicz 1988; Hayunga 1991; Kennedy 2006). They also differ in their evolutionary history (Weinstein and Kuris 2016); for example, nematodes coevolved alongside vertebrates (Poinar 2011), whereas cestodes have emerged more recently (Baer 1952). Previous studies indicate that variation in life history among the helminth groups may help explain key differences in host specificity (Stunkard 1957; Pedersen et al. 2005). For instance, the generalism observed in acanthocephalans is thought to be due to the presence of a free-living stage in their life cycle (Stunkard 1957). The phylogenetic and ecological diversity of helminths presents a rare opportunity to discern whether there are fundamental processes that determine host specificity across broad taxonomic and phylogenetic scales.

Here, we construct a molecular helminth phylogeny for 249 species of helminths that parasitize free-living mammals. This novel phylogeny is the largest ever built for this group, and thus provides unprecedented insight into the factors that shape the evolution of host specificity. We use this phylogeny to test the evidence for two hypotheses in shaping variation in host specificity: (1) evolutionary history, where closely related helminths would always have similar host specificities, and (2) local ecology, where closely related helminths living in different environments would have very different host specificities. We tested these hypotheses using two metrics of host specificity: the number of hosts infected (taxonomic breadth) and the mean pairwise phylogenetic distance among hosts (MPD, Webb et al 2002). We find that the mean pairwise phylogenetic distance among hosts is shaped by evolutionary history, whereas taxonomic breadth is not. Using this phylogeny, we also assessed which mammal species may be more likely to serve as reservoirs for emerging infectious diseases finding that Old and New World monkeys host a phylogenetically diverse set of helminths making them potential sources of zoonotic disease.

MATERIAL AND METHODS

Host-parasite databases

We obtained records of mammal-helminth interactions from two datasets: Global Mammal Parasite Database (GMPD, Stephens et al. 2017) and a novel dataset assembled from parasitology museums (Botero-Cañola 2020). The GMPD includes taxonomic and trait data obtained from the literature for over 10,000 host-helminth associations of wild populations from four Orders of mammals: carnivores (Carnivora), primates (Primates), and ungulates (Artiodactyla and Perissodactyla). Some limitations of the GMPD are the exclusion of rodent hosts and the lack of expert identification of helminths, which could lead to an overestimation of

the number of specialist helminths due to missing hosts or the misidentification of rare species. To account for these potential sources of bias we also ran all analyses with the dataset assembled from voucher specimen-verified museum collections. This dataset includes host-parasite associations for Nearctic mammals in the Orders Artiodactyla (even-toed ungulates), Carnivora (carnivores), Didelphimorphia (opposums), Eulipotyphla (moles and shrews), Lagomorpha (rabbits), and Rodentia (rodents) obtained from the H.W. Manter Parasitology Collection, the former United States National Parasitology Collection, the Museum of Southwestern Biology, and the Canadian Museum of Nature.

Phylogenetic construction

We searched Genbank for DNA sequences (CO1, 18S, and 28S) of all 920 species of helminths included in the GMPD (Stephens et al. 2017), resulting in 249 species. We combined slow evolving sequences (18S and 28S) that resolve deep phylogenetic relationships with a fast-evolving sequence (CO1) to resolve recent relationships. Sequences were aligned using Multiple Sequence Alignment (MUSCLE v. 3.8.425, Edgar 2004) and manually edited in Geneious Prime 2019.0.4 (Kearse et al. 2012). The sequences for each species included 8848 base pairs—1707, 3550, 3591 base pairs for CO1, 18S, and 28S respectively. We implemented a Bayesian likelihood analysis on the combined data set with a GTR + G + I substitution model (PartitionFinder v.2.1.1; Lanfear et al. 2016) and a relaxed lognormal molecular clock model selected in BEASTv.2.5.2 (Bouckaert et al. 2014). We ran three replicates of the Bayesian Markov chain Monte Carlo (MCMC) analysis for 60 million generations each, sampling every 1000 generations (BEASTv.2.5.2; Bouckaert et al. 2014). We assessed the MCMC log files for convergence (effective sample size, ESS > 200, Tracer v.1.7.1; Rambaut et al. 2018) and

removed the first 44% as burn-in (TreeAnnotator v.2.5.2, Bouckaert et al. 2014). The constructed tree is consistent with the topology of published trees for smaller helminth groups (Acanthocephala—Vereweyen et al. 2011; García-Varela et al. 2013, Nematoda—Blaxter et al. 1998; Nadler and Hudspeth 2000, Platyhelminthes—Knapp et al. 2015; Sharma et al. 2016) and clearly assigns recognized Orders to monophyletic clades (Supplemental Figure 1). Focusing only on the host-parasite associations for which we have phylogenetic information for the parasite, the analyses of the GMPD data included 971 associations among 197 mammal species and 249 helminth species; the analyses of the Nearctic museum data included 363 associations among 67 mammal species and 90 helminth species.

