

# "PRINCIPLES OF PHYLOGENETICS: ECOLOGY AND EVOLUTION"

*Integrative Biology 200B*

University of California, Berkeley  
(Lecture Notes from D.D. Ackerly)

Spring 2011

Ginger Jui

April 26, 2011. **Community Ecology and Phylogenetics**

## Readings:

Cavender-Bares, J., D. D. Ackerly, D. Baum, and F. A. Bazzaz. 2004. Phylogenetic overdispersion in Floridian oak communities. *Amer. Nat.* 163:823-843.

Swenson, N.G., B.J. Enquist, J. Pither, J. Thompson, J.K. Zimmerman. 2006. The problem and promise of scale dependency in community phylogenetics. *Ecology* 87: 2418-2424.

## Background:

Webb, C. O., D. D. Ackerly, M. McPeck, and M. J. Donoghue. 2002. Phylogenies and community ecology. *Annu. Rev. Ecol. Syst.* 33:475-505.

*Ecology*, vol. 87, special issue (July 2006) on phylogenies and community ecology

The field of community ecology asks: what are the processes responsible for the identity and relative abundance of species that cooccur in local assemblages, and how do these vary through time? These processes span a wide range, from ecophysiology and stress tolerance, to the intricacies of biotic interactions including competition, predation, symbioses, etc. The concept of the niche has played a central, though controversial role in community ecology. Two related ideas have shaped the intersection of community ecology with phylogenetics: 1) identical species cannot coexist (the competitive exclusion principle), and 2) related species are ecologically similar (niche conservatism or phylogenetic signal). Therefore, as Darwin argued:

*As species of the same genus have usually, though by no means invariably, some similarity in habits and constitution, and always in structure, the struggle will generally be more severe between species of the same genus, when they come into competition with each other, than between species of distinct genera.*

(Darwin 1859)

The corollary of these two principles is that closely related species should co-occur less than would be expected (though the question of what is *expected* requires careful consideration). Initial efforts to test this hypothesis focused on species:genus ratios, predicted to be lower than expected in local assemblages (e.g., on islands, see citations in Webb et al. 2002). With the elaboration of detailed and time-calibrated phylogenies, these questions have been reframed in terms of phyletic distance among co-occurring species.

Starting with Diamond (1975), the focus on the competitive exclusion principle was expanded to the more general idea of community assembly, and the search for rules and regularities in community structure that might reflect underlying ecological processes. One of the important results of these studies, especially in plant ecology, was the renewed attention to convergence in community assembly, i.e., that co-occurring species may actually be phenotypically similar, reflecting similar functional requirements to survive under shared abiotic and biotic conditions. So it is an open question for any particular trait whether co-occurring species will be more similar or more different from each other than expected. These patterns may be termed phenotypic clustering or and phenotypic evenness.

The initial focus on niche conservatism (or high phylogenetic signal) can also be expanded to consider the full range of possibilities: traits of relevance to community assembly may exhibit a high degree of signal,

no signal (= random), or a significantly low signal (convergent evolution). These different possibilities set up the following table, relating patterns of phylogenetic signal, community assembly and resulting phylogenetic community structure (see Webb et al. 2002, Cavender-Bares et al. 2004):

		<b>Trait similarity within communities</b>	
		Clustering of traits (driven by environmental filtering)	Overdispersion of traits (driven by competitive interactions)
<b>Trait evolution</b>	Conserved	Phylogenetic clustering	Phylogenetic overdispersion
	Convergent	Phylogenetic overdispersion	Phylogenetic clustering or random dispersion

Figure 2 from Cavender-Bares et al. 2004

We have discussed measures of phylogenetic signal previously. Analysis of phenotypic clustering and evenness can be conducted based on trait variance and other statistics applied to the distributions of traits among co-occurring species. Two metrics that have proven useful are tests for reduced trait range, as a measure of phenotypic clustering, and for reduced standard deviation of nearest neighbor distances (in trait space), for phenotypic evenness (Kraft et al. 2008; Cornwell and Ackerly 2009). Null models with randomly assembled communities are used as a basis to test for reduced values of these statistics, relative to the null.

Here I will introduce two measures of phylogenetic clustering and evenness, as applied to community ecology.

The basic data for phylogenetic community structure analysis is a phylogenetic tree for the species of a regional species pool (i.e., the collective species list across a range of habitats or a large area), together with individual species lists for smaller plots or specific habitats within the community.

