Global patterns in helminth host specificity: phylogenetic and functional diversity of regional host species pools matter

Konstans Wells[1,*], David I. Gibson[2], Nicholas J. Clark[3]

1 Department of Biosciences, Swansea University, Swansea SA2 8PP, Wales, UK

2 Department of Life Sciences, Natural History Museum, London SW7 5BD, United Kingdom

3 School of Veterinary Science, University of Queensland, Gatton, QLD 4343, Australia

***Correspondence**:

Konstans Wells, *Department of Biosciences, Swansea University, Swansea SA2 8PP, Wales, UK*,

E-mail: k.l.wells@swansea.ac.uk, Phone: +44 (0)1792 518633

Abstract

Host specificity has a major influence on a parasite's ability to shift between human and animal host species. Yet there is a dearth of quantitative approaches to explore variation in host specificity across biogeographical scales, particularly in response to the varying community compositions of potential hosts. We built a global dataset of intermediate host associations for nine of the world's most widespread helminth parasites (all of which infect humans). Using hierarchical models, we asked if realised parasite host specificity varied in response to regional variation in the phylogenetic and functional diversities of potential host species. Parasites were recorded in 4-10 zoogeographical regions, with some showing considerable geographical variation in observed versus expected host specificity. Parasites

generally exhibited the lowest phylogenetic host specificity in regions with the greatest variation in prospective host phylogenetic diversity, namely the Neotropical, Saharo-Arabian and Australian regions. Globally, we uncovered notable variation in parasite host shifting potential. Observed host assemblages for *Hydatigera taeniaeformis* and *Hymenolepis diminuta* were less phylogenetically diverse than expected, suggesting limited potential to spillover into unrelated hosts. Host assemblages for *Echinococcus granulosus*, *Mesocestoides lineatus* and *Trichinella spiralis* were less functionally diverse than expected, suggesting limited potential to shift across host ecological niches. By contrast, *Hydatigera taeniaeformis* infected a higher functional diversity of hosts than expected, indicating strong potential to shift across hosts with different ecological niches. We show that the realised phylogenetic and functional diversities of infected hosts are determined by biogeographical gradients in prospective host species pools. These findings emphasise the need to account for underlying species diversity when assessing parasite host specificity. Our framework to identify variation in realised host specificity is broadly applicable to other host-parasite systems and will provide key insights into parasite invasion potential at regional and global scales.

Keywords: ecological fitting, environmental niche conservatism, host-parasite interactions, host specificity, invasion potential, parasite global spread, phylogeography

Introduction

The spillover of parasites from one host species to another, and the sharing of parasites among humans, domestic animals and wildlife hosts, are of central public and animal health concern (Hassell et al. 2017, Hatcher et al. 2012). Understanding the underlying drivers of parasite host specificity is crucial for disease risk prediction and mitigation (Sokolow et al. 2015, Wood et al. 2012), as well as for forecasting the establishment of novel host-parasite

associations following biotic invasions (Agosta and Klemens 2008). Yet, whether patterns of host shifting (colonising a new host species by means other than co-speciation, resulting in host range expansion) varies due to biogeographical variation in underlying species pools has been only addressed in few studies (Krasnov et al. 2004), hampering the search for general processes that govern multi-species host parasite interactions (Park et al. 2018, Poulin et al. 2011).

Specialisation of species, in terms of both resource use and biotic interactions, is a crucial determinant of species distributions and community assembly across a range of scales (Devictor et al. 2010). For parasites, some adaptation to a particular host environment is required for within-host survival, reproduction and transmission maintenance, depicting a parasite's fundamental niche as an end product of evolutionary adaptations to its host species. Following contact with novel host species, parasite host shifting requires adherence to the principles of 'ecological fitting' (Janzen 1985), which postulates that the sharing of key characteristics with previous host species is necessary for successful infection (Brooks et al. 2006, Hoberg and Brooks 2008, Wells et al. 2015). The extent of a parasite's specialisation is therefore a key determinant of invasion capacity and the likelihood of establishing novel host-parasite associations (Agosta et al. 2010).

Besides host characteristics that allow parasites to switch hosts, host community composition and contact patterns between different host species are key to host shifting (Begon et al. 2002, Clark et al. 2017). At local scales, spatiotemporal variation in host species presence and abundance can result in altered interspecific contact rates that ultimately drive spatiotemporal gradients in a parasite's realised host range (Canard et al. 2014). This heterogeneity in host composition will have important consequences for determining host specificity for widespread parasites, particularly those that infect wildlife. At a global scale, wildlife communities occur in distinct species communities depending on their

biogeographical history, the timing of speciation events and the distributions of habitat biomes (Holt et al. 2013, Kraft et al. 2007, Wallace 1876). The emergence of parasites into novel environments, particularly along invasion routes of globally distributed host species, such as humans and commensal animals, involves contact with endemic species not previously encountered as prospective hosts. Consequently, contemporary regional opportunities for host-parasite species interactions likely set the stage for shaping a parasite's realised niche.

It is reasonable to assume that the realised host specificity of parasites can vary across regions. Yet, to our knowledge, few studies of parasite host specificity across biogeographical scales account for variation in available hosts (Cooper et al. 2012, Doña et al. 2017, Krasnov et al. 2004, Lootvoet et al. 2013, Poisot et al. 2017). Understanding how host specificity relates to the composition of regional species pools can provide important insights into the global invasion potential of parasites (Murray et al. 2015). A better understanding of natural variation in host specificity could be used to quantify parasite invasion potential or facilitate the identification of novel host shifts over distantly related host species (by identifying relatively high phylogenetic or functional diversity of the regional host spectra).

For helminth parasites that use intermediate host species in their life cycles, including most of the focal tapeworms and nematode species of our study, the principles of ecological fitting apply strongly to trophic relationships. Host shifting commonly occurs within similar host feeding guilds (Hoberg and Brooks 2008), suggesting food web structure plays an important role in facilitating helminth parasite invasions. Indeed, for many widespread helminth parasites, host specificity appears less restricted at the intermediate host stage than the definitive host stage, with non-adult intermediate stages able to infect various organs or body cavities of a diversity of intermediate hosts (Morand et al. 2006). Though colonisation

of new bioregions requires ecological fitting across multiple trophic levels (due to the necessity of both intermediate and definitive hosts, Malcicka et al. (2015)), host specificity at the intermediate stage plays a key role in helminth parasite invasion potential. A prominent example of the consequences of differing specificity strategies at the intermediate host level is the comparison of two widespread tapeworm species of *Echinococcus*. Adults of *E. granulosus* develop only in dogs (*Canis familiaris*) and a few other carnivores, but can utilise a large range of herbivorous and omnivorous mammalian species as intermediate hosts. Within its exotic range in Australia, the presence of dingos and dingo/dog hybrids that feed on a large range of endemic wildlife, has enabled the establishment of stable transmission cycles of *E. granulosus* through wild dogs and endemic wildlife (Jenkins 2006), illustrating how host shifting into novel communities may be facilitated by particular regional conditions. In contrast, *E. multilocularis* usually utilises foxes (*Vulpes vulpes*) as its definitive host species and primarily infects voles/mice as intermediate hosts; although this congeneric tapeworm also covers a large geographical range, its distribution is necessarily confined to regions with a sufficient abundance of the major intermediate host species (i.e. voles) in North America, Europe and northern Asia (Davidson et al. 2012). We propose that large scale assessments of host-parasite interactions in relation to host compositional variation can detect these patterns, providing a better understanding of the mechanisms driving parasite distributions.

