

# **Investigating the Role of Phylogenetic Diversity in Species Invasions**

R. Kirsten Tyler  
Ecological Modeling  
December 11, 2017

## **Introduction**

Invasive species are non-native organisms that are introduced by humans and spread throughout a range, to the detriment of the recipient ecosystem (“COP,” 2002). The adverse consequences of plant invasions are vast, including but not limited to alteration of ecosystem functions such as nutrient cycling, fire regime, and hydrology, as well as threats to biodiversity (Mack et al., 2000). For these reasons, it is important for scientists to identify the factors related to invasions. What attributes or processes make some plant invasions successful while others are not?

The exact mechanisms that facilitate plant invasions remain unclear at this point (Gallien & Carboni, 2016; Rejmánek, 2014; Simberloff et al., 2013). Charles Elton’s (1958) original foci of biological invasions such as biotic resistance (Levine & D’Antonio, 1999; Tilman, 2004), enemy release (Keane & Crawley, 2002), and disturbance (Davis, Grime, & Thompson, 2000; Hobbs & Huenneke, 1992) are still important concepts however they don’t consider evolutionary processes that might aid in invasions. Recently, there has been growing interest in the evolutionary histories of the invader and the invaded recipient community but most research has tended to focus on the phylogenetic relatedness of the invader to the community in which it invades (i.e. Darwin’s naturalization hypothesis (Darwin, 1859; Strauss, Webb, & Salamin, 2006)). The evolutionary imbalance hypothesis is a framework that suggests a Darwinian view of invasion ecology in that pre-adaptations arise in species from phylogenetically rich origins

(Fridley & Sax 2014). Darwin (1859) observed that species more diverse in structure, constitution, and habits will have a greater ability to occupy a wide range of habitats. Species living in stable, consistent conditions with intense competition for extended periods of time would be better suited to novel habitats (Darwin 1859).

These observations are difficult to measure but previous studies have used the proxy of phylogenetic diversity (PD) to characterize the evolutionary history in an area (Gerhold et al. 2011; Fridley & Sax 2014; Whitfeld 2014). PD can be described as the sum of all branch lengths of a phylogenetic tree and represents divergence times between species, inferring the number of unique lineages in a region (Faith 1992; Vellend et al. 2010). This could be reflective of the competitive intensity, environmental consistency, and adaptations accumulated by the species in question, allowing us to use PD as an index of unique phylogenetic lineages in each region. Therefore, if introduced species had origins in phylogenetically rich regions (high PD), we would expect to see them become successful invaders in regions lacking phylogenetic diversity (low PD). Previous studies found that regions with higher PD were more likely to produce invasive species and that regions with lower PD were more likely to be invaded (Gerhold et al. 2011; Fridley & Sax 2014; Whitfeld 2014). However, these studies were often too broad or sample sizes were too small to extrapolate results or make generalized conclusions. Further investigation is needed to understand how the success of plant introductions is affected by the PD of origin and recipient regions.

The purpose of this study is to expand upon the evolutionary imbalance hypothesis and strictly test this theory to understand the role of phylogenetic diversity on the probability of introduced species becoming invasive in a novel region. We focused on 751 invasive woody plants to evaluate their global invasion success and explain it using the differences in PD

between origin regions and introduced regions, or the evolutionary imbalance. We hypothesize that species with origin regions of greater phylogenetic diversity will be more likely to invade regions with less phylogenetic diversity.

## **Methods**

### **Data Collection**

A comprehensive global database of invasive trees and shrubs was used in this study (Rejmanek & Richardson, 2013). It includes 751 invasive woody species, their native regions, and 15 introduced regions. For each species introduced, the database shows where the plant has become invasive and where it has not regarding the 15 introduced regions. A map of 35 global floristic regions (Takhtajan, 1986) was used to assign each species an origin region. The angiosperm flora (using taxonomic level of family) of each origin region was used to build a phylogenetic tree. Then using Faith's (Faith, 1992) method for calculating phylogenetic diversity, a PD value was assigned to each plant species and to the 15 introduced regions.

To compare the PD values between introduced and origin regions, the difference between them was calculated and added to the dataset ( $\Delta PD = \text{Origin\_PD} - \text{Intro\_PD}$ ); this serves as the evolutionary imbalance of the species in question and the regions where they are introduced. The dataset was transformed from wide-format to long-format using the melt function in the reshape2 package in R and NA's were replaced with zeros for this analysis ("R Core Team," 2016).

### **Model Choice and Refinement**

To predict the probability of becoming invasive in an introduced region, we used a binomial generalized linear model to analyze the data using JAGS and the Rjags package in R because the response variable is binary, in that it represents "yes invasive" or "no invasive" for

each introduced region (see model information / code in supplementary material). The parameter being predicted is a single probability of becoming invasive (the “yes” outcome) in an introduced region. To do this, we set up the predictor function as a linear model dependent on the fixed-effects parameter of evolutionary imbalance (Delta\_PD). The logit link function constrains probability values between 0 and 1. We used the Bernoulli distribution as the likelihood function for our analysis because it is a discrete distribution that specifies the probability of binary data, given the values of parameters. To express prior beliefs over theta (in the interval  $[0, 1]$ ), we used a beta distribution because it is conjugate to the Bernoulli likelihood function. In our model string, we used beta priors that were normally distributed with low precision and high variance. The model string was written to the working directory.

To call the Jags model, we entered the pertinent data into a list. The predictor variable was Delta\_PD (difference of PD between origin region and introduced region). The y variable was “Inv” (Invaded). We used four chains with various initial starting values that represented different priors because we didn’t have strong prior knowledge. To test for efficiency and adjust when necessary, we instructed Jags to adjust and adapt for 100 time steps. To update the model, we added a burn-in period of 1000 steps.

We collected posterior samples using the coda package. To assess if the chains were producing representative samples from the posterior distribution, we used a density plot which shows the smoothed histograms of parameter values that are sampled in each chain. We used the Gelman-Rubin statistic (Gelman & Rubin, 1992) to assess for chain convergence. This check estimates the variance between chains compared to variance within chains. We performed all analysis in R and R Studio (“R Core Team,” 2016; RStudio Team, 2015).

## **Results**

We observed that Delta\_PD predicted the probability of becoming invasive in an introduced region. Our fixed effect variable of Diff\_PD was significant in explaining the probability of becoming invasive. The estimate of Diff\_PD was observed as -0.4997, so a one unit increase in Diff\_PD can be associated with a -0.4997 decrease in the expected log odds of becoming invasive. This was an unexpected result and requires further investigation.

The summary statistics associated with the model output are found in table 1.

*Table 1: Summary statistics for JAGS GLM model*

Iterations = 1101:6100					
Thinning interval = 1					
Number of chains = 4					
Sample size per chain = 5000					
1. Empirical mean and standard deviation for each variable, plus standard error of the mean:					
	Mean	SD	Naive SE	Time-series SE	
beta.0	-1.6705	0.02856	0.000202	0.0002932	
beta.1	-0.4997	0.25256	0.001786	0.0026463	
2. Quantiles for each variable:					
	2.5%	25%	50%	75%	97.5%
beta.0	-1.7271	-1.6895	-1.6704	-1.6513	-1.615268
beta.1	-0.9917	-0.6702	-0.5002	-0.3278	-0.002409

We analyzed diagnostics using the coda, mcmcplots, and lattice packages in R. To assess whether the model converged, we used the Gelman-Rubin diagnostic (Gelman & Rubin, 1992), shown in figure 1. Visually, the plot suggests that the model did converge. Given that the potential scale reduction factors were all less than 1.1 for both parameters, we concluded that the chains converged. In addition, we looked at the overlaid density plot (figure 2) which shows the smoothed histograms of parameter values that are sampled in each chain.

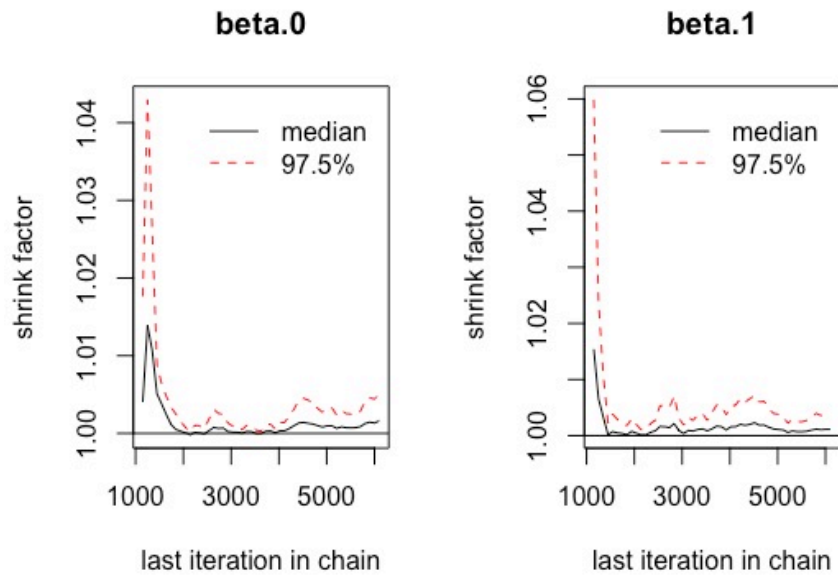


Figure 1: Gelman-Rubin diagnostic plots

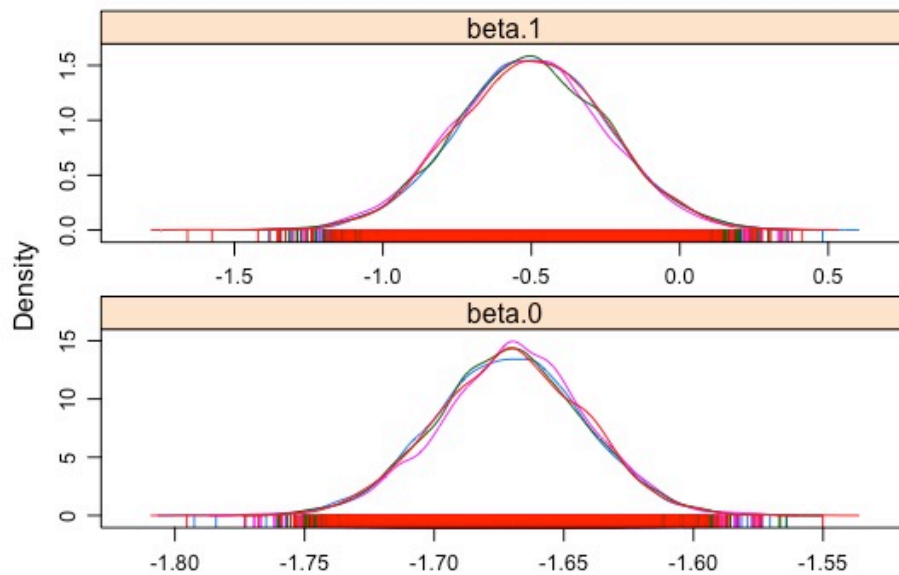


Figure 2: Density plot (overlaid)

## Discussion

We expanded upon the evolutionary imbalance hypothesis in this study to strictly test this theory and understand the role of phylogenetic diversity on the probability of introduced

species becoming invasive in a novel region. We focused on 751 species and 15 introduced regions and compared the PD of the origin regions to the PD of the introduced regions to quantify evolutionary imbalance. We hypothesized that species with origin regions of greater PD would be more likely to invade regions with a lower PD value. We found that PD plays a role in the probability of an introduced species becoming invasive a new range. However, we did not find that this was the case. We observed a negative relationship between the difference in PD and the probability of invasion. Further investigation is necessary to fully understand the effects of the variance between regions. We feel that another analysis (a GLMM) that considers the mixed effects of different regions is essential to understand the role of PD in predicting invasiveness.

Evolutionary imbalance cannot be considered as the only factor in invasiveness, by any means. In fact, this study had many limitations and further investigation is needed on this front. There are many factors involved that could explain invasiveness of a species and we are suggesting that evolutionary imbalance is one of them. In the future, we plan to expand this model to include many other factors such as propagule number, genome size, origin habitat range size, as well as others to describe in more detail how a species becomes invasive in a novel range.

## **References:**

- “COP” (2002). Review and consideration of options for the implementation of Article 8(h) on alien species that threatens ecosystems, habitats or species. In *Conference of the parties to the Convention on Biological Diversity (COP)*.
- Darwin, C. (1859). *On the Origin of the Species*. Darwin (Vol. 5).
- Davis, M. A., Grime, J. P., & Thompson, K. (2000). Fluctuating resources in plant communities:

- A general theory of invasibility. *Journal of Ecology*, 88(3), 528–534.
- Elton, C. S. (1958). *The ecology of invasion by animals and plants*. London, UK: Methuen.
- Faith, D. P. (1992). Conservation evaluation and phylogenetic diversity. *Biological Conservation*, 61(1), 1–10.
- Fridley, J. D., & Sax, D. F. (2014). The imbalance of nature: Revisiting a Darwinian framework for invasion biology. *Global Ecology and Biogeography*.
- Gallien, L., & Carboni, M. (2016). The community ecology of invasive species: where are we and what's next? *Ecography*, (October), In press.
- Gelman, Andrew; Rubin, Donald B. Inference from Iterative Simulation Using Multiple Sequences. *Statist. Sci.* 7 (1992), 4, 457--472.
- Hobbs, R. J., & Huenneke, L. F. (1992). Disturbance, Diversity, and Invasion: Implications for Conservation. *Conservation Biology*.
- Keane, R. M., & Crawley, M. J. (2002). Exotic plant invasions and the enemy release hypothesis. *Trends in Ecology and Evolution*.
- Mack, R. N., Simberloff, D., Mark Lonsdale, W., Evans, H., Clout, M., & Bazzaz, F. A. (2000). Biotic invasions: causes, epidemiology, global consequences, and control. *Ecological Applications*, 10(3), 689–710.
- “R Core Team.” (2016). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.
- Rejmánek, M. (2014). Invasive trees and shrubs: Where do they come from and what we should expect in the future? *Biological Invasions*, 16(3), 483–498.
- Rejmanek, M., & Richardson, D. M. (2013). Trees and shrubs as invasive alien species — 2013 update of the global database . Divers database, (August).



RStudio Team. (2015). RStudio: Integrated Development for R. Boston, MA.

Simberloff, D., Martin, J.-L., Genovesi, P., Maris, V., Wardle, D. A., Aronson, J., Shlisky, A.

(2013). Impacts of biological invasions: what's what and the way forward. *Trends in Ecology & Evolution*, 28(1), 58–66.

Strauss, S. Y., Webb, C. O., & Salamin, N. (2006). Exotic taxa less related to native species are more invasive. *Proceedings of the National Academy of Sciences of the United States of America*, 103(15), 5841–5.

Takhtajan, A. (1986). *Floristic Regions of the World*. Los Angeles, CA: University of California Press.