Host-generalism as a macroparasite transmission strategy: theory and meta-analysis

**Keywords:**

|  |  |  |
| --- | --- | --- |
| Name | Affiliations | ORCID |
| Joanne Cable |  |  |
| Clay Cressler |  |  |
| Amy Ellison |  |  |
| Amy Hurford |  |  |
| Stephen Price |  |  |
| 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 biologial 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 a combination of mathematical models and analysis of host-parasite data (focusing on macroparaistes of fish) to identify host and parasite traits that have the potential to drive the evolution of host generality. Firstly, using a model of a host-parasite system we have identifed traits that select for parasites that have one or multiple species of hosts. Secondly, we calculated indices of host generality using an existing database of fish host-parasite associations, and then combine these indices with data on host size and parasite life history traits. Looking across parasite taxonomic groups, we evaluate whether such traits are correlated with increased or decreased generalism. By combining these two approaches, we develop a theoretical basis for interpreting empirical data on host-parasite specificity.

Introduction

The community of hosts that a parasite infects plays a large role in determining parasite transmission patterns. For a parasite that is able to infect multiple species of host species, there are implications for the evolution of virulence [1], maintenance of transmission [2,3], and host-switching or disease emergence in humans [4].

From an ecological perspective, a parasite that infects multiple hosts species – a generalist parasite – will be able to thrive in a variety of conditions by using a variety of resources. In contrast, a parasite infecting only a single host speices, while having a more limited ranage of infection possibilities be able to evolve great specialism at exploitign that one host species. Quantifying hos host speficic or generla a parasite speceis is, isnot straightforward – frequent measures inlcude the number of recorded host species, the parasite’s abundance and/or prevalence within those hosts, or the phylogenetic relationships between hosts [5].

According to research related to the emergence of infectious disease (ranging from microbes to parasites), most pathogens are generalists; 60% of human infectious diseasess are zoonotic and 80% of pathogens of domestic animals infect multiple hosts [4,6,7]. However, other work suggests that parasites are generally resource specialists. There apears to be soemthign of a paradox, in that parasites commonly being resource specialists with restricted host ranges, but that in parasite evolutionary lineages that are commonly shifts rom one hsot 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 rnage). 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 and diversification of host-parasite systems, with geographical isolation driving host/parasite co-evolution and so specialization towards a single host speices, and geographical expansion driving opportunities for host switching, and thus a wide rhost range [10].

Characteristics of both the parasite and the host will interact with external factors to drive the potential for evolving generalism (Table 1). Parasites: Gender/sexual reproduction, motility and environmental transmission, size, fecundity. Hosts: Reproduction, habitat, diet, group size, moulting.

In this paper we use a theoretical approach to explore relationships among host and parasite traits, and how this relates to parasite generalism. We then use an extensive data set of macroparasites of fish to test the predictions arising from this theoretical apprpoach. In particular, we focus on parasite traits related to transmission route and life cycle, and host traits that are allometrically related to body size, such as carrying capacity and mortality rate. 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]. |

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]. We have added additional records for 105 crustacean parasite species, and incorporated data on parasite life history traits including reproductive strategy, life cycle stages, and transmission routes. Host names were validated by comparison to the internet. To avoid counting intermediate host diversity as for parasites with a complex life cycle, all non-definitive hosts were excluded. After data cleaning, we were left with 23,360 unique host-parasite associations between 8,847 parasite species and 4,243 fish hosts.

**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, or degree), and phylogenetic host specificity (the average phylogenetic distinctiveness, in this case the number of taxonomic steps required to reach a node common to each pair of host species, SPD, and its variance, VarSPD) [5,15]. SPD is given a value of 0 when the parasite has only one host, and VarSPD is only calculated for hosts that have more than one host.

**Data analysis**

The generalism metric for each parasite species was compared to measures of the characteristics described in Table 1. Metrics for endoparasites with direct life cycles were compared separately to summary measures (mean, maximum, standard deviation (SD), and coefficient of variation (CV)) of the maximum host length reported for each of their hosts using linear models. Note that SD and CV of the host length are only calculated for parasites with more than one host.

Due to the distribution of degree for ectoparasites in this case we used a negative binomial generalized linear model (GLM) with a log link function; standard linear models were used for SPD and VarSPD. The mean maximum length was additionally divided into a categorical variable according to quartiles.

The effect of geographical range on parasites’ with a direct life cycle was calculated for ectoparasites only, due to the small number of endoparasites with a direct life cycle. As above, a linear model/ANOVA was used to assess the effect of geographical category on SPD and VarSPD, while a negative binomial GLM was used for degree. Geographical regions are 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, and in these cases the generalism metrics were calculated separately for each region.

Whether parasites has complex or simple life cycles, and theethr transisiosn was rtiohci or not, were each treated as binary parasite characteristics and compared to the parasite indices as for geographical categories. In this case, endo and ectoparasites are considered in the same analysis.

The statistical analysis was conducted using R v3.2.3 [16]. A cut off of P < 0.01 was used for significance.

Model Derivation and Predictions

**Host traits**

We begin by defining a simple host-parasite system with two hosts and two environmentally transmitted parasites. 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. The dynamics of the system can be described by the following system of equations:

We follow the dynamics of susceptible primary and susceptible 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 . In the absence of infection, primary and secondary hosts attain the host species-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 species-specific rates and . Parasites are shed from infected hosts at the host-specific rates and . 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 specialist-only equilibrium host abundances. 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 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 is,

One way to explore how host and parasite characteristics influence the likelihood of generalism evolving () is to look at the derivative of with respect to system parameters. However, this analysis is not particularly revealing, as the model in its current state does not include any relevant trade-offs that constrain the relationships between parameters (see Appendix A). It is also challenging to connect the parameters of such a general model with empirical data on host-parasite associations.

To facilitate a compariosn 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- and temperature-dependent criteria for the evolution of generalism. In particular, 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. Thus we can 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*).

To do this, we ask how changes with changes in these parameters (that is, we look at the partial 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, **there should be a positive correlation among parasites’ generalism index and both the maximum host body size and the mean or median body size across all host species**.