Here, we assess geographical variation in phylogenetic and functional host specificity for nine of the most globally widespread mammalian helminth parasites (all of which infect humans). By accounting for regional variation in the compositions of prospective intermediate host species, we extend commonly used host specificity metrics to provide a relatively unbiased, global-scale assessment of parasite host specificity and invasion potential. We expect the phylogenetic and functional diversities of infected hosts to show

high regional variation for parasites whose transmission cycles are poorly maintained by wildlife within their exotic range (such as *E. multilocularis*). Accordingly, we expect different levels of parasite host specificity to be reflected by the particular host species attributes that predict host association frequencies.

Materials and methods

Mammalian helminth-host database

We compiled a global database of mammalian host-parasite associations from the publicly available Host-Parasite Database of the Natural History Museum (NHM), London (Gibson *et al.*, 2005). Data extraction is described elsewhere, resulting in a database of 24,486 unique combinations of host–parasite–country records for selected helminth taxa (Nematoda, Cestoda, Trematoda), totalling 4,507 parasite species recorded from 1,366 mammalian host species (Wells et al. 2018). Location names were standardised to country names of the current world geopolitical map and assigned to one of 11 zoogeographical regions according to Holt et al. (2013). We focussed on nine focal parasite species: *Calodium hepaticum* (Nematoda), *Echinococcus granulosus* (Cestoda), *Echinococcus multilocularis* (Cestoda), *Hydatigera taeniaeformis* (Cestoda), *Hymenolepis diminuta* (Cestoda), *Mesocestoides lineatus* (Cestoda), *Taenia hydatigena* (Cestoda), *Trichinella spiralis* (Nematoda) and *Versteria mustelae* (Cestoda). These species all infect humans, exhibit large mammalian host ranges (each has been recorded in > 45 host species in our database), are globally distributed and were sufficiently covered in our database for statistical inference.

To elucidate regional patterns in the composition of potential intermediate host species, we gathered lists of all unique mammal species that have been sampled for parasites (i.e. all host species recorded in the database) in each zoogeographical region. Unique lists of potential hosts were generated for each parasite in each region where the parasite has been

recorded, and were constrained to those host species belonging to the same taxonomic orders as the recorded host species (typically involving small mammals and ungulates). This was done because we conservatively consider mammal species from the same orders to be the most likely potential host species. For analysis (as outlined below), we excluded all species belonging to the Carnivora from these lists to focus only on potential intermediate host species (carnivores are typically definitive host species for the focal parasites). Note that this selection comprises definitive hosts for *Hymenolepis diminuta*, which uses arthropods as intermediate hosts. For each mammal species included in these parasite- and region-specific selections (hereafter referred as mammalian species pools), associations with each of the focal parasites were recorded as binary variables (presence-absence) for use as response variables. We are aware that our dataset is erroneous in that it lacks recent records of host parasite interactions (i.e. false zeros if true interactions are not recorded in the dataset and/or novel species are missed out) and also may include accidental hosts in which parasites have been recorded but cannot reproduce (i.e. false positives); while this limits inference on important measures such as host breadth or transmission potential, we believe that our dataset provides meaningful insights into the relative strength of phylogenetic and ecological signals in host specificity, which were the focus of this study.

Mammalian host phylogeny and ecological trait data

A central goal of this study was to assess whether variation in the phylogenetic and ecological similarities of mammalian species predict patterns of parasite sharing across regions. We proceeded by gathering ecological trait data from the PanTHERIA (Jones et al. 2009) and EltonTraits 1.0 (Wilman et al. 2014) databases to characterise all of the sampled mammals using a range of traits likely to impact on their suitability as hosts for parasites with different life histories. Selected traits were: body mass, which is a key feature of mammals in

terms of their metabolism and adaptation to environments; average longevity, litter size and the average number of litters per year as demographic parameters that could be relevant for enabling parasites to complete parts of their life cycles in a host; diet breadth (calculated as a Shannon diversity index based on the proportional use of 10 diet categories as presented in EltonTraits); range area, which we expect to affect the exposure to other mammalian host species; average temperature and average precipitation within a host's distribution as an indicator of climatic niche; latitudinal centroid of distribution as an indicator of the general habitat and climate within which hosts are occurring across a gradient from tropical to polar biotas; and habitat as multiple binary indicators of whether a species uses 1) forest, 2) open vegetation, and/or 3) artificial/anthropogenic habitats. Information on specific habitat utilisation was compiled from the International Union for the Conservation of Nature (IUCN) database (IUCN, 2014). We did not include a larger set of ecological traits in our analysis to avoid trait collinearity issues.

Sampling bias is likely to influence host-parasite occurrences in our database. We queried the number of published references for each binomial wildlife species name from the 'Scopus' literature database (accessed 25/02/2017) as a measure of research effort (used as a covariate in multiple regression models of host associations); we used this measure, since more refined searches, such as the number of references linked only to parasites, included large proportions of zeros and information on the true number of sampled individuals (which should determine the chance that parasites are detected if prevalence is low) was not available.

Phylogenetic relationships between sampled mammal species were estimated from a recent mammalian supertree (Fritz et al. 2009). We used this tree to compute pairwise phylogenetic distances based on a correlation matrix of phylogenetic branch lengths (Paradis et al. 2004). We also quantified pairwise ecological distance between sampled mammal

201 species based on a generalised form of Gower's distance matrices (Gower 1971) using
202 weighted variables based on all of the ecological trait variables described above, following
203 methods in Pavoine et al. (2009). Phylogenetic and ecological distance matrices were scaled
204 (dividing by the maximum for each distance matrix), so all distance measures ranged from
205 zero to one. Data formatting and analyses were conducted in R version 3.4.3 (R Development
206 Core Team 2017) and relied mainly on the packages *ape* (phylogenetic distance calculations)
207 (Paradis et al. 2004), *ade4* (ecological distance calculations) (Dray and Dufour 2007) and
208 *phytools* (phylogenetic tree plotting) (Revell 2011).
209
210 ***Functional and phylogenetic host specificity across biogeographical gradients***
211 To examine whether realised host specificities of focal parasite species varied in relation to
212 the composition of prospective host species pools, we explored variation in ecological and
213 phylogenetic distances among all infected pairs of host species (observed host diversity)
214 versus those in the available mammalian species pools (expected host diversity) for each
215 focal parasite in each region. For this, we used hierarchical linear regression analysis.
216 With $\mathcal{N}(\mu, \sigma^2)$ denoting normal distributions with mean μ and variance σ^2, we write our
217 models as

218 $$dist \sim \mathcal{N}(\mu_{region} + \beta_{region} host, \sigma^2)$$

219 $$\mu_{region} \sim \mathcal{N}(H_\mu, \sigma_\mu^2); \beta_{region} \sim \mathcal{N}(H_\beta, \sigma_\beta^2).$$

220 Here, μ_{region} denotes the region-specific average of either the functional or phylogenetic
221 distances *dist* for mammalian species pools, whereas coefficient β_{region} is the region-specific
222 estimate of pairwise differences (*dist*) between observed and expected host species (i.e.
223 binary indicator variable *host* = 1 if mammal species is infected, 0 otherwise). H_μ and H_β are
224 hyperpriors (i.e. global 'average' values) for the parameters μ and β; all parameters were

estimated independently for phylogenetic (indexed as "*phyl*") and functional (indexed as "*funct*") diversity (Supplementary Information, **Box S1**).