#### 1. Phylogenetic diversity (Net relatedness index, Nearest taxon index):

To determine if the species in a particular plot are more closely related than expected by chance, the mean phylogenetic distance (MPD) is calculated as the sum of the pairwise phyletic distances among all pairs of taxa in the community:

$$MPD = \frac{\sum_{i=1}^{N-1} \sum_{j=i+1}^N d_{i,j} p_i p_j}{\sum_{i=1}^{N-1} \sum_{j=i+1}^N p_i p_j}$$

where  $d_{i,j}$  is phyletic distance between taxa  $i$  and  $j$ , and  $p_i, p_j$  are 0,1 for presence/absence of species. Relative abundance may also be used for  $p_i$  values, and the numerator alone is then known as Rao's entropy, and is closely related to the Simpson diversity index used in ecology.

The expected value for this statistic under a null model can be calculated by randomly drawing

communities of the same species richness from the regional species pool, and calculating the mean MPD across a large number of random draws. If  $MPD_{obs} < MPD_{exp}$ , then the observed community is phylogenetically clustered, and conversely if  $MPD_{obs} > MPD_{exp}$ . Webb et al. (2002, 2008) define the Net Relatedness Index as:

$$NRI = -1 * \frac{MPD_{obs} - mn(MPD_{exp})}{sd(MPD_{exp})}$$

where  $mn$  and  $sd$  are the mean and standard deviation of MPD values obtained from a large number (usually 999+) of random draws. NRI is positive for clustered communities and negative for evenly spread communities, and significance can be determined by ranking the observed value in comparison with the distribution of null values.

A second measure of community phylogenetic structure is whether the most closely related co-occurring species in a community is more or less closely related than expected. This measure is more directly related to the calculation of species:genus ratios, as it focuses on whether closely related species tend to co-occur or not. It can be determined by first calculating the mean nearest neighbor distance:

$$NNPD = \frac{\sum_{i=1}^N \min \left( \sum_{j=1}^{N(j \neq i)} d_{i,j} p_i p_j \right)}{N}$$

and then comparing that to the expected distribution under a null model, to obtain the Nearest Taxon Index:

$$NTI = -1 * \frac{NNPD_{obs} - mn(NNPD_{exp})}{sd(NNPD_{exp})}$$

Again, positive values of NTI indicate that species co-occur with more closely related species than expected, and negative values indicate that closely related species do not co-occur.

Here I have only mentioned the simplest null model, involving a random draw of species from the overall species pool. There are a variety of more sophisticated null models, in particular ones that keep the frequency of occurrence of species constant across an entire data set (see Gotelli and Graves 1996, Kembel 2006).

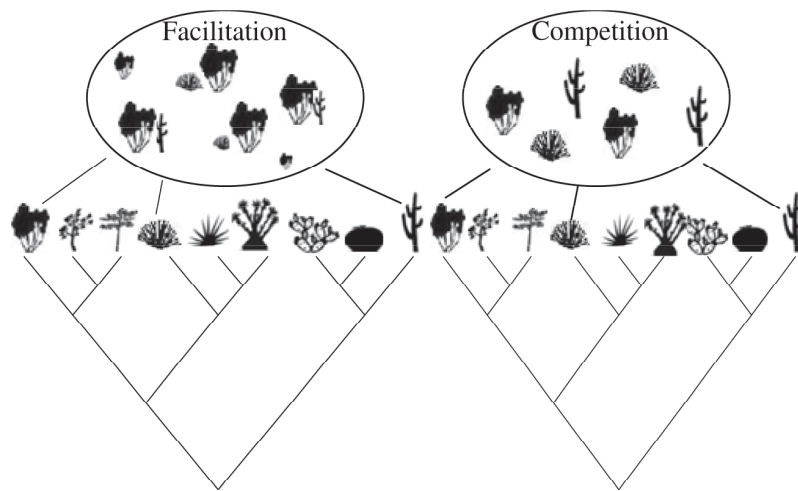
## 2. Mantel test approaches

The second major class of methods to assess phylogenetic community structure is based on Mantel tests of correlations between distance matrices, and is analogous to the use of Mantel tests for assessing phylogenetic signal. In this case the data set would be a series of plots of communities across a landscape. For each pair of taxa in the entire landscape, you can calculate phyletic distance ( $d_{ij}$ ) and a measure of co-occurrence or overlap in distribution, such as:

$$c_{i,j} = \frac{\sum_{x=1}^P p_{x,i} p_{x,j}}{\sqrt{\sum_{x=1}^P p_{x,i} \sum_{x=1}^P p_{x,j}}}$$

where  $P$  is the total number of plots in the data set, and  $p_{x,i}$  is presence/absence (or abundance) of species  $i$  in plot  $x$ .  $c_{i,j}$  ranges from 0, for species that never co-occur, to 1 for species that are always found together. A plot of  $c_{i,j}$  vs.  $p_{i,j}$  then provides a measure of whether closely related species tend to co-occur or to have complementary distributions (see Cavender-Bares et al. 2004).

As in discussions of phylogenetic signal, there is a many-to-one problem in the relationship between community assembly processes and community phylogenetic structure. If we use a single metric, such as NRI, it can either be positive, not different from 0, or negative. Yet, there are many different processes that may shape this statistic, depending on the scale of the study, including filtering, competition, facilitation, biotic interactions, biogeographic sorting, speciation, etc. Not surprisingly, calculating the community phylogenetics signals alone will not reveal the underlying community assembly processes. These approaches must be combined with an appreciation for the scale of the samples and pools, the makeup of the biota, evidence from trait-based studies, and experimental analysis of the factors influencing species establishment and distributions. One of the main reasons that community phylogenetics is valuable is that it is relatively easy, requiring only plot level distribution data and some phylogenetic information.



**Figure 1** Two different ways how phylogenetic overdispersion in communities can be produced. In the case of competition, overdispersion is achieved by the exclusion of close-related species producing a checker-board spatial pattern. In the case of facilitation, overdispersion is achieved because nurses add distant-related species to the community producing a clumped association pattern among species.

Figure from Valiente-Banuet, A. & Verdú, M 2007.

#### Literature Cited:

- Cornwell, W.K. and D.D. Ackerly. 2009. Community assembly and shifts in the distribution of functional trait values across an environmental gradient in coastal California. *Ecological Monographs* 79: 109-126.
- Kraft, N.J.B., R. Valencia, and D.D. Ackerly. 2008. Functional traits and niche-based tree community assembly in an Amazonian forest. *Science* 322: 580-582.
- Kembel, S.W. & Hubbell, S.P. (2006) The phylogenetic structure of a neotropical forest tree community. *ECOLOGY*, **87**, S86-S99.
- Diamond, J.M. (1975). Assembly of species communities. In *Ecology and evolution of communities* (M. L. Cody and J. M. Diamond, eds.) pp. 342-373, Belknap Press, Cambridge, MA.
- Gotelli, N., and G. Graves. 1996. Null models in ecology. Smithsonian Inst. Press, Washington DC.
- Webb, C.O., D.D. Ackerly, and S.W. Kembel. 2008. Phylocom: software for the analysis of phylogenetic community structure and trait evolution. *Bioinformatics* 24: 2098-2100.

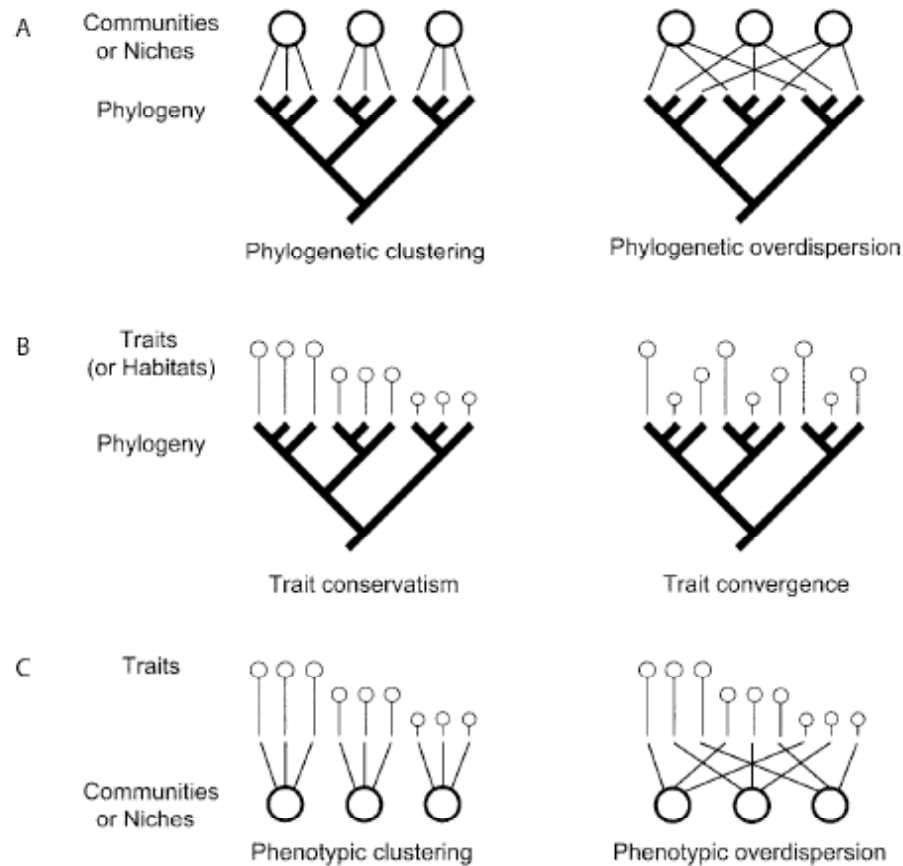
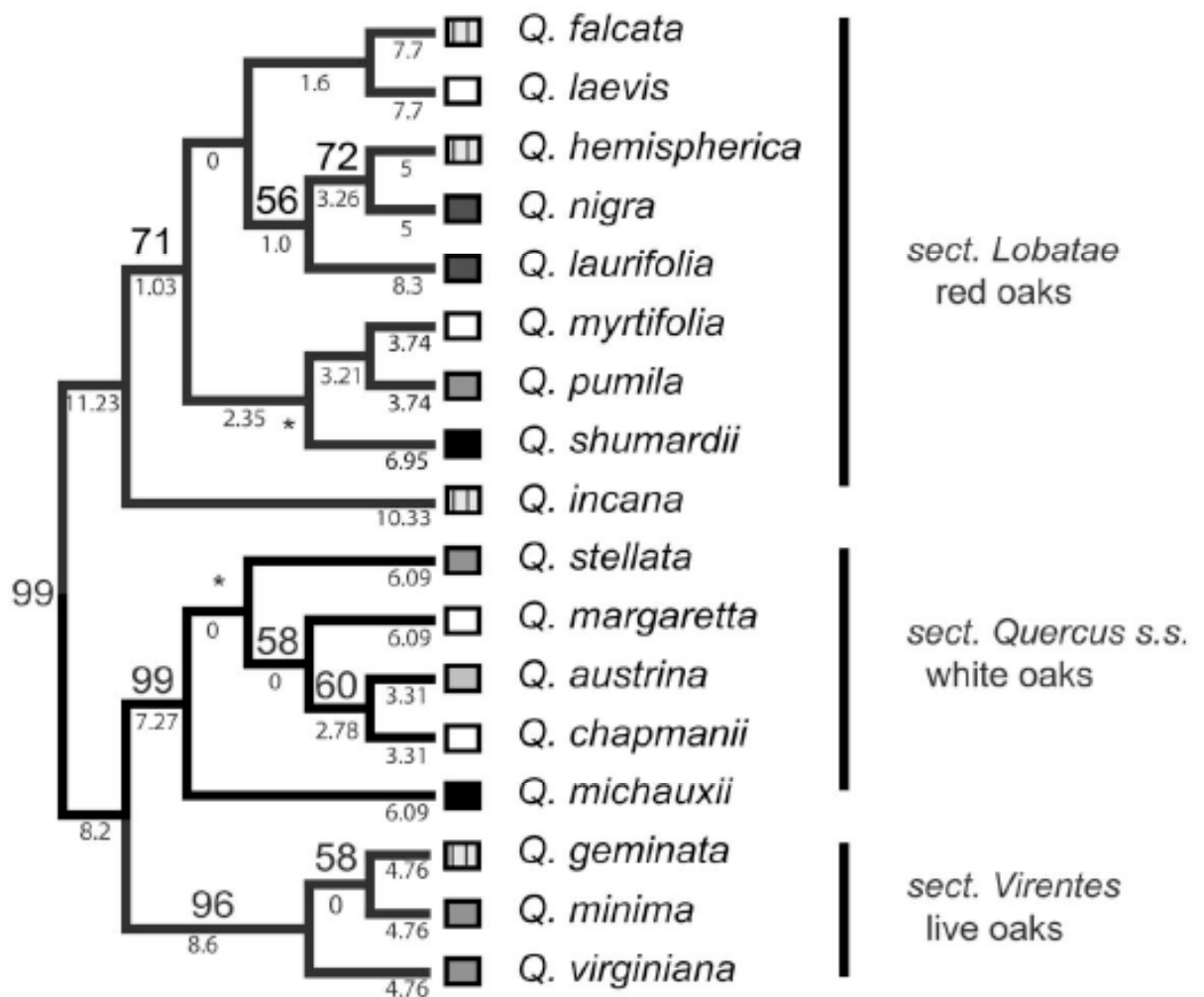
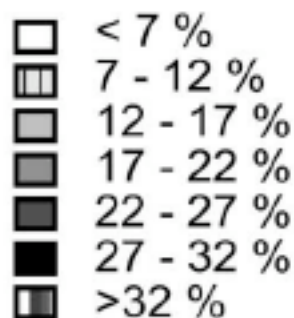


Figure 1: A, Phylogenetic structure of communities or niches can show patterns of clustering (*left*) at one end of the spectrum or overdispersion (*right*) at the other end of the spectrum. B, Evolution of species traits may show patterns of conservatism (*left*) or convergence (*right*). The interaction of phenotypes of sympatric species (competition) and the environmental tolerances of phenotypes (filtering) determine the phenotypic structure of communities. C, Environmental filtering can cause trait values of species within communities to be more similar than expected (phenotypic clustering; *left*). Competitive interactions may cause traits of species to be more different than expected (phenotypic overdispersion; *right*). Metrics to test for these patterns are presented in the study.



## Species habitat preferences

Mean soil moisture  
Vol. H<sub>2</sub>O [Vol. soil]<sup>-1</sup>



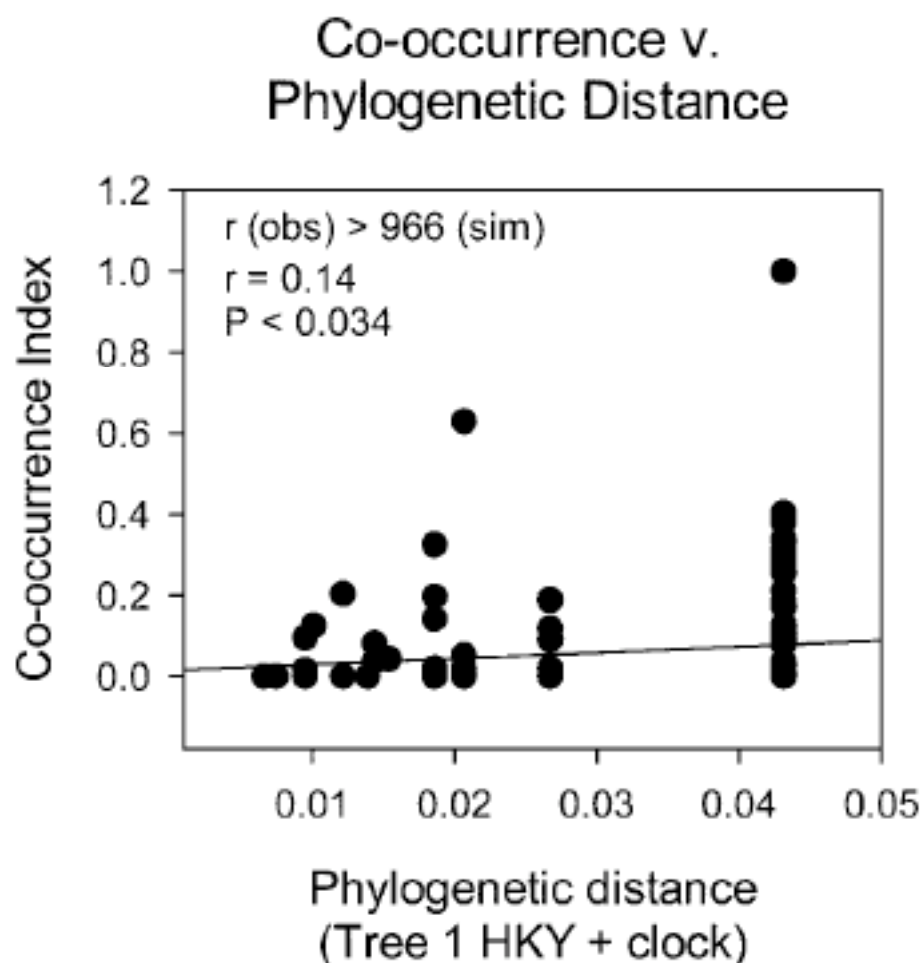


Figure 4: Species pairs with the least genetic distance between them show the lowest degree of co-occurrence resulting in a significant positive correlation between the co-occurrence index, calculated from basal area of oak trees within plots, and phylogenetic distance when compared with a null model in which species are randomized on the phylogeny. Phylogenetic distance is calculated using the HKY + clock model of evolution. Results of randomization tests for both co-occurrence and niche overlap based on basal area matrices are shown in table 2 and are based on presence/absence data matrices in table A1 in the online edition of the *American Naturalist*.

