Ecological Archives E091-234-A1

Meghan A. Duffy, Carla E. Cáceres, Spencer R. Hall, Alan J. Tessier, and Anthony R. Ives. 2010. Temporal, spatial, and between-host comparisons of patterns of parasitism in lake zooplankton. *Ecology* 91:3322–3331.

Appendix A. Additional details regarding sampling methods.

We sampled 15 lakes in Southwest Michigan: Baker, Bassett, Bristol, Cloverdale, Deep, Hall, Lawrence, Little Mill, Long, Pine, Shaw, Sherman, Three Lakes Two, Warner, and Whitford Lakes (Barry and Kalamazoo Counties). During our study, *D. pulicaria* were rare in Baker, Deep, Hall, Long, Shaw, Sherman and Whitford, so we could not obtain accurate estimates of infection prevalence, so our analyses of infections in *D. pulicaria* are restricted to the remaining eight lakes. See Cáceres et al. (2006) for a brief description of these lakes.

LITERATURE CITED

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.

[Back to E091-234]

Ecological Archives E091-234-A2

Meghan A. Duffy, Carla E. Cáceres, Spencer R. Hall, Alan J. Tessier, and Anthony R. Ives. 2010. Temporal, spatial, and between-host comparisons of patterns of parasitism in lake zooplankton. *Ecology* 91:3322–3331.

Appendix B. Computing total variance in generalized kinear mixed models (GLMM) ANOVAs.

Although our analyses have focused on comparing contributions to the variance in parasite prevalence made by different factors (parasite species, host species, lakes, and years), it is sometimes useful to compare the variance explained by a given model to the total variance in prevalence. Here, we present two methods for making this comparison. The first is simple and can easily be performed in the manner of the analyses described in the main text. The second is more complicated and requires more computing.

For the analysis presented in Eq. 2 in the main text, the simplest way to estimate the total variance is with the model

$$\begin{aligned} &\mu_{phlt} = \mathrm{logit}^{-1}(\beta_0 + \varepsilon_A) \\ &\varepsilon_A \sim \mathrm{N}(0, \sigma^2_A) \end{aligned} \qquad \text{(B.1)}$$

where ε_A is a random variable that takes on different values for each μ_{phlt}, so that σ^2_A is the total variance. However, because the data may not be independently distributed (which will be the case when the variance components, e.g., ε_{pt}, have non-zero variance), it is possible for the total variance estimated by σ^2_A to be less than variance obtained by partitioning the model into component sources of variance (e.g., $\sigma^2_p + \sigma^2_h + \sigma^2_{ph} + \sigma^2_l + \sigma^2_t$ in the model given by Eq. 2). This only occurs, however, when the vast majority of the variance is explained by the partitioned model, that is, when there is little residual variance not explained by the model.

A second approach is to explicitly include a residual variance in the model

$$\begin{aligned} &y_{phlt} \sim \mathrm{binomial}(\mu_{phlt}, n_{phlt}) \\ &\mu_{phlt} = \mathrm{logit}^{-1}(\beta_0 + \varepsilon_p + \varepsilon_h + \varepsilon_l + \varepsilon_t + \varepsilon_{phlt}) \\ &\varepsilon_p \sim \mathrm{N}(0, \sigma^2_p) \\ &\varepsilon_h \sim \mathrm{N}(0, \sigma^2_h) \\ &\varepsilon_l \sim \mathrm{N}(0, \sigma^2_l) \\ &\varepsilon_t \sim \mathrm{N}(0, \sigma^2_t) \\ &\varepsilon_{phlt} \sim \mathrm{N}(0, \sigma^2_{phlt}) \end{aligned} \qquad \text{(B.2)}$$

where σ^2_{phlt} is the estimate of the residual variance. This is a GLMM ANOVA as formulated by Gelman and Hill (2007) and Qian and Shen (2007). The difficulty with this model is that it cannot be estimated by the current version of lmer(). Nonetheless, Bayesian approaches can be used, as illustrated by Qian and Shen (2007, Appendix D). We provide computer code modified from Qian and Shen (2007) to estimate parameters in this model (see Supplement).

When applied to the same model (Eq. 1 in the main text), lmer() and Bayesian estimates of the variance components are similar (Table B1A), especially when comparing the proportion of variance explained. The total variance computed from equation B1 using lmer() is slightly less than the summed variance components, indicating that the component model (Eq. 2 in the main text) explains most of the variance in prevalence. This is confirmed by the analysis including residual variance (Eq. B.2) in which the residual variance is only 8% of the total. There are slight differences in the variance components calculated using the model with residuals (Eq. B.2) vs. those calculated using the model without residuals (Eq. 2 in the main text); in particular, the variance of the Parasite × Lake × Year effect is smaller. It is unclear why this occurs. Nonetheless, the overall conclusions about the major sources of variation in the data are similar.

Finally, these analyses allow us to compare the results from lmer() to those from MCMC. Gelman and Hill (2007) note that for analyses with small sample sizes, the Bayesian approach may give better (less-biased) estimates of the variance components of a model. The results in Table B1B suggest that lmer() and MCMC give very similar partitions of the sources of variance in the data sets (for the case when residual variation is not included). To give an explicit test of the bias in the lmer() analysis, we performed a simulation (parametric bootstrap) using the variance parameters estimated from the data (Table B1C). The close match between the "true" parameter values used to simulate the data and the mean parameter values from 100 simulations demonstrates the good performance of lmer() in estimating the variance components of the model.

TABLE B1. Variance components for the best-fitting model corresponding to Table 1A in the main text.

A. Variance estimates. In the column labeled "lmer," variances for the model given by Eq. 2 in the main text were computed using lmer(), with the "summed variance" giving the sum of the partitioned variances and the "total variance" estimated from the model given by Eq. B.1. The column labeled "MCMC" gives Bayesian estimates for the same model. The column labeled "MCMC with residuals" gives Bayesian estimates for the model given by Eq. B.2.

Effect	lmer	MCMC	MCMC with residuals
Parasite × Host	11.5	14.4	13.7
Parasite × Lake	1.76	2.0	1.7
Parasite × Lake × Year	0.829	1.0	0.0
Host × Lake × Year	0.666	0.8	0.5
Host × Lake	0.516	0.5	0.5
Parasite × Year	0.423	0.5	0.6
Residual			1.4
Summed variance	15.67		
Total variance	15.59		

B. Proportion variance explained.

Effect	lmer	MCMC	MCMC with residuals
Parasite × Host	0.73	0.75	0.74
Parasite × Lake	0.11	0.10	0.09
Parasite × Lake × Year	0.052	0.05	0.0
Host × Lake × Year	0.043	0.04	0.03
Host × Lake	0.033	0.03	0.03
Parasite × Year	0.027	0.03	0.03
Residual			0.08
Summed variance	15.67		
Total variance	15.59		

C. Parametric bootstrap.

Effect	"true" values	mean estimates from 100 simulations
Parasite × Host	11.5	11.4
Parasite × Lake	1.76	1.61
Parasite × Lake × Year	0.829	0.81
Host × Lake × Year	0.666	0.68
Host × Lake	0.516	0.54
Parasite × Year	0.423	0.43

LITERATURE CITED

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

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

[Back to E091-234]

Ecological Archives E091-234-A4

Meghan A. Duffy, Carla E. Cáceres, Spencer R. Hall, Alan J. Tessier, and Anthony R. Ives. 2010. Temporal, spatial, and between-host comparisons of patterns of parasitism in lake zooplankton. *Ecology* 91:3322–3331.

Appendix D. Infection prevalences for individual host-parasite species pairings.

In Figs. D1–D8, we present the individual box plots for each of the eight parasites found infecting *D. pulicaria* in the eight lake populations. These figures contain the data from Fig. 1A; here, the data for each parasite are plotted separately to make the distributions of each parasite species clearer. In Figs. D9–D14, we present the individual box plots for each of the six parasites found infecting *D. dentifera* in the 15 lake populations. These figures contain the data from Fig. 1B; here, the data for each parasite are plotted separately to make the distributions of each parasite species clearer.

In Figs. D15–D22, we present the data for each of the eight parasites found infecting *D. pulicaria* in the eight lake populations. These figures contain the data from the top panel of Fig. 2; here, the data for each parasite are plotted separately to show the differences among lakes more clearly. In Figs. D23–D28, we present the data for each of the six parasites found infecting *D. dentifera* in the fifteen lake populations. These figures contain the data from the bottom panel of Fig. 2.

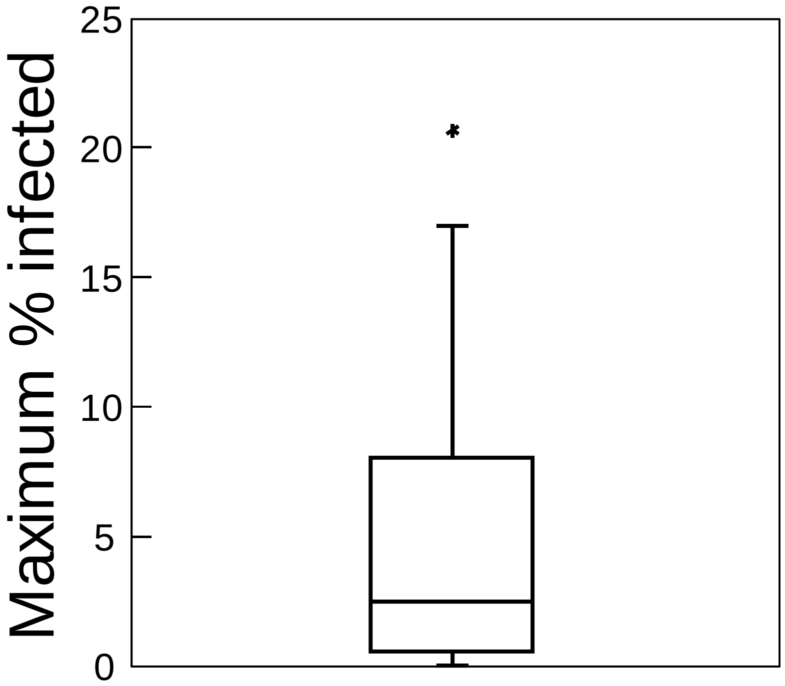

FIG. D1. Box plot of maximum infection prevalences of *Polycaryum laeve* in *D. pulicaria* in eight lakes. This figure contains data from Fig. 1A.

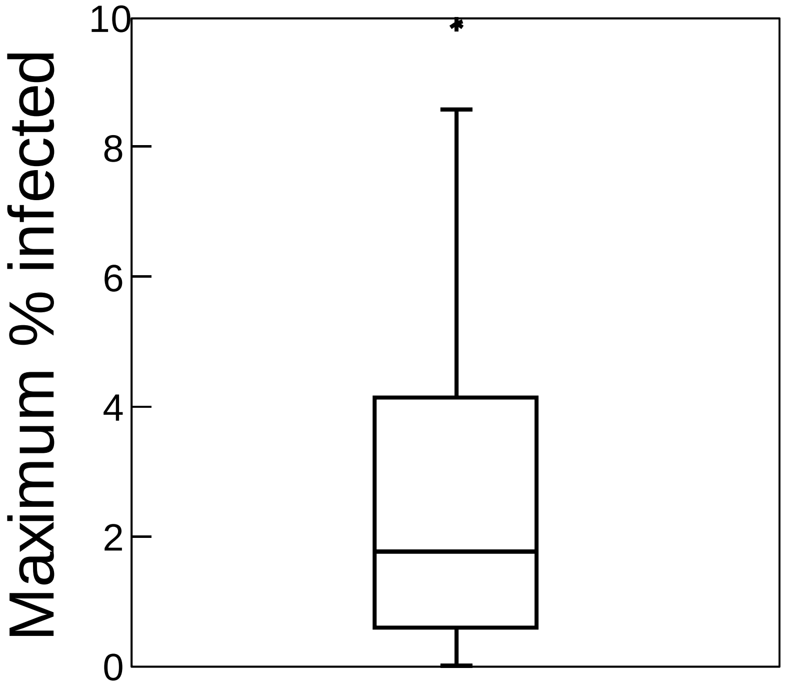

FIG. D2. Box plot of maximum infection prevalences of the Burkholderia-type bacterium (“BB”) in *D. pulicaria* in eight lakes. This figure contains data from Fig. 1A.

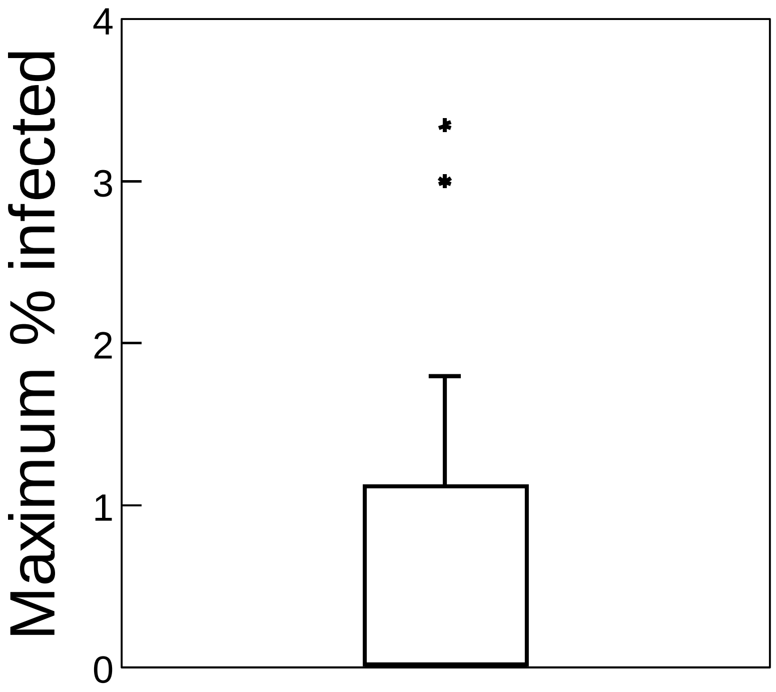

FIG. D3. Box plot of maximum infection prevalences of the brood parasite in *D. pulicaria* in eight lakes. This figure contains data from Fig. 1A.

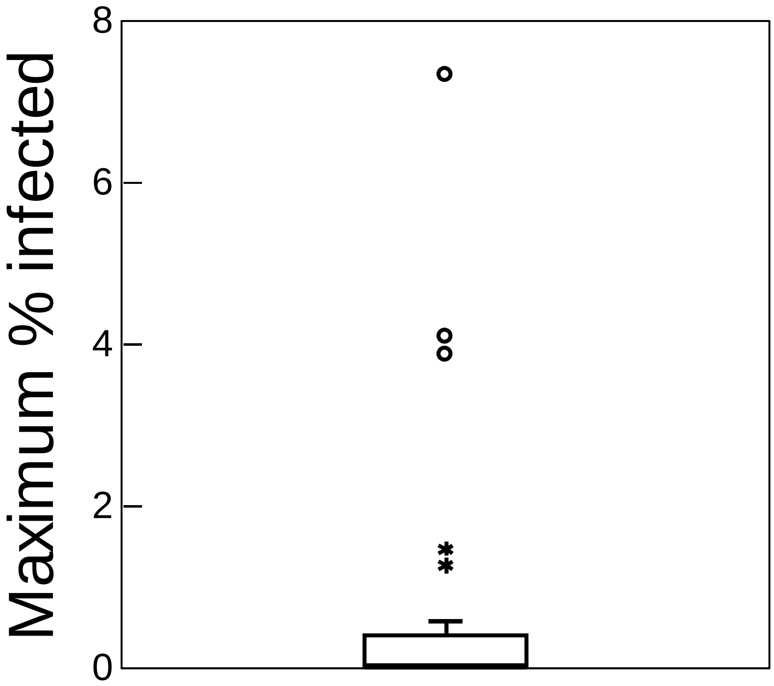

FIG. D4. Box plot of maximum infection prevalences of *Gurleya sp.* in *D. pulicaria* in eight lakes. This figure contains data from Fig. 1A.

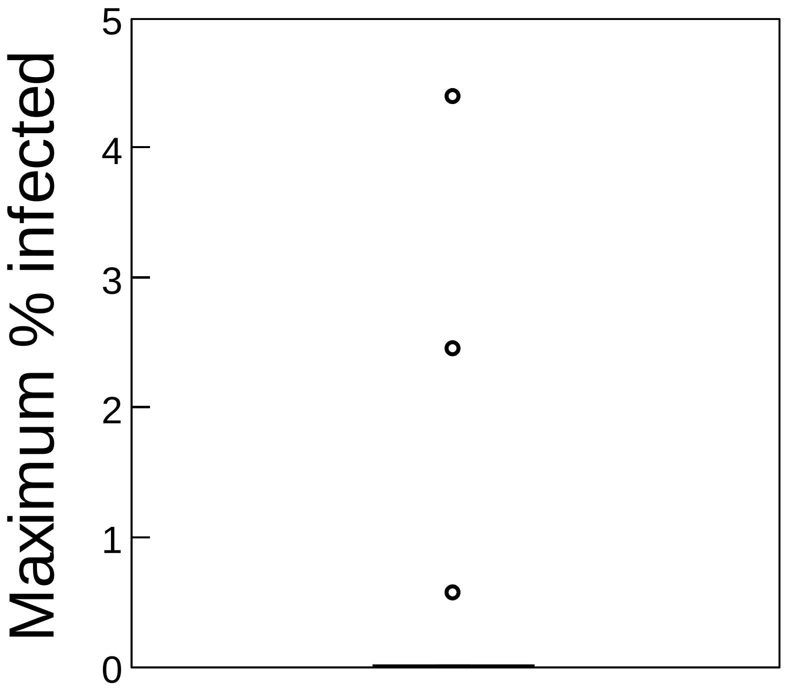

FIG. D5. Box plot of maximum infection prevalences of *Larssonia obtusa* in *D. pulicaria* in eight lakes. This figure contains data from Fig. 1A.

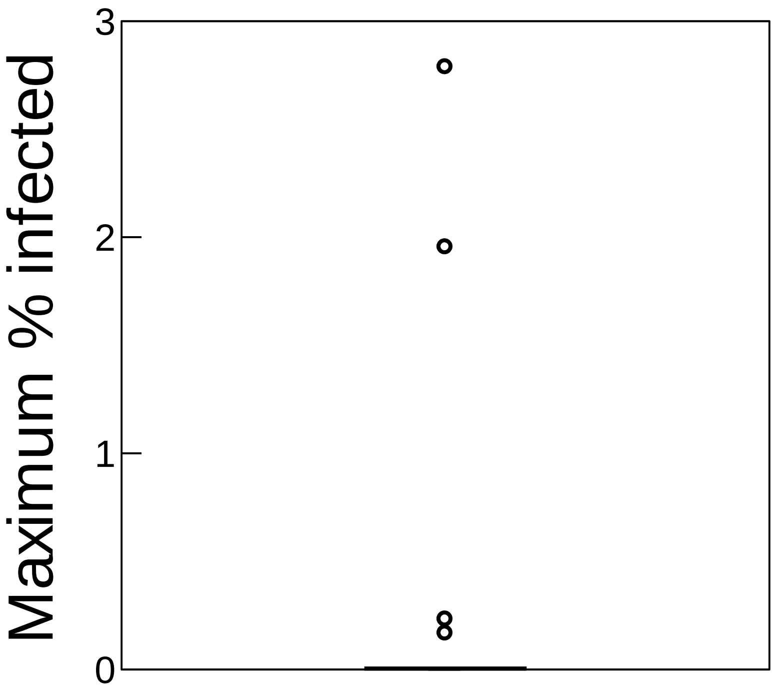

FIG. D6. Box plot of maximum infection prevalences of *Spirobacillus cienkowskii* in *D. pulicaria* in eight lakes. This figure contains data from Fig. 1A.

