2007 NORTHEASTERN NATURALIST 14(3):439–446

Ecological Dissimilarity Analysis: A Simple Method of Demonstrating Community-Habitat Correlations for Frequency Data

Sean F. Werle[1,*], Norman A. Johnson[2], Elizabeth R. Dumont[1],
and Piotr Parasiewicz[3]

Abstract - We introduce an analysis method to demonstrate correlation between biota and the physical habitats that they occupy. Using the same calculations as does Nei's genetic distance index, this method builds independent dissimilarity matrices for both habitat and fauna, which can then be compared in a common statistical framework. An important advantage of this method is that only frequency data are necessary to perform the analysis. We demonstrate the utility of this method using fish community and habitat data from the Eightmile and Pomperaug rivers in Connecticut. In both cases, there is a significant correlation between biota and habitat. Not only is ecological dissimilarity analysis a useful technique for testing community-to-habitat correlation, it is also an excellent tool for communicating this information to the many non-scientists who shape conservation policy.

Introduction

The interrelationship between biological communities and the suite of physical habitats they occupy is often taken for granted. That such a correlation exists has been demonstrated for ecosystems ranging from rivers to tropical forests and urban landscapes (Poff and Ward 1990). Techniques for demonstrating habitat/biota correlation vary, but there are none in common use that can be performed based solely upon frequency data (i.e., relative frequencies ranging from zero to 1, not counts) (e.g., Ahmandi-Nedushan et al. 2006, McGarigal et al. 2000, Pielou 1984). This report introduces a new method for demonstrating such relationships that requires only frequency data and is based on the well-understood mathematics of Nei's genetic distances (Nei 1972). We refer to this method as ecological dissimilarity analysis and introduce a new use of Nei's D_A (Nei et al. 1983), which, in this context, we call Nei's dissimilarity measure.

Ecological dissimilarity analysis involves generating two independently derived Nei's dissimilarity matrices for an ecosystem, one based entirely upon the biota of the system, and the other entirely upon the physical habitat. These matrices can then be compared using a Mantel test (Mantel 1967). A small p-value derived from such a comparison demonstrates that the biological community and the suite of available habitats are significantly

---

[1]Department of Biology, University of Massachusetts, Amherst, MA 01003. [2]Department of Plant, Soil, and Insect Sciences, University of Massachusetts, Amherst, MA 01003. [3]Northeast Instream Habitat Program, Department of Natural Resources Conservation University of MA, Amherst, MA 01003. [*]Corresponding author - swerle@bio.umass.edu.

correlated. Such associations provide an indicator of balance in an ecosystem, whereas a lack of correlation suggests that an ecosystem is disturbed in some way that causes a mismatch between the biological community and the physical habitat template. Used in this way, community-habitat correlations can be applied to monitor ecosystem recovery or to compare the relative health of different localities.

The method of ecological dissimilarity analysis we propose is generalizable across a wide range of biological systems, easy to calculate, and provides an intuitively satisfying indicator of the relationship between the biota and the physical habitat that make up the first building blocks of an ecosystem. An advantage of the method is that it can be calculated for any system for which frequency data are available, allowing it to be used in meta-analyses of disparate data sets. Finally, in contrast to more abstract statistical techniques, the results of ecological dissimilarity analysis can be expressed in a graphic format that is easily accessible to the general public, government employees, and legislators who enact conservation policies. To illustrate the utility of this method, we present analyses of riverine habitat and freshwater fish communities in two Connecticut rivers, the Eightmile and the Pomperaug. We selected these particular examples for three reasons: riverine habitat types are relatively clearly defined, there are rich biotic and abiotic datasets for these rivers (Parasiewicz et al. 2006, Walden and Parasiewicz 2005), and associations between biota and habitat are increasingly used to asses the health of rivers and inform watershed management (for recent examples see: Jansen et al. 2000, Koel 2004, Pegg and McClelland 2004).

Methods

Ecological dissimilarity analysis is accomplished in two steps. First, two dissimilarity matrices must be constructed: a biotic dissimilarity index for all pairwise comparisons among multiple sites, and a habitat dissimilarity index for all pairwise comparisons among the same sites. Second, the correlation between these matrices is evaluated statistically.

To calculate the between-site dissimilarity matrices, the terminology used for Nei's genetic distance (D, and ultimately Nei's modified genetic distance [D_A]; Nei 1972, Nei et al. 1983) needs to be redefined to put the computation into an ecological context. We replace the genetic concepts of locus and allele with the ecological concepts of category and value, respectively. For biological data, the categories are typically species, although any level of taxonomic resolution may be used, and the values are reciprocal frequencies (present and absent). For habitat data, any number of categories can be defined. The values are then the frequencies of the defined subcategories within a category. The same approach could also be used for biological data; for example, frequencies of fish species within a family may be used rather than using the species themselves as categories.

The following description is modified from Nei's original papers (Nei 1972, Nei et al. 1983), and the reader should consult these publications to understand the genetic analog of what we are suggesting here. The following example applies to fish in rivers, but any system of habitat and biota can be similarly treated.

Consider two sites X and Y within a river. Let x_i and y_i represent the frequencies of the *i*th fish species at each site respectively. The probability of randomly choosing the same fish species in two samples from site X is $j_x = \sum x_i^2$, while in site Y this probability is $j_y = \sum y_i^2$. The probability of randomly selecting the same species from site X and from site Y is $j_{xy} = \sum x_i y_i$. The normalized identity (an indicator of the similarity between species frequencies that has no non-genetic analog) of a given species between sites can be expressed as:

$$I_j = {j_{xy}} \Big/ {\sqrt{j_x j_y}} \, .$$

To generalize this expression across all species, we consider J_X, J_Y, and J_{XY}, as the arithmetic means of j_X, j_Y, and j_{XY}, respectively, and then the normalized identity between sites across all species is:

$$I = {J_{xy}} \Big/ {\sqrt{J_x J_y}} \, .$$

This index I can be used to calculate a dissimilarity measure between sites (Nei's original genetic distance), which is defined as $D = -\log_e I$. A problem with this measure is that it is sensitive to rare (low frequency) species and thus changes with increasing sample size, and so it was later modified to reduce this sensitivity (Nei et al. 1983). This is called Nei's modified genetic distance (D_A), which we will refer to here as Nei's dissimilarity measure (M. Nei, Pennsylvania State University, University Park, PA, pers. comm.). The resulting formula is:

$$D_A = \frac{1}{r}\sum_{j=1}^{r}\left(1 - \sum_{i=1}^{m_j}\sqrt{x_{ij}y_{ij}}\right) \, ,$$

where r is the number of categories studied (species in the special case of fish), and m_j is the number of possibilities for the *j*th category. In the case of species frequencies, there are only 2 possibilities, present or absent, so m_j is always 2 and the frequencies always sum to unity (i.e., a frequency at site X for species A of 0.89 will have a corresponding frequency for "lack of species A" equal to 0.11), but when we extend this idea to physical habitat, m_j will be able to take larger values. When this is the case, the species that were considered for fish are replaced with physical categories for habitat. One example is hydromorphologic units (HMUs), which can be defined in various ways. A minimal version is simply pools, riffles, and runs, for which m_j would be equal to 3. Another example is substrate, which could have several possibilities (sand, small gravel, large gravel, boulders, etc.). Other categories can be included with 3 "alleles:" absent, present, or abundant. For our examples, we included canopy shading, woody debris, and large boulders as "3-allele" categories. In all cases, the sum of the frequencies or proportions must be unity (i.e., all possibilities are represented).

Between-site Nei's dissimilarity measure matrices can now easily be calculated for multiple sites using both biotic community and physical habitat data. A number of software packages are available to do this; one program (which we used for the analyses shown here) called DISPAN (Ota 1993) is freely available and simple to use. In the data input file, one only needs to substitute habitat categories or species for "loci," and the possible values for each category as "alleles." (DISPAN also requires the "number of genes sampled," which isn't necessary in our context, and so for the data presented here, we used a constant [40] in the input data file. This value has no effect on the matrix calculation.) The dissimilarity matrices thus generated are mathematically realistic, and correlations between them are indicative of a real correlation between community and habitat. These dissimilarity matrices can also be used to build Unweighted Pair Group Method with Arithmetic mean (UPGMA; Sneath and Sokal 1973) trees (phenetic trees) that give a graphical representation of the correlation or lack thereof between sites (see Figs. 1 and 2). The dissimilarity matrices behind the trees can be compared statistically using a Mantel test (Mantel 1967), which is commonly used to demonstrate correlation between distance matrices.

