Ecology, 91(11), 2010, pp. 3322–3331
© 2010 by the Ecological Society of America

Temporal, spatial, and between-host comparisons of patterns of parasitism in lake zooplankton

Meghan A. Duffy,[1,6] Carla E. Cáceres,[2] Spencer R. Hall,[3] Alan J. Tessier,[4] and Anthony R. Ives[5]

[1]*School of Biology, Georgia Institute of Technology, Atlanta, Georgia 30332-0230 USA*
[2]*School of Integrative Biology, University of Illinois, Urbana, Illinois 61801 USA*
[3]*Department of Biology, Indiana University, Bloomington, Indiana 47405 USA*
[4]*Division of Environmental Biology, National Science Foundation, 4201 Wilson Boulevard, Arlington, Virginia 22230 USA*
[5]*Department of Zoology, University of Wisconsin, Madison, Wisconsin 53706 USA*

Abstract. In nature, multiple parasite species infect multiple host species and are influenced by processes operating across different spatial and temporal scales. Data sets incorporating these complexities offer exciting opportunities to examine factors that shape epidemics. We present a method using generalized linear mixed models in a multilevel modeling framework to analyze patterns of variances and correlations in binomially distributed prevalence data. We then apply it to a multi-lake, multiyear data set involving two *Daphnia* host species and nine microparasite species. We found that the largest source of variation in parasite prevalence was the species identities of host–parasite pairs, indicating strong host–parasite specificity. Within host–parasite combinations, spatial variation (among lakes) exceeded interannual variation. This suggests that factors promoting differences among lakes (e.g., habitat characteristics and species interactions) better explain variation in peak infection prevalence in our data set than factors driving differences among years (e.g., climate). Prevalences of parasites in *D. dentifera* were more positively correlated than those for *D. pulicaria*, suggesting that similar factors influenced epidemic size among parasites in *D. dentifera*. Overall, this study demonstrates a method for parsing patterns of variation and covariation in infection prevalence data, providing greater insight into the relative importance of different underlying drivers of parasitism.

Key words: Daphnia dentifera; Daphnia pulicaria; *generalized linear mixed model (GLMM); host–parasite systems;* Metschnikowia bicuspidata; *Michigan, USA;* Pasteuria ramosa; Polycaryum laeve; Spirobacillus cienkowskii.

Introduction

Recent ecological studies have identified many possible drivers of infectious diseases. For example, climate (Pascual et al. 2000, Thomson et al. 2006), physical habitat characteristics (Cáceres et al. 2006, Johnson et al. 2006*b*), community context (Ostfeld and Holt 2004, Duffy et al. 2005), host species identity (LoGiudice et al. 2008, Hall et al. 2009), and parasite species identity (Mitchell-Olds and Bradley 1996, Ebert et al. 2000) have all been identified as potentially important processes explaining patterns of variation in host–parasite systems. Yet, because these processes operate at many spatial and temporal scales, it has proved difficult to evaluate their relative importance or manipulate many of them experimentally.

The fact that different drivers of infectious diseases operate across different temporal and spatial scales, however, also provides an opportunity: by studying patterns of variation across scales, we can gain insight into the relative importance of different processes that influence the intensity of parasitism experienced by a given host species. For example, if interannual variation in climate strongly shapes disease, one could anticipate large variation among years. However, if some host species are more susceptible to parasitism than others, then the greatest variation should occur among host species.

While these analyses are likely to be informative, they are not trivial. One challenge involves the distribution of these kinds of parasitism data: prevalence data often are distinctly nonnormally distributed. For example, given the commonness of host–parasite specificity (Poulin 2007) and rarity of some parasites, multi-host and multi-parasite data on infection prevalence will likely contain many zeroes. Here we illustrate a method for analyzing binomially distributed prevalence data and apply it to a data set on parasitism in lake *Daphnia* populations. This data set summarizes maximal infection prevalence of nine different parasites in two *Daphnia* host species in 15 lakes from 2003 to 2007. We had previously studied different drivers of parasite prevalence of two parasites of *D. dentifera*, the bacterium *Spirobacillus* and especially the yeast *Metschnikowia* (e.g., physical habitat

Manuscript received 5 September 2009; revised 20 January 2010; accepted 17 March 2010. Corresponding Editor: K. L. Cottingham.
[6] E-mail: duffy@gatech.edu

characteristics, selective vertebrate predation, and invertebrate predation: Duffy et al. 2005, Cáceres et al. 2006, 2009, Hall et al. 2010). Here, we broaden the scope of these previous studies by looking at a broader set of parasites and an additional host and by using an analysis that allows us to gain insight into the relative importance of these different possible drivers of parasitism.

Methods

Study system

We studied *Daphnia* populations in 15 lakes in southwestern Michigan, USA, near the Kellogg Biological Station (Appendix A). The two focal host species, *Daphnia dentifera* (formerly *Daphnia galeata mendotae* and *Daphnia rosea*) and *Daphnia pulicaria*, are common and often dominant grazers in stratified lakes in North America. *Daphnia pulicaria* is more common in spring and summer in these Michigan lakes, whereas *D. dentifera* is more common in summer and autumn (Hu and Tessier 1995, Cáceres and Tessier 2004). During our sampling period, *D. pulicaria* were too rare in seven of these lakes to obtain good infection data. Therefore, our analysis only includes data on infections in *D. pulicaria* in the remaining eight lakes.

We monitored infection prevalence of nine common parasites in *D. dentifera* from 2003 to 2006 and in *D. pulicaria* from 2004 to 2007, following the methods in Cáceres et al. (2006; also see Appendix A). Briefly, on each sampling date we collected 3–4 bottom-to-surface samples of zooplankton from the deep basin of each lake using a 153-μm Wisconsin net. On average, each lake was sampled 9.6 times/year for *D. dentifera* parasites and 5.2 times/year for *D. pulicaria* parasites. On each sampling date, live samples were used to determine the prevalence of infection in each lake by scanning at least 400 *Daphnia* or until the entire sample was searched. Because *Daphnia* are normally transparent, infections are relatively easy to identify by observing hosts under a stereomicroscope at 25–50× magnification. Higher magnification (100–1000×) was used in cases in which the parasite identity was ambiguous at lower levels of magnification. Our estimates of parasite prevalence are likely to be underestimates, since infections can only be detected once symptoms become visually apparent. This bias, which is a common problem in infectious disease research (e.g., Holmstad et al. 2003, O'Meara et al. 2007), is likely to be consistent within individual parasite species, but almost surely varies among different parasite species. Thus, comparisons of infection prevalence across different parasite species should be interpreted with this potential source of bias in mind.

