Host allometry influences the evolution of host range: theory and meta-analysis

Josephine G. Walker\*, Amy Hurford, Jo Cable, Amy R. Ellison, Stephen J. Price, Mark Viney, and Clayton E. Cressler

**Keywords:** fish, macroparasites, generalism, specialism, evolutionary invasion

|  |  |  |
| --- | --- | --- |
| Name | Affiliations | ORCID |
| Jo Cable | School of Biosciences, Cardiff University, Cardiff CF10 3AX, UK | 0000-0002-8510-7055 |
| Clay Cressler | School of Biological Sciences, University of Nebraska-Lincoln, 424 Manter Hall, 1104 T St., Lincoln, NE 68588-0118, USA |  |
| Amy Ellison | School of Biosciences, Cardiff University, Cardiff CF10 3AX, UK |  |
| Amy Hurford | - Department of Biology, Memorial University of Newfoundland, St. John's, NL, A1B 3X9, Canada  - Department of Mathematics and Statistics, Memorial University of Newfoundland, St. John's, NL, A1C 5S7, Canada |  |
| Stephen J. Price | UCL Genetics Institute, Gower Street, London WC1E 6BT, UK | 0000-0001-6983-6250 |
| Mark Viney | - School of Biological Sciences, University of Bristol, Life Sciences Building, 24 Tyndall Avenue, Bristol, BS8 1TQ, UK |  |
| Josephine G. Walker | - School of Biological Sciences, University of Bristol, Life Sciences Building, 24 Tyndall Avenue, Bristol, BS8 1TQ, UK  - School of Social and Community Medicine, University of Bristol, Oakfield House, Oakfield Grove, Bristol, BS8 2BN, UK | 0000-0002-9732-5738 |

\*Author for correspondence (j.g.walker@bristol.ac.uk).

†Present address: School of Social and Community Medicine, University of Bristol, Oakfield House, Oakfield Grove, Bristol, BS8 2BN, UK

Summary

Parasites vary widely in host specificity – some parasite species infect just a single host species, while others are capable of infecting many, thereby being generalists. Understanding the factors that drive parasites' host specificity is of basic biological interest, but also directly relevant to predicting disease emergence in new host species, identifying parasites that are likely to have unidentified additional hosts, and assessing transmission risk. Here, we use mathematical models to investigate how variation in host body size, and environmental temperature affect the evolution of parasite host generalism. Parasites are more likely to evolve a generalist strategy when hosts are large-bodied, when variation in host body size is small, and in cooler environments. We then explore these predictions using a database of over 20,000 fish-macroparasite associations. Within the database we see some evidence supporting these predictions, but also highlight mismatches between theory and data. By combining these two approaches, we both establish a theoretical basis for interpreting empirical data on parasites' host specificity and identify key areas for future work.

Introduction

The range of hosts infected by a parasite species is a key factor affecting transmission. Parasites that have a diversity of host species may be able to persist in a population, where certain reservoir hosts are crucial for the maintenance of transmission [1,2]. In addition, parasites' host range is predicted to affect parasite virulence through mechanisms including relative host availability, maladaptive virulence, and fitness costs associated with infecting novel hosts [3]. The ability of a parasite to infect multiple host species, particularly across taxonomic orders, is a risk factor for emerging infectious diseases (EID) of humans and livestock [4].

Despite evidence for the importance of a parasite’s host range, it remains unclear whether most parasites are generalists or specialists, and what factors might influence the evolution of host range. EID studies, ranging from microbes to macroparasites, suggest that most parasites are generalists: 60% of human infectious diseases are zoonotic and 80% of pathogens of domestic animals infect multiple host species [4–6]. On the other hand, theory suggests that fitness trade-offs between host range and parasite performance on each host can lead to the evolution of host specialization [7], and such trade-offs are common (although not universal) for parasites [8-10]. Ecological specialization is often thought to represent an “evolutionary dead-end,” such that specialist parasites have a reduced potential to adapt to novel hosts, with correspondingly higher extinction and lower speciation rates than generalist parasites [11]. However, empirical evidence suggests that evolutionary transitions between specialism and generalism are bidirectional [12,13]. Rather few studies have examined whether specialism or generalism is the ancestral state for macroparasites of animals, but for both feather lice (arthropods) of doves and gill monogeneans (platyhelminthes) of African freshwater fish, host generalism appears to have derived from ancestral specialism [14,15].

Attempts to understand the ecological drivers of host range evolution have focused on a number of parasite or host traits and environmental factors (Table 1). For example, environmental change could lead to geographical isolation of a population, driving host/parasite co-evolution and so specialization towards a single host species, or to geographical expansion, driving opportunities for shifts to new hosts, and thus a wider host range [16] (and see Cable et al. this volume). In this study, we focus on understanding how host body size and environmental temperature affect host range.

The relationship between host size and parasitism has been explored in depth, often with reference to island biogeography theory (IBT). IBT predicts that the number of parasite species infecting a host will increase with host body size, as larger-bodied hosts represent larger habitat patches with more niches [17-19]. However, most work on this relationship has been host-centric. Few studies have considered the question from the parasite perspective, that is, do parasites that infect large-bodied hosts infect a greater diversity of hosts [20-22]? There are, however, several reasons to suspect that host body size might influence host range. Larger hosts support higher within-host parasite abundances [23], which may influence between-host transmission, for example, by positively or negatively affecting parasite shedding [24]. Host body size also affects key host characteristics, such as longevity and carrying capacity [25], that may affect host availability to parasites.

Temperature may also influence host range through a number of processes. For example, species diversity of both hosts and parasites tends to increase near the tropics [26], an increase that can be explained by increased temperature [27]. Host range may therefore increase with environmental temperature simply because there are more host species available to parasitize. Temperature can also affect parasite survival and infectivity during free-living stages, with different species and stages preferring different temperatures [28]. Given the importance of free-living stages to transmission for many parasites, temperature may then have an important affect on host range evolution. Finally, as with body size, temperature can affect important host characteristics that might affect host availability [25].

General predictions regarding correlations between host, parasite, and environmental characteristics and parasites' host specificity largely come from simple verbal models, and empirical tests of these predictions are often equivocal [10,29]. Here,,we use evolutionary invasion analysis [30] to predict when specialism/generalism is the evolutionarily stable strategy for parasites infecting a host population consisting of different host species. We use allometric scaling relationships to characterize the body size- and temperature-dependence of key host traits, and use the model to predict how host body size, temperature, and transmission mode affect the evolution of host range. We then calculate structural and phylogenetic generalism metrics [31] from an extensive data set of macroparasites of fish [32] to test these predictions. With this approach, we aim to improve our understanding of the ecological and evolutionary factors that contribute to parasite generalism.

