

Ancestral niche reconstruction and tests for conservatism

Shubhi Sharma

December 2020

Introduction

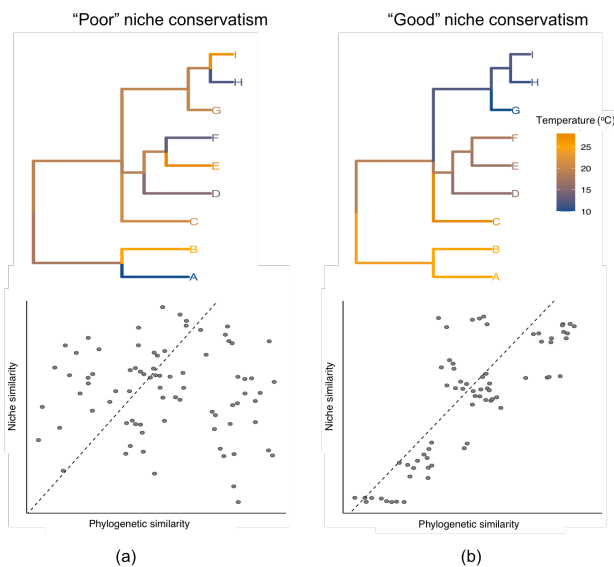


Figure 1: The relationship between phylogenetic and niche similarity in a case where the niche is not conserved (a) and where it is (b). In this example figure, the temperature is mapped onto a monophyletic clade. Niche conservatism determines the environmental conditions a species can tolerate and therefore which regions they can disperse into. By contrast, niche evolution allows members of a clade to adapt and disperse to new environmental conditions.

members of a clade through phylogenetic descent, niche evolution operates on individual species. Thus, changes in niche breadth in one species in one part of the range of a clade might only have a limited impact on the overall distribution of the clade yet a significant impact on the one species. The niche conservatism constrains species to their ancestral niche whereas niche evolution allows species to inhabit new environmental spaces. In this way, the strength of phylogenetic niche conservatism can be a crucial factor in explaining large-scale patterns of species

A fundamental niche describes the abiotic conditions in which a species is able to survive and reproduce (Hutchinson, 1957). This acceptable niche space can be determined by intrinsic traits such as physiology and can be maintained over evolutionary-relevant timescales (Donoghue and Wiens, 2004). *Niche conservatism* is the tendency of a species to retain part of its fundamental niche (Wiens and Graham, 2005). Through phylogenetic descent and shared traits, species will always inhabit similar environments to their close ancestors. The extent to which this phenomenon occurs is of critical importance to understanding the limits of species distributions.

A persistent challenge in biogeography is to explain why species have been able to disperse to some areas but not others- climatic niche specialization has been put forth as a possibility (Smith *et al.*, 2007, Wiens *et al.*, 2007, Li *et al.*, 2009). Phylogenetic niche conservatism determines which environmental conditions a clade can inhabit. However, this generalization does not apply uniformly across clades. A contradictory force, niche evolution, facilitates species dispersal outside of ancestral conditions. Even though certain niche characteristics might be shared by all

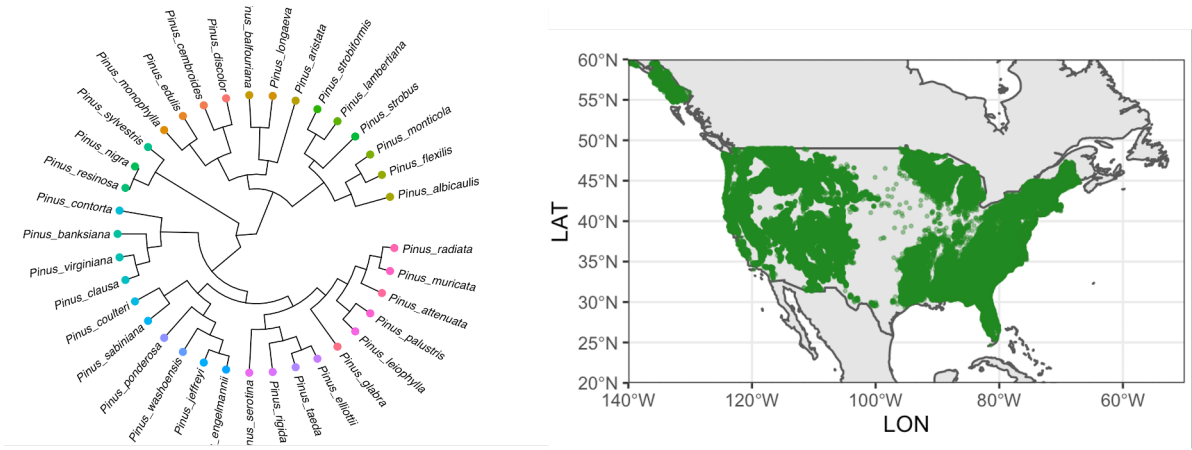


Figure 2: The pine phylogeny used in this analysis and a map of the geographic spread of all pine species in North America

distributions.

