Appendix S1: From feasible sets to the Poisson distribution

The Poisson distribution is the canonical null hypothesis for host-parasite distributions. While there are many ways to obtain a Poisson distribution, the most common derivation in parasite ecology is via a death-immigration process (Anderson and Gordon, 1982). However, one can also obtain a Poisson distribution using the feasible set approach as described below.

As discussed in the main text, the shape of all observed host-parasite distributions are inherently constrained by the total number of parasites in a sample P and the total number of hosts in a sample H. These are hard constraints on the system (Haegeman and Etienne, 2010). Assume now that both hosts and parasites are labeled. For parasites, this means that a given host can have the same parasite intensity (e.g. 1 parasite) in different ways depending on the identity of the individual parasites (e.g. the host is infected with parasite individual A or infected with parasite individual B). For example, given $P = 3$ parasites and $H = 3$ hosts the possible macrostates for the feasible set constrained by $P = 3$ and $H = 3$ are: $\{\{3, 0, 0\}, \{2, 1, 0\}, \{1, 1, 1\}\}$. Given labeled hosts and labeled parasites the probability of seeing the macrostate $\{3, 0, 0\}$ is 3 / 27, the probability of seeing (2, 1, 0) is 18 / 27, and the probability of seeing (1, 1, 1) is 6 / 27. The probabilities of seeing a single host with $x = 0, 1, 2,$ or 3 parasites are $p(0) = 8/27$, $p(1) = 12/27$, $p(2) = 6/27$ and $p(3) = 1/27$.

More generally, given H labeled hosts and P labeled parasites the total number of possible configurations with labeled parasites is $M = H^P$. Given a macrostate m from the feasible set defined by H and P, there are $\mu_m = b_m * \binom{P}{p_1, p_2, \ldots, p_H}$ ways to observed this macrostate with labeled parasites and labeled hosts. b_m is number of ways that macrostate can be realized given unlabeled parasites and labeled hosts (see Methods in main text), $\binom{P}{p_1, p_2, \ldots, p_H}$ is a multinomial coefficient, and p_i is number of parasites in host i. Taken together, the weight on any particular macrostate m given labeled hosts and labeled parasites is μ_m / M.

Haegeman and Etienne (2010) showed that the probability of observing any particular con-

figuration X of labeled parasites is

$$X \sim \text{Multinomial}(P, p_1 = \frac{1}{H}, p_2 = \frac{1}{H}, \ldots, p_H = \frac{1}{H}) \tag{1}$$

From this result, it is easy to see that the probability of a single host having $x = 0, \ldots, P$ parasites is given by a binomial distribution such that

$$x \sim \text{Binomial}(P, \frac{1}{H}) \tag{2}$$

Now as H increases (via a larger sample size of hosts) P will also increase (because you are finding more parasites). Therefore, when H is reasonably large and P is reasonably large, the binomial distribution will be well-approximated by a Poisson distribution based on the standard relationship between the Binomial and the Poisson (Zillio and He, 2010).

Appendix S2: The central tendency of a feasible set

The central tendency of a feasible set could be found in many different ways, including identifying the mean, median, or mode of the feasible set. While there are advantages and disadvantages to each of these measures of central tendency, we chose to use the vector of marginal medians as our measure for central tendency (Niinimaa and Oja, 2006). Take a feasible set F where F is a matrix with H columns and N rows, where H is the number of hosts and N is the number of macrostates in the feasible set (or sampled macrostates in the sampled feasible set, Figure 2). The marginal median is computed by taking the median of each of the H columns.

We chose this measure for two reasons. First, because we are dealing with finite samples from the feasible set, we found that the marginal medians were less variable with sample size than the marginal modes (as used by Locey and White, 2013). That being said, using the marginal modes or medians had no effects on our conclusions and tended to yield very similar predictions (e.g. Figure 2). Moreover, the marginal medians were also less sensitive to skew than the mean

of the feasible set. Second, we found that the using the marginal median for the composition model yield nearly identical predictions as the predicted rank abundance distribution from the analytical solution for composition model (equation 1 in the main text; Figure 3). This provided further evidence that the marginal medians were a good measure of central tendency for this study.

Appendix S3: Goodness-of-fit tests for constraint-based null models

While the R^2 statistic described in the main text provides an easy-to-understand metric of how well a constraint-based model predicts the observed data, it does not account for fact that one of the constraint-based models could have "generated" the data, but the theory and the data might have a low R^2 simply due to sampling error (Xiao et al., 2015). Therefore, we also explored two additional goodness-of-fit tests to determine whether the observed host-parasite distribution differed from a constraint-based null model.

The first test we used was a two-sample Anderson-Darling test which compares the observed RAD and predicted RAD and determines whether these two RADs come from the same distribution (10,000 bootstraps with 'kSamples' R package; Scholz and Zhu, 2015). The null hypothesis of this test is that the two distributions come from the same distribution and is rejected if $p < \alpha$ where α is the Type 1 error rate. Here we set $\alpha = 0.05$.

Using this test, we found that 93% percent of observed host-parasite distributions were not significantly different than the partition model and 90% were not significantly different than the composition model. In contrast, 58% were not significantly different than the Poisson distribution. Appendix S6: Figure 3 shows the proportion of distributions not rejected by the constraint-based models for each host-parasite combination in this study. Across host-parasite combinations, the proportion not rejected by the Anderson-Darling test is largely above 80%. This is a similar pattern to Appendix S6: Figure 2 showing the median R^2 values for each host parasite combination.

While the non-parametric two-sample Anderson-Darling test has more power than other goodness-of-fit tests such as the Kolmogorov-Smirnov test (Engmann and Cousineau, 2011), non-parametric tests inherently lack power compared to parametric alternatives. Therefore, we also implemented a parametric bootstrap test in which we did the following.

