Host allometry and parasite transmission model influence the evolution of host rangeq: theory and meta-analysis

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

**Keywords:**

|  |  |  |
| --- | --- | --- |
| Name | Affiliations | ORCID |
| Jo Cable | School of Biosciences, Cardiff University, Cardiff CF10 3AX, UK | 0000-0002-8510-7055 |
| Clay Cressler |  |  |
| Amy Ellison | School of Biosciences, Cardiff University, Cardiff CF10 3AX, UK |  |
| Amy Hurford |  |  |
| 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 ().

†Present address:

Summary

Parasites vary widely in how many host species they infect, with some parasite species able to only infect a single species, while others are capable of infecting many. Understanding the factors that drive host-parasite specificity is of basic biological interest, but also directly relevant to predicting disease emergence in new host species, and identifying parasites that are likely to have unidentified alternative hosts. Here, we use mathematical models to investigate how parasite transmission strategy, variation in host body size, and environmental temperature affect the evolution of parasite generalism. The models make several general predictions, including that parasites are more likely to evolve a generalist strategy when hosts are large-bodied, when variation in body size is small, and in cooler environments. We explore these predictions using a database of over 20,000 host-parasite associations focused on macroparasites of fish. This analysis shows some evidence for these simple model predictions, but also highlights mismatches between theory and data. By combining these two approaches, we both establish a theoretical basis for interpreting empirical data on host-parasite specificity and identify key areas for future work.

Introduction

The ability of a parasite to infect multiple hosts has implications for the evolution of virulence [1], maintenance of transmission [2,3], and host-switching or disease emergence in humans [4]. (**EXAMPLES**)

Despite evidence for the importance of a parasite’s host range, it remains unclear whether most parasites are generalist or specialist, and what factors might influence the evolution of host range. For example, the emerging infectious disease literature (ranging from microbes to macroparasites) suggests that most pathogens are generalists: 60% of human infectious diseases are zoonotic and 80% of pathogens of domestic animals infect multiple hosts [4,6,7]. However, other work suggests that parasites are generally resource (host) specialists [8]. There appears to be something of a paradox, in that parasites are commonly resource specialists with restricted host ranges, but that in parasite evolutionary lineages there are commonly shifts from one host species to relatively unrelated new host species [8]. In other words, while in theory interspecific competition selects for an increase in ecological specialization (which for parasites means a narrow host range), there are other evolutionary drivers that can lead to the evolution of generalism (for parasites, having a wide host range). Indeed, in some cases parasites’ host specialism appears to be the ancestral state with generalism being derived, for example, in feather lice of doves [9]. Environmental change could drive both the persistence (specialism) and diversification (generalism) of a parasite within its host(s), with geographical isolation driving host/parasite co-evolution and so specialization towards a single host species, and geographical expansion driving opportunities for host switching, and thus a wider host range [10].

Previous studies have suggested that whether a parasite is generalist or specialist is influenced by both parasite and host traits and environmental factors (Table 1; see Cable et al. this volume). Parasites: Gender/sexual reproduction, motility and environmental transmission, size, fecundity. Hosts: Reproduction, habitat, diet, group size, moulting. (THE FOLLOWING SENTENCE NEEDS WORK, but I think we need something here to justify what we are doing.) While each of these is likely important for at least some parasites, some of these factors, such as host body size or parasite transmission mode, will influence parasite evolution more generally. Mathematical models are a useful tool for predicting the affect of such factors on the evolution of host range.

Here, we develop a simple theoretical framework for studying host range evolution, and use simple allometric scaling relationships to investigate how variation in host body size and temperature, as well as parasite transmission mode, affect whether specialism or generalism is the evolutionarily stable parasite strategy. We then use an extensive data set of macroparasites of fish to test the predictions arising from these models. Overall, we aim to improve our understanding of the ecological and evolutionary factors that contribute to parasite generalism.