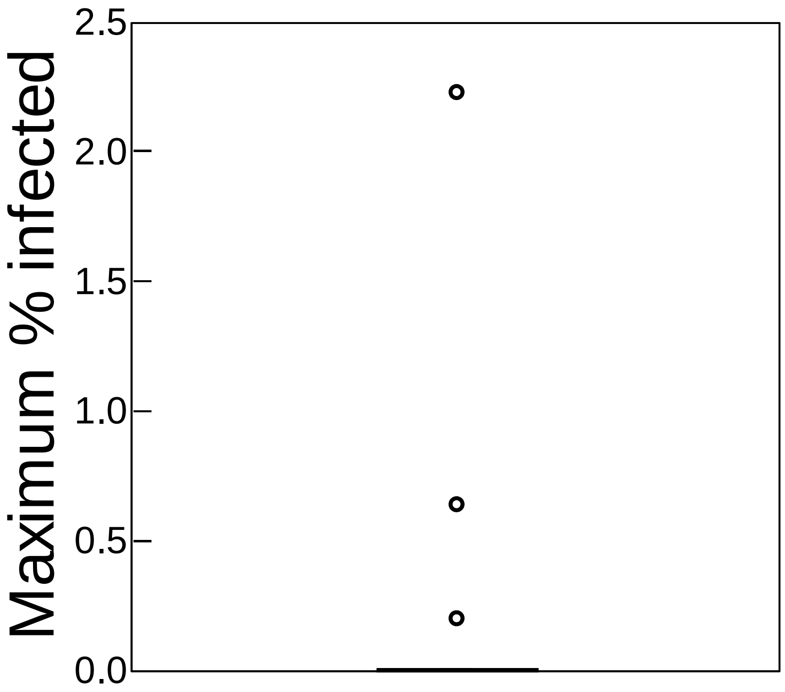

FIG. D7. Box plot of maximum infection prevalences of *Pasteuria ramosa* in *D. pulicaria* in eight lakes. This figure contains data from Fig. 1A.

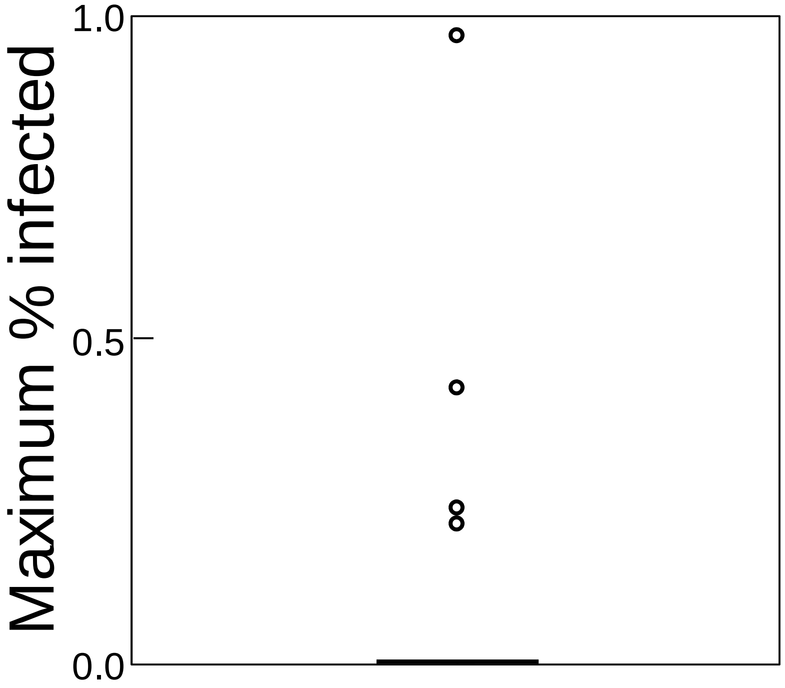

FIG. D8. Box plot of maximum infection prevalences of oomycete parasites in *D. pulicaria* in eight lakes. This figure contains data from Fig. 1A.

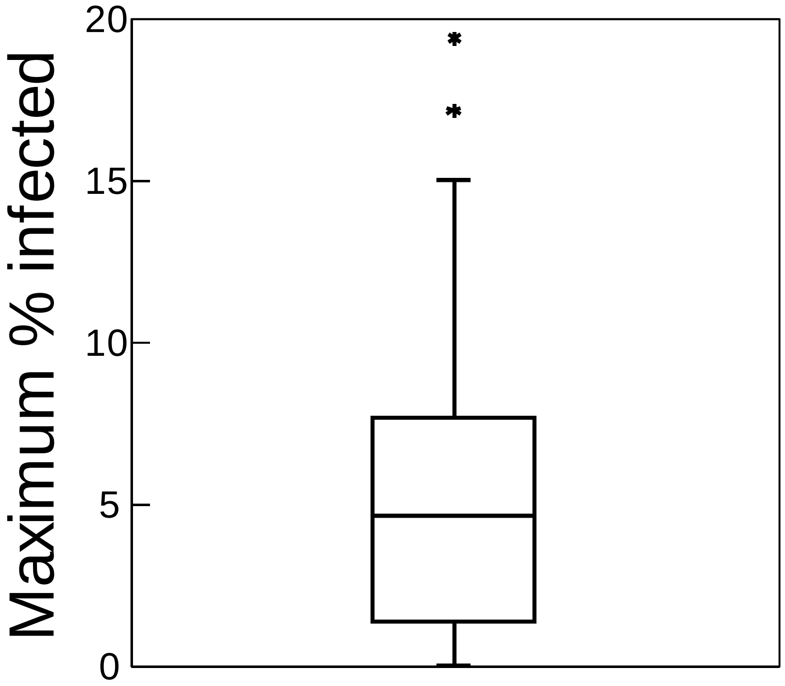

FIG. D9. Box plot of maximum infection prevalences of the brood parasite in *D. dentifera* in 15 lakes. This figure contains data from Fig. 1B.

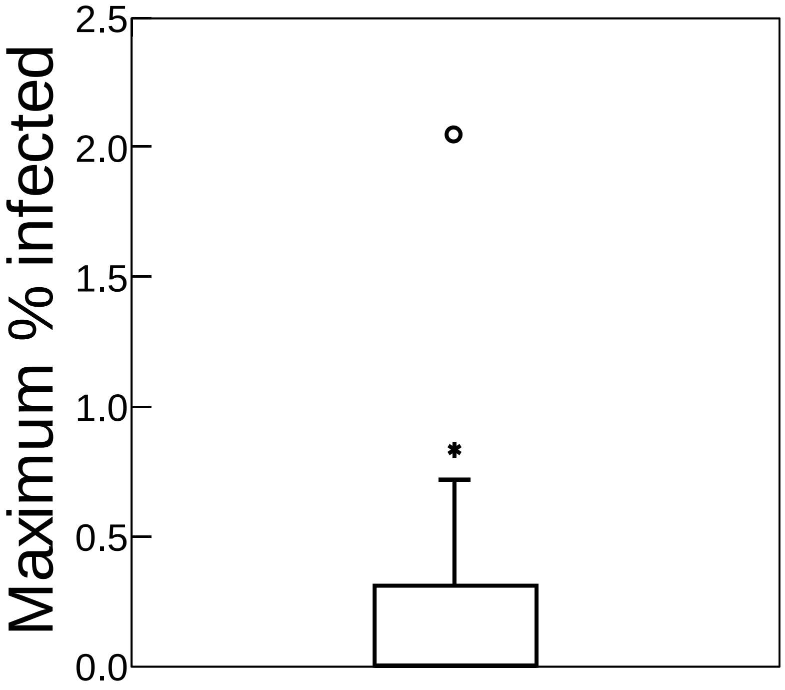

FIG. D10. Box plot of maximum infection prevalences of *Larssonia obtusa* in *D. dentifera* in 15 lakes. This figure contains data from Fig. 1B.

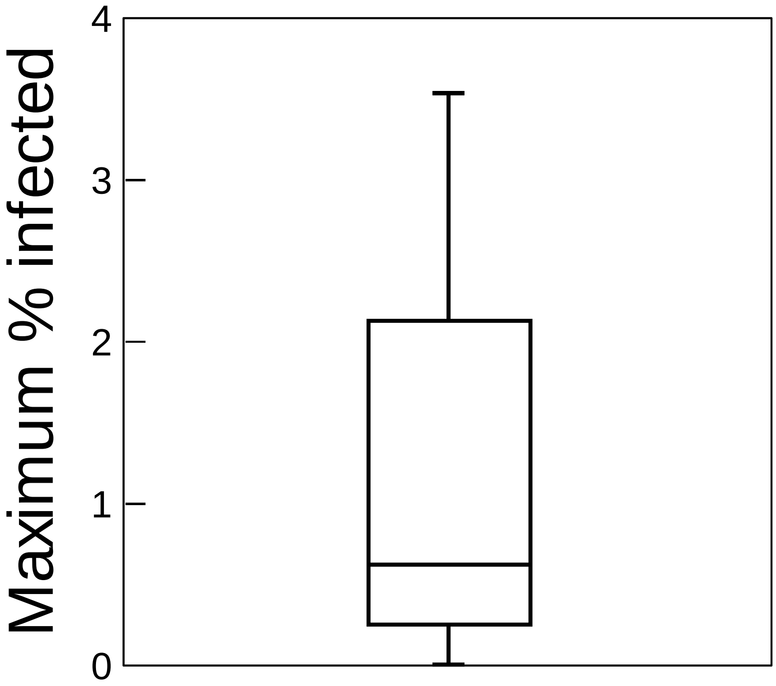

FIG. D11. Box plot of maximum infection prevalences of *Spirobacillus cienkowskii* in *D. dentifera* in 15 lakes. This figure contains data from Fig. 1B.

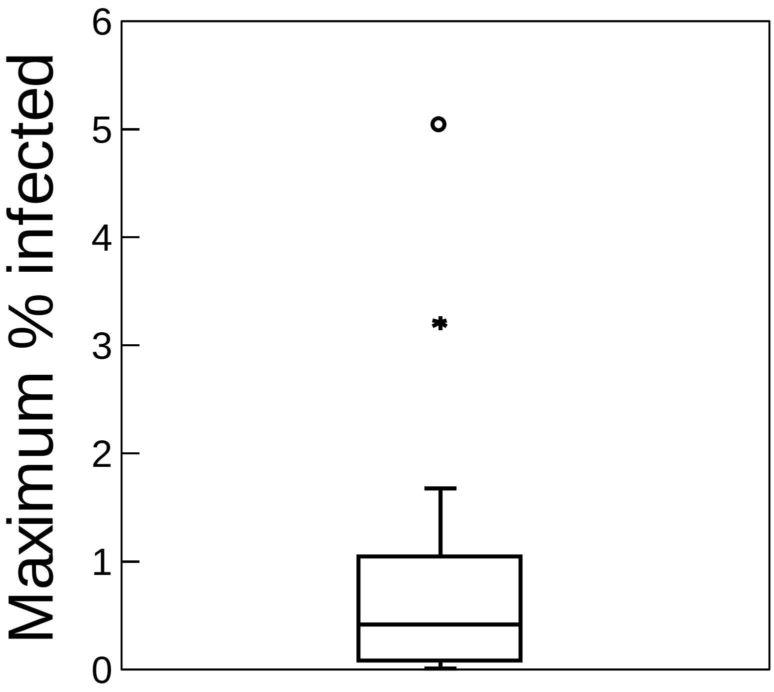

FIG. D12. Box plot of maximum infection prevalences of *Pasteuria ramosa* in *D. dentifera* in 15 lakes. This figure contains data from Fig. 1B.

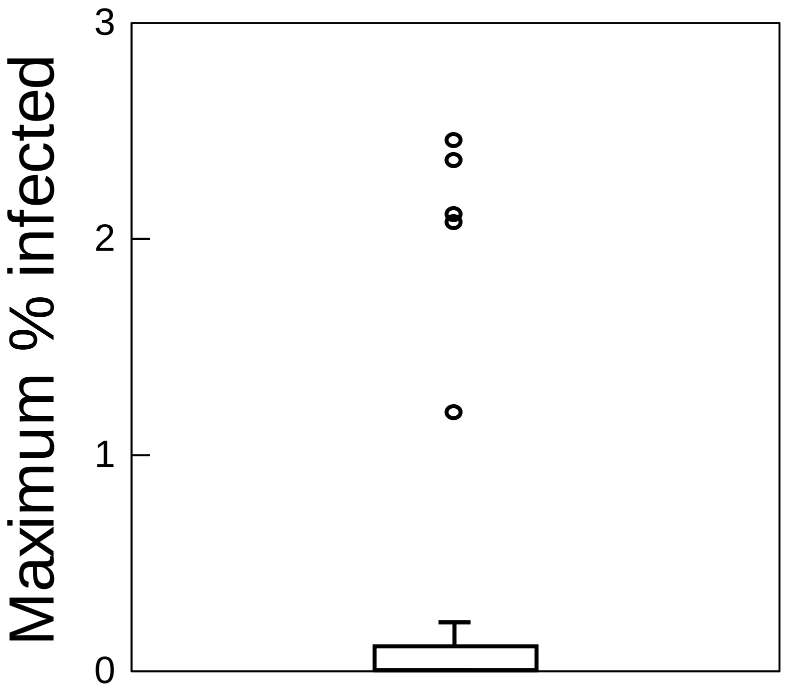

FIG. D13. Box plot of maximum infection prevalences of oomycete parasites in *D. dentifera* in 15 lakes. This figure contains data from Fig. 1B.

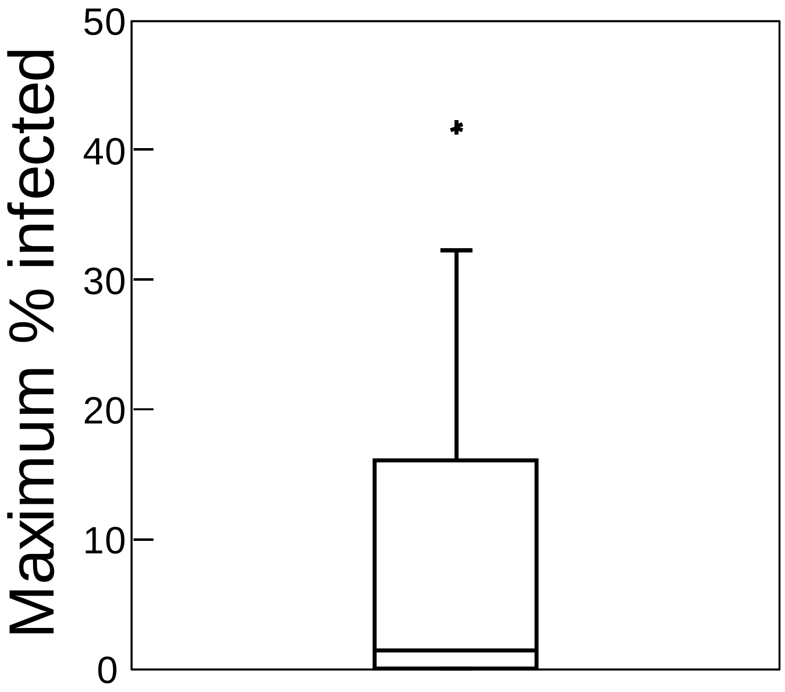

FIG. D14. Box plot of maximum infection prevalences of *Metschnikowia bicuspidata* in *D. dentifera* in 15 lakes. This figure contains data from Fig. 1B.

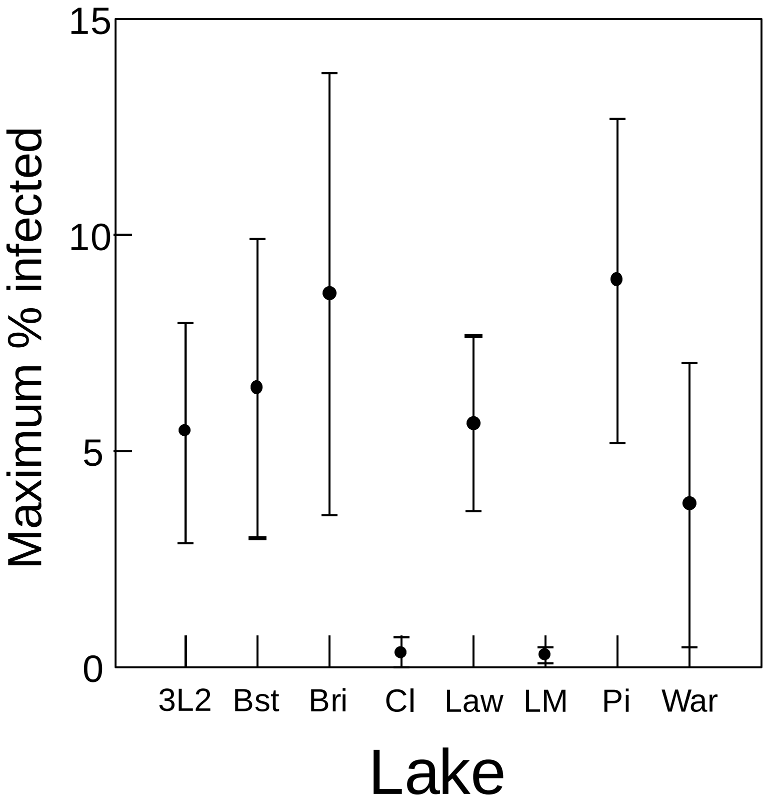

FIG. D15. Prevalence of *Polycaryum laeve* infections in *D. pulicaria* in eight lake populations. Lake abbreviations are: “3L2” = Three Lakes Two, “Bst” = Bassett, “Bri” = Bristol, “Cl” = Cloverdale, “Law” = Lawrence, “LM” = Little Mill, “Pi” = Pine, “War” = Warner. These data are the same as those shown in Fig. 2. Here, we plot each parasite separately to improve clarity. Error bars represent ±1 SE.

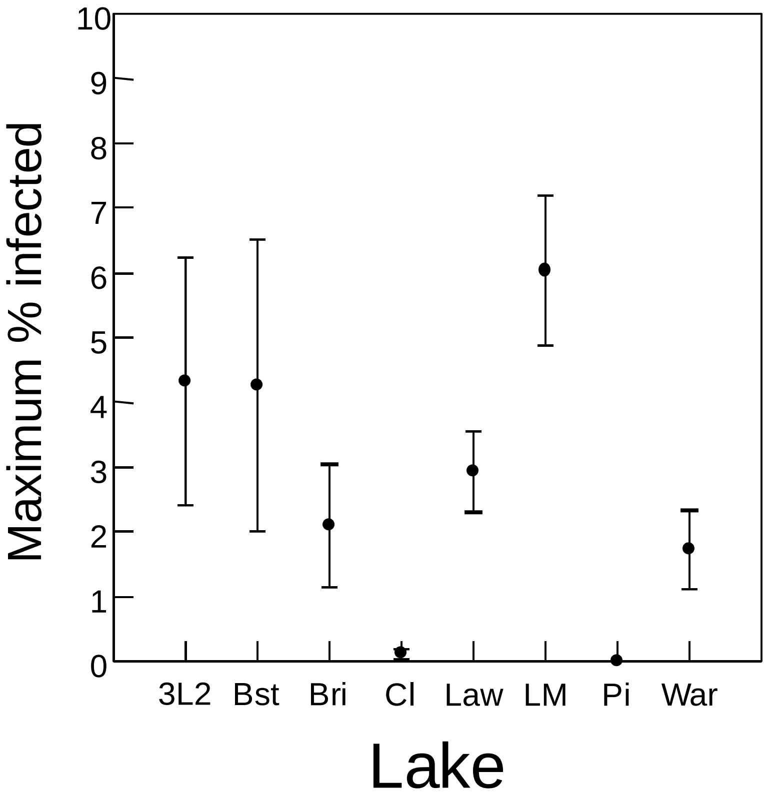

FIG. D16. Prevalence of “BB” (Burkholderia-type bacterium) infections in *D. pulicaria* in eight lake populations. Lake abbreviations are: “3L2” = Three Lakes Two, “Bst” = Bassett, “Bri” = Bristol, “Cl” = Cloverdale, “Law” = Lawrence, “LM” = Little Mill, “Pi” = Pine, “War” = Warner. These data are the same as those shown in Fig. 2. Here, we plot each parasite separately to improve clarity. Error bars represent ±1 SE.