Phylogenetic signal in host specificity metrics

Metrics of host specificity used here—taxonomic breadth and mean pairwise phylogenetic distances (MPD)—were calculated following Park et al. (2018). Taxonomic breadth refers to the number of hosts a parasite can infect (Rhode 1980); whereas mean pairwise phylogenetic distance refers to the mean branch length of all pairs of host species infected by a given parasite (Harnos et al. 2017). We used the Phylogenetic Atlas of Mammal Macroecology’s mammal phylogeny (PHYLACINE; Faurby and Svenning 2015; Faurby et al 2018) to determine the MPD of hosts for each helminth using the R package *picante* (Kembel et al. 2010). For single host parasites, we set host taxonomic breadth equal to one and MPD equal to 0. However, these parasites may have a disproportionate influence on our phylogenetic conclusions: for example, moving from one host to two can have a dramatic effect on MPD, in particular, and high host specificity may be a result of under sampling rather than true specificity. To determine how

sensitive the analyses were to the inclusion of single host parasites we reran all analyses, excluding single host parasites.

We assessed whether MPD and taxonomic breadth exhibited phylogenetic signal by testing whether a non-phylogenetic model (white noise) or Pagel's lambda best explained the data (Pagel 1999). The non-phylogenetic model (white noise) treats each species MPD or taxonomic breadth as an independent sample drawn from null distribution with shared mean and variance. High support for the non-phylogenetic model would indicate that the trait is not constrained by phylogenetic relationships. Under Pagel's lambda, the strength of MPD or taxonomic breadth's phylogenetic signal is estimated between 0 and 1 (where 0 indicates that the measure has been evolving as if the species were related by a "star" phylogeny, and 1 indicates that the measure has been evolving under Brownian motion, BM; Felsenstein 1973). High support for a Pagel's lambda model greater than 0 would indicate that closely related helminths have more similar MPD or taxonomic breadth than distantly related helminths.

Evolutionary model fitting

Based on the results of the phylogenetic signal analyses, we used phylogenetic comparative analysis to further explore the evolutionary history of MPD, with and without single host parasites. Specifically, we fit the MPD data to the parasite phylogeny under different models of trait evolution: Brownian motion (BM), a single regime Ornstein-Uhlenbeck (OU1), and a multiple regime OU based on Phyla. The BM model assumes neutral trait evolution, with a variance between taxa that is proportional to the phylogenetic distance (i.e., the branch lengths) between the taxa (Felsenstein 1973; Felsenstein 1985). In contrast, the OU1 model incorporates a deterministic trend towards a single trait value, often interpreted as stabilizing selection

towards an optimum (Hansen 1997; Butler and King 2004). The OU model can also account for multiple selective regimes, allowing for different trait optima in each regime (Hansen 1997; Butler and King 2004; Beaulieu et al 2012); this is often used to test whether ecological or environmental differences among species influence trait evolution after accounting for phylogeny (e.g., the influence of trophic transmission on the life histories of helminth parasites: Benesh et al. 2011, 2014) . We tested an OU model with three regimes corresponding to the different parasite Phyla. All models were fit using the R package ouch, with fit assessed using ω AICc (R v.3.5.2, R Core Team 2018; Burnham and Anderson 2004; Butler and King 2004; Cressler et al. 2015).

To assess the support for the best-fitting evolutionary model, we simulated 1000 datasets with the best-fitting parameter estimates for each model and then re-fit the competing models to the simulated data to create distributions of likelihood values under different generating models (Boettiger et al. 2012). We used a Phylogenetic Monte Carlo (Boettiger et al. 2012) approach to calculate distributions of the test statistic

$$\delta = -2 (\log \mathcal{L}_0 - \log \mathcal{L}_1)$$

where $\mathcal{L}_0$ is the likelihood of the simpler model and $\mathcal{L}_1$ is the likelihood of the more complex model. We computed this test statistic for the BM:OU3 comparison and the OU1:OU3 comparison for the GMPD and Nearctic datasets with and without specialist parasites.

To determine whether the observed value of $\hat{\delta}$ from fitting the real data was significantly different from a null expectation, we approximated a p-value (the probability of observing δ if the simpler model were true). We simulated 1000 datasets using the simpler model at its MLE parameter estimates and then we fit both the simple model and OU3 to the simulated dataset and computed the values of δ . This produces a null distribution of δ under the simpler model. We

calculate an approximate p-value as the fraction of δ values in this distribution that are as extreme as the observed value, $\hat{\delta}$.

To determine the power of each comparison, we computed the probability of rejecting the simpler model when the data were generated by the more complex (OU3) model. We simulated 1000 datasets using the OU3 model at its MLE parameter estimates and then we fit both models to the simulated dataset and computed the values of δ . The fraction of δ values that are larger than the 95th percentile of the null distribution generated under the simpler model above gives an estimate of power.