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 negative correlation between generalism index and the variance in host body size**. Putting these two results together, **there should be a very strong positive correlation between generalism index and the coefficient of variation in host body size**.

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.

For ectoparasites, 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 predcits 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).

The effect of temperature will be the same for both endo- and ectoparasites: **generalists will have an easier time invading when temperatures are colder.**

**Parasite traits**

In this analysis we examine whether a parasite that has a complex life cycle or that is trophially transmitted is also likely to be a generalist. Our approach is to extend Choisy’s model for the evolution of trophic transmission [17] to include generalist parasites and to consider complex life cycles that do not necessarily involve trophic transmission. The Choisy model assumes that parasites are shed into the environment from an infected terminal host. Parasites in the environment may either infect only the terminal host (a simple life cycle) or may infect an intermediate host that is consumed by the terminal host (trophic transmission; [Figure 1, 17]).

To extend the Choisy model for generalist parasites, we assume that there are two terminal host species. The rate that parasites are shed from the terminal host i into the environment is λi, however, for simplicity in interpreting the conditions for generalism to evolve, we make several simplifying assumptions one of which is that the shedding rate from both terminal hosts are equal, λ1=λ1=λ. We also assume that the mortality rates of both terminal hosts are equal, μ11=μ12=μ1 and we let the mortality rate of an infected intermediate host and of the parasite in the environment be μ2 and μ3 respectively. The difference between a parasite with a complex life cycle and a parasite with trophic transmission is that the intermediate hosts are killed when trophic transmission occurs but not for a complex life cycle. As such, we introduce a parameter, τ, such that when τ=0 a complex life cycle occurs and when τ=1 trophic transmission occurs. For simplicity it is assumed that if trophic transmission occurs in one host it also occurs in the other host.

N1i is the abundance of the terminal host i and N2 is the abundance of the intermediate host. The remaining parameters are the ones that are allowed to evolve parasites that have some combination of generalism, complex life cycles, and/or trophic transmission. The rate that the parasite infects the terminal host i from the environment is β1i. The rate that the parasite infects the intermediate host from the environment is β2 and the rate that the intermediate host interacts with the terminal host i leading to the transmission of the infection is pi.

The system of differential equations describing the spread of the parasite between the different hosts is,

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

For the system of equations (1)-(4), the basic reproduction number of the parasite is,

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

where the next generation matrix is,



where the elements in the next generation matrix are: the average number of intermediate hosts infected from a parasite in the environment,

the average number of terminal hosts infected from a parasite in the environment,

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

Note that this transmission is indirect since the infected intermediate host does not shed directly into the environment, but rather infects the terminal host, which sheds the parasite into the environment. Finally, the average number of parasites shed into the environment from an infected terminal host is,

There are three key aspects of the model formulation (equations 1-4) that affect our conclusions regarding generalism, complex life cycles, and trophic transmission. These are:

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 a complex life cycle or trophic transmission, the same intermediate host can infect either terminal host

**Are parasites with complex life cycles more likely to be generalists – it is unclear**

To determine whether parasite that have complex life cycles 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 A 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, β11=A- β2, then the condition for generalism to evolve is N12 > N11, 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 Si-SiSi, meaning that a specialist parasite with a simple life cycle (Si) evolves to infect two hosts and has a simple life cycle with each host (SiSi).

Given the model formulation (equations 1-4), there are several ways a generalist parasite with a complex life cycle might evolve. Firstly, the parasite may have a complex life cycle with host 1 and evolve the ability to be transmitted from the intermediate host to the second terminal host: this is denoted by the scenario Cx-CxCx. Alternatively, the parasite may evolve the ability to infect the intermediate host and since the model formulation assumes that the same intermediate host can infect both terminal hosts (A3), a generalist parasite with a complex lifecycle evolves from a parasite with a simple life cycle, Si-CxCx. While there are other possibilities (i.e., a parasite that has a complex life cycle evolves to have a simple life cycle with the second host, Cx-SiCx, resulting in a generalist parasite with a complex life cycle) these possibilities do not lead to different conclusions and for simplicity we do not discuss them.

Our analysis (summarized in Table 1) reveals that a parasite with a complex life cycle is equally likely to evolve to be a generalist (Cx-CxCx) as a parasite with a simple life cycle (Si-SiSi). Both of these evolutionary transitions (Si-SiSi and Cx-CxCx) are more likely to occur than a parasite with a simple life cycle evolving to be a generalist with a complex life cycle (Si-CxCx) which requires that the mortality rate of the intermediate host, μ2, be less than,

|  |  |
| --- | --- |
|  | (6) |

in the absence of a life history tradeoff and less than,

|  |  |
| --- | --- |
|  | (7) |

when there is a linear life history tradeoff involving the parasite’s ability to infect the intermediate host and the terminal hosts (see Table 1 for further details). Both of the conditions (6) and (7) were derived under the assumptions that p1 = p2 = p and β11= β12= β1. The reason that μ2 must be sufficiently small for Si-CxCx to evolve is because infecting the intermediate host removes parasites from the environment that might have otherwise infected the terminal host via the simple life cycle pathway (A1). As such, Si-CxCx is less likely to evolve than Si-SiSi. If the parasite already has a complex life cycle, no additional parasites are removed from the environment when generalism evolves because the same intermediate host infects both terminal hosts (A3). As such, Cx-CxCx and Si-SiSi require the same conditions to evolve.

So far, we have considered parasites that first evolve a complex life cycle and then evolve generalism, however, we might also ask whether a parasite that is a generalist is more likely to evolve a complex life cycle than a parasite that is a specialist. For the specialist parasite, the condition for a complex life cycle to evolve (Si-Cx) is that the mortality rate of the intermediate host, μ2, must be less than,

|  |  |
| --- | --- |
|  | (8) |

If we compare this condition to the condition for a parasite to evolve to be a generalist and have a complex life cycle (Si-CxCx), the parasite that is a generalist with a complex life cycle is more likely to evolve since C1 > C2. The reason that the condition for Si-CxCx to evolve is less strict is that there is no cost to being able to infect the second terminal host (A3) because infected intermediate hosts are not destroyed when transmission occurs for a complex life cycle. When a life history trade-off between the ability to infect the intermediate and the terminal hosts from the environment occurs the condition for Si-Cx to evolve is that μ2 must be less than,

|  |  |
| --- | --- |
|  | (9) |

Overall, we conclude that there is no clear relationship between whether a parasite that has a complex life cycle is more likely to be a generalist than a parasite with a simple life cycle. We found mixed results for each of the scenarios that we examined (Table 1). The evolution of a generalist parasite with complex life cycles (Si-CxCx) was more likely when compared to one reference scenario (Si-Cx), but less likely when compared to another (Si-SiSi). For the other scenarios considered both simple and complex life cycles were equally likely to be associated with generalism.

**Parasites that are trophically transmitted are less likely to be generalists**

The conditions for a parasite that has trophic transmission to evolve to be a generalist (Tr-TrTr) are the same as the conditions for parasites with simple or complex life cycles to evolve to be generalist (Si-SiSi or Cx-CxCx; Table 1). The scenario Tr-TrTr assumes that a parasite, which is consumed by the second host, develops the ability to infect that host (p2 evolves to be positive). Alternatively, a parasite with a simple life cycle might evolve the ability to infect an intermediate host that then infects the two terminal hosts that consume it. This scenario, Si-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,

|  |  |
| --- | --- |
|  | (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. Of all the scenarios we consider, this one is the least likely to evolve since C1>C2>C3 and C1t > C2t> C3t > C4t. For a parasite that has a simple 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), and for these reasons Si-TrTr is least likely 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 (Si-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 (Si-TrTr; μ2 < C4t; equation (11)) is that the abundance of the intermediate host must be less than the sum of the abundances of both terminal hosts N2 < N11 + N12. 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 (Choisy et al. 2003).

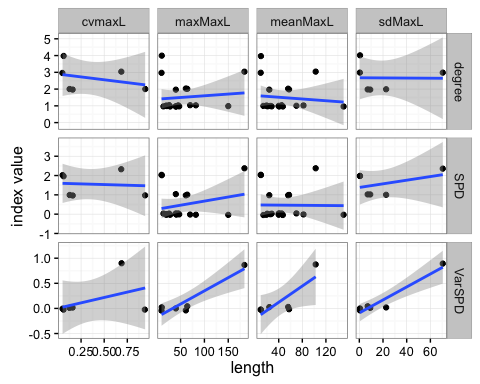
Overall, we conclude that parasites that have trophic transmission are less likely to be generalists than parasites with simple or complex life cycles. This is because for all the scenarios we considered the evolution of trophically transmitted generalist parasites was either equally likely or less likely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

Table 2: Are parasites that have complex life cycles or trophic transmission more likely to be generalists than parasites that have simple 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 | | | | | |
| Si-Cx | β2 | β12=0  p2=0  τ=0 | - | μ2 < C2 | Reference scenario |
| β11=A- β2 | μ2 < C2t | Reference scenario |
| Si-SiSi | β12 | β2=0 | - | Always satisfied | Reference scenario |
| β11=A- β12 | N11<N12 | Reference scenario |
| Cx-CxCx | p2 | τ=0 | - | Always satisfied | Si-SiSi is equally likely |
| p1=A-p2 | N11<N12 | As above |
| Si-CxCx | β2 | τ=0  p1=p2=p  β11 = β12= β1 | - | μ2 < C1 | Si-SiSi is more likely than Si-CxCx due to A1.  Si-Cx is equally likely because A1 affects both Si-Cx and Si-CxCx equally. |
| β1=A- β2 | μ2 < C1t | Si-SiSi - the conditions are not directly comparable  Si-Cx is less likely than Si-CxCx due to A3. |
| Trophic transmission | | | | | |
| Si-Tr | β2 | β12=0  p2=0  τ=1 | - | μ2 < C3 | Reference scenario |
| β11=A- β2 | μ2 < C3t | Reference scenario |
| Tr-TrTr | p2 | τ=1 | - | Always satisfied | Si-SiSi is equally likely  Cx-CxCx is equally likely |
| p1=A-p2 | N11<N12 | As above |
| Si-TrTr | β2 | τ=1  p1=p2=p  β11 = β12= β1 | - | μ2 < C3 | Si-SiSi is more likely than Si-TrTr due to A1.  Si-CxCx is more likely than Si-TrTr due to A2.  Si-Tr is equally likely |
| β1=A- β2 | μ2 < C4t | Si-SiSi – the conditions are not comparable  Si-CxCx is more likely than Si-TrTr due to A2.  Si-Tr is more likely than Si-TrTr. |

Results (Database)

## Endoparasites & Host Body Size



summary(lm(data=hp.endoDirect,cvmaxL~degree))

##   
## Call:  
## lm(formula = cvmaxL ~ degree, data = hp.endoDirect)  
##   
## Residuals:  
## 7 11 16 17 18 20   
## -0.23790 0.50639 -0.27655 -0.31015 0.39853 -0.08031   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 0.7336 0.6118 1.199 0.297  
## degree -0.1486 0.2210 -0.673 0.538  
##   
## Residual standard error: 0.4034 on 4 degrees of freedom  
## (14 observations deleted due to missingness)  
## Multiple R-squared: 0.1016, Adjusted R-squared: -0.123   
## F-statistic: 0.4523 on 1 and 4 DF, p-value: 0.5381

summary(lm(data=hp.endoDirect,cvmaxL~SPD))

##   
## Call:  
## lm(formula = cvmaxL ~ SPD, data = hp.endoDirect)  
##   
## Residuals:  
## 7 11 16 17 18 20   
## -0.2642 0.5764 -0.2065 -0.2401 0.3896 -0.2552   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 0.41857 0.50588 0.827 0.455  
## SPD -0.05221 0.30557 -0.171 0.873  
##   
## Residual standard error: 0.4241 on 4 degrees of freedom  
## (14 observations deleted due to missingness)  
## Multiple R-squared: 0.007245, Adjusted R-squared: -0.2409   
## F-statistic: 0.02919 on 1 and 4 DF, p-value: 0.8726

summary(lm(data=hp.endoDirect,cvmaxL~VarSPD))

