Supplementary Material for: What controls the range of hosts a fish parasite infects?

Tad Dallas[1,2], Andrew Park [1], and John M. Drake[1]

April 20, 2015

Does parasite taxonomy or host specificity influence predictive accuracy?

We investigated if model predictive accuracy was biased by either parasite specificity (number of hosts), or type (taxonomic group). We found no evidence that parasite taxonomic group influenced predictive accuracy for models trained on host traits, parasite community structure, geographic variables, or the full model (Figure 1). Further, there was no percievable difference in the relative contributions of each predictor variable across parasites of different taxonomic group (Figure 3). Lastly, parasite specificity, defined as the number of unique hosts a single parasite species infects, only slightly influenced model predictive accuracy (Figure 2). Predictive accuracy became more variable with decreasing parasite specificity, but the mean accuracy remained unchanged over the range of parasite specificity values.
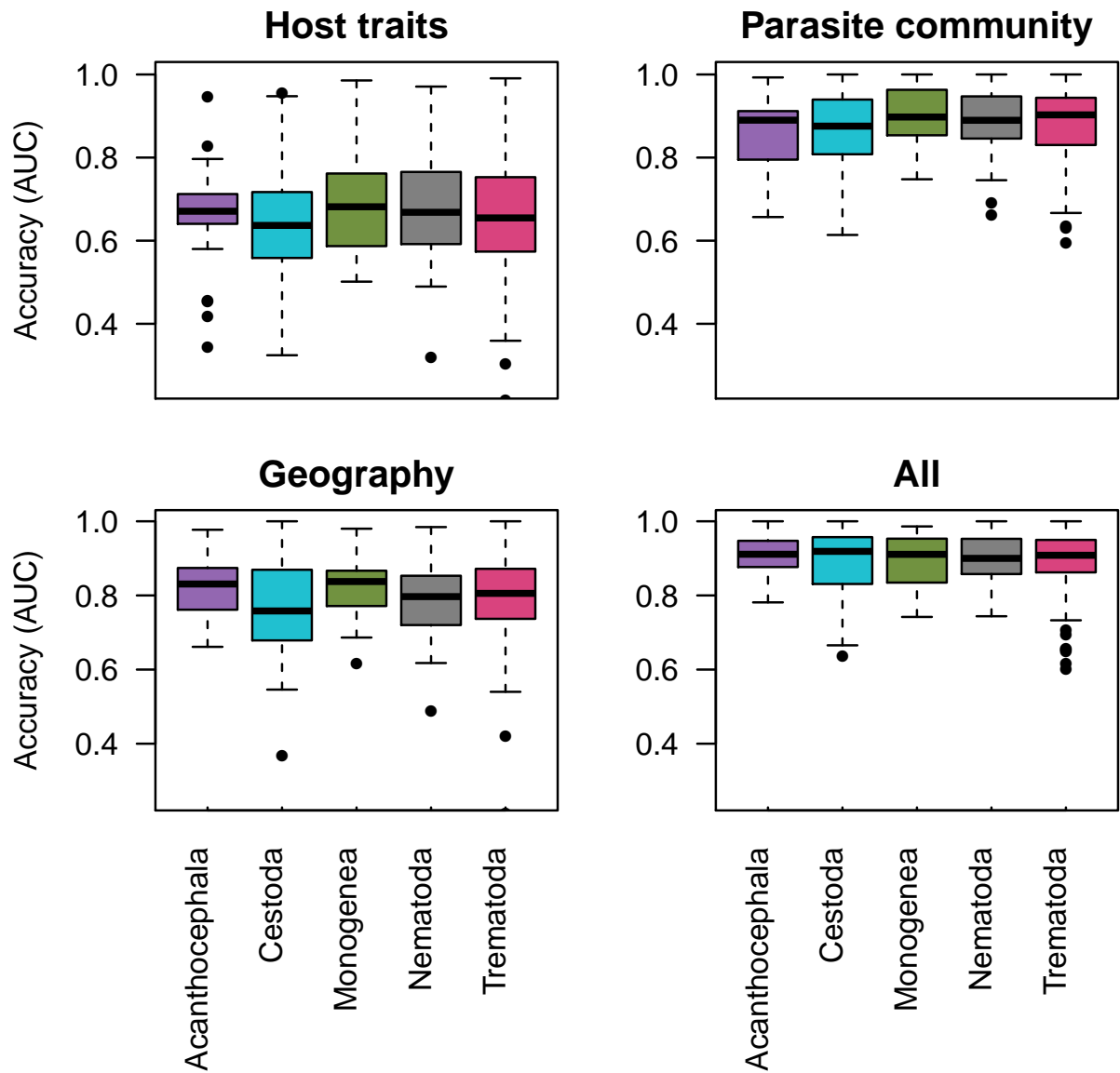
Figures