Six of the focal parasites observed have already been described: the microsporidians *Gurleya* sp. and *Larssonia obtusa*, the yeast *Metschnikowia bicuspidata*, the chytrid *Polycaryum laeve*, and the bacteria *Pasteuria ramosa* and *Spirobacillus cienkowskii* (see Plate 1) (Green 1974, Ebert 2005, Johnson et al. 2006*a*, Rodrigues et al. 2008). Two of the remaining parasites have not been positively identified taxonomically; one of these is a fungal brood parasite ("brood"; Hall et al. 2005*b*), while the other is an unidentified oomycete (Green 1974, Wolinska et al. 2008). The final parasite is a Burkholdaria-type bacterium ("BB") that we are currently working to describe.

Our first objective was to parse the sources of variation among host species, parasite species, lake, and year. We used the data from the eight lakes in which both hosts were common and the three years (2004–2006) in which we monitored infections in both host species. Our second objective was to examine the pattern of variation within each individual host–parasite pairing, which allowed us to focus on the variation among lakes and years. For *D. dentifera*, this analysis included data from all 15 lakes and 2003–2006; for *D. pulicaria*, this analysis included data from eight lakes and 2004–2007.

Analyses

The primary challenge in analyzing these data is that the maximum prevalences of infection values are not normally distributed. Instead, they follow a binomial process (1 = infection, 0 = not) in which n individuals are sampled at any one time point but only a fraction are infected with a given parasite. Sampling according to a binomial process introduces measurement error (error that would disappear if the sample sizes n were very large). This measurement error may confound the partitioning of sources of "true" variation and reduce or obscure correlations. Therefore, accounting for the binomial sampling process should improve the analysis of patterns of variances and correlations among parasite infections among lakes, among years, and between host species.

The approach we used takes advantage of the statistical structure of generalized linear mixed models (GLMMs). Mixed models in general, and GLMMs in particular, are often used to estimate regression coefficients when covariation exists among samples, for example, when repeated measurements are taken on the same individual (Gelman and Hill 2007, McCulloch et al. 2008); in these cases, the covariance is considered a nuisance that must be extracted to correctly estimate regression coefficients. However, GLMMs can also be used to focus on the variance–covariance structure of non-Gaussian (i.e., nonnormal) data. To illustrate this, let y_{phlt} denote the number of individual hosts of species h infected by parasite species p in lake l during year t, so for example y_{1hlt} and y_{2hlt} would give the number of hosts of species h in lake–year l–t infected with parasites 1 and 2, respectively. A statistical model can be formulated as follows:

$$y_{phlt} \sim \text{binomial}(\mu_{phlt}, n_{phlt})$$

$$\mu_{phlt} = \text{logit}^{-1}(\beta_0 + \varepsilon_p + \varepsilon_h + \varepsilon_l + \varepsilon_t)$$

$$\varepsilon_p \sim \mathcal{N}(0, \sigma_p^2)$$

$$\varepsilon_h \sim \mathcal{N}(0, \sigma_h^2)$$

$$\varepsilon_l \sim \mathcal{N}(0, \sigma_l^2)$$

$$\varepsilon_t \sim \mathcal{N}(0, \sigma_t^2). \quad (1)$$

Here, y_{phlt} is binomially distributed with μ_{phlt} giving the probability (between 0 and 1) that an individual host is infected (i.e., the prevalence of infection), and n_{phlt} is the number of hosts in a given parasite–host–lake–year sample. The probability of a host being infected is itself given by an inverse logit function of normally distributed random variables ε_p, ε_h, ε_l, and ε_t that each have their own variances. Thus, this model treats the probability μ_{phlt} in the binomial distribution as itself a random variable. Because μ_{phlt} takes on values between 0 and 1, the sum $\beta_0 + \varepsilon_p + \varepsilon_h + \varepsilon_l + \varepsilon_t$ is logit-transformed to bound its value accordingly; although a different function than the logit could have been used, a logit function is a natural and common choice for a binomial distribution. This formulation is essentially the same as the multilevel ANOVA presented by Gelman and Hill (2007) and applied to ecological examples by Qian and Shen (2007).

We have presented Eq. 1 in the style of a multilevel model (Gelman and Hill 2007), although it would be written equally well using GLMM formalism involving fixed and random effects (McCulloch et al. 2008). The key to the model formulation is that the probabilities of being infected in the binomial distribution are assumed to be given by Gaussian distributions that are transformed through a logit link function; therefore, the error terms ε_p, ε_h, ε_l, and ε_t represent random effects in the GLMM. Analyses of the variance and correlation patterns of prevalence are thus performed on the underlying prevalences given by μ_{phlt} after accounting for (extracting) the variance occurring in the binomial sampling process. Variance and covariance patterns in prevalence are then assessed in terms of the variances and covariances in logit(μ_{phlt}). As we illustrate below, this provides a flexible framework for analyzing prevalence data.

