From Genes to Ecosystems: Measuring Evolutionary Diversity and Community Structure with Forest Inventory and Analysis (FIA) Data

Kevin M. Potter

Abstract: Forest genetic sustainability is an important component of forest health because genetic diversity and evolutionary processes allow for the adaptation of species and for the maintenance of ecosystem functionality and resilience. Phylogenetic community analyses, a set of new statistical methods for describing the evolutionary relationships among species, offer an innovative approach for assessing the health of forest communities from an evolutionary perspective. Forest Inventory and Analysis data are ideal for conducting phylogenetic community analyses for forest tree species at broad scales. FIA data from 100,000 plots across the conterminous United States were used to investigate the evolutionary characteristics of forest tree communities. This required generating a phylogenetic “evolutionary tree” of the 311 forest tree species inventoried by FIA, based on recent gene sequencing studies of several plant groups. Phylogenetic diversity was quantified for each plot; this statistic sums the evolutionary age of the species present, using the phylogenetic tree. This is a more meaningful way to quantify biodiversity than species richness because, rather than weighting all species on a plot equally regardless of relatedness, it measures their cumulative evolutionary age. General patterns at the ecoregion section scale were similar between mean plot-level species richness and phylogenetic diversity, but important differences also existed. Additionally, the analyses described each plot’s phylogenetic community structure, or whether the species were more clustered or dispersed across the 311-species reference evolutionary tree than expected by chance. The most phylogenetically overdispersed sections were located in the Interior West, while the most phylogenetically clustered included several in the Upper Midwest, New England, California and the Southeastern Coastal Plain. Communities that are phylogenetically clustered consist of more closely related species and therefore may be more susceptible to threats such as pests and climate change. Phylogenetically dispersed communities may be more resilient to these pressures, because greater evolutionary diversity is expected to translate into a greater likelihood that more species will be unaffected by, or will be adaptable to, environmental changes.

Keywords: Biodiversity, evolutionary biology, landscape genetics, forest health monitoring.

1 North Carolina State University; Department of Forestry and Environmental Resources; Forest Health Monitoring Program; 3041 Cornwallis Road, Research Triangle Park, NC 27709; kevinpotter@fs.fed.us.
Introduction

The integration of three diverse scientific disciplines – landscape ecology, conservation biology, and evolutionary genetics – can offer innovative and valuable insights about the health of forested ecosystems, particularly at large scales (figure 1). Approaching conservation from a landscape ecology perspective is vitally important because large-scale processes affect biodiversity and because management decisions are often made at regional levels. At the same time, it is imperative to consider genetic processes in a landscape ecology context because ecological dynamics at a variety of scales can affect the evolution of species. Understanding conservation biology in the light of evolutionary processes, meanwhile, is necessary to measure and manage forest biodiversity. Forest health monitoring work that integrates all three disciplines of landscape ecology, conservation, and genetics is now possible because of the availability of recent advancements in computing, statistical analysis, and molecular biology. This work is particularly relevant to assessing forest genetic sustainability across large scales using U.S. Forest Service Forest Inventory and Analysis (FIA) Phase 2 inventory data.

Figure 1: Relationships among landscape ecology, conservation biology and evolutionary genetics within the context of monitoring forest health and sustainability.
Forest genetic sustainability is an important component of forest health because genetic diversity and evolutionary processes allow for the adaptation of species and for the maintenance of ecosystem functionality and resilience. Genetic diversity within species, for example, reflects the recent and existing integrity of evolutionary and ecological processes (Brown et al. 2000). At the same time, existing evolutionary lineages will generate future biodiversity, and are therefore a cornerstone of environmental health (Erwin 1991). The importance of forest genetic diversity is reflected in the Criteria and Indicators for the Conservation and Sustainable Management of Temperate and Boreal Forests, which includes three indicators of genetic diversity under Criterion 1 (conservation of biological diversity):

1. number and geographic distribution of forest associated species at risk of losing genetic variation and locally adapted genotypes,

2. population levels of selected representative forest associated species to describe genetic diversity, and

3. status of on-site and off-site efforts focused on conservation of genetic diversity (Montréal Process Working Group 2008).