We fitted regressions and estimated β coefficients for each parasite in a Bayesian framework with Markov Chain Monte Carlo (MCMC) sampling based on the Gibbs sampler in the software JAGS version 4.3.0, operated via the R package *rjags* (Plummer 2016), which conveniently allowed us to account for the hierarchical model structure (see Supplementary Information, **Box S1**). Priors were specified with $H_{\mu} \sim N(0, 100)$ and $\sigma \sim dexp(0.5)$. We ran two chains of 100,000 iterations each for parameter adaptation, then sampled 5,000 posterior parameter estimates. Chain mixing was inspected both visually and with the Gelman-Rubin diagnostic (all values < 1.2). Given our hierarchical model structure, we interpreted effect sizes of β_{region} as potential evidence that functional/phylogenetic distances between host species differ from random draws of expected species from the respective mammalian species pool; negative values of β_{region} indicate a higher functional/phylogenetic similarity between observed host species than expected (i.e. smaller distances than expected), indicating higher host specificity (Clark and Clegg 2017). Positive values of β_{region} indicate parasites infect more distantly related host species than expected, indicating generalism. Estimates of μ_{region} give information on the regional averages of observed – expected distances, whereas variance terms σ_{μ}^2 and σ_b^2 indicate global variation in μ and β across regions where focal parasites occur. We considered β effects as 'significant' if 95% credible intervals did not include zero. We then used these β_{region} coefficients as response variables in linear models, including parasite species and region as categorical covariates, to explore patterns of overall variation in β_{region}.

We next gathered insights into the spread of parasites across regional host communities by computing probabilistic estimates of the proportion of prospective host species that a parasite infects (referred to herein as 'host association rates') within each

regional species pool (we did this separately for non-carnivoran and for carnivoran hosts, i.e. species from the order Carnivora). Likely host association rates were generated from a binomial distribution based on the number of observed host species and the number of species in the mammalian species pools (Supplementary Information, **Box S2)**. For these estimates, we again used species-level 'average' hyperpriors and fitted the model in a Bayesian framework to obtain posterior distributions. MCMC chain lengths and model checking procedures were as above.

Elucidating drivers of host associations with multiple regression

To further explore intermediate host traits that may act as drivers of parasite infection, we used hierarchical logistic regression to test which host attributes most likely predict host association probability with any of the focal parasites. Association with the focal parasite was included as a binary response ('1' if a species has been recorded as host species; '0' if no association has been recorded). Predictor variables included ecological trait variables and numbers of published papers, which were log-transformed if featuring overdispersion (body mass, range area, number of publications) and scaled (dividing centred values by one SD) to facilitate comparison of effect sizes. To account for underlying phylogenetic relationships between host species, we modelled variance-covariance relationships based on phylogenetic distance matrices using a multivariate-normal error structure (i.e. a phylogenetic generalised linear model; see Supplementary Information, **Box S3**). We also fitted these regression models in a Bayesian framework using MCMC sampling. We chose this approach as hierarchical models can combine dissimilar types of data (i.e. multiple numerical and categorical covariates, together with phylogenetic distance matrices) in a consistent probabilistic framework, and can accommodate for missing data through imputation. Specific trait data, for example, are currently not available for a considerable diversity of mammalian

species; we imputed missing values of ecological trait covariates during MCMC updates, randomly drawing values from priors according to the mean and variance of all observed trait values (considering all information in the trait databases) from species in the same taxonomic orders. Model fitting and assessment was conducted as specified above.

Results

The nine focal parasite species were recorded in 52 – 80 different mammalian host species and across 4 – 10 different zoogeographical regions (**Table S1**). All parasite species have been recorded infecting humans and, apart from *Hymenolepis diminuta* and *Versteria mustelae*, were also recorded infecting domestic dogs (*Canis familiaris*). *Hymenolepis diminuta*, which uses insects as intermediate hosts, was the only focal parasite species not recorded infecting host species from the order Carnivora.

We found considerable evidence that host-parasite interactions vary across zoogeographical regions. For all species, we detected significant variation (represented as non-overlapping credible intervals) in the estimated host association rates of non-carnivoran species pools between different regions (**Figure 1**). Host association rates of carnivoran hosts also exhibited some regional variation, notably including relatively high host association rates of up to 36% (95% CI: 25 – 52%) for *Mesocestoides lineatus* in the Neotropical region, as well as host association rates of 35% in the Saharo-Arabian and 38% in the Neotropical regions (both 95% CI: 24 – 51%) for *Trichinella spiralis*. Regional host association rates for non-carnivoran and carnivoran host species were positively correlated for *Echinococcus granulosus*, *Hydatigera taeniaeformis* and *Trichinella spiralis* (all Spearman rank correlations $r > 0.7$, $p < 0.05$), respectively, indicating that increasing regional numbers of definitive host species resulted also in increasing numbers of intermediate host species.

Across all regions, estimated host association rates were highest for *Hymenolepis diminuta*, with an overall average of 11% (95% CI: 5 – 21%) (**Figure 1**).

Regional variation in phylogenetic and functional host diversity

Differences between observed and expected phylogenetic diversity of non-carnivoran host assemblages showed considerable variation across regions. Overall, the strongest evidence for phylogenetic host specificity was found for *Hydatigera taeniaeformis* and *Hymenolepis diminuta* (negative values for global hyperpriors H_β(*phyl*) of -0.08 with 95% CI of -0.12 – -0.03 and -0.10 with 95% CI of -0.14 – -0.06, respectively; suggesting these parasites infected hosts that were more closely related than expected) (**Figure 2**). In contrast, we found higher than expected phylogenetic host diversity (indicating broadly generalist parasites) for *E. granulosus* in the Oriental region and for *Trichinella spiralis* in the Nearctic, Neotropical and Australian regions (**Figure 2**). Overall, observed – expected regional phylogenetic host diversity estimates were higher (closer to zero, given that most effects are reported to be < 0) in the Saharo-Arabian, Neotropical and Australian regions than the global average (across all parasites, according to results from linear regression analysis), indicating that host specificity appears to be lowest in these regions. Notably, these regions comprised those with the highest variation in the phylogenetic diversity of regional mammalian species pools (**Figure 3**).

Comparisons of observed and expected functional host diversity indicated that the highest host specificities were attributed to *E. granulosus*, *M. lineatus*, *Taenia hydatigena* and *Trichinella spiralis* (all H_β(*funct*) estimates smaller than zero, suggesting these parasites infect hosts that are more functionally similar than expected). In contrast, low functional host specificity was recorded for *Hydatigera taeniaeformis* (H_β(*funct*) of 0.03 with 95% CI of 0.01 – 0.06) (**Figure 2**). Observed – expected functional host diversity exhibited statistically significant regional variation for some species (i.e. *E. multilocularis*, *Hydatigera*

taeniaeformis, *Hymenolepis diminuta* and *Versteria mustelae*) (**Figure 2**), but no regional trends were detectable when considering all parasites together (according to the linear regression model). The overall functional diversity of regional mammalian species pools exhibited particularly low variation in the Panamanian and Oceanian regions (**Figure 3**).

Host attributes driving helminth parasite association probability

We identified only a few host attributes that influenced parasite association probabilities (**Figure 4**). The probability of a potential host to be associated with *E. granulosus* increased with increasing body mass. Host association probability for *E. multilocularis* increased with host longevity, but decreased with increasing average temperature within host ranges. For *Taenia hydatigena*, association probability increased with increasing latitudinal centroid of host species distributions. Large credible intervals for most parameters suggest that accurate trait-based prediction of association frequency is limited using the current data and model. Collectively, a much broader range of host attributes would have revealed 'statistically significant' effects if underlying host phylogenetic relationships had been ignored (see **Figure S1**). This suggests that, despite propensities for some parasites to infect a high diversity of host species, host association probabilities are still strongly driven by host phylogenetic relationships (**Figure 5**).