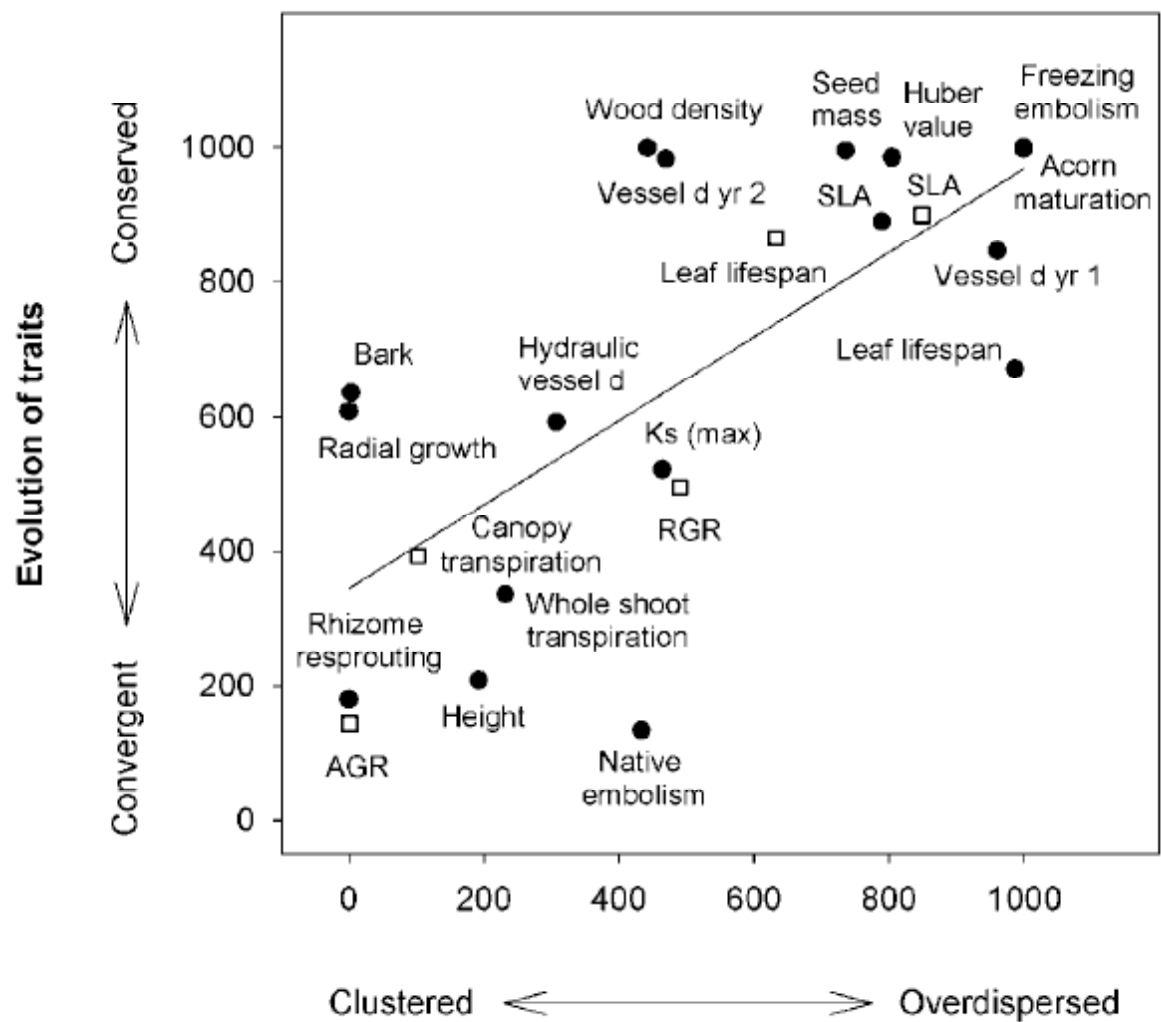


Figure 6: Conservatism of trait evolution in relation to the similarity of traits within communities. The X-axis indicates the number of simulations from the null model that are less than the observed value for the correlation of species pairwise trait differences with degree of co-occurrence. The Y-axis indicates the number of simulations from the null model that are greater than the observed value for the correlation of species pairwise trait differences with their phylogenetic distances. Axes are ordered with these orientations for comparison with figure 2. More convergent traits tend