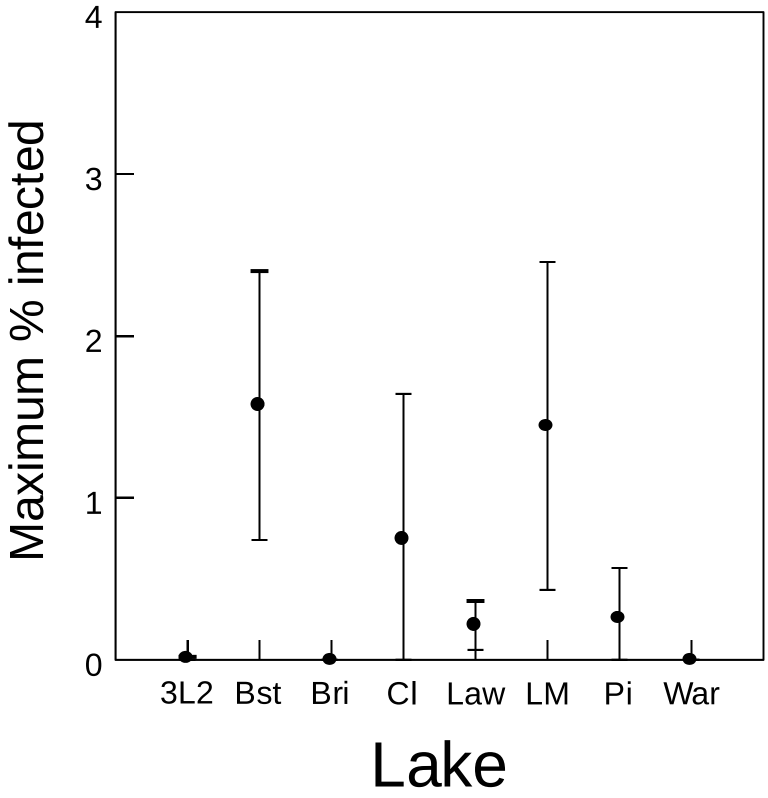

FIG. D17. Prevalence of brood parasite infections in *D. pulicaria* in eight lake populations. Lake abbreviations are: “3L2” = Three Lakes Two, “Bst” = Bassett, “Bri” = Bristol, “Cl” = Cloverdale, “Law” = Lawrence, “LM” = Little Mill, “Pi” = Pine, “War” = Warner. These data are the same as those shown in Fig. 2. Here, we plot each parasite separately to improve clarity. Error bars represent ±1 SE.

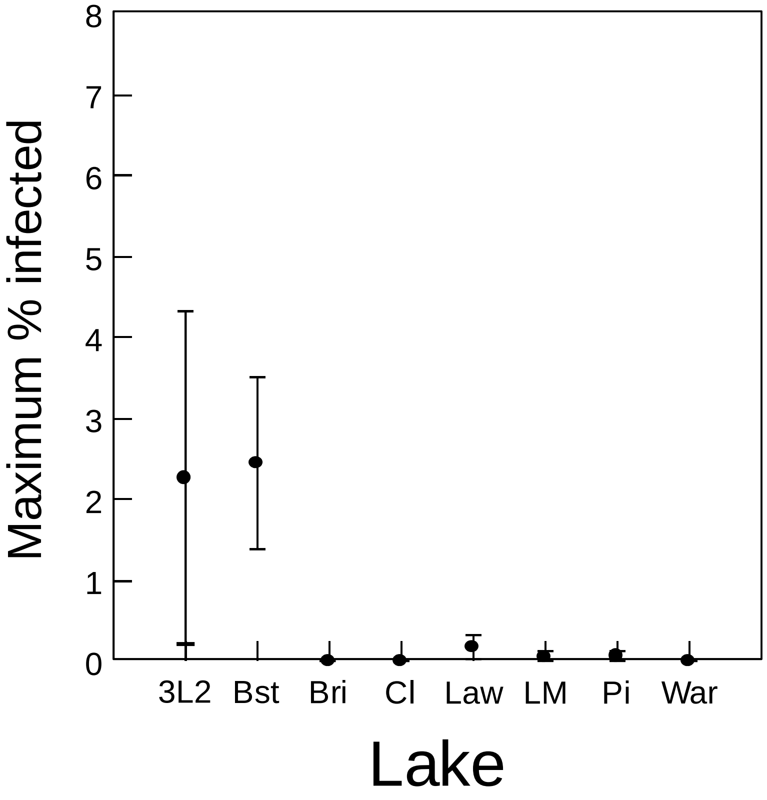

FIG. D18. Prevalence of *Gurleya sp.* infections in *D. pulicaria* in eight lake populations. Lake abbreviations are: “3L2” = Three Lakes Two, “Bst” = Bassett, “Bri” = Bristol, “Cl” = Cloverdale, “Law” = Lawrence, “LM” = Little Mill, “Pi” = Pine, “War” = Warner. These data are the same as those shown in Fig. 2. Here, we plot each parasite separately to improve clarity. Error bars represent ±1 SE.

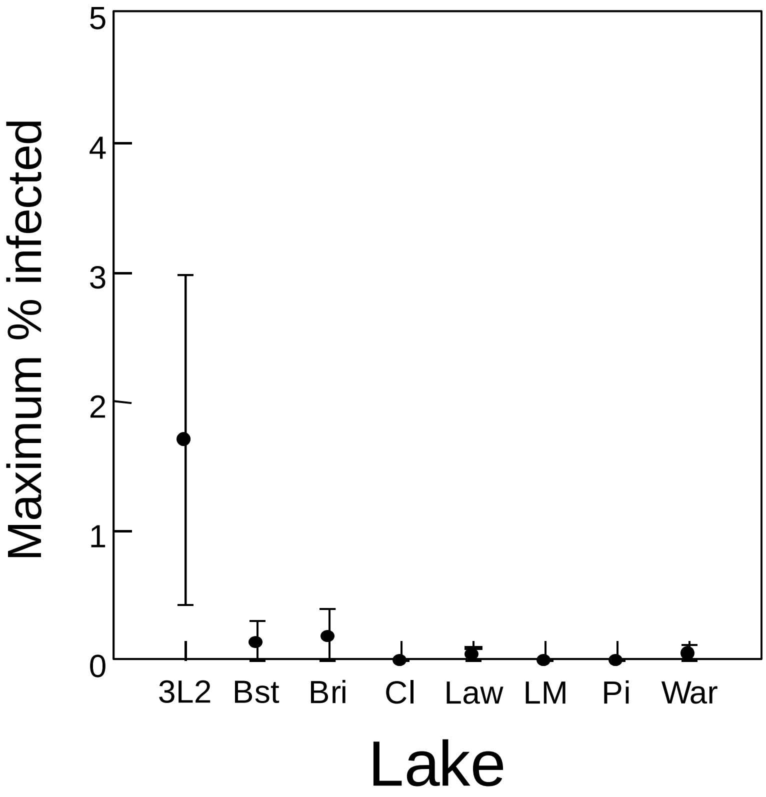

FIG. D19. Prevalence of *Larssonia obtusa* infections in *D. pulicaria* in eight lake populations. Lake abbreviations are: “3L2” = Three Lakes Two, “Bst” = Bassett, “Bri” = Bristol, “Cl” = Cloverdale, “Law” = Lawrence, “LM” = Little Mill, “Pi” = Pine, “War” = Warner. These data are the same as those shown in Fig. 2. Here, we plot each parasite separately to improve clarity. Error bars represent ±1 SE.

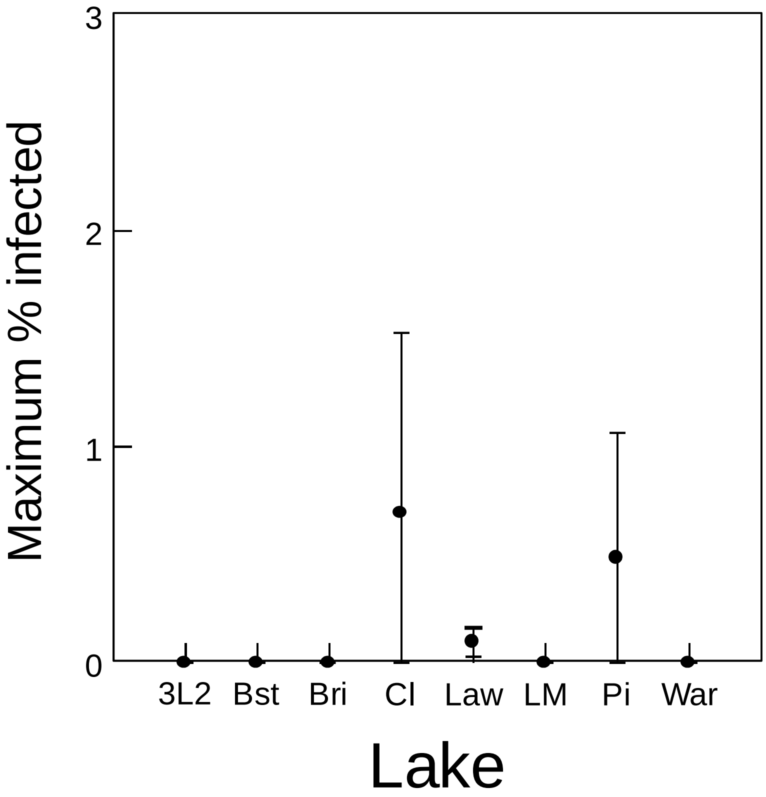

FIG. D20. Prevalence of *Spirobacillus cienkowskii* infections in *D. pulicaria* in eight lake populations. Lake abbreviations are: “3L2” = Three Lakes Two, “Bst” = Bassett, “Bri” = Bristol, “Cl” = Cloverdale, “Law” = Lawrence, “LM” = Little Mill, “Pi” = Pine, “War” = Warner. These data are the same as those shown in Fig. 2. Here, we plot each parasite separately to improve clarity. Error bars represent ±1 SE.

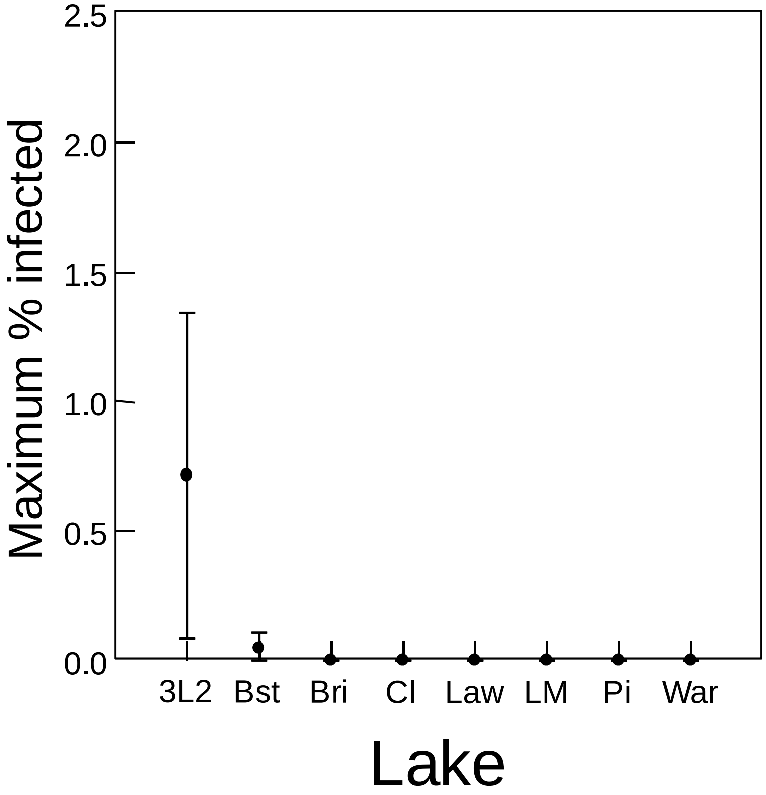

FIG. D21. Prevalence of *Pasteuria ramosa* infections in *D. pulicaria* in eight lake populations. Lake abbreviations are: “3L2” = Three Lakes Two, “Bst” = Bassett, “Bri” = Bristol, “Cl” = Cloverdale, “Law” = Lawrence, “LM” = Little Mill, “Pi” = Pine, “War” = Warner. These data are the same as those shown in Fig. 2. Here, we plot each parasite separately to improve clarity. Error bars represent ±1 SE.

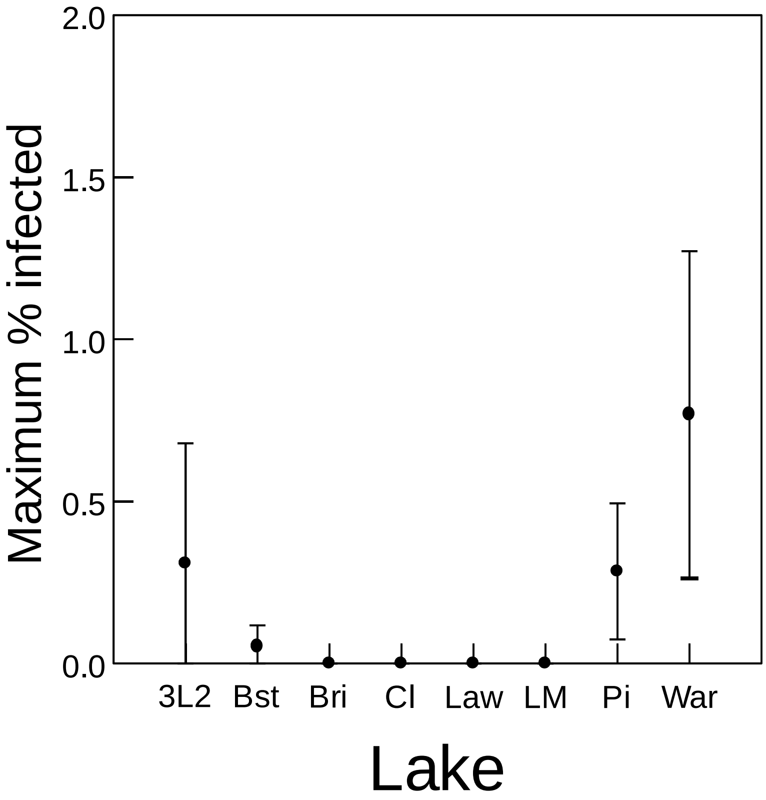

FIG. D22. Prevalence of oomycete infections in *D. pulicaria* in eight lake populations. Lake abbreviations are: “3L2” = Three Lakes Two, “Bst” = Bassett, “Bri” = Bristol, “Cl” = Cloverdale, “Law” = Lawrence, “LM” = Little Mill, “Pi” = Pine, “War” = Warner. These data are the same as those shown in Fig. 2. Here, we plot each parasite separately to improve clarity. Error bars represent ±1 SE.

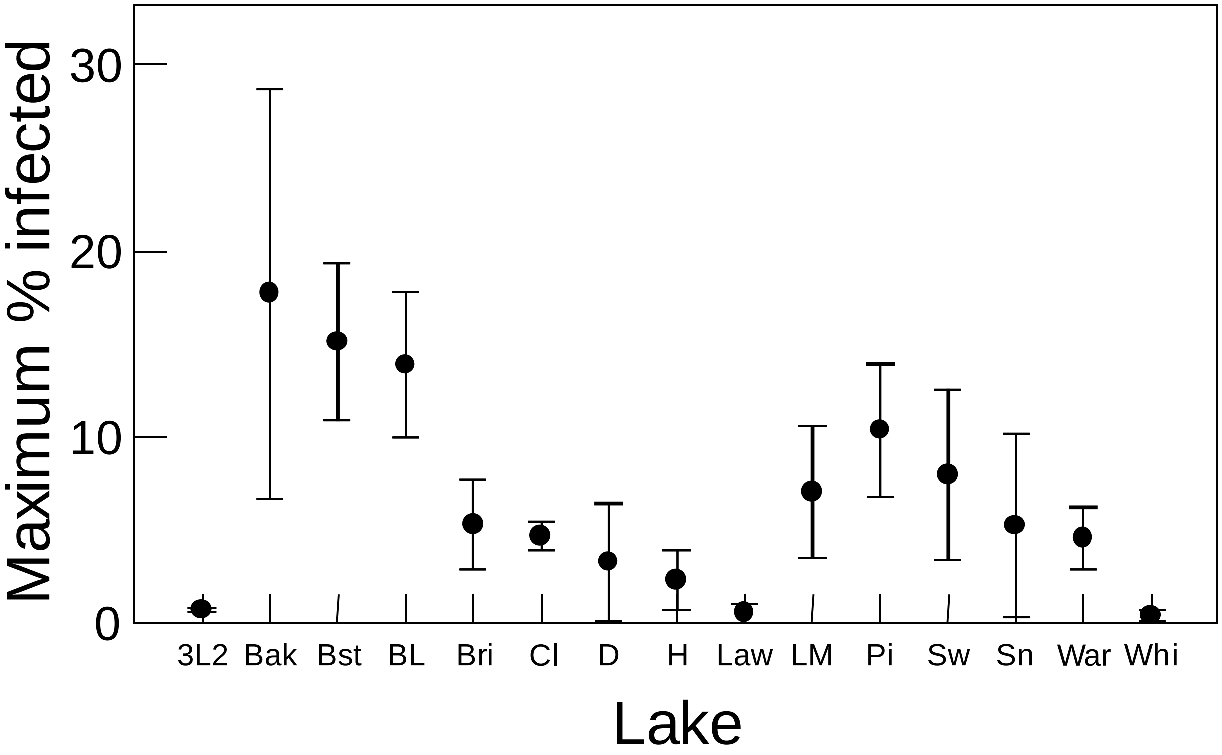

FIG. D23. Prevalence of brood parasite infections in *D. dentifera* in 15 lake populations. Lake abbreviations are: "3L2" = Three Lakes Two, "Bak" = Baker, "Bst" = Bassett, "BL" = "Big Long", "Bri" = Bristol, "Cl" = Cloverdale, "D" = Deep, "H" = Hall, "Law" = Lawrence, "LM" = Little Mill, "Pi" = Pine, "Sw" = Shaw, "Sn" = Sherman, "War" = Warner, and "Whi" = Whitford. These data are the same as those shown in the bottom panel of Fig. 2. Here, we plot each parasite separately to improve clarity. Error bars represent ±1 SE.

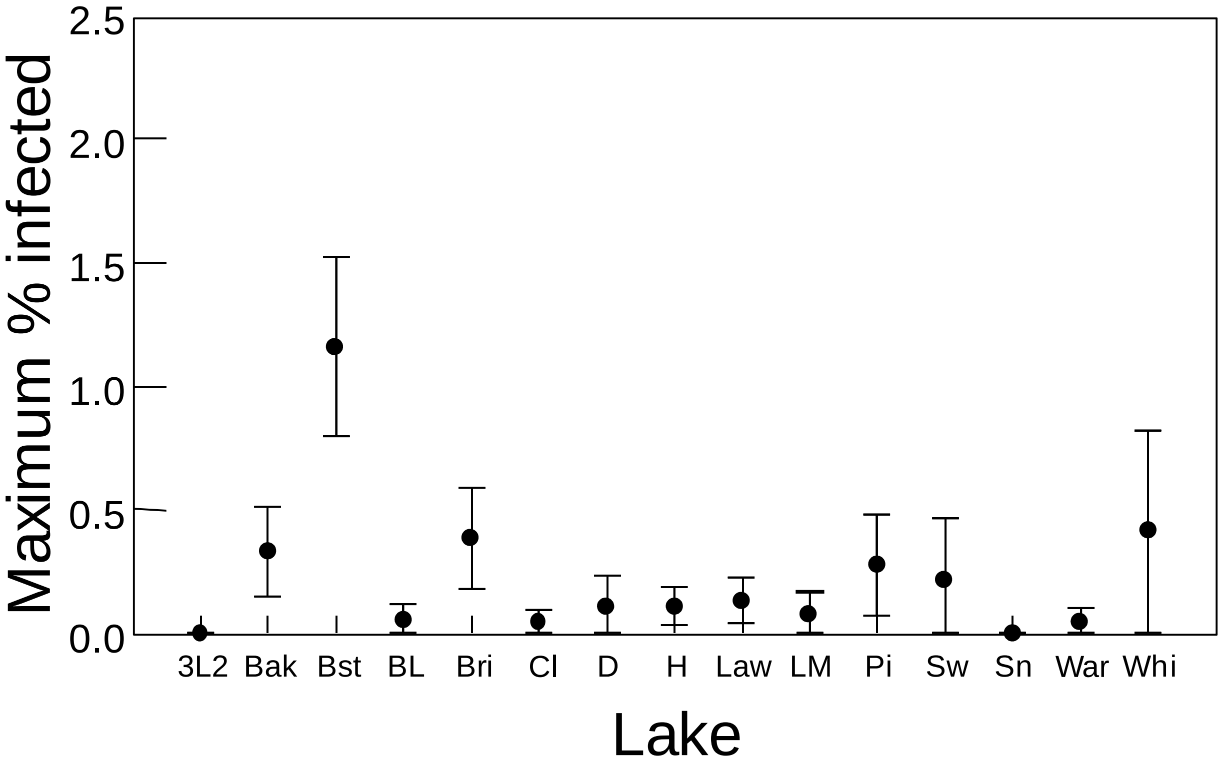

FIG. D24. Prevalence of *Larssonia obtusa* infections in *D. dentifera* in 15 lake populations. Lake abbreviations are: “3L2” = Three Lakes Two, “Bak” = Baker, “Bst” = Bassett, “BL” = “Big Long”, “Bri” = Bristol, “Cl” = Cloverdale, “D” = Deep, “H” = Hall, “Law” = Lawrence, “LM” = Little Mill, “Pi” = Pine, “Sw” = Shaw, “Sn” = Sherman, “War” = Warner, and “Whi” = Whitford. These data are the same as those shown in the bottom panel of Fig. 2. Here, we plot each parasite separately to improve clarity. Error bars represent ±1 SE.

