

Re-considering the agenda for biotic homogenization research

David J. Harris

Introduction

As human influence creates the conditions for some species to spread globally and for others to go extinct, community ecologists have become increasingly concerned about the loss of local biotic distinctiveness, or biotic homogenization. Homogenization is typically quantified as an increase in the average Jaccard similarity among pairs of sites on a landscape (\bar{J} , defined more precisely in Equation 1 below). Community ecologists have struggled, however, to scale site-level biotic turnover into predictable changes in this quantity. While several frameworks have been proposed for understanding these effects, they are cumbersome to use in practice. For example, Olden and Poff (2003)'s framework involves classifying biotic turnover events into 14 categories, while Rosenblad and Sax (2016)'s framework involves six. In either framework, a researcher may need to keep track of thousands of such events across thousands of pairs of sites. Moreover, these pairs of sites are not statistically independent of one another, as each site belongs to many pairs. Worse, many event types are compatible with either homogenization or differentiation, depending on the prior state of the community. This complexity makes it difficult to test researchers' claims that their framework can explain commonly-observed patterns.

The Jaccard similarity between site i and j (\bar{J}) is defined as the proportion of species that are shared between them. More precisely, it is the number of species that are shared between the two sites (S_{ij}) divided by the number of species that occur in at least one (T_{ij}). The landscape-level mean Jaccard similarity value that homogenization research has largely focused on is therefore given by

$$\bar{J} = \frac{1}{\binom{n}{2}} \sum_{i \neq j} \left(\frac{S_{ij}}{T_{ij}} \right), \quad (1)$$

where n is the number of sites on the landscape (e.g. islands in an archipelago) and $\binom{n}{2}$ is the number of site pairs, equal to $n(n-1)/2$. Most existing frameworks focus on tracking biotic turnover's effects on individual S_{ij} and T_{ij} values (e.g. Olden and Poff 2003, Rosenblad and Sax 2016). However, there is good reason to believe that this level of detail is not necessary for understanding \bar{J} . Using 47 large data sets from the USDA PLANTS database (one

for each of the contiguous US states except Delaware; USDA NRCS 2010), I found that independently shuffling the list of sites occupied by every species only affected \bar{J} by a nearly imperceptible amount (an average of 0.006 on a scale from 0 to 1; Appendix 1).

The role of occupancy

If similarity and homogenization do not depend strongly on which sites are occupied by which species, then what does matter? In Harris et al. (2011), two colleagues and I showed (empirically and with an appeal to the law of large numbers) that average similarity depends primarily on the proportion of sites occupied by each species. Specifically, we defined an approximation to mean Jaccard similarity, J^* , given by the average value of S_{ij} divided by the average value of T_{ij} . Substituting in formulas for these two averages, we derived

$$J^* = \sum_k \binom{p_k n}{2} / \sum_k \left[\binom{n}{2} - \binom{(1-p_k)n}{2} \right], \quad (2)$$

where p_k is the proportion of sites occupied by species k .

Despite the omission of any information at the level of individual sites or site pairs, we showed that the approximation explained 99.8% of the variance in \bar{J} across the 47 USDA PLANTS data sets. It also explained an average of 98.8% of the variance in species-level effects on average similarity. The paper also introduced the *blender* package (Harris 2014) for R (R Core Team 2015), which allows users to easily perform these calculations. Similar results were presented around the same time by Chase et al. (2011) and by Vergara et al. (2011).

The above equation can be simplified further using Harris et al. (2011)‘s notion of “effective occupancy,” denoted p^* . This (approximate?) weighted average of the species’ p_k values acts as a “center of gravity” for average similarity: species whose p_k values exceed p^* pull its value up, while species with smaller p_k value pull p^* down. For the same reason, the existence of exotic species will generally cause net differentiation until they occupy at least as many sites as their native counterparts; from then on, their spread will make the landscape more homogenous than it would have been if they hadn’t invaded at all (Figure 1). The existence of such a threshold has been clear to homogenization researchers for more than a decade (refs; Rosenblad and Sax 2016), but its location has not been discussed outside of Harris et al. (2011) (nor has the location of an equally important point at $p^*/2$, shown in Figure 1).

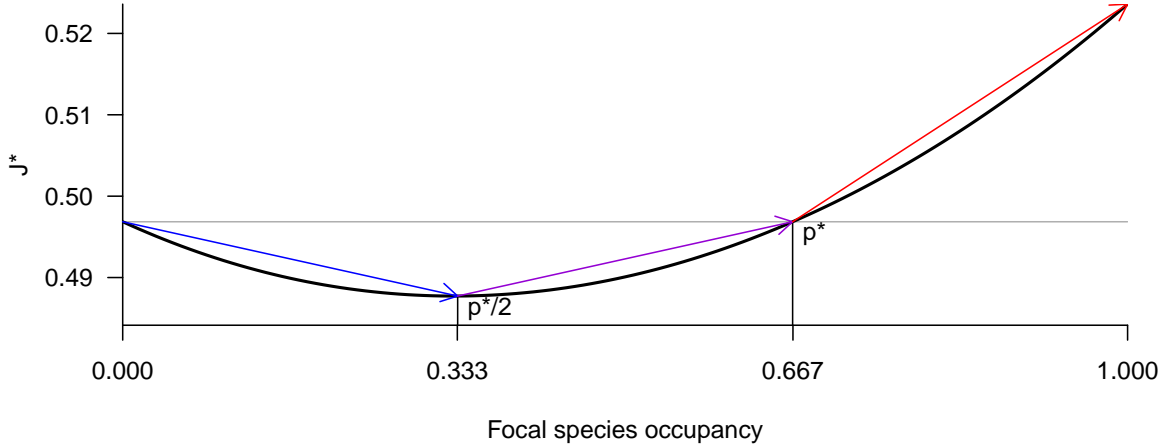


Figure 1: Effect of an exotic species on J^* , given a landscape with 120 sites and 20 native species whose occupancy rates produce $p^* = 2/3$. When an exotic species begins spreading, it will decrease mean similarity until its own occupancy rate reaches $p^*/2$ (blue arrow). When exotic occupancy reaches p^* , its occupancy will match the rest of the community, and J^* will return to its initial value (purple arrow). Only once the species becomes more widespread than its native counterparts will it be able to raise J^* above its initial value (red arrow).

A large portion of the homogenization literature can be explained by this single value. Species that have had longer to spread, and species whose local occupancy rates did not start at zero do not have any unique biological property that causes them to exert larger homogenizing effects on landscapes; they will simply tend to be farther along the path shown in Figure 1 (c.f. [refs]). Precisely the same argument could be made for species with high dispersal rates [refs], [etc.].

Residual variation

One point that Harris et al. (2011) made but did not emphasize is that the residuals from the J^* approximation are given by an identity from Welsh et al. (1988):

$$\bar{J} = J^* - \frac{\text{cov}(T_{ij}, S_{ij}/T_{ij})}{\text{mean}(T_{ij})}, \quad (3)$$

where cov refers to the population covariance (rather than the more familiar sample covariance). In other words, \bar{J} can be exactly decomposed into an occupancy component and a covariance component¹. In 2011, we largely disregarded the covariance component

¹Note that our 2011 treatment of this decomposition switched J^* and \bar{J} in Appendix 2's Equation B1

because it was usually small (and approaches zero as the size of the data set increases under some fairly realistic assumptions), but it has important consequences for the way we think about homogenization. In general, it implies that landscapes whose local species compositions are determined by a small number of important factors (which will tend to have negative covariances) will have higher mean similarity values than would be implied by their occupancy rates (Figure 2a). The covariance term also implies a possible discrepancy between an intuitive view of beta diversity and the one given by \bar{J} (Figure 2); this apparent conflict deserves further scrutiny. If \bar{J} remains an important metric in future research, then a deeper understanding of this covariance term will be important.²