Partitioning sources of variation.—Eq. 1 gives a model of the simple case in which the variance in logit(μ_{phlt}) can be divided simply into separate components for parasites (p), hosts (h), lakes (l), and years (t). It can be extended to include interactions among these factors. For example, there are often differences among hosts in their infection by different parasites (Poulin 2007). This type of host-specific susceptibility to different parasites can be included in the model given by Eq. 1 with an additional normal random variable, ε_{ph}, in the distribution of μ_{phlt}

$$\mu_{phlt} = \text{logit}^{-1}(\beta_0 + \varepsilon_p + \varepsilon_h + \varepsilon_{ph} + \varepsilon_l + \varepsilon_t)$$

$$\varepsilon_{ph} \sim \mathcal{N}(0, \sigma_{ph}^2). \quad (2)$$

The random variable ε_{ph} takes a different value for each parasite–host pair. If the variance in ε_{ph} is zero ($\sigma_{ph}^2 = 0$), then the variability among parasite–host pairs is given solely by $\varepsilon_p + \varepsilon_h$ with corresponding variance $\sigma_p^2 + \sigma_h^2$. If, in contrast, the variance in ε_{ph} is not zero ($\sigma_{ph}^2 > 0$), then the variability among parasite–host pairs given by $\sigma_p^2 + \sigma_h^2 + \sigma_{ph}^2$ implies that some parasite species are associated with some host species. This is conceptually analogous to an interaction effect between parasite species and host species that might be detected in fixed effects, but here we place this interaction in the random effects component of the GLMM.

There are numerous ways in which the GLMM could be constructed to explore different ways of partitioning variances. In addition to the partitioning illustrated above, we investigated interactions between the other random effects; for example, there could be an interaction between the random effects for lakes and years (ε_{lt}). To adjudicate among different models that partition variances in different ways, we use the Akaike information criterion (AIC) to select the variance–covariance structure that best fits the data. For an explicit example of model construction, see the Supplement. For all analyses we present in the main text, we use the routine lmer( ) (Bates et al. 2008) in the R programming language (R Core Development Team 2008; see Appendix E for code).

Finally, it may be useful to compare the variance explained by a given model to the total variance in prevalence, or equivalently, to compute the residual variance in a model. For GLMMs, this involves some technical issues and therefore is addressed in Appendix B. This appendix presents a Bayesian (Markov chain Monte Carlo [MCMC]) approach to estimating GLMMs (Gelman and Hill 2007), which allows greater flexibility in formulating models to partition the variance in GLMMs than lmer( ). Appendix B also shows that the performance of lmer( ) and MCMC when applied with the same statistical model are similar, and we use simulations to demonstrate the good performance of lmer( ) for the data set we analyze here.

Correlations between species.—GLMMs can also be used to estimate the correlation between infection levels of different parasite species on the same host species or between hosts for the same parasite species. These correlations can help to identify similarities between parasite species in drivers of parasite prevalence. To illustrate this, we consider the question of whether two parasite species have prevalences that are positively correlated among lakes and years for the same host. The Gaussian component of the GLMM could be formu-

