Multilevel models for the distribution of hosts and symbionts

Maxwell B. Joseph[1], William E. Stutz[1], and Pieter T. J. Johnson[1]

[1]*Ecology and Evolutionary Biology, University of Colorado, Boulder*

E-mail addresses:

maxwell.b.joseph@colorado.edu*

william.stutz@colorado.edu

pieter.johnson@colorado.edu

* corresponding author

Mailing address: CB 334, University of Colorado, Boulder, CO, USA, 80309

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1502v1 | CC-BY 4.0 Open Access | rec: 12 Nov 2015, publ: 12 Nov 2015

Abstract

Symbiont occurrence is influenced by host occurrence and vice versa, which leads to correlations in host-symbiont distributions at multiple levels. Interactions between co-infecting symbionts within host individuals can cause correlations in the abundance of two symbiont species across individual hosts. Similarly, interactions between symbiont transmission and host population dynamics can drive correlations between symbiont and host abundance across habitat patches. If ignored, these interactions can confound estimated responses of hosts and symbionts to other factors. Here, we present a general hierarchical modeling framework for distributions of hosts and symbionts, estimating correlations in host-symbiont distributions at the among-site, within-site, among-species, and among-individual levels. We present an empirical example from a multi-host multi-parasite system involving amphibians and their micro- and macroparasites. Amphibian hosts and their parasites were correlated at multiple levels of organization. Macroparasites often co-infected individual hosts, but rarely co-infected with the amphibian chytrid fungus. Such correlations may result from interactions among parasites and hosts, joint responses to environmental factors, or sampling bias. Joint host-symbiont models account for environmental constraints and species interactions while partitioning variance and dependence in abundance at multiple levels. This framework can be adapted to a wide variety of study systems and sampling designs.

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1502v1 | CC-BY 4.0 Open Access | rec: 12 Nov 2015, publ: 12 Nov 2015

Introduction

Symbiotic organisms - those that live with, in, or on free living hosts - play important roles in disease dynamics, food production, and host health (Bashan 1998; Jones *et al.* 2008). However, host-symbiont interactions complicate efforts to explain symbiont occurrence and abundance for several reasons. First, symbiont distributions depend on host distributions. In the extreme, obligate symbionts cannot exist without hosts (Moran & Baumann 2000). Symbionts also influence host distributions through effects on fitness and population dynamics (Ebert *et al.* 2000; Lloyd-Smith *et al.* 2005). Further complexity arises in systems with multi-host symbionts, and host individuals infected with multiple co-infecting symbionts. Symbionts occupying the same host individual can interact, such that one symbiont may directly affect the distribution of another symbiont at the individual level (Telfer *et al.* 2010). Useful models of symbiont occurrence and abundance should accommodate these bidirectional influences and the hierarchical nature of host-symbiont interactions (Mihaljevic 2012).

Multilevel modeling provides a promising avenue to understand patterns in host and symbiont abundance at different levels of biological organization (Gelman & Hill 2007). A general host-symbiont modeling framework must be multivariate: any interaction between a host and a symbiont involves at least two species. Further, useful methods should make use of observable host and symbiont data which often consist of discrete counts, but may also include binary measurements of habitat use or continuous measures of density. Continuous and discrete multivariate observations can be modeled by combining univariate distributions with multivariate linear predictors, leading to a multivariate probit for binary data, multivariate Poisson for counts, and multivariate lognormal for continuous positive observations (Ashford & Sowden 1970; Aitchison 1982; Aitchison & Ho 1989). Such models are increasingly being used to model distributions of free-living species while accounting for species interactions (Wisz *et al.* 2013; Clark *et al.* 2014; Pollock *et al.* 2014).

Here we expand upon existing methods to develop a hierarchical, multivariate framework for modeling host and symbiont distributions that accounts for multiple levels of correlation,

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1502v1 | CC-BY 4.0 Open Access | rec: 12 Nov 2015, publ: 12 Nov 2015

level-specific covariates, and flexible likelihood specifications. We begin by outlining the general features and logic of this approach. We then present an empirical case study of amphibian hosts and their parasites, revealing correlation among species at multiple levels and demonstrating the types of insights gained in practice. We conclude by discussing limitations and potential extensions.

Methods

While ecologists often seek to estimate the effects of one species on another species, this requires strong causal assumptions when working with observational data (Pearl 2000). Instead, correlations in species abundance and occurrence - potentially resulting from species interactions - can be modeled as a proxy, helping to generate hypotheses about interactions that ideally can be pursued experimentally (Ovaskainen *et al.* 2010). Due to the hierarchical nature of host-symbiont interactions, these correlations can occur at multiple levels (Mideo *et al.* 2008). Symbionts may be correlated at the level of host individuals, positively if two symbiont species often co-infect hosts (e.g., Puoti *et al.* 2002). Symbionts may also be correlated at the level of host species, positively if two symbionts tend to infect the same species (e.g., Johnson *et al.* 2012). Hosts and symbionts might also be correlated within and among spatial locations (hereafter "sites"). While such correlations can arise through species interactions, they can also emerge from simultaneous responses to extrinsic factors or sampling bias. These alternative drivers of correlations are not guaranteed to be differentiable from observational data alone (Pearl 2000; Dorazio & Connor 2014), emphasizing the importance of methods that limit causal assumptions.

We consider a landscape with discrete habitat patches (sites) containing multiple species of hosts and symbionts. At each site, replicate surveys are conducted to measure host density, and symbiont abundance is observed by sampling individual hosts. We assume each host species $h = 1, ..., H$ is present or absent at each site $i = 1, ..., N$, with occurrence constant

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1502v1 | CC-BY 4.0 Open Access | rec: 12 Nov 2015, publ: 12 Nov 2015

across surveys. If they are present, they have some non-zero site-level average density μ_{ih}. The probability of occurrence ψ_{ih} and expected density within a site are assumed to be proportional (He & Gaston 2003). Hosts may be present at a site but unobserved (MacKenzie *et al.* 2002). Conditional on occurrence, the probability of detection increases with density (Royle *et al.* 2005). In other words, sites favoring high density are commonly occupied, and abundant hosts are easier to detect than rare hosts. At site i, $J_i > 1$ repeat surveys are conducted, leading to the following likelihood or sampling distribution for host abundance observations:

$$y_{ih} \sim \begin{cases} \psi_{ih} \prod_{j=1}^{J_i} f(y_{ihj}|\theta_{ihj}), & \sum_{j=1}^{J_i} y_{ihj} > 0 \\ \psi_{ih} \prod_{j=1}^{J_i} f(0|\theta_{ihj}) + 1 - \psi_{ih}, & \text{otherwise} \end{cases} \quad \text{(eqn 1)}$$

Where y_{ih} is a vector of length J_i with elements consisting of abundance measurements (e.g., counts) of species h at site i in each survey. This is a mixture model with components representing cases in which species h is present or absent from site i with probabilities ψ_{ih} and $1 - \psi_{ih}$, respectively. Further, $f(y|\theta_{ihj})$ is a probability density or mass function with parameter(s) θ_{ihj} potentially varying across sites, species, and surveys (Ver Hoef & Boveng 2007). If species h is not observed at site i, then it was absent with probability $1 - \psi_{ih}$ or present but unobserved with probability $\psi_{ih} \prod_{J_i} f(0|\theta_{ihj})$. False absences are more likely for species with low densities and those highly aggregated within sites. For simplicity we assume that detection implies species presence, but a likelihood could be specified to account for false positives (Royle & Link 2006).