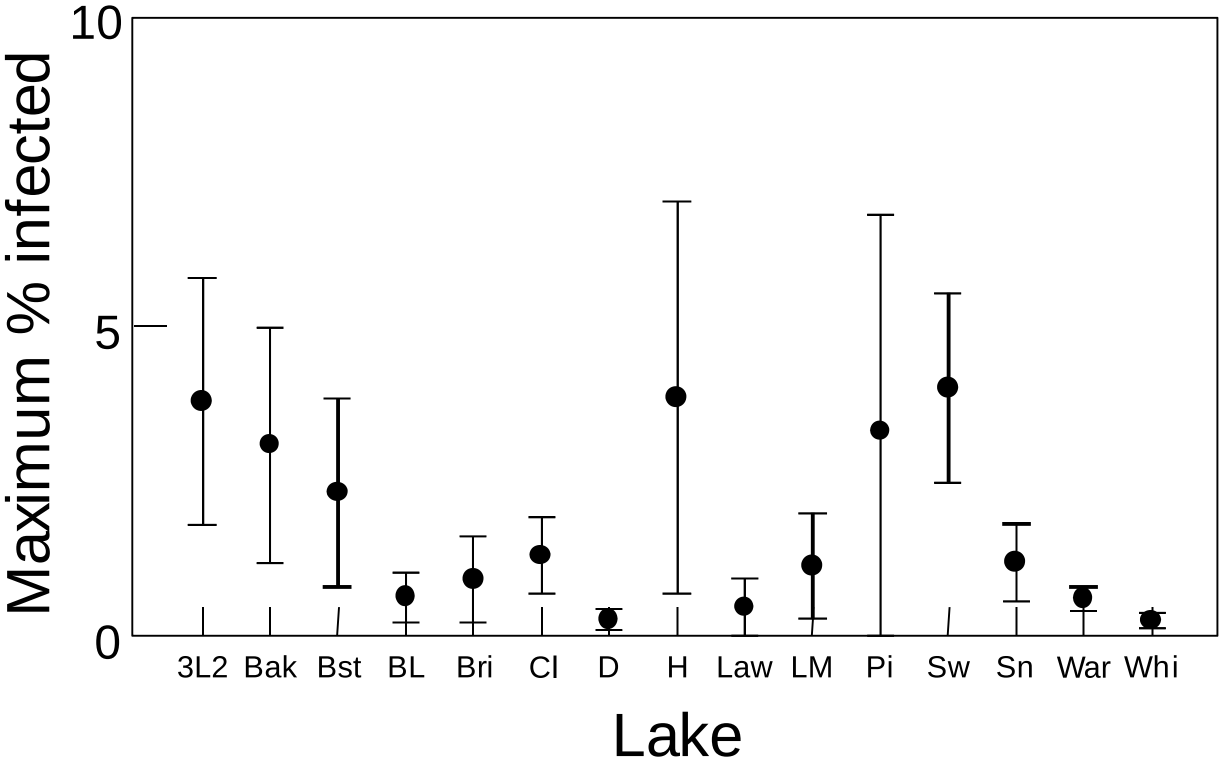

FIG. D25. Prevalence of *Spirobacillus cienkowskii* infections in *D. dentifera* in 15 lake populations. Lake abbreviations are: “3L2” = Three Lakes Two, “Bak” = Baker, “Bst” = Bassett, “BL” = “Big Long”, “Bri” = Bristol, “Cl” = Cloverdale, “D” = Deep, “H” = Hall, “Law” = Lawrence, “LM” = Little Mill, “Pi” = Pine, “Sw” = Shaw, “Sn” = Sherman, “War” = Warner, and “Whi” = Whitford. These data are the same as those shown in the bottom panel of Fig. 2. Here, we plot each parasite separately to improve clarity. Error bars represent ±1 SE.

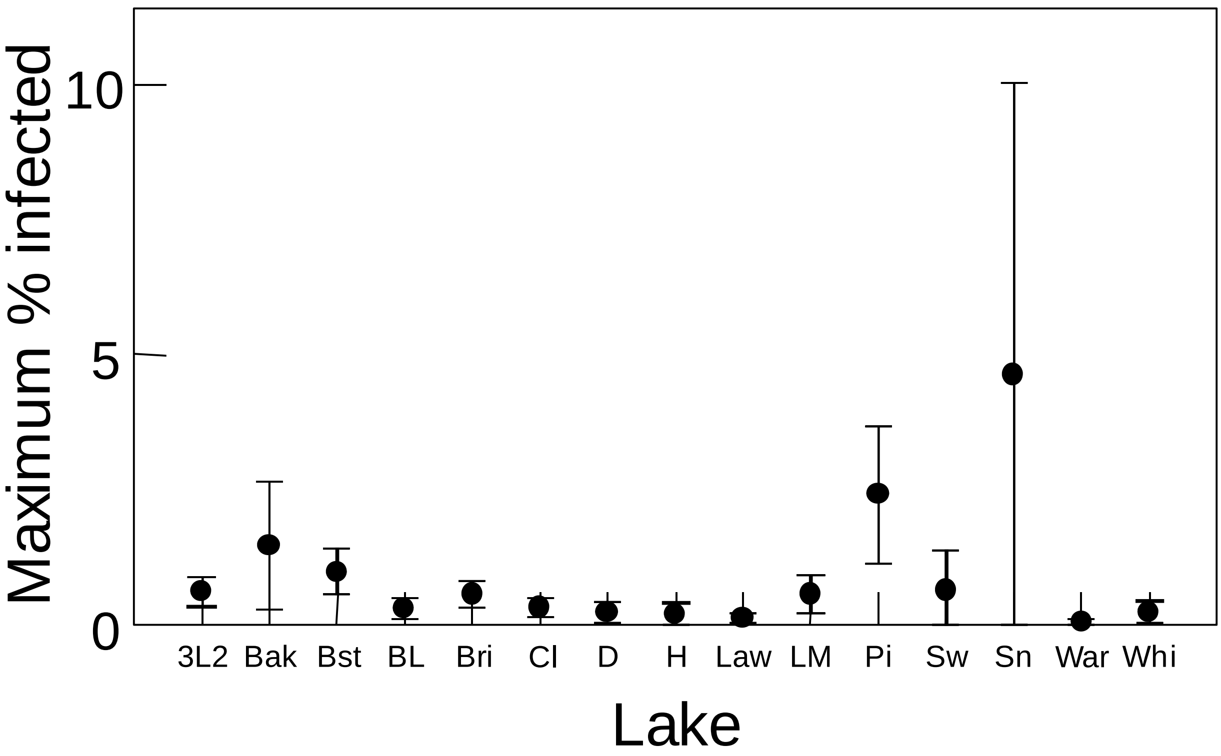

FIG. D26. Prevalence of *Pasteuria ramosa* infections in *D. dentifera* in 15 lake populations. Lake abbreviations are: “3L2” = Three Lakes Two, “Bak” = Baker, “Bst” = Bassett, “BL” = “Big Long”, “Bri” = Bristol, “Cl” = Cloverdale, “D” = Deep, “H” = Hall, “Law” = Lawrence, “LM” = Little Mill, “Pi” = Pine, “Sw” = Shaw, “Sn” = Sherman, “War” = Warner, and “Whi” = Whitford. These data are the same as those shown in the bottom panel of Fig. 2. Here, we plot each parasite separately to improve clarity. Error bars represent ±1 SE.

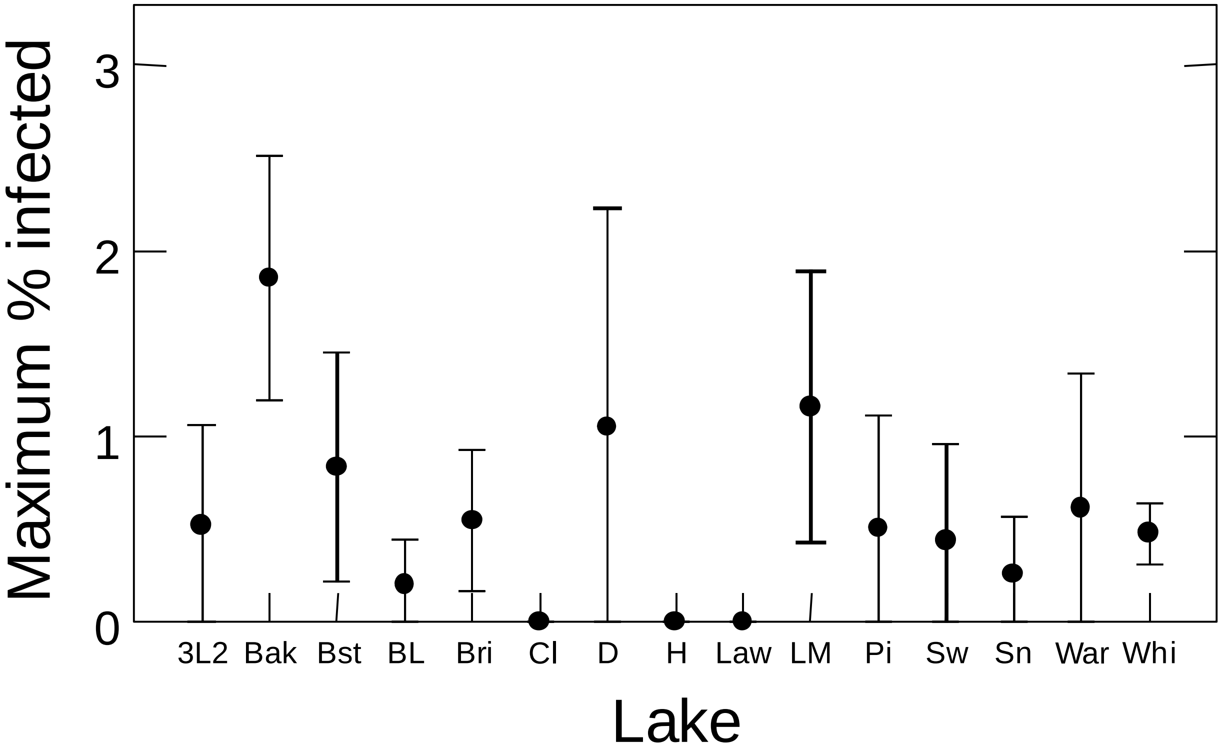

FIG. D27. Prevalence of oomycete infections in *D. dentifera* in 15 lake populations. Lake abbreviations are: “3L2” = Three Lakes Two, “Bak” = Baker, “Bst” = Bassett, “BL” = “Big Long”, “Bri” = Bristol, “Cl” = Cloverdale, “D” = Deep, “H” = Hall, “Law” = Lawrence, “LM” = Little Mill, “Pi” = Pine, “Sw” = Shaw, “Sn” = Sherman, “War” = Warner, and “Whi” = Whitford. These data are the same as those shown in the bottom panel of Fig. 2. Here, we plot each parasite separately to improve clarity. Error bars represent ±1 SE.

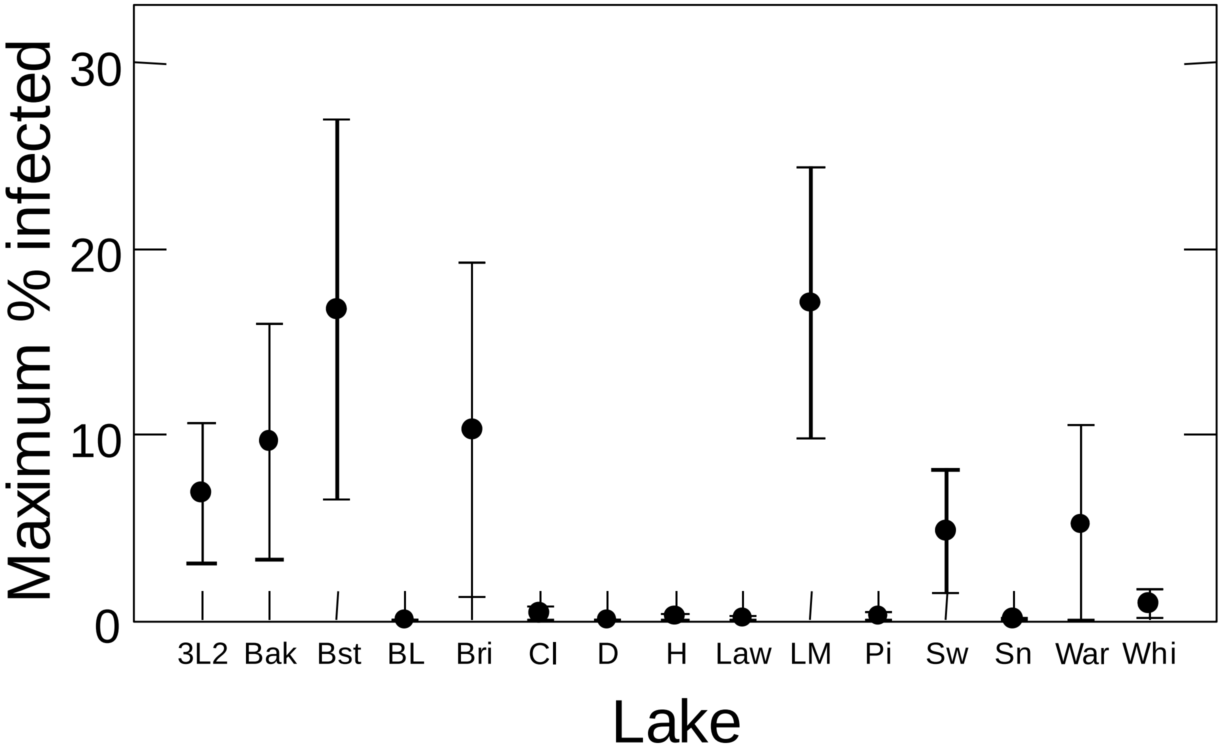

FIG. D28. Prevalence of *Metschnikowia bicuspidata* infections in *D. dentifera* in 15 lake populations. Lake abbreviations are: “3L2” = Three Lakes Two, “Bak” = Baker, “Bst” = Bassett, “BL” = “Big Long”, “Bri” = Bristol, “Cl” = Cloverdale, “D” = Deep, “H” = Hall, “Law” = Lawrence, “LM” = Little Mill, “Pi” = Pine, “Sw” = Shaw, “Sn” = Sherman, “War” = Warner, and “Whi” = Whitford. These data are the same as those shown in the bottom panel of Fig. 2. Here, we plot each parasite separately to improve clarity. Error bars represent ±1 SE.

Ecological Archives E091-234-A4

Meghan A. Duffy, Carla E. Cáceres, Spencer R. Hall, Alan J. Tessier, and Anthony R. Ives. 2010. Temporal, spatial, and between-host comparisons of patterns of parasitism in lake zooplankton. *Ecology* 91:3322–3331.

Appendix D. Infection prevalences for individual host-parasite species pairings.

In Figs. D1–D8, we present the individual box plots for each of the eight parasites found infecting *D. pulicaria* in the eight lake populations. These figures contain the data from Fig. 1A; here, the data for each parasite are plotted separately to make the distributions of each parasite species clearer. In Figs. D9–D14, we present the individual box plots for each of the six parasites found infecting *D. dentifera* in the 15 lake populations. These figures contain the data from Fig. 1B; here, the data for each parasite are plotted separately to make the distributions of each parasite species clearer.

In Figs. D15–D22, we present the data for each of the eight parasites found infecting *D. pulicaria* in the eight lake populations. These figures contain the data from the top panel of Fig. 2; here, the data for each parasite are plotted separately to show the differences among lakes more clearly. In Figs. D23–D28, we present the data for each of the six parasites found infecting *D. dentifera* in the fifteen lake populations. These figures contain the data from the bottom panel of Fig. 2.

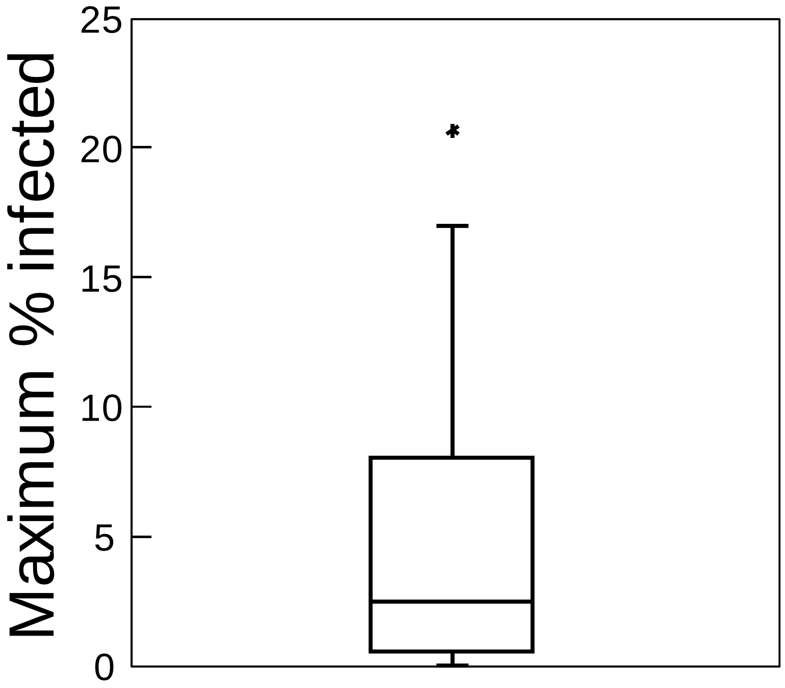

FIG. D1. Box plot of maximum infection prevalences of *Polycaryum laeve* in *D. pulicaria* in eight lakes. This figure contains data from Fig. 1A.

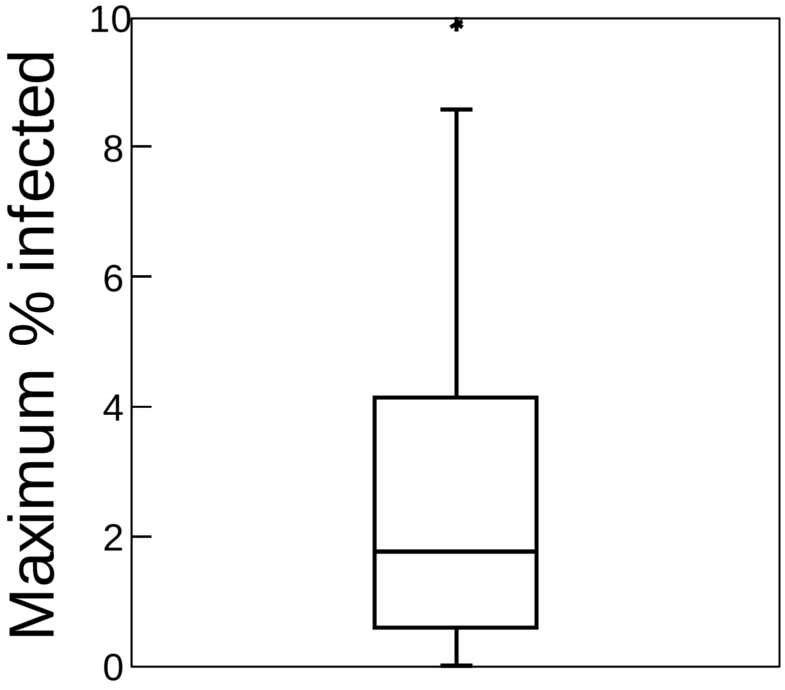

FIG. D2. Box plot of maximum infection prevalences of the Burkholderia-type bacterium (“BB”) in *D. pulicaria* in eight lakes. This figure contains data from Fig. 1A.

