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, allometry

|  |  |  |
| --- | --- | --- |
| 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 how many host species they infect, with some parasite species infecting 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, identifying parasites that are likely to have unidentified alternative 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 generalism. 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 then explore these predictions using a database of over 20,000 fish- macroparasite associations. Within the database we see some evidence for these simple model 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 host-parasite specificity and identify key areas for future work.

Introduction

The range of hosts infected by a parasite species is a key factor affecting transmission: alternative hosts may enable parasite persistence in a population, with certain reservoir hosts crucial for the maintenance of transmission [1,2]. In addition, 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 hosts, 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. The EID literature, covering microbes to macroparasites, suggests that most parasites are generalists: 60% of human infectious diseases are zoonotic and 80% of pathogens of domestic animals infect multiple hosts [4–6]. However, theory suggests that parasites are generally resource (host) specialists, with interspecific competition selecting for an increase in ecological specialization (a narrow host range) [7]. Moreover, studies suggest that evolutionary transitions between specialism and generalism are bidirectional, despite the ‘dead-end’ theory predicting that parasites that evolve to become more specialist will have reduced potential to infect additional hosts [8,7]. There are few examples of studies that examine whether specialism or generalism is the ancestral state for macroparasites of animals, but for both feather lice of doves and gill monogeneans of African freshwater fish, host generalism appears to have derived from ancestral specialism [9,10].

Previous research on ecological drivers of parasite generalism or specialism have examined the relationship between host range and parasite or host traits and environmental factors (Table 1). Changes in these traits or factors could theoretically drive either the persistence (specialism) or diversification (generalism) of a parasite within its host(s), or lead to new species evolving [11]. For example, environmental change could lead to geographical isolation, driving host/parasite co-evolution and so specialization towards a single host species, or to geographical expansion, driving opportunities for host switching, and thus a wider host range [12] (and see Cable et al. this volume). In this study, we focus on the role of host body size and environmental temperature on host range evolution. The relationship between host size and parasite richness has been explored in depth, often with reference to island biogeography theory such that larger-bodied hosts represent larger habitat patches with more niches to support a range of parasite species [13]. By this theory, parasite species richness is predicted to increase with host body size, and this relationship has been observed in a variety of animal hosts [14], including for ectoparasitic monogeneans of freshwater fish [15]. Temperature is hypothesized to affect the host range of parasites due to temperature-dependent development of parasite free-living stages. Recent studies testing this hypothesis in gastro-intestinal parasites of primates reached opposite conclusions, with one finding a positive relationship between parasite species richness and hot weather (with a time lag of 4 weeks) [16] and the other finding a negative relationship [17]. Richness of metazoan ectoparasites on marine fish were observed to correlate with water temperature after controlling for phylogeny and sampling [18].

General predictions regarding correlation of traits with parasite host range may not apply equally across all host-parasite systems, and alternative hypotheses may lead to ambiguous predictions when interactions between traits are considered. Here, we use a mathematical model based on a framework of ecological theory to derive and justify predictions regarding the relationships between host body size and temperature and the evolution of parasite host range. Specifically, we develop a simple theoretical framework for studying host range evolution in a heterogeneous host population when there is a cost of generalism. We use allometric scaling relationships to investigate how variation in host body size and temperature affect whether specialism or generalism is the evolutionarily stable parasite strategy. We then calculate structural and phylogenetic generalism metrics [19] from an extensive data set of macroparasites of fish [20] to test the predictions arising from these models. 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** |
| 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 [21].  Higher number of host species per parasite and network connectance observed for endoparasites compared to ectoparasites of fish [22]. |
| 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].  Complex life cycles may constrain parasite diversification [23]. |
| 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 [12]. |
| 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 [24] 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 Monogeneans [21]. |
| Host body size\* | Continuous (here, maximum length of fish host) | Parasites of ungulates: parasite species richness found to increase with host body size across all parasites groups [14].  In monogeneans infecting fish, host body size is hypothesized to affect parasite attachment, so that those affecting larger fish are more specialized, due to lower diversity of large-bodied fish [25]  Guegan 1992 found strong association between host body size and parasite richness in tropical freshwater fish |

Model Derivation and Predictions*:* **Effect of host body size and temperature on generalism**

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. 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, implying that contact is parasite-controlled. 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 [26] 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 [27], 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. Substituting 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 . For this simple model, this analysis makes very intuitive 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. In particular, host carrying capacities and mortality rates will scale with host body size [28] 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.**

The mass of the secondary host is given by , where is the mass of the primary host, and we assume that the secondary host is smaller than the primary host (). This essentially assumes that the specialist parasite has evolved to exploit the host that maximizes its fitness, since the fitness of the specialist () is an increasing function of host body size. 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 the generalism index of each parasite and both the body size of its largest host.** Moreover, **there will be a positive correlation between generalism index and mean body size** () as , and for all *W*. However, because , the correlation between generalism index and maximum body size should be larger than that between generalism and mean body size.