We assume that the occupancy probability of species h at site i increases with expected density μ_{ih} as follows (He *et al.* 2002; Smith *et al.* 2012):

$$logit(\psi_{ih}) = \gamma_{0h} + \gamma_{1h} log(\mu_{ih}) \quad \text{(eqn 2)}$$

Here, γ_{0h} is the probability of host species h occurring at site i on a logit scale when

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1502v1 | CC-BY 4.0 Open Access | rec: 12 Nov 2015, publ: 12 Nov 2015

the mean density is one individual per unit area of habitat (e.g. per square meter), and γ_{1h} is a parameter that describes the scaling between expected density and the probability of occupancy, which we expect to be positive. This occupancy submodel could also include covariates such as habitat area.

Symbiont species $s = 1, ..., S$ are present or absent at each site. At site i, K_i host individuals are sampled and their infections quantified. Non-detection of symbiont s at site i can result from true absence or failure to sample an infected host, and sites favoring high symbiont abundance are more likely to be occupied:

$$y_{is} \sim \begin{cases} \psi_{is} \prod_{k=1}^{K_i} f(y_{isk}|\theta_{isk}), & \sum_{k=1}^{K_i} y_{isk} > 0 \\ \psi_{is} \prod_{k=1}^{K_i} f(0|\theta_{isk}) + 1 - \psi_{is}, & \text{otherwise} \end{cases} \quad \text{(eqn 3)}$$

$$logit(\psi_{is}) = \gamma_{0s} + \gamma_{1s} log(\mu_{is}) \quad \text{(eqn 4)}$$

Every host and symbiont species has a site-level mean density, and these densities may be correlated e.g., if an abundant reservoir host increases infection in other hosts (Ashford 2003). Species have some among-site variance in their abundances, and these variance parameters may differ across species. Species that are always at low or high abundance will have low variance, and species that are abundant in some sites, and absent from others will have higher variance. These correlation and variance parameters are used to construct a covariance matrix $\mathbf{\Sigma}_{site}$ with elements $\rho_{mn}\sigma_m\sigma_n$ in the m^{th} row, n^{th} column, where ρ_{mn} is the correlation between species m and n, and σ_m is the among site standard deviation for species m. Each site has a random effect vector $\boldsymbol{\alpha}_i$ of length $H + S$: $\boldsymbol{\alpha}_i \sim N_{H+S}(\mathbf{0}, \mathbf{\Sigma}_{site})$, where $N_d(\mathbf{0}, \mathbf{\Sigma})$ represents a multivariate normal distribution with dimension d, mean vector $\mathbf{0}$, and covariance matrix $\mathbf{\Sigma}$.

Within sites, hosts and symbiont density can vary among survey locations. Uniformly distributed species have low variance, and spatially aggregated species have high variance. Species are correlated within sites if they tend to be observed together in the same surveys more or less often then expected by chance, for example. We can represent these survey

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1502v1 | CC-BY 4.0 Open Access | rec: 12 Nov 2015, publ: 12 Nov 2015

level correlations and variance parameters in a covariance matrix $\mathbf{\Sigma}_{survey}$, which gives rise to to $J_{tot} = \sum_i j_i$ survey level random effect vectors $\boldsymbol{\alpha}_j$, each with length $H + S$: $\boldsymbol{\alpha}_j \sim N_{H+S}(\mathbf{0}, \mathbf{\Sigma}_{survey})$. Random effects may be adapted to alternative sampling designs. For instance, if hosts are sampled for symbionts independently from host density surveys, then symbionts are not associated with particular surveys and the survey-level random effects may instead have dimension H.

Differences in overall mean abundance are represented with a host species specific random effect α_{0h} which is univariate normally distributed around a community mean, with among species variance. Together, these random effects contribute to the expected number of individual hosts of species h detected in a survey j at site i, here with a log-link:

$$log(\mu_{ihj}) = \alpha_{0h} + \alpha_{jh} + \alpha_{ih} \qquad \text{(eqn 5)}$$

Depending on survey design, this expectation might include an offset that accounts for among-survey variation in sampling time intervals or area (Gelman & Hill 2007).

The expected density of symbionts also includes an intercept α_{0s} and elements from the site-level and survey-level random effects. However, because of the nature of host-symbiont interactions, symbionts have the potential for correlation at additional levels. Specifically, symbionts may be correlated at the individual host level, e.g., if two symbionts commonly co-infect host individuals. We represent these host individual differences with $K_{tot} = \sum_i K_i$ multivariate normal random effects with mean zero and covariance matrix $\mathbf{\Sigma}_{indiv}$ including correlation terms and symbiont species specific variance terms representing how variable host individuals are in their infection abundances: $\boldsymbol{\alpha}_k \sim N_S(\mathbf{0}, \mathbf{\Sigma}_{indiv})$.

Finally, hosts may vary in their symbiont infection abundances at the species level. This variation may be correlated if two host species are functionally alike, e.g., they tend to be similarly susceptible to infection across a range of symbiont species. To allow for species level variation we consider $h = 1, ..., H$ multivariate normal random vectors, each with S elements: $\boldsymbol{\alpha}_h \sim N_S(\mathbf{0}, \mathbf{\Sigma}_{species})$.

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1502v1 | CC-BY 4.0 Open Access | rec: 12 Nov 2015, publ: 12 Nov 2015

Together, these random effects contribute to the expected infection load of symbiont s at site i in host individual k of species h sampled in survey j:

$$log(\mu_{isk}) = \alpha_{0s} + \alpha_{is} + \alpha_{j[k]s} + \alpha_{h_k s} + \alpha_{ks} \quad \text{(eqn 6)}$$

If host sampling for symbionts occurs separately from host abundance surveys, then sampled hosts are not associated with surveys, simplifying the random effects:

$$log(\mu_{isk}) = \alpha_{0s} + \alpha_{is} + \alpha_{h_k s} + \alpha_{ks} \quad \text{(eqn 7)}$$

Case study: amphibian communities and their parasites

Amphibians in the San Francisco Bay Area of California are infected with a diverse suite of parasites, including macroparasitic helminth worms (*Ribeiroia ondatrae* Looss, 1907, *Echinostoma* sp., *Cephalogonimus* sp., *Alaria* sp.), and microparasites such as *Ranavirus* sp. and the amphibian chytrid fungus *Batrachochytrium dendrobatidis*, Longcore, Pessier & D.K. Nichols (1999), hereafter referred to as *Bd*.