The first two of these indicators attempt to quantify the number and population levels of individual species at risk of losing genetic variability, while the third indicator seeks to evaluate the success gene conservation efforts for forest species, presumably those at risk. All three are valuable measurements, but each is limited to a handful of indicator species and therefore does not address the evolutionary diversity present across entire forest communities. A new approach, however, can now combine nationwide FIA forest inventory data, an ecoregion-scale perspective, and recent advancements in gene sequencing and molecular systematics to quantify evolutionary diversity for communities of species across large spatial scales.

This approach, community phylogenetic analyses (Webb and others 2002), offers the ability to synthesize evolutionary biology and landscape ecology in the context of assessing forest community health, both in terms of biodiversity and community resilience to stress. Community phylogenetic analyses are now possible across broad taxonomic species groupings (forest trees, for example) and across large spatial scales because of recent advances in the ability to create robust phylogenies describing the relationships among species. These phylogenies are hypothesized “family trees” of species (figure 2) that are developed by surveying molecular systematic studies, which compare differences and similarities in gene sequences to determine the evolutionary relationships among the species and to estimate how long ago species diverged from each other. Fossil evidence from palaeobotanical studies can further calibrate the dates at which species or species groups diverged. The branches within these phylogenetic trees often are measured in millions of years. Constructing such
phylogenetic trees does not always require new gene sequencing work, because molecular systematic studies now have been published for a wide variety of plant groups.

Figure 2: A phylogenetic tree of the hypothesized evolutionary relationships among true fir (Abies) species of North America inventoried by FIA. The scale bar at the bottom of the figure depicts time in millions of years. Such phylogenetic trees are hypotheses of the relationships among species, and can be improved over time with additional gene sequencing studies and with additional information unearthed in the fossil record.

Such phylogenetic trees are necessary for two types of phylogenetic community analyses that are useful in describing evolutionary relationships within forest tree communities at broad spatial scales: the generation of phylogenetic diversity statistics and the quantification of phylogenetic community structure.

Phylogenetic diversity statistics (Faith 1992, Webb and others 2006) are particularly meaningful as a measure of biodiversity because they quantify the cumulative evolutionary age and evolutionary potential of all the species in the community of interest. This is typically done by first generating a phylogenetic tree encompassing the species present in a community, based on existing molecular systematics research. Next, the lengths of the branches of that phylogenetic tree, usually measured in millions of years of evolutionary time, are measured and summed. This approach may be more meaningful than traditional biodiversity metrics such as species richness and abundance, which weight the value of all species equally regardless of their relatedness (figure 3). Maintaining the evolutionary potential of groups of species measured in this way has become an increasingly important conservation goal (Rodrigues and Gaston 2002, Sechrest and others 2002, Soltis and Gitzendanner 1999).
Figure 3: A comparison between two measurements of biodiversity, species richness and phylogenetic diversity. Species richness is the number of species present in the community, while phylogenetic diversity is the summed evolutionary age of all the species in the community, as measured by the phylogenetic tree encompassing the species present. Community (A) has greater species richness than community (B), but community (B) has greater phylogenetic diversity because the species in the community contain greater evolutionary distance. The branch lengths shown on the phylogenetic trees are for demonstration purposes, and do not depict actual evolutionary relationships among the species.

A second set of phylogenetic analyses focuses on determining whether the species within a specific community are more phylogenetically clustered or dispersed on the evolutionary tree of life than expected by chance (Webb 2000). Such an analysis would, for example, compare the phylogenetic tree of the species within a single FIA plot to the phylogenetic tree that includes all the tree species in North America. If the species on the FIA plot are more closely related than expected by random chance, when compared to the larger reference phylogenetic tree, then the community is phylogenetically clustered (figure 4a). If the species on the plot are less closely related than expected by chance, then the community is phylogenetically overdispersed (figure 4b). This type of analysis is useful from
a forest health perspective because it offers a way to quantify a community’s potential evolutionary resilience in the face of pests, pathogens, climate change and other stresses. That is because communities made up of species that are overdispersed on the phylogenetic tree possess greater-than-expected evolutionary diversity, and may therefore encompass a higher proportion of species unaffected by a given stressor or able to adapt to it. Phylogenetically clustered communities, meanwhile, contain less evolutionary diversity and may be at greater cumulative risk from stressors that might affect several closely related species, such as sudden oak death (*Phytophthora ramorum*).

**Figure 4:** Phylogenetic trees of hypothetical tree communities representing the two types of phylogenetic community structure, (A) phylogenetic clustering and (B) phylogenetic overdispersion. The species in red are those present in the community; the species in black and red together encompass the phylogenetic reference tree of species that could exist in the community.