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 . Note that if we allowed (secondary host larger than primary host), this would make invasion by the generalist even easier. 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 B 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.**

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 [20]. 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 [29]. 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 [30] 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* [31]*.* 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 [30] 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 [32] 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 [33] in the R package *ape* [34]. 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 2).

Table 2: Generalism metrics calculated from host-parasite database [19]

|  |  |  |
| --- | --- | --- |
| 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 [35], 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 [36], 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 [37] 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) [19,36] was generated for all parasites using *picante* in R [38]. 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 that generalism should be positively correlated with host body size and negatively correlated with temperature.

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. y fparasites 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.

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 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).

Results

*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 endoparasites, 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.

For some measures we do see a positive correlation (Table 3), 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).

The 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 (Table 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.

*Temperature/geographic range of all parasites*

The model predicts generalism is more likely in colder environments. We see higher degree, G, and SPD in cool regions.

Table 3: Model coefficients and 95% confidence intervals from generalised linear models of metrics by traits

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Infection site | Response (metric) | Predictor (trait) | | Coefficient | Confidence Interval | Unit |
| Endoparasite | degree | Mean host body length | | -0.00152 | -0.00211, -0.000931 | unit change in log degree per cm increase in length |
| Max host body length | | 0.00545 | 0.00507, 0.00583 |
| CV host body length | | 1.43589 | 1.27125, 1.60134 |
| G | Mean host body length | | -0.00105 | -0.00194, -0.000183 | log odds ratio per cm increase in length |
| Max host body length | | 0.00652 | 0.00567, 0.00740 |
| PD | Mean host body length | | -0.00180 | -0.00264, -0.000956 | unit change in log(PD+0.01) per cm increase in length |
| Max host body length | | 0.00209 | 0.00170, 0.00249 |
| CV host body length | | 1.16029 | 1.02587, 1.29472 |
| SPD | Mean host body length | | -0.000093 | -0.000138, -0.0000477 | unit change in SPD per cm increase in length |
| Max host body length | | 0.000240 | 0.000209, 0.000271 |
| CV host body length | | 0.06276 | 0.05289, 0.07262 |
| SES-PD | Mean host body length | | -0.0108949 | -0.0150664, -0.0067235 | unit change in SES-PD per cm increase in length |
| Max host body length | | -0.0102986 | -0.0122554, -0.0083417 |
| CV host body length | | 0.7280691 | -0.0079981, 1.4641362 |
| Ectoparasite | degree | Mean host body length | Q2 | 0.2235999 | 0.1423384, 0.3049629 | unit change in log degree compared to 1st quartile category |
| Q3 | 0.406262 | 0.334761, 0.4780085 |
| Q4 | 0.0995918 | -0.0302969, 0.2285828 |
| Max host body length | Q2 | 0.2568647 | 0.1742667, 0.339628 |
| Q3 | 0.5386372 | 0.4648116, 0.6128243 |
| Q4 | 1.019527 | 0.9210028, 1.1183049 |
| CV host body length | Q2 | 0.1901761 | 0.0808269, 0.2996433 |
| Q3 | 0.674365 | 0.5707044, 0.7783339 |
| Q4 | 0.2631751 | 0.1548948, 0.3716116 |
| G | Mean host body length | Q2 | 0.3274065 | 0.1475052, 0.5078821 | log odds ratio compared to 1st quartile category |
| Q3 | 0.466491 | 0.3068503, 0.6274147 |
| Q4 | -0.2616669 | -0.5719861, 0.0389363 |
| Max host body length | Q2 | 0.7788821 | 0.590882, 0.968715 |
| Q3 | 1.0020415 | 0.8291696, 1.1774999 |
| Q4 | 1.0857277 | 0.8338062, 1.3381528 |
| PD | Mean host body length | Q2 | 0.0292508 | -0.0882624, 0.1467641 | unit change in log(PD+0.01) compared to 1st quartile category |
| Q3 | 0.0802839 | -0.0241731, 0.1847409 |
| Q4 | -0.4862968 | -0.6968441, -0.2757497 |
| Max host body length | Q2 | 0.0143388 | -0.1141987, 0.1428764 |
| Q3 | 0.110301 | -0.0082226, 0.2288246 |
| Q4 | 0.3628297 | 0.2027306, 0.5229288 |
| CV host body length | Q2 | 0.11333 | 0.0033182, 0.2233418 |
| Q3 | 0.5111788 | 0.4013089, 0.6210488 |
| Q4 | 0.5073291 | 0.3973884, 0.6172698 |
| SPD | Mean host body length | Q2 | 0.010031 | 0.0022415, 0.0178204 | unit change in SPD compared to 1st quartile category |
| Q3 | 0.012963 | 0.0060482, 0.0198778 |
| Q4 | -0.0244058 | -0.0366981, -0.0121134 |
| Max host body length | Q2 | 0.0255724 | 0.0179175, 0.0332273 |
| Q3 | 0.0291506 | 0.0221589, 0.0361422 |
| Q4 | 0.0384315 | 0.0276104, 0.0492526 |
| CV host body length | Q2 | -0.0019701 | -0.012041, 0.0081007 |
| Q3 | 0.0210478 | 0.01099, 0.0311057 |
| Q4 | 0.0550058 | 0.0449414, 0.0650701 |
| SES-PD | Mean host body length | Q2 | -0.9680191 | -1.6079538, -0.3280843 | unit change in SES-PD compared to 1st quartile category |
| Q3 | -1.4574457 | -2.0262805, -0.8886109 |
| Q4 | -3.6327163 | -4.77928, -2.4861527 |
| Max host body length | Q2 | -0.8936799 | -1.5891787, -0.1981812 |
| Q3 | -2.2616954 | -2.90301, -1.6203809 |
| Q4 | -2.3225065 | -3.1887804, -1.4562327 |
| CV host body length | Q2 | -0.7854338 | -1.394371, -0.1764966 |
| Q3 | -0.4861553 | -1.0943073, 0.1219967 |
| Q4 | 1.9586667 | 1.350123, 2.5672105 |
| Both | degree | Geographic group (ref=”warm”) | | 0.4052638 | 0.361268, 0.4492741 | unit change in log degree for cool group |
| G | 0.2775012 | 0.187575, 0.3676597 | Log odds ratio for cool group |
| PD | 0.2072072 | 0.1407433, 0.2736711 | unit change in log(PD+0.01) for cool group |
| SPD | 0.0027182 | -0.0027517, 0.008188 | unit change in SPD for cool group |
| SES-PD | -0.7006804 | -1.0391793, -0.3621815 | unit change in SES-PD for cool group |

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 host, parasite, and environmental traits that influence host range [14–18]. These studies have suggested a number of important factors thought to influence the evolution of host range (Table 1).

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).

there is another study which uses allometric scaling to think about parasites (abundance and size but not diversity, would be good to add this as contrast to our approach but supporting use of allometry) [http://dx.doi.org/10.1017/S0031182015001444](http://dx.doi.org/10.1017/S0031182015001444" \t "_blank)

The model investigates the influence of host body size, temperature, and host infection site (endo- vs. ectoparasitism, making 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. In general, these predictions match hypotheses proposed in previous work that did not use explicit allometric modelling (Table 1), but allow for more nuanced predictions such as the non-linear relationship between ectoparasite richness and host body size.

The model uses evolutionary invasion analysis to determine when a parasite that infects two hosts will be able to invade a system where another parasite infects only one host. This theoretical invasion possibility was assumed to translate to observed generalism in a database of host-parasite interactions. We measured structural generalism based on number of hosts (degree and G), and also looked at phylogenetic generalism in which a parasite which infects more distantly related hosts is considered more generalist than a parasite infecting closely related hosts. Therefore, we assumed that phylogenetic generalism is more likely under the same conditions as structural generalism. Phylogenetic generalism tends to correlate with number of hosts [19]. In addition, we included a measure of phylogenetic generalism that is scaled to remove the association with number of hosts (SES-PD), which tests whether phylogenetic generalism in itself matches the model predictions.

In general, the observed relationships were fairly consistent across all the generalism metrics. However, for geography/temperature, SES-PD had the opposite relationship to the other metrics, perhaps revealing a difference between phylogenetic generalism and parasite richness.

Some other limitations of the model:

* + for ectotherms doesn’t the immune response vary with temperature, such that they are immune supressed when cold, immune active when warm. If so, increased temperature might (via immunity) protect from infection and so, comparatively, reduce mortality
  + 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.
  + Difficulty of testing predictions for ectoparasites because those predictions are more sensitive to other parameter values.
  + Limitations of endo/ecto division when GI is possibly kind of ecto
    - But Poulin also found a difference in the relationship between parasite richness and ecological traits for endos vs ectos (ectos showing no relationship with the traits examined) http://www.jstor.org/stable/2937061?seq=1#page\_scan\_tab\_contents

When model predictions were tested in a large database of host-parasite interactions, we found generally good support for the predictions (Table 3).