1. We computed the log-likelihood of seeing a given observed host-parasite distribution based on the probability mass function $p(x|p, H)$ of seeing a single host with $x = 0, \ldots, P$ parasites for the partition model and the composition model. For the composition model, we used in the analytical formula given in equation 1 in the manuscript. For the partition model, we approximated the probability mass function by randomly drawing 500 macrostates and computing $p(x|P, H) = \sum_{m \in \hat{F}} p(x|m, P, H)p(m|P, H)$ where m is a macrostate in the sampled feasible set $\hat{F}$.

2. For each observed host-parasite distribution with P and H, we then sampled 500 host-parasite distributions of length H by drawing from $p(x|P, H)$ from the corresponding constraint-based model.

3. For each of these 500 sampled host-parasite distributions, we computed the log-likelihood of seeing the simulated data under the model. This gave us a distribution of 500 log-likelihoods.

4. We then compared the observed log-likelihood to the distribution of sampled log-likelihoods. If the observed log-likelihood did not fall between the the 0.025 and 0.975 quantiles on the sampled log-likelihoods, we rejected that the observed host-parasite distribution was "generated" by the constraint-based model (Rominger and Merow, 2016).

Using this parametric bootstrap approach we found that 99% of the observed host-parasite distributions were not significantly different than the partition model and 96% were not significantly different than the composition model, similar to the values we saw for the Anderson-Darling test. In contrast, only 5% were not significantly different than the Poisson model using

this parametric test. This test provides even further support that the observed-host parasite distributions are highly constrained by P and H.

Appendix S4: Extending the constraint-based models to include parasite-induced host mortality

Ribeiroia has well-documented negative effects on amphibian survival in the lab and in the field (Johnson, 1999; Johnson et al., 2012). Therefore, we might expect that incorporating the effect of *Ribeiroia*-induced host mortality as an additional constraint on a predicted host-*Ribeiroia* distribution will improve the overall fit of a given constraint-based model to the observed parasite distribution.

To account for *Ribeiroia*-induced mortality, we use the data from the laboratory infection experiments described in Johnson (1999) and Johnson et al. (2012) to estimate an intensity-dependent survival curve for *Pseudacris regilla* infected with *Ribeiroia*. We use a standard logistic survival curve given by

$$\text{logit}(s(x)) = a + bx \tag{3}$$

where logit is the logistic function, $s(x)$ is the probability of amphibian survival given a *Ribeiroia* intensity of x, b is the effect of *Ribeiroia* intensity on the log-odds of amphibian survival, and a is the "threshold" at which the host begins to experience parasite-induced mortality. Using a generalized linear model with a binomial response and a logistic link, we estimated the parameters of the *P. regilla-Ribeiroia* survival curve to be $a = 1.67$ and $b = -0.05$. See the file `manuscript_analysis_parasite_mortality.py` or Figure 2 in the main text for the data used to fit this GLM.

Using this estimated survival curve, we used a Metropolis-Hastings algorithm to draw a weighted feasible set that accounted for the additional constraint of parasite-induced host mor-

tality. We did this using the following algorithm:

1. Calculate the total number of *Ribeiroia* parasites P and *Pseudacris* hosts H in given empirical host-parasite distribution.

2. Draw an initial candidate macrostate with P and H using the algorithms provided by Locey and McGlinn (2013).

3. For the candidate macrostate, calculate the likelihood of observing this macrostate given the host-survival curve described above (Equation 3). To do this, we assumed that each host in a macrostate was independent and calculated the likelihood of observing the macrostate by multiplying the probabilities of observing each host with a given parasite abundance under the estimated survival curve. The assumption of independence is conditional on observing the macrostate, not deriving the macrostate where each host is inherently non-independent given that the total number of parasites in the system is fixed.

4. Propose a new macrostate and calculate its likelihood from equation 3. The proposal distribution for drawing a new macrostate is symmetric due to fact that the basic partition model assumes that each macrostate is equally likely (Locey and White, 2013).

5. Take the ratio, r, of the proposed likelihood over the candidate likelihood. If r is greater than 1, accept the proposed macrostate. Otherwise, accept the proposed macrostate with probability r and accept the candidate macrostate with probability $1 - r$.

6. Set the accepted macrostate as your new candidate macrostate and repeat steps 3-5 a large number of times. Discard the first half of the iterations as warm-up/burn-in samples.

The remaining samples give the weighted feasible set with the additional constraint of parasite-induced mortality.

To sample from the composition model with parasite-induced host mortality we used the same procedure described above, but in addition to assigning each proposed macrostate a likelihood based on the survival function, we also assign each proposed macrostate a likelihood based

on composition weighting described in the main text. This amounts to multiplying the two likelihoods. As above, the likelihood ratio of the proposed macrostate and the candidate macrostate determines whether to accept or reject the proposed macrostate.

We could also sample from a binomial distribution with a parasite-induced mortality constraint. This is useful as the binomial distribution is the finite equivalent of a Poisson distribution in this case (see Appendix A), which is the standard null model used in parasite ecology (Anderson and Gordon, 1982). We could do this by changing our proposal distribution to a multinomial distribution where the probability of any one of the H hosts encountering a parasite is $1/H$ (Haegeman and Etienne, 2010). The multinomial distribution from which we propose a new configuration X is then given by

$$X \sim \text{Multinomial}(P, p_1 = \frac{1}{H}, p_2 = \frac{1}{H}, \ldots, p_H = \frac{1}{H}) \tag{4}$$

We can draw proposal configurations from this multinomial model and, as described above, assign them a likelihood based on both 1) their likelihood given by the estimated survival function and 2) their probability under the multinomial model. Then we accept or reject our proposed configuration based on the ratio of the likelihoods for the proposed and candidate configuration times the probability ratio of the candidate and proposed configurations under the multinomial model. This is the additional weighting on the acceptance ratio imposed by the Metropolis-Hastings algorithm. In summary, this is a long-winded way of saying that if the survival function likelihood is 1 and there is no effect of parasite mortality, the algorithm will sample from a multinomial distribution whose predicted rank abundance distribution is equal to the predicted RAD from a binomial distribution with P parasites and H hosts. These algorithms are implemented and tested in the accompanying code.