Table 1: Examples of host and parasite traits predicted to affect the evolution of parasite generalism. Traits investigated in this paper are indicated by \*.

|  |  |  |
| --- | --- | --- |
| **Trait** | **Levels** | **Previous Hypotheses or Observations** |
| Host seeking behaviour\* | *Active*: e.g., mobile parasites that seek out hosts;  *Passive*: e.g., parasites transmitted during host-host contact or via ingestion | Parasites that actively seek out hosts should be more specific than parasites that are transmitted by direct contact between hosts. Parasites transmitted via ingestion should be less specific than parasites infecting through other routes [33]. |
| Infection site\* | *Endoparasite*:lives inside the host; *Ectoparasite*: lives on the surface of the host | Infection site will give different opportunities for transmission mode, for example the mobility of infective stages may affect the evolution of generalism [33].  Higher number of host species per parasite and network connectance observed for endoparasites compared to ectoparasites of fish [34]. |
| Life cycle | *Complex* - Transmission involves one or more intermediate hosts  *Direct* – no intermediate hosts | Parasites with complex life cycles exhibit more range in acceptable hosts and may be more likely to evolve generalism [33].  Direct life cycle parasites of primates are less host-specific than complex life cycle parasites [35]. |
| Trophic transmission\* | *Yes* - For parasites that have complex life cycles, trophic transmission occurs when the intermediate host is ingested by the terminal host  *No* - Transmission to the terminal host does not involve ingestion | Trophic transmission will restrict exposure of life stages to guilds within trophic levels, such that host-parasite associations track broadly and predictably across trophic levels because the completion of transmission in a complex system is dependent on the structure of food webs [36]. |
| Host geographic range as proxy for temperature\* | Geographic regions: Africa; Antarctica; Australia; Indopacific; Nearctic; Neotropical; Palearctic. | Allometric relationships exist between temperature and life history parameters [25]  Higher species diversity in the tropics [26,27]  Digenean parasites of marine fish in tropical seas infect fewer hosts than those that parasitize fish in colder seas [26].  No relationship is observed between latitude and generalism for Monogeneans [37]. |
| Host body size\* | Continuous (here, maximum length of fish host) | Specialist Monogenean parasites tend to be found on large-bodied fish hosts [20,21,38].  Variance in phylogenetic diversity of host species infected by fleas is negatively correlated with mean host body size [22]. |

**Model Derivation and Predictions*:***

We begin by defining a simple host-parasite system with two hosts and two environmentally transmitted parasites, since transmission is through the environment for macroparasites of fish. For ease of presentation, the model assumes a parasite with a direct life cycle; Appendix A shows a similar analysis for a trophically transmitted parasite. The first parasite is a specialist, infecting only the first (primary) host. The second parasite in a generalist, infecting both the primary host and the secondary host. In Appendix A, we also consider cases where there are three parasites: two parasites that are specialized on the primary and secondary hosts and a generalist parasite. We follow the dynamics of susceptible primary and secondary hosts , primary hosts infected by the specialist parasite , primary and secondary hosts infected by the generalist parasite , and specialist and generalist parasites in the environment . The dynamics of the system can be described by the following system of equations:

In the absence of any infection, the population sizes for primary and secondary hosts will be equal to the host-specific carrying capacities and . Infection occurs by parasites in the environment contacting hosts at the *per capita* rate , which is assumed to be equal across both hosts and parasites. Contact results in the removal of parasites from the environment – note that we assume that specialist parasites do not contact the secondary host. Moreover, both the specialist and generalist parasite only come in contact with susceptible hosts. Both of these assumptions suggest that the parasite controls the contact process. In Appendix A, we consider cases where we relax these assumptions. Infected primary and secondary hosts die at the host-specific rates and . Parasites are shed from infected hosts at the host-specific rates and . Note that the mortality and shedding rates do not depend on whether the host is infected by the specialist or generalist parasite. However, we assume that the cost of parasite generalism is that shedding rate by generalists is a fraction *a* of the shedding rate of specialist parasites.

We use evolutionary invasion analysis [30] to determine the conditions under which a generalist parasite can invade a system where the specialist parasite is present at equilibrium. Mathematically, this corresponds to investigating the stability of the equilibrium point , where the ^ denotes equilibrium values set by the specialist parasite. The Jacobian for that equilibrium is the block triangular matrix,

The stability of this system is determined by the eigenvalues of and . Because we have assumed that the specialist-only system reaches a stable equilibrium, all of the eigenvalues of have negative real part, so stability is determined entirely by the eigenvalues of .

Applying the next-generation matrix theorem [39], the specialist-only system will be unstable (i.e., generalism will evolve) whenever the invasion fitness of the generalist (which we will express as ) satisfies,

where and are the equilibrium host abundances when only the resident parasite is present. These terms have intuitive biological meanings: is the probability that a parasite in the environment infects a susceptible primary host and is the expected number of new generalist parasites produced per infected primary host; the second set of terms have an analogous interpretation for the secondary host. Thus, for generalism to evolve, each generalist parasite in the environment must be expected to produce more than one new generalist parasite in the environment; that is, a successfully invading generalist’s *R*0 will be >1.

While the expression in Eq. (1) is easy to understand, substituting in the equilibrium abundance of the primary and secondary host simplifies the expression to

One way to explore the consequences of host, parasite, or environmental characteristics for the evolution of generalism is to ask whether increasing the value of a parameter representing such a characteristic increases or decreases the value of , that is, to calculate the sign of the partial derivative of with respect to that parameter. If the sign is positive, then increasing the value of the parameter makes generalism more likely to evolve. For this simple model, this analysis makes very intuitive predictions – for example, increasing the abundance of the secondary host increases whereas increasing the environmental mortality rate decreases it (see Appendix A for more details). However, in its current state the model does not include any biologically relevant trade-offs that constrain the relationships between parameters. And, more to the point of this paper, it is challenging to connect the parameters of such a general model with empirical data on host-parasite associations.

To facilitate a comparison between the model and data, we take advantage of the fact that many key parameters of the model are likely to be allometric functions of host body size and temperature. In particular, host carrying capacities and mortality rates will scale with host body size [25] as

and

where is the Boltzmann factor, which describes how temperature affects reaction kinetics (e.g., metabolic rate), is the body mass of host *i,* and are proportionality constants. *E* is the average activation energy of rate-limiting biochemical metabolic reactions, *k* is Boltzmann’s constant, and *T* is temperature. Since our dataset deals with parasites of ectotherms, we assume that *T* is the temperature of the environment, and is the same for both hosts. Increasing mass will decrease the carrying capacity and mortality rate, whereas increasing temperature decreases carrying capacity and increases mortality rate.

Host body size and temperature should also affect parasite abundance, either within-host (for endoparasites) or on host surfaces (for ectoparasites), though the scaling of abundance with body size differs in these two cases [23]. We assume that shedding rate scales linearly with parasite abundance, giving

We add these expressions into the expression above to attain host body size-, temperature-, and infection site-dependent criteria for the evolution of generalism. In particular, it is immediately clear that, all else equal, will be larger for endoparasites than ectoparasites because shedding rate will be higher. Thus, generalism is more likely to evolve for endoparasites than ectoparasites.