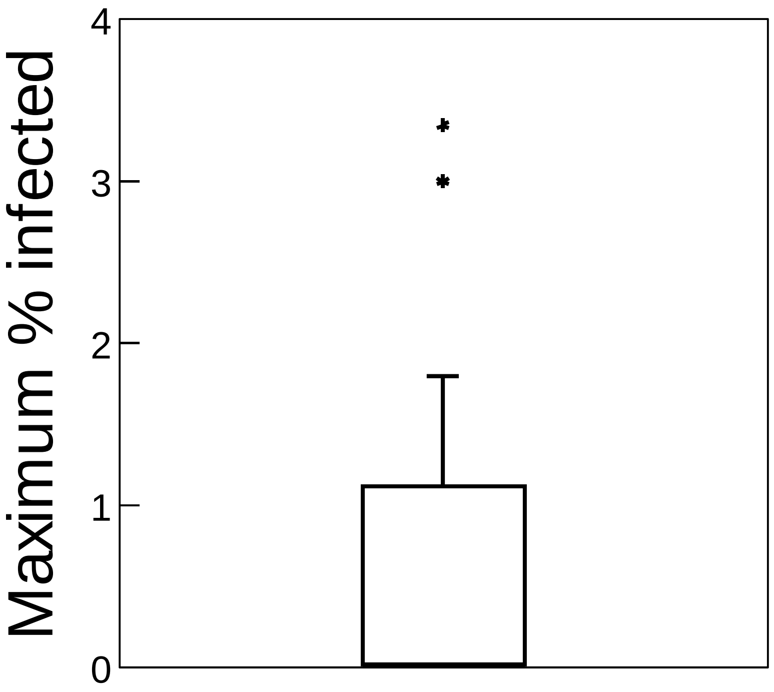

FIG. D3. Box plot of maximum infection prevalences of the brood parasite in *D. pulicaria* in eight lakes. This figure contains data from Fig. 1A.

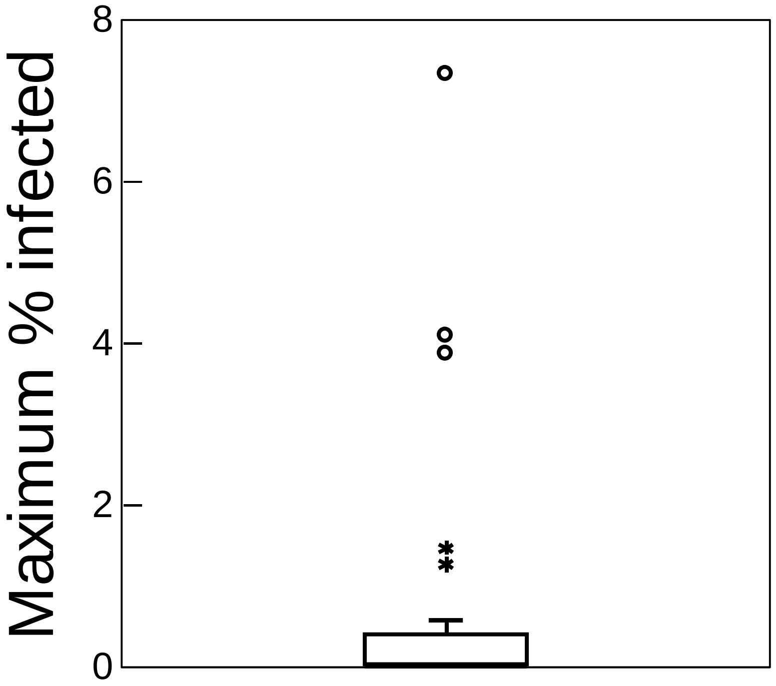

FIG. D4. Box plot of maximum infection prevalences of *Gurleya sp.* in *D. pulicaria* in eight lakes. This figure contains data from Fig. 1A.

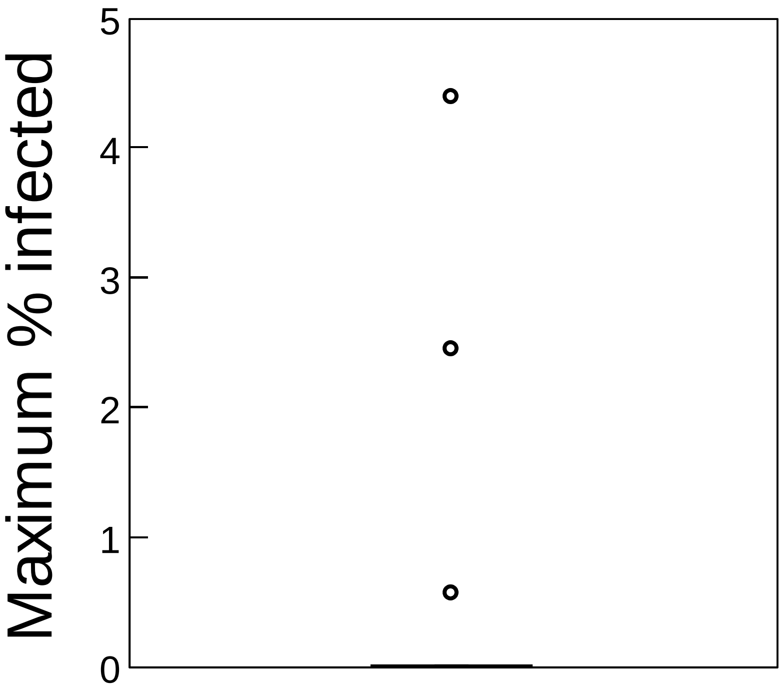

FIG. D5. Box plot of maximum infection prevalences of *Larssonia obtusa* in *D. pulicaria* in eight lakes. This figure contains data from Fig. 1A.

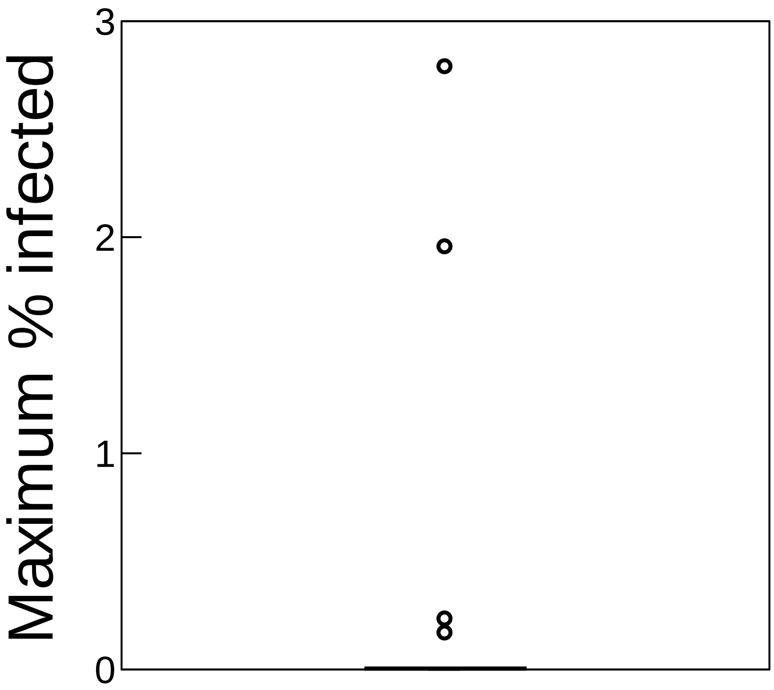

FIG. D6. Box plot of maximum infection prevalences of *Spirobacillus cienkowskii* in *D. pulicaria* in eight lakes. This figure contains data from Fig. 1A.

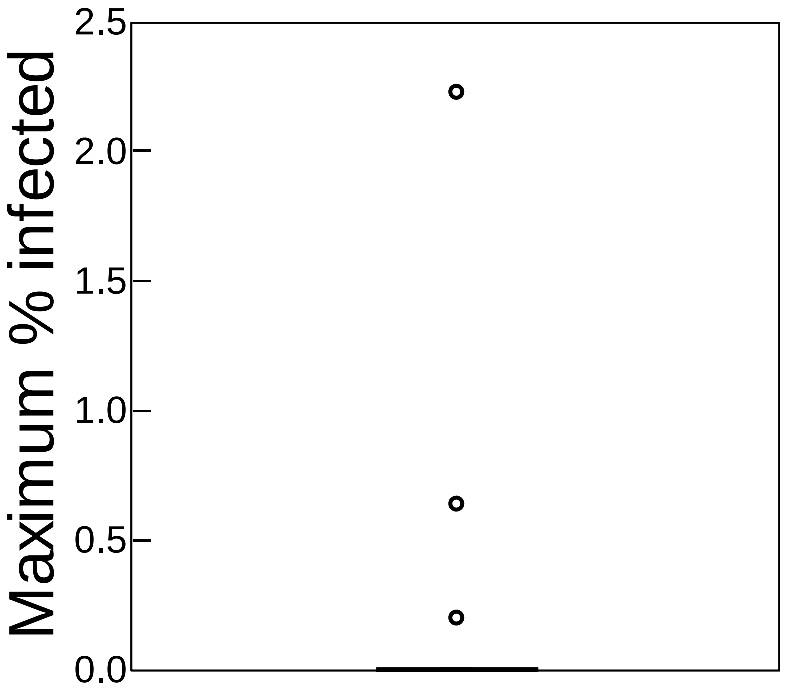

FIG. D7. Box plot of maximum infection prevalences of *Pasteuria ramosa* in *D. pulicaria* in eight lakes. This figure contains data from Fig. 1A.

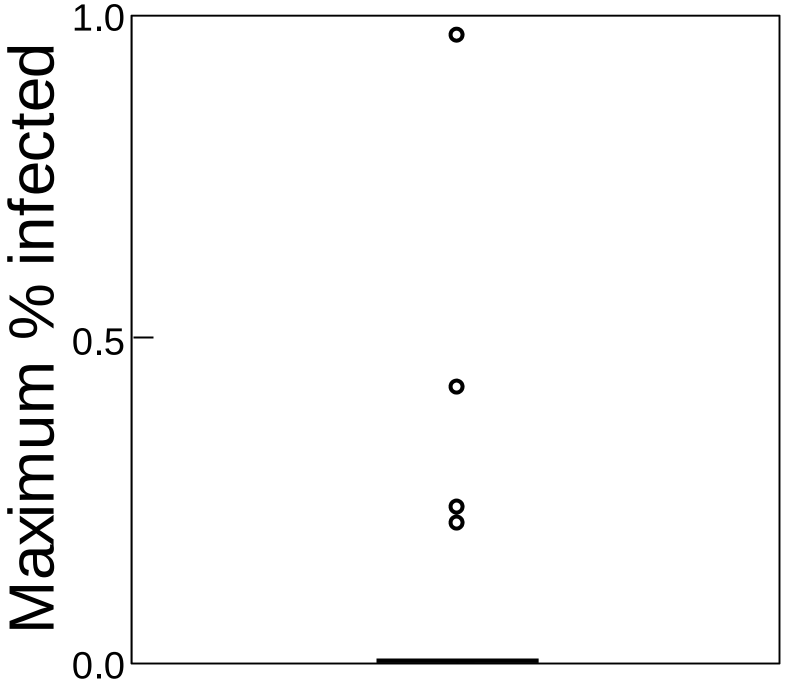

FIG. D8. Box plot of maximum infection prevalences of oomycete parasites in *D. pulicaria* in eight lakes. This figure contains data from Fig. 1A.

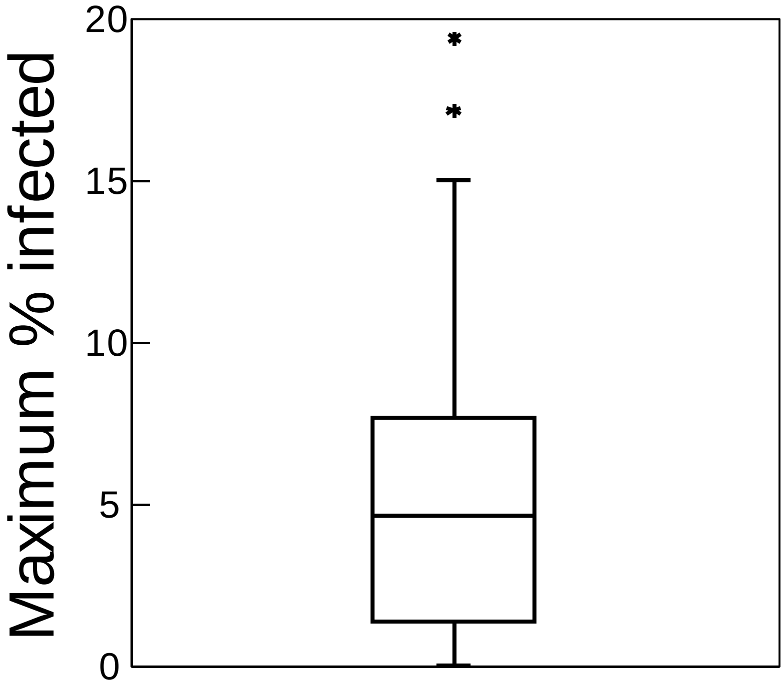

FIG. D9. Box plot of maximum infection prevalences of the brood parasite in *D. dentifera* in 15 lakes. This figure contains data from Fig. 1B.

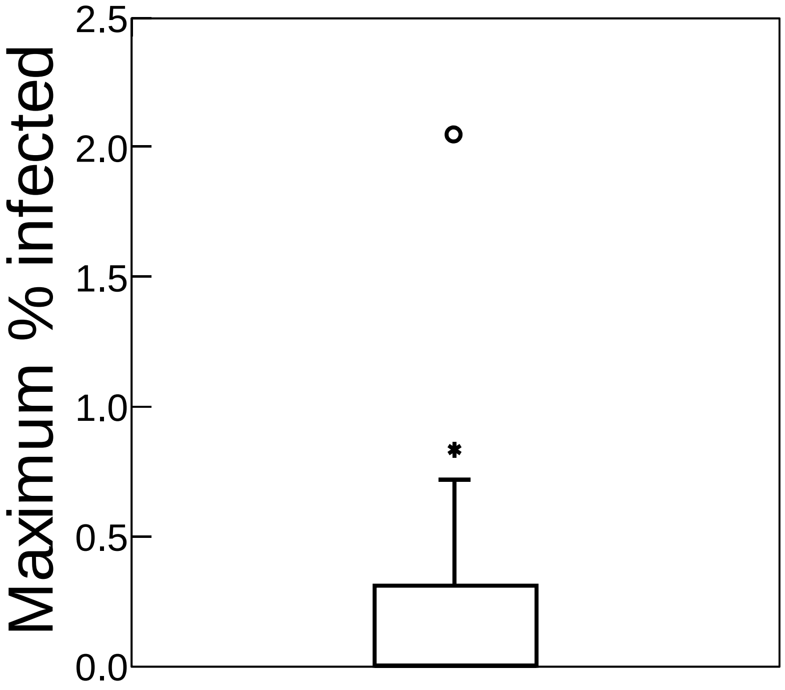

FIG. D10. Box plot of maximum infection prevalences of *Larssonia obtusa* in *D. dentifera* in 15 lakes. This figure contains data from Fig. 1B.

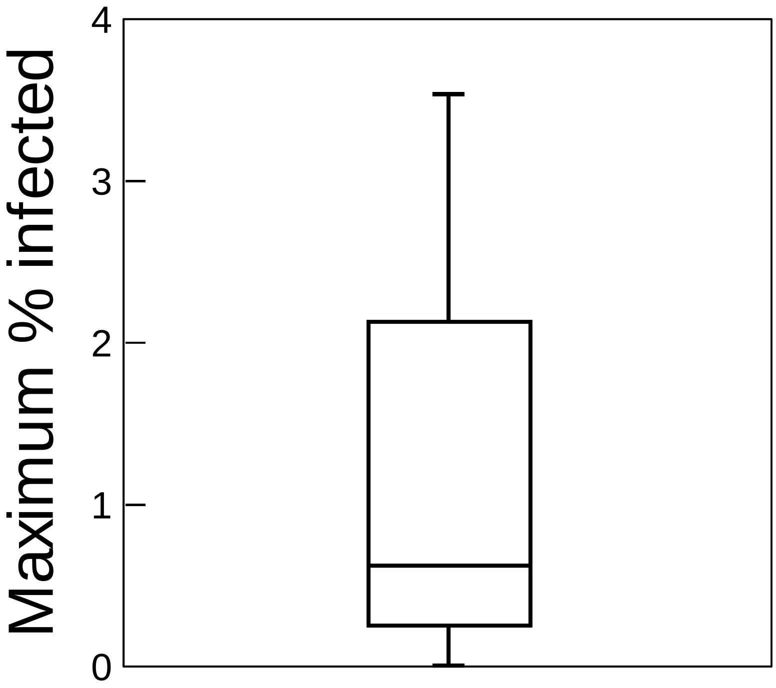

FIG. D11. Box plot of maximum infection prevalences of *Spirobacillus cienkowskii* in *D. dentifera* in 15 lakes. This figure contains data from Fig. 1B.

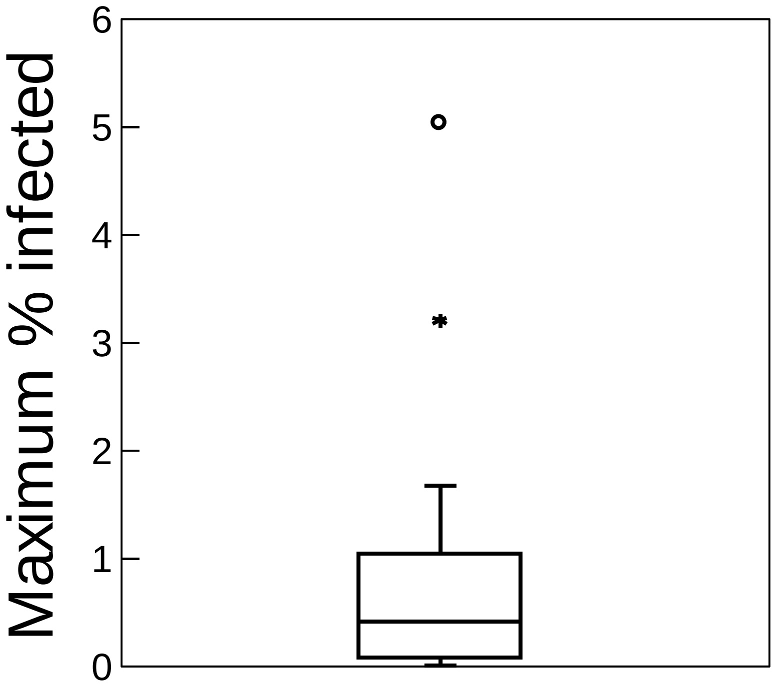

FIG. D12. Box plot of maximum infection prevalences of *Pasteuria ramosa* in *D. dentifera* in 15 lakes. This figure contains data from Fig. 1B.

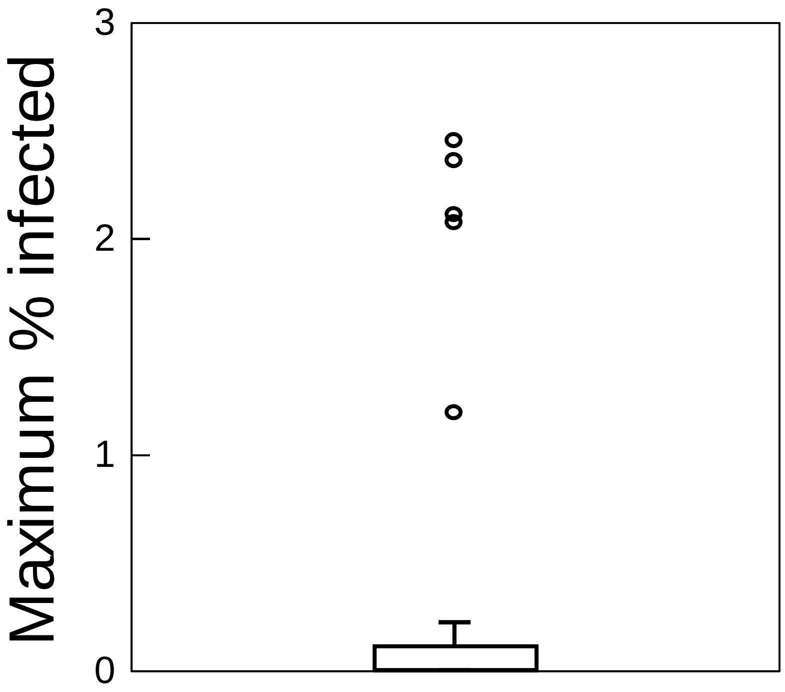

FIG. D13. Box plot of maximum infection prevalences of oomycete parasites in *D. dentifera* in 15 lakes. This figure contains data from Fig. 1B.