Results (Demonstration of Principle)

The following two examples demonstrate the principle and application of ecological dissimilarity analysis. These are meta-analyses of data reported in earlier publications; detailed information surrounding data collection can be found in the cited reports. Raw data sets from those reports are also available from P. Parasiewicz upon request.

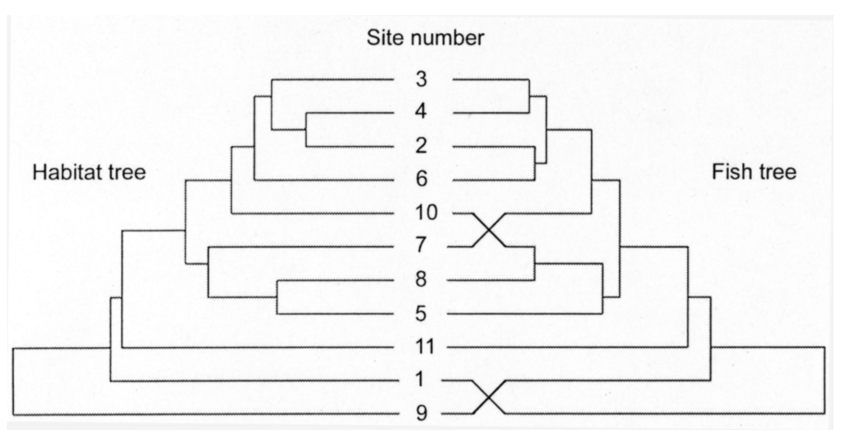

Figure 1. Independent UPGMA trees based on data from physical habitat (on the left) and fish community (on the right) from 11 sites in the Eightmile River. Both trees are based on matrices of Nei's dissimilarity measures (D_A) calculated between sites and are significantly correlated (Mantel test, $p < 0.01$). Only the horizontal component of diagonal lines should be counted in the branch lengths.

Eightmile River

The Eightmile River is a fourth-order tributary of the Connecticut River that drains a 62-square-mile watershed in southeastern Connecticut. Data summarizing fish communities and physical habitat from 11 sites in the Eightmile watershed (Walden and Parasiewicz 2005) were compiled to construct the biotic and habitat ecological dissimilarity matrices. The fish matrix was calculated from frequency data for each of the 19 fish species that were collected from the 11 sites within the Eightmile watershed. Collections from each site comprised some subset of these 19 species. The habitat matrix was calculated from 5 physical categories: hydromorphologic unit (HMU), which had 8 subcategories (pool, sidearm, backwater, glide, riffle, ruffle, rapid, and run); substrate, also with 8 subcategories (sapropel, pelal, psammal, akal, microlithal, mesolithal, macrolithal, megalithal [Austrian Standard ÖNORM 6232 1995]); woody debris; canopy shading; and boulders. The frequency of these last three habitat variables (woody debris, canopy shading, and boulders) at each locality was recorded as absent, present, or abundantly present.

The UPGMA trees built from these two matrices are shown in Figure 1. The lengths of the branches of these trees are proportional to the degree of dissimilarity (either biotic or physical) between sites. Diagonally crossed branches do not represent increased branch length; they simply serve to visually link crossed branches within a single figure. We used a Mantel test to determine that the matrices behind these trees are significantly correlated ($p < 0.01$; R Development Core Team 2004).

Pomperaug River

The Pomperaug River drains a 90-square-mile watershed in western Connecticut. Using the same methods cited above for the Eightmile River,

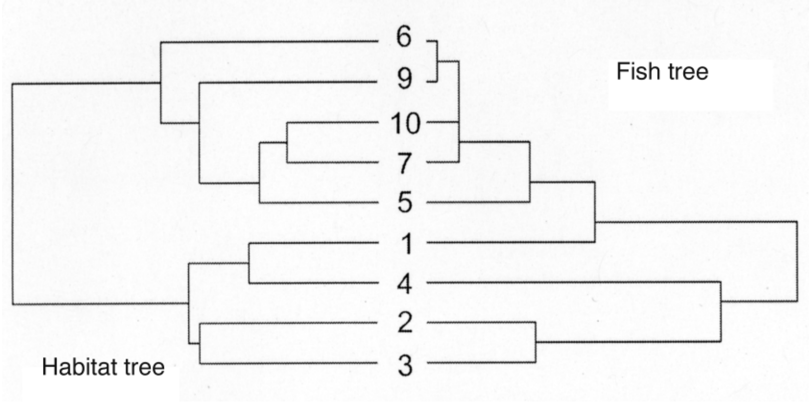

Figure 2. Independent UPGMA trees based on data from physical habitat (on the left) and fish community (on the right) from 9 sites in the Pomperaug River (site 8 was not sampled). As above, both trees are based on matrices of Nei's dissimilarity measures (D_A) calculated between sites and are significantly correlated (Mantel test, $p < 0.05$).