Figure 1: Accuracy (Area under Receiver operating characteristic (ROC) curves) for boosted regression models trained using host traits (top left), parasite community similarity (top right), geographic variables (bottom left), and all available data (bottom right) as a function of parasite type (x-axis).
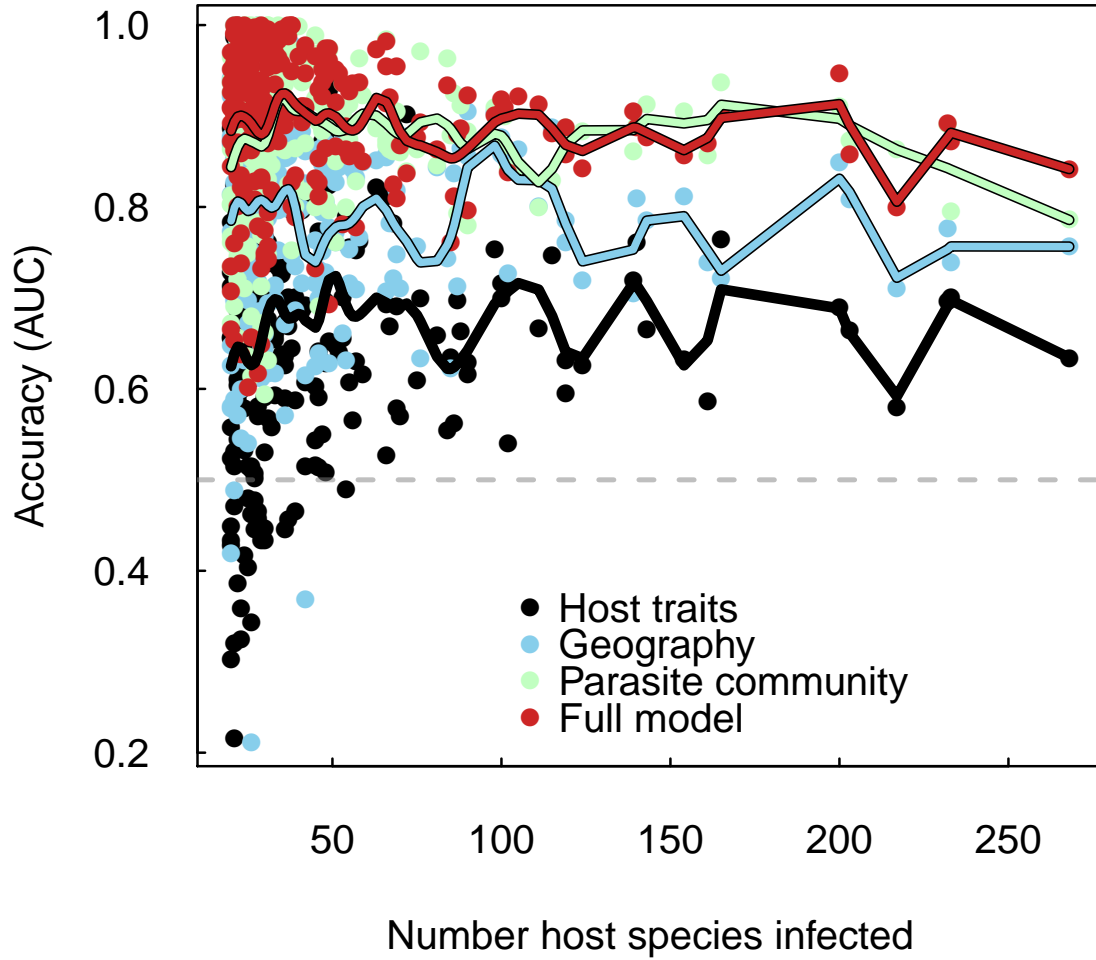
Figure 2: Accuracy (Area under Receiver operating characteristic (ROC) curves) for boosted regression models trained using host traits, parasite community similarity, geographic variables, and all available data as a function of the number of hosts the parasite species was collected on (x-axis). This host range estimate was uncorrelated to accuracy, though some models trained on host traits (black dots) performed poorly when parasite species were specific to a smaller number of host species. Lines correspond to fitted splines (spar = 0.5, to restrict possibility of "oversmoothing").
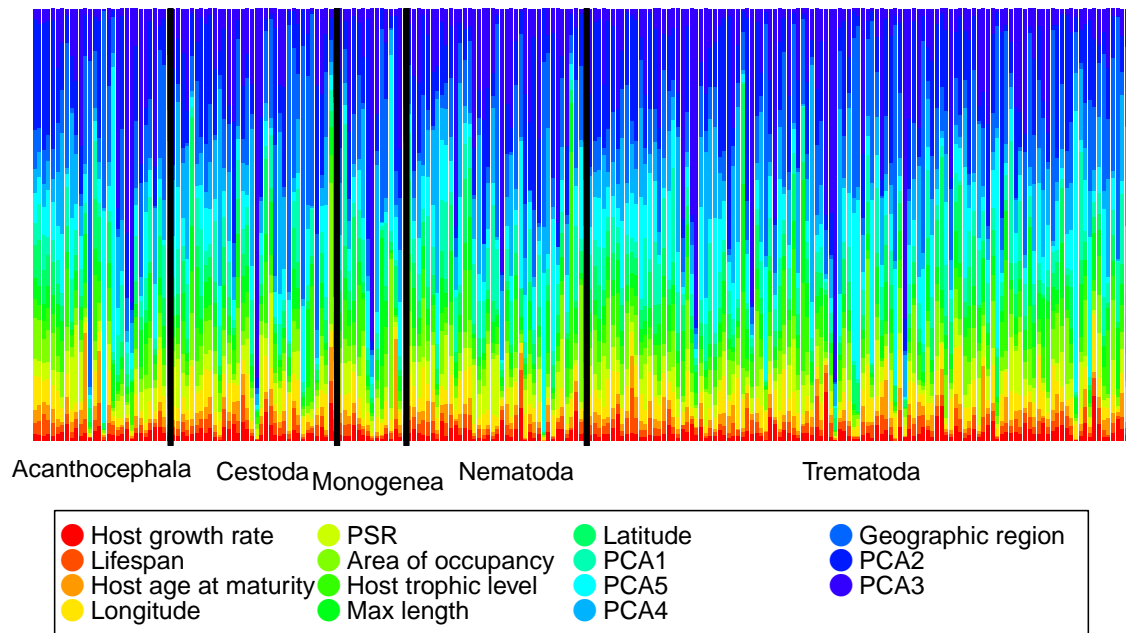
Figure 3: Variation in variable importance as a function of parasite type. Each column represents a model trained on occurrence data for a single parasite species. Variables are sorted by their mean relative importance values across parasite species, but these values vary among parasite species.