Discussion

Understanding drivers of parasite spillover is key to mitigating parasite transmission and the health impacts of parasitic disease. Using a global database of helminth parasite interactions with mammalian hosts, we show that realised phylogenetic and functional host specificities differ between zoogeographical regions and across parasite species. Our study sheds valuable light on the extent to which host selection and specificity vary depending on regional species

pools. We provide a framework to study host selection from databases of host-parasite associations, offering quantitative insights into host shifting patterns for widespread parasites. This information will provide useful new insights into how different parasite species may spread across global scales and in response to distinct regional host species pools.

Our findings provide a greater understanding of host specificity for parasites of veterinary and medical significance. The cestodes *Hydatigera taeniaeformis* and *Hymenolepis diminuta*, for example, exhibited clear phylogenetic host specificity. In contrast, higher than expected phylogenetic diversity of non-carnivoran hosts for the nematode *Trichinella spiralis* suggests that host shifting by this parasite can involve distantly related species in some regions. Overall, we found the lowest phylogenetic host specificity in regions with the greatest variation in prospective host phylogenetic diversity, namely the Neotropical, Saharo-Arabian and Australian regions (**Figure 3**). Due to the presence of taxonomically unique endemic species, these regions contain pairs of closely, as well as distantly, related species.

Functional host specificity exhibited significant variation among regions for some parasites (*Echinococcus multilocularis*, *Hydatigera taeniaeformis*, *Hymenolepis diminuta* and *Versteria mustelae*), but no common geographical trends. We found lower than expected host functional diversity for *E. granulosus*, *M. lineatus*, *Taenia hydatigena* and *Trichinella spiralis*, indicating some degree of host specificity and suggesting a general tendency to switch to novel hosts with similar ecological niches. In contrast, higher than expected functional host diversity for *Hydatigera taeniaeformis* suggests that, for this cosmopolitan tapeworm, the ecological niche of a prospective host species is not a strong determinant of host shifting. This is in line with the broad diversity of intermediate hosts recorded previously for this parasite (Lavikainen et al. 2016). Notably, given the positive correlation in host association rates between non-carnivoran and carnivoran hosts observed for this tapeworm, a

plausible explanation for this pattern is that the diversity of carnivoran hosts (mainly felids) with varying prey species in their diet facilitates host shifting to intermediate hosts across different ecological niches. Alternatively, a large diversity of intermediate host species from different ecological niches is also possible if the definitive host has access to a large diversity of prey items, particularly if transmission involves only a single carnivore host species in certain regions.

Transmission cycles through novel host communities – feral or endemic species?

Feral and invasive species are increasingly recognised as major agents for large-scale parasite spread (Adlard et al. 2015, Blackburn and Ewen 2017, Hulme 2014, Wells et al. 2015). In our study, domestic and commensal animal species, such as dogs, cats and commensal rats (genus *Rattus*), serve as important host species and likely play major roles in spreading helminth parasites worldwide. While it is difficult to identify the geographical origins of parasites and their ancestral/original host species, the exploration of novel host communities in exotic ranges allows us to ask whether transmission cycles are maintained by feral or endemic host species. If endemic wildlife strongly contributed to increased phylogenetic diversity in prospective host pools, parasites infecting a lower than expected host phylogenetic diversity (i.e. *Hydatigera taeniaeformis*, *Hymenolepis diminuta* and *C. hepaticum*, in some regions) may not be capable of regularly shifting to endemic wildlife species. Interestingly, we found a lower than expected host phylogenetic diversity for *E. granulosus* in the Nearctic, Palaearctic, Panamanian and Afrotropical regions, but an opposite pattern in the Oriental region (**Figure 2**). In this particular region, domestic dogs represent the only recorded carnivoran host of *E. granulosus*, whereas a broad diversity of herbivorous/omnivorous mammals are recorded as intermediate hosts (including small mammals such as the squirrel *Ratufa indica*, ungulates such as *Bos taurus*, and the Asian elephant *Elephas maximus*).

Given the well-established transmission cycle of *E. granulosus*, which involves canine definitive hosts consuming infected tissues of intermediate hosts, this example provides a clear illustration that host shifting to distantly related intermediate host species is driven by diet diversity of carnivoran host species (i.e. parasites will be unable to complete their transmission if they infect intermediate hosts that are not consumed by an appropriate definitive host). This is especially true for free-roaming domestic dogs and cats, whose prey spectra likely increase with their ongoing encroachments into natural habitats, which is often facilitated by human landscape conversion and environmental modifications (Baker et al. 2005, Doherty et al. 2016, Young et al. 2011). Likewise, domestic pet access to animals sourced by humans, through hunting, the meat industry, agriculture or exotic pets, may facilitate parasite transmission to a diversity of species if intermediate host tissue with vital larval parasite stages are ingested and transmission is enabled (Jones et al. 2013, Salb et al. 2008). Contact and interaction opportunities between feral and endemic animals may therefore play an important role in parasite spread at the human-domestic animal-wildlife interface.

Host attributes driving association risk

Predicting parasite spread requires an understanding of host attributes that enable host shifting (Han et al. 2016, Krasnov et al. 2010, Wells et al. 2015). Notably, if host shifting among intermediate hosts is largely a consequence of sharing the same definitive host predator, attributes that successfully predict infection risk can be linked to a suite of factors that facilitate parasite survival and transmission from intermediate host species (i.e. through ecological fitting), in addition to factors that determine whether these species are suitable prey. We found that host traits predicted infection risk for only three of the nine focal parasites species when accounting for host phylogeny. Positive correlations with body mass

(*E. granulosus*) and longevity (*E. multilocularis*) indicate some impact of host demographic traits, whereas a negative correlation with average temperature within host ranges (*E. multilocularis*) and a positive correlation with latitudinal centroid of host distributions (*Taenia hydatigena*) suggest some geographical constraints. Host longevity has been suggested as an important trait for helminth parasite infection, as sufficiently long lifespans are needed for parasite within-host development and transmission (Morand and Harvey 2000). Surprisingly, features of habitat utilisation were not identified as predictors of infection risks in our study. In previous work, we found that two commensal rat species (*Rattus rattus* species complex and *R. norvegicus*) most intensively share helminth parasites with wildlife species that are of least conservation concern (Wells *et al.*, 2015), which are likely those species well adapted to anthropogenically modified landscapes. In another study, we showed that the risk of being infested by the cosmopolitan cat flea (*Cenocephalides felis*), which is arguably one of the most widespread mammalian ectoparasites, appears to be greater for mammal species inhabiting anthropogenic environments (Clark et al. 2018). However, whether such correlates are chiefly driven by ecological fitting that enables parasites to thrive in different hosts, or, alternatively, by trophic interactions being concentrated within certain habitats, is difficult to resolve using traits-based regression analyses. As most of our focal parasites use dogs and other invasive carnivores as definitive hosts, the vast range of habitats explored by these carnivores might explain the absence of stronger habitat effects for predicting intermediate host infection risk. Further research into this topic is warranted. In particular, unravelling drivers of infection patterns and host shifting for parasites with different life histories, geographical distributions and transmission patterns could be an interesting and fruitful research avenue.

Scaling issues in studying specialisation and realised regional niches

Our study is based on a large global database of host-parasite associations in different regions/countries of the world. Although such data are useful for elucidating large-scale macroecological patterns (Stephens et al. 2016, Wells et al. 2018), the necessary level of data pooling and/or lack of more detailed data comes at the cost of neglecting fine-scale patterns in species occurrences and interactions. Within zoogeographical regions, ranges of species do not necessarily overlap. Therefore, species considered in regional species pools in our study do not necessarily have sympatric occurrences in local communities. Moreover, landscape structuring into different habitat types may further drive possible contacts and interactions between host species. Gradients in habitats from urban to remote natural vegetation are strong drivers of the structure of local host communities and parasite transmission pathways (Liccioli et al. 2015, Wells et al. 2014). While our approach provides insights into regional host assemblages, studies at a finer scale are necessary to determine realised host specificity in local communities with truly sympatric species.

Sampling bias and cryptic species

The compilation of data from a large range of studies and references into quantitative metanalysis does not, unfortunately, allow rigorous evaluation of data accuracy. Nor does it account for sampling bias the way that systematically conducted studies can, particularly those with full records of sampling efforts and depositions of voucher specimens. Parasites are likely to be overlooked in host species with low sampling intensity, especially if parasite prevalence is low (Little 2004, Walther et al. 1995). Some records of host-parasite associations may be accidental records (‘false positives’) of parasites that have been recorded in a host species in which they cannot complete their life cycle. Moreover, references to parasite scientific names may involve some level of misclassification and crypto-diversity (de León and Nadler 2010); in the case of *Hydatigera taeniaeformis*, for example, it has been

suggested that different subspecies circulate in different hosts (Lavikainen et al. 2016). Modern molecular tools may shed further light on different lineages and subspecies of the examined parasite species. In the future, this may also refine our picture of host selection and parasite spread amid the challenge to collect sufficiently large datasets and suitable specimens for global comparative studies.

Data deposition

The data are available from the freely accessible databases cited in the manuscript.

Acknowledgement – We are grateful to the Natural History Museum, London, for access to records of host-parasite associations and for making these data publicly available.

Funding – Parts of the work was supported by a National Geographic Society grant (WW-211R-17) to NJC and KW.

Author contributions – KW and NJC conceived the idea for this study. KW designed the study, carried out the analysis and wrote the first draft. All authors interpreted results and contributed to revisions.

References

Adlard, R. D. et al. 2015. The butterfly effect: parasite diversity, environment, and emerging disease in aquatic wildlife. — Trends Parasitol. 31: 160-166.

Agosta, S. J. et al. 2010. How specialists can be generalists: resolving the "parasite paradox" and implications for emerging infectious disease. — Zoologia 27: 151-162.

497 Agosta, S. J. and Klemens, J. A. 2008. Ecological fitting by phenotypically flexible
498 genotypes: implications for species associations, community assembly and evolution. —
499 Ecol. Lett. 11: 1123-1134.
500 Baker, P. J. et al. 2005. Impact of predation by domestic cats *Felis catus* in an urban area. —
501 Mammal Review 35: 302-312.
502 Begon, M. et al. 2002. A clarification of transmission terms in host-microparasite models:
503 Numbers, densities and areas. — Epidemiol. Infect. 129: 147-153.
504 Blackburn, T. M. and Ewen, J. G. 2016. Parasites as drivers and passengers of human-
505 mediated biological invasions. — EcoHealth 14: 61-73.
506 Brooks, D. R. et al. 2006. Ecological fitting as a determinant of the community structure of
507 platyhelminth parasites of anurans. — Ecology 87: S76-S85.
508 Canard, E. F. et al. 2014. Empirical evaluation of neutral interactions in host-parasite
509 networks. — Am. Nat. 183: 468-479.
510 Clark, N. J. and Clegg, S. M. 2017. Integrating phylogenetic and ecological distances reveals
511 new insights into parasite host-specificity. — Mol. Ecol. 26: 3074-3086.
512 Clark, N. J. et al. 2017. Climate, host phylogeny and the connectivity of host communities
513 govern regional parasite assembly. — Diversity Distrib. 24: 13-23.
514 Clark, N. J. et al. 2018. Parasite spread at the domestic animal - wildlife interface:
515 anthropogenic habitat use, phylogeny and body mass drive risk of cat and dog flea
516 (*Ctenocephalides* spp.) infestation in wild mammals. — Parasites & Vectors 11: 8,
517 doi:10.1186/s13071-017-2564-z.
518 Cooper, N. et al. 2012. Phylogenetic host-specificity and understanding parasite sharing in
519 primates. — Ecol. Lett. 15: 1370-1377.
520 de León, G. P.-P. and Nadler, S. A. 2010. What we don't recognize can hurt us: a plea for
521 awareness about cryptic species. — J. Parasitol. 96: 453-464.

522 Davidson, R. K. et al. 2012. The impact of globalisation on the distribution of *Echinococcus*
523 *multilocularis*. — Trends in Parasitology 28: 239-247.
524 Devictor, V. et al. 2010. Defining and measuring ecological specialization. — J. Appl. Ecol.
525 47: 15-25.
526 Doherty, T. S. et al. 2016. Invasive predators and global biodiversity loss. — Proc. Natl.
527 Acad. Sci. USA 113: 11261-11265.
528 Doña, J. et al. 2017. Host-specificity, infrequent major host switching and the diversification
529 of highly host-specific symbionts: The case of vane-dwelling feather mites. — Global
530 Ecol. Biogeogr. doi: 10.1111/geb.12680.
531 Dray, S. and Dufour, A.-B. 2007. The ade4 package: implementing the duality diagram for
532 ecologists. — J. Stat. Softw. 22: 1-20.
533 Fritz, S. A. et al. 2009. Geographical variation in predictors of mammalian extinction risk:
534 big is bad, but only in the tropics. — Ecol. Lett. 12: 538-549.
535 Gibson, D. I. et al 2005. Host-parasite database of the Natural History Museum, London.
536 http://www.nhm.ac.uk/research-curation/scientific-resources/taxonomy-systematics/host-
537 parasites/index.html
538 Gower, J. C. 1971. A general coefficient of similarity and some of its properties. —
539 Biometrics 857-871.
540 Han, B. A. et al. 2016. Global patterns of zoonotic disease in mammals. — Trends Parasitol.
541 32: 565-577.
542 Hassell, J. M. et al. 2017. Urbanization and disease emergence: dynamics at the wildlife-
543 livestock-human interface. — Trends Ecol. Evol. 32: 55-67.
544 Hatcher, M. J. et al. 2012. Disease emergence and invasions. — Funct. Ecol. 26: 1275-1287.

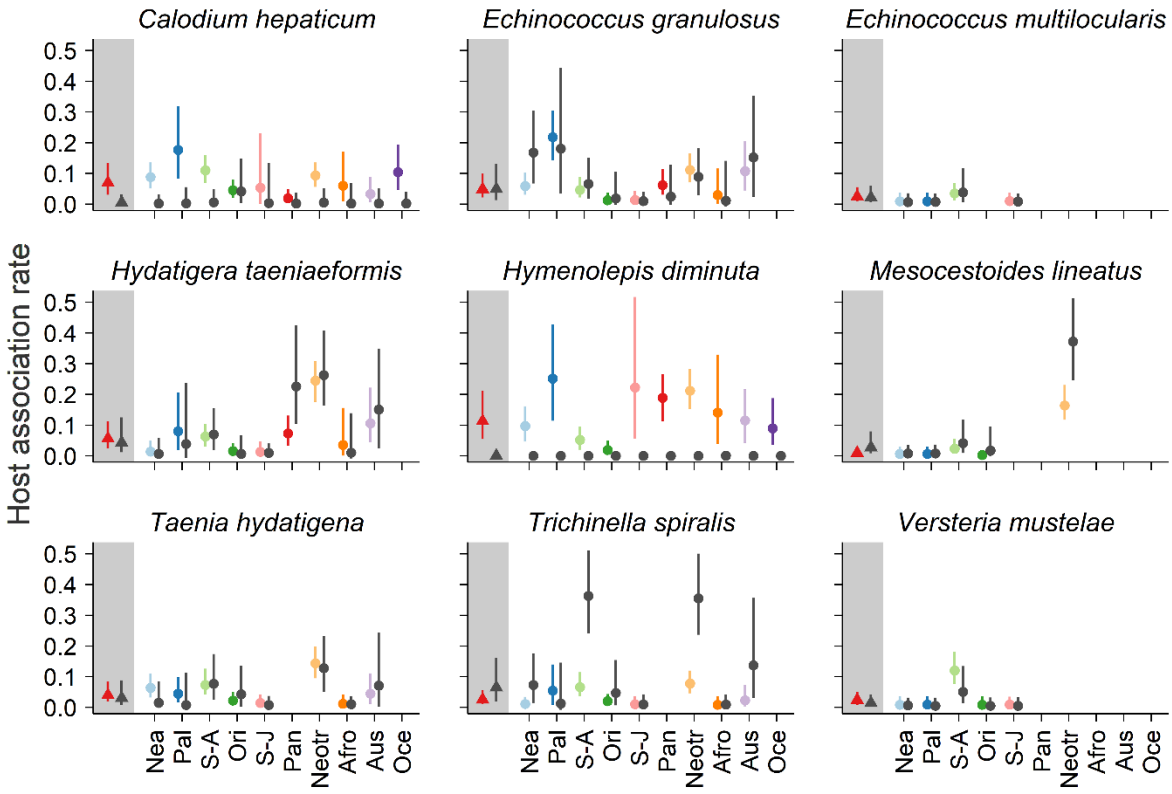
545 Hoberg, E. P. and Brooks, D. R. 2008. A macroevolutionary mosaic: episodic host-switching,
546 geographical colonization and diversification in complex host–parasite systems. — J.
547 Biogeogr. 35: 1533-1550.
548 Holt, B. G. et al. 2013. An update of Wallace's zoogeographic regions of the world. —
549 Science 339: 74-78.
550 Hulme, P. E. 2014. Invasive species challenge the global response to emerging diseases. —
551 Trends Parasitol. 30: 267-270.
552 Janzen, D. H. 1985. On ecological fitting. — Oikos 45: 308-310.
553 Jenkins, D. J. 2006. *Echinococcus granulosus* in Australia, widespread and doing well! —
554 Parasitol. Int. 55: S203-S206.
555 Jones, K. E. et al. 2009. PanTHERIA: a species-level database of life history, ecology, and
556 geography of extant and recently extinct mammals. — Ecology 90: 2648-2648.
557 Jones, B. A. et al. 2013. Zoonosis emergence linked to agricultural intensification and
558 environmental change. — Proc. Natl. Acad. Sci. USA 110: 8399-8404.
559 Kraft, N. J. B. et al. 2007. Trait evolution, community assembly, and the phylogenetic
560 structure of ecological communities. — Am. Nat. 170: 271-283.
561 Krasnov, B. R. et al. 2004. Geographical variation in host specificity of fleas (Siphonaptera)
562 parasitic on small mammals: the influence of phylogeny and local environmental
563 conditions. — Ecography 27: 787-797.
564 Krasnov, B. R. et al. 2010. Similarity in ectoparasite faunas of Palaearctic rodents as a
565 function of host phylogenetic, geographic or environmental distances: Which matters the
566 most? — Int. J. Parasitol. 40: 807-817.
567 Lavikainen, A. et al. 2016. Reappraisal of *Hydatigera taeniaeformis* (Batsch, 1786) (Cestoda:
568 Taeniidae) sensu lato with description of *Hydatigera kamiyai* n. sp. — Int. J. Parasitol.
569 46: 361-374.

570 Liccioli, S. et al. 2015. Wilderness in the 'city' revisited: different *urbes* shape transmission of
571 *Echinococcus multilocularis* by altering predator and prey communities. — Trends
572 Parasitol. 31: 297-305.
573 Little, R. J. 2004. To model or not to model? Competing modes of inference for finite
574 population sampling. — J. Am. Stat. Assoc. 99: 546-556.
575 Lootvoet, A. et al. 2013. Patterns and processes of alternative host use in a generalist parasite:
576 insights from a natural host–parasite interaction. — Funct. Ecol. 27: 1403-1414.
577 Malcicka, M. et al. 2015. Multi level ecological fitting: indirect life cycles are not a barrier to
578 host switching and invasion. — Global Change Biol. 21: 3210-3218.
579 Morand, S. and Harvey, P. H. 2000. Mammalian metabolism, longevity and parasite species
580 richness. — Proc. R. Soc. B 267: 1999-2003.
581 Morand, S. et al. 2006. Micromammals and macroparasites. From evolutionary ecology to
582 management. — Springer-Verlag, Tokyo.
583 Murray, K. A. et al. 2015. Global biogeography of human infectious diseases. — Proc. Natl.
584 Acad. Sci. USA 112: 12746-12751.
585 Paradis, E. et al. 2004. APE: analyses of phylogenetics and evolution in R language. —
586 Bioinformatics 20: 289-290.
587 Park, A. W. et al. 2018. Characterizing the phylogenetic specialism–generalism spectrum of
588 mammal parasites. — Proc. R. Soc. B 285: 20172613.
589 Pavoine, S. et al. 2009. On the challenge of treating various types of variables: application for
590 improving the measurement of functional diversity. — Oikos 118: 391-402.
591 Plummer, M. 2016. rjags: Bayesian Graphical Models using MCMC. R package version 4-6.
592 https://CRAN.R-project.org/package=rjags
593 Poisot, T. et al. 2017. Hosts, parasites and their interactions respond to different climatic
594 variables. — Global Ecol. Biogeogr. 26: 942-951.

595 Poulin, R. et al. 2011. Host specificity in phylogenetic and geographic space. — Trends
596 Parasitol. 27: 355-361.
597 R Development Core Team 2017. R: A language and environment for statistical computing.
598 R Foundation for Statistical Computing. https://www.R-project.org/
599 Revell, L. J. 2011. phytools: an R package for phylogenetic comparative biology (and other
600 things). — Methods Ecol. Evol. 3: 217-223.
601 Salb, A. L. et al. 2008. Dogs as sources and sentinels of parasites in humans and wildlife,
602 northern Canada. — Emerg. Infect. Dis. 14: 60-63.
603 Sokolow, S. H. et al. 2015. Reduced transmission of human schistosomiasis after restoration
604 of a native river prawn that preys on the snail intermediate host. — Proc. Natl. Acad. Sci.
605 USA 112: 9650-9655.
606 Stephens, P. R. et al. 2016. The macroecology of infectious diseases: a new perspective on
607 global-scale drivers of pathogen distributions and impacts. — Ecol. Lett. 19: 1159–1171.
608 Wallace, A. R. 1876. The geographical distributions of animals, with a study of the relations
609 of living and extinct faunas as elucidating the past changes of the Earth's surface. —
610 Macmillan.
611 Walther, B. A. et al. 1995. Sampling effort and parasite species richness. — Parasitol. Today
612 11: 306-310.
613 Wells, K. et al. 2014. Shifts from native to invasive small mammals across gradients from
614 tropical forest to urban habitat in Borneo. — Biodiv. Conserv. 23: 2289-2303.
615 Wells, K. et al. 2015. The importance of parasite geography and spillover effects for global
616 patterns of host–parasite associations in two invasive species. — Diversity Distrib. 21:
617 477-486.
618 Wells, K. et al. in press. Global spread of helminth parasites at the human – domestic animal
619 – wildlife interface. — Global Change Biol. doi: 10.1111/gcb.14064

620 Wilman, H. et al. 2014. EltonTraits 1.0: Species-level foraging attributes of the world's birds
621 and mammals. — Ecology 95: 2027-2027.
622 Wood, J. L. N. et al. 2012. A framework for the study of zoonotic disease emergence and its
623 drivers: spillover of bat pathogens as a case study. — Philosoph. Trans. R. Soc. B 367:
624 2881-2892.
625 Young, J. K. et al. 2011. Is wildlife going to the dogs? Impacts of feral and free-roaming dogs
626 on wildlife populations. — BioScience 61: 125-132.

627
628
629
630
631
632



633

Figure 1. Relative host association rates of regional mammalian species pools by some of the most globally widespread helminth parasites. Regional host association rates represent the proportion of sampled mammalian species that are infected, estimated separately for non-carnivoran (various colours for different regions) and carnivoran hosts (black). Points are posterior modes (estimated using a MCMC sampling from an underlying binomial distribution), bars are 95% credible intervals. In the shaded sections, triangles represent global 'averages' (hyperprior) for each parasite (red: non-carnivoran hosts, black: order Carnivora). Zoogeographical regions are denoted as Nea: Nearctic, Pal: Palaearctic, S-A: Saharo-Arabian, Ori: Oriental, S-J: Sino-Japanese, Pan: Panamanian, Neotr: Neotropical, Afro: Afrotropical, Aus: Australian, Oce: Oceanian.

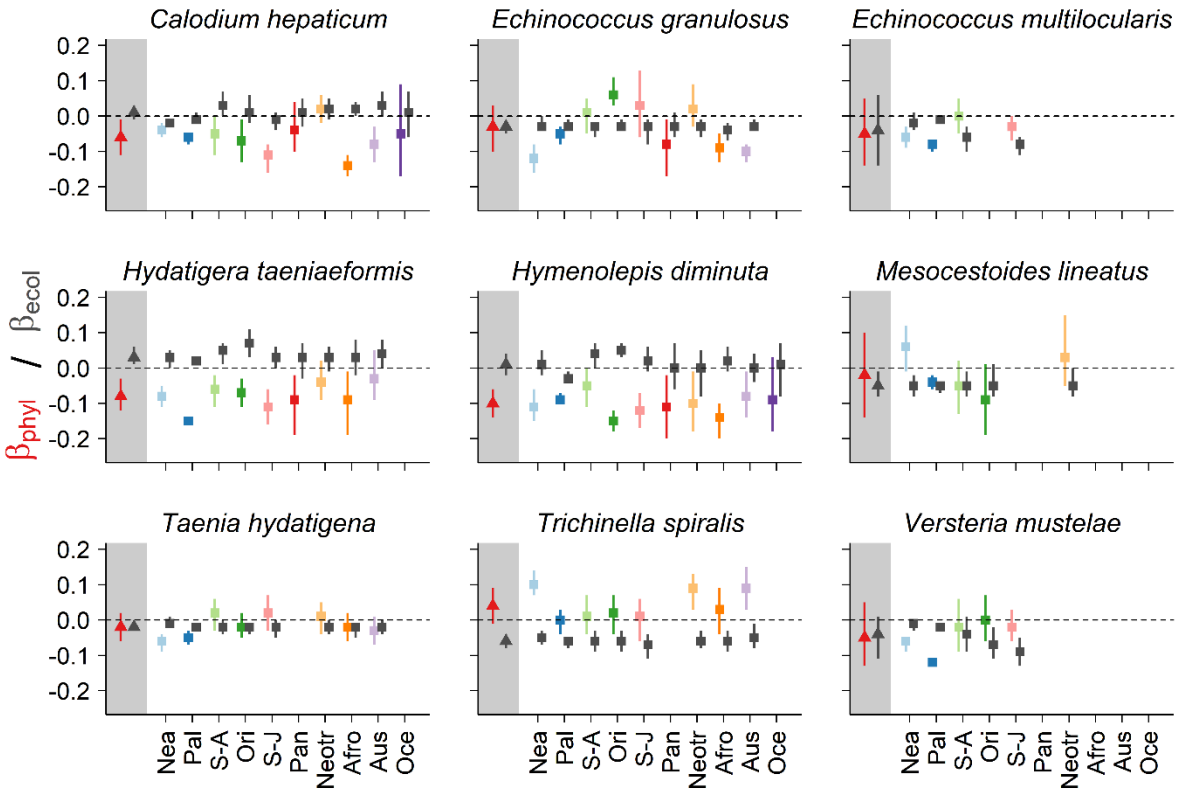


Figure 2. Relative difference between observed and expected phylogenetic (various colours for different regions) and functional (black) diversity of intermediate (non-carnivoran) host species as estimated from regression coefficients. Values < 0 indicate pairs of infected hosts were more phylogenetically/functionally similar than expected based on random draws from regional mammalian species pools, indicating high host specificity. Values > 0 indicate pairs of infected host species were more distantly related than expected, suggesting host generalism. Boxes represent posterior modes, bars 95% credible intervals. Triangles are global 'averages' (hyperprior) for each parasite (red: host phylogenetic diversity, black: host functional diversity). Zoogeographical regions are denoted as Nea: Nearctic, Pal: Palaearctic, S-A: Saharo-Arabian, Ori: Oriental, S-J: Sino-Japanese, Pan: Panamanian, Neotr: Neotropical, Afro: Afrotropical, Aus: Australian, Oce: Oceanian.