we computed fish and habitat dissimilarity matrices for data collected from nine sites in the Pomperaug River watershed (Parasiewicz et al. 2006). Again we found a significant correlation between the biotic (fish) and habitat matrices (Mantel test, $p < 0.05$; Fig. 2).

Discussion

An important goal of environmental science is the development of simple and intuitive methods of gauging the health status of ecosystems. Such information is critical to the development and implementation of effective conservation strategies as well as to public education. Assessing ecosystem health is a worthy goal, but certainly not an easy one; even defining what is meant by "health" is often challenging. We suggest that the strength of correlation between habitat and biota provides an opportunity to obtain an indicator of ecosystem "health" that is statistically testable. Well-balanced systems typically contain a variety of species that are associated with different physical resources as well as a few more broadly distributed habitat generalists. In these systems, the correlation between habitat and biota will be strong. In impacted ecosystems, the correlation should be weaker as empty niches either go unfilled or are taken over by generalists.

The two river ecosystems we analyzed are excellent examples of one application of the ecological dissimilarity analysis we propose. Both the Eightmile and Pomperaug rivers are considered to be relatively healthy, but are recovering from significant historic environmental perturbations. Originally traversing wooded terrains, the Eighmile and Pomperaug watersheds were deforested in the 19th century as the region was settled and converted to farmland. Now both rivers flow mainly through rural new-growth (<50 years old) forest and have recently experienced only minimal anthropogenic modification. For these and other historically impacted rivers, ecological dissimilarity analysis provides a method for monitoring recovery.

Although we used river ecosystems as examples, ecological dissimilarity analysis can be used in any situation that poses questions about associations between species and habitats. Rather than documenting ecosystem recovery, the method could be used to evaluate the strength of association between species and habitats in permanently modified urban landscapes (e.g., Sandstrom et al. 2006). Similarly, by sampling sites that represent altered and unimpacted environments, one could determine whether species-habitat associations are disrupted by habitat modification (e.g., Bobo et al. 2006). By surveying in same site(s) repeatedly over time, one could use ecological dissimilarity analysis to evaluate the strength of association between changes in species composition and habitat changes following perturbation. Recent examples of studies in which this may have been of interest include the recovery of spider assemblages after burning (Brennan et al. 2006) and changes in bird community structure following conversion of forest to agriculture (Aratrakorn et al. 2006). Finally, investigations of species-habitat correlations among sites spread across geographic gradients could be useful in evaluating the consequences of habitat fragmentation (e.g., Pinkus-Rendon et al. 2006) or, in rivers, patterns of species turnover from upstream to downstream (e.g., Bhat 2004).

Correlations between communities and habitats are usually shown using sophisticated analyses such as principle component analysis (PCA) or canonical correspondence analysis (CCA). While these analyses are powerful and fairly easy to perform, they suffer from two limitations: they are not suitable for analysis of frequency data alone, and they are difficult for the layperson to understand intuitively. The method of ecological dissimilarity analysis that we present here overcomes both of these problems. The first solution is inherent in the method, and the second is solved because the visual presentation provided by UPGMA trees is something that most people can intuitively grasp. This is important, because studies of habitat and biota are often directed toward the goal of ecologically sound resource management, and decisions about management are almost never made by scientists alone, but rather by politicians, voters, and citizens' groups—people who may be unfamiliar with scientific methods. For studies where results are intended to inform management decisions, a simple, intuitive way of describing the health of an ecosystem is a valuable tool. We suggest that ecological dissimilarity analysis provides such a tool. The community-habitat trees that are generated by the technique are both easy for a general audience to understand and are backed by rigorous statistical testing. From a management perspective, these are valuable advantages of ecological dissimilarity analysis over other, more complex methods of correlation.

Acknowledgments

The authors would like to thank the National Park Service and the Eightmile River Wild and Scenic Study Committee for funding the Eightmile River study and the Pomperaug River Watershed Coalition and the Connecticut Department of Environmental Protection (CT-DEP) for funding the Pomperaug River study. We are also grateful to Professor Masatoshi Nei of Pennsylvania State University for helpful comments on the manuscript. We also thank the University of MA, Natural History Collections, for covering page charges for this manuscript.

Literature Cited

Ahmandi-Nedushan, B., A. St.-Hilaire, M. Berube, E. Robichaud, N. Thiemonge, and B. Bobee. 2006. A review of statistical methods for the evaluation of aquatic habitat suitability for instream flow assessment. River Research and Applications 22:503–523.

Aratrakorn, S., S. Thunhikorn, and P.F. Donald. 2006. Changes in bird communities following conversion of lowland forest to oil palm and rubber plantations in southern Thailand. Bird Conservation International 16:71–82.

Austrian Standard ÖNORM 6232. 1995. Richtlinien für die oekologische Untersuchung und Bewertung von Fleissgewaessern. P. 38. Oesterreichische Normungsinstitut, Vienna, Austria.

Bhat, A. 2004. Patterns in the distribution of freshwater fishes in rivers of central western Ghats, India and their associations with environmental gradients. Hydrobiologia 529:83–97.

Bobo, K.S., M. Waltert, H. Fermon, J. Njokagbor, and M. Muhlenberg. 2006. From forest to farmland: Butterfly diversity and habitat associations along a gradient of forest conversion in southwestern Cameroon. Journal of Insect Conservation 10:29–42.

Brennan, K.E.C., L. Ashby, J.D. Majer, M.L. Moir, and J.M. Koch. 2006. Simplifying assessment of forest management practices for invertebrates: How effective are higher taxon and habitat surrogates for spiders following prescribed burning? Forest Ecology and Management 231:138–154.

Jansen, W., J. Bohmer, B. Kappus, T. Beiter, B. Breitinger, and C. Hock. 2000. Benthic invertebrate and fish communities as indicators of morphological integrity in the Enz River (southwest Germany). Hydrobiologia 422:331–342.

Koel, T.M. 2004. Spatial variation in fish species richness of the upper Mississippi River system. Transactions of the American Fisheries Society 133:984–1003.

Mantel, N. 1967. The detection of disease clustering and a generalised regression approach. Cancer Research 27:209–220.

McGarigal, K., S. Cushman, and S. Stafford 2000. Multivariate Statistics for Wildlife and Ecology Research. Springer-verlag, New York, NY.

Nei, M. 1972. Genetic distance between populations. American Naturalist 106:283–292.

Nei, M., F. Tajima, and Y. Tateno. 1983. Accuracy of estimated phylogenetic trees from molecular data. II. Gene frequency data. Journal of Molecular Evolution 19:153–170.

Ota, T. 1993. DISPAN. Pennsylvania State University, University Park, PA.

Parasiewicz, P., J. Dodge , J. Legros, J. Rogers, and M. Wirth. 2006. Assessment and instream habitat restoration planning for Pomperaug, Nonnewaug, and Weekeepeemee rivers in Connecticut. Technical Report for the Pomperaug River Watershed Coalition, Southbury, CT 06488. 120 pp.

Pegg, M.A., and M.A. McClelland. 2004. Spatial and temporal patterns in fish communities along the Illinois River. Ecology of Freshwater Fish 13:125–135.

Pielou, E.C. 1984. The Interpretation of Ecological Data: A Primer on Classification and Ordination. John Wiley and Sons, New York, NY.

Pinkus-Rendon, M.A., J.L. Leon-Cortes, and G. Ibarra-Nunez. 2006. Spider diversity in a tropical habitat gradient in Chiapas, Mexico. Diversity and Distributions 12:61–69.

Poff, N.L., and J.V. Ward. 1990. The physical habitat template of lotic systems: Recovery in the context of historical pattern of spatio-temporal heterogeneity. Environmental Management 14:629–646.

R Development Core Team. 2004. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Sandstrom, U.G., P. Angelstam, and G. Mikusinski. 2006. Ecological diversity of birds in relation to the structure of urban green space. Landscape and Urban Planning 77:39–53.

Sneath, P.H.A., and R.R. Sokal 1973. Numerical Taxonomy: The Principles and Practice of Numerical Classification. W.H. Freeman and Co., San Francisco, CA.

Walden, D.L., and P. Parasiewicz. 2005. Integrative assessment of biological and hysical attributes of the Eightmile River. Northeast Instream Habitat Program, Amherst, MA.