Consider the relationship between phylogenetic and niche similarity for a hypothetical species (Figure 1). When niche conservatism signal at the sister-species and clade level is strong, there is a clear linear relationship between niche similarity and the phylogenetic structure - as niche similarity decreases, so does phylogenetic distance (Figure 1b). On the other hand, consider a case of poor niche conservatism or conversely, rapid niche evolution (Figure 1a). The linear relationship between niche and phylogenetic similarity breaks down. There is an inherent relationship between an environment a species exists in and it's phylogenetic descent.

A method for testing for niche conservatism is to fit and compare alternative models of stasis and divergence to traits relevant to species' climatic distributions and thereby test whether phylogenetic niche conservatism is an important ecological factor in influencing the geographic distribution of the species (Kozak and Wiens, 2010). Wiens *et al.*, (2010) fit three models-Ornstein-Uhlenbeck, Brownian Motion and white noise models, comparing the fits amongst them. Each model of evolution supports a different hypothesis - for example, the Ornstein-Uhlenbeck model supports evidence for stabilizing selection or evolutionary stasis and therefore provides strong support for niche conservatism by proving that traits are constrained by some optimum value.

Kozak and Wiens (2010) used white noise, Brownian Motion and Ornstein-Uhlenbeck to model climatic niches of Appalachian salamanders (Plethodontidae) to test for niche conservatism. The study found Ornstein-Uhlenbeck models that assumed separate adaptive optimal for climatic regimes to fit the observed data best therefore providing strong support for niche conservatism in the study clade.

This analysis is an exploratory study testing for abiotic niche conservatism in North American Pine trees. To test for niche conservatism, three abiotic and climatic variables important to the distribution of Pines in North America will be selected and reconstructed. To reconstruct ancestral niches, two models of trait evolution, Brownian Motion and Ornstein-Uhlenbeck, will be tested. To confirm niche conservatism, tests of phylogenetic signal from the structure of the traits will be conducted. The methods of this study are detailed below.

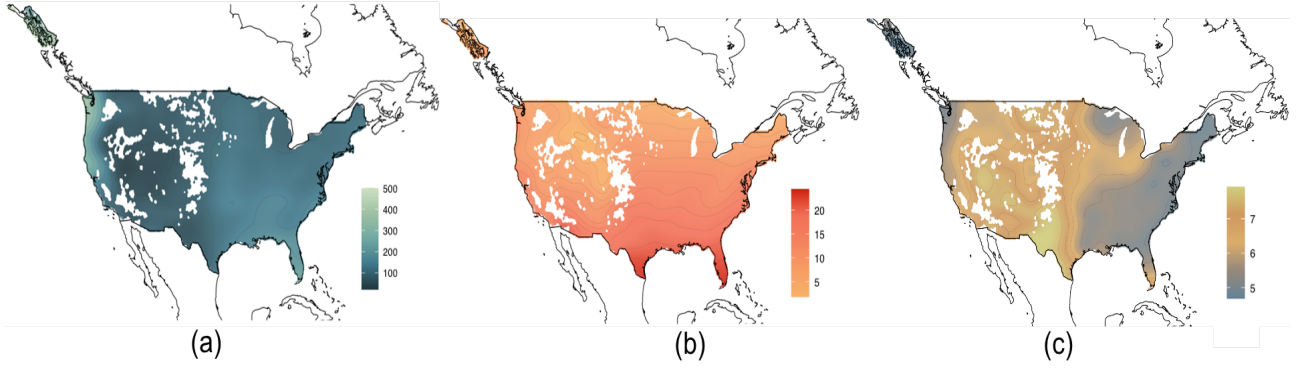


Figure 3: Climate maps for precipitation (mm; left) and temperature (°C middle) and mean soil pH (right) in the United States. Data from WorldClim gridded dataset. The gridded dataset is spatially interpolated using splines.

Methods

Data

Phylogeny and inventory data. Pines (*Pinus*, under Family *Pinaceae*) describe a genus of Coniferous trees in North America. Pines, although found globally, are native to northern temperate regions. In North America, Pines are distributed widely across the country (Figure 2). To characterize the distribution of this genus, inventory data was used in the analysis. The United States Department of Agriculture’s Forest Inventory Analysis (FIA) holds data for periodic forest inventories dating back to 1969. The forest inventory has more than 20,000 plots and more than 347 tree species. While the inventories hold extensive information about each species (diameter at breast height, crown, canopy cover, height etc.), presence/absence was used in this analysis. Each species belonging to genus was collected and plots where measurements for the species existed (did not exist), were marked as ‘presences’ (‘absences’). The phylogeny was taken from a published time-calibrated phylogeny of all tree species in the FIA dataset (Potter and Woodall, 2012). The pine phylogeny was created by pruning the whole phylogeny to species belonging to *Pinus*.

Climate and soil variables. Two climate variables - annual average temperature and precipitation - emerge as important variables for determining and modeling the distribution of tree species (Waltari *et al.*, 2014). Climate data was downloaded from the WorldClim database. In addition to the two climatic variables, a soil layer with average soil pH for the top 4 strata was included in the analysis. This was done to include a variable that is more spatially-variable than simply temperature and precipitation (which are spatially coarse variables; Figure 3). The tips of the phylogeny were annotated by the averages of each of these variables.

Models

Brownian Motion Model. The simplest macroevolutionary model for phylogenetic niche conservatism is that species inherit their niches from their ancestors and then slowly diverge over time. This is essentially the Brownian motion (BM) model of trait evolution. Under this model, traits evolve up the phylogeny via random walk and trait differences accumulate over time (Felsenstein, 1973, 1985).

$$BM \sim N(0, \sigma^2)$$

$$\sigma^2 \sim Unif(L, U) \text{ where } (L, U) \in \mathbb{R}^+$$

where σ^2 is the only random parameter in the model and represents rate of evolution. In this Bayesian set-up of the Brownian Motion, σ^2 is drawn from a uniform distribution with the hyperprior parameters L, U . All models were implemented in RevBayes. To test whether or not niche traits evolve under the BM model, Pagel's λ was estimated (Pagel, 1999). λ is a multiplier of the off-diagonal's of the variance-covariance matrix which best fits the distribution of data at the tips of a given phylogeny. It is estimated using Maximum Likelihood and varies from 0 (traits have no phylogenetic structure) to 1 (phylogeny perfectly explains trait evolution under Brownian process). An alternative to Pagel's λ is Blomberg's K (Blomberg *et al.*, 2003). When K is between 0 and 1, it can be interpreted the same as λ . When K is greater than 1, it indicates that the traits are more similar than would be expected under the BM model.

Ornstein-Uhlenbeck model. A generalization of the BM model is described by the Ornstein-Uhlenbeck (OU) model of evolution. Under this model, the traits are evolving towards an optimum which itself is evolving by a Brownian process.

$$\begin{aligned} OU &\sim N(\alpha\theta, \sigma^2) \\ \sigma^2 &\sim Unif(L, U) \text{ where } (L, U) \in \mathbb{R}^+ \\ \theta &\sim Unif(A, B) \text{ where } (A, B) \in \mathbb{R} \\ \alpha &\sim Exp(\lambda) \end{aligned}$$

θ is the central tendency for the trait with a uniform prior with limits (A, B) . α is the "weight" of this central tendency- it can sometimes be referred to as a 'rubber band' parameter as it determines how much to pull trait evolution back to the central tendency. It has an exponential prior distribution with hyperprior parameter λ . Finally, σ^2 is the rate of trait evolution with uniform prior and hyperprior parameters (U, L) . The OU model provides more support for the phylogenetic niche conservatism hypothesis- this model favors 'stabilizing selection' i.e. evolution occurs by Brownian motion but around a central tendency for a given trait. The Brownian motion is a special case of the OU model, where the central tendency is equal to 0, so there is no expected net change in the trait over time.

Multivariate Relaxed Clock Brownian Motion. A third model was considered in this analysis. As an attempt to model niche evolution and reconstruction as closely, a multivariate Brownian motion with variable branch evolution rates (a relaxed clock) model was developed.

$$\begin{aligned} MBN &\sim MVN(0, t\mathbf{\Sigma}_i) \\ \mathbf{\Sigma}_i &= \sigma_i^2 \mathbf{R} \gamma^2 \text{ where } \mathbf{R} \text{ is the correlation matrix} \\ \sigma_i^2 &= \sigma_r^2 \cdot r_i \\ \rho &\sim LKJ(\eta) \\ \gamma^s &\sim Dir(\alpha) \\ \sigma_r^2 &\sim Unif(L, U) \text{ where } U, L \in \mathbb{R}^+ \\ r_i &\sim mN(\mu_R, \sigma_R) + (1 - m)r_a \\ r_a &\sim logN(a, b) \end{aligned}$$

where the multivariate brownian motion model is modelled as a multivariate normal model with mean zero. The variance-covariance matrix of this model, $\mathbf{\Sigma}$ is factored into \mathbf{R} , the correlation matrix parameterized by ρ , σ_i^2 , the branch-specific rate parameter and γ^2 , the proportional rate of evolution per character trait. Parameters that are indexed by i , are allowed to vary with the branches. The rate multiplier is modelled as a mixture-distribution. The tree is given a prior probability of a shift in rate occurring. The rate multiplier is drawn from a mixture distribution weighted by this probability. A directed-acyclic graph (DAG) represents this model (Supplemental figures).