For simplicity, we let the mass of the secondary host be , where is the mass of the primary host, and we assume that the secondary host is smaller than the primary host (). To investigate how the evolution of generalism is affected by host body size (*W*), the difference in body size between the two hosts (*f*), and the temperature of the environment (*T*), we ask how changes with changes in these parameters (that is, we look at the derivatives of with respect to *W*, *f*, and *T*). We will consider these derivatives for both endoparasites and ectoparasites.

For endoparasites, is an increasing function of host body size:

Thus we predict that parasites infecting large-bodied hosts are more likely to be generalists than parasites infecting small-bodied hosts. That is, when looking across a large number of host-parasite associations, **there should be positive correlations between a parasite’s generality and the body size of its largest host.** Moreover, **there will be a positive correlation between generalism index and mean body size**: mean body size is , and , where ., which must be positive.

Similarly, is an increasing function of *f*, the relative difference in body size between hosts:

This result is intuitive: increasing *f* increases the size of the secondary host; is the sum of terms dealing with infection in the primary and secondary host, and, as we have already shown, increasing host mass increases . Thus we predict that **there should be a strong positive correlation between generalism index and the coefficient of variation in host body size.** The coefficient of variation is a better metric for this prediction than the raw variance because the variance in body sizes among hosts will be positively correlated with mean body size among hosts.

For ectoparasites, the response of to changes in body size is more complicated. The effects of increasing host mass or increasing the difference in mass between hosts are given by the derivatives

and

For both of these derivatives, the sign is determined by . Plugging in the scaling functions for and , we find that increasing host body size will increase if

This indicates that it will be easier for a generalist ectoparasite to invade when host body size increases, but only up to a point. Put another way, **this predicts that there should be few generalist parasites of either very small bodied or very large bodied hosts**. If the primary host is very large, then it will be easier for a generalist to invade if the secondary host is much smaller (i.e., *f* is small). However, it is important to note that both of these predictions now depend on the values of the parameters, making these predictions somewhat more challenging to address.

The effect of temperature will be the same for both endo- and ectoparasites, as parasite infection site has no effect on temperature scaling. For both, increasing temperature decreases :

Thus we predict that **generalism should be more likely in colder environments than in warmer ones**. A corollary of this (which we cannot address in our current dataset) is that generalism should be more common among parasites of ectotherms than endotherms.

This model is intentionally simple. In Appendix A, we investigate the sensitivity of our predictions to the assumptions made by this model by considering alternative models that differed in 4 key ways: the number of specialist parasites; the effect of parasitism on host population growth; the control of parasite transmission; and the parasite’s life cycle. In the model presented above, we assumed that the secondary host was unexploited by any parasite, so we considered model variants that assumed that both the primary and secondary hosts were infected by a specialist parasite. The model above also assumes that the host population size is regulated by the parasite, so we considered models where total host population size is constant. In the model above, we assumed that the parasite has complete control of the infection process. For example, the specialist parasite is removed from the environment only by susceptible primary hosts. This assumes active host seeking by the parasite and that the parasite can detect and avoid already-infected hosts. We consider models that relax both of these assumptions. Finally, we also considered how the predictions change for a trophically transmitted parasite, when there is a single intermediate host that consumes parasites in the environment, and then transmits those parasites to either of two definitive hosts.

For all of these models, it is possible to write down the generalist’s in a form analogous to Eq. (1) above. The models differ in how that expression simplifies when the various equilibria are substituted. Table 2 shows these simplified expressions as well as our predictions for how host body size and temperature affect for both endo- and ectoparasites for models dealing with direct life cycle parasites. Table 3 shows similar results for trophically transmitted parasites. For this analysis, we only looked at endoparasites, since there are no trophically transmitted ectoparasites in our dataset. However, the complexity of this model prevents simple analytical results for most cases.

Table 2. Effect of increasing body size and temperature on the evolution of generalism for direct life cycle parasites under a range of assumptions.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Number of specialist parasites | Parasite regulates population growth? | Active or passive host seeking? | Avoidance of already infected hosts? | Generalist | Effect of increased body size on | Effect of increased temperature |
| 1 | Yes | Active | Yes |  | Increase (endoparasites)  Intermediate peak (ectoparasites) | Decrease |
| 2 | Yes | Active | Yes |  | Increase (both) | None |
| 2 | No | Active | Yes |  | Increase (both) | None |
| 2 | No | Active | No |  | Increase (both) | Increase |
| 2 | No | Passive | No |  | None (both) | None |

Table 3. Effect of increasing body size and temperature on the evolution of generalism for trophically transmitted parasites under a range of assumptions

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Number of specialist parasites | Parasite regulates population growth? | Active or passive host seeking? | Avoidance of already infected hosts? | Generalist | Effect of increased body size on | Effect of increased temperature |
| 1 | Yes | Active | Yes | No simple expression | Increase | Decrease |
| 2 | Yes | Active | Yes | No simple expression | Variable | Variable |
| 2 | No | Active | Yes | No simple expression | Variable | Variable |
| 2 | No | Active | No |  | None | None |
| 2 | No | Passive | No |  | None | None |

What these analyses reveal is that, for direct life cycle parasites, the effect of host body size is almost always to increase the value of , thereby making it easier for generalists to invade. This is because larger hosts support a larger parasite population size, thereby increasing shedding, and larger hosts have lower mortality rates, reducing parasite virulence. The effect of temperature on direct life cycle parasites is more complicated, and depends on the modelling assumptions.

For trophically transmitted parasites, on the other hand, the results are much more variable. This suggests that general patterns may be difficult to ascertain for trophically transmitted parasites.

Methods

*Data collection*

The Fish Parasite Ecology Database contains more than 38,000 records of associations between 4,650 host fish species and 11802 helminth parasites, as well as ecological, biogeographical, and phylogenetic information on the host species, including host body size and geographic region [32]. As the number of ectoparasite species was low, additional parasite-host records were included for 105 crustacean parasite species, and we included data on parasite life history traits including reproductive strategy, life cycle stages, and transmission routes from a range of primary literature sources. If there was any ambiguity regarding the taxonomic status of the parasites they were excluded from the database. To remove synonyms and other inconsistencies, host species names were quality-checked by Entrez Direct queries ([www.ncbi.nlm.nih.gov/books/NBK179288/)](http://www.ncbi.nlm.nih.gov/books/NBK179288/)) to the NCBI taxonomy database and FishBase [40]. Parasite species names were checked against NCBI taxonomy database in the same way and also checked against the NHM Host-parasite database (<http://www.nhm.ac.uk/research-curation/scientific-resources/taxonomy-systematics/host-parasites/database)> using a custom script and the World Register of Marine Species (WoRMS), Catalogue of Life (CoL), Integrated Taxonomic Information System (ITIS) and Global Names Index (GNI) databases through the Lifewatch Taxonomic Backbone (<http://www.lifewatch.be/data-services/)>. All intermediate hosts were excluded from the calculation of generalism metrics, such that generalism in parasites with complex life cycles was based on the definitive hosts only. After data cleaning, we were left with 23,331 unique host-parasite associations between 8,846 parasite species and 4,237 fish hosts.