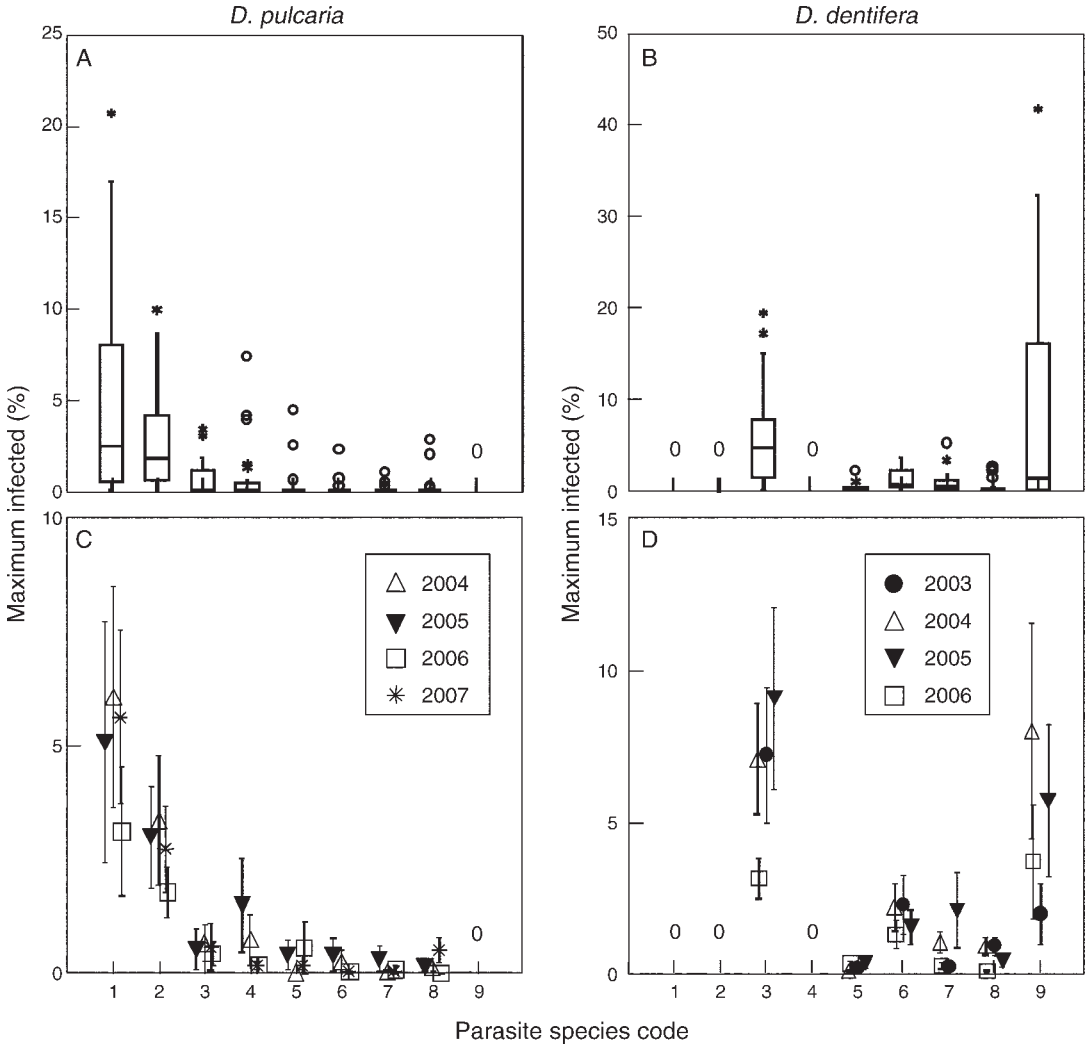

FIG. 1. Maximum infection prevalence of nine parasite species in two hosts (*Daphnia pulicaria* and *D. dentifera*) in 15 lakes in southwestern Michigan, USA. (A, B) Box plots of maximum infection prevalence of nine parasites in eight lakes from 2004 to 2006. The data shown in this figure correspond to those used in the analysis presented in Table 1. Note that the *y*-axis scales differ between the two panels. In Appendix D, we split the data for each parasite into a separate panel. The horizontal line in the center of the boxes represents the median, while the box encompasses the central 50% of the values. The whiskers extend to 1.5 times the interquartile range. Stars indicate observed values that fall between 1.5 and 3 times the interquartile range, while circles represent observed values that are >3 times the interquartile range. (C) Yearly averages of maximum infection prevalence in *D. pulicaria* in eight lakes from 2004 to 2007. (D) Yearly averages of maximum infection prevalence in *D. dentifera* in 15 lakes from 2003 to 2006. In panels (C) and (D), each point represents the average maximum infection prevalence of a given parasite in a given year. Error bars represent ±SE. Within each parasite species, points for different years are offset slightly along the *x*-axis. For all panels, the parasites include: (1) the chytrid *Polycaryum laeve*, (2) a Burkholderia-type bacterium ("BB"), (3) a fungal brood parasite ("brood"), (4) the microsporidians *Gurleya* sp. and (5) *Larssonia obtusa*, (6) the bacteria *Spirobacillus cienkowskii* and (7) *Pasteuria ramosa*, (8) an oomycete, and (9) the yeast *Metschnikowia bicuspidata*. Cases in which a given host species was never observed to be infected with a given parasite species (e.g., *Metschnikowia* in *D. pulicaria*) are indicated with a zero.

lated as

$$\mu_{plt} = \text{logit}^{-1}(\beta_1 x_1 + \beta_2 x_2 + \varepsilon_{p|lt})$$

$$\varepsilon_{p|lt} \sim \text{independent } \mathcal{N}(0, \boldsymbol{\Sigma}_p). \quad (3)$$

Here, β_1 and β_2 are the means of the first and second parasite prevalences. The values of the categorical variable x_1 are one for data points from parasite species 1 and zero for data points from parasite species 2; categorical variable x_2 is defined similarly but with values of zero given for parasite 1 and values of one given for parasite 2. The random variable $\varepsilon_{p|lt}$ is normally distributed, but to account for possible correlations between parasites nested within each lake–year (indicated by $p \mid lt$), the covariance matrix for $\varepsilon_{p|lt}$ is

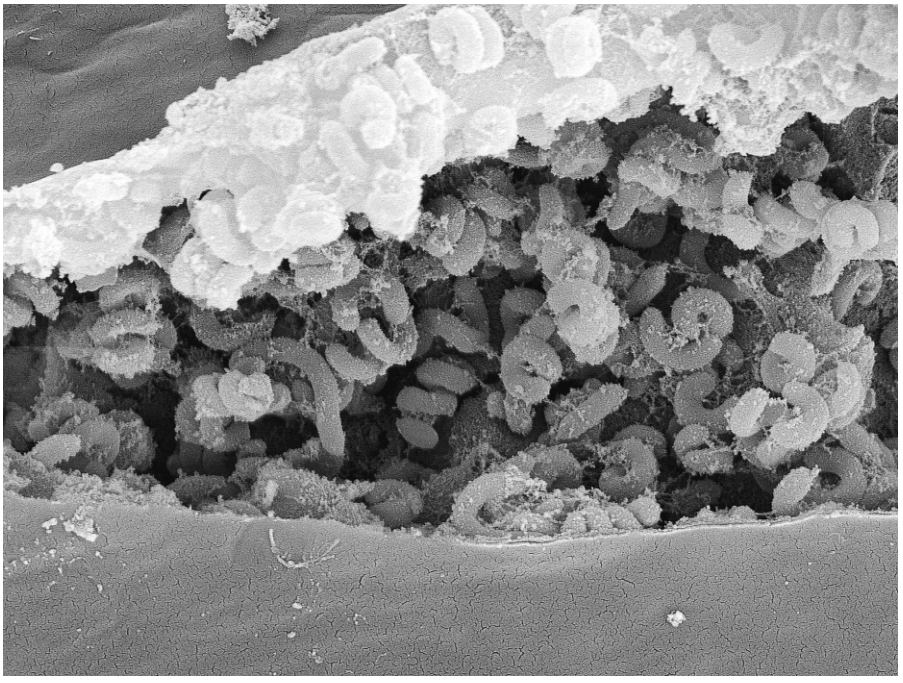

PLATE 1. Cells of the bacterial parasite *Spirobacillus cienkowskii* filling the carapace of a *Daphnia* individual. All of the spiral-shaped cells are the bacterial parasite. Photo credit: M. A. Duffy and Carol Flegler.

$$\mathbf{\Sigma}_p = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}. \quad (4)$$

The correlation between prevalences of the two parasite species is given by the estimate of ρ. To assess the statistical significance of the estimate of ρ, we used parametric bootstrapping in which the fitted model was used to simulate 1000 data sets, the GLMM was fit to each simulation data set, and the resulting bootstrap distribution of ρ was used to approximate the distribution of the estimator of ρ. We also asked whether infections by the same parasite are correlated between hosts in the same lake–years. For this analysis, the two parasites attacking the same host in Eqs. 3 and 4 are replaced by the two hosts being attacked by the same parasite. If the resulting estimate of the correlation ρ is positive, then lake–years in which the parasite has high prevalence in *D. pulicaria* are also likely to have high prevalence in *D. dentifera*.

RESULTS

We found substantial variation in maximal infection prevalence (Fig. 1A, B; Appendix D). The parasites that were most common in *D. pulicaria*, "BB" and *Polycaryum*, were never observed infecting *D. dentifera*. Conversely, one of the most common *D. dentifera* parasites, *Metschnikowia*, was never observed infecting *D. pulicaria*; in addition, the brood parasite was common in *D. dentifera* but rare in *D. pulicaria*.