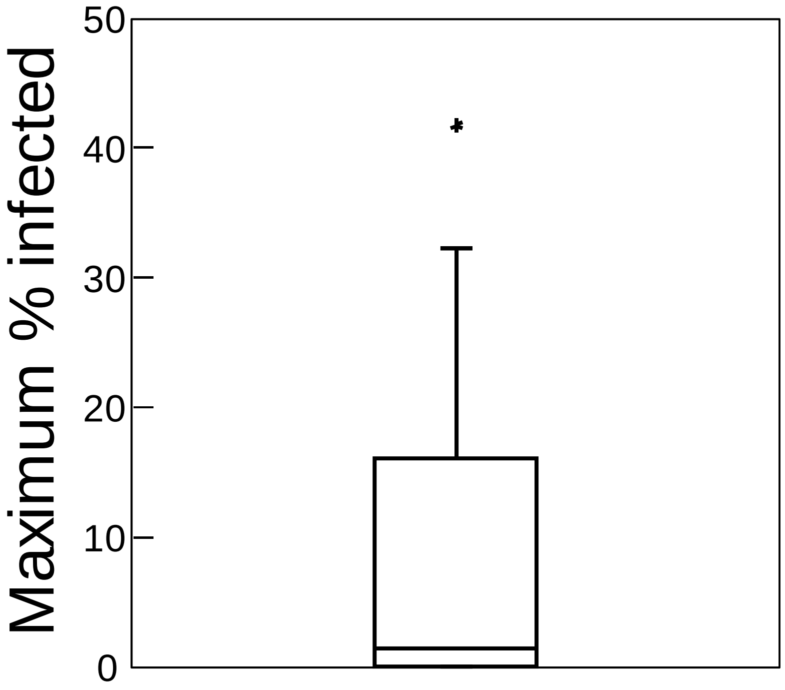

FIG. D14. Box plot of maximum infection prevalences of *Metschnikowia bicuspidata* in *D. dentifera* in 15 lakes. This figure contains data from Fig. 1B.

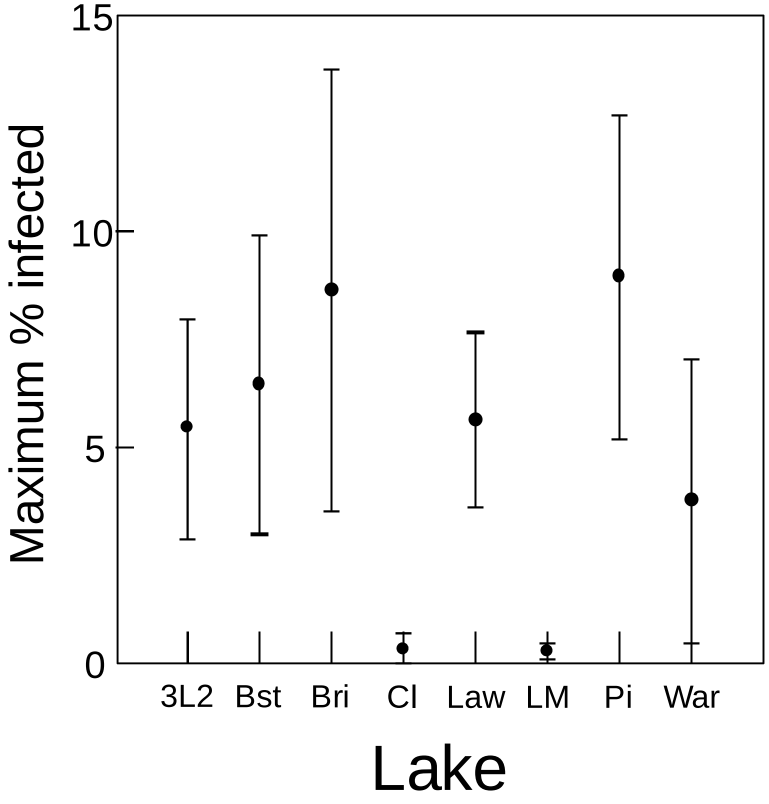

FIG. D15. Prevalence of *Polycaryum laeve* infections in *D. pulicaria* in eight lake populations. Lake abbreviations are: “3L2” = Three Lakes Two, “Bst” = Bassett, “Bri” = Bristol, “Cl” = Cloverdale, “Law” = Lawrence, “LM” = Little Mill, “Pi” = Pine, “War” = Warner. These data are the same as those shown in Fig. 2. Here, we plot each parasite separately to improve clarity. Error bars represent ±1 SE.

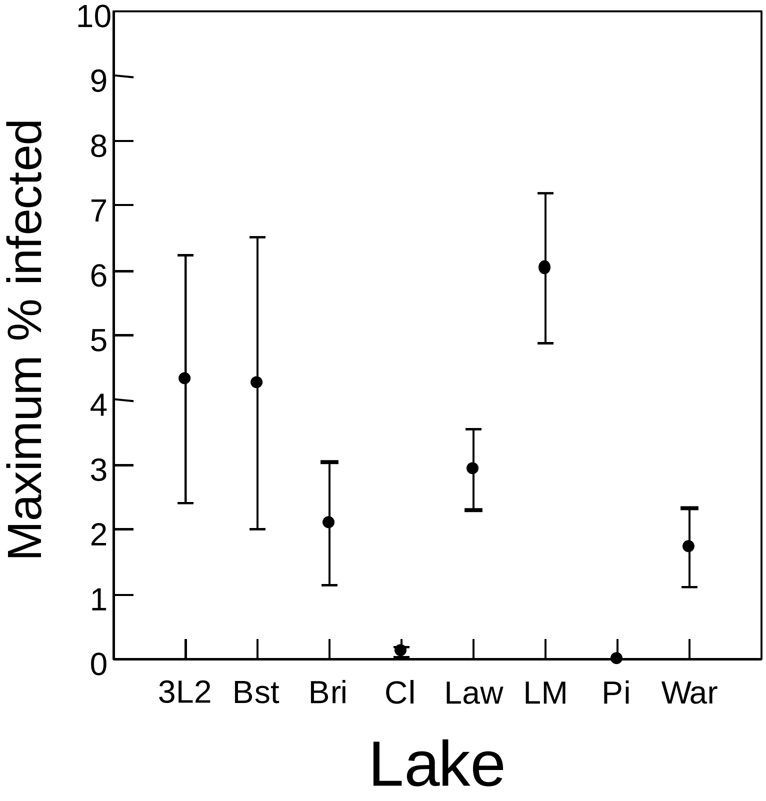

FIG. D16. Prevalence of “BB” (Burkholderia-type bacterium) infections in *D. pulicaria* in eight lake populations. Lake abbreviations are: “3L2” = Three Lakes Two, “Bst” = Bassett, “Bri” = Bristol, “Cl” = Cloverdale, “Law” = Lawrence, “LM” = Little Mill, “Pi” = Pine, “War” = Warner. These data are the same as those shown in Fig. 2. Here, we plot each parasite separately to improve clarity. Error bars represent ±1 SE.

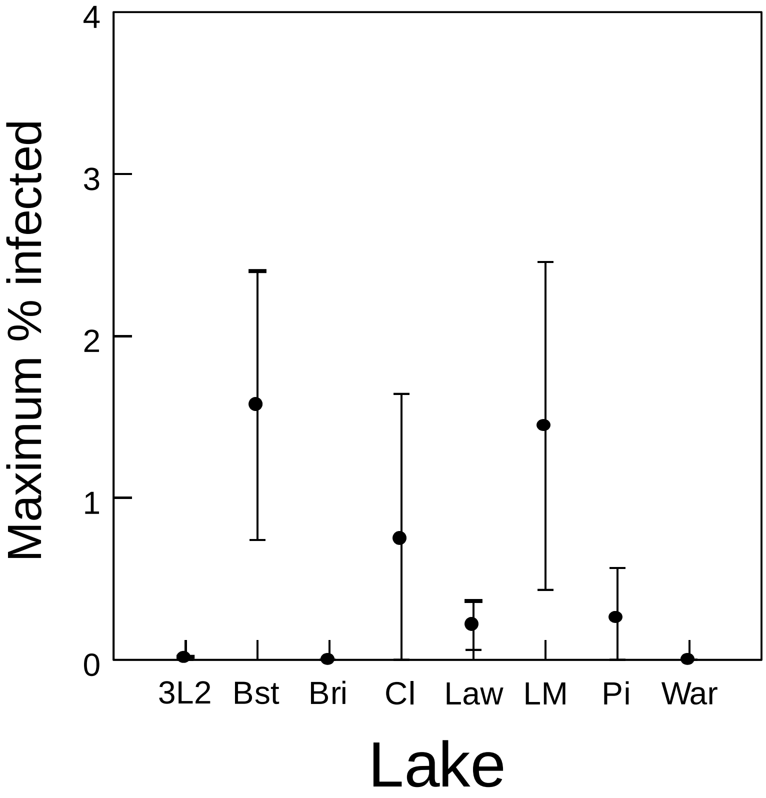

FIG. D17. Prevalence of brood parasite infections in *D. pulicaria* in eight lake populations. Lake abbreviations are: “3L2” = Three Lakes Two, “Bst” = Bassett, “Bri” = Bristol, “Cl” = Cloverdale, “Law” = Lawrence, “LM” = Little Mill, “Pi” = Pine, “War” = Warner. These data are the same as those shown in Fig. 2. Here, we plot each parasite separately to improve clarity. Error bars represent ±1 SE.

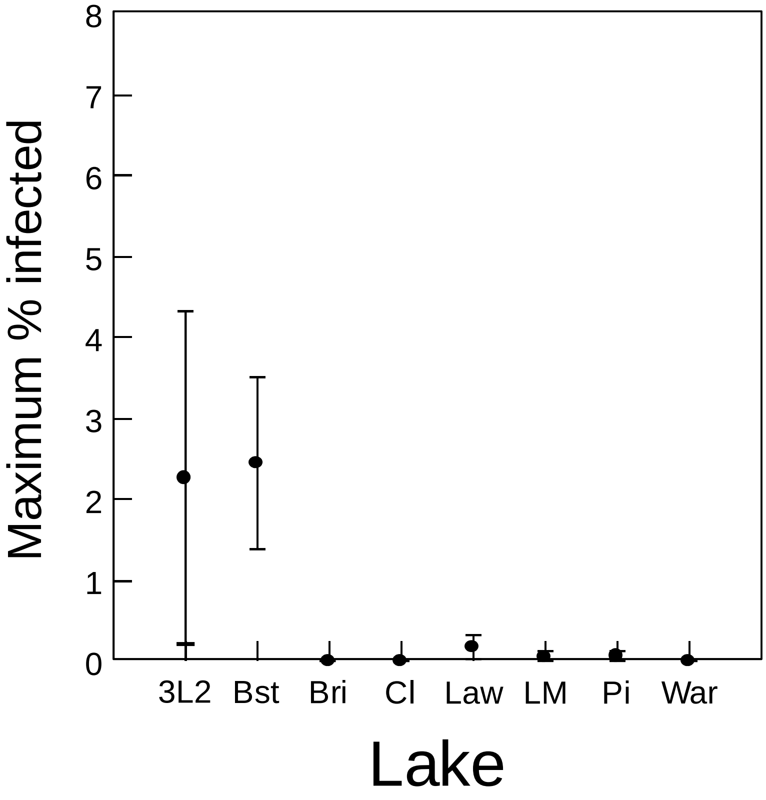

FIG. D18. Prevalence of *Gurleya sp.* infections in *D. pulicaria* in eight lake populations. Lake abbreviations are: "3L2" = Three Lakes Two, "Bst" = Bassett, "Bri" = Bristol, "Cl" = Cloverdale, "Law" = Lawrence, "LM" = Little Mill, "Pi" = Pine, "War" = Warner. These data are the same as those shown in Fig. 2. Here, we plot each parasite separately to improve clarity. Error bars represent ±1 SE.

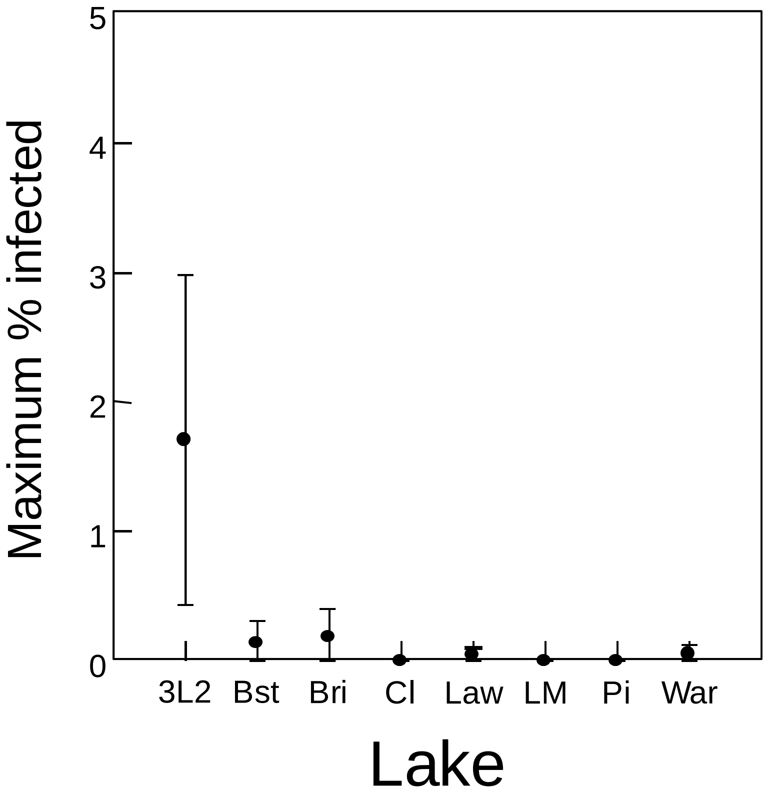

FIG. D19. Prevalence of *Larssonia obtusa* infections in *D. pulicaria* in eight lake populations. Lake abbreviations are: “3L2” = Three Lakes Two, “Bst” = Bassett, “Bri” = Bristol, “Cl” = Cloverdale, “Law” = Lawrence, “LM” = Little Mill, “Pi” = Pine, “War” = Warner. These data are the same as those shown in Fig. 2. Here, we plot each parasite separately to improve clarity. Error bars represent ±1 SE.

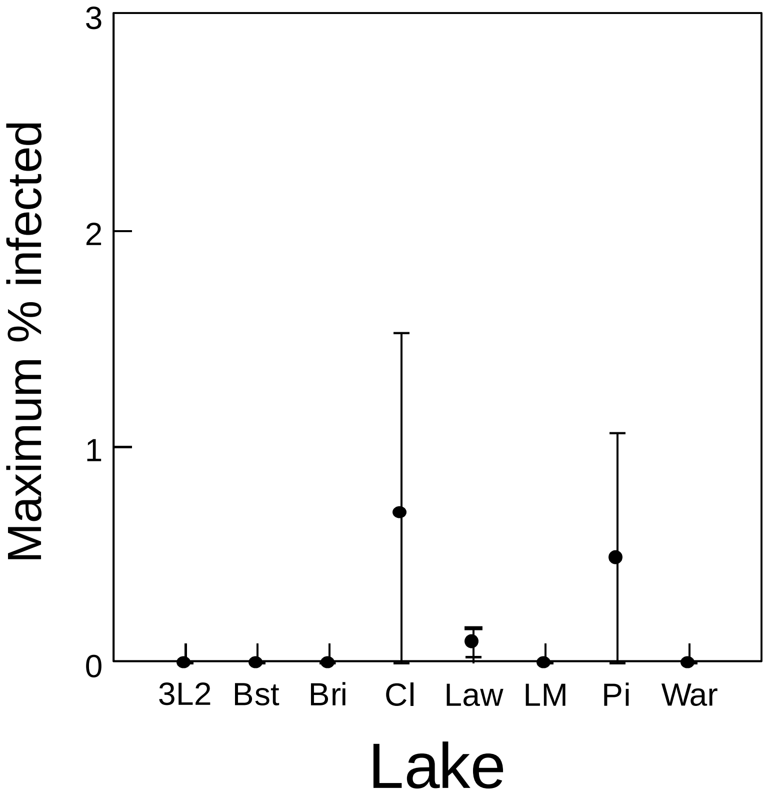

FIG. D20. Prevalence of *Spirobacillus cienkowskii* infections in *D. pulicaria* in eight lake populations. Lake abbreviations are: “3L2” = Three Lakes Two, “Bst” = Bassett, “Bri” = Bristol, “Cl” = Cloverdale, “Law” = Lawrence, “LM” = Little Mill, “Pi” = Pine, “War” = Warner. These data are the same as those shown in Fig. 2. Here, we plot each parasite separately to improve clarity. Error bars represent ±1 SE.

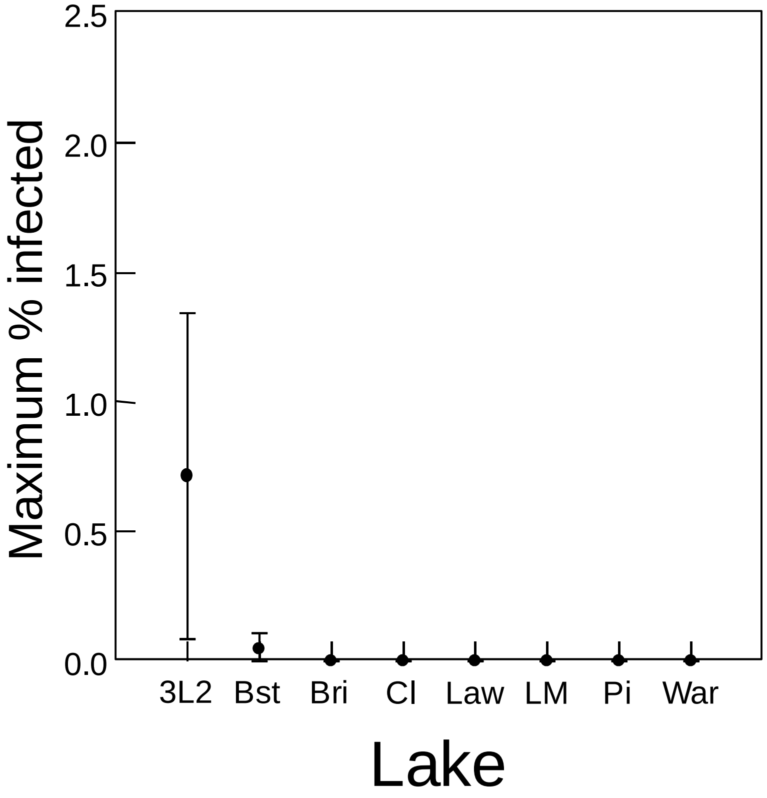

FIG. D21. Prevalence of *Pasteuria ramosa* infections in *D. pulicaria* in eight lake populations. Lake abbreviations are: “3L2” = Three Lakes Two, “Bst” = Bassett, “Bri” = Bristol, “Cl” = Cloverdale, “Law” = Lawrence, “LM” = Little Mill, “Pi” = Pine, “War” = Warner. These data are the same as those shown in Fig. 2. Here, we plot each parasite separately to improve clarity. Error bars represent ±1 SE.