Phylogenetic Generalized Least Squares

We assessed whether variation host specificity (MPD) is shaped by parasite and host life history traits. Specifically, we used phylogenetic generalized least squares (PGLS) to test for if parasite length (Benesh et al. 2017), transmission mode (close-contact, environmental, and trophic transmission; Stephens et al. 2017), and average host mass (in grams, PHYLACINE, Faurby and Svenning 2015; Faurby et al 2018) explain the variation in MPD for both datasets with and without single host helminths. PGLS analyses were conducted with 52 species for which parasite body length was well documented using the R package *caper* (Orme et al. 2013). To account for phylogenetic signal, we fit each dataset's PGLS using Pagel's λ (Pagel 1999).

Helminth phylogenetic diversity within mammal clades

We calculated the mean pairwise phylogenetic distances (MPD) of helminths that infect each host to identify patterns of helminth diversity within host clades. The helminth phylogeny constructed in this study was used to determine the MPD of helminths that infect each host,

using the R package *picante* (Kembel et al. 2010). Host MPD was visualized on the mammal phylogeny using the R package *phytools* (Revell 2012). We also determined the pairwise distance (PD) of each mammal host to humans to assess which hosts are closely related to humans and are also infected by a broad range of helminths, potentially identifying host species at high risk for emerging infectious disease.

RESULTS

Phylogenetic signal in host specificity metrics

We assessed whether the host specificity metrics, MPD and taxonomic breadth, exhibited phylogenetic signal and found that MPD had phylogenetic signal and taxonomic breadth did not, regardless of the dataset or the inclusion of single host parasites. For MPD, datasets that excluded single host parasite had higher λ values, indicating stronger phylogenetic signal (Nearctic with single host parasites: $\lambda = 0.38$; Nearctic without single host parasites: $\lambda = 0.92$; GMPD with single host parasites: $\lambda = 0.61$; GMPD without single host parasites: $\lambda = 0.90$). For taxonomic breadth, the λ value did not change with the exclusion of single host parasites (Nearctic with single host parasites: $\lambda = 0.09$; Nearctic without single host parasites: $\lambda = 0.09$; GMPD with single host parasites: $\lambda = 0.00$; GMPD without single host parasites: $\lambda = 0.00$).

Given that MPD showed phylogenetic signal, MPD values were visualized on the helminth phylogeny, including single host helminths, for the GMPD and Nearctic dataset (Figures 1 and 2, respectively). Within the GMPD phylogeny, there are clades of specialist helminths (e.g., *Mansonella sp.* and *Cylicocyclus sp.*, Figure 1) and clades of generalist helminths (e.g., *Alaria sp.* and *Trichinella sp.*, Fig. 1). There are also clades in which one or a few species' host specificity varies from the majority of the clade (e.g., *Taenia madoquae* is a

specialist in a predominantly generalist clade, Figure 1). Overall, there is no obvious sign of evolution either towards or away from specialism across the tree. The Nearctic phylogeny show similar variation, including clades of only specialist helminths (e.g., *Trichuris sp.*, Figure 2) and clades with both extreme specialists and generalists (e.g., acanthocephalans, Figure 2).

Evolutionary model fitting

We modelled the evolutionary history of host specificity (MPD) and established that an OU3 model with regimes for each of the parasite Phyla best described the Nearctic and GMPD dataset, with and without the exclusion of single host parasites (using ϖ AICc; Table 1; full details in Supplemental Tables 1-4). We assessed the support for the OU3 model using a Phylogenetic Monte Carlo approach. There was high support for the OU3 model compared to the BM model for both datasets, with and without single host parasites (Supplemental Figure 2). The OU3 model had high support compared to the OU1 model for both datasets, with single host parasites. However, the GMPD dataset without single host parasites had lower power, but still maintained a high p-value, suggesting that it was unlikely that the data was generated by an OU1 model (Supplemental Figure 3).

The OU3 model parameter estimates for the GMPD and Nearctic datasets, with and without single host parasites, can be found in Table 2. For both datasets, with and without single host parasites, the nematode optimum is at a lower MPD value, indicating higher host specificity for nematodes than Platyhelminthes and Acanthocephala. In the GMPD, with and without single host parasites, the Platyhelminthes optimum is the largest, whereas in the Nearctic dataset, the Acanthocephala optimum is the largest. However, regardless of dataset, the stationary variance

of the OU process (given by $\sigma^2/2\alpha$) is very large, indicating that, within each taxonomic clade, there is considerable variation in MPD among parasite species.

Life-history traits and host specificity

For each dataset, with and without specialist helminths, we assessed whether variation host specificity (MPD) was shaped by parasite and host life history traits (parasite length, close-contact transmission, environmental transmission, trophic transmission, and average host mass) using PGLS. In the Nearctic dataset, with specialists, the best model only included host mass as a predictor ($R^2 = 0.18$, $p = 0.01$). We found a significant negative effect of host mass indicating that more specialist helminths are found in large-bodied mammals ($p = 0.01$, See Supplemental File). In the Nearctic dataset, without specialists, the best model included parasite length, close transmission, environmental transmission, and trophic transmission ($R^2 = 0.07$, $p = 0.27$); however, none of these predictors were significant. In the GMPD, with and without specialists, the best model only included close transmission ($R^2 = 0.01$, $p = 0.24$; $R^2 = 0.05$, $p = 0.09$). In neither model did close transmission have a significant effect on MPD.