##   
## Call:  
## lm(formula = cvmaxL ~ VarSPD, data = hp.endoDirect)  
##   
## Residuals:  
## 7 11 16 17 18 20   
## -2.176e-01 6.753e-01 -1.077e-01 -1.413e-01 2.498e-16 -2.086e-01   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 0.2676 0.1701 1.573 0.191  
## VarSPD 0.4711 0.4686 1.005 0.372  
##   
## Residual standard error: 0.3803 on 4 degrees of freedom  
## (14 observations deleted due to missingness)  
## Multiple R-squared: 0.2017, Adjusted R-squared: 0.00214   
## F-statistic: 1.011 on 1 and 4 DF, p-value: 0.3716

summary(lm(data=hp.endoDirect,maxMaxL~degree))

##   
## Call:  
## lm(formula = maxMaxL ~ degree, data = hp.endoDirect)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -53.147 -27.207 -8.057 8.497 125.550   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 42.860 20.516 2.089 0.0512 .  
## degree 5.197 11.845 0.439 0.6661   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 45.87 on 18 degrees of freedom  
## Multiple R-squared: 0.01058, Adjusted R-squared: -0.04439   
## F-statistic: 0.1925 on 1 and 18 DF, p-value: 0.6661

summary(lm(data=hp.endoDirect,maxMaxL~SPD))

##   
## Call:  
## lm(formula = maxMaxL ~ SPD, data = hp.endoDirect)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -61.273 -23.530 -4.380 3.924 108.245   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 44.38 11.68 3.800 0.00131 \*\*  
## SPD 13.45 12.88 1.044 0.31032   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 44.78 on 18 degrees of freedom  
## Multiple R-squared: 0.05709, Adjusted R-squared: 0.004708   
## F-statistic: 1.09 on 1 and 18 DF, p-value: 0.3103

summary(lm(data=hp.endoDirect,maxMaxL~VarSPD))

##   
## Call:  
## lm(formula = maxMaxL ~ VarSPD, data = hp.endoDirect)  
##   
## Residuals:  
## 7 11 16 17 18 20   
## -2.710e+01 2.900e+00 2.690e+01 2.390e+01 5.329e-15 -2.660e+01   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 37.10 11.72 3.167 0.03397 \*   
## VarSPD 165.26 32.28 5.119 0.00689 \*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 26.2 on 4 degrees of freedom  
## (14 observations deleted due to missingness)  
## Multiple R-squared: 0.8676, Adjusted R-squared: 0.8345   
## F-statistic: 26.2 on 1 and 4 DF, p-value: 0.006892

summary(lm(data=hp.endoDirect,meanMaxL~degree))

##   
## Call:  
## lm(formula = meanMaxL ~ degree, data = hp.endoDirect)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -33.162 -25.490 -7.162 13.228 102.838   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 51.177 15.933 3.212 0.00483 \*\*  
## degree -4.015 9.199 -0.436 0.66768   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 35.63 on 18 degrees of freedom  
## Multiple R-squared: 0.01047, Adjusted R-squared: -0.0445   
## F-statistic: 0.1905 on 1 and 18 DF, p-value: 0.6677

summary(lm(data=hp.endoDirect,meanMaxL~SPD))

##   
## Call:  
## lm(formula = meanMaxL ~ SPD, data = hp.endoDirect)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -34.958 -24.526 -5.376 11.475 104.624   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 45.3761 9.3397 4.858 0.000126 \*\*\*  
## SPD -0.4756 10.3001 -0.046 0.963676   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 35.81 on 18 degrees of freedom  
## Multiple R-squared: 0.0001185, Adjusted R-squared: -0.05543   
## F-statistic: 0.002132 on 1 and 18 DF, p-value: 0.9637

summary(lm(data=hp.endoDirect,meanMaxL~VarSPD))

##   
## Call:  
## lm(formula = meanMaxL ~ VarSPD, data = hp.endoDirect)  
##   
## Residuals:  
## 7 11 16 17 18 20   
## -2.190e+01 -7.363e+00 2.614e+01 2.464e+01 4.441e-15 -2.151e+01   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 31.36 10.69 2.933 0.0427 \*  
## VarSPD 80.22 29.47 2.722 0.0529 .  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 23.91 on 4 degrees of freedom  
## (14 observations deleted due to missingness)  
## Multiple R-squared: 0.6495, Adjusted R-squared: 0.5618   
## F-statistic: 7.412 on 1 and 4 DF, p-value: 0.05285

summary(lm(data=hp.endoDirect,sdMaxL~degree))

##   
## Call:  
## lm(formula = sdMaxL ~ degree, data = hp.endoDirect)  
##   
## Residuals:  
## 7 11 16 17 18 20   
## -17.778 3.924 -9.511 -11.632 52.214 -17.218   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 19.6076 45.3709 0.432 0.688  
## degree -0.4523 16.3860 -0.028 0.979  
##   
## Residual standard error: 29.92 on 4 degrees of freedom  
## (14 observations deleted due to missingness)  
## Multiple R-squared: 0.0001905, Adjusted R-squared: -0.2498   
## F-statistic: 0.000762 on 1 and 4 DF, p-value: 0.9793

summary(lm(data=hp.endoDirect,sdMaxL~SPD))

##   
## Call:  
## lm(formula = sdMaxL ~ SPD, data = hp.endoDirect)  
##   
## Residuals:  
## 7 11 16 17 18 20   
## -25.6984 13.9379 0.5029 -1.6185 38.4669 -25.5908   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) -8.792 32.627 -0.269 0.801  
## SPD 17.482 19.708 0.887 0.425  
##   
## Residual standard error: 27.35 on 4 degrees of freedom  
## (14 observations deleted due to missingness)  
## Multiple R-squared: 0.1644, Adjusted R-squared: -0.04453   
## F-statistic: 0.7868 on 1 and 4 DF, p-value: 0.4252