Five amphibian hosts comprise the majority of non-threatened (available for sampling) amphibian species: the Pacific chorus frog *Pseudacris regilla* (Baird & Girard, 1852), California newt *Taricha torosa* (Rathke, in Eschscholtz, 1833), rough-skinned newt *Taricha granulosa* (Skilton, 1849), western toad *Anaxyrus boreas* Baird & Girard, 1852, and the non-native American bullfrog *Lithobates catesbeianus* (Shaw, 1802) (Johnson *et al.* 2013). Previous studies in this system have revealed correlations between parasites at the host individual and site levels (Johnson & Buller 2011; Hoverman *et al.* 2013).

In 2013, field crews visited 87 wetland sites in Contra Costa, Alameda, and Santa Clara counties. At each site, crews conducted dip net sweep surveys ($\bar{J}_i = 9.9$, standard deviation $(s_{J_i}) = 1.2$, range $= [2, 15]$, $J_{tot} = 914$) to quantify amphibian density, recording the numbers and species identities of all amphibians observed. Crews collected hosts at each site to quantify parasite infections ($\bar{K}_i = 17.8$, $s_{K_i} = 12.7$, range $= [1, 82]$, $K_{tot} = 1550$), and

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1502v1 | CC-BY 4.0 Open Access | rec: 12 Nov 2015, publ: 12 Nov 2015

these collection events were separate from the sweep surveys. Collected hosts were larval or recently metamorphosed. We assessed macroparasite infection abundance via dissection (Johnson *et al.* 2013), and infection loads of Bd and *Ranavirus* using quantitative polymerase chain reaction of skin swabs and organ tissue, respectively (Hyatt *et al.* 2007; Hoverman *et al.* 2010). This work was approved by the University of Colorado Institutional Animal Care and Use Committee, protocol 1302.02.

Parameter estimation

We used a Bayesian approach to estimate parameters, combining prior information with a Poisson likelihood to generate a posterior distribution for unknown quantities. We simulated samples from the posterior using Markov chain Monte Carlo (MCMC) sampling in the probabilistic programming language Stan (Hoffman & Gelman 2014) . All data and code required to reproduce the analysis is available in the code supplement.

Results

We uncovered correlations between hosts and parasites at every level in the model. At the site level, we detected multiple correlations between hosts and parasites (Figure 1). Sites with high densities of Pacific chorus frogs had high densities of California newts and western toads, possibly due to similar habitat requirements (Joseph *et al.* 2015). Sites with high densities of chorus frogs had higher Bd infection loads, consistent with this species' role as a reservoir host (Reeder *et al.* 2012). Sites with high levels of infection of *Cephalogonimus* tended to have lower levels of infection with Bd. Macroparasites were positively correlated across sites, probably due to availability of planorbid snails that release macroparasite infective stages (cercariae), and deposition of parasite eggs in feces of carnivorous definitive hosts.

Within sites at the survey level, California newts correlated positively with rough-skinned newts and Pacific chorus frogs (Figure 2). These correlations imply that these species tend

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1502v1 | CC-BY 4.0 Open Access | rec: 12 Nov 2015, publ: 12 Nov 2015

to be co-aggregated within sites, potentially due to similar microhabitat preferences.

At the host species level, among-parasite correlations were estimated with low precision as we would expect when trying to estimate a correlation with five points (host species). However, some posteriors leaned toward positive correlations e.g., between Bd and *Alaria* (Figure 3). This was driven by high infection abundances of most parasites in Pacific chorus frogs, consistent with these fast-lived hosts investing little in parasite defense (Johnson *et al.* 2012). More host species are needed to make reliable inference at this level.

At the individual host level macroparasite loads correlated positively, so that if an individual was heavily infected with one macroparasite, it was more likely to be heavily infected with other macroparasites (Figure 4). These positive correlations can occur despite negative within-host interactions (Johnson & Hoverman 2012). For instance *Ribeiroia* and *Echinostoma* both have negative effects on the persistence of one another within host individuals, and the positive correlation may result from these parasites having similar niche requirements and host preferences (Johnson & Buller 2011). In contrast, Bd correlated negatively with three macroparasites: *Ribeiroia*, *Alaria*, and *Echinostoma*. Parasite interactions could drive these correlations or they could result from confounding variables. For example, host age increases cumulative exposure, confounding inference on parasite interactions derived from correlations. Such correlations may disappear after including the confounding trait as a covariate, contingent on the validity of the model with respect to the true latent processes (Pearl 2000). Last, correlations could arise from sampling bias (Berkson 1946). For instance, if Bd or *Echinostoma* infection increases catchability, then these two parasites will correlate negatively in our sample even if they are not correlated within the population.

We partitioned variation in host and parasite abundance among model levels to better understand the relative strength of processes operating at different scales. This analysis aims to summarize the correlations and extra-Poisson variance induced by the random effects. We considered effective variance $V_e(X) := |\mathbf{\Sigma}_X|^{1/d}$, the $d-$th root of the determinant of a covariance matrix $\mathbf{\Sigma}_X$ with dimension d, which represents the average scatter in any direction

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1502v1 | CC-BY 4.0 Open Access | rec: 12 Nov 2015, publ: 12 Nov 2015

(Peña & Rodríguez 2003). We also considered effective dependence $D_e(X) := 1 - |\boldsymbol{R}_X|^{1/d}$, where $\boldsymbol{R}_X$ is a correlation matrix, which captures the stochastic dependence among species (Peña & Rodríguez 2003). If species tend to be highly correlated, this parameter will be close to one. With no correlation among species, effective dependence is zero. Within-site, among-survey variation accounted for less variation in host abundance than among site random effects (Figure 5). For parasites, variation among host individuals exceeded among-site variation. This is striking, but consistent with the notion that parasites are overdispersed and aggregated among host individuals (Anderson & May 1978). Despite high variance, parasite abundance showed relatively low dependence at the individual level. Effective dependence was comparable across other model levels, which might be expected if species interactions and/or joint responses to covariates similarly influence patterns of co-aggregation at these levels.

Discussion

We presented a general hierarchical modeling framework to understand correlations and drivers of host and symbiont abundance. This builds upon existing multi-species abundance models and specifically extends a two symbiont abundance model by Stutz et al. *in review*, allowing for more than two species of symbionts, inclusion of hosts (any number of species), partially observed occurrence states, and greater flexibility in likelihood specification. Many host-symbiont distributions could be investigated with this method beyond host-parasite associations, including commensal and mutualistic symbionts of plants and animals.

The primary benefits of this approach include a lack of causal assumptions and the ability to decompose variation and dependence across multiple levels of organization. If there is a clear causal direction, e.g., in an experimental setting where host abundance is fixed, then one could extend this method to model the effect of host density on symbiont abundance, rather than their correlation alone. Effective variance and dependence may reflect the relative

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1502v1 | CC-BY 4.0 Open Access | rec: 12 Nov 2015, publ: 12 Nov 2015

importance of processes at different levels of organization. For instance, we found variation among host species in parasite abundance comparable to variation among spatial locations, both of which exceeded variation within sites. Generally, the contribution of model levels to effective variance will differ among study systems, and the ability to compare across levels should be valuable in determining how to begin model expansion. In our case study for instance, a logical next step would be inclusion of site and host individual level covariates.

Alternative likelihood functions, including those accounting for measurement error, can be readily combined with this method. Here we made use of a Poisson likelihood, but some situations may call for the use of zero-inflated probability distributions with support for all real postive values, such as a zero-inflated lognormal or gamma (Miller *et al.* 2012). This would allow for direct modeling of observations generated via quantitative polymerase chain reaction, typical of applications to viruses and bacteria, and environmental DNA of free-living species. Continuous distributions would circumvent the need to round values for use with Poisson or negative binomial distributions with integer support. Last, we have assumed that infections are detected without error, but a rich set of methods could be applied to account for error in this measurement process (Lachish *et al.* 2012; Miller *et al.* 2012).

We assumed that sites favoring high density are more likely to be occupied (Smith *et al.* 2012). However, if different processes drive species occurrence and abundance, then alternative occurrence submodels could be developed. In particular, spatial and temporal dependence may be useful for representing limits to species occurrence (Holt & Keitt 2000). Future developments of this approach might prioritize inclusion of spatiotemporally explicit colonization dynamics that account for occupancy status of neighboring sites, habitat quality, and dispersal functions (Broms *et al.* 2015). These approaches will prove useful to understand how much of the spread of an invasive symbiont may be due to changes in the host distribution vs. changes in the symbiont distribution alone, with potential applications to the management of emerging infectious diseases (Mitchell *et al.* 2006).

Symbionts have received an increased appreciation over past decades as the field of disease

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1502v1 | CC-BY 4.0 Open Access | rec: 12 Nov 2015, publ: 12 Nov 2015

ecology has gained momentum and as modern genetic methods have increased our ability to sample unculturable communities (Schrag & Wiener 1995; Riesenfeld *et al.* 2004). However, the development of methods to understand the distribution of symbionts has not kept pace with developments in free living species (Bailey *et al.* 2014). The approach presented here draws upon these developments with the goal of producing a general approach that can be readily adapted to other host-symbiont systems. Simultaneously modeling hosts and their symbionts in this hierarchical framework provides a powerful method to dissect patterns of occurrence and abundance for free living and symbiotic organisms.

Acknowledgments

We thank Melina Allahverdian, Dana Calhoun, Kelly DeRolf, Jackie Gregory, Emily Hannon, Jeremy Henderson, Megan Housman, Aaron Klingborg, Bryan LaFonte, Keegan McCaffrey, Travis McDevitt-Galles, Mary Toothman, and Vanessa Wuerthner for assistance with data collection. We also thank Tad Dallas, Helen McCreery, Joseph Mihaljevic, Lauren Shoemaker, Timothy Szewczyk, and Topher Weiss-Lehman for feedback on the manuscript. Finally, we gratefully acknowledge contributions of data and intellectual suggestions from Cheryl Briggs, Jason Hoverman, Jason Rohr, and Andrew Blaustein. This research was supported by funding from the National Institutes of Health, Ecology and Evolution of Infectious Diseases Program (R01GM109499), the National Science Foundation (DEB-1149308), and the David and Lucile Packard Foundation. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Data accessibility

R scripts and data files: uploaded as online supporting information

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1502v1 | CC-BY 4.0 Open Access | rec: 12 Nov 2015, publ: 12 Nov 2015

References

(2015) RStan: the R interface to Stan.

(2015) Stan: A C++ Library for Probability and Sampling, Version 2.8.0.

Aitchison, J. & Ho, C.H. (1989) The multivariate Poisson-log normal distribution. *Biometrika*, **76**, 643–653.

Aitchison, J. (1982) The statistical analysis of compositional data. *Journal of the Royal Statistical Society: Series B*, **44**, 139–177.

Anderson, R.M. & May, R.M. (1978) Regulation and stability of host-parasite population interactions: I. regulatory processes. *The Journal of Animal Ecology*, **47**, 219.

Ashford, J.R. & Sowden, R.R. (1970) Multi-variate probit analysis. *Biometrics*, **26**, 535–546.

Ashford, R.W. (2003) When is a reservoir not a reservoir? *Emerging infectious diseases*, **9**, 1495–1496.

Bailey, L.L., MacKenzie, D.I. & Nichols, J.D. (2014) Advances and applications of occupancy models. *Methods in Ecology and Evolution*, **5**, 1269–1279.

Bashan, Y. (1998) Inoculants of plant growth-promoting bacteria for use in agriculture. *Biotechnology Advances*, **16**, 729–770.

Berkson, J. (1946) Limitations of the application of fourfold table analysis to hospital data. *Biometrics Bulletin*, pp. 47–53.

Broms, K.M., Hooten, M.B., Johnson, D.S., Altwegg, R. & Conquest, L.L. (2015) Dynamic occupancy models for explicit colonization processes. *Ecology*, **in press**, 150716143727003.

Clark, J.S., Gelfand, A.E., Woodall, C.W. & Zhu, K. (2014) More than the sum of the parts: forest climate response from joint species distribution models. *Ecological Applications*, **24**, 990–999.

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1502v1 | CC-BY 4.0 Open Access | rec: 12 Nov 2015, publ: 12 Nov 2015

Dorazio, R.M. & Connor, E.F. (2014) Estimating abundances of interacting species using morphological traits, foraging guilds, and habitat. *PLoS ONE*, **9**, e94323.

Ebert, D., Lipsitch, M. & Mangin, K.L. (2000) The effect of parasites on host population density and extinction: experimental epidemiology with *Daphnia* and six microparasites. *The American Naturalist*, **156**, 459–477.

Gelman, A. & Hill, J. (2007) *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Analytical Methods for Social Research. Cambridge University Press.

He, F. & Gaston, K.J. (2003) Occupancy, spatial variance, and the abundance of species. *The American naturalist*, **162**, 366–375.

He, F., Gaston, K.J. & Wu, J. (2002) On species occupancy-abundance models. *Ecoscience*, **9**, 119–126.

Hoffman, M.D. & Gelman, A. (2014) The No-U-Turn Sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *The Journal of Machine Learning Research*, **15**, 1593–1623.

Holt, R.D. & Keitt, T.H. (2000) Alternative causes for range limits: a metapopulation perspective. *Ecology Letters*, **3**, 41–47.

Hoverman, J.T., Gray, M.J. & Miller, D.L. (2010) Anuran susceptibilities to ranaviruses: role of species identity, exposure route, and a novel virus isolate. *Diseases of Aquatic Organisms*, **89**, 97–107.

Hoverman, J.T., Hoye, B.J. & Johnson, P.T.J. (2013) Does timing matter? How priority effects influence the outcome of parasite interactions within hosts. *Oecologia*, **173**, 1471–1480.