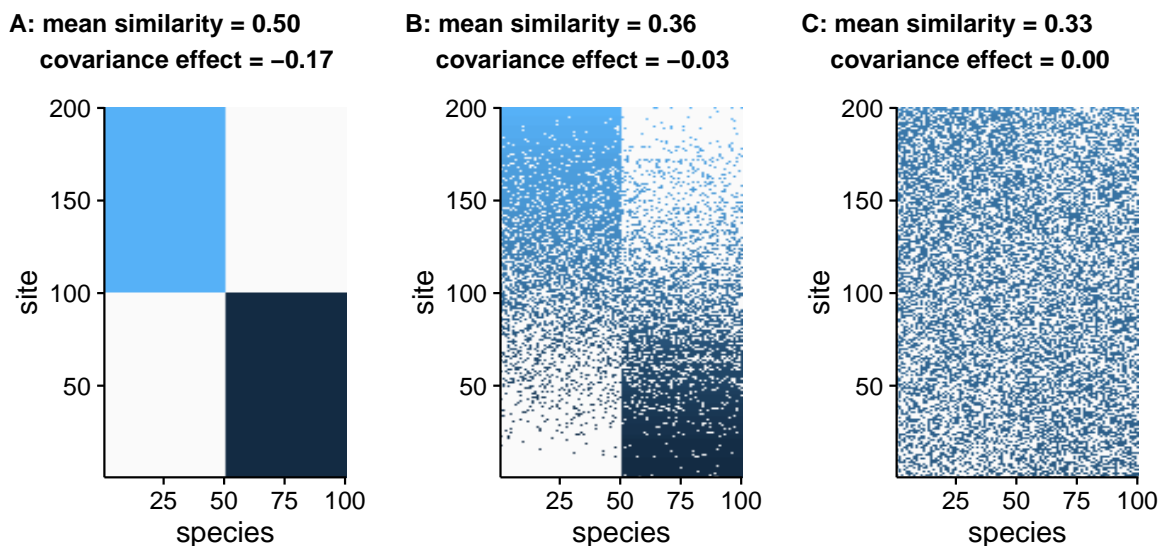


Figure 2: Visualizing the covariance component of \bar{J} (Equation 3), holding each species' occurrence rate constant at 50%. In **A**, there are only two kinds of communities: light blue communities that contain species 1-50 and dark blue communities that contain species 51-100. Here, the covariance between T_{ij} and S_{ij}/T_{ij} is strongly negative. In **B**, the two community types are not as distinct, and the absolute value of the covariance is smaller. In **C**, there are no distinct community types: each site is approximately the same intermediate blue shade and the covariance is not distinguishable from zero. Counter-intuitively, the process of randomly spreading species across the landscape decreases mean similarity from 0.5 to 0.33 rather than increasing it.

and some of the subsequent discussion in that Appendix is backwards.

²Welsh et al. (1988) provides additional decompositions, including an approximate based on the variance of T_{ij} that may be especially useful in this regard.

Avenues for future research

If most of the variance studied in homogenization research can be explained by p^* or J^* and the rest can be explained by a simple covariance term, where does this leave the field? At the end of Harris et al. (2011), we listed three paths forward, each of which remains promising five years later.

1. To the extent that p^* explains most of the variance, homogenization researchers should focus on explaining and predicting changes in species' occupancy rates. These are already important for other reasons in basic and applied community ecology.
2. Researchers should increase their focus on the covariance effects that cause deviations from J^* or from permutation-based null distributions (Chase 2007, Chase et al. 2011).
3. The fact that average Jaccard similarity can be calculated with so little ecological information (just occupancy rates and a covariance term) suggests that a great deal of information is lost during the averaging process. To that end, homogenization researchers should continue to think about similarity at scales below the landscape level (e.g. relationships between pairwise similarity and geographic distance). As the focus of homogenization research shifts toward sub-landscape scales, a deeper understanding of biotic turnover on individual pairs of sites (as provided by the frameworks of Olden and Poff (2003) and Rosenblad and Sax (2016)) will become increasingly valuable.

Acknowledgements

This work was funded by the Gordon and Betty Moore Foundation's Data-Driven Discovery Initiative through Grant GBMF4563 to E. P. White.

References

- Chase, J. M. 2007. Drought mediates the importance of stochastic community assembly. *Proceedings of the National Academy of Sciences* 104:17430–17434.
- Chase, J. M., N. J. Kraft, K. G. Smith, M. Vellend, and B. D. Inouye. 2011. Using null models to disentangle variation in community dissimilarity from variation in α -diversity.

Ecosphere 2:art24.

Harris, D. J. 2014. Blender: Analyze biotic homogenization of landscapes.

Harris, D. J., K. G. Smith, and P. J. Hanly. 2011. Occupancy is nine-tenths of the law: Occupancy rates determine the homogenizing and differentiating effects of exotic species. *The American naturalist* 177:535.

Olden, J. D., and N. L. Poff. 2003. Toward a mechanistic understanding and prediction of biotic homogenization. *The American Naturalist* 162:442–460.

R Core Team. 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rosenblad, K. C., and D. F. Sax. 2016. A new framework for investigating biotic homogenization and exploring future trajectories: Oceanic island plant and bird assemblages as a case study. *Ecography*:n/a–n/a.

USDA NRCS. 2010. The PLANTS database, accessed 2010-11-23 and retrieved from the blender package, (version 0.1.2; harris 2014). <http://plants.usda.gov>.

Vergara, P. M., J. Pizarro, and S. A. Castro. 2011. An island biogeography approach for understanding changes in compositional similarity at present scenario of biotic homogenization. *Ecological Modelling* 222:1964–1971.

Welsh, A. H., A. T. Peterson, and S. A. Altmann. 1988. The fallacy of averages. *The American Naturalist* 132:277–288.