*Host genetic distances*

Host mitochondrial DNA sequences (complete mitochondrial genomes and full or partial sequences from mitochondrial loci) were downloaded in fasta format from the NCBI nucleotide database using an Entrez Direct query. Sequences were discarded if the sequence header did not contain the species name (either full or abbreviated scientific name). They were then sorted by locus (limited to cytochrome oxidase 1, cytochrome B, 12s and 16s) based on regular expression matches to the sequence header and the assignment to loci checked by local BLASTN [41] searches (default settings, version 2.2.29) to databases of representative sequences from the relevant locus. When multiple sequences were available for a given locus and host, a consensus sequence was generated using the EMBOSS program *cons* [42]*.* The consensus sequence was used in downstream analyses except in cases when just two sequences had contributed to the consensus and the result contained more than 1% of variable positions. In such cases the first sequence was used instead of the consensus if a megaBLAST [41] search against the NCBI nucleotide database hit a member of the same genus with percentage identity greater than 90%, and the second sequence used otherwise. For hosts without sequence data for a particular locus, sequences were extracted from mitochondrial complete genomes when available using BLASTN. Consensus sequences for all available host species were combined and aligned - adjusting for direction - using MAFFT for each locus. Alignments were then trimmed with trimAl [43] to include only those columns where less than 50% of taxa had a gap and those taxa where 50% of the nucleotide positions had the same ‘element’ (e.g. a gap or a residue) as more than half of the other taxa in the alignment.

Trimmed alignments were used to compute pairwise genetic distances using the dist.dna function and the K80 model of DNA evolution [44] in the R package *ape* [45]. Since different taxa were represented among loci, the consistency of pairwise distances among loci across a range divergence times was assessed. The pairwise genetic distances between those host taxa with sequences for all loci were extracted. For each locus these pairwise genetic distances were plotted against their corresponding pairwise distance generated from cytochrome oxidase 1 sequences. The point patterns were compared to the one-to-one line (complete correspondence), and correspondence to this line used to select the loci to be concatenated using *ape*. Pairwise genetic distances were recalculated on this final concatenated alignment. Missing pairs were imputed using a custom R script by averaging according to the following relationships between taxa: 1) pairs from different genera – the mean of the genetic distances of one member of the pair (determined by the data available) to congeners of the remaining pair member was calculated, 2) pairs from the same genera – the mean genetic distance of all the other pairwise comparisons within that genus was calculated. Where no data were available for any member of a genus, the mean within genus average for all genera was used. If suitable data were unavailable the same principals were applied at increasing taxonomic levels (family, order, class) until values were obtained for all pairwise comparisons.

*Generalism metrics*

No information on abundance or prevalence of parasites within hosts was available within the original database (ref), so parasites’ host generalism metrics were defined according to structural and phylogenetic metrics, with phylogeny based on the pairwise genetic distances between hosts (Table 4).

Table 4: Generalism metrics calculated from host-parasite database [32]

|  |  |  |
| --- | --- | --- |
| Metric | Description | Facet |
| Degree | Number of hosts | structural |
| G | Binary measure, G=1 if degree > 1 | structural |
| SPD | Mean pairwise phylogenetic distance between all hosts [46], SPD = 0 for G=0. | phylogenetic |
| PD | Faith’s phylogenetic distance, minimum total length of all the phylogenetic branches required to span all hosts on the phylogenetic tree [47], PD = NA for G=0 | phylogenetic |
| SES-PD | Standardized Effect Size of PD based on 1000 runs, with a negative value indicating that the observed tree length is smaller (the hosts are more closely related) than what you might find by chance | phylogenetic |

A UPGMA tree was calculated from the full distance matrix using *phangorn* in R [48] and the tree reordered from root to tip so that edges from the root node were listed first (‘cladewise’ reordering). Our final list of host-parasite associations was converted to a binary ‘community data matrix’ with parasites as rows and hosts as columns (1 indicates a recorded association, 0 indicates no record).

Faith’s phylogenetic diversity index (PD; here, the length of the host tree with the root excluded) [31,47] was generated for all parasites using *picante* in R [49]. We used the ses.pd function implemented in *picante* to generate the standardized effect size of phylogenetic diversity (SES-PD) based on 1000 runs. SES-PD compares the actual Faith’s PD value for each parasite to a summary of the metric calculated after repeatedly shuffling taxa labels of all taxa in the phylogeny in order to assess if phylodiversity is high or low for a given number of hosts.

*Data analysis*

The generalism metrics for each parasite species were compared to parasite traits to test the model predictions. We separated out the analysis by parasite life cycle (direct vs. trophic), since the modeling results suggests that life cycle strong affects model predictions. We note, however, that life cycle and infection site are confounded in the dataset as nearly all of the direct life cycle parasites are ectoparasites (4216/4226) whereas nearly all of the trophically transmitted parasites are endoparasites (n/x).

Metrics for parasites with direct or trophic life cycles were compared to summary measures (mean, maximum, and coefficient of variation (CV)) of the lengths reported for each of their hosts. Note that CV of the host length is only calculated for parasites with more than one host.

As a rough proxy for temperature, the effect of geographic region on parasites with direct or trophic life cycles was calculated for endoparasites and ectoparasites together. Regions were assessed as defined in Table 1 and also divided into two groups, where Antarctica, Nearctic, and Palearctic were assumed to be colder than Africa, Australia, Indopacific, and Neotropical regions. Some host-parasite associations were reported in more than one region, so for the analysis based on geographic regions the generalism metrics were calculated separately for each region.

Because the generalism metrics come from very different distributions, we used GLMs with different error distributions for statistical analyses. For degree, we used negative binomial regression with a log link function (glm.nb() in the MASS package in R); for G, logistic regression (glm(family="binomial") in R); and for PD, SPD, and SES-PD, linear regression (lm() in R), with PD log transformed (log PD+0.01).

Comparing Model Predictions to Empirical Data

*Pairwise genetic distances between hosts*

The total number of host species in our dataset was 4621. Pairwise genetic distances calculated using the cytochrome oxidase 1 gene corresponded to those from cytochrome B (Fig. S1) and these loci were selected for concatenation and subsequent calculation of pairwise genetic distances. 3253 hosts were represented by cytochrome oxidase 1 sequences and 2193 hosts by cytochrome B. 1915 hosts had sequences from both loci whilst a total of 3531 hosts had representative sequences from one or both loci. The mean pairwise genetic distance between hosts was 0.263 (standard deviation=0.034; fig. S2).

