Effects of Incorporating Genetic Differentiation and Competition on Species Distribution Models Under Climate Change

Anna Calderon

Matthew Lau

Harvard Forest

**Abstract**

It will be vital to enhance the predictability of species distribution models to better gauge species response to changing conditions and better asses the shifting distributions of species. However, most species distribution models do not consider intraspecific genetic variation or competition. Including genetic heterogeneity and intraspecific competition of a species into ENMs may result in a model that more accurately predicts current and future species distributions. We propose developing two conceptual models—an ENM and a gENM that will account for genetic differentiation and competition within populations—in order to compare the differences in predictability. We anticipate two primary results: we expect to see that including intraspecific variability and competition will enhance our model’s predictability, climatic suitability ranges will be more pessimistic when intraspecific competition is included. If our anticipated results agree with the hypothesis, then future applications of species distribution models are encouraged to incorporate intraspecific differentiation and competition theory into their models.

**Introduction**

Species distribution models are tools used to spatially visualize and predict the presence of a species based on correlative variables such as geography and environmental variables. However, there are many limitations in regards to species distribution models such as the exclusion of biotic and abiotic interactions (Oney et al., 2012). Recently, the idea of including genetic differentiation in ENMs has taken root. It is now widely accepted that a species’ ability to occupy a wide range of habitats largely depends on genetic modifications within populations optimized by natural selection (Antonovics 1971). Despite this recognized phenomenon, few studies have incorporated intraspecific genetic variation into their models. Studies have shown correlations between genetic differentiation within a species and geographical heterogeneity as a result of natural selection (Linhart and Grant, 1996). To a lesser extent, other studies also argue that predicting species distributions is dependent on species interactions--such as the local competition and demographic variables (Clark et al., 2011).

It has been documented that including intraspecific genetic differentiation may help improve the predictability of ENMs under climate change. For example, Valladares et al. (2014) found that by accounting for dispersal limitations and population differentiation, conventional ENMs underestimated the changes in distributions under climate change. Similarly, Oney et al.  (2013) found that conventional SDMs underestimated habitat expansions and overestimated habitat range contractions as opposed to models which included intraspecific variability.  Garzón et al., similarly found that including local adaptation and plasticity dampens the reduction in species range distribution under projected climate change conditions. Other studies report that working with localized occurrence data increases the accuracy of ENMs (Marcer et al., 2016). Little to no research has been done on the effects of competition on the predictability of a species distribution model.

We propose developing two conceptual models involving the use and exclusion of intraspecific genetic variation and competition on arbitrary populations projected thirty-four years into the future. Competition will also be simulated among genetic clusters. Furthermore, we plan to utilize both of these models to compare the differences in predictability by including and excluding intraspecific theory. Finally, we propose an analogous model using real data on different species of Aphaenogaster. [End this section broadly]

**Methods**

Environmental Niche Model

Using data provided by the Global Biodiversity Information Facility, we retrieve record data of a species of Aphaenogaster. Due to the lack of absence data, we will generate pseudo-absence points following the protocol in Yoder 2013. We will overlay both presence and absence data on a map and overlay climate variables using the WORLDCLIM database.

In order to build our species distribution models we plan to use MaxEnt and GLM algorithms, since the competing algorithm—MaxLike—uses presence-only data. There are some limitations to using MaxEnt, however. Gotelli and Stanton-Geddi report that MaxEnt models poorly estimate the range expansions and contractions under climate change projections (Gotelli and Stanton-Geddi, 2015). Our decision to run simulations using generalized linear models also stems from Gotelli and Stanton-Geddi, who report that GLMs seem to accurately predict probability of occurrence when working with presence and absence data.

Genetic Environmental Niche Model

Due to the lack of genetic data, we will employ an equation that calculates the genetic distances in order to make genetic clusters. These genetic distances will be a function of environmental, geographic, and spatial variables. We are including geographical and spatial variables because geography and location often determines gene flow (Cushman and McRae, 2015), and thus determine how our species may be genetically clustered.  Using the GRAINSCAPE package in R (Galpern, 2013), we will be able to simulate what barriers species may encounter and include that information into our equation to produce our clusters. Using our genetic cluster, we will project our species distribution model thirty-four years into the future using the similar protocol from Valladares et al. (2014).

Intraspecific Competition

Using both species distribution models (ENM and gENM), we will overlay the maps and highlight any areas where two or more populations overlap. We will determine which population has better competitive abilities to estimate what population would predominate in the area of overlap.

R-Data Tracker and Data Provenance

* Data Provenance provides useful meta-data about “where” a data object has come from (i.e., what ways data has been altered as its been processed by various functions)
* It has not been widely tested in the realm of ENM or simulation studies, even though such projects use or produce multiple data streams.
* Our gENM study provides both an excellent application of Data Provenance to be used for developing simulation studies by using the debugging and code optimization functionality of RDataTracker (a Data Provenance software package).

Timeline