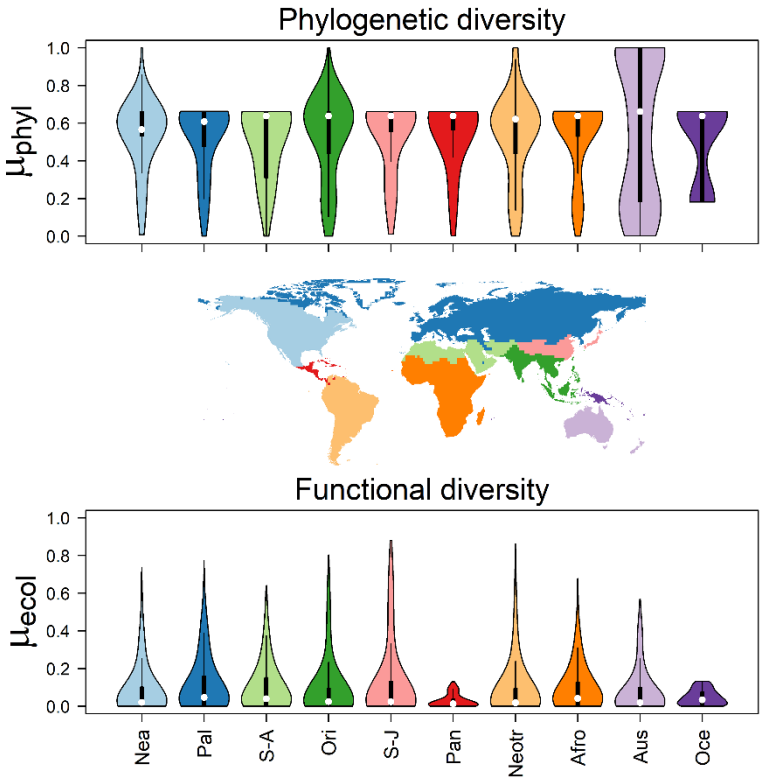
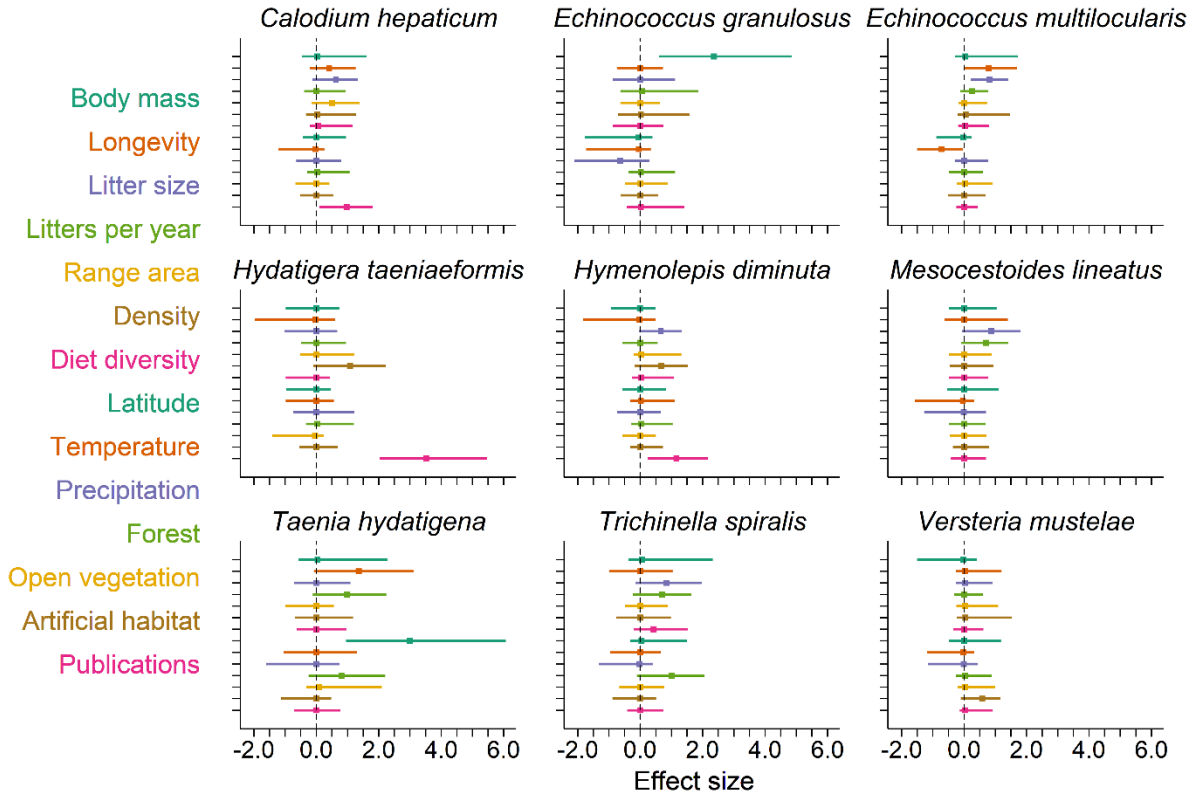


Figure 3. Distribution of the phylogenetic (μ_{phyl}) and functional diversities (μ_{ecol}) of prospective mammalian intermediate host species assemblages (excluding Carnivora) for some of the most globally widespread helminth parasite species. Each violin plot shows the range of respective diversity measures calculated from pairwise phylogenetic and functional distances for all combination of sampled mammal species in different zoogeographical regions (Nea: Nearctic, Pal: Palaearctic, S-A: Saharo-Arabian, Ori: Oriental, S-J: Sino-

673 Japanese, Pan: Panamanian, Neotr: Neotropical, Afro: Afrotropical, Aus: Australian, Oce:
674 Oceanian). Note that measures are restricted to species recorded in our host-parasite
675 databases and do not fully represent true distribution of entire communities.
676
677
678
679
680
681



682
683 **Figure 4.** Relative effect sizes of different covariates on the association probability of
684 intermediate (non-carnivoran) mammalian host species with some of the most invasive and
685 globally widespread helminth species. Note that carnivores were not included in the analysis
686 to focus on species that most likely serve as intermediate hosts. Points are posterior modes,
687 bars are 95% credible intervals.
688
689
690
691
692

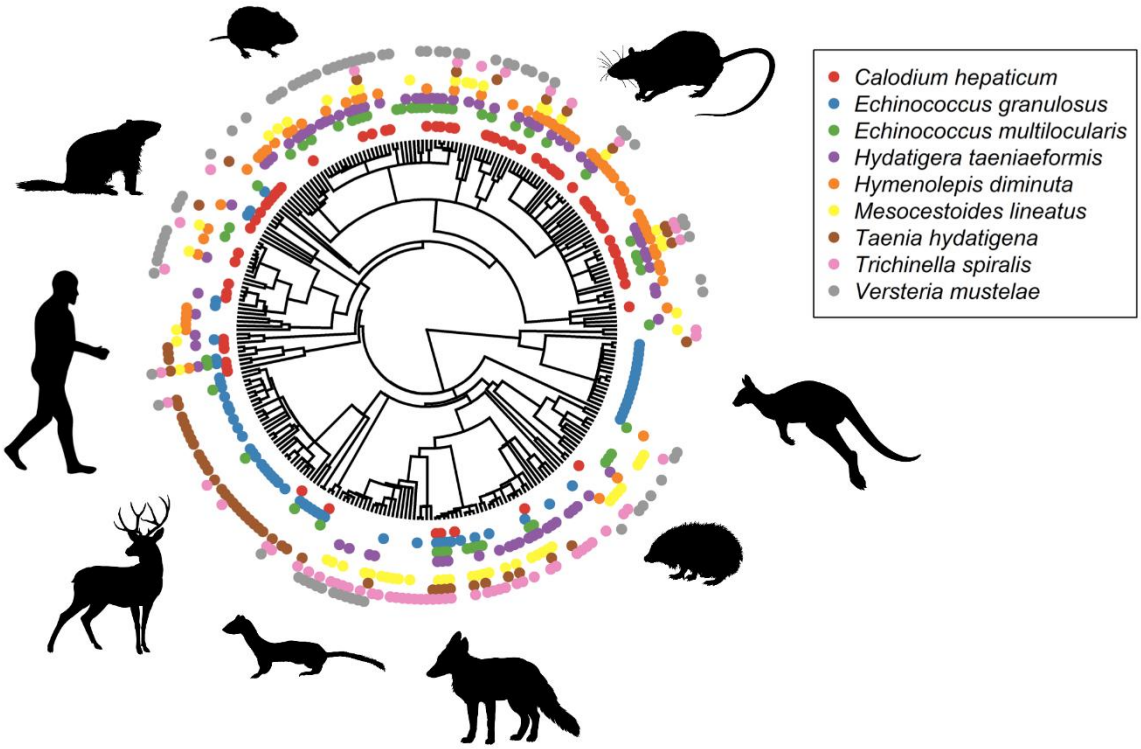


Figure 5. Infection of different mammalian host species with globally widespread helminth species across a phylogeny of the 276 species that have been recorded to be infected with at least one of the focal parasites. The position of major host groups in the phylogenetic tree are indicated with silhouette images (clockwise from top: rats (Muridae), kangaroos (Potoroidae), hedgehogs (Erinaceidae), foxes (Canidae), martens (Mustelidae), deers (Cervidae), man (*Homo sapiens*), marmots (Sciuridae), voles (Cricetidae). Images were sourced from http://www.supercoloring.com under a creative commons license.