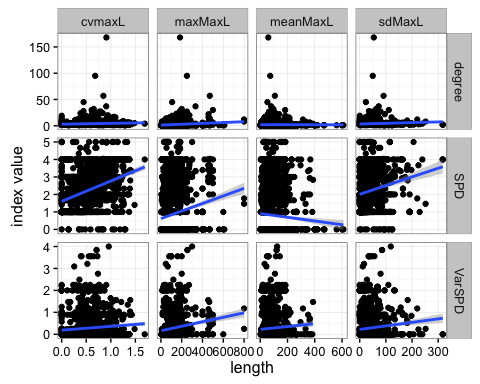
summary(lm(data=hp.endoDirect,sdMaxL~VarSPD))

##   
## Call:  
## lm(formula = sdMaxL ~ VarSPD, data = hp.endoDirect)  
##   
## Residuals:  
## 7 11 16 17 18 20   
## -7.516e+00 1.464e+01 1.204e+00 -9.177e-01 7.550e-15 -7.409e+00   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 7.989 4.049 1.973 0.11979   
## VarSPD 70.286 11.159 6.299 0.00325 \*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 9.055 on 4 degrees of freedom  
## (14 observations deleted due to missingness)  
## Multiple R-squared: 0.9084, Adjusted R-squared: 0.8855   
## F-statistic: 39.67 on 1 and 4 DF, p-value: 0.003247

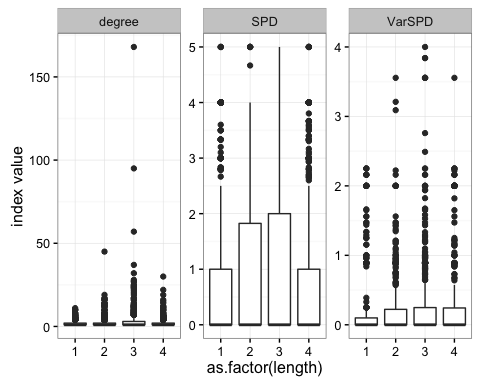
## Ectoparasites & Host Body Size

## Warning: Removed 21278 rows containing non-finite values (stat\_smooth).

## Warning: Removed 21278 rows containing missing values (geom\_point).



## Warning: Removed 2653 rows containing non-finite values (stat\_boxplot).



Negative binomial GLM seems like the best way to compare Ecto degree to quartiles of length categories. The estimates are log values, so exp() to get the estimate. Except, it's strictly positive so should use a zero truncated model?

summary(lm(data=hp.ectoDirect,cvmaxL~SPD))

##   
## Call:  
## lm(formula = cvmaxL ~ SPD, data = hp.ectoDirect)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.8175 -0.2446 -0.0123 0.2220 1.0840   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.236676 0.018157 13.04 <2e-16 \*\*\*  
## SPD 0.116167 0.007408 15.68 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.3327 on 1574 degrees of freedom  
## (2647 observations deleted due to missingness)  
## Multiple R-squared: 0.1351, Adjusted R-squared: 0.1346   
## F-statistic: 245.9 on 1 and 1574 DF, p-value: < 2.2e-16

summary(lm(data=hp.ectoDirect,cvmaxL~VarSPD))

##   
## Call:  
## lm(formula = cvmaxL ~ VarSPD, data = hp.ectoDirect)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.61554 -0.27229 -0.04986 0.21649 1.22153   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.468859 0.009929 47.220 < 2e-16 \*\*\*  
## VarSPD 0.065194 0.015474 4.213 2.66e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.3543 on 1554 degrees of freedom  
## (2667 observations deleted due to missingness)  
## Multiple R-squared: 0.01129, Adjusted R-squared: 0.01066   
## F-statistic: 17.75 on 1 and 1554 DF, p-value: 2.662e-05

summary(lm(data=hp.ectoDirect,maxMaxL~SPD))

##   
## Call:  
## lm(formula = maxMaxL ~ SPD, data = hp.ectoDirect)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -118.84 -61.57 -31.36 62.36 699.55   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 86.359 1.529 56.466 <2e-16 \*\*\*  
## SPD 9.621 1.016 9.466 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 83.25 on 4210 degrees of freedom  
## (11 observations deleted due to missingness)  
## Multiple R-squared: 0.02084, Adjusted R-squared: 0.02061   
## F-statistic: 89.61 on 1 and 4210 DF, p-value: < 2.2e-16

summary(lm(data=hp.ectoDirect,maxMaxL~VarSPD))

##   
## Call:  
## lm(formula = maxMaxL ~ VarSPD, data = hp.ectoDirect)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -142.38 -66.59 -22.83 61.87 690.91   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 102.828 2.511 40.952 < 2e-16 \*\*\*  
## VarSPD 25.177 3.921 6.422 1.78e-10 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 89.89 on 1565 degrees of freedom  
## (2656 observations deleted due to missingness)  
## Multiple R-squared: 0.02567, Adjusted R-squared: 0.02505   
## F-statistic: 41.24 on 1 and 1565 DF, p-value: 1.781e-10

summary(lm(data=hp.ectoDirect,meanMaxL~SPD))

##   
## Call:  
## lm(formula = meanMaxL ~ SPD, data = hp.ectoDirect)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -83.40 -54.40 -26.54 45.60 525.60   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 84.4009 1.3372 63.117 < 2e-16 \*\*\*  
## SPD -3.5445 0.8887 -3.989 6.76e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 72.78 on 4210 degrees of freedom  
## (11 observations deleted due to missingness)  
## Multiple R-squared: 0.003765, Adjusted R-squared: 0.003528   
## F-statistic: 15.91 on 1 and 4210 DF, p-value: 6.759e-05

summary(lm(data=hp.ectoDirect,meanMaxL~VarSPD))

##   
## Call:  
## lm(formula = meanMaxL ~ VarSPD, data = hp.ectoDirect)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -82.65 -46.50 -19.99 27.25 308.00   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 74.002 1.756 42.135 < 2e-16 \*\*\*  
## VarSPD 7.672 2.742 2.798 0.00521 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 62.87 on 1565 degrees of freedom  
## (2656 observations deleted due to missingness)  
## Multiple R-squared: 0.004977, Adjusted R-squared: 0.004341   
## F-statistic: 7.828 on 1 and 1565 DF, p-value: 0.005208

summary(lm(data=hp.ectoDirect,sdMaxL~SPD))