We applied the algorithms and *P. regilla-Ribeiroia* survival curve to all 133 *P. regilla-Ribeiroia* distributions in the dataset. For each constraint-based model with parasite mortality, we ran the Metropolis-Hastings algorithm for 2000 iterations, discarding the first 1000 iterations as warm-

up/burn-in samples. We ran this analysis multiple times from different random starting points to ensure the chains were converging to the same stationary distribution. In general, visual inspection of the trace plots of the mortality-constrained feasible set chains showed consistent convergence, good mixing, and generally had acceptance rates above 50%. This high acceptance rate was expected as these chains were designed to have an acceptance rate of 1 (i.e. sampling from the unconstrained distribution) if parasite-induced mortality was not important. The chains of the constrained maximum entropy model had an average acceptance rate of 0.34 and the majority of the chains showed good mixing. However, 9 of the 133 constrained composition model chains had acceptance rates of less than 10% and showed high autocorrelation between samples. Both excluding the distributions resulting from these chains from the analysis and running the chains for longer had no effect on the conclusions we drew about parasite-induced mortality improving the fit of the maximum entropy model. Moreover, these chains were not problematic in the constrained partition model for which we also concluded that parasite-induced host mortality improved the fit of the constraint-based model. Because these under-sampled chains did not affect our inference or conclusions, we included the distributions resulting from these chains in the analysis presented in the main paper.

Appendix S5: Randomization test for heterogeneity

To test whether the host-parasite distribution predicted by the heterogeneity model was better than a host-parasite distribution generated by a randomly grouping hosts, we randomly permuted hosts with their corresponding parasite intensities into groups with the same number of hosts H_j as predicted by the regression tree analysis on a given host-parasite distribution. We then calculated the total number of parasites in each group j and used the procedure described in Figure 3 in the main text to generate the predicted RAD for these randomly permuted mixture distributions. More specifically, this is equivalent to finding the rank abundance distribution for the mixture model $g(x) = \sum_{j=1}^{G} \frac{H_i}{H} p(x|P_{j,rand}, H_j)$ where $P_{j,rand}$ indicates that the total number of

parasites in each group j varies with each random permutation. We repeated this randomization 200 times for every empirical host-parasite distribution for both the partition model and composition model. This generated 400 permuted host-parasite distributions (200 using the partition model and 200 using the composition model) for every observed host-parasite distribution. We could then use these permuted samples to determine whether the host groupings produced by the regression tree analysis improved the fit of the constraint-based models to the empirical distributions more than we would expect by randomly permuting hosts into groups. If the mixture model from the regression tree provided a significantly better fit to the host-parasite distribution then randomly grouping hosts, we expected its R^2 value to be significantly higher than the upper bound of the 95% quantile of the R^2 values from the randomly generated groupings.

Appendix S6: The levels of aggregation and predictions of constraint-based null models for specific amphibian host-trematode parasite distributions

References

Anderson, R. M. and D. M. Gordon, 1982. Processes influencing the distribution of parasite numbers within host populations with special emphasis on parasite-induced host mortalities. *Parasitology* **85**:373–398.

Engmann, S. and D. Cousineau, 2011. Comparing distributions: the two-sample Anderson-Darling test as an alternative to the Kolmogorov-Smirnoff test. *Journal of Applied Quantitative Methods* **6**:1–17.

Haegeman, B. and R. S. Etienne, 2010. Entropy maximization and the spatial distribution of species. *The American Naturalist* **175**:E74–90.

Johnson, P. T., 1999. The effect of trematode infection on amphibian limb development and survivorship. *Science* **284**:802–804.

Johnson, P. T. J., J. R. Rohr, J. T. Hoverman, E. Kellermanns, J. Bowerman, and K. B. Lunde, 2012. Living fast and dying of infection: Host life history drives interspecific variation in infection and disease risk. *Ecology Letters* **15**:235–242.

Locey, K. J. and D. J. McGlinn, 2013. Efficient algorithms for sampling feasible sets of macroecological patterns. *PeerJ* pages 1–23.

Locey, K. J. and E. P. White, 2013. How species richness and total abundance constrain the distribution of abundance. *Ecology Letters* **16**:1177–85.

Niinimaa, A. and H. Oja, 2006. Multivariate Median. In *Encyclopedia of Statistical Sciences*, pages 1 – 9. John Wiley & Sons, Ltd, Hoboken, New Jersey.

Rominger, A. J. and C. Merow, 2016. meteR : an r package for testing the maximum entropy theory of ecology. *Methods in Ecology and Evolution* pages 1–7.

Scholz, F. and A. Zhu, 2015. kSamples: K-sample rank tests and their combinations.

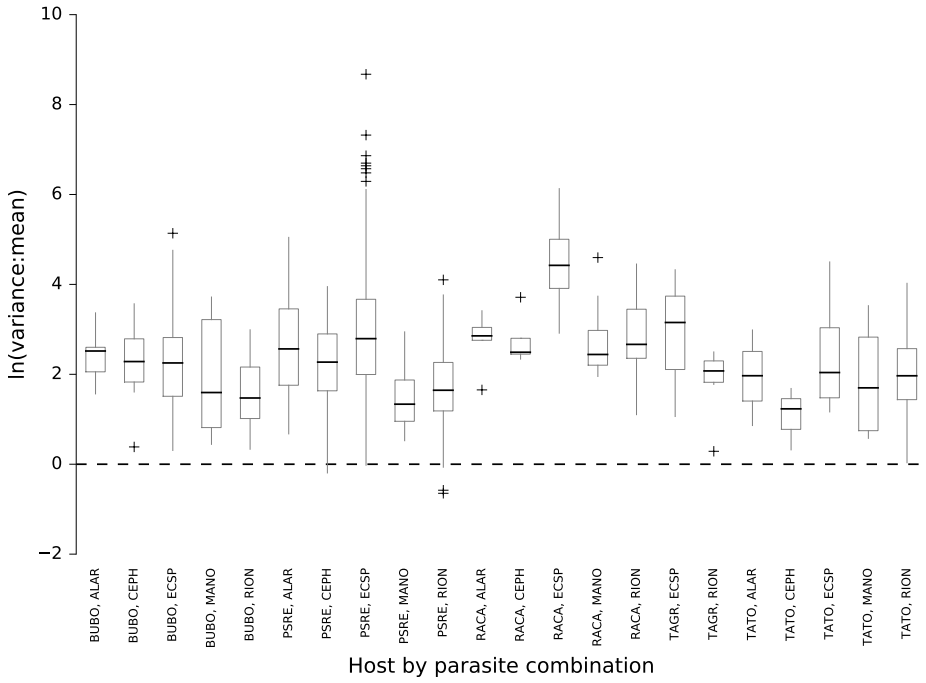

Figure 1: Boxplots of the ln(variance to mean ratio) for each host by parasite combination for all 842 host-parasite distributions used in this analysis. The black dashed line indicates where the ln(variance to mean ratio) is zero, consistent with an unaggregated Poisson distribution. As expected, nearly all of the host-parasite distributions have a ln(variance to mean ratio) greater than 0, indicating an aggregated distribution.

228 Xiao, X., D. J. McGlinn, and E. P. White, 2015. A strong test of the Maximum Entropy Theory of Ecology. *The American Naturalist* **185**:E70–80.

Zillio, T. and F. He, 2010. Modeling spatial aggregation of finite populations. *Ecology* **91**:3698–
231 3706.

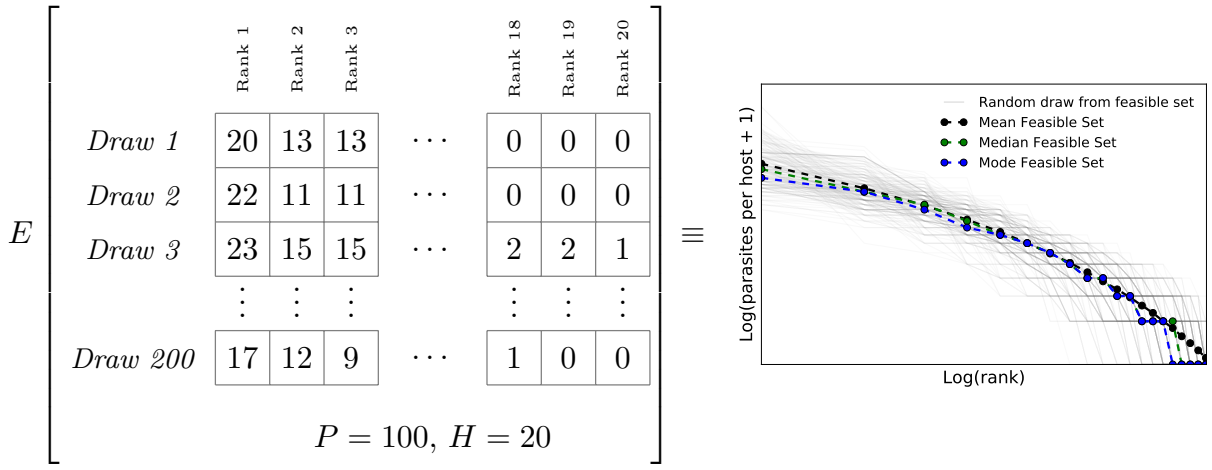

Figure 2: Given a host-parasite system with $P = 100$ parasites and $H = 20$ hosts, the feasible set can be approximated by drawing some number of random macrostates from the full feasible set (200 in this example) and ranking the hosts in each drawn configuration where the host with the most parasites has a rank of 1 and the host with the fewest individuals has a rank of H. The plot shows the graphical representation of this procedure where each gray line is a sampled configuration from the feasible set and the dashed lines are different measures of the center of the sampled feasible set.

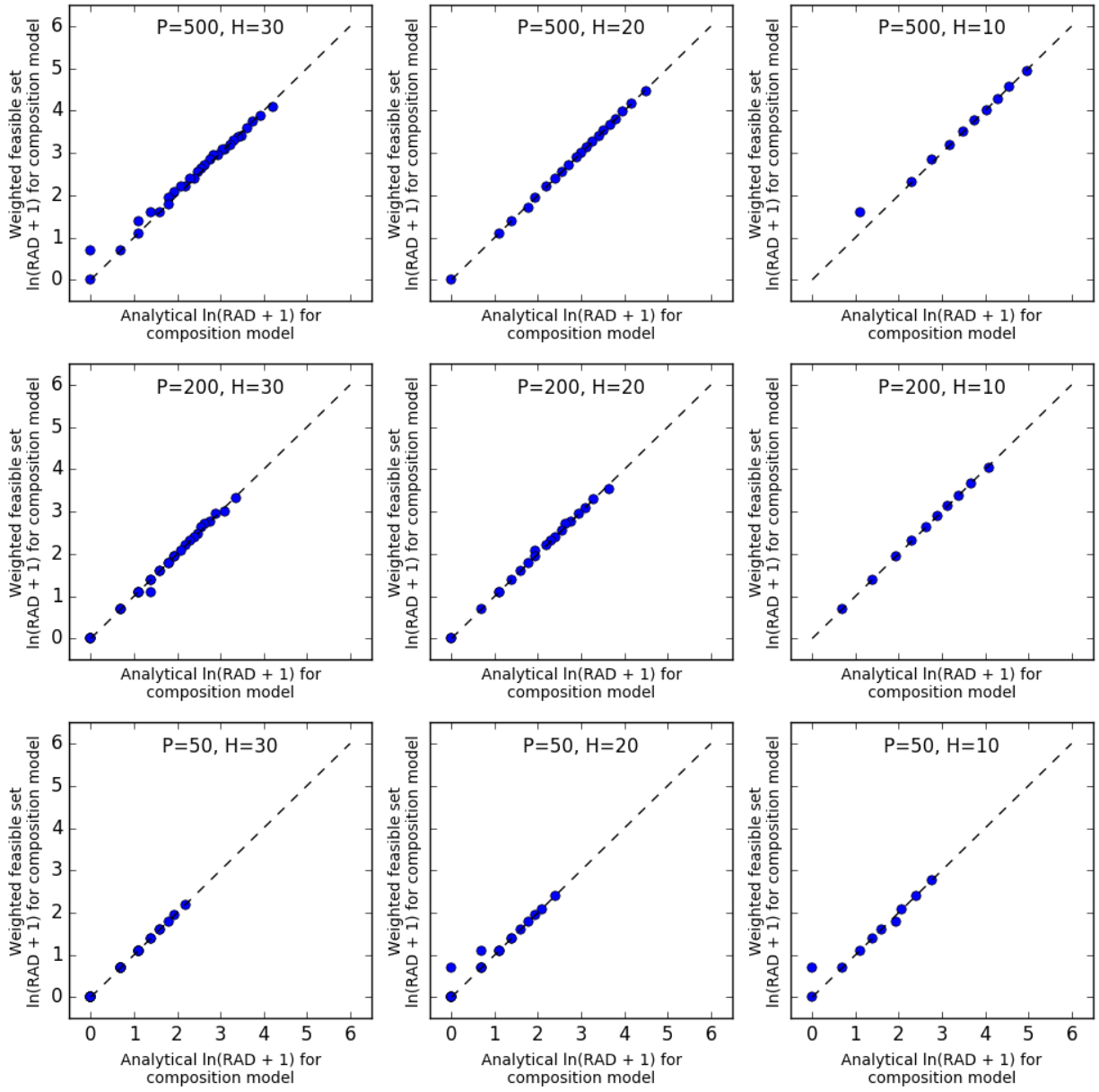

Figure 3: Comparing rank abundance distributions (RAD) from the analytical solution to the composition model (equation 1 in the main text) and the rank abundance distribution (RAD) predicted from the marginal medians of the weighted feasible set of the composition model. Each plot shows a different combination of P and H. Each blue point is a particular host in the RAD with some ln(parasite abundance + 1). The black dashed line is the one to one line. Other than some occasional sampling error, these two methods of predicting the RAD give nearly identical results.

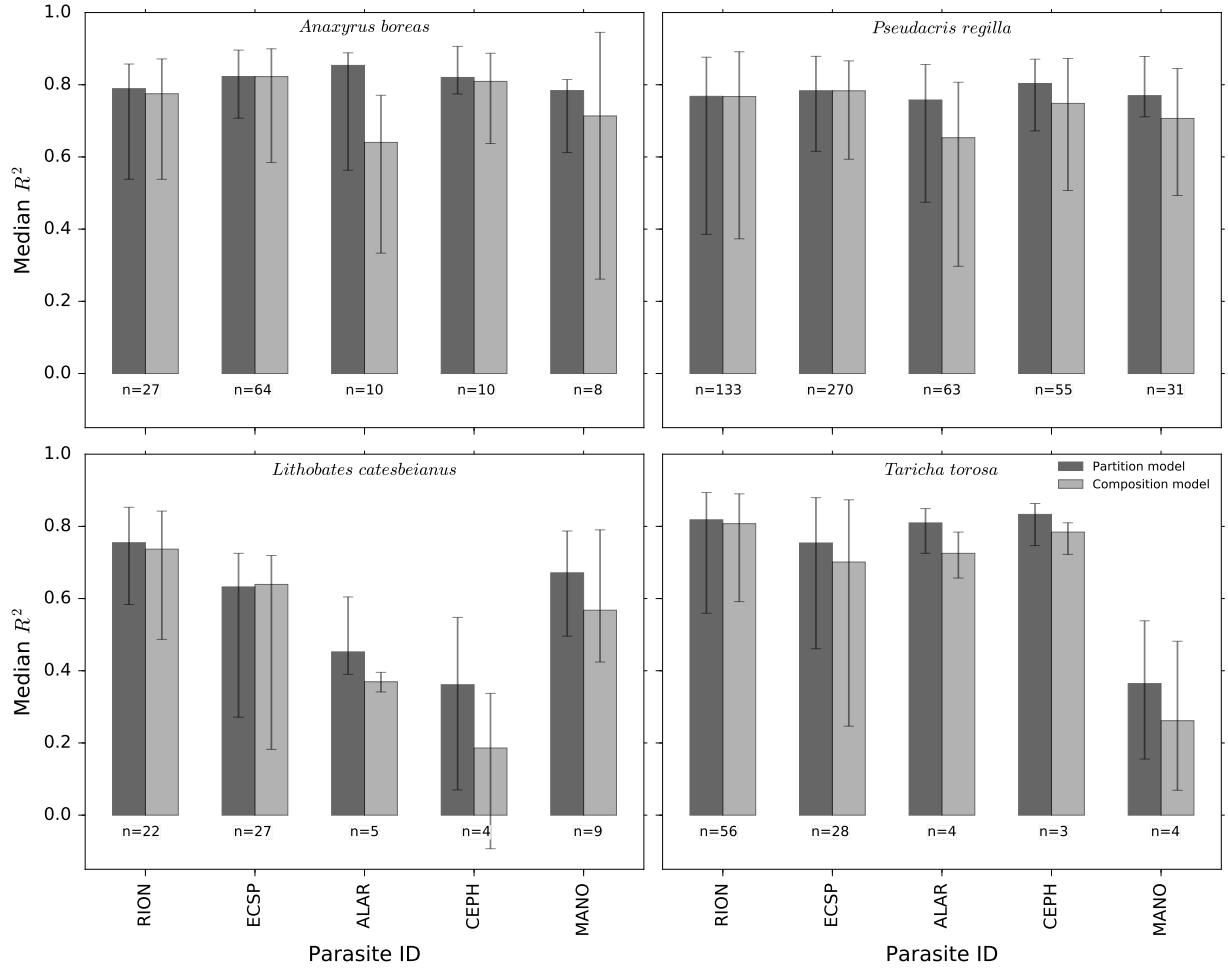

Figure 4: The height of each bar gives the median R^2 about the 1:1 line comparing the observed and predicted rank abundance distributions for various host-parasite combinations. The error bars give the first and third quartiles of the distribution of R^2 values for each distribution of a species host-parasite distribution. The number of distributions for each host-parasite combination that were used to compute this median R^2 are shown in the figure. The x-axis gives the 5 trematode parasites examined in this analysis: *Ribeiroia ondatrae* (RION), *Echinostoma* sp. (ECSP), *Alaria* sp. (ALAR), *Cephalogonimus* sp. (CEPH), and *Manodistomum* sp. (MANO). *Taricha granulosa* is not shown in this plot as it was never infected with ALAR, CEPH or MANO.

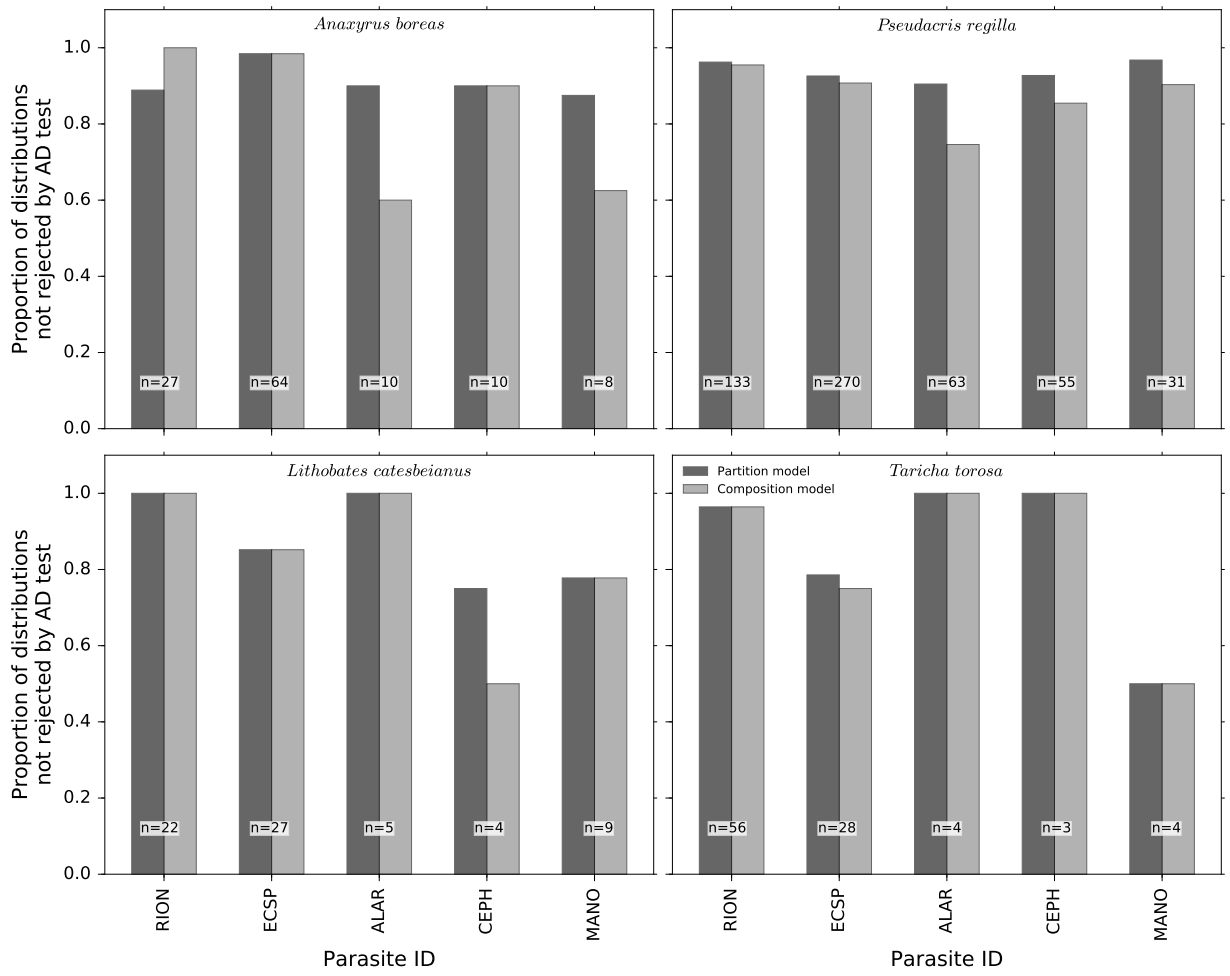

Figure 5: The proportion of predicted host-parasite distributions that were not rejected by the Anderson-Darling test ($\alpha = 0.05$) when compared to the partition model and the composition model. The number of distributions tested for any given host-parasite combination are also displayed on the figure. The x-axis gives the 5 trematode parasites examined in this analysis: *Ribeiroia ondatrae* (RION), *Echinostoma* sp. (ECSP), *Alaria* sp. (ALAR), *Cephalogonimus* sp. (CEPH), and *Manodistomum* sp. (MANO). *Taricha granulosa* is not shown in this plot as it was never infected with ALAR, CEPH or MANO.

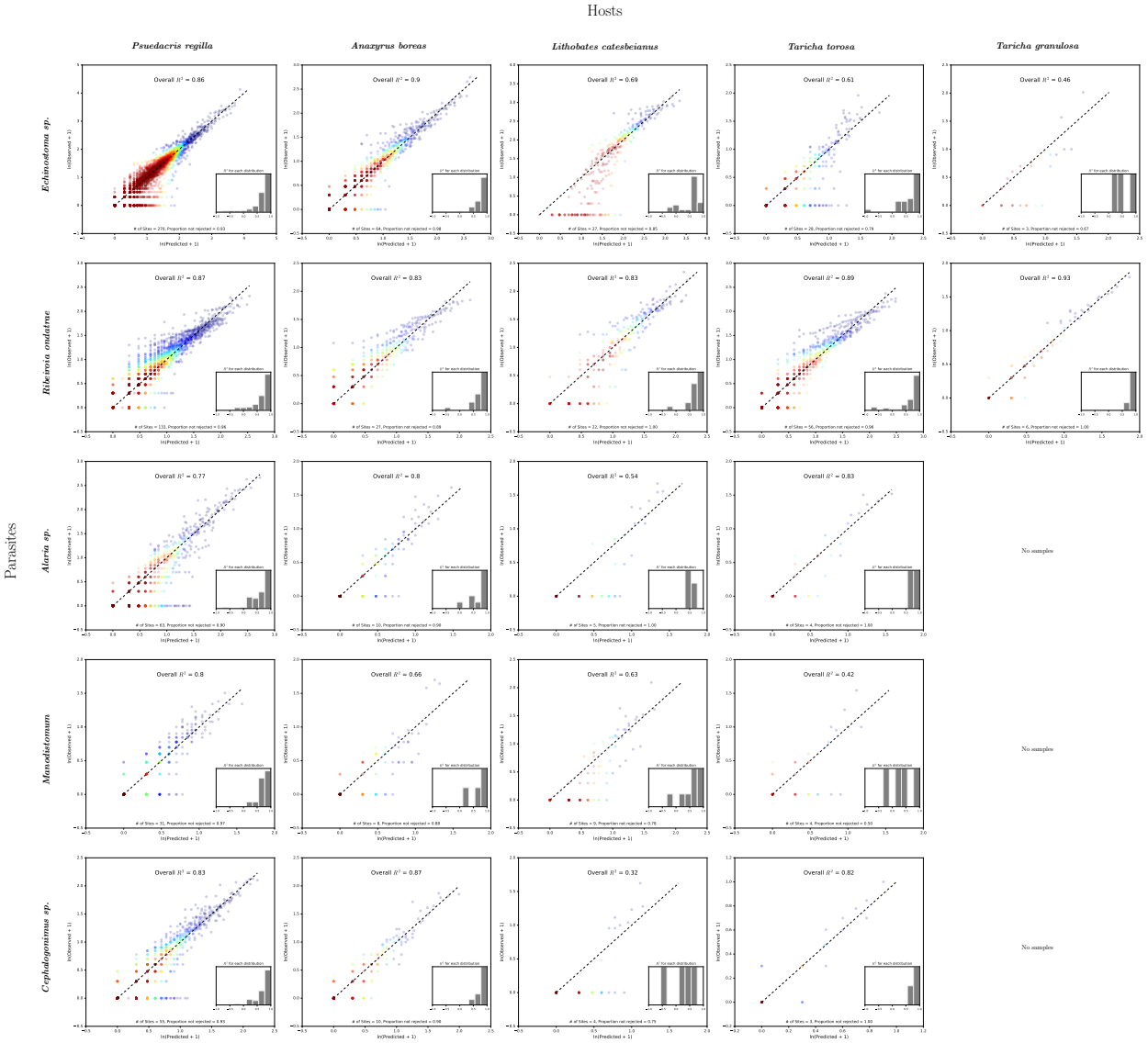

Figure 6: Comparison of observed and predicted host-parasite distributions from the partition model for each host-parasite combination. In each subplot, the x-axis gives the model ln(predicted parasite abundance + 1) and the y-axis gives the ln(observed parasite abundance + 1). Each point gives a particular host's predicted and observed parasite abundance. The color of the points represent the density of points in that region. "Hotter" colors mean there are more points in that region while "cooler" colors mean there are less points in that region. The dashed black line gives the 1:1 line, along which we would expect the points to fall if the model was a perfect fit. The overall R^2 measures the proportion of variance in the data described by the model. The text at the bottom of the plot gives the total number of distributions that were tested and the proportion of them that were not rejected by an Anderson-Darling test at $\alpha = 0.05$. Finally, the histogram in the lower right hand corner of each plot gives the histogram of the R^2 values for each of the host-parasite distributions for a given host-parasite combination.