The high variation in prevalence across host–parasite pairings is supported by the statistical analysis, which found that most variation in prevalence among parasite–host–lake–years was contained in the host × parasite interaction term σ^2_{ph} (Table 1). The best-supported model does not include a host effect (indicating that the hosts do not differ substantially in the maximal infection prevalences they suffer), nor does it include a parasite effect (indicating that maximal prevalences of parasites do not differ significantly). Taken together, these indicate that the large variance contained in the host × parasite interaction term is caused by parasites having different prevalences on different hosts (Fig.

TABLE 1. Variance estimates for best-fitting models.

Effect	Variance
A) Full data set†	
Parasite × host	11.5
Parasite × lake	1.76
Parasite × lake × year	0.829
Host × lake × year	0.666
Host × lake	0.516
Parasite × year	0.423
B) Data set using only parasites that attack both hosts	
Parasite × host	1.27
Host	1.26
Host × lake × year	1.03
Parasite × year	0.983
Parasite × lake	0.865
Parasite × lake × year	0.413

† Best-fitting model for the full data set (including both hosts and all parasites) using a generalized linear model assuming binomially distributed data.

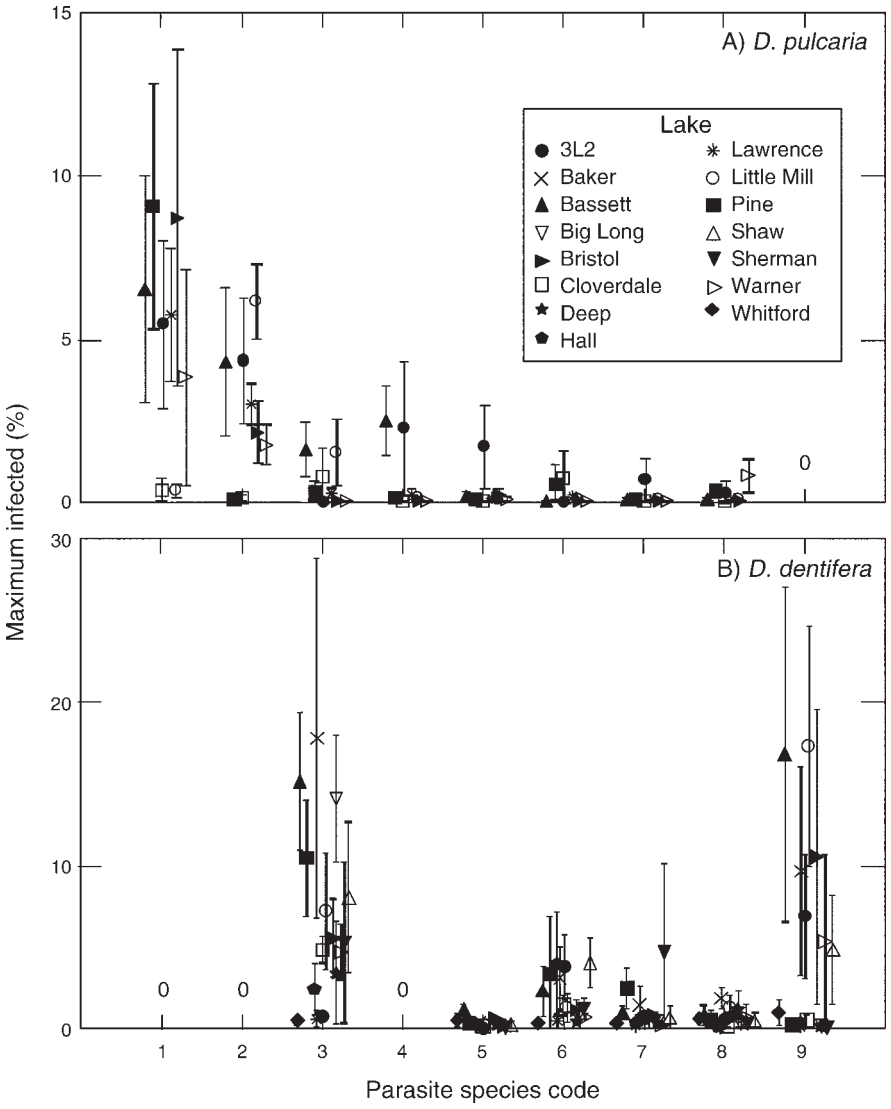

FIG. 2. Lake averages of maximum infection prevalence in (A) *Daphnia pulicaria* and (B) *D. dentifera*; each point represents the maximum infection prevalence of a given parasite in a given host species in a given lake, averaged across years. Data for *D. pulicaria* are from eight lakes from 2004 to 2007. Data for *D. dentifera* are from 15 lakes from 2003 to 2006; each point represents the average maximum infection prevalence of a given parasite in a given lake. Error bars represent ±SE. Parasites are numbered as in Fig. 1. Within each parasite species, points for different lakes are offset slightly along the *x*-axis to make the symbols more apparent. In Appendix D, we split the data for each parasite into a separate panel, to make the differences among lakes more apparent.

1A, B). Given that some of the parasites were only observed infecting one of the two host species, some of this result is due to strict host specificity. When we restrict our analysis to the five parasites that infect both host species, we find that the host × parasite interaction and host species identity explain the most variation (Table 1).

Because the host × parasite interaction was so dominant, it obscured our ability to look at other sources of variation. We therefore repeated the analyses separately for each individual host–parasite pairing. This allowed us to analyze the variation at the lake and year scales. For all host–parasite pairings, the greatest variation was explained by lake, not year (Fig. 1C, D, Fig. 2, Table 2; Appendix D). In most cases (e.g., for *D. dentifera* host: brood, *Metschnikowia*, and *Spirobacillus*; all parasites of *D. pulicaria*), the lake effect accounted for considerably more variation than the year effect. In a few instances, the year effect rivaled the lake effect (e.g., for *D. dentifera*: oomycete, *Pasteuria*).

Numerous drivers could act at the lake level. To gain insight into whether parasite prevalence is influenced by the same drivers, we looked for correlations among parasites. We found that maximal infection prevalences

TABLE 2. Variance among lakes and years for individual host–parasite pairings (data shown in Figs. 1C, D and 2).