Table 1: Host and parasite characteristics predicted to affect generalism that are explored in this paper.

|  |  |  |
| --- | --- | --- |
| **Variable** | **Levels** | **Previous Hypotheses** |
| Infection site | *Endoparasite* lives inside the host, *Ectoparasite* lives on the surface of the host | Infection site will give different opportunities for transmission mode, and mobility of infective stages may affect the evolution of generalism [11] |
| 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 [Nobel 1989 cited in 11] |
| 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 [10] This is also related to host diet… |
| Host geographic range | Geographic regions: Africa; Antarctica; Australia; Indopacific; Nearctic; Neotropical; Palearctic. | Allometric relationships exist between temperature and life history parameters [12] that appear in mathematical models describing the evolution of generalism. While in general diversity is higher in the tropics, Digenean parasites of marine fish in tropical seas are more host specific than those that parasitize fish in colder seas (Rohde 1993-book). No relationship is observed between latitude and generalism for Monogenans [11]. |
| Host body size | Continuous (maximum length of fish) | Parasites of ungulates: parasite species richness found to increase with host body size across all parasites groups [13]. |

Model Derivation and Predictions

**Effect of host body size and temperature**

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 trophically transmitted parasites, at least one host is infected through the environment). 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. 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 is caused by hosts contacting parasites in the environment at the *per capita* rate , which is assumed to be equal across both hosts and parasites. 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 are assumed to be independent of which parasite strain the host is infected with. 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 (Geritz et al 1998) 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 of the full system where and where all other state variables are at the equilibrium set by the specialist parasite. The Jacobian for that equilibrium is the block-diagonal matrix,

The stability of this system is given 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 (Hurford et al. 2010), 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 the 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 generalists to be able to invade, 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 this form of is easy to understand, analytically it will be easier to work with a slightly different expression. Plugging in the equilibrium abundance of the primary and secondary host, the expression for simplifies to

At this point, we could specify different parameter values (corresponding to different host and parasite traits) and ask whether . However, we can get more general insights into the evolution of generalism by asking whether increasing the value of each parameter increases or decreases the value of , that is, by looking at the derivatives of with respect to each parameter. For this simple model, this analysis makes very intuitive (and boring) predictions (see Appendix A), as the model in its current state does not include any biologically relevant trade-offs that constrain the relationships between parameters. It is also 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 [28]. In particular, host carrying capacities and mortality rates will scale with host body size (Savage et al. 2004) 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 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. 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 life cycle-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. This makes it easier for a generalist endoparasite to invade than a generalist ectoparasite.

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 assume that the primary host is larger than the secondary host, since the fitness of a specialist parasite (given by 1/) is an increasing function of body size, and that the body size of the secondary host is , where *W* is the size of the primary host.

We then 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 the generalism index of each parasite and the both body size of its largest host and the mean body size across all its hosts.**

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, and as we have already shown, increasing host mass increases the likelihood of invasion. 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.

Temperature likewise has a consistent effect on , but in the opposite direction: 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.

We point out that the majority of endoparasites in the fish-macroparasite database are actually trophically transmitted. Thus, in Appendix 2 we show that these predictions hold under a model of trophic transmission as well. This trophic transmission model is more challenging to analyse, so we are forced to rely on numerical exploration to validate this prediction.

For ectoparasites (the majority of which have direct life cycles), the response of to changes in traits is more complicated. For example, 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 , this expression will be negative, making both derivatives positive, as they were for endoparasites, whenever

That is, 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: **generalists will have an easier time invading when temperatures are colder.**

**Effect of parasite transmission mode**

In this analysis we examine how parasite transmission mode (direct versus trophic transmission) affects the evolution of generalism. We extend Choisy et al.’s model for the evolution of trophic transmission [27] to include generalist parasites by assuming that there are two terminal host species. Parasites are shed into the environment from an infected terminal host, where they may either infect only the terminal hosts (a direct life cycle) or may infect a single intermediate host that is consumed by the terminal hosts (trophic transmission). To facilitate analysis, we assume that the host abundances are constant, with the abundance of terminal host *i* and the abundance of the intermediate host. The fraction of terminal and intermediate hosts that are infected are and , respectively. The system of differential equations describing the spread of the parasite between the different hosts is,

|  |  |
| --- | --- |
|  | (1)-(4) |

Parasites are shed from the terminal hosts at a rate The mortality rates of infected terminal hosts, infected intermediate hosts, and parasites in the environment are , , and , respectively. The remaining parameters control the route of transmission. The rate that the parasite infects terminal host *i* from the environment is , the rate that it infects the intermediate host from the environment is , and the rate that the rate of contact between the intermediate host and terminal host *i* is . When and , the model is of a direct life cycle parasite; when the opposite condition holds, the model is of trophic transmission.

For this system of equations, the basic reproduction number of the parasite is

|  |  |
| --- | --- |
|  | (5) |

where is the average number of intermediate hosts infected from a parasite in the environment,

are the average number of terminal hosts of species 1 and 2 , respectively, infected from a parasite in the environment,

is the average number of parasites shed into the environment from an infected intermediate host,

and are the average number of parasites shed into the environment from infected terminal hosts of species 1 and 2, respectively,

There are three key aspects of the model formulation that affect our conclusions regarding generalism and trophic transmission:

A1) Parasites that infect an intermediate or terminal host are removed from the environment,

A2) When trophic transmission occurs, the intermediate host is killed, and

A3) For trophically transmitted parasite, the same intermediate host can infect either terminal host.

To determine whether parasite that are trophically transmitted are more likely to be generalists, we compare the conditions for the evolution of generalism for parasites with simple and complex life cycles. For a parasite that has a simple life cycle, generalism (the ability to infect the second host, i=2) will always evolve unless there is a trade-off between acquiring the ability to infect the second host and another life history parameter (Table 1; see Appendix 3 for details of how the analysis was performed). Assuming a linear trade-off between the rate that parasites in the environment can infect either host, , then the condition for generalism to evolve is , i.e., that the abundance of the new host must exceed the abundance of the host infected when the parasite is a specialist. In Table 1, we denote this scenario as Di-DiDi, meaning that a specialist parasite with a direct life cycle (Di) evolves to infect two hosts and has a direct life cycle with each host (DiDi).

Given the model formulation, there are several ways a generalist parasite that is trophically transmitted might evolve. For example, the parasite may be trophically transmitted and evolve the ability to be transmitted from the intermediate host to the second terminal host - this is denoted by the scenario Tr-TrTr and involves evolving to be greater than 0. Our analysis (summarized in Table 1) reveals that a trophically transmitted parasite is equally likely to evolve to be a generalist (Tr-TrTr) as a parasite with a direct life cycle (Di-DiDi).

Alternatively, a specialist parasite with a direct life cycle might evolve the ability to infect an intermediate host that then infects the two terminal hosts that consume it. This scenario, Di-TrTr, evolves if the mortality rate of the intermediate host, μ2, is less than,

|  |  |
| --- | --- |
|  | (10) |

in the absence of a life history trade-off or if it is less than,

|  |  |
| --- | --- |
|  | (11) |

assuming a linear life history tradeoff between the rate that parasites infect the intermediate host from the environment and the rate they infect the terminal host. For a parasite that has a direct life cycle but evolves to be trophically transmitted, parasites are removed from the environment when they infect the intermediate host (A1) and infected intermediate hosts are destroyed by predation due to both terminal hosts (A2), making Di-TrTr unlikely to evolve.