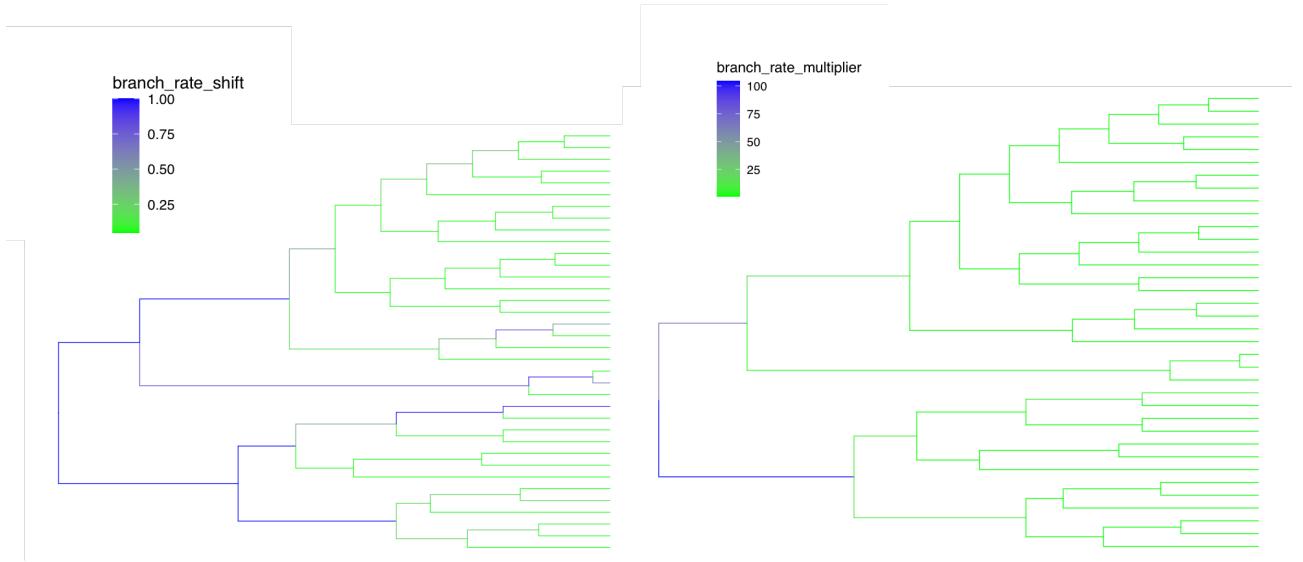


Figure 4: Posterior distribution values for branch rate shift probability m (left) and branch rate multiplier r_i (right). The branch rate shift probability determines how often a change in rate σ_i^2 occurs whereas the branch rate multiplier determines the magnitude of that shift.

Results

Model fits. To compare the fits between BM and OU models, phylogenetic simulations were used. The niche trait data for the phylogeny was simulated under the Brownian motion, then the Ornstein-Uhlenbeck model. Patterns that follow the one-to-one line would indicate a perfect fit under this model. To further test these fits, a log-likelihood ratio test was done. The log-likelihood for BM versus OU ($\log\text{Lik}_{BM}/\log\text{Lik}_{OU}$) for temperature was found to be 15.33, $\log\text{Lik}_{BM}/\log\text{Lik}_{OU}$ for precipitation 197.06 and finally for soil pH 12.43. All log-likelihood ratio tests suggest a preference for Brownian motion for all three variables over Ornstein-Uhlenbeck models (scatter plots between simulated and observed data are in the supplemental figures).

Phylogenetic signal tests. Two tests for phylogenetic signal were conducted - Pagel's λ and Blomberg's K . Both statistics were tested for statistical significance by the 0.05 and 0.01 significance levels. All tests were found to be significant. For temperature, $\lambda = 0.64$ ($p < 0.05$), $K = 0.54$ ($p < 0.05$), precipitation, $\lambda = 0.74$ ($p < 0.05$), $K = 0.629$ ($p < 0.01$) and finally for soil pH, $\lambda = 0.078$ ($p < 0.01$), $K = 0.64$ ($p < 0.01$). From these results, we find a strong phylogenetic signal in the distribution of the traits and further, a support for trait evolution under the Brownian Motion model.

Posterior distributions. The posterior distribution for probability of rate shifts indicates the branches where a shift in rates are likely to occur (Figure 4). We see that shifts are likely to occur in some branches but not in most. This doesn't tell us anything about the size of the shift. Therefore, the posterior distribution for the rate shift multiplier indicates the magnitude of the shift. We see that the two diverging branches leading into this clade have the highest magnitude of shift (Figure). This tells us that two sub-clades of *Pinus* might be different but there is not a lot of variation within them.

Reconstructions. Reconstruction from the Brownian motion model for each variable, temperature, precipitation and soil pH are shown in Figure 5.

From Figure 5, we can visually see a distinction in the two sub-clades, where one with *Pinus longaeva*, *Pinus monophylla*, *Pinus edulis*, *Pinus cembroides* and *Pinus discolor* occupies a high soil pH, low temperatures and low precipitation. A second sub-clade including *Pinus glabra*, *Pinus attenuata*, *Pinus radiata*, *Pinus palustris*, *Pinus rigida*, *Pinus taeda* and *Pinus*

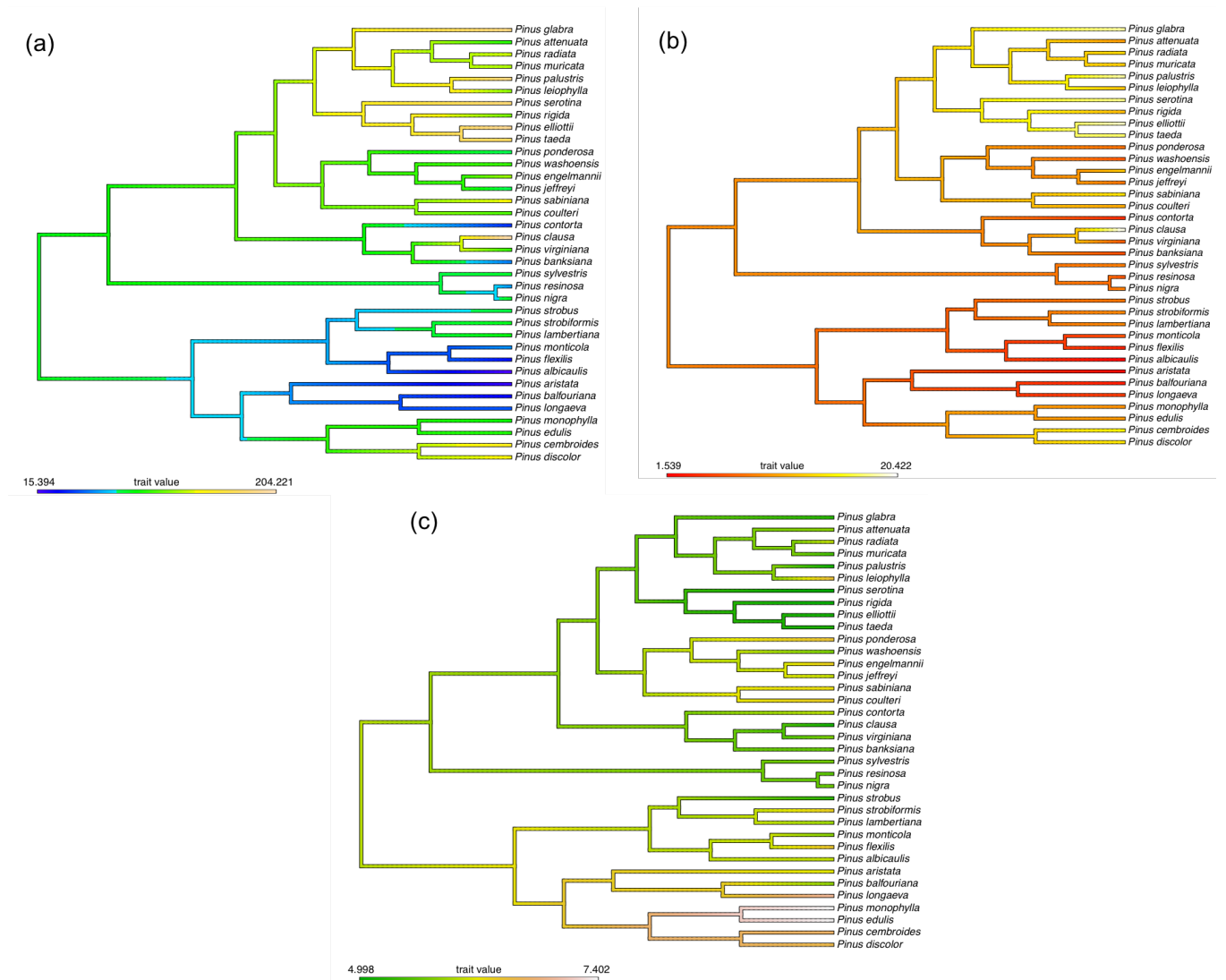


Figure 5: Ancestral state reconstruction for (a) precipitation, (b) temperature and (c) soil pH. Reconstructions were done by Retriected Maximum Likelihood methods

elliottii occupies a niche with high precipitation, high temperatures and low soil pH. There is not much variation within these sub-clades but a clear difference between them.

Discussion and Conclusion

Given the results presented in this analysis, we see that there is no reason to believe traits are more similar than what would be expected under the Brownian Motion model. There is some controversy surrounding thresholds for niche conservatism - does finding a phylogenetic signal in a niche variable indicate niche conservatism? Losos (2008) argued that although a phylogenetic signal among trait distributions is necessary for proving niche conservatism, traits must be more similar than expected under the Brownian Motion model to truly indicate a stabilizing selection and be conserved.

On the other hand, Wiens *et al.*, (2010) argues that any phylogenetic signal, even a weak phylogenetic signal, demonstrates a case of niche conservatism ($\lambda, K > 0$).

The results of this study provide a phylogenetic signal alone. Under the argument, the traits in clade *Pinus* are not conserved. However, under the second, a phylogenetic signal provides evidence for phylogenetic niche conservatism under the drift model (Cooper *et al.*, 2010).

Limitations and future directions. A consideration that should be taken into account is one of phylogenetic inertia. When the rate of evolution is too slow and is outpaced by the rate of change of an external driver (such as environmental change), species can take a long time to reach a new trait optimum in the rapidly-changing environment, therefore retain ancestral characteristics. Therefore putting the evolution, speciation and diversification of *Pinus* in the context of historical environmental change would allow a direct comparison of rates of evolutionary and environmental change providing an opportunity to detect phylogenetic inertia. Further, a way to expand this study would be to test different values of θ and α in OU models and check how sensitive the model may be to a change in these parameters.

References

- Blomberg S.P., Garland T., and Ives A.R., (2003), Testing for phylogenetic signal in comparative data: behavioral traits are more labile, *Evolution*, 57: 717-745
- Cooper N., Jetz W., and Freckleton R.P., (2010), Phylogenetic comparative approaches for studying niche conservatism, *Journal of Evolutionary Biology*, 23, 2529-2539
- Li J., He Q., Hua X., Zho J., Xu H., Chen J., and Fu C., (2009), Climate and hisotry explain the species richness peak at mid-elevation for *Schizothorax* fishes distributed in the Tibetan Plateau and its adjacent regions, *Global Ecology and Biogeography*, 18:264-272
- Kozak K.H., and Wiens J.J., (2010), Niche conservatism drives elevational diversity patterns in Appalachian Salamanders, *American Naturalist*, 176(1), 40-54
- Losos J.B., Leal M., Glor R.E., De Queiroz K., Hertz P.E., Schettino L.R., Lara A.C., Jackman T.R., and Larson A., (2003), Niche lability in the evolution of a Caribbean lizard community, *Nature*, 424, 542-545
- Pagel M., (1997) Inferring evolutionary processes from phylogenies, *Zool. Scr*, 26, 331-348

Peterson A.T., Soberon J., and Sanchez-Cordero V., (1999) Conservatism of ecological niches in evolutionary time, *textitScience*, 285, 1265-1267

Potter K.M. and Woodall C.W., (2012) Trends over time in tree and seedling phylogenetic diversity indicate regional differences in forest biodiversity change, *textitEcological applications*, 22(2), 517-531

Smith, S.A., Nieto Montes de Oca, T.W., Reeder and J.J. Wiens (2007), A phylogenetic perspective on elevational species richness patterns in Middle American treefrogs: why so few species in lowland tropical rainforests? *Evolution*, 61: 1188-1207

Waltari E., Schroeder R., McDonald K., Anderson R.P., Carnaval A., (2014), Bioclimatic variables derived from remote sensing: assessment and application for species distribution modelling, *Methods in Ecology and Evolution*, 5, 1033-1042

Wiens J.J. and Graham C.H., (2005) Niche conservatism: integrating evolution, ecology and conservation Biology, *Annual Review of Ecology and Evolution Syst.*, 36, 516-539

Wiens J.J., Parra-Olea G., Garcia-Paris M., and Wake D.B., (2007) Phylogenetic history underlies elevational patterns of biodiversity in tropical salamanders, *Proceedings of the Royal Society B: Biological Sciences*, 274:919-928

Wiens J.J., Ackerly D.D., Allen A.P., Anacker B.L., Buckley L.B., Cornell H.V., Damschen E.I., Davies T.J., Grytnes J., Harrison A., Hawkins S.P., Holt R.D., McCain C.M. and Stephens P.R., (2010) Niche conservatism as an emerging principle in ecology and conservation biology, *Ecology Letters*, 13, 1310-1324