Host	Parasite	Variance	
		Lake	Year
Daphnia dentifera	Brood	1.88	0.269
	Oomycete	1.35	1.29
	Larssonia	1.13	0.608
	Metschnikowia	6.52	0.187
	Pasteuria	1.22	1.07
	Spirobacillus	0.950	0.0577
Daphnia pulicaria	BB	3.48	0.0513
	Brood	1.36	0.110
	Oomycete	5.70	1.79
	Gurleya	6.16	0.973
	Polycaryum	2.35	0.0329
	Spirobacillus	5.10	1.60

Note: "BB" stands for Burkholdaria-type bacterium.

of parasites of *D. dentifera* were generally positively correlated (Table 3), suggesting similar drivers. In general, there were weaker correlations among *D. pulicaria* parasites (Table 4). Parasite prevalences were not strongly correlated between the two host species (Table 5).

To investigate how our GLMM approach differed from a more conventional approach that does not explicitly account for the binomial sampling of infected individuals, we repeated these analyses by logit-transforming the prevalence data and treating them as normally distributed. As presented in Appendix C, the "standard" approach left more variance unexplained than the GLMM approach and therefore was less likely to identify patterns (specifically, the host × lake × year and parasite × lake × year interactions in Table 1) in the data. Also, the "standard" approach generally gave lower estimates of correlations and was less likely to identify correlations as significantly different from zero (Table 3). Thus, the GLMM approach, by factoring out measurement error associated with binomial sampling, gives more informative results.

DISCUSSION

Many factors have been proposed as important drivers of parasitism in natural populations. However, the difficulties inherent in simultaneously studying processes across multiple scales (space, time, and species) have severely limited our understanding of the relative importance of different factors. Here we present a method for partitioning sources of variation in large-scale data sets and apply it to a data set of multiple host and parasite species. By identifying the dominant sources of variation in the data, we obtain insight into the relative importance of different drivers of parasitism.

We found that by far the greatest variance in maximum prevalence in our multi-parasite–host–lake–year data was contained in the host–parasite interaction. This indicates substantial differences among the two host species in the maximal infection prevalences of different parasites. Indeed, several of the parasites were only observed infecting one or the other of the two host species (Fig. 1). This suggests strong species-level differences in the susceptibility of hosts to different parasites and/or species-level differences in the infectivity of different parasites on different hosts. Cases in which closely related species differed strongly in their susceptibility to a given parasite have been observed in numerous systems previously (e.g., *Sellaphora* diatoms and chytrids, Mann 1989, 1999; anther smuts in Caryophyllacea, Le Gac et al. 2007), though, in other systems (e.g., primates, Pedersen et al. 2005), such strong specificity was found to be relatively rare. Our data suggest that strong specificity is the dominant

TABLE 3. Correlations between parasites in *Daphnia dentifera* populations.

Parasite	Brood	Oomycete	Larssonia	*Metschnikowia*	*Pasteuria*
Oomycete	0.220				
Larssonia	0.307*	−0.079			
Metschnikowia	0.338	0.480†	0.507†		
Pasteuria	0.492†	−0.058	−0.068	0.368	
Spirobacillus	0.302†	0.313*	0.064	0.487†	0.299*

Notes: Tests of the null hypothesis $\rho = 0$ were performed using parametric bootstrapping (1000 simulations). Note that the statistical significance of ρ does not necessarily correspond to its magnitude; this is due to differences in prevalence among parasite species, with lower prevalence giving less statistical power to reject the null hypothesis.

* $P < 0.05$; † $P < 0.02$.

TABLE 4. Correlations between parasites in *Daphnia pulicaria* populations.

Parasite	BB	Brood	Oomycete	*Gurleya*	*Polycaryum*
Brood	0.407*				
Oomycete	−0.146	−0.001			
Gurleya	0.633†	0.530*	−0.213		
Polycaryum	0.175	−0.003	0.062	0.440	
Spirobacillus	−0.318	−0.098	−0.068	−0.270	0.029

Notes: Tests of the null hypothesis $\rho = 0$ were performed using parametric bootstrapping (1000 simulations). BB stands for Burkholdaria-type bacterium.
* $P < 0.05$; † $P < 0.02$.

driver of patterns of parasitism in our *Daphnia*–parasite system.

Looking within individual host–parasite pairings, we found that there was always more variation among lakes than among years in peak prevalence. This suggests that factors influencing differences among lakes (such as physical habitat characteristics and community context) have more of an effect on parasitism than factors that cause differences among years (such as climate). One way in which physical habitat characteristics may influence parasitism is via effects on free-living infective stages ("spores"). Many of the *Daphnia* parasites are known to produce spores (Ebert 2005, Johnson et al. 2006*a*), and physical habitat characteristics (such as basin shape) can influence the ability of the spores to be resuspended and remain in the water column (though the strength of the effect may vary with spore size and motility). In the *Daphnia dentifera–Metschnikowia* system, we have previously found strong correlations between lake basin shape and maximal infection prevalence (Cáceres et al. 2006, Hall et al. 2010). This may be due, in part, to effects of basin shape on currents that drive the movement of particles (such as parasite spores) from nearshore to offshore (Hall et al. 2010), where they can be ingested by *Daphnia*.

Physical habitat characteristics also strongly influence community context (Tessier and Woodruff 2002, Hall et al. 2010), creating consistent differences among lakes in the densities of predators including fish (Mittelbach 1984) and predatory invertebrates (Duffy et al. 2004, Garcia and Mittelbach 2008, Cáceres et al. 2009). Bluegill sunfish selectively prey upon infected *Daphnia* (*Spirobacillus*, Duffy et al. 2005; *Polycaryum*, Johnson et al. 2006*b*; *Metschnikowia*, Duffy and Hall 2008; *Pasteuria* and oomycete, M. A. Duffy, *unpublished data*), presumably because infections increase the opacity of the normally transparent hosts. While we have not measured the selectivity of fish predation on the remaining four parasites, these also increase host opacity, and, therefore, we expect that fish would also prey selectively on *Daphnia* infected with those parasites. Selective predation should have large effects on parasitism (Packer et al. 2003, Ostfeld and Holt 2004, Hall et al. 2005*a*), including on maximal infection prevalence (Duffy and Hall 2008). Lakes with higher fish predation would be expected to have lower maximal infection prevalences. That, in turn, should lead to positive correlations between maximal infection prevalences of different parasites, since high levels of fish predation should depress all parasites (albeit to varying degrees, depending on the degree of selectivity). These positive correlations should be stronger in *D. dentifera* than in *D. pulicaria*, since *D. dentifera* live higher in the water column and are subject to higher levels of fish predation (Leibold and Tessier 1997). It is therefore interesting that, in our data set, we generally observed positive correlations between parasites of *D. dentifera* (Table 3).