We might also consider a parasite that is already a generalist and ask if this parasite is more likely to evolve trophic transmission than a parasite that is a specialist. The condition for a specialist with a simple life cycle to evolve trophic transmission (Di-Tr) is μ2 < C3 in absence of a life history trade-off, and that μ2 must be less than,

|  |  |
| --- | --- |
|  | (12) |

when there is a linear life history trade-off between the ability of the parasite in the environment to infect the intermediate host and the terminal host. This condition is closely related to the Choisy et al.’s finding that a necessary condition for the evolution of trophic transmission (in the presence of a linear trade-off) is that the abundance of the intermediate host must exceed the abundance of the terminal host. This can be seen in equation (12) because if N2 < N11 then C3t is negative and the condition for trophic transmission to evolve can never be satisfied. In addition, the necessary condition for trophic transmission to evolve when the parasite also evolves generalism (Di-TrTr; μ2 < C4t; equation (11)) is that the abundance of the intermediate host must be greater than the sum of the abundances of both terminal hosts . As such, our result that the condition for trophic transmission to evolve with generalism is more strict than the condition for trophic transmission to evolve for a specialist parasite makes sense in light of previous results that report the abundance of the intermediate host must exceed the abundance of the terminal host(s) for trophic transmission to evolve [27].

Overall, we conclude that parasites that have trophic transmission are less likely to be generalists than parasites with simple or complex life cycles. This is because for all the scenarios we considered the evolution of trophically transmitted generalist parasites was either equally likely or less likely than the evolution of generalist parasites with other types of life cycles (Table 2).

Table 2: Are parasites that are trophic transmitted more likely to be generalists than parasites that have direct life cycles? The scenario Si-SiSi means that a specialist parasite with a simple life cycle (Si) evolves to become a generalist with a simple life cycle (SiSi). We compare this reference scenario to that of a parasite with a complex life cycle evolving to be a generalist (Cx-CxCx) and find that the conditions that must be satisfied are the same. ‘As above’ indicates that considering a life history trade-off does not impact the conclusions.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Scenario | Evolving parameter | Assumptions | Tradeoff | Condition | More or less likely to evolve compared to reference scenario (underlined) |
| Complex life cycles | | | | | |
| Di-DiDi | β12 | β2=0 | - | Always satisfied | Reference scenario |
| β11=A- β12 | N11<N12 | Reference scenario |
| Di-Tr | β2 | β12=0  p2=0  τ=1 | - | μ2 < C3 | Reference scenario |
| β11=A- β2 | μ2 < C3t | Reference scenario |
| Tr-TrTr | p2 |  | - | Always satisfied | Di-DiDi is equally likely |
| p1=A-p2 | N11<N12 | As above |
| Di-TrTr | β2 | p1=p2=p  β11 = β12= β1 | - | μ2 < C3 | Di-DiDi is more likely than Di-TrTr due to A1.  Di-Tr is equally likely |
| β1=A- β2 | μ2 < C4t | Di-DiDi – the conditions are not comparable  Di-Tr is more likely than Di-TrTr. |

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 [14]. Additional records were included for xx monogenean and 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 [15]. Parasite species names were checked against NCBI taxonomy database in the same way and were 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 23,360 unique host-parasite associations between 8,847 parasite species and 4,243 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 [16] 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* [17]*.* 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 [16] 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 [18] 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 [19] in the R package *ape* [20]. 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 1sequences. The point patterns were compared to the one to one line, and correspondence to this line used to select the loci to be concatenated using *ape*. Pairwise genetic distances for missing pairs were then imputed using a custom R script by averaging according to the following relationships between taxa: 1) pairs from different genera – the mean genetic distances of one member of the pair (depending on available data) 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, so parasites’ host generalism metrics were defined according to structural specificity (the number of hosts, referred to as “degree” and a binary variable G that is 0 if the parasite has a single host and 1 otherwise) and phylogenetic host specificity (the average pairwise genetic distance between all hosts, SPD) [21,22]. SPD is given a value of 0 when the parasite has only one host.

The pairwise genetic distances between hosts were used to calculate additional measures of generalism. A UPGMA tree was calculated from the full distance matrix using *phangorn* in R [23] 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 (PS; here, the length of the host tree with the root excluded) [5,24] was generated for all parasites using *picante* in R [25]. We used the ses.pd function implemented in *picante* to generate the standardized effect size of phylogenetic diversity (SPS) based on 1000 runs [5]. SPS 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 measures of host and parasite characteristics (Table 1 in main document) for which theoretical predictions were made. Metrics for parasites with direct or trophic life cycles were compared to summary measures (mean, maximum, and coefficient of variation (CV)) of the maximum length reported for each of their hosts. Note that CV of the host length is only calculated for parasites with more than one host. Endoparasites and ectoparasites were assessed separately, with an additional categorical length measure used for ectoparasites, whereby the length summary measures were divided into a categorical variable according to quartiles. This was done to facilitate comparison with the model prediction that generalism should be unlikely for parasites of very large, or very small, hosts.

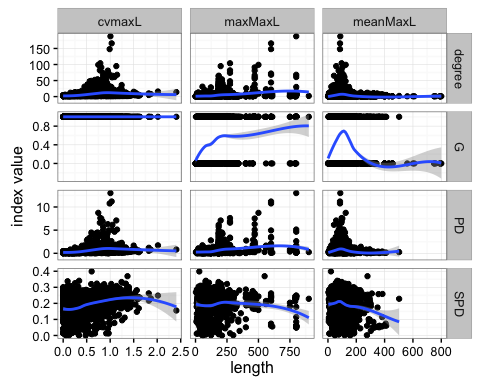
The effect of geographic region on parasites with direct or trophic life cycle was calculated for endoparasites and ectoparasites together. Regions were assessed as defined in Table 1 and also divided into two groups, where Antarctica, Nearctic, and Palearctic were assumed to be colder than Africa, Australia, Indopacific, and Neotropical regions. Some host-parasite associations were reported in more than one region, so for this analysis the generalism metrics were calculated separately for each region. Unfortunately, it is not possible to more precisely specify the temperatures, as each region typically spans a wide range of temperatures.

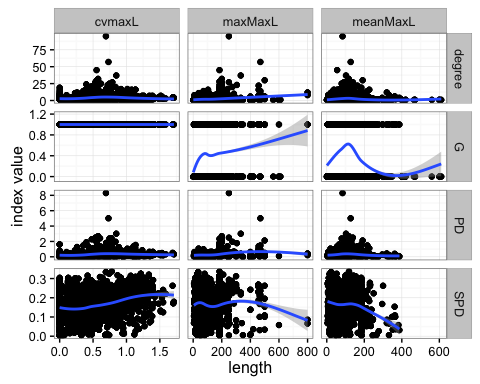
Because the generalism metrics come from very different distributions, we used GLMs with different error distributions for statistical analyses. For degree, we used negative binomial regression with a log link function (glm.nb() in R); for G, logistic regression (glm(family="binomial") in R); and for PD and SPD, linear regression (lm() in R), with PD log transformed. For lm(), assumptions of homoscedasticity and normality of residuals were assessed visually, and in some cases there was a bit of heteroscedasticity so there may be a bias in the standard error estimates, it has lower variance at higher values (but actually it has very few points at higher values, and so these have a lot of leverage in the estimates). [I considered doing bootstrapping to adjust p-values, or weighted least squares but not sure how to weight, but haven't done this...thoughts on better ways to do this analysis welcome!]

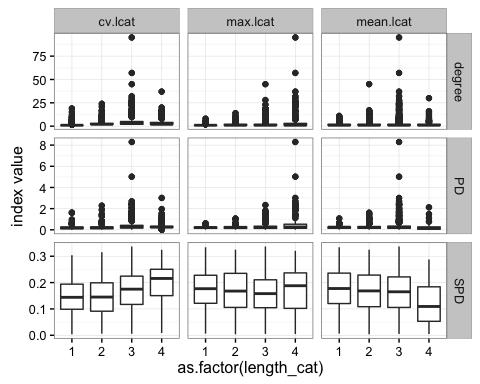
Results

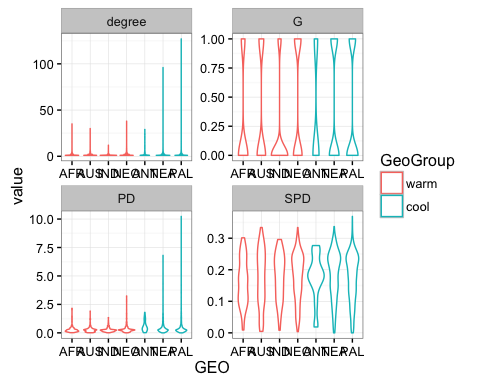
The allometric scaling model predicts that, for endoparasites, there should be a positive correlation among parasites’ generalism metrics and both the maximum host body size and the mean and CV body size across all host species. For some measures we do see a positive correlation (Fig. 1), and in particular there is a strong positive correlation between the coefficient of variation of host length and the mean phylogenetic distance between hosts (SPD). Plots are shown with loess ("locally weighted regression" smoothed lines to demonstrate the shape of the relationship). (HOW DO WE WANT TO DISPLAY THE REGRESSION RESULTS? JUST STICK THEM IN AN APPENDIX? ORGANIZE THEM INTO A TABLE?)

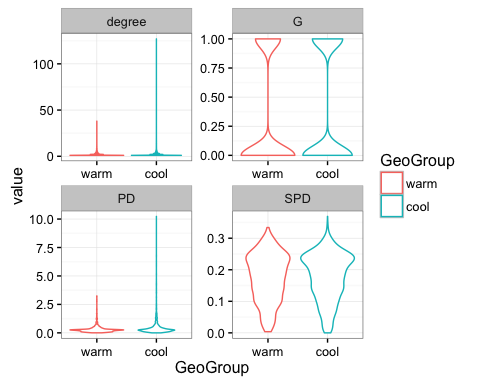
The allometric scaling model also predicts that for ectoparasites, there should be few generalist parasites of either very small bodied or very large bodied hosts. This seems to be the case for degree, PD, and PS, but not so clear for SPD (Fig. 2, 3).. However, because we categorized by quartiles, in some cases relatively low host body size metric values end up in the 4th quartile because the distribution of metric values is skewed with a long tail.











Discussion

The number of hosts a parasite can infect has important epidemiological and evolutionary implications (REFS). WE NEED TO SAY SOMETHING THAT CONTRASTS OUR APPROACH TO THE STUDY OF GENERALISM FROM OTHER GROUPS. Something like the following, which I basically just made up, but seems reasonable: “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 host, parasite, and environmental traits that influence host range (REFS). These studies have suggested a number of important factors thought to influence the evolution of host range (Table 1). Blah blah blah”

Here, we take a different approach, deriving simple mathematical models that incorporate host, parasite, and environmental characteristics that are likely to be generally important to the evolution of parasite generalism. In particular, we focus on variation in host body size, temperature, and parasite life cycle (endo- vs. ectoparasite, direct vs. trophic transmission).

The first model investigates the influence of host body size, temperature, and host infection site (endo- vs. ectoparasitism). The model makes several simple predictions. First, for parasites with a direct life cycle, endoparasites should have a wider host range than ectoparasites. Second, for endoparasites (be they directly or trophically transmitted), increasing host body size and increasing the similarity of body size between hosts both increase the likelihood of generalism evolving. Third, for endoparasites, generalism is most likely when the hosts have intermediate body size, or when the variation in body size is very large. Fourth, regardless of transmission mode or life cycle, generalism is more likely to evolve in colder environments.

* Some of the predictions from the model match the data analysis, but others don’t. Should outline these.
* We have a problem of very little data for endoparasites with direct life cycles. Correlation between traits – so generally ectos have direct and endos have complex.
* There is also a ton of missing data on the parasite traits – many of these parasites have hardly been studied. Perhaps this introduces a bias if parasites that are better studied are more likely to be included, and better studied parasites are more likely to have been found in more hosts.
* Issue with interpretation host traits model: in reality, what the model says is that, in a community with more species of similar body size, generalist parasites are likely to be able to invade. This is not the information that we can glean from the host-parasite association database, because it gives no information about the *potential* hosts within an environment.
* Generalism indices would be more informative if had abundance or prevalence data.
* Only include fish hosts.
* How does what we find relate to the literature outlined in the introduction table?
* Some sort of overall message?
* Mean body size is likely to go down as generalism increases in the dataset because there are more small-bodied hosts than large-bodied hosts.
* Difficulty of testing predictions for ectoparasites because those predictions are more sensitive to other parameter values.

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.

**Ethics**

Research on humans must include a statement detailing ethical approval and informed consent. Research using animals must adhere to local guidelines and state that appropriate ethical approval and licences were obtained. Please read our [editorial policies](http://royalsocietypublishing.org/ethics-and-policy)carefully before submission. Please send relevant documentation (e.g. consent forms) to the Editorial Office.

**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, and ARE acquired the data. JGW and SJP calculated the phylogenetic metrics. JGW and CEC wrote the first version of the article and all authors contributed to revisions and final editing.

All submissions, other than those with a single author, must include an Authors’ Contributions section which individually lists the specific contribution of each author. The list of Authors should meet all of the following criteria; 1) substantial contributions to conception and design, or acquisition of data, or analysis and interpretation of data; 2) drafting the article or revising it critically for important intellectual content; and 3) final approval of the version to be published. All contributors who do not meet all of these criteria should be included in the acknowledgements.

**Competing Interests**

All manuscripts must include a competing interests section. If you have no competing interests please state *‘I/We have no competing interests.’*  
  
Competing interests are defined as those that, through their potential influence on behaviour or content or from perception of such potential influences, could undermine the objectivity, integrity or perceived value of publication. For more details please read our [editorial policies](http://royalsocietypublishing.org/ethics-and-policy). You must disclose any relationships, whether professional or personal, with the Guest Editors of the issue.

**Funding**

Please list the source of funding for each author.

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. 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)

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. Buhnerkempe, M. G., Roberts, M. G., Dobson, A. P., Heesterbeek, H., Hudson, P. J. & Lloyd-Smith, J. O. 2015 Eight challenges in modelling disease ecology in multi-host, multi-agent systems. *Epidemics* **10**, 26–30. (doi:10.1016/j.epidem.2014.10.001)

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. 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)

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

7. 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)

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

9. 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)

10. 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)

11. Poulin, R. 1998 Host specificity. In *Evolutionary ecology of parasites: from individuals to communities*, London, UK: Chapman and Hall.

12. Molnár, P. K., Kutz, S. J., Hoar, B. M. & Dobson, A. P. 2013 Metabolic approaches to understanding climate change impacts on seasonal host-macroparasite dynamics. *Ecol. Lett.* **16**, 9–21. (doi:10.1111/ele.12022)

13. 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.

14. 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)

15. Froese, R. & Pauly, D. 2016 Fishbase.

16. 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)

17. 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)

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

19. 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.

20. 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)

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

22. 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)

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

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

25. 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)

26. R Core Team 2014 *R: A Language and Environment for Statistical Computing*. Vienna, Austria. http://www.r-project.org: R Foundation for Statistical Computing.

27. Choisy, M., P. Brown, S., D. Lafferty, K. & Thomas, F. 2003 Evolution of Trophic Transmission in Parasites: Why Add Intermediate Hosts? *Am. Nat.* **162**, 172–181. (doi:10.1086/375681)

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.