Hyatt, A.D., Boyle, D.G., Olsen, V., Boyle, D.B., Berger, L., Obendorf, D., Dalton, A., Kriger, K., Hero, M., Hines, H., Phillott, R., Campbell, R., Marantelli, G., Gleason,

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1502v1 | CC-BY 4.0 Open Access | rec: 12 Nov 2015, publ: 12 Nov 2015

F. & Colling, A. (2007) Diagnostic assays and sampling protocols for the detection of *Batrachochytrium dendrobatidis*. *Diseases of Aquatic Organisms*, **73**, 175–192.

Johnson, P.T.J. & Buller, I.D. (2011) Parasite competition hidden by correlated coinfection: using surveys and experiments to understand parasite interactions. *Ecology*, **92**, 535–541.

Johnson, P.T.J. & Hoverman, J.T. (2012) Parasite diversity and coinfection determine pathogen infection success and host fitness. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 9006–11.

Johnson, P.T.J., Preston, D.L., Hoverman, J.T. & Richgels, K.L.D. (2013) Biodiversity decreases disease through predictable changes in host community competence. *Nature*, **494**, 230–3.

Johnson, P.T.J., Rohr, J.R., Hoverman, J.T., Kellermanns, E., Bowerman, J. & Lunde, K.B. (2012) Living fast and dying of infection: host life history drives interspecific variation in infection and disease risk. *Ecology letters*, **15**, 235–42.

Jones, K.E., Patel, N.G., Levy, M.a., Storeygard, A., Balk, D., Gittleman, J.L. & Daszak, P. (2008) Global trends in emerging infectious diseases. *Nature*, **451**, 990–3.

Joseph, M.B., Preston, D.L. & Johnson, P.T.J. (2015) Integrating occupancy models and structural equation models to understand species occurrence. *Ecology*.

Lachish, S., Gopalaswamy, A.M., Knowles, S.C.L. & Sheldon, B.C. (2012) Site-occupancy modelling as a novel framework for assessing test sensitivity and estimating wildlife disease prevalence from imperfect diagnostic tests. *Methods in Ecology and Evolution*, **3**, 339–348.

Lloyd-Smith, J.O., Cross, P.C., Briggs, C.J., Daugherty, M., Getz, W.M., Latto, J., Sanchez, M.S., Smith, A.B. & Swei, A. (2005) Should we expect population thresholds for wildlife disease? *Trends in ecology & evolution*, **20**, 511–9.

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1502v1 | CC-BY 4.0 Open Access | rec: 12 Nov 2015, publ: 12 Nov 2015

366 MacKenzie, D., Nichols, J. & Lachman, G. (2002) Estimating site occupancy rates when
367 detection probabilities are less than one. *Ecology*, **83**, 2248–2255.

368 Mideo, N., Alizon, S. & Day, T. (2008) Linking within- and between-host dynamics in the
369 evolutionary epidemiology of infectious diseases. *Trends in Ecology and Evolution*, **23**,
370 511–517.

371 Mihaljevic, J.R. (2012) Linking metacommunity theory and symbiont evolutionary ecology.
372 *Trends in Ecology and Evolution*, **27**, 323–329.

373 Miller, D.A.W., Talley, B.L., Lips, K.R. & Campbell Grant, E.H. (2012) Estimating patterns
374 and drivers of infection prevalence and intensity when detection is imperfect and sampling
375 error occurs. *Methods in Ecology and Evolution*, **3**, 850–859.

376 Mitchell, C.E., Agrawal, A.A., Bever, J.D., Gilbert, G.S., Hufbauer, R.A., Klironomos, J.N.,
377 Maron, J.L., Morris, W.F., Parker, I.M., Power, A.G., Seabloom, E.W., Torchin, M.E. &
378 Vazquez, D.P. (2006) Biotic interactions and plant invasions. *Ecology Letters*, **9**, 726–740.

379 Moran, N.A. & Baumann, P. (2000) Bacterial endosymbionts in animals. *Current Opinion*
380 *in Microbiology*, **3**, 270–275.

381 Ovaskainen, O., Hottola, J. & Shtonen, J. (2010) Modeling species co-occurrence by multi-
382 variate logistic regression generates new hypotheses on fungal interactions. *Ecology*, **91**,
383 2514–2521.

384 Pearl, J. (2000) Causality. *New York: Cambridge*.

385 Peña, D. & Rodríguez, J. (2003) Descriptive measures of multivariate scatter and linear
386 dependence. *Journal of Multivariate Analysis*, **85**, 361–374.

387 Pollock, L.J., Tingley, R., Morris, W.K., Golding, N., O'Hara, R.B., Parris, K.M., Vesk, P.a.
388 & Mccarthy, M.a. (2014) Understanding co-occurrence by modelling species simultaneously

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1502v1 | CC-BY 4.0 Open Access | rec: 12 Nov 2015, publ: 12 Nov 2015

with a Joint Species Distribution Model (JSDM). *Methods in Ecology and Evolution*, **5**, 397–406.

Puoti, M., Airoldi, M., Bruno, R., Zanini, B., Spinetti, A., Pezzoli, C., Patroni, A., Castelli, F., Sacchi, P., Filice, G. & Carosi, G. (2002) Hepatitis B virus co-infection in human immunodeficiency virus infected subjects. *AIDS Reviews*, **4**, 27–35.

Reeder, N.M.M., Pessier, A.P. & Vredenburg, V.T. (2012) A reservoir species for the emerging amphibian pathogen *Batrachochytrium dendrobatidis* thrives in a landscape decimated by disease. *PLoS ONE*, **7**, 1–7.

Riesenfeld, C.S., Schloss, P.D. & Handelsman, J. (2004) Metagenomics: genomic analysis of microbial communities. *Annual Review of Genetics*, **38**, 525–552.

Royle, J.A. & Link, W.A. (2006) Generalized site occupancy models allowing for false positive and false negative errors. *Ecology*, **87**, 835–841.

Royle, J.A., Nichols, J.D., Kéry, M. & Ranta, E. (2005) Modelling occurrence and abundance of species when detection is imperfect. *Oikos*, **110**, 353–359.

Schrag, S.J. & Wiener, P. (1995) Emerging infectious disease: what are the relative roles of ecology and evolution? *Trends in Ecology & Evolution*, **10**, 319–324.

Smith, A.N.H., Anderson, M.J. & Millar, R.B. (2012) Incorporating the intraspecific occupancy-abundance relationship into zero-inflated models. *Ecology*, **93**, 2526–2532.

Telfer, S., Lambin, X., Birtles, R., Beldomenico, P., Burthe, S., Paterson, S. & Begon, M. (2010) Species interactions in a parasite community drive infection risk in a wildlife population. *Science*, **330**, 243–246.

Ver Hoef, J.M. & Boveng, P.L. (2007) Quasi-poisson vs. negative binomial regression: How should we model overdispersed count data? *Ecology*, **88**, 2766–2772.