*Infection site*

The allometric scaling model predicts that for parasites with a direct life cycle, generalism should be higher in endoparasites compared to ectoparasites. In the fish data set of macroparasites, there are 4226 parasites with a direct life cycle, of which only 10 (0.2%) are endoparasites. Due to the small sample size for endoparasites, no significant difference is found for generalism metrics by infection site (not shown).

*Host body size*

The model predicts that, for direct life cycle parasites, there should be a positive correlation among parasites’ generalism metrics and both the maximum and mean host body size, with a particularly strong positive correlation between generalism and the coefficient of variation in host body size. We observe a strong and significant positive correlation between many, but not all, body size metrics and generalism metrics (Fig. 1, Table x).

For trophically transmitted parasites, the model makes no definitive predictions. The correlation between generalism and body size can be positive or negative. Interestingly, however, we observe identical patterns of correlation for trophically transmitted parasites as we did for direct life cycle parasites (Fig. 2, Table y).

*Temperature/geographic range of all parasites*

In general, the models make very different predictions about how temperature affects the evolution of generalism. At least some models, however, predict that generalism is more likely in colder environments, and we see higher degree in cool regions for direct life cycle parasites (Fig. 3, Table z) and higher degree, G, and SPD in cool regions for trophically transmitted parasites (Fig. 4, Table a).

Discussion

The number of hosts a parasite can infect has important epidemiological and evolutionary implications [1–4]. Previous authors have approached the study of host range using a comparative approach, analysing groups of closely related parasites that differ in the number of hosts infected by species within the group to attempt to identify the key factors that influence host range [18-22]. On the basis of these studies, verbal models have been developed that suggest how such factors might influence the evolution of host range more generally (Table 1, and see reviews in refs. 10 and 29). For example, host range might be influenced by phylogenetic constraints if the fitness cost of infecting multiple hosts is lower when the hosts are closely related [50]. However, while these verbal models are intuitively appealing, empirical tests of their predictions are often equivocal [10,29].

Here, we take a different approach, deriving simple mathematical models that incorporate host, parasite, and environmental characteristics using principles from metabolic scaling theory [23,25]. This allows us to incorporate biologically feasible constraints on the epidemiological processes included in mathematical models of host-parasite interaction. We then use evolutionary invasion analysis [30] to study how variation in host body size, temperature, infection site, and parasite life cycle influence the evolution of host range, here quantified as the effect of these characteristics on the magnitude of a generalist parasite’s invasion fitness.

This mathematical approach can help illuminate the strengths and the weaknesses of verbal models for the evolution of host range. In particular, the dynamical interaction between hosts and parasites can have counterintuitive outcomes that affect the validity of verbal model predictions. For example, previous authors have suggested that a narrow host range (high host specificity) is more likely to evolve when hosts are abundant because increased abundance increases the probability that a specialist will encounter its host [51,52]. Our model analyses reveal that host abundance is unlikely to be directly relevant to the evolution of host range. This is because parasite fitness depends not on the *total* abundance of hosts, but on the abundance of *susceptible* hosts. The dynamic interaction between the host and parasite causes the abundance of susceptible hosts to depend on parasite traits rather than host traits like carrying capacity. Thus the fitness of the generalist does not depend directly on host abundance, which can be seen from the fact that carrying capacity rarely appears in the generalist values presented in Table 2. The exception is when the parasite actively seeks out hosts, but cannot discriminate between infected hosts (which are a dead-end for those parasites) and susceptible hosts. In that case, generalists are more easily able to invade when there is a high probability that a parasite will be lost from the system (i.e., is large). Turning that prediction around, specialist parasites will be favoured when it is likely that they are able to come in contact with a host, as suggested by the verbal theory. Thus, by analysing the question mathematically, we come to a more complete understanding of when an intuitive verbal prediction is likely to apply.

There are, however, important challenges in attempting to test the predictions of simple mathematical models using data from real host-parasite systems. In particular, theory on the evolution of specialization indicates that the crucial determinant of host range is the trade-off between a parasite’s ability to infect multiple hosts and its fitness on each host [7,53]. Here we quantified that trade-off using the parameter *a*, which reduced the shedding rate of a generalist parasite to a fraction of that of a specialist parasite. Such a reduction in shedding might be caused by a reduction in infection intensity, as other studies have shown that generalist parasites often have lower infection intensities than specialists [54,55]. Indeed, many experimental evolution studies have shown that, as a parasite is forced to adapt to a novel host, it gradually loses its infectiousness and/or replication ability in the original host, such that, when the parasite is able to infect both the original and novel host, its fitness is lower in each than when it is specialized [56]. However, fitness trade-offs are notoriously challenging to measure, so assessing the importance of such trade-offs to the evolution of host range in any large host-parasite dataset is practically impossible. Using allometric scaling relationships to define model parameters in terms of easily measurable host traits like body size and temperature provides us with an opportunity more explicitly connect the model with data.

A second general issue with connecting the model results to data is that of phylogenetic relatedness. The model only makes predictions about the number of hosts that a parasite can infect. In reality, however, we want to distinguish between a parasite that infects *n* hosts within the same genus and a parasite than infects *n* hosts across many genera. Here we addressed that issue by using several measures of host range. We measured “structural” generalism using the number of hosts (degree and G), and we measured “phylogenetic” generalism using metrics that account for the phylogenetic distance between hosts (SPD and SES-PD). SPD, which measures the mean pairwise phylogenetic distance between hosts, has been shown to correlate with degree [31], so we also included a measure of phylogenetic generalism that is scaled to remove the association with number of hosts (SES-PD). SES-PD therefore attempts to measure only the phylogenetic distinctiveness of the host range, so a parasite with only two hosts could have a much higher value of SES-PD than a parasite with ten hosts. The important point is that the model is directly making predictions about structural generalism and we are assuming that phylogenetic generalism is more likely under the same conditions as structural generalism.

The models predict that, for direct life cycle parasites, increasing host body size increases the fitness of the generalist parasite, suggesting that there should be positive correlation between host body size and host range. For trophically transmitted parasites, the model predictions were more complicated, suggesting that this correlation could be positive or negative, depending on model assumptions and the value of other parameters. Interestingly, previous verbal models for host range evolution have suggested the correlation between host range and host body size should work in the opposite direction, with high host specificity evolving when hosts are large-bodied [22], supposedly because large-bodied species are longer-lived, and thus more predictable in their availability. However, the predictability of a resource (in this case, the host) depends on the probability of encountering that resource [52], which is determined not by resource lifespan but by abundance. Thus the observed allometric relationship between body size and abundance would seem to run counter to this verbal model. Nevertheless, a number of studies have shown a negative correlation between mean or maximum host body size and generalism [20-22]. We examined this correlation in our fish-macroparasite database using different metrics of host size (size of a parasite’s largest host, mean size of all hosts, and the coefficient of variation in host size) and of host range (Table 4). For both direct and trophic life cycle parasites, we found a strong and significant positive correlation between the coefficient of variation in host body size and all metrics of host range. The maximum host body size was positively correlated with all generalism metrics except SES-PD. There was a weak negative correlation between mean host body size and all metrics of host range (Fig. 1, 2). The data thus provides some support for the model predictions, especially when looking at structural generalism metrics. The negative correlation between mean host body size and generalism is interesting, as it has been observed in other studies with smaller datasets [21,22]. As it turns out, whether we interpret the model as predicting that the mean host body size for generalist parasites is larger than that for specialist parasites depends on the implicit assumption that, if the generalist parasite can invade (its invasion fitness is greater than one), it displaces the specialist parasite.

If we had instead assumed that the a generalist parasite would coexist with any specialist parasites, our predictions would be affected. To see how, consider the system defined above A generalist parasite can invade if (eq. 1)

whereas a parasite specialized on the smaller secondary host can invade if

It is clear from comparing these two expressions that it is quite likely that a generalist parasite could invade even when a specialist could not because the generalist’s fitness also depends on the primary host. A specialist parasite could invade when a generalist could not only when *a* is very small (the cost of generalism is very high). If generalists and specialists can coexist, this result suggests that both generalist and specialist parasites will infect large-bodied hosts, whereas only generalist parasites will infect small-bodied hosts. This would lead to a prediction that the correlation between mean host body size and host range should be negative, as we observed in our dataset. On the other hand, there would likely be no correlation between the maximum host body size and host range, which is not what we observed. Thus, there is no simple way to reconcile the differences between the model and data analyses, which underscores the importance of understanding how model results are translated into empirically testable predictions.

The models made very inconsistent predictions about the influence of temperature on host range evolution (Table 2, 3). Perhaps unsurprisingly, the data is also somewhat ambivalent on this question. Our analysis suggests that the number of hosts a parasite can infect (‘degree’) is higher in colder regions (Fig. 3,4) for both direct and trophic life cycle parasites, a result which has been observed before [26]. On the other hand, for direct life cycle parasites, the other metrics of host range do not show any significant differences between warm and cold regions (Fig. 3), whereas for trophically transmitted parasites, there are some positive and some negative correlations between host range and temperature. However, it is important to be aware that ectotherm body size also increases with decreasing temperature. If hosts in colder waters are larger, that could be an important confounding influence on these patterns. (Does host size distribution differ by region in our dataset?)

Here we attempted to study the ecological factors that influence host range by combining an evolutionary analysis of a class of simple epidemiological model with analysis of a massive database of host-parasite associations. This revealed a number of places where model and data agree, as well as important areas of disagreement. We suggest that this approach is a valuable approach going forward, and highlight a few of the ways in which the models developed here could be productively extended.

In particular, we have ignored the role of parasite size here, as our database did not include that information. However, parasite life history is likely to be dependent on body size as well [23], and there is a positive correlation between host and parasite body size, both empirically and theoretically [20,21,57,58]. In particular, parasite body size will affect within-host abundance, and thus shedding rate. Moreover, we have assumed that shedding rate is positive correlated with abundance, but for many parasites the opposite is true: increased within-host abundance increases density-dependence, thereby reducing parasite fecundity such that shedding is actually lower [24,59].

Another important simplification is in our assumptions about parasite virulence. Simple verbal models would suggest that more virulent parasites are more likely to be specialists, as the fitness trade-off for infecting multiple hosts should be steeper [29]. In our models, increasing the value of host mortality would always reduce a generalist’s , suggesting the specialism would be favoured. However, we have assumed that virulence depends only on host body size. If instead it depends upon within-host abundance, then shedding and virulence are linked. This might make a more virulent generalist parasite more likely to invade. For example, from Table 2, the of a generalist when there are two specialist parasites is

If shedding rate is a function of virulence, then whether increased virulence increases or decreases the generalist’s depends on the signs of and . If the increase in shedding with virulence is large enough, then the virulent generalist can invade.

Understanding the processes that influence host range evolution is often highlighted as a key challenge for the evolutionary ecology of parasites [10,29,31], especially given that host range is closely linked to transmission, particularly in regards to reservoir hosts, spillover/emergence, and changes in virulence [1-4]. Combining simple mathematical models with analysis of host-parasite databases may help reveal general principles shaping the evolution of host range.

Additional Information

**Information on the following should be included whenever relevant.**

**Acknowledgments**

Please acknowledge anyone who contributed to the study but did not meet the authorship criteria. We thank Oana Paun for assistance in data collection.

**Ethics**

NA

**Data Accessibility**

All manuscripts which report primary data (usually research articles) should include a Data Accessibility section which states where the article's supporting data can be accessed. This section should also include details, where possible, of where to access other relevant research materials such as statistical tools, protocols, software etc. If the data has been deposited in an external repository this section should list the database, accession number and link to the DOI for all data from the article that has been made publicly available, for instance:

DNA sequences: Genbank accessions F234391-F234402 (**http://dx.doi.org/xxxxx**)  
Phylogenetic data, including alignments: TreeBASE accession number S9123 (**http://dx.doi.org/xxxxx**)  
Climate data and MaxEnt input files: Dryad doi:10.5521/dryad.12311 (**http://dx.doi.org/xxxxx**)  
  
If the data is included in the article’s Supplementary Material this should be stated here, for instance:  
The datasets supporting this article have been uploaded as part of the Supplementary Material.

**Authors' Contributions**  
All authors contributed to the conception and design of the article. CEC and AH contributed model derivation and analysis. JGW, JC, ARE acquired the data. JGW and SJP calculated the phylogenetic metrics and cleaned and analysed the data. JGW and CEC wrote the first version of the article and all authors contributed to revisions and final editing.

**Competing Interests**

We have no competing interests.

**Funding**

All authors were supported by the British Ecological Society Parasite and Pathogen Special Interest Group to attend the 2015 Transmission Research Retreat. JC and AE were funded by the Welsh Government and Higher Education Funding Council for Wales through the Sêr Cymru National Research Network for Low Carbon, Energy and Environment. AH was supported by a Discovery Grant (RGPIN 2014-05413) from the National Sciences and Engineering Research Council of Canada. SJP is funded by NERC grants NE/M000338/1, NE/M000591/1 & NE/M00080X/1.

References

1. Authors. Year. Title. *Abbreviated Journal title* **Volume**, page range. (doi)

2. Authors. Year. Title. *Abbreviated Journal title* **Volume**, page range. (doi)3. Authors. Year. Title. *Abbreviated Journal title* **Volume**, page range. (doi)

4. Authors. Year. Title. *Abbreviated Journal title* **Volume**, page range. (doi)5. Authors. Year. Title. *Abbreviated Journal title* **Volume**, page range. (doi)

6. Authors. Year. Title. *Abbreviated Journal title* **Volume**, page range. (doi)

1. Streicker, D. G., Fenton, A. & Pedersen, A. B. 2013 Differential sources of host species heterogeneity influence the transmission and control of multihost parasites. *Ecol. Lett.* **16**, 975–984. (doi:10.1111/ele.12122)

2. Viana, M., Mancy, R., Biek, R., Cleaveland, S., Cross, P. C., Lloyd-Smith, J. O. & Haydon, D. T. 2014 Assembling evidence for identifying reservoirs of infection. *Trends Ecol. Evol.* **29**, 270–279. (doi:10.1016/j.tree.2014.03.002)

3. Leggett, H. C., Buckling, A., Long, G. H. & Boots, M. 2013 Generalism and the evolution of parasite virulence. *Trends Ecol. Evol.* **28**, 592–596. (doi:10.1016/j.tree.2013.07.002)

4. Cleaveland, S., Laurenson, M. K. & Taylor, L. H. 2001 Diseases of humans and their domestic mammals: pathogen characteristics, host range and the risk of emergence. *Philos. Trans. R. Soc. B Biol. Sci.* **356**, 991–999. (doi:10.1098/rstb.2001.0889)

5. Woolhouse, M. E. J., Taylor, L. H. & Haydon, D. T. 2001 Population Biology of Multihost Pathogens. *Science (80-. ).* **292**, 1109–1112. (doi:10.1126/science.1059026)

6. Taylor, L. H., Latham, S. M. & Woolhouse, M. E. J. 2001 Risk factors for human disease emergence. *Philos. Trans. R. Soc. B Biol. Sci.* **356**, 983–9. (doi:10.1098/rstb.2001.0888)

7. Futuyma, D. A. & Moreno, G. 1988. The evolution of ecological specialization. *Ann. Rev. Ecol. Syst.* **19**: 207-233.

8, Poulin, R. 1998. Large-scale patterns of host use by parasites of freshwater fishes. *Ecol. Lett.* **1**: 118-128.

9. Garamszegi, L. Z. 2006. The evolution of virulence and host specialization in malaria parasites of primates. *Ecol. Lett.* **9**: 933-940.

10. Poulin, R. 2011. *Evolutionary Ecology of Parasites*. 2nd edition. Princeton University Press.

11. Day, E. H., Hua, X., & Bromham, L. 2016. Is specialization an evolutionary dead-end? Testing for differences in speciation, extinction, and trait transition rates across diverse phylogenies of specialists and generalists. *J. Evol. Biol.* **29**: 1257-1267.

12. Agosta, S. J., Janz, N. & Brooks, D. R. 2010 How specialists can be generalists: resolving the ‘parasite paradox’ and implications for emerging infectious disease. *Zoologia* **27**, 151–162. (doi:10.1590/S1984-46702010000200001)

13. Nosil, P. & Mooers, A. Ø. 2005 Testing hypotheses about ecological specialization using phylogenetic trees. *Evolution (N. Y).* **59**, 2256. (doi:10.1554/05-169.1)

14. Johnson, K. P., Malenke, J. R. & Clayton, D. H. 2009 Competition promotes the evolution of host generalists in obligate parasites. *Proc. Biol. Sci.* **276**, 3921–6. (doi:10.1098/rspb.2009.1174)

15. Mendlová, M. & Šimková, A. 2014 Evolution of host specificity in monogeneans parasitizing African cichlid fish. *Parasit. Vectors* **7**, 69. (doi:10.1186/1756-3305-7-69)

16. Hoberg, E. P. & Brooks, D. R. 2008 A macroevolutionary mosaic: episodic host-switching, geographical colonization and diversification in complex host-parasite systems. *J. Biogeogr.* **35**, 1533–1550. (doi:10.1111/j.1365-2699.2008.01951.x)

17. Kuris, A. M., Blaustein, A. R. & Alio, J. J. 1980 Hosts as Islands. *Am. Nat.* **116**, 570–586.

18. Ezenwa, V. O., Price, S. A., Altizer, S., Vitone, N. D. & Cook, K. C. 2006 Host traits and parasite species richness in even and odd-toed hoofed mammals, Artiodactyla and Perissodactyla. *Oikos* **115**, 526–536.

19. Guégan, J.-F., Lambert, A., Lévêque, C., Combes, C. & Euzet, L. 1992 Can host body size explain the parasite species richness in tropical freshwater fishes? *Oecologia* **90**, 197–204. (doi:10.1007/BF00317176)

20. Sasal, P., Trouve, S., Muller-Graf, C. & Morand, S. 1999. Specificity and host predictability: a comparative analysis among monogenean parasites of fish. *J. Anim. Ecol.* **68**: 437-444.

21. Desdevises, Y., Morand, S, & Legendre, P. 2002. Evolution and determinants of host specificity in the genus *Lamellodiscus* (Monogenea). *Biol. J. Linn. Soc.* **77**: 432-443.

22. Krasnov, B. R., S. Morand, D. Mouillot, G. I. Shenbrot, I. S. Khokhlova, and R. Poulin. 2006. Resource predictability and host specificity in fleas: the effect of host body mass. *Parasitology* 133: 81-88.

23. Hechinger, R. 2013. A metabolic and body-size scaling framework for parasite within-host abundance, biomass, and energy flux. *Am. Nat.* 182: 234-248.

24. Cattadori, I. M., Wagner, B. R., Wodzinski, L. A., Pathak, A. K., Poole, A. & Boag, B. 2014. Infections do not predict shedding in co-infections with two helminths from a natural system. *Ecology* **95**: 1684–1692. (doi:10.1890/13-1538.1)

25. Savage, V. M., Gilloly, J. F., Brown, J. H. & Charnov, E. L. 2004 Effects of body size and temperature on population growth. *Am. Nat.* **163**, 429–41. (doi:10.1086/381872)

26. Rohde, K. 1993. *Ecology of Marine Parasites: An Introduction to Marine Parasitology*. 2nd ed. CAB International, Wallingford, UK.