As expected in the model, for endoparasites, the coefficient of variation of host body size had the strongest relationship with generalism. The effect of maximum host body size was also positively correlated with generalism, while mean body size tended to have a small negative correlation with generalism. This may be due to mean body size decreasing as generalism increases in the dataset because there are more small-bodied hosts than large-bodied hosts.

There are a number of potential limitations or confounders affecting the observed relationships in the data.

* Potential limitations/confounders for data
  + Amy: Also, are there interactions going on, i.e., endoparasites (wider host range) in warmer waters (narrower host range) – does this types of thing partly explain why the results are a bit noisy?
  + (identifying what is a species of parasite is difficult to begin with which would obviously affect measures of parasite host range, and see Walker et al in this issue)

Phenotypic plasticity is important characteristic of generalist parasites

Smith et al 2008 Costa Rica 171 morphs but COI revealed another 142 spp – appreciation of species diversity switched perception on the level of host specialism within a taxa <http://www.ncbi.nlm.nih.gov/pubmed/18716001> [39]

* + Could temperature effect be confounded with available hosts and higher host diversity in warm regions? Obvious roughness of using region as a proxy for temperature.
  + 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.
  + Generalism indices would be more informative if had abundance or prevalence data. What is the meaning of phylogenetic vs structural diversity/generalism metrics in relation to the model?
  + Only include fish hosts.
  + Categorizing ectos by quartiles potentially limiting.

The model predictions are somewhat supported by a dataset of fish parasites, but also demonstrate what additional data would be useful to collect in these types of datasets. If we had information on parasite abundance then this could be investigated alongside generalism as they are likely to be linked, and we could also then look at the relative differences in abundance as part of the measure of generalism.

Overall: using an allometric scaling model, we find support for previous hypotheses that parasite richness increases with body size, but contradict the hypothesis that parasite richness increases with temperature. The interactions between temperature and parasite infection are likely to be complex and may have different effects on different life stages, or interact with other factors like latitudinal differences in host availability (similar argument about complexity of effect of temperature is outlined in review on host-parasite cospeciation [11]). By using a model which uses allometric scaling theory, we can demonstrate a mechanistic effect of certain traits on parasite host range and strengthen or challenge existing hypotheses that are based on other areas of ecological theory.

How does this relate to transmission? The evolution of host range is closely linked to transmission, particularly in regards to reservoir hosts, spillover/emergence, and changes in virulence. As outlined in the beginning of the introduction.

Clay’s contribution to the Discussion

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. 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)

8. 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)

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. 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)

11. Brunner, F. S. & Eizaguirre, C. 2016 Can environmental change affect host-parasite mediated speciation? *Zoology* **in press**, 1–11. (doi:10.1016/j.zool.2016.04.001)

12. 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)

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

14. 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.

15. 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)

16. Benavides, J. A., Huchard, E., Pettorelli, N., King, A. J., Brown, M. E., Archer, C. E., Appleton, C. C., Raymond, M. & Cowlishaw, G. 2012 From parasite encounter to infection: Multiple-scale drivers of parasite richness in a wild social primate population. *Am. J. Phys. Anthropol.* **147**, 52–63. (doi:10.1002/ajpa.21627)

17. Poirotte, C., Basset, D., Willaume, E., Makaba, F., Kappeler, P. M. & Charpentier, M. J. E. 2016 Environmental and individual determinants of parasite richness across seasons in a free-ranging population of Mandrills ( *M* *andrillus sphinx)*. *Am. J. Phys. Anthropol.* **159**, 442–456. (doi:10.1002/ajpa.22888)

18. Poulin, R. & Rohde, K. 1997 Comparing the richness of metazoan ectoparasite communities of marine fishes: controlling for host phylogeny. *Oecologia* **110**, 278–283. (doi:10.1007/s004420050160)

19. 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)

20. 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)

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

22. 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)

23. Poulin, R. & Morand, S. 2004 *Parasite Biodiversity*. Washington, D.C.: Smithsonian Institution.

24. 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)

25. 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)

26. 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)

27. 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) http://yahoo.com/

28. Savage, V. M., Gillooly, J. F., Woodruff, W. H., West, G. B., Allen, A. P., Enquist, B. J. & Brown, J. H. 2004 The predominance of quarter-power scaling in biology. *Funct. Ecol.* **18**, 257–282.

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

30. 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)

31. 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)

32. 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)

33. 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.

34. 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)

35. 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)

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

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

38. 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)

39. Smith, M. A., Rodriguez, J. J., Whitfield, J. B., Deans, A. R., Janzen, D. H., Hallwachs, W. & Hebert, P. D. N. 2008 Extreme diversity of tropical parasitoid wasps exposed by iterative integration of natural history, DNA barcoding, morphology, and collections. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 12359–12364. (doi:10.1073/pnas.0805319105)

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???