to be spatially clustered, and conserved traits tend to be overdispersed ( $r = 0.71$ ). Filled symbols represent traits measured on mature trees; open symbols represent traits measured on seedlings in a common garden.



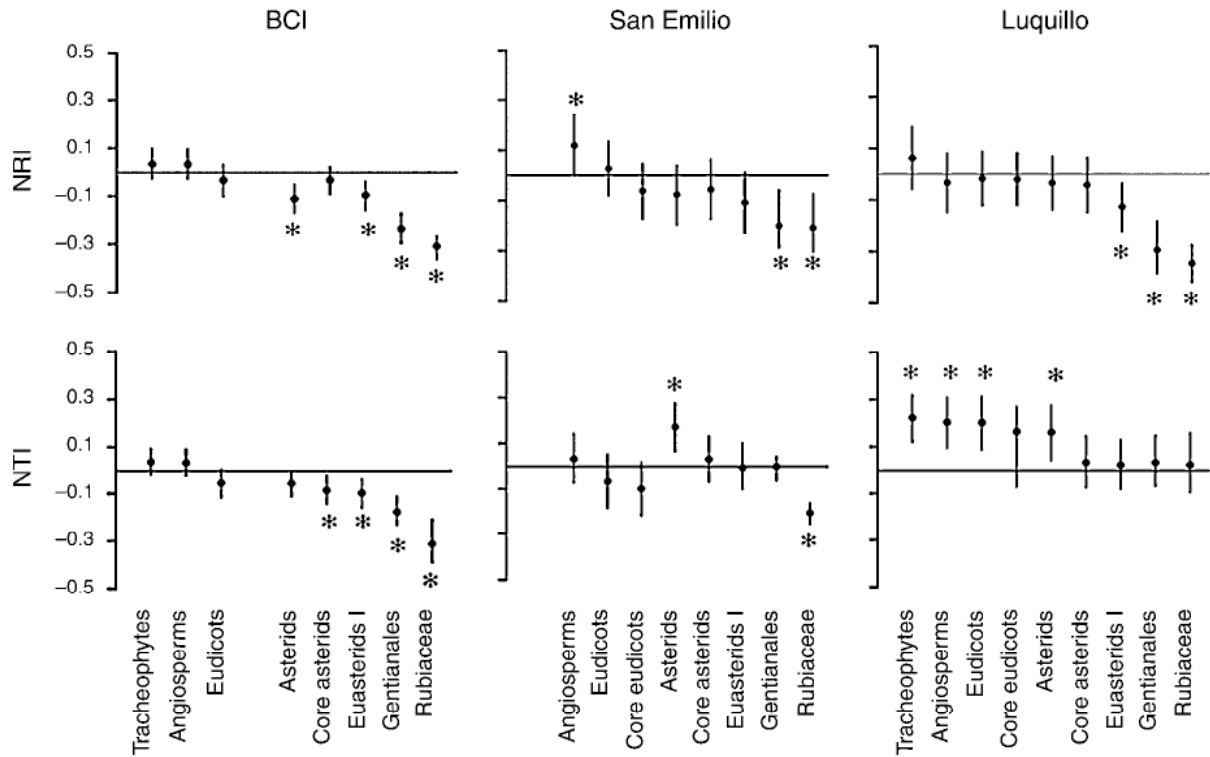


FIG. 1. The median net relatedness index (NRI) and nearest taxon index (NTI) scores for Barro Colorado Island (BCI), San Emilio, and Luquillo forest dynamics plots (FDPs) using eight or nine different taxonomic scales. Positive values indicate phylogenetic clustering, and negative values indicate phylogenetic overdispersion. The bars represent 95% confidence intervals.

\*  $P < 0.05$  (Wilcoxon test).

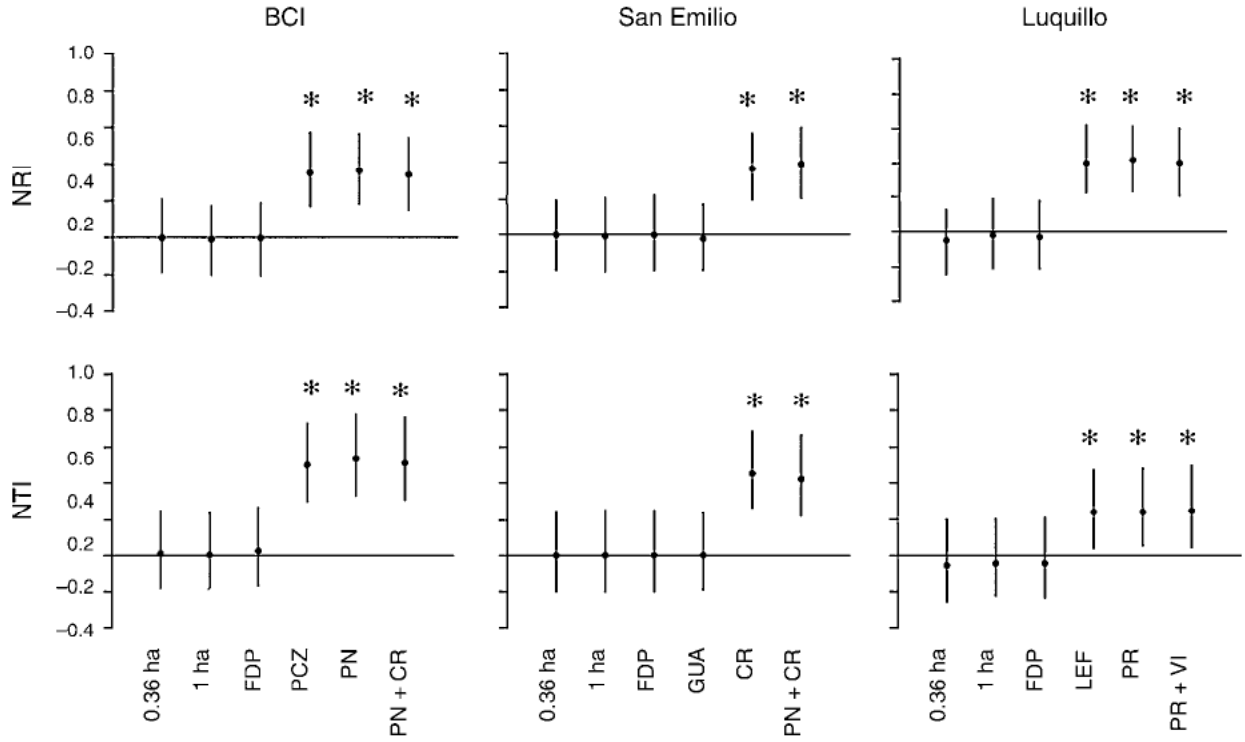


FIG. 2. The median NRI and NTI scores for BCI, San Emilio, and Luquillo FDPs using six different species pool sizes. Positive values indicate phylogenetic clustering, and negative values indicate phylogenetic overdispersion. The bars represent 95% confidence intervals. Key to abbreviations: FDP, forest dynamics plot; PCZ, Panama Canal Zone; PN, Panama; PN + CR, Panama and Costa Rica; LEF, Luquillo Experimental Forest; PR, Puerto Rico; PR + VI, Puerto Rico and the Virgin Islands; GUA, Santa Rosa and Palo Verde National Parks; CR, Costa Rica.

\*  $P < 0.05$  (Wilcoxon test).