Helminth phylogenetic diversity within mammal clades

For each dataset, we calculated the mean pairwise phylogenetic distance of helminths infecting each mammal and assessed patterns of helminth diversity within mammal clades. In the Nearctic dataset, 28% of the mammals were only infected by one helminth (MPD = 0), and thirteen species (19%) were infected by multiple phylogenetically distant helminths (MPD value > 0.9; since the helminth tree is scaled to a height of one, a host with parasite MPD > 0.9 indicates that its parasites diverged deep in the tree). Of mammals with high parasite MPDs, 15% are

ungulates, 54% are carnivores, and 31% are rodents (Supplemental Figure 4). In the GMPD, 38% of the mammals were only infected by one helminth (MPD = 0), and 27 species (14%) were infected by multiple phylogenetically distant helminths (MPD value > 0.9). Of these, 22% are ungulates, 63% are carnivores, and 11% are primates (Supplemental Figure 5). There are five carnivores with high MPD (> 0.9) values in both datasets (*Lynx rufus*, *Urocyon cinereoargenteus*, *Lontra canadensis*, *Procyon lotor*, and *Ursus americanus*). *Urocyon cinereoargenteus* (gray fox) and *Procyon lotor* (raccoon) have MPD values greater than 0.95 in both datasets.

We plotted the pairwise distance (PD) of each mammal host to humans against the MPD of helminths infecting each mammal to identify mammals that might pose a higher risk for a spillover event. In the GMPD dataset, which does not include rodents, we found that new world monkeys, which are closely related to humans, have the highest MPD values (Figure 3). In the Nearctic dataset, the species with the highest MPD values (i.e., *Procyon lotor* and *Peromyscus gossypinus*) are not closely related to humans (Figure 3).

DISCUSSION

Prior work has assessed the variation in host specificity in taxonomically restricted groups of parasites (Desdevises et al. 2002; Krasnov et al. 2005, 2008; Hellgren et al. 2009; Loiseau et al. 2012; Dallas and Presley 2014; Clark and Clegg 2017). These and other studies on host specificity are necessary to address specific hypotheses about the determinants of host specificity within a parasite group (e.g., determining whether monogenean host specificity varies with host body size; Desdevises et al. 2002). Krasnov et al. have studied the effects of specific environmental factors on the host specificity of fleas that infect mammals finding that fleas with a larger geographic range also have a larger host breadth and that these increase with latitude

(2005; 2008). Additionally, work on two species of fleas, *Listropsylla agrippinae* and *Chiastopsylla rossi*, has indicated that host breadth varies with time spent on the host, suggesting that life history traits shape host specificity (Van der Mescht et al. 2015). In this study, we asked whether we could identify any general patterns of variation in host specificity. The general patterns indicated in this study would benefit from additional work within specific clades to identify the specific evolutionary or ecological interactions shaping host specificity.

We used a novel helminth phylogeny and two helminth-mammal databases to test competing hypotheses about the determinants of variation in host specificity. The first hypothesis we considered suggests that host specificity is determined by local environmental factors, such as host diversity and abundance or abiotic factors. Under this hypothesis, closely related parasites living in different environments would have very different host specificities, whether measured using the phylogenetic distance between hosts (MPD) or the number of hosts (taxonomic breadth). The second hypothesis suggests that host specificity is determined by evolutionary constraint. Under this hypothesis, closely related parasites would always have similar host specificities. Finally, host specificity could be determined by some combination or neither of these hypotheses.

We found, regardless of the dataset used or the inclusion of single host parasites, that MPD is constrained by evolutionary history and taxonomic breadth is not. This suggests that the phylogenetic breadth of hosts a parasite can infect is shaped by evolution, and possibly coevolution (Hadfield et al. 2014), whereas the number of hosts a parasite can infect is not. Similar results have been found in other helminths (monogeneans, Desdevises et al. 2002; trematodes, cestodes, and nematodes; Mouillet et al. 2006). This difference between host specificity metrics suggests that host specificity is determined by local environment *and*

evolutionary history. One possible interpretation of these results is that the phylogenetic range of hosts a parasite can infect is constrained its life history traits (given the results of PGLS, it appears that the life history traits that effect MPD likely vary with parasite species), whereas the number of hosts that a parasite can infect within that range is determined by the availability of those hosts within the extent of the parasite's geographic range. Some studies have shown support for this idea, finding that the host range of a parasite varies across different areas of their geographic distribution (Thompson 2005; Ricklefs 2010; Huang et al. 2018). Additional evidence for this interpretation could come from studies counting how many hosts within a parasite's phylogenetic range have an overlapping geographic range with the parasite: this number should correspond to the parasite's taxonomic breadth. Together, these results suggest that, like free-living species, parasites have a fundamental (shaped by MPD) and a realized (taxonomic breadth) niche.

MPD was best described by an OU3 model with regimes specified by Phyla for both datasets, with and without single host helminths, indicating that host specificity has evolved differently within clades of helminths. The OU3 model parameters for each dataset indicate that Nematodes are the most host specific phylum of helminths. However, the most generalist phylum of helminths varies with the dataset used; in the GMPD Platyhelminthes are the most generalist and in the Nearctic dataset the Acanthocephala are the most generalist. The Acanthocephala are trophically transmitted, with both a free-living life stage and life stages inside their intermediate and definitive hosts (Kennedy, 2006); given this variability in habitats and hosts, it makes some sense that they would be generalist, able to infect a broad range of hosts (Pedersen et al. 2005). Given that both Platyhelminthes and Nematodes contain a diverse range

of life cycles, habitats, and hosts, further research is needed to understand why Platyhelminthes appear more generalist and Nematodes appear more specialist.

Mammals are considered common reservoirs for zoonotic pathogens because they are closely related to and have frequent interactions with humans (Brook and Dobson 2015; Han et al. 2015). Prior research has used mammal traits and parasite host specificity to predict which mammal-parasite interactions are most likely to lead to spill-over events (Stephens et al. 2016; Olival et al. 2017). These and other research indicate that parasites with low host specificity, i.e., generalists, are a greater zoonotic threat because they are more likely to infect novel hosts than specialists (Woolhouse and Gowtage-Sequeria 2005; Johnson et al. 2015).

Here we used the helminth phylogeny to add a new perspective to this discussion, by considering the phylogenetic diversity of parasites infecting the host, in addition to the host range of the parasites. Hosts that are closely related to humans and are infected by phylogenetically distance parasites may be more likely to harbor zoonotic pathogens. We found that both Old and New World monkeys serve as hosts to many distantly related helminths (Figure 3). This is consistent with prior work indicating that zoonotic diseases ‘jump’ from hosts that are closely related to humans (Pederson and Davies 2009; Han et al. 2016). This helminth phylogeny provides a unique approach for addressing which hosts are likely to harbor zoonotic diseases by allowing for the identification of host clades that are infected by phylogenetically distant parasites.

355 **Acknowledgements**

356 We thank the members of the Macroecology of Infectious Disease Research Coordination
357 Network (NSF-DEB 131223), especially J. Davies, M. Farrell, J. Herrera, A. Park, and P.
358 Stephens, for helpful discussion and development of methods, and S. Huang for their comments
359 on the manuscript.
360

References

- Baer, J.G. (1952). Ecology of Animal Parasites. University of Illinois Press, Urbana, pp. 224.
- Benesh, D.P., Lafferty, K.D. & Kuris, A. (2017). A life cycle database for parasitic acanthocephalans, cestodes, and nematodes. *Ecology*, 98, 882.
- Beaulieu, J.M., Jhvueng, D.C., Boettiger, C. & O'Meara, B.C. (2012). Modeling stabilizing selection: expanding the Ornstein–Uhlenbeck model of adaptive evolution. *Evolution*, 66, 2369-2383.
- Benesh, D.P., Chubb, J.C. & Parker, G.A. (2011). Exploitation of the same trophic link favors convergence of larval life-history strategies in complex life cycle helminths. *Evolution*, 65(8), 2286-2299.
- Benesh, D.P., Chubb, J.C. & Parker, G.A. (2014). The trophic vacuum and the evolution of complex life cycles in trophically transmitted helminths. *Proc. R. Soc. B.*, 281(1793), 20141462.
- Bernays, E. & Graham, M. (1988). On the evolution of host specificity in phytohagous arthropods. *Ecology*, 69, 886-892.
- Bikhovski, B.E. (1957). Monogenetic trematodes, their classification and phylogeny. Moscow & Leningrad: Isdatelsvo Akad, pp. 509
- Blaxter, M.L., De Ley P., Garey, J.R., Liu, L.X., Scheldeman, P., Vierstraete, A. *et al.* (1998). A molecular evolutionary framework for the phylum Nematoda. *Nature*, 392, 71-75.
- Boettiger, C., Coop, G. & Ralph, P. (2012). Is your phylogeny informative? Measuring the power of comparative methods. *Evolution*, 66(7), 2240-2251.

Botero-Cañola, S. (2020). Parasite geography: exploring drivers of parasite biodiversity distribution at species and community scales. PhD Dissertation, College of Arts and Science, University of Nebraska, Lincoln, Nebraska, USA

Bouckaert, R.R., Heled, J., Kühnert, D., Vaughan T, Wu, C-H, Xie, D., *et al.* (2014). BEAST 2: A software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.*, 10, e1003537.

Brook, C.E. & Dobson, A.P. (2015). Bats as ‘special’ reservoirs for emerging zoonotic pathogens. *Trends in Microbiol.*, 23, 172-180.

Burnham, K.P. & Anderson, D.R. (2004). Multimodel inference: understanding AIC and BIC in model selection. *Sociol. Methods Res.*, 33, 261-304.

Butler, M.A. & King, A.A. (2004). Phylogenetic comparative analysis: a modeling approach for adaptive evolution. *Am. Nat.*, 164, 683-695.

Clark, N.J. & Clegg, S.M. (2017). Integrating phylogenetic and ecological distances reveals new insights into parasite host specificity. *Mol. Ecol.*, 26, 3074-3086.

Cressler, C.E., Butler, M.A. & King, A.A. (2015). Detecting adaptive evolution in phylogenetic comparative analysis using the Ornstein-Uhlenbeck model. *Syst. Biol.*, 64, 953-968.

Dallas, T.A., Aguirre, A.A., Budischak, S., Carlson, C., Ezenwa, V., Han, B., *et al.* (2018). Gauging support for macroecological patterns in helminth parasites. *Global Ecol. Biogeogr.*, 27(12), 1437-1447.

Dallas, T. & Presley, S.J. (2014). Relative importance of host environment, transmission potential and host phylogeny to the structure of parasite metacommunities. *Oikos*, 123, 866-874.

404 Dentzien-Dias, P.C., Poinar Jr, G., de Figueiredo, A.E.Q., Pacheco, A.C.L., Horn, B.L. &
 405 Schultz, C.L. (2013). Tapeworm eggs in a 270 million-year-old shark coprolite. *PLoS*
 406 *One*, 8, e55007.

407 Desdevises, Y., Morand, S. & Legendre, P. (2002). Evolution and determinants of host
 408 specificity in the genus *Lamellodiscus* (Monogenea). *Biol. J. Linnean Soc.*, 77, 431-443.

409 Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high
 410 throughput. *Nucleic Acids Res.* 32, 1792-1797.

411 Fallon, S.M., Bermingham, E. & Ricklefs, R.E. (2005). Host specialization and geographic
 412 localization of avian malaria parasites: a regional analysis in the Lesser Antilles. *Am.*
 413 *Nat.*, 165(4), 466-480.

414 Faurby, S., Davis, M., Pedersen, R.Ø., Schowaneck, S.D., Antonelli, A. & Svenning, J-C. (2018).
 415 PHYLACINE 1.2: The Phylogenetic Atlas of Mammal Macroecology. *Ecology*, 99,
 416 2626.

417 Faurby, S. & Svenning, J.C. (2015). A species-level phylogeny of all extant and late Quaternary
 418 extinct mammals using a novel heuristic-hierarchical Bayesian approach. *Mol.*
 419 *Phylogenet. Evol.*, 84, 14-26.

420 Felsenstein, J. (1973). Maximum-likelihood estimation of evolutionary trees from continuous
 421 characters. *Am. J. Hum. Genet.*, 25, 471-492.

422 Felsenstein, J. (1985). Phylogenies and the comparative method. *Am. Nat.*, 125, 1-15.

423 García-Varela, M, Pérez-Ponce de León, G., Aznar, F.J. & Nadler, S.A. (2013). Phylogenetic
 424 relationship among genera of Polymorphidae (Acanthocephala), inferred from nuclear
 425 and mitochondrial gene sequences. *Mol. Phylo. Evo.*, 68, 176-184.

426 Han, B.A., Kramer, A.M. & Drake, J.M. (2016). Global patterns of zoonotic diseases in
427 mammals. *Trends in Parasitol.*, 32, 565-577.

428 Han, B.A., Schmidt, J.P., Bowden, S.E. & Drake, J.M. (2015). Rodent reservoirs of future
429 zoonotic diseases. *PNAS*, 112, 7039–7044.

430 Hansen, T.F. (1997). Stabilizing selection and the comparative analysis of adaption. *Evolution*,
431 51(5), 1341-1351.

432 Harnos, A., Lang, Z., Petrás, D., Bush, S.E., Szabó, K. & Rózsa, L. (2017). Size matters for lice
433 on birds: coevolutionary allometry of host and parasite body size. *Evolution*, 71, 421-431.

434 Hayunga, E.G. (1991). Morphological adaptations of intestinal helminths. *J. Parasitol.*, 77, 865-
435 873.

436 Hellgren, O., Pérez-Tris, J. & Bensch, S. (2009). A jack-of-all-trades and still a master of some:
437 prevalence and host range in avian malaria and related blood parasites. *Ecology*, 90(10),
438 2840-2849.

439 Huang, X., Ellis, V.A., Jönsson, J. & Bensch, S. (2018). Generalist haemosporidian parasites are
440 better adapted to a subset of host species in a multiple host community. *Mol.*
441 *Ecol.*, 27(21), 4336-4346.

442 Johnson, C.K., Hitchens, P.L., Evans, T.S., Goldstein, T., Thomas, K., Clements, A., *et al.*
443 (2015). Spillover and pandemic properties of zoonotic viruses with high host plasticity.
444 *Sci. Rep.*, 5, 14830.

445 Kearns, G.C. (1994). Evolutionary expansion of the Monogenea. *Int. J. Parasitol.*, 24, 1227-1271.

446 Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., *et al.* (2012).
447 Geneious Basic: An integrated and extendable desktop software platform for the
448 organization and analysis of sequence data. *Bioinformatics*, 28, 1647–1649.

Kembel, S.W., Cowan, P.D., Helmus, M.R., Cornwell, W.K., Morlon, H., Ackerly, D.D., *et al.* (2010). Picante: R tools for integrating phylogenies and ecology. *Bioinformatics*, 26, 1463–1464.

Kennedy, C.R. (2006). Ecology of the Acanthocephala. Cambridge University Press, Cambridge.

Knapp, J., Gottstein, B., Saarma, U. & Millon, L. (2015). Taxonomy, phylogeny and molecular epidemiology of *Echinococcus multilocularis*: from fundamental knowledge to health ecology. *Vet. Parasitol.*, 213, 85-91.

Krasnov, B.R., Poulin, R., Shenbrot, G.I., Mouillot, D. & Khokhlova, I.S. (2005). Host specificity and geographic range in haematophagous ectoparasites. *Oikos*, 108(3), 449-456.

Krasnov, B.R., Shenbrot, G.I., Khokhlova, I.S., Mouillot, D. & Poulin, R. (2008). Latitudinal gradients in niche breadth: empirical evidence from haematophagous ectoparasites. *J. Biogeogr.*, 35(4), 592-601.

Lanfear, R., Frandsen, P.B., Wright, A.M., Senfeld, T. & Calcott, B. (2016). PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Mol. Biol. Evol.*, 34, 772-773.

Loiseau, C., Harrigan, R.J., Robert, A., Bowie, R.C., Thomassen, H.A., Smith, T.B., *et al.* (2012). Host and habitat specialization of avian malaria in Africa. *Mol. Ecol.*, 21, 431-441.

Mackiewicz, J.S. (1988). Cestode transmission patterns. *J. Parasitol.*, 74, 60-71.

McCoy, K.D., Boulinier, T., Tirard, C. & Michalakis, Y. (2001). Host specificity of a generalist parasite: genetic evidence of sympatric host races in the seabird tick *Ixodes uriae*. *J. Evol. Biol.*, 14(3), 395-405.

- Mouillot, D., Krasnov, B.R., Shenbrot, G.I., Gaston, K.J. & Poulin, R. (2006). Conservatism of host specificity in parasites. *Ecography*, 29, 596-602.
- Nadler, S.A. & Hudspeth, D.S.S. (2000). Phylogeny of the Ascaridoidea (Nematoda: Ascaridida) based on three genes and morphology: hypotheses of structural and sequence evolution. *J. Parasitol.*, 86, 380-393.
- Olival, K.J., Hosseini, P.R., Zambrana-Torrel, C., Ross, N., Bogich, T.L. & Daszak, P. (2017). Host and viral traits predict zoonotic spillover from mammals. *Nature*, doi:10.1038/nature22975.
- Orme, D., Freckleton, R., Thomas, G. & Petzoldt, T. (2013). The caper package: comparative analysis of phylogenetics and evolution in R. *R package version*, 5, 1-36.
- Pagel, M. (1999). Inferring the historical patterns of biological evolution. *Nature*, 401, 877–884.
- Park, A.W., Farrell, M.J., Schmidt, J.P., Huang, S., Dallas, T.A., Pappalardo, P., *et al.* (2018). Characterizing the phylogenetic specialism-generalism spectrum of mammal parasites. *Proc. R. Soc. B.*, 285, 20172613.
- Pedersen, A.B., Altizer, S., Poss, M., Cunningham, A.A. & Nunn, C.L. (2005). Patterns of host specificity and transmission among parasites of wild primates. *Int. J. Parasitol.*, 35, 647-657.
- Pederson, A.B. & Davies, T.J. (2009). Cross-species pathogen transmission and disease emergence in primates. *EcoHealth*, 6, 496–508.
- Poinar Jr, G.O. (2011). The evolutionary history of nematodes—as revealed in stone, amber and mummies. Brill, Leiden, the Netherlands.
- Poulin, R. (1992). Determinants of host-specificity in parasites of freshwater fishes. *Int. J. Parasitol.*, 22, 753-758.

Poulin, R., Krasnov, B.R. & Mouillot, D. (2011). Host specificity in phylogenetic and geographic space. *Trends Parasitol.*, 27(8), 355-361.

R Core Team. (2018). R: A language and environment for statistical computing.

Rambaut, A., Drummond, A.J., Xie, D., Baele, G. & Suchard, M.A. 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst. Biol.*, 67, 901-904.