##   
## Call:  
## lm(formula = sdMaxL ~ SPD, data = hp.ectoDirect)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -50.165 -24.441 -12.788 8.708 273.403   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 19.7782 2.1296 9.287 < 2e-16 \*\*\*  
## SPD 6.0773 0.8689 6.994 3.93e-12 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 39.03 on 1574 degrees of freedom  
## (2647 observations deleted due to missingness)  
## Multiple R-squared: 0.03014, Adjusted R-squared: 0.02953   
## F-statistic: 48.92 on 1 and 1574 DF, p-value: 3.929e-12

summary(lm(data=hp.ectoDirect,sdMaxL~VarSPD))

##   
## Call:  
## lm(formula = sdMaxL ~ VarSPD, data = hp.ectoDirect)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -47.06 -25.52 -12.79 10.84 287.03   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 30.465 1.096 27.787 < 2e-16 \*\*\*  
## VarSPD 7.376 1.709 4.317 1.68e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 39.12 on 1554 degrees of freedom  
## (2667 observations deleted due to missingness)  
## Multiple R-squared: 0.01185, Adjusted R-squared: 0.01122   
## F-statistic: 18.64 on 1 and 1554 DF, p-value: 1.681e-05

# these LMs don't have straight Q-Q plots. Need to think about the implications.  
  
library(MASS)

##   
## Attaching package: 'MASS'

## The following object is masked from 'package:dplyr':  
##   
## select

summary(glm.nb(data=hp.ectoDirect,degree ~ cvmaxL))

##   
## Call:  
## glm.nb(formula = degree ~ cvmaxL, data = hp.ectoDirect, init.theta = 2.764163414,   
## link = log)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.5676 -0.6411 -0.4357 -0.0065 11.7204   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 1.08921 0.03469 31.397 <2e-16 \*\*\*  
## cvmaxL 0.46109 0.05513 8.364 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for Negative Binomial(2.7642) family taken to be 1)  
##   
## Null deviance: 1403.5 on 1575 degrees of freedom  
## Residual deviance: 1341.9 on 1574 degrees of freedom  
## (2647 observations deleted due to missingness)  
## AIC: 7269.7  
##   
## Number of Fisher Scoring iterations: 1  
##   
##   
## Theta: 2.764   
## Std. Err.: 0.143   
##   
## 2 x log-likelihood: -7263.732

summary(glm.nb(data=hp.ectoDirect,degree ~ maxMaxL))

##   
## Call:  
## glm.nb(formula = degree ~ maxMaxL, data = hp.ectoDirect, init.theta = 2.928829563,   
## link = log)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.2294 -0.5662 -0.3934 0.1606 15.0265   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 0.3657332 0.0216160 16.92 <2e-16 \*\*\*  
## maxMaxL 0.0032613 0.0001509 21.61 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for Negative Binomial(2.9288) family taken to be 1)  
##   
## Null deviance: 3558.3 on 4211 degrees of freedom  
## Residual deviance: 3105.6 on 4210 degrees of freedom  
## (11 observations deleted due to missingness)  
## AIC: 15085  
##   
## Number of Fisher Scoring iterations: 1  
##   
##   
## Theta: 2.929   
## Std. Err.: 0.118   
##   
## 2 x log-likelihood: -15078.712

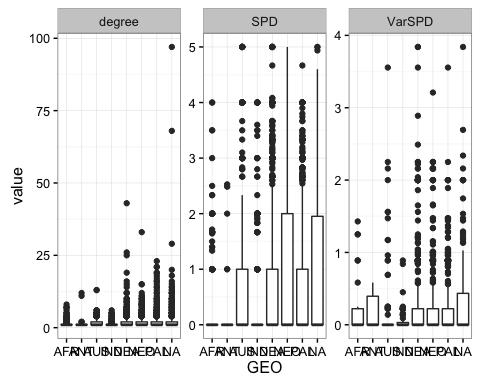
summary(glm.nb(data=hp.ectoDirect,degree ~ meanMaxL))

##   
## Call:  
## glm.nb(formula = degree ~ meanMaxL, data = hp.ectoDirect, init.theta = 2.52344473,   
## link = log)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -0.6446 -0.6381 -0.6148 -0.0257 15.6915   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 0.7292920 0.0217387 33.548 <2e-16 \*\*\*  
## meanMaxL -0.0001822 0.0002000 -0.911 0.362   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for Negative Binomial(2.5234) family taken to be 1)  
##   
## Null deviance: 3296.1 on 4211 degrees of freedom  
## Residual deviance: 3295.3 on 4210 degrees of freedom  
## (11 observations deleted due to missingness)  
## AIC: 15522  
##   
## Number of Fisher Scoring iterations: 1  
##   
##   
## Theta: 2.5234   
## Std. Err.: 0.0967   
##   
## 2 x log-likelihood: -15516.0440

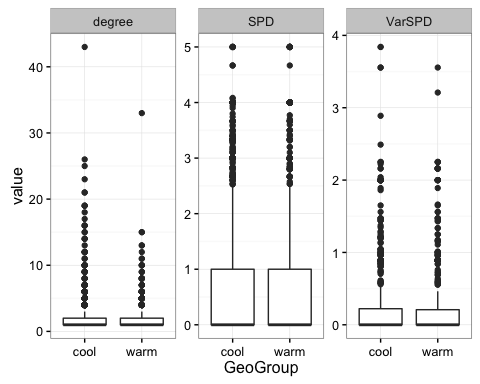
summary(glm.nb(data=hp.ectoDirect,degree ~ as.factor(mean.lcat)))

##   
## Call:  
## glm.nb(formula = degree ~ as.factor(mean.lcat), data = hp.ectoDirect,   
## init.theta = 2.763942524, link = log)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -0.9668 -0.5946 -0.4342 0.0486 14.2012   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 0.48940 0.03024 16.184 < 2e-16 \*\*\*  
## as.factor(mean.lcat)2 0.17390 0.04179 4.161 3.17e-05 \*\*\*  
## as.factor(mean.lcat)3 0.56872 0.03987 14.265 < 2e-16 \*\*\*  
## as.factor(mean.lcat)4 0.05284 0.04255 1.242 0.214   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for Negative Binomial(2.7639) family taken to be 1)  
##   
## Null deviance: 3459.3 on 4222 degrees of freedom  
## Residual deviance: 3191.8 on 4219 degrees of freedom  
## AIC: 15297  
##   
## Number of Fisher Scoring iterations: 1  
##   
##   
## Theta: 2.764   
## Std. Err.: 0.109   
##   
## 2 x log-likelihood: -15287.051

## Warning: Removed 3639 rows containing non-finite values (stat\_boxplot).



## Warning: Removed 5018 rows containing non-finite values (stat\_boxplot).



summary(lm(data=GeoDirectEcto,SPD ~ GEO))

##   
## Call:  
## lm(formula = SPD ~ GEO, data = GeoDirectEcto)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.0253 -0.7426 -0.6403 0.3597 4.2574   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.35421 0.06917 5.121 3.16e-07 \*\*\*  
## GEOANT -0.15829 0.18385 -0.861 0.389294   
## GEOAUS 0.45001 0.09124 4.932 8.41e-07 \*\*\*  
## GEOIND -0.06451 0.08562 -0.753 0.451222   
## GEONEA 0.38842 0.07677 5.059 4.36e-07 \*\*\*  
## GEONEO 0.67107 0.07953 8.438 < 2e-16 \*\*\*  
## GEOPAL 0.28606 0.07518 3.805 0.000143 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.155 on 4827 degrees of freedom  
## (581 observations deleted due to missingness)  
## Multiple R-squared: 0.03534, Adjusted R-squared: 0.03414   
## F-statistic: 29.47 on 6 and 4827 DF, p-value: < 2.2e-16

summary(lm(data=GeoDirectEcto,VarSPD ~ GEO))

##   
## Call:  
## lm(formula = VarSPD ~ GEO, data = GeoDirectEcto)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.2775 -0.2603 -0.1959 -0.0381 3.5625   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.14653 0.06485 2.259 0.0240 \*  
## GEOANT 0.04904 0.23915 0.205 0.8376   
## GEOAUS 0.02000 0.07835 0.255 0.7985   
## GEOIND -0.07752 0.08797 -0.881 0.3783   
## GEONEA 0.13095 0.06969 1.879 0.0604 .  
## GEONEO 0.11381 0.07006 1.625 0.1045   
## GEOPAL 0.04933 0.06885 0.716 0.4739   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.5147 on 1552 degrees of freedom  
## (3856 observations deleted due to missingness)  
## Multiple R-squared: 0.01126, Adjusted R-squared: 0.007442   
## F-statistic: 2.947 on 6 and 1552 DF, p-value: 0.007303

summary(glm.nb(data=GeoDirectEcto,degree ~ GEO))

##   
## Call:  
## glm.nb(formula = degree ~ GEO, data = GeoDirectEcto, init.theta = 6.922037142,   
## link = log)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -0.6612 -0.5933 -0.4841 0.1268 9.9275   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 0.36772 0.05476 6.715 1.89e-11 \*\*\*  
## GEOANT 0.05213 0.14292 0.365 0.715299   
## GEOAUS 0.12002 0.07079 1.696 0.089974 .   
## GEOIND -0.16152 0.06948 -2.325 0.020093 \*   
## GEONEA 0.22264 0.05983 3.721 0.000198 \*\*\*  
## GEONEO 0.28537 0.06138 4.650 3.33e-06 \*\*\*  
## GEOPAL 0.20177 0.05883 3.430 0.000604 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for Negative Binomial(6.922) family taken to be 1)  
##   
## Null deviance: 3392.6 on 4833 degrees of freedom  
## Residual deviance: 3291.8 on 4827 degrees of freedom  
## (581 observations deleted due to missingness)  
## AIC: 15430  
##   
## Number of Fisher Scoring iterations: 1  
##   
##   
## Theta: 6.922   
## Std. Err.: 0.487   
##   
## 2 x log-likelihood: -15414.026

summary(glm.nb(data=GeoDirectEcto,degree ~ GeoGroup))

##   
## Call:  
## glm.nb(formula = degree ~ GeoGroup, data = GeoDirectEcto, init.theta = 6.649502335,   
## link = log)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -0.5761 -0.5761 -0.4802 0.1434 9.8901   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 0.57628 0.01598 36.052 < 2e-16 \*\*\*  
## GeoGroupwarm -0.09070 0.02510 -3.614 0.000301 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for Negative Binomial(6.6495) family taken to be 1)  
##   
## Null deviance: 3355.2 on 4833 degrees of freedom  
## Residual deviance: 3342.1 on 4832 degrees of freedom  
## (581 observations deleted due to missingness)  
## AIC: 15507  
##   
## Number of Fisher Scoring iterations: 1  
##   
##   
## Theta: 6.650   
## Std. Err.: 0.459   
##   
## 2 x log-likelihood: -15501.253

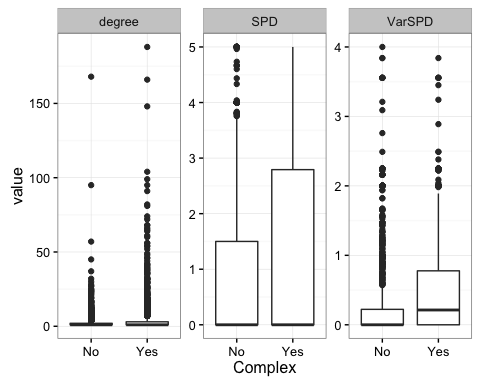
summary(lm(data=GeoDirectEcto,SPD ~ GeoGroup))

##   
## Call:  
## lm(formula = SPD ~ GeoGroup, data = GeoDirectEcto)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.7046 -0.7046 -0.6771 0.3229 4.3229   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.67710 0.02226 30.411 <2e-16 \*\*\*  
## GeoGroupwarm 0.02755 0.03422 0.805 0.421   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.176 on 4832 degrees of freedom  
## (581 observations deleted due to missingness)  
## Multiple R-squared: 0.0001341, Adjusted R-squared: -7.286e-05   
## F-statistic: 0.6479 on 1 and 4832 DF, p-value: 0.4209

summary(lm(data=GeoDirectEcto,VarSPD ~ GeoGroup))