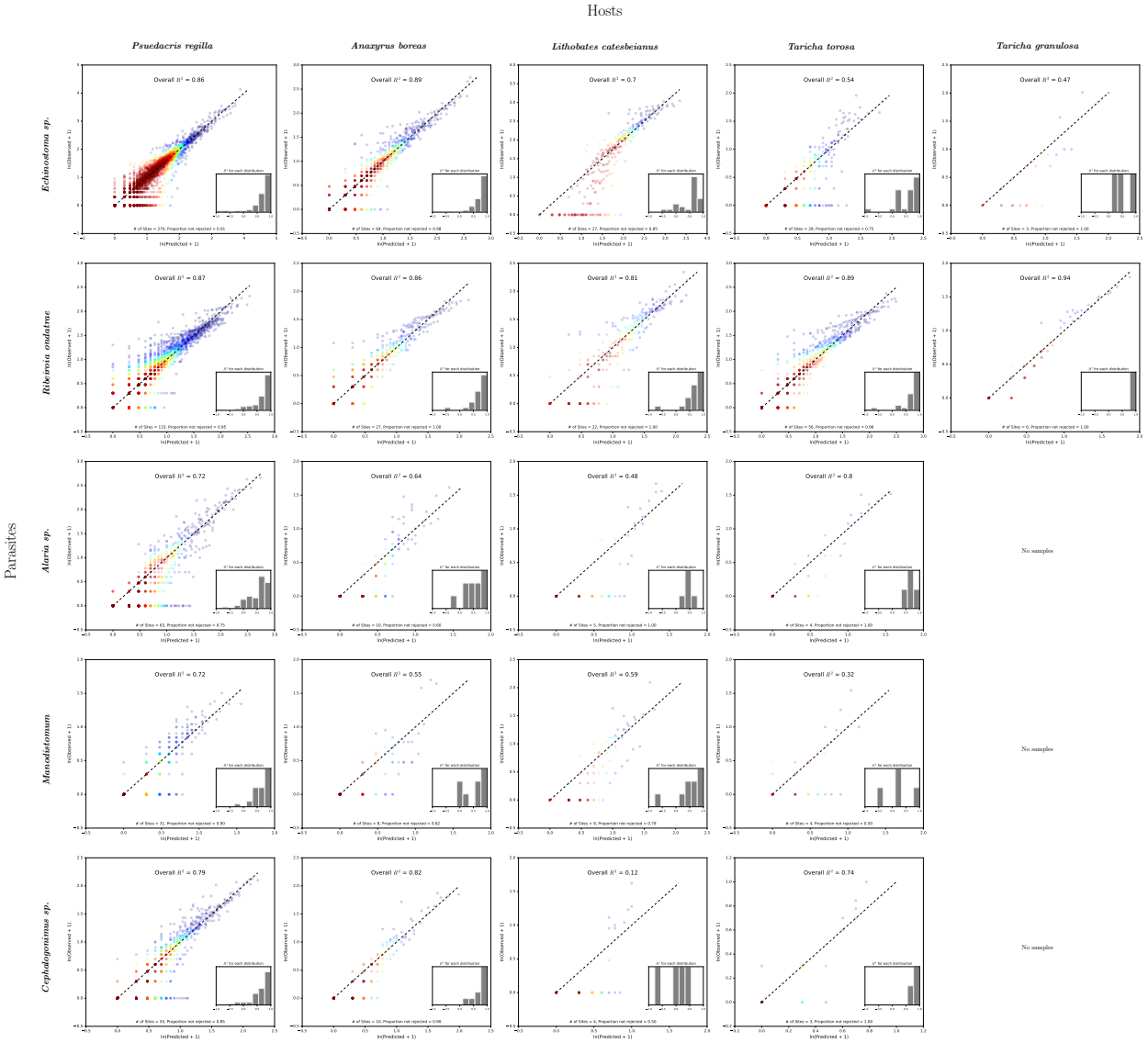

Figure 7: Comparison of observed and predicted host-parasite distributions from the composition model for each host-parasite combination. In each subplot, the x-axis gives the model ln(predicted parasite abundance + 1) and the y-axis gives the ln(observed parasite abundance + 1). Each point gives a particular host's predicted and observed parasite abundance. The color of the points represent the density of points in that region. "Hotter" colors mean there are more points in that region while "cooler" colors mean there are less points in that region. The dashed black line gives the 1:1 line, along which we would expect the points to fall if the model was a perfect fit. The overall R^2 measures the proportion of variance in the data described by the model. The text at the bottom of the plot gives the total number of distributions that were tested and the proportion of them that were not rejected by an Anderson-Darling test at $\alpha = 0.05$. Finally, the histogram in the lower right hand corner of each plot gives the histogram of the R^2 values for each of the host-parasite distributions for a given host-parasite combination.

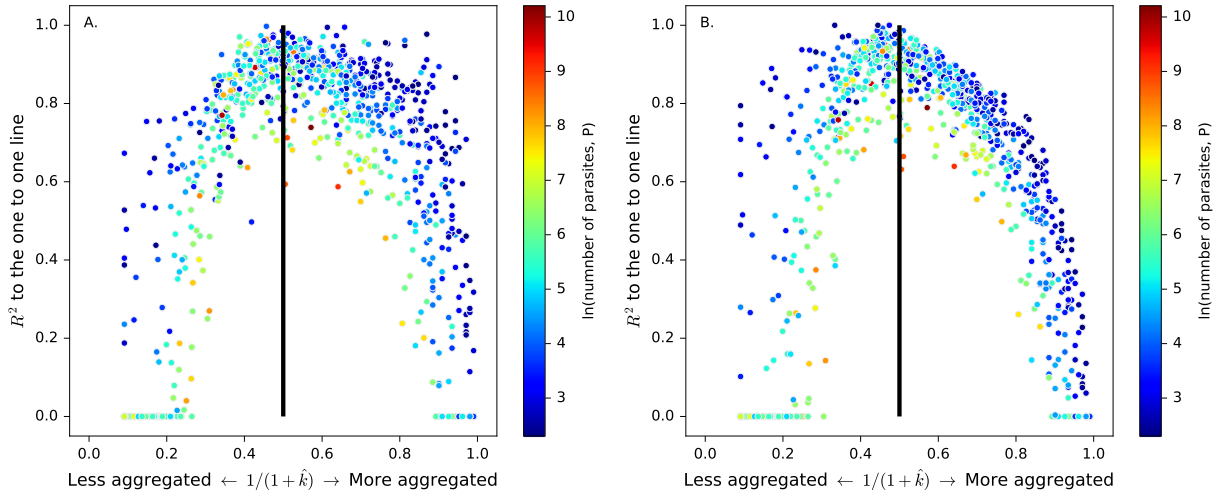

Figure 8: **A.** gives the partition model R^2 values plotted against the maximum likelihood estimates of the k parameter from the negative binomial distribution ($\hat{k}$) for all 842 host-parasite distributions used in this study. Each point is a host-parasite distribution and the color of each point indicates how many parasites (P) were in that distribution. The bold vertical line indicates where $\hat{k} = 1$ or $1/(1+\hat{k}) = 0.5$, which corresponds to the composition model. $\hat{k}$ was transformed for visual clarity. **B.** provides the same information, except that the R^2 values are from the composition model. Note that because of error in the estimate of $\hat{k}$ due to finite sampling, $\hat{k}$ may not equal one and R^2 can still be well above 0.5. Also note that the partition model, when compared to the composition model, tends to yield in higher R^2 values for distributions that are more aggregated than $\hat{k} = 1$ or $1/(1+\hat{k}) = 0.5$.