Revell, L.J. (2012). Phytools: An R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.*, 3, 217-223.

Ricklefs, R. E. (2010). Evolutionary diversification, coevolution between populations and their antagonists, and the filling of niche space. *PNAS*, 107(4), 1265-1272.

Rohde, K. (1980). Host specificity indices of parasites and their application. *Experientia*, 36, 1369.

Sharma, S., Lyngdoh, D., Roy, B. & Tandon, V. (2016). Molecular phylogeny of Cyclophyllidea (Cestoda: Eucestoda): an in-silico analysis based on mtCOI gene. *Parasitol. Res.*, 115, 3329-3335.

Stephens, P.R., Altizer, S., Smith, K.F., Alonso Aguirre, A., Brown, J.H., Budischak, S.A., *et al.* (2016). The macroecology of infectious diseases: a new perspective on global drivers of pathogen distributions and impacts. *Eco. Lett.*, 19, 1159-1171.

Stephens, P.R., Pappalardo, P., Huang, S., Byers, J.E., Farrell, M.J., Gehman, A., *et al.* (2017). Global mammal parasite database version 2.0. *Ecology*, 98, 1476.

Stunkard, H.W. (1957). Host-specificity and parallel evolution of parasitic flatworms. *Z. Tropenmed. Parasitol.*, 8, 254-263.

Thompson, J.N. (2005). The geographic mosaic of coevolution. University of Chicago Press.

517 Van der Mescht, L., Matthee, S., & Matthee, C. A. (2015). Comparative phylogeography
518 between two generalist flea species reveal a complex interaction between parasite life
519 history and host vicariance: parasite-host association matters. *BMC Evol. Biol.*, 15(1),
520 105.

521 Vereweyen, L., Klimpel, S. & Palm, H.W. (2011). Molecular phylogeny of the Acanthocephala
522 (class Palaeacanthocephala) with a paraphyletic assemblage of the orders Polymorphida
523 and Echinorhyncida. *PLoS One*, 6, e28285.

524 Webb, C.O., Ackerly, D.D., McPeck, M.A. & Donoghue, M.J. (2002). Phylogenies and
525 community ecology. *Annu. Rev. Ecol. Syst.*, 33, 475-505.

526 Weinstein, S.B. & Kuris, A.M. (2016). Independent origins of parasitism in Animalia. *Biol. Lett.*,
527 12, 20160324.

528 Woolhouse, M.E.J. & Gowtage-Sequeria, S. (2005). Host range and emerging and reemerging
529 pathogens. *Emerg. Infect. Dis.*, 11, 1842-1847.

530 Zhang, Z., Strotz, L.C., Topper, T.P., Chen, F., Chen, Y., Liang, Y., *et al.* (2020). An encrusting
531 kleptoparasite-host interaction from the early Cambrian. *Nat. Commun.*, 11, 1-7.

Figures

Figure 1. GMPD helminth phylogeny colored by MPD (warmer colors indicate higher specialism). Single host helminths are included.

Figure 2. Nearctic helminth phylogeny colored by MPD (single host helminths included).

Table 1. Weighted AICc scores for each of the three evolutionary models (a three-regime OU model [OU3], a single-regime OU model [OU1], and a Brownian motion model [BM]) fit to the MPD data for each dataset, with and without specialist helminths.

Model	Nearctic	Nearctic (no specialists)	GMPD	GMPD (no specialists)
OU3	0.99	0.99	0.99	0.88
OU	9.9×10^{-3}	4.1×10^{-4}	6.5×10^{-3}	0.11
BM	2.2×10^{-7}	5.3×10^{-5}	1.6×10^{-20}	5.9×10^{-6}

Table 2. Parameter estimates (confidence intervals) for three-regime OU models (OU3) fit to the MPD data for each dataset, with and without specialist helminths.

Dataset	Alpha	s2	Acanthocephala	Nematoda	Platyhelminthes
Nearctic	21.2 (5.94, 1100)	167000 (40400, 7930000)	134 (85.6, 184)	42 (23, 61.7)	95.1 (73.3, 116)
Nearctic (no specialists)	486 (6.9, 1630)	2630000 (38700, 9770000)	201 (147, 254)	52.9 (35.5, 69)	119 (99.3, 139)
GMPD	16.8 (11, 33.7)	42500 (27700, 76800)	24.8 (6.19, 43.1)	21.7 (14.6, 28.6)	47.6 (39.1, 55.8)
GMPD (no specialists)	6.49 (4.56, 17.1)	13800 (10200, 37200)	67.7 (34.4, 99.8)	37.5 (24.6, 50.7)	68.2 (54.4, 81.7)

551 Figure 3. Mean pairwise phylogenetic distance (MPD) of helminths infecting each mammal
 552 compared to their phylogenetic distance to humans in the GMPD and Nearctic dataset. *Didelphis*
 553 *virginiana* and *Didelphis marsupialis* are most distantly related to humans (PD = 380), thus were
 554 excluded from the figure.

555