##   
## Call:  
## lm(formula = VarSPD ~ GeoGroup, data = GeoDirectEcto)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.2325 -0.2325 -0.2076 -0.0103 3.6075   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.23248 0.01716 13.551 <2e-16 \*\*\*  
## GeoGroupwarm -0.02486 0.02653 -0.937 0.349   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.5167 on 1557 degrees of freedom  
## (3856 observations deleted due to missingness)  
## Multiple R-squared: 0.0005635, Adjusted R-squared: -7.835e-05   
## F-statistic: 0.8779 on 1 and 1557 DF, p-value: 0.3489

## Warning: Removed 4610 rows containing non-finite values (stat\_boxplot).



summary(glm.nb(data=hp.comp,degree ~ Complex))

##   
## Call:  
## glm.nb(formula = degree ~ Complex, data = hp.comp, init.theta = 1.347301944,   
## link = log)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -0.9710 -0.9710 -0.5436 -0.0173 12.5390   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 0.71240 0.01704 41.8 <2e-16 \*\*\*  
## ComplexYes 0.57178 0.02433 23.5 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for Negative Binomial(1.3473) family taken to be 1)  
##   
## Null deviance: 7230.4 on 7621 degrees of freedom  
## Residual deviance: 6673.4 on 7620 degrees of freedom  
## AIC: 32424  
##   
## Number of Fisher Scoring iterations: 1  
##   
##   
## Theta: 1.3473   
## Std. Err.: 0.0277   
##   
## 2 x log-likelihood: -32418.3850

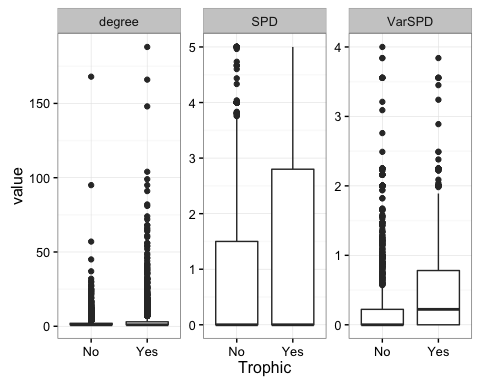
summary(lm(data=hp.comp,SPD ~ Complex))

##   
## Call:  
## lm(formula = SPD ~ Complex, data = hp.comp)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.1924 -1.1924 -0.8166 1.1167 4.1834   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.81663 0.02139 38.17 <2e-16 \*\*\*  
## ComplexYes 0.37577 0.03213 11.70 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.394 on 7620 degrees of freedom  
## Multiple R-squared: 0.01763, Adjusted R-squared: 0.0175   
## F-statistic: 136.8 on 1 and 7620 DF, p-value: < 2.2e-16

summary(lm(data=hp.comp,VarSPD ~ Complex))

##   
## Call:  
## lm(formula = VarSPD ~ Complex, data = hp.comp)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.4343 -0.2725 -0.2725 0.1467 3.7275   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.27253 0.01469 18.546 < 2e-16 \*\*\*  
## ComplexYes 0.16180 0.02128 7.603 3.85e-14 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.5834 on 3010 degrees of freedom  
## (4610 observations deleted due to missingness)  
## Multiple R-squared: 0.01884, Adjusted R-squared: 0.01852   
## F-statistic: 57.8 on 1 and 3010 DF, p-value: 3.851e-14

## Warning: Removed 4457 rows containing non-finite values (stat\_boxplot).



summary(glm.nb(data=hp.trop,degree ~ Trophic))

##   
## Call:  
## glm.nb(formula = degree ~ Trophic, data = hp.trop, init.theta = 1.328485495,   
## link = log)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -0.9842 -0.5550 -0.5550 -0.0331 12.3777   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 0.73020 0.01696 43.05 <2e-16 \*\*\*  
## TrophicYes 0.57808 0.02490 23.22 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for Negative Binomial(1.3285) family taken to be 1)  
##   
## Null deviance: 7014.3 on 7369 degrees of freedom  
## Residual deviance: 6468.7 on 7368 degrees of freedom  
## AIC: 31441  
##   
## Number of Fisher Scoring iterations: 1  
##   
##   
## Theta: 1.3285   
## Std. Err.: 0.0276   
##   
## 2 x log-likelihood: -31435.2590

summary(lm(data=hp.trop,SPD ~ Trophic))

##   
## Call:  
## lm(formula = SPD ~ Trophic, data = hp.trop)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.1875 -0.8269 -0.8269 1.1052 4.1731   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.82695 0.02121 38.98 <2e-16 \*\*\*  
## TrophicYes 0.36057 0.03283 10.98 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.39 on 7368 degrees of freedom  
## Multiple R-squared: 0.01611, Adjusted R-squared: 0.01598   
## F-statistic: 120.7 on 1 and 7368 DF, p-value: < 2.2e-16

summary(lm(data=hp.trop,VarSPD ~ Trophic))

##   
## Call:  
## lm(formula = VarSPD ~ Trophic, data = hp.trop)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.4392 -0.2716 -0.2716 0.1322 3.7284   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.27162 0.01455 18.668 < 2e-16 \*\*\*  
## TrophicYes 0.16759 0.02175 7.704 1.8e-14 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.5838 on 2911 degrees of freedom  
## (4457 observations deleted due to missingness)  
## Multiple R-squared: 0.01998, Adjusted R-squared: 0.01964   
## F-statistic: 59.35 on 1 and 2911 DF, p-value: 1.798e-14

with(hp.def,table(Trophic,Complex))

## Complex  
## Trophic No Yes Yes\_No  
## No 4233 54 0  
## Yes 10 3008 0

# Trophic and Complex are almost perfectly correlated in this subset of the dataset.

Discussion

* 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?

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 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:http://dx.doi.org/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 (80-. ).* **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. 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)

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

17. 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.