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1502v1 | CC-BY 4.0 Open Access | rec: 12 Nov 2015, publ: 12 Nov 2015

412 Wisz, M.S., Pottier, J., Kissling, W.D., Pellissier, L., Lenoir, J., Damgaard, C.F., Dormann,
413 C.F., Forchhammer, M.C., Grytnes, J.A., Guisan, A., Heikkinen, R.K., Høye, T.T., Kühn,
414 I., Luoto, M., Maiorano, L., Nilsson, M.C., Normand, S., Öckinger, E., Schmidt, N.M.,
415 Termansen, M., Timmermann, A., Wardle, D.A., Aastrup, P. & Svenning, J.C. (2013)
416 The role of biotic interactions in shaping distributions and realised assemblages of species:
417 Implications for species distribution modelling. *Biological Reviews*, **88**, 15–30.

418 # Figure 1

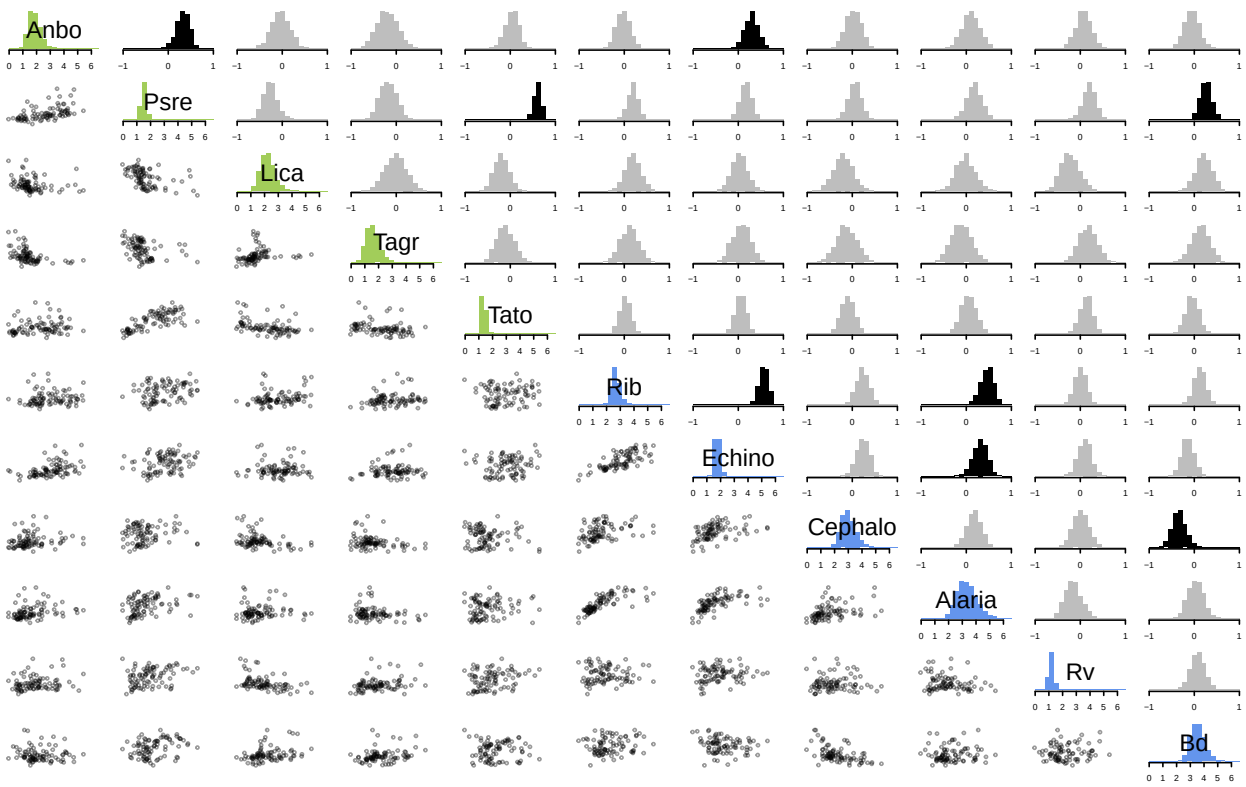

Figure 1: Site level variance covariance matrix and random effect posteriors. Diagonal elements display the among-site standard deviation in abundance for all host and parasite species (Anbo = *Anaxyrus boreas*, Psre = *Pseudacris regilla*, Lica = *Lithobates catesbeianus*, Tagr = *Taricha granulosa*, Tato = *Taricha torosa*, Rib = *Ribeiroia ondatrae*, Echino = *Echinostoma* sp., Cephalo = *Cephalogonimus* sp., Alaria = *Alaria* sp., Rv = *Ranavirus* sp., Bd = *Batrachochytrium dendrobatidis*). Green indicates hosts and blue parasites. Upper triangular elements show among-species correlation parameters. Black indicates correlations that are probably positive or probably negative (95% of posterior probability mass greater than or less than zero); grey indicates otherwise. Lower triangular elements show bivariate scatter plots of the posterior means of the site-level random effects corresponding to the intersection of the species in the rows and columns.

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1502v1 | CC-BY 4.0 Open Access | rec: 12 Nov 2015, publ: 12 Nov 2015

Figure 2

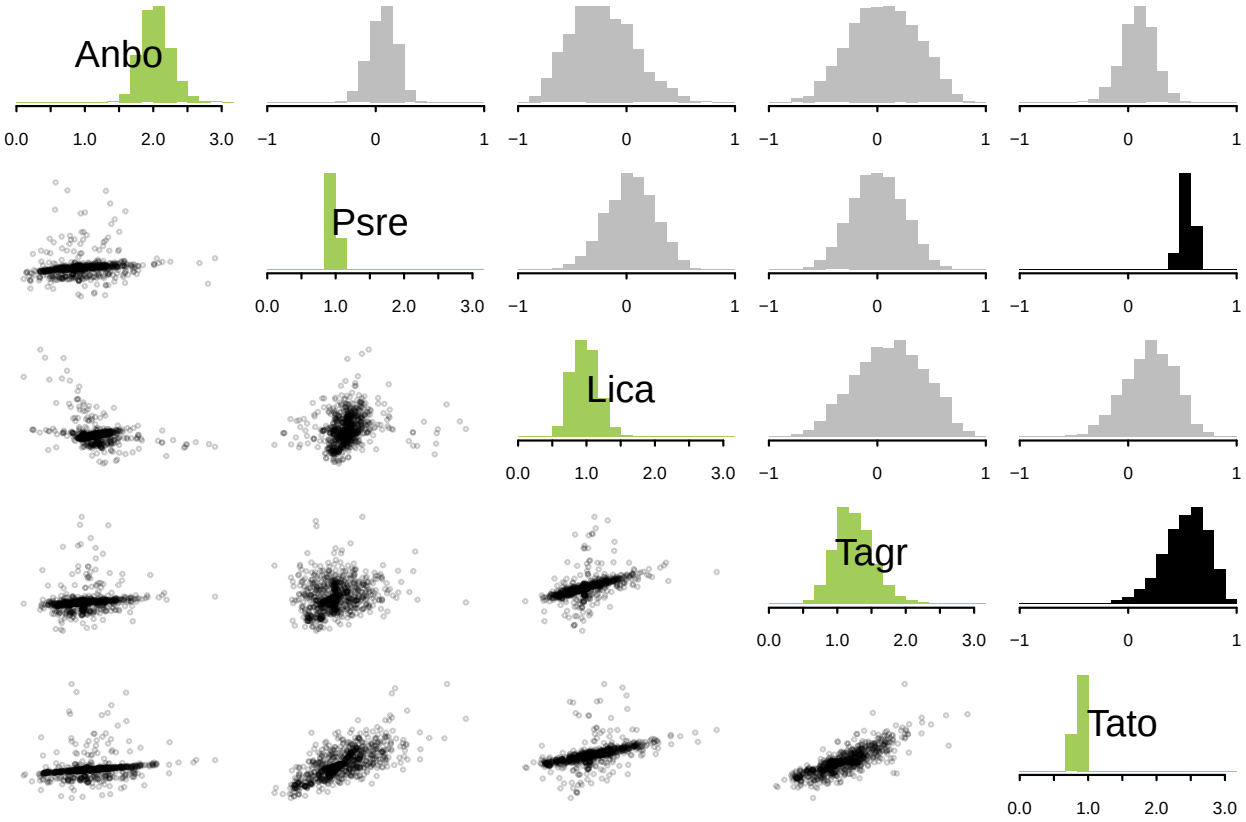

Figure 2: Survey level variance covariance matrix and random effect posteriors. Diagonal elements display the among-survey standard deviation in abundance for host species (Anbo = *Anaxyrus boreas*, Psre = *Pseudacris regilla*, Lica = *Lithobates catesbeianus*, Tagr = *Taricha granulosa*, Tato = *Taricha torosa*). Upper triangular elements show among-species correlation parameters. Black indicates correlations that are probably positive or probably negative (95% of posterior probability mass greater than or less than zero); grey indicates otherwise. Lower triangular elements show bivariate scatter plots of the posterior means of the survey-level random effects corresponding to the intersection of the species in the rows and columns.

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1502v1 | CC-BY 4.0 Open Access | rec: 12 Nov 2015, publ: 12 Nov 2015

Figure 3

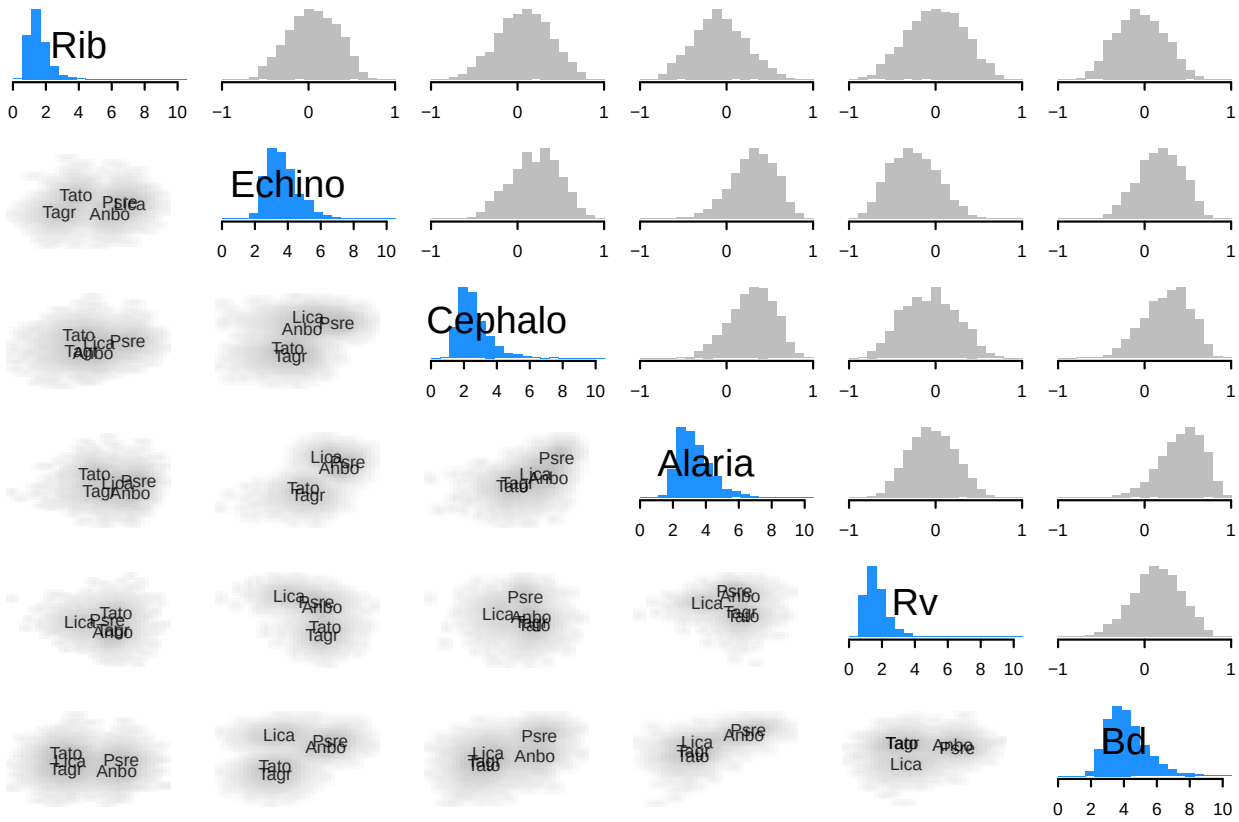

Figure 3: Host species level variance covariance matrix and random effect posteriors. Diagonal elements display the among host species standard deviation in abundance for parasite species (Rib = *Ribeiroia ondatrae*, Echino = *Echinostoma* sp., Cephalo = *Cephalogonimus* sp., Alaria = *Alaria* sp., Rv = *Ranavirus* sp., Bd = *Batrachochytrium dendrobatidis*). Upper triangular elements show among-species correlation parameters. Black indicates correlations that are probably positive or probably negative (95% of posterior probability mass greater than or less than zero); grey indicates otherwise. Lower triangular elements show bivariate smoothed scatter plots of species-level random effects, with host species codes printed at the posterior means (Anbo = *Anaxyrus boreas*, Psre = *Pseudacris regilla*, Lica = *Lithobates catesbeianus*, Tagr = *Taricha granulosa*, Tato = *Taricha torosa*).