Supplemental figures

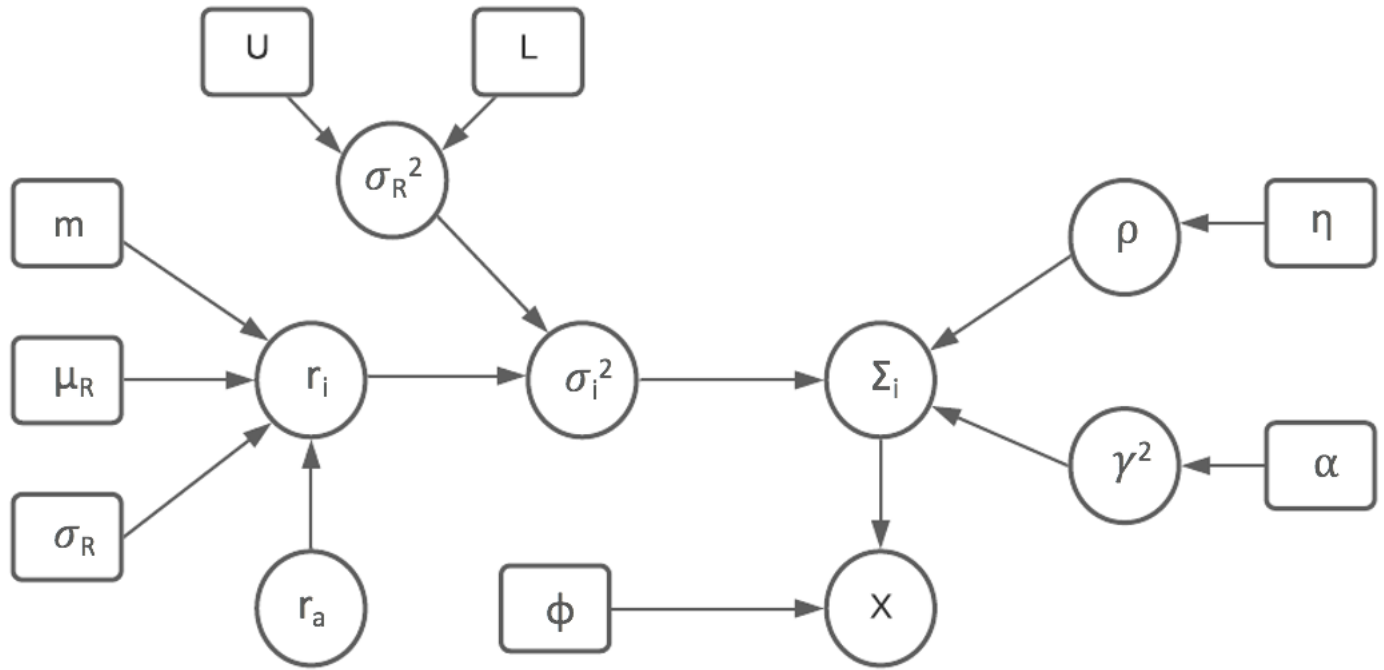


Figure 6: The DAG represents the Multivariate relaxed clock brownian motion model. The model has two components - relaxed clock component where parameters are indexed by i and are allowed to vary with branches and the multivariate component where the model estimates the correlation and proportional rates of evolution for each character trait

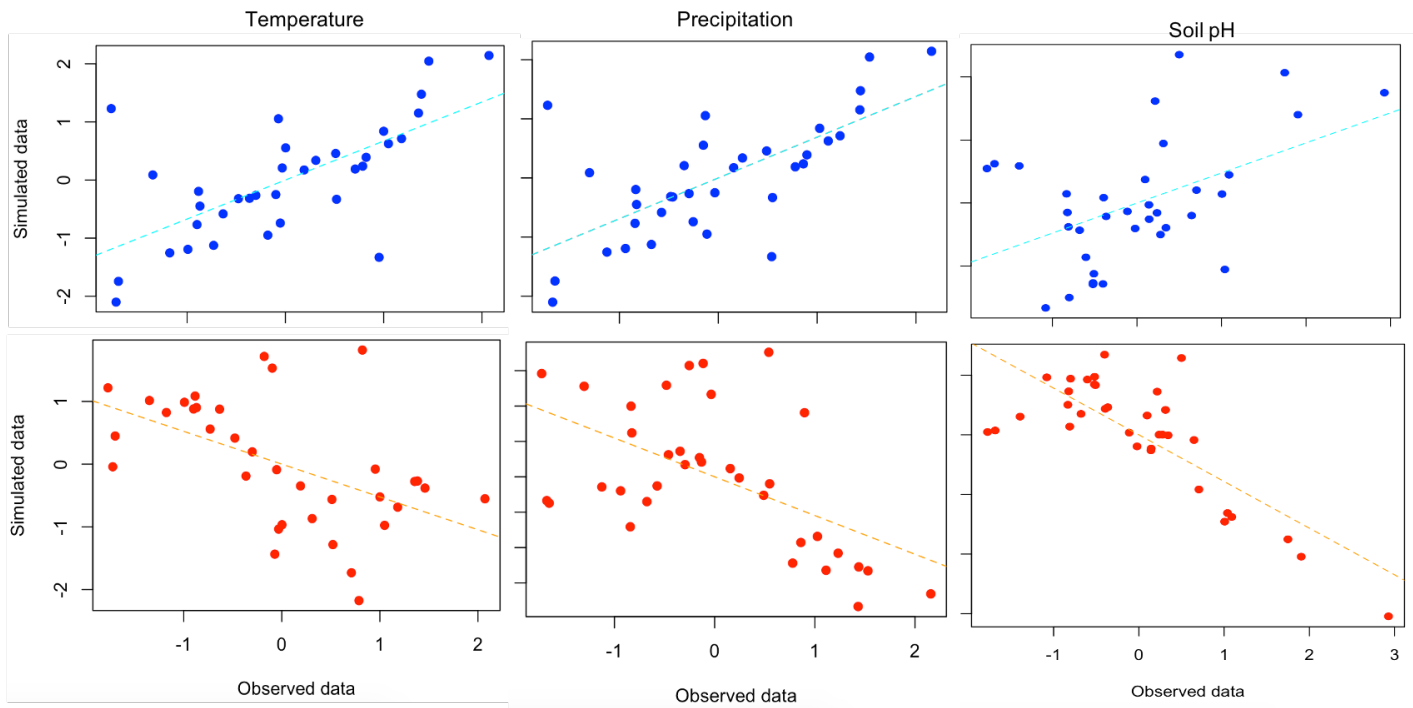


Figure 7: Comparisons of simulated data and observed data for all three abiotic variables under the Brownian Motion Model (BM) and Ornstein-Uhlenbeck (OU) model. If patterns follow the linear line closely, it indicates a reasonably good simulation of the observed data. All data was scaled for these plots