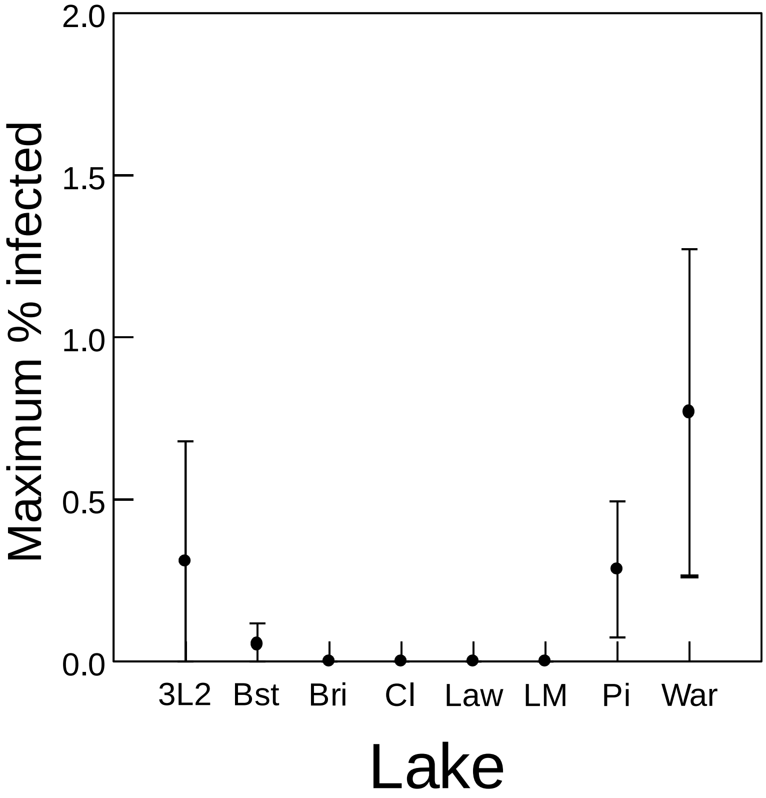

FIG. D22. Prevalence of oomycete infections in *D. pulicaria* in eight lake populations. Lake abbreviations are: “3L2” = Three Lakes Two, “Bst” = Bassett, “Bri” = Bristol, “Cl” = Cloverdale, “Law” = Lawrence, “LM” = Little Mill, “Pi” = Pine, “War” = Warner. These data are the same as those shown in Fig. 2. Here, we plot each parasite separately to improve clarity. Error bars represent ±1 SE.

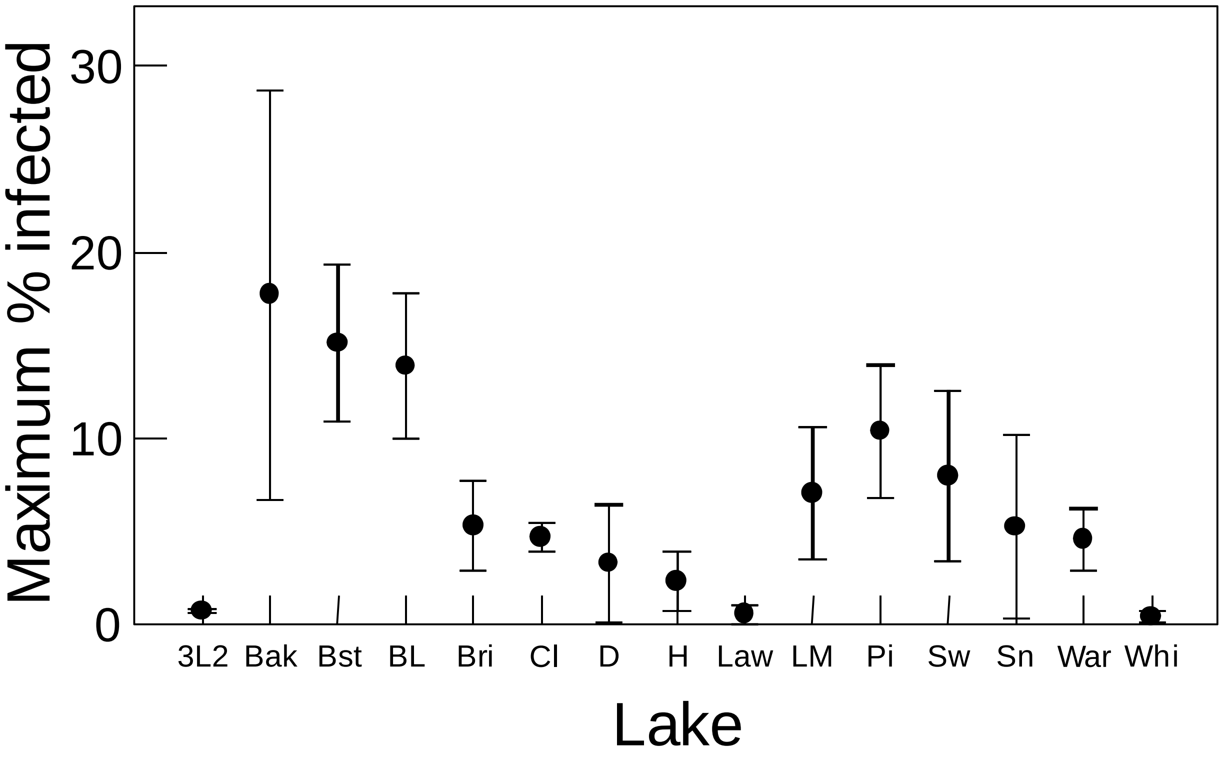

FIG. D23. Prevalence of brood parasite infections in *D. dentifera* in 15 lake populations. Lake abbreviations are: "3L2" = Three Lakes Two, "Bak" = Baker, "Bst" = Bassett, "BL" = "Big Long", "Bri" = Bristol, "Cl" = Cloverdale, "D" = Deep, "H" = Hall, "Law" = Lawrence, "LM" = Little Mill, "Pi" = Pine, "Sw" = Shaw, "Sn" = Sherman, "War" = Warner, and "Whi" = Whitford. These data are the same as those shown in the bottom panel of Fig. 2. Here, we plot each parasite separately to improve clarity. Error bars represent ±1 SE.

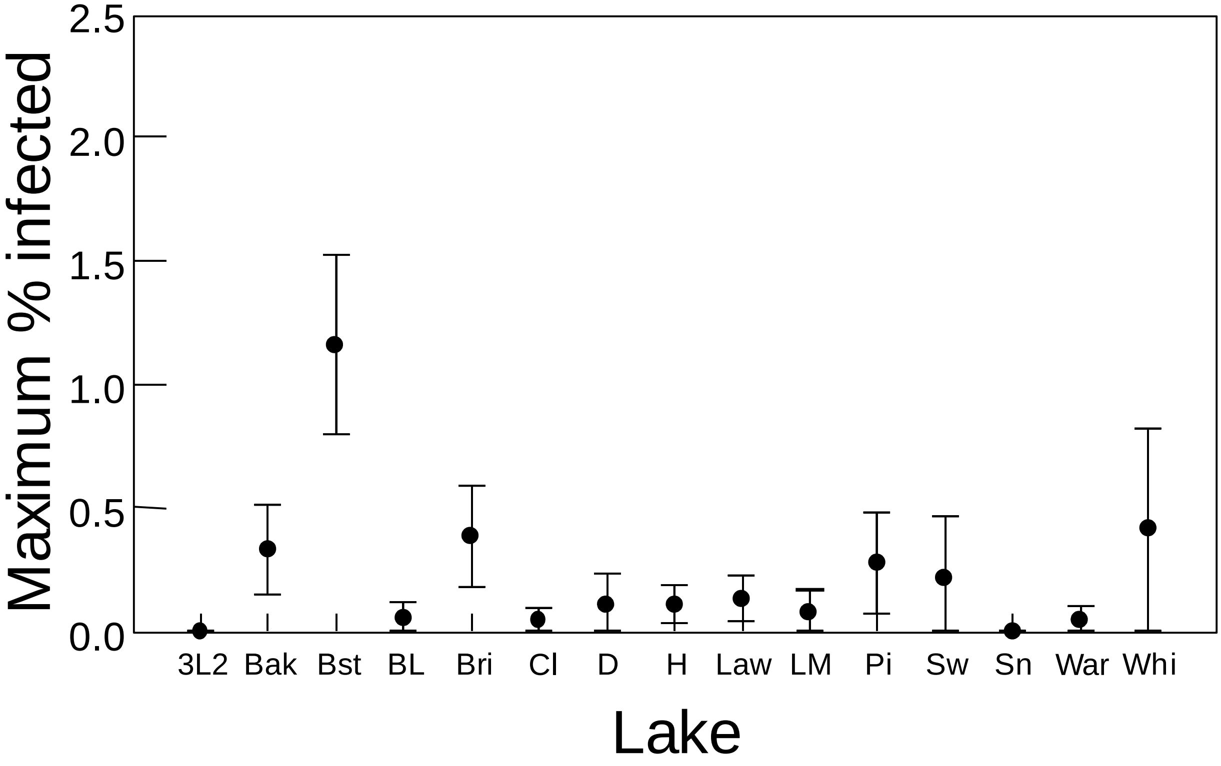

FIG. D24. Prevalence of *Larssonia obtusa* infections in *D. dentifera* in 15 lake populations. Lake abbreviations are: "3L2" = Three Lakes Two, "Bak" = Baker, "Bst" = Bassett, "BL" = "Big Long", "Bri" = Bristol, "Cl" = Cloverdale, "D" = Deep, "H" = Hall, "Law" = Lawrence, "LM" = Little Mill, "Pi" = Pine, "Sw" = Shaw, "Sn" = Sherman, "War" = Warner, and "Whi" = Whitford. These data are the same as those shown in the bottom panel of Fig. 2. Here, we plot each parasite separately to improve clarity. Error bars represent ±1 SE.

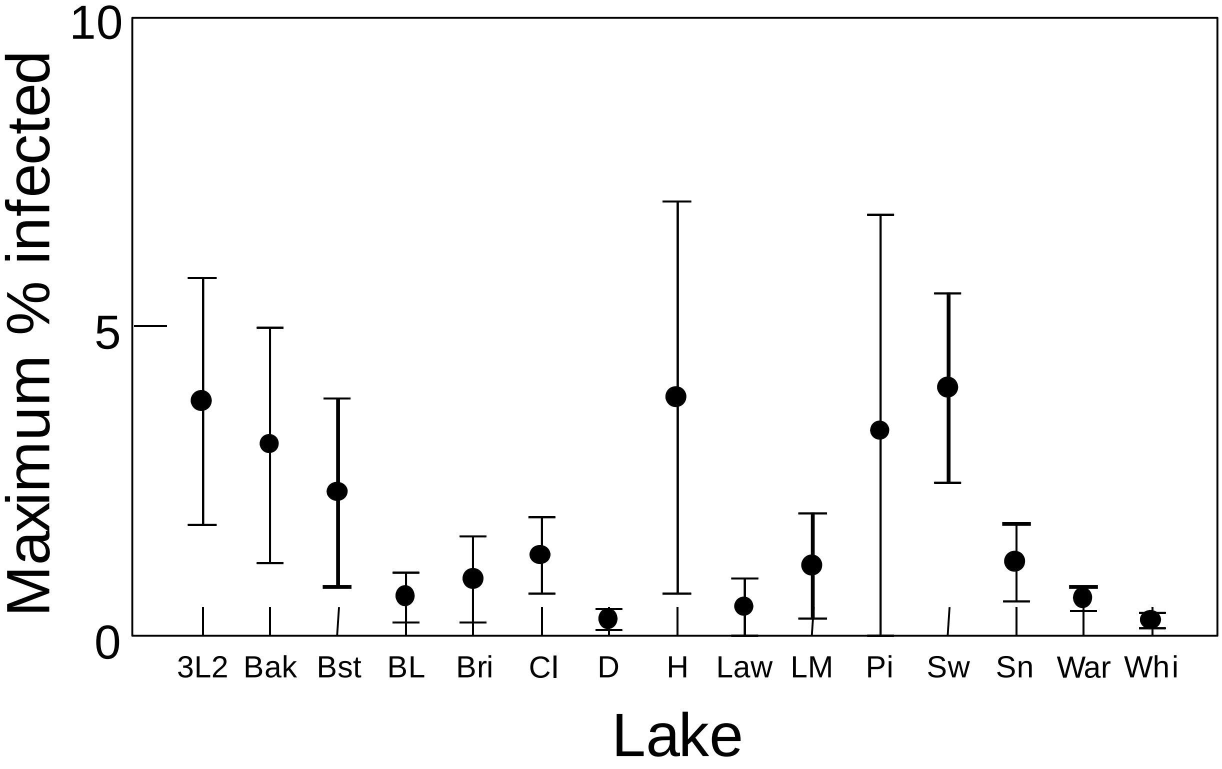

FIG. D25. Prevalence of *Spirobacillus cienkowskii* infections in *D. dentifera* in 15 lake populations. Lake abbreviations are: “3L2” = Three Lakes Two, “Bak” = Baker, “Bst” = Bassett, “BL” = “Big Long”, “Bri” = Bristol, “Cl” = Cloverdale, “D” = Deep, “H” = Hall, “Law” = Lawrence, “LM” = Little Mill, “Pi” = Pine, “Sw” = Shaw, “Sn” = Sherman, “War” = Warner, and “Whi” = Whitford. These data are the same as those shown in the bottom panel of Fig. 2. Here, we plot each parasite separately to improve clarity. Error bars represent ±1 SE.

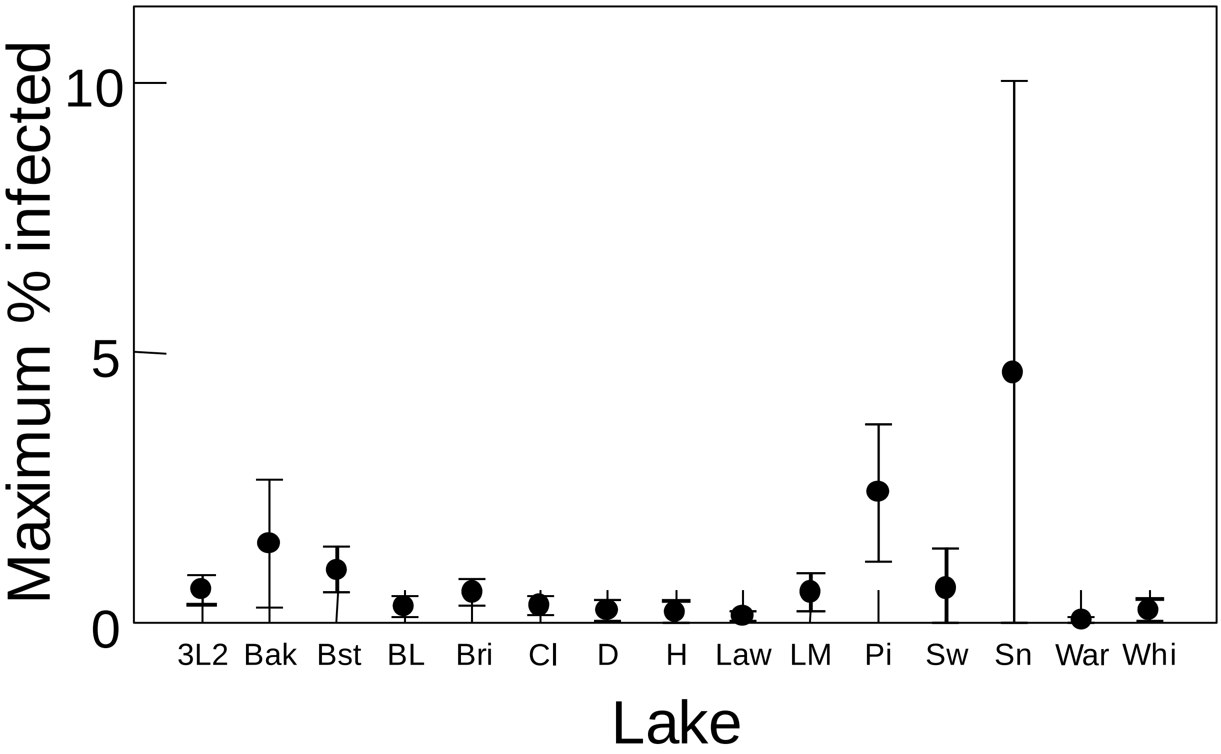

FIG. D26. Prevalence of *Pasteuria ramosa* infections in *D. dentifera* in 15 lake populations. Lake abbreviations are: “3L2” = Three Lakes Two, “Bak” = Baker, “Bst” = Bassett, “BL” = “Big Long”, “Bri” = Bristol, “Cl” = Cloverdale, “D” = Deep, “H” = Hall, “Law” = Lawrence, “LM” = Little Mill, “Pi” = Pine, “Sw” = Shaw, “Sn” = Sherman, “War” = Warner, and “Whi” = Whitford. These data are the same as those shown in the bottom panel of Fig. 2. Here, we plot each parasite separately to improve clarity. Error bars represent ±1 SE.

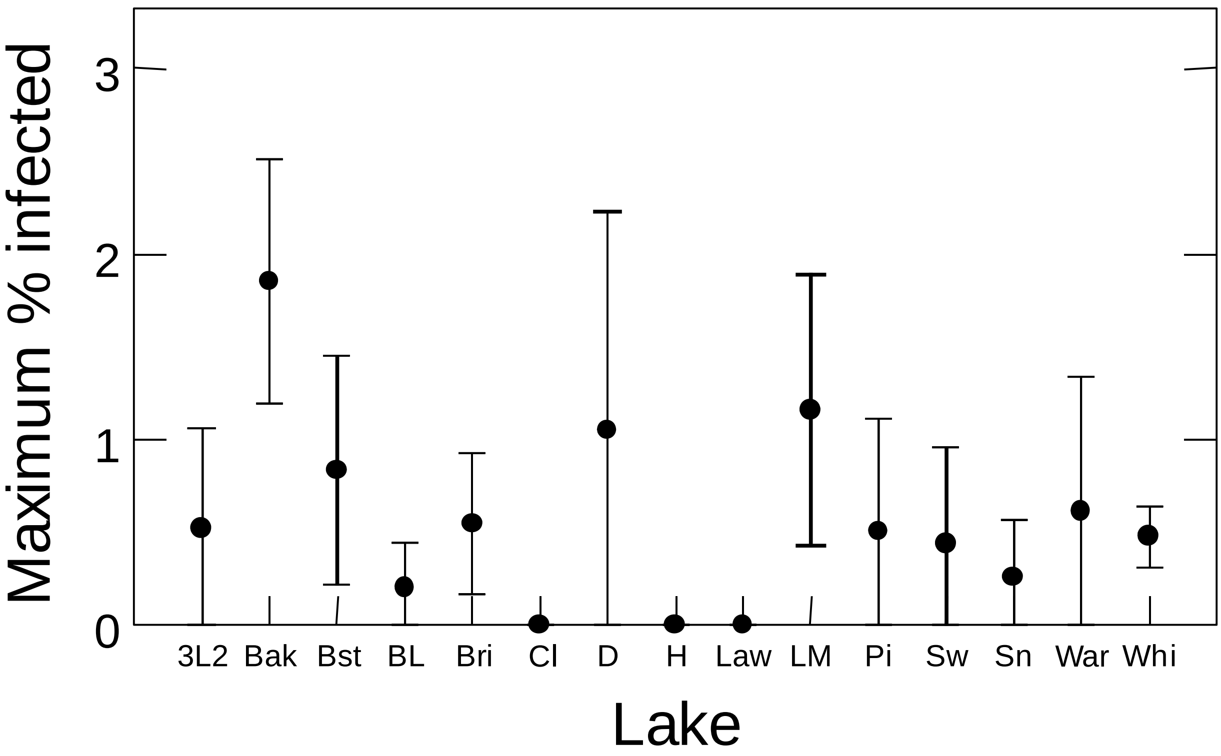

FIG. D27. Prevalence of oomycete infections in *D. dentifera* in 15 lake populations. Lake abbreviations are: “3L2” = Three Lakes Two, “Bak” = Baker, “Bst” = Bassett, “BL” = “Big Long”, “Bri” = Bristol, “Cl” = Cloverdale, “D” = Deep, “H” = Hall, “Law” = Lawrence, “LM” = Little Mill, “Pi” = Pine, “Sw” = Shaw, “Sn” = Sherman, “War” = Warner, and “Whi” = Whitford. These data are the same as those shown in the bottom panel of Fig. 2. Here, we plot each parasite separately to improve clarity. Error bars represent ±1 SE.

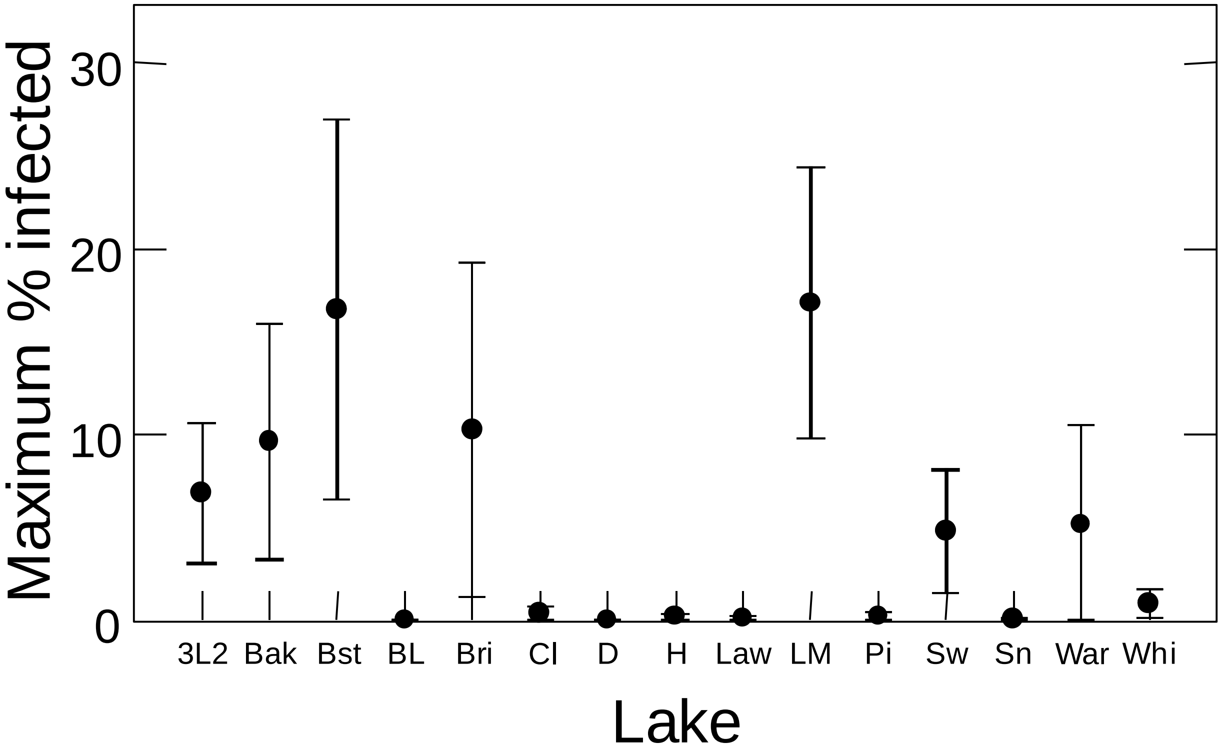

FIG. D28. Prevalence of *Metschnikowia bicuspidata* infections in *D. dentifera* in 15 lake populations. Lake abbreviations are: “3L2” = Three Lakes Two, “Bak” = Baker, “Bst” = Bassett, “BL” = “Big Long”, “Bri” = Bristol, “Cl” = Cloverdale, “D” = Deep, “H” = Hall, “Law” = Lawrence, “LM” = Little Mill, “Pi” = Pine, “Sw” = Shaw, “Sn” = Sherman, “War” = Warner, and “Whi” = Whitford. These data are the same as those shown in the bottom panel of Fig. 2. Here, we plot each parasite separately to improve clarity. Error bars represent ±1 SE.

Ecological Archives E091-234-A5

Meghan A. Duffy, Carla E. Cáceres, Spencer R. Hall, Alan J. Tessier, and Anthony R. Ives. 2010. Temporal, spatial, and between-host comparisons of patterns of parasitism in lake zooplankton. *Ecology* 91:3322–3331.

Appendix E. Annotated R code and output for analyses presented in the article.

This appendix presents example R computer code used for the analyses. We have organized the code to correspond to each table in the article, with code in blue, R output given in black, and annotation in red. R version 2.8.1 and lmer() in package lme4 version 0.999375-28 were used throughout. The data sets are available in the Supplement and are identified in the annotation of the computer code.

TABLE E1. Best-fitting model for full data set (including both hosts and all parasites). The full data are shown in Fig. 1A and B.

```
# load necessary libraries
library(lme4)

# input data set: DentPulInfectionsStacked_9Jun09.txt
d <- read.table(file.choose(), header=T)
d$Year <- as.factor(d$Year)

# define new categorical variables
d$LakeYear <- as.factor(100*as.numeric(d$Year) + as.numeric(d$Lake))

d$HostLakeYear <- as.factor(1000*as.numeric(d$Year) + 10*as.numeric(d$Lake) +
as.numeric(d$HostSp))

d$ParaLakeYear <- as.factor(1000*as.numeric(d$Year) + 10*as.numeric(d$Lake) +
as.numeric(d$ParaSp))

d$ParaHostYear <- as.factor(1000*as.numeric(d$Year) + 10*as.numeric(d$HostSp) +
as.numeric(d$ParaSp))

d$ParaHostLake <- as.factor(1000*as.numeric(d$Lake) + 10*as.numeric(d$HostSp) +
as.numeric(d$ParaSp))

d$ParaHostLake <- as.factor(1000*as.numeric(d$Lake) + 10*as.numeric(d$HostSp) +
as.numeric(d$ParaSp))

d$HostYear <- as.factor(100*as.numeric(d$Year) + as.numeric(d$HostSp))
d$ParaYear <- as.factor(100*as.numeric(d$Year) + as.numeric(d$ParaSp))

d$HostLake <- as.factor(10*as.numeric(d$Lake) + as.numeric(d$HostSp))
d$ParaLake <- as.factor(10*as.numeric(d$Lake) + as.numeric(d$ParaSp))

d$ParaHost <- as.factor(10*as.numeric(d$ParaSp) + as.numeric(d$HostSp))

d$ParaHostLakeYear <- as.factor(100000*as.numeric(d$Year) + 10000*as.numeric(d$Lake) +
100*as.numeric(d$HostSp) + as.numeric(d$ParaSp))

d$NN <- as.real(d$N)
d$z <- d$Ninf/d$NN
```

```
# perform analysis
lmer(z ~ 1 + (1 | ParaHost) + (1 | ParaLakeYear) + (1 | HostLakeYear) + (1 | ParaLake)
+ (1 | ParaYear) + (1 | HostLake), data=d,weights=NN,family=binomial,method="Laplace")
Generalized linear mixed model fit by the Laplace approximation
Formula: z ~ 1 + (1 | ParaHost) + (1 | ParaLakeYear) + (1 | HostLakeYear) +      (1 |
ParaLake) + (1 | ParaYear) + (1 | HostLake)
   Data: d
   AIC   BIC logLik deviance
 918.7 947.2 -452.4    904.7
Random effects:
 Groups       Name        Variance Std.Dev.
 ParaLakeYear (Intercept)  0.82868 0.91032
 ParaLake     (Intercept)  1.76350 1.32797
 HostLakeYear (Intercept)  0.66606 0.81613
 ParaYear     (Intercept)  0.42249 0.64999
 ParaHost     (Intercept) 11.47281 3.38715
 HostLake     (Intercept)  0.51627 0.71852
Number of obs: 432, groups: ParaLakeYear, 216; ParaLake, 72; HostLakeYear, 48;
ParaYear, 27; ParaHost, 18; HostLake, 16

Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -8.0255     0.8747  -9.175   <2e-16 ***
```

TABLE E2. Best-fitting model for data set using only parasites that attack both hosts (continued immediately with the data set generated for Table 1).

```
# perform analysis
dd=d[d$ParaSp == "Brood" | d$ParaSp == "Larssonia" | d$ParaSp == "Pasteuria" |
d$ParaSp == "Spiro" | d$ParaSp == "Oomycete",]

lmer(z ~ 1 + (1 | HostSp) + (1 | ParaHost) + (1 | HostLakeYear) + (1 | ParaLakeYear) +
(1 | ParaLake) + (1 | ParaYear), data=dd,weights=NN,family=binomial,method="Laplace")
Generalized linear mixed model fit by the Laplace approximation
Formula: z ~ 1 + (1 | HostSp) + (1 | ParaHost) + (1 | HostLakeYear) +      (1 |
ParaLakeYear) + (1 | ParaLake) + (1 | ParaYear)
   Data: dd
   AIC   BIC logLik deviance
 530.1 554.4 -258.0    516.1
Random effects:
 Groups       Name        Variance Std.Dev.
 ParaLakeYear (Intercept) 0.41319  0.64280
 HostLakeYear (Intercept) 1.03392  1.01682
 ParaLake     (Intercept) 0.86484  0.92997
 ParaYear     (Intercept) 0.98245  0.99119
 ParaHost     (Intercept) 1.26775  1.12595
 HostSp       (Intercept) 1.25450  1.12005
Number of obs: 240, groups: ParaLakeYear, 120; HostLakeYear, 48; ParaLake, 40;
ParaYear, 15; ParaHost, 10; HostSp, 2

Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -6.9813     0.9387  -7.437 1.03e-13 ***
```

TABLE E3. Variance among lakes and years for individual host-parasite pairings. These data are shown in Fig. 1C–F.

```
# input data set: DentInfectionsStacked20032006_9Jun09.txt
d <- read.table(file.choose(), header=T)
d$Year <- as.factor(d$Year)

# define new categorical variables
d$LakeYear <- as.factor(100*as.numeric(d$Year) + as.numeric(d$Lake))

d$ParaLakeYear <- as.factor(1000*as.numeric(d$Year) + 10*as.numeric(d$Lake) +
as.numeric(d$ParaSp))
```

```
d$ParaYear <- as.factor(100*as.numeric(d$Year) + as.numeric(d$ParaSp))

d$ParaLake <- as.factor(10*as.numeric(d$Lake) + as.numeric(d$ParaSp))

d$NN <- as.real(d$N)
d$z <- d$Ninf/d$NN

# perform analysis for Brood parasite on dentifera
dp <- d[d$ParaSp == "Brood",]
lmer(z ~ 1 + (1 | Lake) + (1 | Year), data=dp, weights=NN, family=binomial,
method="Laplace")
Generalized linear mixed model fit by the Laplace approximation
Formula: z ~ 1 + (1 | Lake) + (1 | Year)
   Data: dp
   AIC   BIC logLik deviance
 524.8 531.1 -259.4    518.8
Random effects:
 Groups Name        Variance Std.Dev.
 Lake   (Intercept) 1.88475  1.37286
 Year   (Intercept) 0.26857  0.51824
Number of obs: 60, groups: Lake, 15; Year, 4

Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -3.2828     0.4444  -7.387 1.50e-13 ***
```

TABLE E4. Correlations between parasites in *D. dentifera* populations (continued from Table E3 for *dentifera*).

```
# select pair of parasites
pd <- d[(d$ParaSp == "Brood" | d$ParaSp == "Oomycete"),]

# perform analysis for Brood and Oomycete on dentifera
h1 <- lmer(z ~ 0 + ParaSp + (0 + ParaSp | LakeYear), data=pd, weights=NN,
family=binomial, method="Laplace")

show(h1)
Generalized linear mixed model fit by the Laplace approximation
Formula: z ~ 0 + ParaSp + (0 + ParaSp | LakeYear)
   Data: pd
   AIC   BIC logLik deviance
 385.6 399.5 -187.8    375.6
Random effects:
 Groups   Name           Variance Std.Dev. Corr
 LakeYear ParaSpBrood    2.4378   1.5614
          ParaSpOomycete 6.6847   2.5855   0.220
Number of obs: 120, groups: LakeYear, 60

Fixed effects:
               Estimate Std. Error z value Pr(>|z|)
ParaSpBrood     -3.4889     0.2142  -16.29   <2e-16 ***
ParaSpOomycete  -7.4470     0.4096  -18.18   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
            PrSpBr
ParaSpOmyct 0.171
```

TABLE E5. Correlations between parasites in *D. pulicaria* populations. This is produced in the same way as Table E4.

TABLE E6. Correlations (ρ) of infection prevalence between *D. dentifera* and *D. pulicaria* for parasite species shared by the two host species.

```
library(lme4)
```

```
# input data set: DentPulInfectionsStacked_20Mar09.txt
d <- read.table(file.choose(), header=T)

# Alphabetize ParNum
d$aParNum <- d$ParNum
d$aParNum[d$ParNum==8] <- 1
d$aParNum[d$ParNum==9] <- 2
d$aParNum[d$ParNum==3] <- 3
d$aParNum[d$ParNum==7] <- 4
d$aParNum[d$ParNum==5] <- 5
d$aParNum[d$ParNum==1] <- 6
d$aParNum[d$ParNum==6] <- 7
d$aParNum[d$ParNum==2] <- 8
d$aParNum[d$ParNum==4] <- 9

# new variables
d$Year <- as.factor(d$Year)
d$LakeYear <- as.factor(100*as.numeric(d$Year) + as.numeric(d$Lake))
d$NN <- as.real(d$N)
d$z <- d$Ninf/d$NN

# perform analysis for Brood
pd <- d[d$aParNum == 3,]
lmer(z ~ 0 + HostSp + (0 + HostSp | LakeYear),
data=pd,weights=NN,family=binomial,method="Laplace")
```

TABLE E7. Variance components for the best-fitting model corresponding to Table 1A in the main text.

This Bayesian analysis requires using WinBUGS through R. Gelman and Hill (2007) and Qian and Shen (2007) give excellent introductions to this approach. The code below was adapted from Qian and Shen (2007).

The WinBUGs model below is first created and stored in the working directory of R.

```
model {
  for (i in 1:n){
    y[i] ~ dbin (p[i], n.y[i])
        ## response variable distribution is binomial
    logit(p[i]) <- Xbeta[i]
        ## logit transformation of the probability of success
    Xbeta[i] <- b.0 + b.t1[t1[i]] + b.t2[t2[i]] + b.t3[t3[i]] + b.t4[t4[i]] +
b.t5[t5[i]] + b.t6[t6[i]] + b.tA[tA[i]]
  }
  b.0 ~ dnorm(0, 0.0001)
  for (i.t1 in 1:n.t1){
    b.t1[i.t1] ~ dnorm (0, tau.t1)
  }
  for (i.t2 in 1:n.t2){
    b.t2[i.t2] ~ dnorm (0, tau.t2)
  }
  for (i.t3 in 1:n.t3){
    b.t3[i.t3] ~ dnorm (0, tau.t3)
  }
  for (i.t4 in 1:n.t4){
    b.t4[i.t4] ~ dnorm (0, tau.t4)
  }
  for (i.t5 in 1:n.t5){
    b.t5[i.t5] ~ dnorm (0, tau.t5)
  }
  for (i.t6 in 1:n.t6){
    b.t6[i.t6] ~ dnorm (0, tau.t6)
  }
  for (i.tA in 1:n.tA){
    b.tA[i.tA] ~ dnorm (0, tau.tA)
```

```
  }
        ## model coefficient priors
  sigma.t1 ~ dunif(0,100)
  sigma.t2 ~ dunif(0,100)
  sigma.t3 ~ dunif(0,100)
  sigma.t4 ~ dunif(0,100)
  sigma.t5 ~ dunif(0,100)
  sigma.t6 ~ dunif(0,100)
  sigma.tA ~ dunif(0,100)
        ## standard deviation prior
  tau.t1<- pow(sigma.t1, -2)
  tau.t2<- pow(sigma.t2, -2)
  tau.t3<- pow(sigma.t3, -2)
  tau.t4<- pow(sigma.t4, -2)
  tau.t5<- pow(sigma.t5, -2)
  tau.t6<- pow(sigma.t6, -2)
  tau.tA<- pow(sigma.tA, -2)

  s.t1 <- sd(b.t1[])
  s.t2 <- sd(b.t2[])
  s.t3 <- sd(b.t3[])
  s.t4 <- sd(b.t4[])
  s.t5 <- sd(b.t5[])
  s.t6 <- sd(b.t6[])
  s.tA <- sd(b.tA[])

  gamma.0 <- b.0 + mean(b.t1[]) + mean(b.t2[]) + mean(b.t3[]) + mean(b.t4[]) +
mean(b.t5[]) + mean(b.t6[]) + mean(b.tA[])
  for (i.t1 in 1:n.t1){
    gamma.t1[i.t1] <- b.t1[i.t1] - mean(b.t1[])
  }
  for (i.t2 in 1:n.t2){
    gamma.t2[i.t2] <- b.t2[i.t2] - mean(b.t2[])
  }
  for (i.t3 in 1:n.t3){
    gamma.t3[i.t3] <- b.t3[i.t3] - mean(b.t3[])
  }
  for (i.t4 in 1:n.t4){
    gamma.t4[i.t4] <- b.t4[i.t4] - mean(b.t4[])
  }
  for (i.t5 in 1:n.t5){
    gamma.t5[i.t5] <- b.t5[i.t5] - mean(b.t5[])
  }
  for (i.t6 in 1:n.t6){
    gamma.t6[i.t6] <- b.t6[i.t6] - mean(b.t6[])
  }
  for (i.tA in 1:n.tA){
    gamma.tA[i.tA] <- b.tA[i.tA] - mean(b.tA[])
  }
}
```

The WinBUGs model is then run through R using the library “arm”.

```
# Data are first uploaded into R using the code for Table 1 (above)

library (arm)

bugs.in <- function(infile=d){
    y <- infile$Ninf
    n.y <- infile$N
    n <- length(y)
    t1 <- as.numeric(ordered(infile$ParaHost))
    n.t1 <- max(t1)
    t2 <- as.numeric(ordered(infile$ParaLakeYear))
    n.t2 <- max(t2)
    t3 <- as.numeric(ordered(infile$HostLakeYear))
    n.t3 <- max(t3)
    t4 <- as.numeric(ordered(infile$ParaLake))
    n.t4 <- max(t4)
    t5 <- as.numeric(ordered(infile$ParaYear))
```

```
    n.t5 <- max(t5)
    t6 <- as.numeric(ordered(infile$HostLake))
    n.t6 <- max(t6)
    tA <- as.numeric(ordered(infile$ParaHostLakeYear))
    n.tA <- max(tA)

    bugs.dat <- list(n=n, n.y=n.y, n.t1=n.t1, n.t2=n.t2, n.t3=n.t3, n.t4=n.t4,
n.t5=n.t5, n.t6=n.t6, n.tA=n.tA, y=y, t1=t1, t2=t2, t3=t3, t4=t4, t5=t5, t6=t6, tA=tA)

    inits1 <- list(b.t1 = rep(0.00, n.t1),
        b.t2 = rep(0.00, n.t2),
        b.t3 = rep(0.00, n.t3),
        b.t4 = rep(0.00, n.t4),
        b.t5 = rep(0.00, n.t5),
        b.t6 = rep(0.00, n.t6),
        b.tA = rep(0.00, n.tA),
        b.0 = 0.00,
        sigma.t1=1,
        sigma.t2=1,
        sigma.t3=1,
        sigma.t4=1,
        sigma.t5=1,
        sigma.t6=1,
        sigma.tA=1)

    inits2 <- list(b.t1 = rep(0.50, n.t1),
        b.t2 = rep(1.00, n.t2),
        b.t3 = rep(0.50, n.t3),
        b.t4 = rep(1.00, n.t4),
        b.t5 = rep(0.50, n.t5),
        b.t6 = rep(1.00, n.t6),
        b.tA = rep(1.00, n.tA),
        b.0 = 1.00,
        sigma.t1=1,
        sigma.t2=2,
        sigma.t3=3,
        sigma.t4=2,
        sigma.t5=2,
        sigma.t6=2,
        sigma.tA=3)

    inits3 <- list(b.t1 = rep(1.50, n.t1),
        b.t2 = rep(0.50, n.t2),
        b.t3 = rep(1.00, n.t3),
        b.t4 = rep(0.50, n.t4),
        b.t5 = rep(1.00, n.t5),
        b.t6 = rep(0.50, n.t6),
        b.tA = rep(1.00, n.tA),
        b.0 = 0.50,
        sigma.t1=1,
        sigma.t2=2,
        sigma.t3=3,
        sigma.t4=2,
        sigma.t6=3,
        sigma.tA=2)

    inits <- list (inits1, inits2, inits3)
    parameters <-
c("s.t1","s.t2","s.t3","s.t4","s.t5","s.t6","s.tA","gamma.0","gamma.t1","gamma.t2","ga
mma.t3","gamma.t4","gamma.t5","gamma.t6","gamma.tA")
    return(list(para=parameters, data=bugs.dat, inits=inits))
}
input.to.bugs <- bugs.in()
bugs.out <- bugs(input.to.bugs$data,
input.to.bugs$inits, input.to.bugs$para,
model.file="duffy1.bug",n.chains=3, n.iter=100000, DIC=F)

print(bugs.out)
```