

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

**Authors:** Tad Dallas<sup>1,2</sup>, Andrew Park<sup>1</sup>, and John M. Drake<sup>1</sup>

**Affiliations:**

1. University of Georgia, Odum School of Ecology, 140 E. Green Street, Athens GA, 30602.
2. Corresponding author: [tdallas@uga.edu](mailto:tdallas@uga.edu)

## Abstract

### Take-home messages

1. It is possible to predict parasite niche breadth using either host traits/geographic location, or parasite community similarity (best).
2. Predictive accuracy does not vary as a function of host specificity (though I only consider parasites with 20 or more occurrence records, but these could all be on the same host species).
3. It may be possible to predict parasite spillover from invasive hosts to native host communities, or to predict biotic resistance of a community to invasion.

## Keywords

## Introduction

What has been done in fish parasite stuff? Cite some Kennedy, some Poulin, and then talk about what Strona has done with regards to predicting parasite co-occurrence. Point out that his predictive power wasn't great, and that co-occurrence is a different question from predicting parasite niche breadth. Discuss the differences in my approach from Strona's approach, and set up the question as both a prediction question, and a "what's important" question. Specifically, I trained models on geographic variables and

host traits, and on just parasite information. Parasite information performed better, regardless of geographic information exclusion. This suggests that patch quality and geographic location may not matter as much as the existing community of parasites you harbor. In other words, prediction is better when only considering the parasite community of a host, rather than anything about the host.

**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? How about using solely information on parasite community structure?

Does the importance of different host functional traits or parasite community information differ with parasite type? (supplement)

Since geographic variables are important, what if we try to predict parasite niche breadth in a specific biogeographic region? (supplement)

**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, geographic location, and parasite community information. Parasite occurrence on a given host could be constrained by space (geographic location), patch quality (host characteristics), or through interactions with competing parasites (parasite community structure). It's important to note right up front that the importance of parasite community structure cannot be interpreted as evidence for community interactions, as parasites could infect hosts based on their traits, and the parasite community information could just be serving as a proxy for unmeasured host trait variation. However, predicting parasite occurrence based solely on parasite community information does remove some importance of the patch (host), and is easy to sell, as it may be possible to predict spillover of parasites, or the degree of biotic resistance a community offers to a potential invader, simply by having presence-absence data on parasite communities.

## Methods

**Data and processing** We use an existing global database of fish-parasite associations [Strona et al., 2013] 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 [Strona and Lafferty, 2012, Strona et al., 2013], and FishBase [Froese and Pauly, 2010]. 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)

Or do I focus on BRT? BRT is a bit old hat. Perhaps I could report on BRT, but include analyses with SVM, LR, and RF in the supplement?

## 73 Results

## 74 Discussion

## 75 Acknowledgements

## 76 References

77 Rainer Froese and Daniel Pauly. Fishbase, 2010.

78 Giovanni Strona and Kevin D Lafferty. FishPEST: an innovative software suite for fish  
79 parasitologists. *Trends in parasitology*, 28(4):123, 2012.

80 Giovanni Strona, Maria Lourdes D Palomares, Nicolas Bailly, Paolo Galli, and Kevin D  
81 Lafferty. Host range, host ecology, and distribution of more than 11 800 fish parasite  
82 species: Ecological archives e094-045. *Ecology*, 94(2):544–544, 2013.

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

<b>Variable</b>	<b>Units</b>	<b>Description</b>	<b>Range</b>
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 <sup>-1</sup>	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 ?????

## 84 **Figures**

Figure 1: Image plot with brt model results sorted by parasite type.

Figure 2: BRT RC contributions for each variable, or some kind of plot comparing the results from models trained with host traits/geography, and those trained with parasite community measures?

Figure 3: