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 to predict the effects of including genetic variability and intraspecific competition will have on each population. We anticipate two primary results: we expect to see that including intraspecific variability and competition will enhance our model’s predictability by increasing our model’s AUC values. Second, we expect that climatic suitability ranges depend on the combination plasticity and competitive abilities of each population. In general, we expect to find that populations portraying high amounts of plasticity and aggressive competitive behaviors will experience the high range expansions. If our results agree with our expectations, then we encourage future research 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 with regards to species distribution models such as the exclusion of biotic and abiotic interactions (Oney et al., 2012). 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 competition and demographic variables (Clark et al., 2011).

It has been documented that including intraspecific genetic differentiation may 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. (2011), 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).

We propose two conceptual models with and without intraspecific genetic variation and competition projected thirty-four years into the future. Competition will also be examined among genetic clusters. Furthermore, we plan to utilize these models to compare the differences in their predictability by exploring AUC values and estimating differences in area between models.

**Methods**

Environmental Niche Model

Using data provided by the Global Biodiversity Information Facility, we will 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. MaxEnt is the most commonly used algorithm for species distribution models; 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 develop and apply an equation that calculates 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. We will then use the STRUCTURE package to create genetic clusters. Using our genetic clusters, 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.

Data Provenance

Because we will use data from multiple data streams, we plan to use R Data Tracker to help us keep track of our data sources and and track alterations to the data sets. R Data Tracker is a Data Provenance software package that is able to track what functions have been applied to the data set, which makes it a powerful tool when considering the reproducibility of a study. Moreover, it is particularly useful to this project as little to no ENM studies have used this tool. Here, our study provides an excellent opportunity to show the usefulness of R Data Tracker by using the debugging and code optimization functions. We hope that including Data Provenance in our study will also inspire other research to implement this useful tool.

Timeline

Week four and part of week five will be dedicated to learning how to use the GRAINSCAPE and STRUCTURE packages in order to make our genetic clusters. The remaining days of week five, week six will be set aside to incorporating the genetic clusters into our species distribution models using the protocol from Valladares et al. (2014). From week seven to week eight, time will be spent incorporating intraspecific competition into our model. Week nine will also be dedicated to analyzing the differences ENM

../../time%20line%20w4:5.pdf

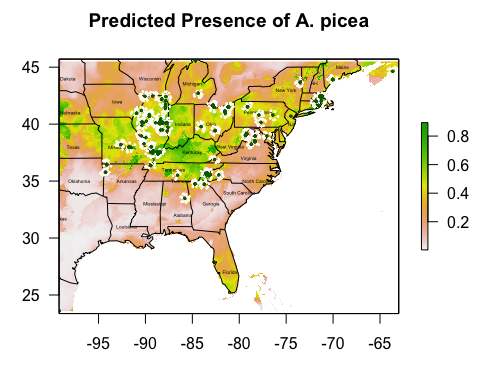
../../Presentation1.pdf

**../../Presentation1.pdf**

**Results**

Preliminary Results

Environmental niche modeling on A. picea shows the future distribution of the species under climate change scenarios projected thirty-four years into the future. Figure 1 shows that A. picea is more likely to occur between latitudes 35 to 42.5 and longitudes -95 to -70. However, the species distribution model does not consider genetic clusters



**Figure 1a** shows the species distribution of *Aphaenogaster picea*. The green colors represent high occurrence probability while the warmer and lighter colors represent lower occurrence probability.

Expected Results

We expect that including genetic clustering and intraspecific competition will increase our model’s ability to predict occurrence probability and habitat suitability. Specifically, we expect to see greater AUC values in comparison to our conventional ENM. We expect that our gENM will have higher predictability since the data will not be homogeneous, and dispersal limitations will be accounted for by the GRAINSCAPE R package.

We also expect that phenotypic plasticity and competitive abilities among populations will play a major role in determining the distributions of populations. Under competitive exclusion, we expect to see that a combination of high amounts of plasticity and high competitive abilities within a population, will result in the highest range expansions. Conversely, a combination of low amounts of plasticity and low competitive abilities will result in the highest range contractions. For example, for a scenario with high margin plasticity and high margin competition, we would expect populations that lie anywhere between that spectrum to show range contractions. Figure 2 a shows that the green population, which does not compete aggressively and has the least amount of plasticity, is expected to experience compaction. On the other hand, when central plasticity and competition is high (figure 2b), we expect the marginal populations to experience range contractions. However, what population dominates what area will depend on the combination of plasticity scenarios and competition.

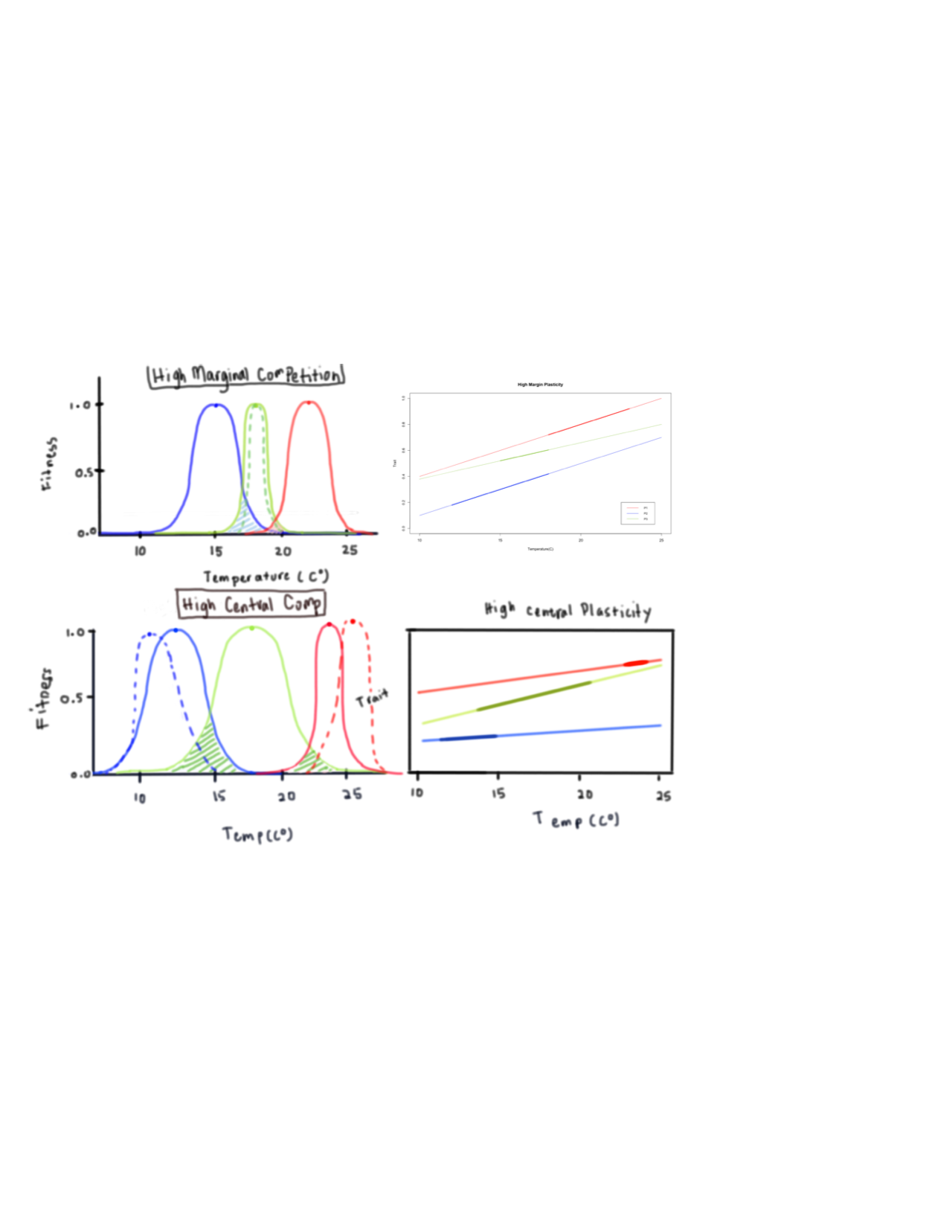


Figure 2a (top), 2b (bottom) shows different plasticity scenarios and fitness curves. Each color represents a single population. The graphs on the right are the plasticity of traits, such as tolerance, plotted against increasing temperatures, and the thick portion of the lines represent the range of temperatures where the population is most fit. The graphs to the left show each population’s fitness and the dotted curves represent how fitness is expected to change when intraspecific competition is included.

**Discussion**

We believe that by including intraspecific variation, in the form of genetic clustering, and intraspecific competition will enhance the predictability of our species distribution models. We plan to achieve this by using GRAINSCAPE and STRUCTURE packages to create our genetic clusters, which will then be projected thirty-four years into the future and under climate change scenarios. We expect range contractions and expansions to be dependent on the combination of amounts of plasticity and competitive aggressiveness of each population. If the results are consistent with our conceptual models, future research using SDMs is encouraged to include intraspecific variation and to consider variables that may lead to range expansions and contractions, such as competition and dispersal.