Habitat characteristics have been found to be important in a variety of other disease systems. For example, a survey of parasitism in eels in Nova Scotia found strong effects of pH on the abundance of different parasites (Marcogliese and Cone 1996); in some cases, this effect was thought to be mediated by effects of pH on other members of the food web (specifically mollusks). With Lyme disease, infection prevalence in tick nymphs is strongly correlated with habitat size (Allan et al. 2003). Specifically, larger habitats have lower infection prevalences due to increased densities of hosts that are less competent reservoirs of the Lyme pathogen, *Borrelia burgdorferi*.

Our understanding of the relative importance of different factors that influence parasitism in natural populations has been limited by the difficulty of performing manipulative experiments at sufficiently large spatial scales and sufficiently long temporal scales. Furthermore, it is even more difficult to conduct experimental studies that encompass several potential drivers. It is generally much more feasible (though admittedly still quite labor intensive) to conduct observational studies that encompass multiple scales. However, these data are not easily analyzed, as they are likely to be distinctly nonnormal. In this study, we

TABLE 5. Correlations (ρ) of infection prevalence between *Daphnia dentifera* and *D. pulicaria* for parasite species shared by the two host species.

Parasite	ρ
Brood	0.509†
Oomycete	0.043
Larssonia	0.006
Pasteuria	0.038
Spirobacillus	−0.337

Note: Tests of the null hypothesis $\rho = 0$ were performed using parametric bootstrapping (1000 simulations).
† $P < 0.02$.

presented a method for analyzing these data and show that doing so can quantitatively reveal patterns of variation in parasitism, suggesting the most likely drivers of infection prevalence. As disease ecology matures and moves beyond studying a single factor in a single host–parasite combination, this approach permits a more general understanding of the processes influencing parasitism in nature.

Acknowledgments

We thank the Albert, Brehm, Burdick, Champion, and Gregg families for access to private lakes. Claes Becker, Wiebke Boeing, Bob Duffy, Cindy Hartway, Ally Morgan, Jessica VanDyke, Johnny Tenegra, and Pam Woodruff all provided much-appreciated assistance in the field. Comments from an anonymous reviewer and Song Qian greatly improved this manuscript. Funding for this research was provided by NSF (OCE-0235119, OCE-0235039, DEB-0816613, DEB-0613510, DEB-0614316, and DEB-0827396) and the Georgia Institute of Technology. Findings and conclusions expressed herein are those of the authors and do not necessarily reflect the views of the National Science Foundation. This is KBS contribution 1558.

Literature Cited

Allan, B. F., F. Keesing, and R. S. Ostfeld. 2003. Effect of forest fragmentation on Lyme disease risk. Conservation Biology 17:267–272.

Bates, D., M. Maechler, and B. Dai. 2008. lme4: Linear mixed-effects models using S4 classes. R package. ⟨http://lme4.r-forge.r-project.org/⟩

Cáceres, C. E., S. R. Hall, M. A. Duffy, A. J. Tessier, C. Helmle, and S. MacIntyre. 2006. Physical structure of lakes constrains epidemics in *Daphnia* populations. Ecology 87: 1438–1444.

Cáceres, C. E., C. J. Knight, and S. R. Hall. 2009. Predator spreaders: predation can enhance parasite success in a planktonic host–parasite system. Ecology 90:2850–2858.

Cáceres, C. E., and A. J. Tessier. 2004. To sink or swim: variable diapause strategies among *Daphnia* species. Limnology and Oceanography 49:1333–1340.

Duffy, M. A., and S. R. Hall. 2008. Selective predation and rapid evolution can jointly dampen effects of virulent parasites on *Daphnia* populations. American Naturalist 171: 499–510.

Duffy, M. A., S. R. Hall, A. J. Tessier, and M. Huebner. 2005. Selective predators and their parasitized prey: Are epidemics in zooplankton under top-down control? Limnology and Oceanography 50:412–420.

Duffy, M. A., A. J. Tessier, and M. Kosnik. 2004. Testing the ecological relevance of *Daphnia* species designations. Freshwater Biology 49:55–64.

Ebert, D. 2005. Ecology, epidemiology and evolution of parasitism in *Daphnia*. National Library of Medicine, National Center for Biotechnology Information, Bethesda, Maryland, USA.

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

Garcia, E. A., and G. G. Mittelbach. 2008. Regional coexistence and local dominance in *Chaoborus*: species sorting along a predation gradient. Ecology 89:1703–1713.

Gelman, A., and J. Hill. 2007. Data analysis using regression and multilevel/hierarchical models. Cambridge University Press, Cambridge, UK.

Green, J. 1974. Parasites and epibionts of Cladocera. Transactions of the Zoological Society of London 32:417–515.

Hall, S. R., C. R. Becker, J. L. Simonis, M. A. Duffy, A. J. Tessier, and C. E. Cáceres. 2009. Friendly competition: evidence for a dilution effect among competitors in a planktonic host–parasite system. Ecology 90:791–801.

Hall, S. R., M. A. Duffy, and C. E. Cáceres. 2005*a*. Selective predation and productivity jointly drive complex behavior in host–parasite systems. American Naturalist 165:70–81.

Hall, S. R., M. A. Duffy, A. J. Tessier, and C. E. Cáceres. 2005*b*. Spatial heterogeneity of daphniid parasitism in lakes. Oecologia 143:635–644.

Hall, S. R., R. Smyth, C. R. Becker, M. A. Duffy, C. J. Knight, S. MacIntyre, A. J. Tessier, and C. E. Cáceres. 2010. Why are *Daphnia* in some lakes sicker? Disease ecology, habitat structure, and the plankton. BioScience 60:363–375.

Holmstad, P. R., A. Anwar, T. Lezhova, and A. Skorping. 2003. Standard sampling techniques underestimate prevalence of Avian Hematozoa in willow ptarmigan (*Lagopus lagopus*). Journal of Wildlife Diseases 39:354–358.

Hu, S. S., and A. J. Tessier. 1995. Seasonal succession and the strength of intra- and interspecific competition in a *Daphnia* assemblage. Ecology 76:2278–2294.

Johnson, P. T. J., J. E. Longcore, D. E. Stanton, R. B. Carnegie, J. D. Shields, and E. R. Preu. 2006*a*. Chytrid infections of *Daphnia pulicaria*: development, ecology, pathology and phylogeny of *Polycaryum laeve*. Freshwater Biology 51:634–648.

Johnson, P. T. J., D. E. Stanton, E. R. Preu, K. J. Forshay, and S. R. Carpenter. 2006*b*. Dining on disease: how interactions between infection and environment affect predation risk. Ecology 87:1973–1980.

Le Gac, M., M. E. Hood, E. Fournier, and T. Giraud. 2007. Phylogenetic evidence of host-specific cryptic species in the anther smut fungus. Evolution 61:15–26.

Leibold, M. A., and A. J. Tessier. 1997. Habitat partitioning by zooplankton and the structure of lake ecosystems. Pages 3–30 *in* B. Streit, T. Städler, and C. M. Lively, editors. Evolutionary ecology of freshwater animals. Birkhäuser Verlag, Basel, Switzerland.

LoGiudice, K., S. T. K. Duerr, M. J. Newhouse, K. A. Schmidt, M. E. Killilea, and R. S. Ostfeld. 2008. Impact of host community composition on Lyme disease risk. Ecology 89:2841–2849.

Mann, D. G. 1989. The species concept in diatoms: evidence for morphologically distinct, sympatric gamodemes in 4 epipelic species. Plant Systematics and Evolution 164:215–237.

Mann, D. G. 1999. The species concept in diatoms. Phycologia 38:437–495.

Marcogliese, D. J., and D. K. Cone. 1996. On the distribution and abundance of eel parasites in Nova Scotia: influence of pH. Journal of Parasitology 82:389–399.

McCulloch, C. E., S. R. Searle, and J. M. Neuhaus. 2008. Generalized, linear and mixed models. Wiley and Sons, Hoboken, New Jersey, USA.

Mitchell-Olds, T., and D. Bradley. 1996. Genetics of *Brassica rapa* 3. Costs of disease resistance to three fungal pathogens. Evolution 50:1859–1865.

Mittelbach, G. G. 1984. Predation and resource partitioning in two sunfishes (Centrarchidae). Ecology 65:499–513.

O'Meara, W. P., W. E. Collins, and F. E. McKenzie. 2007. Parasite prevalence: a static measure of dynamic infections. American Journal of Tropical Medicine and Hygiene 77:246–249.

Ostfeld, R. S., and R. D. Holt. 2004. Are predators good for your health? Evaluating evidence for top-down regulation of zoonotic disease reservoirs. Frontiers in Ecology and the Environment 2:13–20.

Packer, C., R. D. Holt, P. J. Hudson, K. D. Lafferty, and A. P. Dobson. 2003. Keeping the herds healthy and alert: implications of predator control for infectious disease. Ecology Letters 6:797–802.

Pascual, M., Xavier Rodo, S. P. Ellner, R. Colwell, and M. J. Bouma. 2000. Cholera dynamics and El Nino-Southern Oscillation. Science 289:1766–1769.

Pedersen, A. B., S. Altizer, M. Poss, A. A. Cunningham, and C. L. Nunn. 2005. Patterns of host specificity and transmission among parasites of wild primates. International Journal for Parasitology 35:647–657.

Poulin, R. 2007. Evolutionary ecology of parasites. Second edition. Princeton University Press, Princeton, New Jersey, USA.

Qian, S. S., and Z. Shen. 2007. Ecological applications of multilevel analysis of variance. Ecology 88:2489–2495.

R Core Development Team. 2008. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rodrigues, J. L. M., M. A. Duffy, A. J. Tessier, D. Ebert, L. Mouton, and T. M. Schmidt. 2008. Phylogenetic characterization and prevalence of *Spirobacillus cienkowskii*: a red-pigmented, spiral-shaped bacterial pathogen of freshwater *Daphnia* species. Applied and Environmental Microbiology 74:1575–1582.

Tessier, A. J., and P. Woodruff. 2002. Cryptic trophic cascade along a gradient of lake size. Ecology 83:1263–1270.

Thomson, M. C., F. J. Doblas-Reyes, S. J. Mason, R. Hagedorn, S. J. Connor, T. Phindela, A. P. Morse, and T. N. Palmer. 2006. Malaria early warnings based on seasonal climate forecasts from multi-model ensembles. Nature 439:576–579.

Wolinska, J., K. C. King, F. Vigneux, and C. M. Lively. 2008. Virulence, cultivating conditions, and phylogenetic analyses of oomycete parasites in *Daphnia*. Parasitology 135:1667–1678.

APPENDIX A

Additional details regarding sampling methods (*Ecological Archives* E091-234-A1).

APPENDIX B

Computing total variance in generalized linear mixed models (GLMM) ANOVAs (*Ecological Archives* E091-234-A2).

APPENDIX C

Comparison between GLMM and "standard" methods (*Ecological Archives* E091-234-A3).

APPENDIX D

Infection prevalences for individual host–parasite species pairings (*Ecological Archives* E091-234-A4).

APPENDIX E

Annotated R code and output for analyses presented in the article (*Ecological Archives* E091-234-A5).

SUPPLEMENT

Data files to be used with annotated R code presented in Appendix E (*Ecological Archives* E091-234-S1).