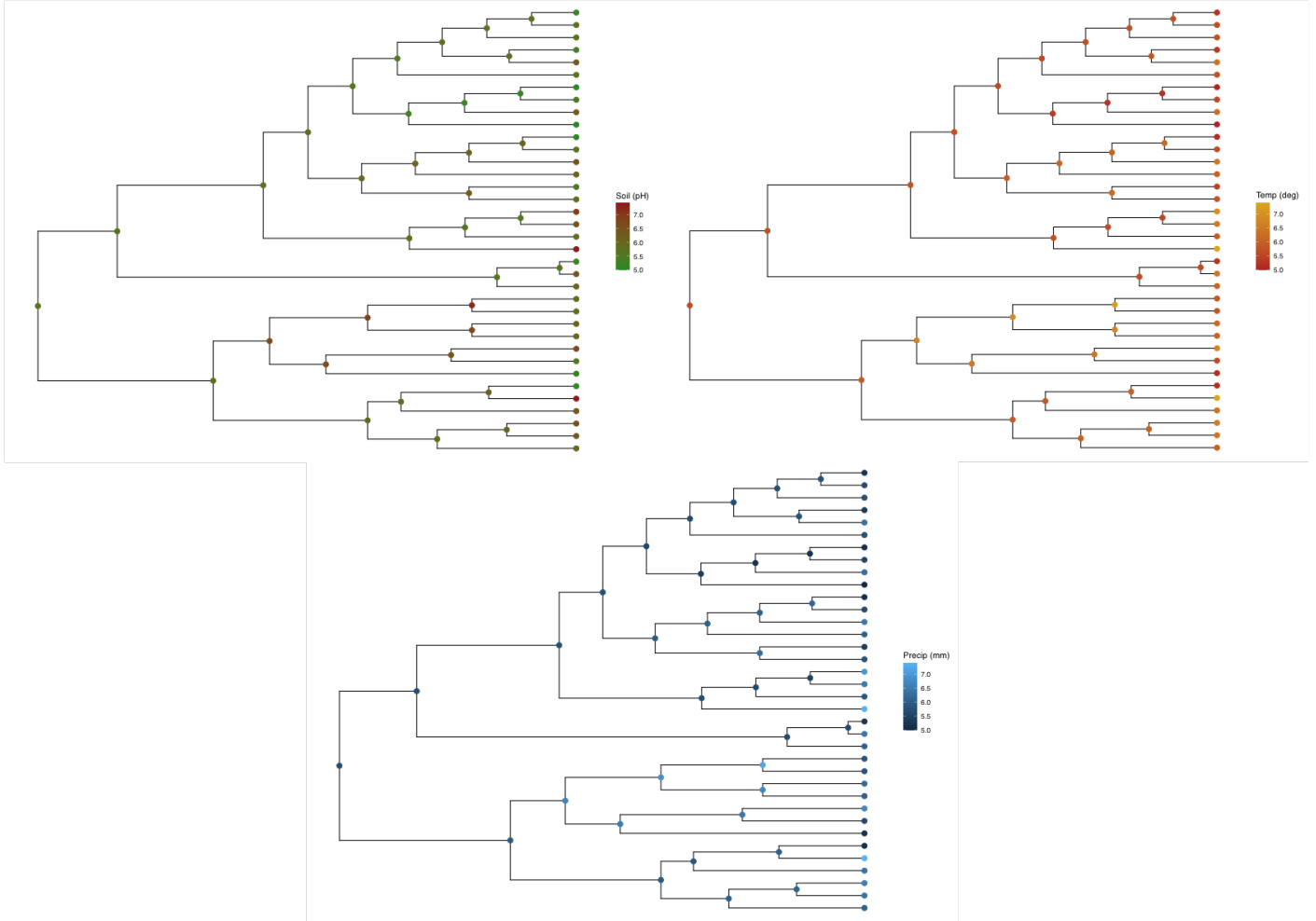


Figure 8: Another reconstruction using the relaxed clock multivariate brownian motion model. To do this reconstruction, the ancestral states were not inferred jointly with the other random parameters. The estimated branch rate multipliers from the posterior were multiplied by to the branch lengths of the pine phylogeny in order to lengthen branches where more shifts occurred, and shorten branches where not. This re-estimated phylogeny was then used for ancestral state reconstruction using restricted maximum likelihood methods.