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1502v1 | CC-BY 4.0 Open Access | rec: 12 Nov 2015, publ: 12 Nov 2015

421 # Figure 4

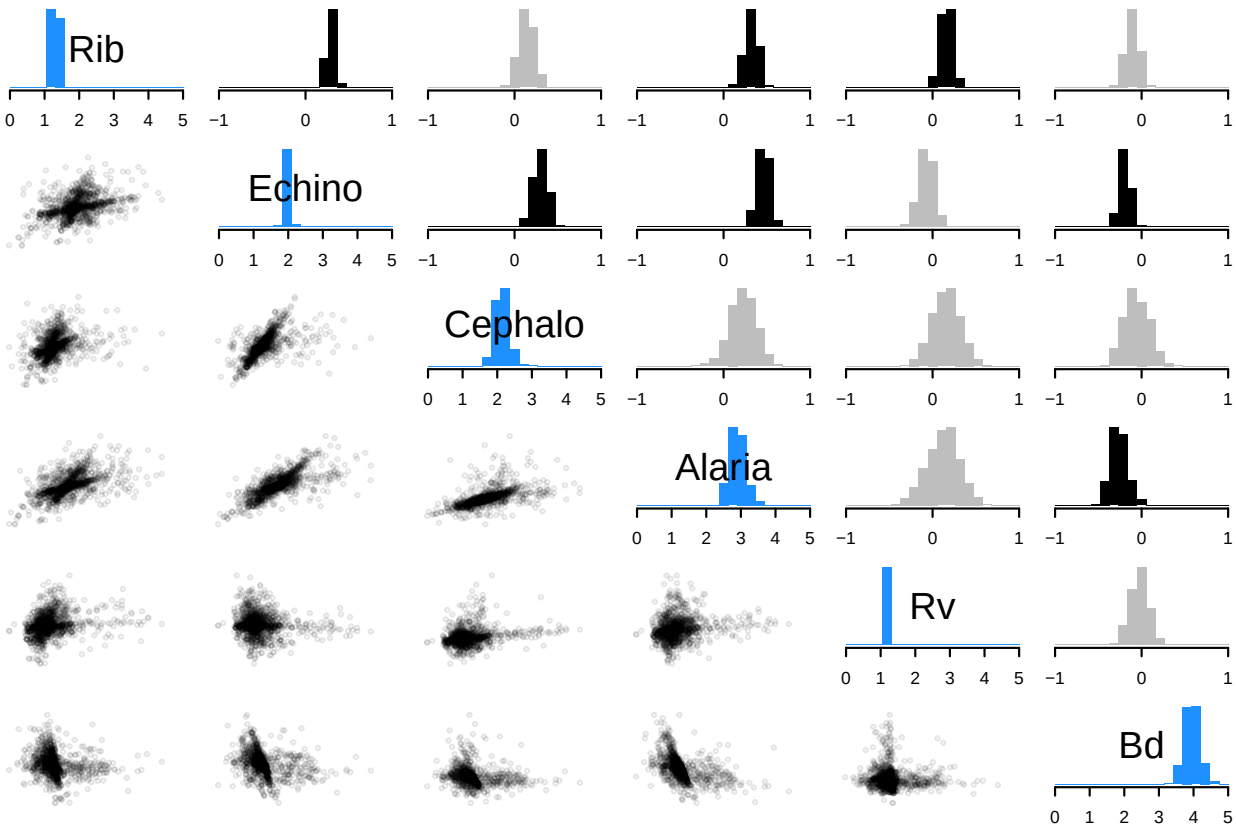

Figure 4: Individual level variance covariance matrix and random effect posteriors. Diagonal elements display the among host individual standard deviation in abundance for parasite species (Rib = *Ribeiroia ondatrae*, Echino = *Echinostoma* sp., Cephalo = *Cephalogonimus* sp., Alaria = *Alaria* sp., Rv = *Ranavirus* sp., Bd = *Batrachochytrium dendrobatidis*). Upper triangular elements show among-individual correlation parameters. Black indicates correlations that are probably positive or probably negative (95% of posterior probability mass greater than or less than zero); grey indicates otherwise. Lower triangular elements show bivariate scatter plots of the posterior means of the individual-level random effects corresponding to the intersection of the species in the rows and columns.

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1502v1 | CC-BY 4.0 Open Access | rec: 12 Nov 2015, publ: 12 Nov 2015

Figure 5

,

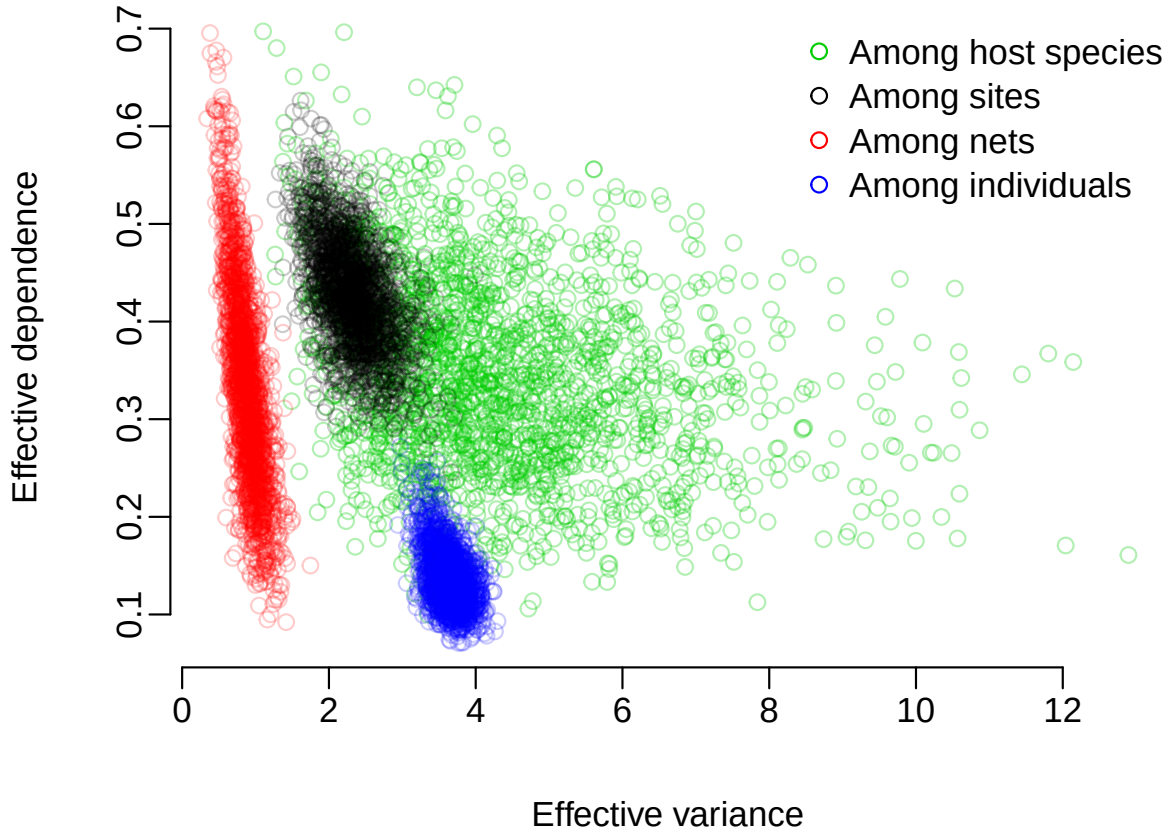

Figure 5: Bivariate posterior distributions of the effective variance and dependence for the multivariate random effects. Each point represents a simulated draw from the posterior. Effective variance measures the magnitude of spread in any direction of the random effects, and effective dependence measures the magnitude of among-species correlation.

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1502v1 | CC-BY 4.0 Open Access | rec: 12 Nov 2015, publ: 12 Nov 2015