27. Allen, A. P., Brown, J. H., & Gillooly, J. F. 2002. Global biodiversity, biochemical kinetics, and the energetic-equivalence rule. *Science* **297**: 1545-1548. (doi: 10.1126/science.1072380)

28. Pietrock, M. & Marcogliese, D. J. 2003. Free-living endohelminth stages: at the mercy of environmental conditions. *Trends Parasitol.* **19**: 293-299. (doi: 10.1016/S1471-4922(03)00117-X)

29. Schmid-Hempel, P. 2011. *Evolutionary Parasitology.* Oxford University Press.

30. Geritz, S. A. H., Kisdi, É., Meszéna, G. & Metz, J. A. J. 1998 Evolutionarily singular strategies and the adaptive growth and branching of the evolutionary tree. *Evol. Ecol.* **12**, 35–57. (doi:10.1023/A:1006554906681)

31. Poulin, R., Krasnov, B. R. & Mouillot, D. 2011 Host specificity in phylogenetic and geographic space. *Trends Parasitol.* **27**, 355–361. (doi:10.1016/j.pt.2011.05.003)

32. Strona, G., Palomares, M. L. D., Bailly, N., Galli, P. & Lafferty, K. D. 2013 Host range, host ecology, and distribution of more than 11 800 fish parasite species. *Ecology* **94**, 544. (doi:10.1890/12-1419.1)

33. Noble, E. R., Noble, G. A., Schad, G. A. & MacInnes, A. J. 1989. *Parasitology. The Biology of Animal Parasites.* 6th edn. Lea & Febig, Philadelphia, PA, USA.

34. Bellay, S., Oliveira, E. F., Almeida-Neto, M., Mello, M. A. R., Takemoto, R. M. & Luque, J. L. 2015 Ectoparasites and endoparasites of fish form networks with different structures. *Parasitology* **142**, 901–909. (doi:10.1017/S0031182015000128)

35. 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.

36. Hoberg, E. P. & Brooks, D. R. 2008. A macroevolutionary mosaic: episodic host-switching, geographical colonization and diversification in complex host-parasite systems. *J. Biogeo.* **35**: 1533-1550. (doi: 10.1111/j.1365-2699.2008.01915.x)

37. Rohde, K. 2002. Ecology and biogeography of marine parasites. *Adv. Marine Biol.* **43**: 1-86.

38. Morand, S., Simková, A., Matejusová, I., Plaisance, L., Verneau, O. & Desdevises, Y. 2002 Investigating patterns may reveal processes: evolutionary ecology of ectoparasitic monogeneans. *Int. J. Parasitol.* **32**, 111–119. (doi:10.1016/S0020-7519(01)00347-2)

39. Hurford, A., Cownden, D. & Day, T. 2010 Next-generation tools for evolutionary invasion analyses. *J. R. Soc. Interface* **7**, 561–571. (doi:10.1098/rsif.2009.0448)

40. Froese, R., Pauly, D. & Editors 2015 FishBase.

41. McGinnis, S. & Madden, T. L. 2004 BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Res.* **32**, W20–W25. (doi:10.1093/nar/gkh435)

42. Rice, P., Longden, I. & Bleasby, A. 2000 EMBOSS: The European Molecular Biology Open Software Suite. *Trends Genet.* **16**, 276–277. (doi:10.1016/S0168-9525(00)02024-2)

43. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. 2009 trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973. (doi:10.1093/bioinformatics/btp348)

44. Kimura, M. 1980 A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* **16**, 111–120.

45. Paradis, E., Claude, J. & Strimmer, K. 2004 APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics* **20**, 289–290. (doi:10.1093/bioinformatics/btg412)

46. Poulin, R. & Mouillot, D. 2003 Parasite specialization from a phylogenetic perspective: a new index of host specificity. *Parasitology* **126**, 473–480. (doi:10.1017/S0031182003002993)

47. Faith, D. P. 1992 Conservation evaluation and phylogenetic diversity. *Biol. Conserv.* **61**, 1–10. (doi:10.1016/0006-3207(92)91201-3)

48. Schliep, K. P. 2011 phangorn: phylogenetic analysis in R. *Bioinformatics* **27**, 592–593. (doi:10.1093/bioinformatics/btq706)

49. Kembel, S. W., Cowan, P. D., Helmus, M. R., Cornwell, W. K., Morlon, H., Ackerly, D. D., Blomberg, S. P. & Webb, C. O. 2010 Picante: R tools for integrating phylogenies and ecology. *Bioinformatics* **26**, 1463–1464. (doi:10.1093/bioinformatics/btq166)

50. Poulin, R. 1992. Determinants of host-specificity in paraistes of freshwater fishes. *Int. J. Parasit.* **22**: 753-758.

51. Combes, C. 2001. *Parasitism: The Ecology and Evolution of Intimate Interactions*. University of Chicago Press, Chicago, IL, USA.

52. Ward, S. A. 1992. Assessing functional explanations of host-specificity. *Am Nat* 139: 883-891.

53. Gudelj, I., Van den Bosch, F. & Gilligan, C. A. 2004 Transmission rates and adaptive evolution of pathogens in sympatric heterogeneous plant populations. *Proc. R. Soc. London B Biol. Sci.* **271**, 2187–2194. (doi:10.1098/rspb.2004.2837)

54. Poulin, R. 1998. Large-scale patterns of host use by parasites of freshwater fishes. *Ecol. Lett.* **1**: 118-128.

55. Garamszegi, L. Z. 2006. The evolution of virulence and host specialization in malaria parasites of primates. *Ecol. Lett.* **9**: 933-940.

56. Ebert, D. 1998. Experimental evolution of parasites. *Science* **282**: 1432-1435.

57. Morand, S., Legendre, P., Garner, S. L., & Hugot, J. P. 1996. Body size evolution of oxyurid (Nematoda) parasites: the role of hosts. *Oecologia* **107**: 274-282.

58. De Leo, G. A., Dobson, A. P. & Gatto, M. 2016. Body size and meta-community structure: the allometric scaling of parasitic worm communities in their mammalian hosts. *Parasitol.* **143**: 880-893.

59. Keymer, A. E. & Slater, A. F. 1987. Helminth fecundity: density dependence or statistical illusion? *Parasitol. Today* **3**: 56–58.

Tables

Tables should be inserted at the end of the document, unless they are provided in another format (e.g. Excel) in which case they should be supplied as a separate file. Please provide all tables in an editable format (rather than embedded as an image).

Figure and table captions

Table and figure captions should be included at the end of the manuscript file and should be brief and informative. Ensure that permission has been obtained for all use of third party or previously published figures, and include full credit information. If publishing an open access paper, permission must be cleared for this use. Please let the Editorial Office know of any copyright issues.

Figures

For final submissions, figures should be uploaded as separate, high resolution, figure files.

Supplementary material

Supplementary material can be used for supporting data sets, movies, figures and tables, and any other supporting material. The main article, however, should stand on its own merit. Where possible, supplementary material should be combined into one Word document or PDF. A template is available on our website, or on request.



Figure S1. Comparison of pairwise genetic distances between hosts generated from cytochrome oxidase 1 (CO1) sequences and cytochrome B (cytB) sequences plotted against one to one line.



Figure S2. Distribution of genetic distances from all pairwise comparisons of hosts.

* Appendix A & B: Additional model derivation detail
* Appendix C: data
* Appendix D: additional results???