

Title: What controls the range of hosts a fish parasite infects?

Authors: Tad Dallas^{1,2}, Andrew Park¹, and John M. Drake¹

Affiliations:

1. University of Georgia, Odum School of Ecology, 140 E. Green Street, Athens GA, 30602.

2. Corresponding author: tdallas@uga.edu

Abstract

Keywords

Introduction

Knowledge gap What constrains the range of hosts that a parasite can infect? Is there a simple range of host functional traits that can determine the likelihood that a parasite infects a given host species? How well can we predict parasite occurrences given *only* host life history traits? Does the importance of different host functional traits differ with parasite type?

Thesis Here, I apply a series of predictive models in order to predict parasite occurrence across a range of potential host species for a large set of parasites of freshwater fish, using host functional traits, and geographic location. I examine the role of parasite group (Acanthocephala, Cestoda, Monogenea, Nematoda, Trematoda) on model performance, and on the relative contribution of different variables to prediction.

Methods

Data and processing We use an existing global database of fish-parasite associations [?] consisting of over 38000 parasite records spanning a large diversity of parasites (Acanthocephala, Cestoda, Monogenea, Nematoda, Trematoda). In order to allow for

cross-validation and accurate prediction, we constrained our analyses to parasites with a minimum of 20 host records. In other words, we only examined parasites that had been recorded more than 20 times, but these occurrences could be on fewer than 20 host species. The inclusion of duplicate occurrences was only permitted if the parasite was recorded on a host in a different geographic location, based on latitude and longitude values. Our response variable was parasite occurrence (binary), and was predicted using only host life history traits, and geographic location of host capture. Host trait information was obtained through the FishPest database [??], and FishBase [?]. Host traits descriptions are provided in Table 1.

Predictor variables Areas of occupancy were calculated as follows: for each species, we plotted all available point records on a global grid of 1x1 degrees Lat-Lon and then we counted the number of grid cells where the species is known to occur.

Model formulation We trained a series of models in order to compare predictive performance of different techniques. Each model was trained on 70% of the data, and accuracy was determined from the remaining 30%. We generated background data by randomly sampling host species where parasite i was not recorded. To maintain proportional training data, the number of random samples was selected to be five times greater than the occurrence records.

Models used Discuss null predictions scenario, and then go into other algorithms used (brt, svm, lr, rf)

³⁸ **Results**

³⁹ **Discussion**

⁴⁰ **Acknowledgements**

⁴¹ **References**

Tables

Table 1: Description and units of variables used to predict parasite occurrences.

Variable	Units	Description	Range
Max length	cm	Maximum fish species length	1 – 2000
Trophic level	–	1 + mean trophic level of food items	2 – 5
Age at maturity	years	Age at sexual maturity	0.1 – 34
Life span	years	Estimated maximum age	0 – 145
Growth rate	years ⁻¹	Rate to approach asymptotic length	0.02 – 9.87
Marine	–	Is host found in marine habitat?	binary
Freshwater	–	Is host found in freshwater habitat?	binary
Brackish	–	Is host found in brackish habitat?	binary
Geographic region	–		7 unique regions
Area of occupancy	–	Area of suitable habitat occupied	
Latitude	degrees		1 – 148 ????
Longitude	degrees		1 – 359 ?????

43 **Figures**

Figure 1:

Figure 2:

Figure 3: