ECOLOGY LETTERS

Ecology Letters, (2010) **13**: 1290–1299 doi: 10.1111/j.1461-0248.2010.01523.x

LETTER

Metacommunity phylogenetics: separating the roles of environmental filters and historical biogeography

Mathew A. Leibold,[1]* Evan P. Economo[1†] and Pedro Peres-Neto[2]

[1]Section of Integrative Biology, 1 University Station C0930, University of Texas at Austin, Austin, TX 76712, USA

[2]Department of Biological Sciences, Université du Québec à Montréal, Case Postale 8888, Succursale Centre-Ville Montréal, QC H3C 3P8, Canada

**Correspondence: E-mail: mleibold@mail.utexas.edu*

[†]Present address: University of Michigan, 2014 Kraus Natural Science Building, 830 N. University, Ann Arbor, MI 48109-1048, USA

Abstract

Biogeographical, evolutionary and ecological processes interact to regulate patterns in metacommunities. However, as there are few quantitative methods for evaluating their joint effects, resolving this interaction is difficult. We develop a method that aims to evaluate the interaction between phylogenetic structure, historical biogeographic events and environmental filtering in driving species distributions in a large-scale metacommunity. Using freshwater zooplankton as a case study, we contrast the phylogenetic metacommunity structure of calanoid copepods and an ecologically similar but more vagile group, daphniids, in the northeastern US. We find that legacies of historical biogeographical events have strongly constrained calanoid distributions within this area, but that adaptation to different water chemistry and lake morphology drives the metacommunity structure of daphniids. Our findings show that biogeographic history and metacommunity processes jointly regulate community structure in these lakes and suggest that this also depends on factors that affect the colonization rate of different types of organisms.

Keywords

Calanoida, community phylogeny, daphniidae, historical biogeography, metacommunity, variation decomposition, zooplankton.

Ecology Letters (2010) 13: 1290–1299

INTRODUCTION

The factors that regulate how species are distributed into communities involve a complex mixture of ecological, evolutionary and biogeographical processes (Leibold *et al.* 2004; Ackerly *et al.* 2006; Ricklefs 2007). Traditional ecological approaches focus on processes that result in associations between community composition and environmental gradients via the process of species sorting (see Chase & Leibold 2003). More recent work highlights additional spatial effects due to variation in dispersal rates across taxa and areas, the effects of patch size and isolation on biogeographic rates, and stochastic equilibria driven by spatial structure (Hubbell 2001; Leibold *et al.* 2004). Evolutionary processes including allopatric, parapatric and sympatric speciation that may or may not be associated with ecological speciation or adaptive changes, and post-speciation divergences (Gavrilets & Losos 2009) may also be important. Finally, historical contingencies may affect species distributions into local assemblages (Ricklefs 2007). These should produce patterns that are neither explained by current patterns of spatial landscape structure or environmental heterogeneity, but instead reflect their configuration at some point in the past.

Most past ecological studies of community composition have focused on relatively small spatial scales. Typically they focus on purely ecological hypotheses with the assumption that all sites share a common species pool of potential colonizers. Similarly, most studies of species ranges have ignored or simplified ecological hypotheses and focused on larger spatial scales where sites do not share a common species pool. Recently, however, the difference in scale between these two types of studies have become less distinct because metacommunity ecology has provided ways for ecologists to study ecological phenomena over increasingly larger spatial scales (Leibold *et al.* 2004), because evolutionary ecologists have begun to study how phylogeny affects patterns of species coexistence into local communities within metacommunities (reviewed in Cavender-Bares *et al.* 2009), and because evolutionary biologists have increasingly

incorporated ecological thinking into large-scale biogeographic studies (e.g., McPeek & Brown 2000; Gillespie 2004; Revell *et al.* 2007; McPeek 2008; Moen *et al.* 2009; Rabosky 2009; Stephens & Wiens 2009) even though they do not provide a general quantitative framework for doing so.

Work to date shows that species distributions can have important spatial and environmental determinants involving metacommunity processes (e.g., Cottenie 2005; Beisner *et al.* 2006) and that phylogenetic relationships can be associated with patterns of local coexistence in ways that are thought to reflect ecological processes such as habitat differentiation vs. competitive niche differentiation (see Cavender-Bares *et al.* 2009). These inferences, however, are limited by possible confounding factors. For example, purely spatial pattern that is normally interpreted as reflecting patchiness, mass effects or neutral dynamics in metacommunity analyses, could be due to historical biogeographic legacies or environmental predictors that are themselves spatially structured (Peres-Neto & Legendre 2010). Similarly, low coexistence of closely related species that are often interpreted as being linked to adaptation to spatially structured environmental filters could instead be due to recent effects of allopatric or parapatric speciation or effects of historical barriers. While some of these effects may have originally been caused by environmental barriers (Wiens & Graham 2005), they can result in current patterns of distribution that do not correlate well with environmental conditions because species are absent from sites where they would otherwise be predicted to occur.

Although statistical approaches that integrate historical biogeography and phylogeny are becoming more popular (e.g., Ree & Smith 2008), attempts to bring in an ecological perspective are still in development (e.g., Peres-Neto 2004; Davies *et al.* 2007; Helmus *et al.* 2007). Our ability to detect, differentiate and understand the roles of different metacommunity assembly processes including dispersal limitation, environmental filtering and historical biogeographical events and constraints, has been particularly challenging because of the possibly complex ways that they may simultaneously affect patterns.

In this article, we describe a quantitative framework for linking species distributions to patterns of environmental sorting and the legacy of historical biogeography in a phylogenetic framework that allows us to identify correlations between the distribution of species in a metacommunity and these possible regulating factors. A key parameter is likely to be colonization rate. Vagile organisms are expected to more quickly overcome the legacy of previous historical biogeographic events and sort into appropriate environments than less vagile clades. As a means of illustrating our approach, we applied our framework to test this hypothesis by contrasting the distribution of two different clades of freshwater zooplankton in the northeastern US under the *a priori* hypothesis that one clade (calanoid copepods) would be much more likely to be constrained by historical biogeography than another more vagile group (daphniid cladocerans).

MATERIAL AND METHODS

Our method is a generalization and expansion of the method developed by Legendre *et al.* (1997) called the 'fourth corner' method which has until now only been used to test purely ecological hypotheses. Their approach utilizes a trait matrix, an incidence matrix, and an environment matrix to derive a correlation matrix between species traits and environmental variables which they call the 'fourth corner' correlation matrix. Shortcomings of their method include the fact that it does not allow incorporation into a broader variation decomposition framework and it has a problematic significance test (Dray & Legendre 2008).

Although inspired by the fourth corner method, our method is different in several important ways. First, we incorporate a phylogenetic analysis, allowing us to estimate how environmental and biogeographic constraints on species distributions change through evolutionary events (i.e., across the phylogeny). Second, in addition to examining environmental controls on species distributions, we also evaluate the role of historical biogeography by incorporating *a priori* hypothetical relationships among areas using what we call the 'biogeography matrix'. Finally, we extend the original method (Legendre *et al.* 1997), which used simple correlations, by incorporating partial correlation analysis (akin to partial regression slopes) to control for possible confounding effects of environment on correlations with biogeography. The method does not intend to provide statistical descriptions of the phylogenetic structure of an individual community, such as over- or underdispersion (Webb *et al.* 2002; Cavender-Bares *et al.* 2009). Rather, we use the distribution of species into communities to map the influence of environmental and spatial factors on limiting site occupancy onto the phylogeny (see Materials and methods and Discussion for further details). We use a node-by-node scheme to quantify how environmental variation and biogeography explains phylogenetic differences in community composition across sites. A significant result for a given node–variable pair implies that the two lineages emanating from a node have distributions within the survey that differ in relation to particular site variables. In principle any site attribute could be used, but here we consider environmental and biogeographic ones only (Fig. 1). The method is based on a four-step procedure: (1) derivation of data matrices, (2) standardization of these matrices, (3) calculation of semi-partial correlation matrices and (4) test of these correlations by permutation procedures.

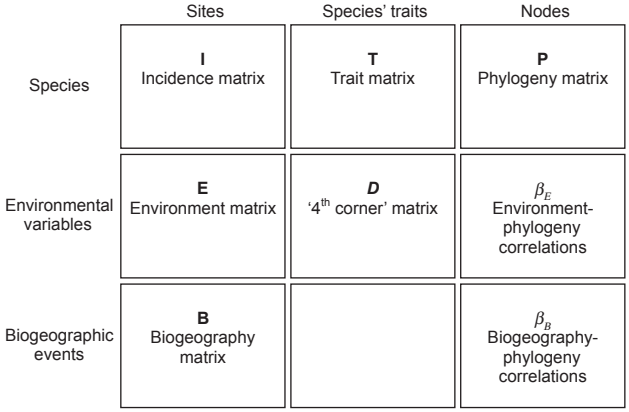

Figure 1 Schematic arrangement of data matrices **I**, **E**, **B**, **P** and **T** and the three association ($\boldsymbol{D}$, $\boldsymbol{\beta_E}$, $\boldsymbol{\beta_B}$) matrices discussed in the text.

Data matrices

We use four data matrices in our analyses that can be derived from data as follows:

The incidence matrix **I** describes which species are found at which sites. Each entry is coded with a 1, if a given species (column) is found at a given site (row) and 0 elsewhere.

The species-by-node matrix **P** contains phylogenetic information. In principle it could be coded in various ways (see Supporting information) but we focus on a node-by-node approach inspired by Felsenstein's phylogenetic independent contrasts (PIC, Felsenstein 1985). We code the species-by-node **P** matrix so as to produce a set of node-by-community statistics contained in matrix $\mathbf{P}_{\mathbf{\Sigma}}$. In matrix **P**, all species descending from one of the branches emanating from a node are all arbitrarily given negative values, and those descended from the other branch are given positive values (which branch is given the negative sign is arbitrary). If a species is not a descendent of a node, it is given a value of 0. The sum of species codes from the two branches equal 1 and −1, thus species located in more species-rich branches are downweighted. The weights each of the species can be adjusted to reflect the phylogenetic topology and branch lengths. For the analyses in this article, we used a topological coding that does not account for branch lengths (which are not available for our study system), but we describe in the Supporting information a method that does so. In our scheme, codes begin at 1 (or −1, depending on the arbitrary sign of the branch) at the origin of the branch, and are reduced by half at each bifurcating node until the species are reached. Thus, a species on a branch with one species has a code of 1, a branch with two species each have a code of (0.5, 0.5), three species gives (0.5, 0.25, 0.25), and so forth (e.g., see Fig. 2). Therefore, the coding associated to any particular node can rebuild the entire original topology for that node.

The entries of the node-by-community $\mathbf{P}_{\mathbf{\Sigma}}$ matrix are simply the sums of **P** values for all species that are both descendents of a given node and occupants of the community (note $\mathbf{P}_{\mathbf{\Sigma}} = \mathbf{IP}$). The $\mathbf{P}_{\mathbf{\Sigma}}$ values reflect the differential representation of species in the community along the two different branches emanating from a node in the phylogeny (i.e., phylogenetic composition or balance of species distributions across the two sides of any give node). The values in $\mathbf{P}_{\mathbf{\Sigma}}$ range between 1, where all daughter species of the one of the branches are present but none of the other are present and −1 in the opposite case. If all species that are daughter of a given node are present then the value is 0, which means there is no evidence from that community that the two branches have a difference in their propensity to occur under the local conditions. Looking across many communities, correlations between these node-community values $\mathbf{P}_{\mathbf{\Sigma}}$ and the site variables of those communities reveals that the two branches emanating from the node have diverged in their response to an environmental filter (e.g., different temperature tolerances) or biogeographic event (e.g., on different sides of a historical dispersal barrier). In the Supporting information, we provide extensive details and discussion regarding our coding scheme and its quantitative equivalence to the PIC approach for reconstructing ancestral states.

The biogeography matrix **B** is meant to identify contrasts between pairs of biogeographic areas that should be differentially affected by historical effects based on some independent evidence (e.g., geological, paleoclimate, evidence from other clades). These are expressed as Helmert contrasts (see Legendre & Legendre 1998) in which sites that occur in one area are coded +1, those in the other area are coded −1 and sites that occur in neither area are coded 0.

The environment matrix **E** is derived by measuring environmental attributes of each site and entering values for each variable (column) at each site (row).

Semi-partial correlation matrices

The first step is to calculate the weighted sum of contrasts across species in each node per community $\mathbf{P}_{\Sigma\text{std}} = \mathbf{IP}_{\text{std}}$. The matrix $\mathbf{P}_{\text{std}}$ is the standardized version of matrix **P**. Given that sites containing a greater number of species or species with a greater number of occurrences provide more reliable estimates regarding possible effects of environment and biogeography on distributions, we applied a weighting factor to standardize the data prior to analysis to avoid effects of rare species and unusual sites (similar rationale were used by Legendre *et al.* 1997 and Layou 2007; and are common in gradient analysis; Peres-Neto *et al.* 2006).

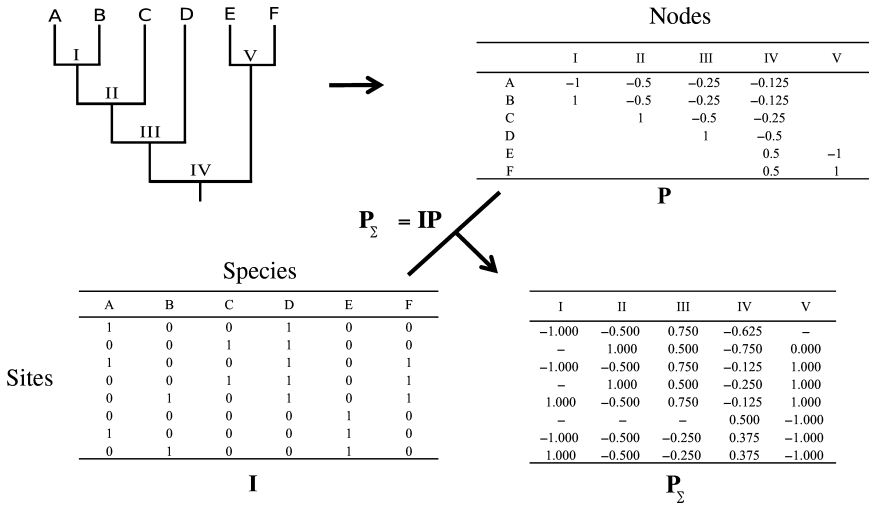

Figure 2 An example depicting the coding of a phylogeny in the species-by-node **P** matrix, and the calculation of the $\mathbf{P}_\Sigma$ statistic for each node-community pair. The p_{ij} value for a given species (i) and node (j) pair reflects the species' contribution to its branch emanating from the node. If a species is not a daughter of a given node, it has no value and does not enter in the calculation. For each node, all species are given negative values on one branch and positive values on the other branch. The $\mathbf{P}_\Sigma$ matrix, whose values are the equivalent of phylogenetic independent contrasts with initial states being the entries of the incidence matrix (1 – present, 0 – absent), is calculated with a simple matrix multiplication $\mathbf{P}_\Sigma = \mathbf{IP}$. Note that in our analyses the **P** is standardized with a method described in the text, and is not depicted above. For more information on this coding scheme, see the Materials and methods in the text and Supporting information.

Computational details of the standardization procedure are provided as Supporting information.

Note that species (columns of **I**) that are not descendents of a particular node in **P** are removed while calculating the sum value for communities based on that node in question. In addition, sites that are occupied only by species that are not descendents of a particular node in **P** are also removed. For example, in Fig. 2, node III, only species from A to D are considered the metacommunity analysis of that node (i.e., species E and F are not included). Therefore, when the species of a particular node are absent from particular sites, these sites are not considered in the analyses for that particular node.

We use semi-partial correlations (Peres-Neto *et al.* 2006; Peres-Neto & Legendre 2010) instead of partial correlations. In partial correlations, both the response and the particular predictor of interest are controlled for the other environmental variables; whereas in semi-partial correlations only the predictor variable is controlled for the other predictors. Therefore, semi-partial correlations correspond to the total amount of variation explained by a particular environmental variable in relation to the total amount of variation in the response matrix $\mathbf{P}_{\Sigma\text{std}}$ instead of the total amount of variation left to be explained in $\mathbf{P}_{\Sigma\text{std}}$ once other environmental variables were already taken into account, which is the case of partial correlations. The semi-partial correlations have a clearer interpretation when compared to partial correlations in which they represent the amount of variance accounted for by each predictor uniquely and it is the statistic used in the classic variation partitioning framework (Peres-Neto *et al.* 2006).

There is no algebraic solution for simultaneously calculating a matrix of semi-partial correlations between response (i.e., $\mathbf{P}_{\Sigma\text{std}}$) and each single predictor variables (i.e., environmental and biogeographical predictors). Therefore, to calculate the contribution of the ith to the jth node, we need to calculate three regression models: (1) using all predictors ($\mathbf{E}_{\text{std}}$ stands for the standardized environmental matrix; see Supporting information): $\beta_{E(j,\text{all})} = (\mathbf{E}^{T}_{\text{std}(j,\text{all})}\mathbf{W}_n\mathbf{E}_{\text{std}(j,\text{all})})^{-1}\mathbf{E}^{T}_{\text{std}(j,\text{all})}\mathbf{W}_n\mathbf{P}_{\Sigma\text{std}(j,\text{all})}$, where T stands from transposed and $\mathbf{W}_n$ is a $(n \times n)$ diagonal matrix with elements equal to the sum of the rows of the incidence matrix **I**, where the first non-zero element of $\mathbf{W}_n$ (i.e., $[\mathbf{W}_n(1,1)]$) is the sum of the occurrences across all species for patch 1, and so on (i.e., trace ($\mathbf{W}_n$) equals the total sum of all species' occurrences); (2) using all predictors but i: $\beta_{E(j,-i)} = (\mathbf{E}^{T}_{\text{std}(-i)}\mathbf{W}_n\mathbf{E}_{\text{std}(-i)})^{-1}\mathbf{E}^{T}_{\text{std}(-i)}\mathbf{W}_n\mathbf{P}_{\Sigma\text{std}}$; and (3) using only predictor i: $\beta_{E(j,i)} = (\mathbf{E}^{T}_{\text{std}(i)}\mathbf{W}_n\mathbf{E}_{\text{std}(i)})^{-1}\mathbf{E}^{T}_{\text{std}(i)}\mathbf{W}_n\mathbf{P}_{\Sigma\text{std}}$. Next, calculate the predicted values for regressions 1 and 2 $\hat{\mathbf{Y}}_{(j,\text{all})} = \mathbf{E}_{\text{std}(j,\text{all})}\beta_{E(j,\text{all})}$ and $\hat{\mathbf{Y}}_{(j,-i)} = \mathbf{E}_{\text{std}(j,-i)}\beta_{E(j,-i)}$, respectively. Finally, the weighted semi-partial correlation for the ith environmental variable for any particular node j can be calculated as:

$$\rho_{(j,i)} = \frac{\beta_{E(j,i)}\sqrt{\frac{\mathrm{trace}(\hat{\mathbf{Y}}^{T}_{(j,\mathrm{all})}\hat{\mathbf{Y}}_{(j,\mathrm{all})})}{\mathrm{trace}(\mathbf{P}^{T}_{\Sigma\mathrm{std}}\mathbf{P}^{T}_{\Sigma\mathrm{std}})} - \frac{\mathrm{trace}(\hat{\mathbf{Y}}^{T}_{(j,-i)}\hat{\mathbf{Y}}_{(j,-i)})}{\mathrm{trace}(\mathbf{P}^{T}_{\Sigma\mathrm{std}}\mathbf{P}^{T}_{\Sigma\mathrm{std}})}}}{|\beta_{E(j,i)}|} = \frac{\beta_{E(j,i)}\sqrt{R^{2}_{E_{(j,\mathrm{all})}} - R^{2}_{E_{(j,-i)}}}}{|\beta_{E(j,i)}|}$$

where $\rho_{(j,i)}$ is the semi-partial correlation between node j and environmental predictor i. $\rho_{(j,i)}$ is based on the difference between the total amount of variation in node j explained by all environmental variables $R^{2}_{E_{(j,\mathrm{all})}}$ and the total amount of variation in node j explained by all environmental variables except i $R^{2}_{E_{(j,-i)}}$. To determine the sign of the correlation, we then multiply the difference by $\beta_{(j,i)}$ and divide it by its absolute value $|\beta_{(j,i)}|$. The same operations are then repeated for the biogeographical matrix **B** by replacing $\mathbf{E}_{\mathrm{std}}$ by $\mathbf{B}_{\mathrm{std}}$ (but see below).

The coefficients based on environmental predictors are based on all sites, but biogeographic events are computed using only sites (columns of **B**) relevant to a given biogeographic event (i.e., sites that do not occur in either area involved in a particular biogeographic comparison are removed from the calculations). To correct for possible correlations of environmental gradients within biogeographic events, we calculate the partial correlation of each biogeographic event including the environmental predictors by concatenating the relevant rows of **B** with the same site rows in environment matrix **E**. Note that the opposite (i.e., control for biogeography while testing for the correlation with the environment) would be analytically impractical as each biogeographic event would end up with a matrix of environmental semi-partial correlations. Moreover, by considering environmental variation within events, one could artificially reduce variation within environmental gradients hence lowering the power to detect important environmental correlates. Thus, biogeographic comparisons are corrected for possible confounding for environmental differences whereas the reverse is not true.

Permutation tests

We determine the significance of semi-partial correlations by permutation based on 9999 randomizations of the **E** and **B** matrices for each comparison. We permute the row position of the **E** or **B** matrix and the column position of the **P** matrix (recalculating $\mathbf{P}_{\Sigma\mathrm{std}}$ for each permutation) to generate a distribution of expected semi-partial correlations and compare these to the actual value to this distribution. A pseudo-F statistic (i.e., pseudo referring to the fact that the statistic used is pivotal under permutation – no degrees of freedom are required as in the traditional F, hence facilitating computation) was calculated as follows: $F_{\mathrm{pseudo}} = (R^{2}_{E_{(j,\mathrm{all})}} - R^{2}_{E_{(j,-i)}})/\mathrm{trace}((\mathbf{P}^{T}_{\Sigma\mathrm{std}} - \hat{\mathbf{Y}}^{T}_{(j,\mathrm{all})})^{T}\ (\mathbf{P}^{T}_{\Sigma\mathrm{std}} - \hat{\mathbf{Y}}^{T}_{(j,\mathrm{all})}))$. The P-values were calculated as: the proportion of random F_{pseudo} equal to or larger than the respective observed F_{pseudo} value. The P-values are then corrected for significance using the false discovery rate (FDR) procedure (Benjamini & Hochberg 1995). All procedures were programmed in Matlab and the functions are available in the Supporting information.

Illustration: freshwater zooplankton in the northeastern US

We applied our method to data on the distribution of zooplankton in the northeastern US and contrasted results for calanoid copepods with those of daphniid cladocerans to test the hypothesis that less vagile clades will have distributions that are more constrained by historical biogeography than more vagile ones. Both clades have generally similar means of dispersal involving the windborne or phoretic dispersal of resting stages via animal vectors. However, the likelihood of post-dispersal establishment is thought to contrast sharply. Cladocerans, including daphniids are thought to have reasonably high colonization rates among lakes because they can establish large populations via asexual reproduction that minimizes mate-finding constraints during annual sexual phases. In contrast, copepods are obligately sexual, generating two problems in establishment. First, they can have strong Allee effects at low densities due to difficulties in finding mates (Sarnelle & Knapp 2004). Second, mating interference by established populations on rare colonists due to behavioural miscues in species recognition may generate strong priority effects (Thum 2007). We used our method to test for the possible role of these general differences in colonizing ability in affecting the influence of biogeographic legacies and environmental adaptation on phylogenetic structure of the metacommunity.

Note that in our case most of the phylogenetic events predate most if not all of the biogeographic events (Colbourne & Hebert 1996; Thum & Harrison 2009), so that associations between them indicate the dynamics of range shifts rather than speciation *per se*. We thus interpret correlations between phylogenetic events and biogeographic events as revealing the effects of phylogeny on the distribution of species into the ancestral pre-Pleistocene refugia hypothesized by Stemberger rather than reflecting actual vicariance events responsible for phylogenetic events. Our method could of course be also used more intuitively in cases where the phylogenetic events are directly associated with biogeography such as would occur under scenarios of allopatric speciation. Regardless of the ambiguity of interpretating such historical effects, our analysis serves as a usefully illustrative and interesting application of the method.

Data on distribution of zooplankton as well as lake morphology and water chemistry came from the EPA

Northeastern Lakes Survey of 1991–1994 (available at: http://www.epa.gov/emap/html/data/surfwatr/data/nelakes.html). We culled this data set to have uniform sampling per lake by selecting the first visit made to each lake. We further culled five lakes out of the 332 in the original data that did not have complete water quality and lake morphology data. We selected environmental variables *a priori* that were likely to be relatively independent of each other to minimize covariances among environmental variables. We retained the following environmental variables in our analysis (codes in Fig. 3 are in parentheses): colour (COLOUR), conductivity (COND), dissolved organic carbon (DOC), total nitrogen (NTL), Ph (PHEQ), total phosphorus (PTL), average depth (AV_DEP), elevation (ELEV) and the ratio of total nitrogen to total phosphorus (TN:TP).

We assembled genera and species-wide composite phylogeny based on molecular studies depicting only the taxa of interest (i.e., species within the studied region) and not all species in these clades. Data on diaptomid calanoid phylogenies came from Thum (2004), the *Daphnia* phylogeny came from Colbourne & Hebert (1996) and Omilian & Taylor (2001). We further assumed that two closely related genera to the diaptomids and *Daphnia* were monophyletic and sister clades (i.e., we assume that the two species of *Simocephalus* are monophyletic and sister to the *Daphnia* clade and that two species of *Epischura* are monophyletic and sister taxa to all other diaptomids).

We used the reconstruction of historical biogeography for the area from Stemberger (1995) to generate six testable *a priori* contrasts attributable to historical biogeography. For example, Stemberger argued that lakes west of the Hudson River were likely colonized from refugia in Mississippi Valley and Appalachian Plateaus whereas those east of the Hudson River were colonized from refugia in the Georges Bank and Sable Bank (areas that were then above sea level). The biotas in these different regions would thus be distinct if the effects of these refugia still affect current distributions and this can be tested as a biogeographical contrast. Although Stemberger's hypothesis was in part shaped by data from the survey (Stemberger 1995), the hypothesis itself was derived from geological inferences about post-Pleistocene changes involving glacial retreats and subsequent uplifting of different land masses. To generate specific contrasts, we examined Fig. 2 in Stemberger (1995) and the associated text and identified six independent comparisons consistent with that hypothesis. These are (codes denoted in Fig. 2 are in parentheses): east vs. west of the Hudson river (S1), Adirondacks vs. other lakes west of the Hudson (S2), east vs. west of the Connecticut River east of the Hudson (S3), Maine vs. other lakes east of the Hudson (S4), New Hampshire vs. other lakes to its west but east of the Hudson (S5),

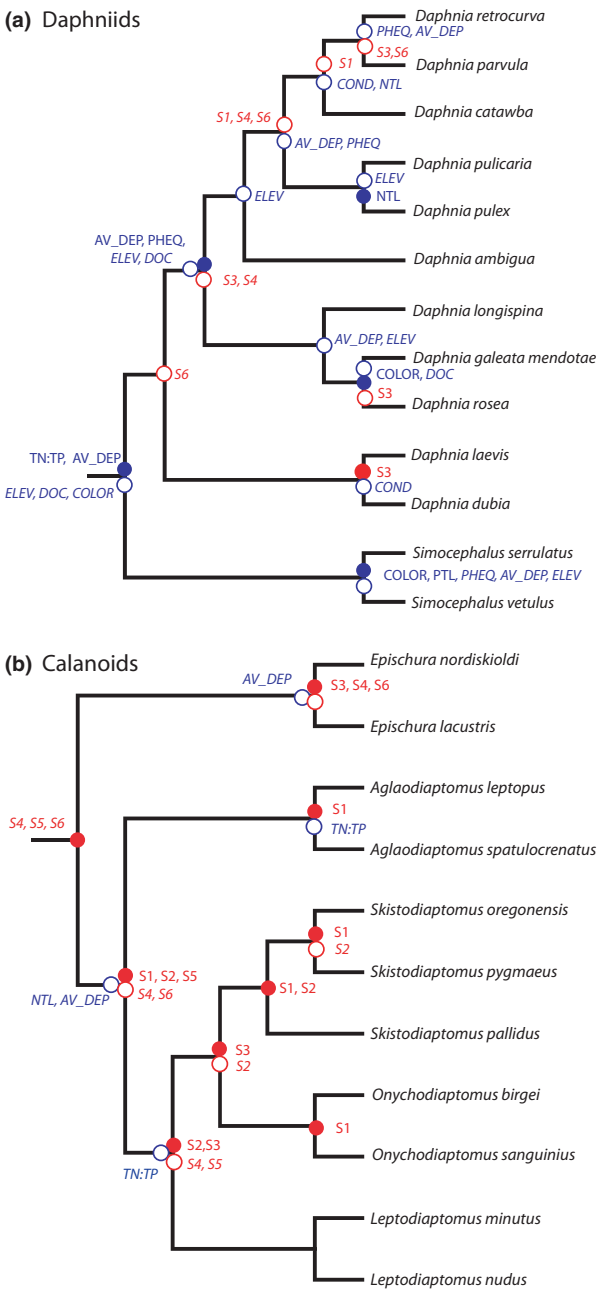

Figure 3 Graphical mapping of significant associations between environment and phylogeny (blue dots) and historical biogeography and phylogeny (red dots) for (a) daphniids and (b) calanoids. The quantitative results are shown in the Supporting information, Table S1. For each significant correlation, we plotted its location (node) on the phylogeny. Significant associations involving environmental factors are shown with blue dots, those for historical biogeography are shown in red. Each is coded to different environmental variables matching codes described in the text. Significant contrasts after FDR correction for multiple comparisons are shown as solid symbols, additional contrasts that are significant only as isolated tests (no correction) are shown as open symbols.

New Hampshire vs. lakes in Rhode Island and Massachusetts (S6).

RESULTS

The results of our analyses (Fig. 3; Table S1 in Supporting information) show strongly contrasting patterns between the two clades. Calanoids show little evidence for any phylogenetic signal on the distribution of species along water chemistry or lake morphology gradients. Only four nodes indicate any likely association and these are not statistically significant when corrected for multiple tests. In contrast to the effects of environment, calanoids show rather strong evidence for effects of historical biogeography in this area with seven of the ten nodes showing significant biogeographic contrasts based on Stemberger's (1995) hypotheses about post-Pleistocene recolonization of the area.

The results for daphniids differ strongly from the pattern found in calanoids. In the daphniids, there was much stronger evidence that past phylogenetic events affect the current distribution of species along water chemistry and lake morphology gradients with nine of the twelve nodes showing likely correlations, four of which remain statistically significant even when corrected for multiple tests. In contrast with calanoids, there were much weaker correlations between phylogenetic events and biogeographic factors, none of which were significant when correcting for multiple tests. The differences in the number of significant associations in the two matrices between daphniids and calanoids are significant (biogeography: χ^2 [d.f., 1] = 157.21, $P < 0.0001$), environment: χ^2 [d.f., 1] = 6.4, $P < 0.05$).

DISCUSSION

We introduced a quantitative framework to evaluate the relative importance of evolutionary events into different environments and the effects of historical biogeographic events on the spatial and phylogenetic structure of metacommunities. We predicted that these effects would vary in relative importance between two clades of zooplankton differing in life history characteristics, and our method detected correlations consistent with our predictions. The distributions of daphniids in this area were largely due to species sorting along environmental gradients, whereas the distribution of calanoids on the same landscape was primarily due to the legacy of historical biogeographic events. Our approach builds on previous work relating community structure to phylogeny (Cavender-Bares *et al.* 2009) by providing a way to incorporate biogeographic effects into the study of phylogenetic processes in community assembly. In doing so, we also introduced a novel way of using phylogenetic variation in analysing metacommunities by determining the relative importance of environmental variables and biogeographical events (or any other patch property such as connectivity, for instance) along temporal evolutionary events (i.e., node codes; Fig. 2). In this way, we can map onto the phylogeny the important of different patch properties to the spatial distribution of the clades involved.

There has been a growing interest in linking community distributional data to phylogenetic structure (Cavender-Bares *et al.* 2009; Cadotte *et al.* 2010) and a parallel increase in numbers of methods to explore these patterns (Pausas & Verdú 2010). Most methods either explore variance (i.e., phylogenetic diversity) or location (phylogenetic composition) components of the metacommunity phylogenetic distribution. For instance, methods exploring assembly patterns such as phylogenetic under and overdispersion (e.g., Peres-Neto 2004; Cavender-Bares *et al.* 2006) can be seen as assessing the variance component whereas methods exploring whether communities having similar phylogenetic composition inhabit similar environments can be seen as assessing the location component (e.g., Emerson & Gillespie 2008; Pillar & Duarte 2010). In a different direction, our method focuses on measuring phylogenetic unbalance within nodes rather than estimating phylogenetic composition and diversity. Our approach brings at least two important innovations over other current approaches. First, we can, in cases where phylogenetic events are concurrent with biogeographic ones, determine whether and how distributional patterns correlate with environmental filtering and/or biogeography as a function of evolutionary events (time) such as speciation. By considering all species at once (i.e., current approaches), one cannot separate these changes in metacommunity assembly processes through evolutionary time (but see Pavoine *et al.* 2010 for a node-by-node application regarding trait diversity). Second, our approach can differentiate between metacommunity processes that may lead to conflicting patterns across species, where species in different clades may be governed by different types of metacommunity processes. For instance, some local communities may be composed of species that are on average more related than a random expectation (i.e., phylogenetic underdispersion), whereas others contain species that are less related than a random expectation (phylogenetic overdispersion; Cavender-Bares *et al.* 2009). If these phylogenetic patterns are driven by environmental filtering or biogeographic events (or other processes such as local competition or patch connectivity), local communities composed by a mix of these patterns (i.e., some species are underdispersed whereas other are overdispersed) will appear as being non-structured. Using a node-by-node approach, we consider variation at all levels of the phylogeny and these issues can be resolved or at least explored. For instance,

imagine that two clades have opposite relationships with an important environmental variable (one clade positive, whereas the other negative). By analysing both clades together, these relationships would cancel each other out and not reveal the nature of their patterns. Similarly, our node-by-node framework could be adapted to explore issues regarding conservatism vs. divergence in habitat affinity in a metacommunity as a parallel to niche conservatism and niche divergence in species traits (Wiens & Graham 2005).

Like many other studies, our study is limited in its inferences by the data available for quantitative analyses. In our case, we ignore species that are not found within our data set and we use composite phylogenies with arbitrary branch lengths that ignore phylogenetic uncertainty (see Nakagawa & Freckleton 2008 for a discussion on these issues and potential ways to address these problems). Although it is beyond the scope of this article, future work could examine the robustness of such inferences to different types of missing data that are common in phylogenetic studies and their applications. In general, while there are data limitations of the current study, these do not reflect general limitations of the method if more complete data were available nor do they prevent us from making conclusions that carefully consider these possible limitations.

The phylogenetic structure of the lake metacommunity, we analysed reveals contrasting roles of biogeographic events and environmental factors in explaining the distribution of daphniids vs. calanoids across the eastern North America. They indicate that two major groups of zooplanktonic organisms that share resources and predators have distributions that have distinct phylogenetic signatures in their spatial distributions. Daphniids appear to have experienced substantial niche divergence that result in affinities for distinct water chemistry and lake attributes but their distributions are unconstrained by biogeographic factors across the Northeast US. Calanoids that coexist both locally and regionally with the daphniids in the same part of the US have distributions that still retain strong historical signal with less differentiation in distinct habitat types.

Our study shows that geographic range shifts and adaptations to different environmental conditions both affect the structure of this aquatic metacommunity. It shows that the relative roles of these two processes contrasts for the two clades in ways related to their differences in their overall dispersal abilities. Our study, however, does not address different but related questions about the overall evolutionary history of these clades. To do so would require more extensive phylogenetic information about other species in each of these clades that are not found in this area as well as information about their distributions in other areas. However, our method could be used more broadly, if such data were available.

Our approach allows the joint quantitative evaluation of ecological, biogeographic and macroevolutionary processes on patterns of metacommunity composition to be mapped onto a phylogeny. Previous work in metacommunity ecology has hypothesized that different dynamics may occur depending on how colonization rate compares to local per capita turnover rates and to background local population extinction rates (Leibold *et al.* 2004). However, this previous work made the assumption that the regional species pool is uniform over the entire metacommunity and that there are no historical effects of biogeography. Our findings indicate that this assumption is not necessarily warranted at larger spatial scales. They suggest that interactions between historical biogeography and ecological factors may be controlled by how fast the colonizing abilities of taxa can overcome the legacies of dynamic biogeographic processes. It is likely that metacommunity dynamics and historical biogeography interface in many other groups of organisms across a wide range of ecosystems.

ACKNOWLEDGEMENTS

We dedicate this article to Richard Stemberger (1945–2007). Thanks to Richard and to Ryan Thum for help with interpreting the NELS database and comments. We also thank Sharon Strauss for initial stimulation of this idea and the NCEAS working group on 'Metacommunity Evolution' and the 'Community Phylogeny' reading group at U. Texas including Dan Bolnick and David Cannatella for discussion. Helpful reviews from J. Wiens and two other referees greatly improved the manuscipt. This research was funded by NSF grants DEB-0717370 and DEB-0640302, by NCEAS, and an NSF GRF to EPE.

REFERENCES

Ackerly, D.D., Schwilk, D.W. & Webb, C.O. (2006). Niche evolution and adaptive radiation: testing the order of trait divergence. *Ecology*, 87, S50–S61.

Beisner, B.E., Peres-Neto, P.R., Lindstrom, E.S., Barnett, A. & Longhi, M.L. (2006). The role of environmental and spatial processes in structuring lake communities from bacteria to fish. *Ecology*, 87, 2985–2991.

Benjamini, Y. & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc.*, 57, 289–300.

Cadotte, M.W., Davies, J., Regetz, J., Kemberl, S.W., Cleland, E. & Oakley, T.H. (2010). Phylogenetic diversity metrics for ecological communities: integrating species richness, abundance and evolutionary history. *Ecol. Lett.*, 13, 96–105.

Cavender-Bares, J., Keen, A. & Miles, B. (2006). Phylogenetic structure of Floridian plant communities depends on taxonomic and spatial scale. *Ecology*, 87, S109–S122.

Cavender-Bares, J., Kozak, K.H., Fine, P.V.A. & Kembel, S.W. (2009). The merging of community ecology and phylogenetic biology. *Ecol. Lett.*, 12, 693–715.

Chase, J.M. & Leibold, M.A. (2003). *Ecological Niches: Linking Classical and Contemporary Approaches: Interspecific Interactions*. University of Chicago Press, Chicago.

Colbourne, J.K. & Hebert, P.D.N. (1996). The systematics of North American Daphnia (Crustacea: Anomopoda): a molecular phylogenetic approach. *Philos. Trans. R. Soc. Lond. B*, 351, 349–360.

Cottenie, K. (2005). Integrating environmental and spatial processes in ecological community dynamics. *Ecol. Lett.*, 8, 1175–1182.

Davies, T.J., Meiri, S., Barraclough, T.G. & Gittleman, J.L. (2007). Species coexistence and character divergence across carnivores. *Ecol. Lett.*, 10, 146–152.

Dray, S. & Legendre, P. (2008). Testing the species traits-environment relationships: the fourth corner problem revisited. *Ecology*, 89, 3400–3412.

Emerson, B.C. & Gillespie, R.G. (2008). Phylogenetic analysis of community assembly and structure over space and time. *Trends Ecol. Evol.*, 23, 621–630.

Felsenstein, J. (1985). Phylogenies and the comparative method. *Am. Nat.*, 125, 1–15.

Gavrilets, S. & Losos, J.B. (2009). Adaptive radiation: contrasting theory with data. *Science*, 323, 732–737.

Gillespie, R. (2004). Community assembly through adaptive radiation in Hawaiian spiders. *Science*, 303, 356–359.

Helmus, M.R., Savage, K., Diebel, M.W., Maxted, J.T. & Ives, A.R. (2007). Separating the determinants of phylogenetic community structure. *Ecol. Lett.*, 10, 917–925.

Hubbell, S.P. (2001). *The Unified Neutral Theory of Biodiversity*. Princeton University Press, Princeton.

Layou, K.M. (2007). A quantitative null model of additive diversity partitioning: examining the response of beta diversity to extinction. *Paleobiology*, 33, 116–124.

Legendre, P. & Legendre, L. (1998). *Numerical Ecology*, 2nd English edn. Elsevier Science, Amsterdam.

Legendre, P., Galzin, R. & HarmelinVivien, M.L. (1997). Relating behavior to habitat: Solutions to the fourth-corner problem. *Ecology*, 78, 547–562.

Leibold, M.A., Holyoak, M., Mouquet, N., Amarasekare, P., Chase, J.M., Hoopes, M.F. *et al.* (2004). The metacommunity concept: a framework for multi-scale community ecology. *Ecol. Lett.*, 7, 601–613.

McPeek, M.A. (2008). The ecological dynamics of clade diversification and community assembly. *Am. Nat.*, 172, E270–E284.

McPeek, M.A. & Brown, J.M. (2000). Building a regional species pool: diversification of the Enallagma damselflies in eastern North American waters. *Ecology*, 81, 904–920.

Moen, D.S., Smith, S.A. & Wiens, J.J. (2009). Community assembly through evolutionary diversification and dispersal in Middle American treefrogs. *Evolution*, 63, 3228–3247.

Nakagawa, S. & Freckleton, R. (2008). Missing inaction: the dangers of ignoring missing data. *Trends Ecol. Evol.*, 23, 592–596.

Omilian, A.R. & Taylor, D.J. (2001). Rate acceleration and long-branch attraction in a conserved gene of cryptic daphniid (Crustacea) species. *Mol. Biol. Evol.*, 18, 2201–2212.

Pausas, J.G. & Verdú, M.(2010). The jungle of methods for evaluating phenotypic and phylogenetic structure of communities. *Bioscience*, 60, 614–625.

Pavoine, S., Baguette, M. & Bonsall, M.B. (2010). Decomposition of trait diversity among the nodes of a phylogenetic tree. *Ecol. Monog.*, 80, 485–507.

Peres-Neto, P.R. (2004). Patterns in co-occurrence of stream fish communities: the role of site suitability, morphology and phylogeny versus species interactions. *Oecologia*, 140, 352–360.

Peres-Neto, P.R. & Legendre, P. (2010). Estimating and controlling for spatial autocorrelation in the study of ecological communities. *Glob. Ecol. Biogeogr.*, 19, 174–184.

Peres-Neto, P.R., Legendre, P., Dray, S. & Borcard, D. (2006). Variation partitioning of species data matrices: estimation and comparison of fractions. *Ecology*, 87, 2614–2625.

Pillar, V.D. & Duarte, L.D.S. (2010). A framework for metacommunity analysis of phylogenetic structure. *Ecol. Lett.*, 13, 587–596.

Rabosky, D.L. (2009). Ecological limits and diversification rate: alternative paradigms to explain variation in species richness among clades and regions. *Ecol. Lett.*, 12, 735–743.

Ree, R.H. & Smith, S.A. (2008). Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Syst. Biol.*, 57, 4–14.

Revell, L.J., Johnson, M.A., Schulte, J.A., Kolbe, J.J. & Losos, J.B. (2007). A phylogenetic test for adaptive convergence in rock-dwelling lizards. *Evolution*, 61, 2898–2912.

Ricklefs, R.E. (2007). History and diversity: explorations at the intersection of ecology and evolution. *Am. Nat.*, 170, S56–S70.

Sarnelle, O. & Knapp, R.A. (2004). Zooplankton recovery after fish removal: limitations of the egg bank. *Limn. Ocean.*, 49, 1382–1392.

Stemberger, R.S. (1995). Pleistocene refuge areas and postglacial dispersal of copepods of the northeastern United States. *Can. J. Fish. Aquat. Sci.*, 52, 2197–2210.

Stephens, P.R. & Wiens, J.J. (2009). Bridging the gap between community ecology and historical biogeography: niche conservatism and community structure in emydid turtles. *Mol. Ecol.*, 18, 4664–4679.

Thum, R.A. (2004). Using 18S rDNA to resolve diaptomid copepod (Copepoda: Calanoida: Diaptomidae) phylogeny: an example with the North American genera. *Hydrobiologia*, 519, 135–141.

Thum, R.A. (2007). Reproductive interference, priority effects and the maintenance of parapatry in Skistodiaptomus copepods. *Oikos*, 116, 759–768.

Thum, R.A. & Harrison, R.G. (2009). Deep genetic divergence among morphologically similar and parapatric Skistodiaptomus (Copepoda: Calanoids: Diaptomidae) challenge the hypothesis of Pleistocen speciation. *Biol. J. Linn. Soc.*, 96, 150–156.

Webb, C.O, Ackerly, D.D, McPeek, M.A. & Donoghy, M.J. (2002). Phylogenies and community ecology. *Ann. Rev. Ecol. Syst.*, 33, 475–505.

Wiens, J.J. & Graham, C.H. (2005). Niche conservatism: integrating evolution, ecology, and conservation biology. *Ann. Rev. Ecol. Evol. Syst.*, 36, 519–539.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Data S1 Quantitative results illustrate in Figure 3, a detailed discussion of the phylogenetic coding.
Data S2 Matlab code used in the analyzes described in this paper.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

Editor, John Wiens
Manuscript received 11 February 2010
First decision made 23 March 2010
Second decision made 17 June 2010
Manuscript accepted 14 July 2010