The work described in this paper had three main objectives:

1. Assess the usefulness of FIA tree inventory datasets for phylogenetic diversity and community structure analyses across large ecoregion scales.

2. Compare mean FIA plot-level tree species richness to mean plot phylogenetic diversity at the ecoregion section scale.

3. Test for correlations between mean plot measures of phylogenetic community structure and environmental variables at the ecoregion section scale.
Methods

The software package Phylocom 3.41 (Webb and others 2007) was used to calculate phylogenetic diversity and phylogenetic community structure statistics for the forest trees present on each of 102,304 one-sixth-acre Forest Inventory and Analysis (FIA) plots across the 48 conterminous United States. These plots represented the latest available FIA Phase 2 tree and sapling inventory data (trees ≥ 1 inch dbh) as of November 2007 (Forest Inventory and Analysis Program 2007). A phylogenetic reference supertree (figure 5), encompassing the 311 forest tree species inventoried by FIA across the conterminous United States, was constructed for the estimation of phylogenetic distance among species in units of millions of years (Potter in review).

Figure 5: The phylogenetic supertree used in this analysis, incorporating 311 North American forest tree species inventoried by FIA. Evolutionary relationships and branch lengths were based on a survey of approximately 70 recent molecular systematic and paleobotanical studies, with the exception of basal angiosperm relationships, which taken from Wikstrom et al. (2001). Branch lengths are measured in millions of years.
Based on its approximate geographic coordinates, each FIA plot was assigned with ArcMap 9.2 (ESRI 2006) to the appropriate ecoregion section (Bailey 1995), using the most recent geographical information system (GIS) mapping of Bailey’s hierarchical system of ecoregion domains, divisions, provinces and sections (Cleland and others 2005). Ecoregion sections containing fewer than 25 plots were excluded from the analyses to ensure an adequate sampling.

Species richness was calculated for each FIA plot, as was Faith’s (1992) index of phylogenetic diversity (PD). This quantifies the total evolutionary history represented by the species in a community by determining the total phylogenetic branch length distance of the species on the plot, divided by the total branch length distance of the reference phylogenetic supertree encompassing all the North American tree species inventoried by FIA. Mean PD and species richness values were calculated for ecoregion sections.

Additionally, two measures of phylogenetic community structure were calculated for each FIA plot. Both statistics measure whether the configuration of the phylogenetic tree, encompassing the species in a given community, is more or less clustered than expected by chance, as compared to a regional species pool. In this analysis, the regional species pool for each plot consisted of all the species inventoried on all the FIA plots within the same ecoregion section as the plot in question. (See Potter (in review) for details about the calculation of these statistics.)

1. The Nearest Taxon Index (NTI) is a standardized measure of the branch-tip phylogenetic clustering or overdispersion of the species on the FIA plot, regardless of the arrangement of the higher level groups in the phylogenetic tree (Webb and others 2006). In other words, NTI measures whether evolutionary diversity among the species on the plot is higher (overdispersed) or lower (clustered) than expected by chance, as compared to the ecoregion section species pool.

2. The Net Relatedness Index (NRI) is a standardized measurement of basal or tree-wide phylogenetic clustering or overdispersion of the species on the FIA plot (Webb and others 2006). In other words, NRI measures whether evolutionary diversity among the deeper phylogenetic ranks (families, orders, classes and divisions) present on the plot is higher (overdispersed) or lower (clustered) than expected by chance, again as compared to the ecoregion section species pool.

NTI and NRI measure different evolutionary characteristics of communities, so the two statistics could show different results for a plot, for example indicating clustering by one metric and overdispersion by the other. NTI and NRI values are negative when species on a plot do not occur together with closely related species (overdispersed), and positive when species occur with other closely related species (clustered) (Kembel and Hubbell 2006). The means for plot-level NRI
and NTI values were calculated across ecoregion sections, with a \( t \) test determining whether the section-level mean values were significantly different from 0.

The possibility of large-scale correlations of these two phylogenetic community structure measures with ecological variables was tested for several climate (PRISM Group 2008), soil (Miller and White 1998) and topographic (United States Geological Survey 1996) variables at the ecoregion section scale. The ecoregion section means of each of these was determined across the forested area of the section, using a forest cover map (1 km\(^2\) resolution) derived from Moderate Resolution Imaging Spectroradiometer (MODIS) imagery by the U.S. Forest Service Remote Sensing Applications Center (United States Department of Agriculture Forest Service 2008).

**Results and Discussion**

Values of both mean plot species richness and phylogenetic diversity tended to be higher in ecoregion sections of the eastern United States than in the western part of the country (figure 6). The sections with the lowest means for both statistics were located in the interior West. While the general pattern was similar between these two measures of biodiversity, important differences were apparent. Specifically, phylogenetic diversity was higher relative to species richness in several ecoregion sections along the Pacific coast, from central California to Washington, as well as in sections in the Northeast and in the Great Lakes region.

Not surprisingly, the 10 ecoregion sections with the highest mean species richness values are all located in the southeastern United States. However, only six of the 10 sections with the highest mean phylogenetic diversity are located in the Southeast, with the other four in the Northeast. Ecoregion sections with high phylogenetic diversity relative to species richness appear to have a combination of both high angiosperm (flowering plant) and gymnosperm (cone-bearing) species richness (Potter in review).

Meanwhile, the analysis of phylogenetic community structure detected regional patterns in phylogenetic overdispersion and clustering (figure 7). Both the branch-tip Nearest Taxon Index (NTI) and tree-wide Net Relatedness Index (NRI) metrics indicated the existence of phylogenetic clustering in some of the most northerly ecoregion sections, including those in the Northeast, the Great Lakes region, the northern Great Plains, and the Pacific Northwest, as well as along the Appalachian and Sierra Nevada mountain chains. Slight to moderate overdispersion for both metrics was present for sections in the Interior West, particularly in the Southwest.
Figure 6: Mean FIA plot-level forest tree species richness (A) and Faith’s index of phylogenetic diversity (B) across ecoregion sections. The two statistics were divided into six equal interval classes for comparison purposes.

At the same time, some differences were apparent between the two measures. Perhaps the most notable difference is in the Southeast, where six sections were overdispersed using NTI, which quantifies clustering at the tips of the phylogenetic tree, but where none were overdispersed using the NRI, which
measures tree-wide clustering. In fact, four of these six sections were clustered using the NRI. This is most likely the situation because gymnosperm diversity is low in these sections, resulting in tree-wide phylogenetic clustering, while their angiosperm diversity is moderate, with angiosperm species evenly spaced across the phylogenetic tree on many plots (Potter in review).

Such differences may have important forest health implications. A community consisting of species that are overdispersed across the phylogenetic tree but clustered toward the branch tips may contain fewer species susceptible to a given threat. That community, however, may be more likely to lose important ecological functions provided by any species that are eliminated. Meanwhile, a community of species clustered across the phylogenetic tree but overdispersed at the tips may better retain its ecological functionality in such a situation, but might encompass more species at risk of elimination.

It is worth noting that it is possible for a community to possess high phylogenetic diversity but to still have a clustered phylogenetic community structure, and, as a result, to be more susceptible to certain threats. This is because of important differences in the two metrics. Phylogenetic diversity measures the total evolutionary diversity present in a community, while NRI and NTI are indices that quantify the configuration of the evolutionary relatedness among the species. Ecoregions section with high mean plot-level phylogenetic diversity and a high degree of clustering include the Eastern Upper Peninsula (212R), the Blue Ridge Mountains (M221D), the Catskill Mountains (211I), the Maine-New Brunswick Foothills and Lowlands (211B), and the Klamath Mountains (M261A).

The patterns of phylogenetic community structure detected in this study may be driven at least in part by ecological interactions among species within the community. Phylogenetic clustering, for example, may be caused by environmental filtering, the process that occurs when closely related species tend to co-occur because they share similar tolerances to the abiotic environment (Cavender-Bares and others 2004, Tofts and Silvertown 2000). Because the species present in such communities share much evolutionary history and an affinity for similar environmental conditions, they may be particularly susceptible collectively to a certain threats, such as insects and diseases targeting specific families or genera of tree species. Phylogenetic overdispersion in a community, meanwhile, may indicate the existence of competitive exclusion. This occurs when closely related species compete for similar environmental niches and exclude each other from a community when they share limiting resources (Cavender-Bares and others 2004, Tofts and Silvertown 2000). The ecological integrity of such communities may be less at risk from changing conditions because they encompass a wider variety of evolutionary adaptations to respond to those changes.
Figure 7: Mean FIA plot-level measures of phylogenetic clustering across ecoregion sections, using (A) the Nearest Taxon Index, a measure of clustering at the tips of the phylogenetic branches, and (B) the Net Relatedness Index, a measure of clustering throughout the phylogenetic tree. A t-test was used to determine whether the mean index values were significantly different from 0, with negative index values overdispersed and positive values clustered.
Several correlations were detected between mean phylogenetic community structure values and environmental characteristics at the ecoregion section scale (table 1). These correlations suggest that competitive exclusion for resources among related species (phylogenetic overdispersion) is an important process in ecoregion sections with less “ hospitable” environments. Meanwhile, environmental filtering, which occurs when more closely related species tend to occupy the same kinds of environments (phylogenetic clustering), might be more common when environmental conditions are more hospitable, such as at sites with moister and less compacted soils, at lower elevations, and with greater amounts of precipitation.

For example, soil available water capacity was the environmental variable with the strongest positive correlation phylogenetic clustering across ecoregion sections, indicating communities with wetter soils were generally more phylogenetically clustered. Greater soil acidity (lower pH) also was positively associated with phylogenetic clustering, as was depth to bedrock. Among the climate variables, annual precipitation had a weak but significant correlation with NRI, indicating that forest tree communities in ecoregion sections with more precipitation were more phylogenetically clustered using the phylogenetic tree-wide metric. Standard deviation of monthly maximum and minimum temperatures was negatively correlated with both NRI and NTI, suggesting that variability in temperatures was associated with greater phylogenetic overdispersion of communities in the ecoregion. Finally, three different topographical measures were important in predicting phylogenetic structure, with ecoregion section elevation mean, elevation variation (standard deviation) and elevation range all correlated with greater phylogenetic overdispersion.

**Table 1:** Correlations between mean plot-level phylogenetic clustering statistics across ecoregion sections and environmental variables at the section scale.

<table>
<thead>
<tr>
<th>Environmental variable</th>
<th>NRI</th>
<th>NTI</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Soils</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Available water capacity (AWC)</td>
<td>0.475***</td>
<td>0.463***</td>
</tr>
<tr>
<td>Bulk density</td>
<td>-0.519***</td>
<td>-0.37***</td>
</tr>
<tr>
<td>pH</td>
<td>-0.355***</td>
<td>-0.249**</td>
</tr>
<tr>
<td>Depth to bedrock</td>
<td>0.34***</td>
<td>0.321***</td>
</tr>
<tr>
<td><strong>Climate</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean annual precipitation (1971-2000)</td>
<td>0.226**</td>
<td>ns</td>
</tr>
<tr>
<td>Monthly maximum temperature (1971-2000) (mean)</td>
<td>ns</td>
<td>ns</td>
</tr>
<tr>
<td>Monthly maximum temperature (1971-2000) (standard deviation)</td>
<td>-0.377***</td>
<td>-0.383***</td>
</tr>
<tr>
<td>Monthly minimum temperature (1971-2000) (mean)</td>
<td>ns</td>
<td>ns</td>
</tr>
<tr>
<td>Monthly minimum temperature (1971-2000) (standard deviation)</td>
<td>-0.345***</td>
<td>-0.347***</td>
</tr>
<tr>
<td><strong>Topography</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elevation (mean)</td>
<td>-0.413***</td>
<td>-0.41***</td>
</tr>
<tr>
<td>Elevation (standard deviation)</td>
<td>-0.274***</td>
<td>-0.281***</td>
</tr>
<tr>
<td>Elevation (range)</td>
<td>-0.315***</td>
<td>-0.339***</td>
</tr>
</tbody>
</table>

*** p<0.001; ** p<0.01; ns, not significant
Conclusions and future work

The work outlined in this paper demonstrates the applicability of phylogenetic community analysis to data sets at regional and even national scales. Generating these phylogenetic diversity and community structure statistics at large scales is one approach for investigating the role evolutionary biology plays in shaping processes and patterns in the natural world. The results of this study indicate that differences exist between phylogenetic diversity and species richness within ecoregion sections, and suggest that phylogenetic diversity may be a more meaningful measure of biodiversity because it accounts for evolutionary relationships among the species in a community. Further, the analyses of phylogenetic community structure detected intriguing regional patterns of phylogenetic clustering and overdispersion. Additional research is necessary to better understand the causes of these patterns, and to further explore the correlations, at regional scales, between phylogenetic clustering and several environmental variables.

The types of analyses described in this paper have potential for investigating the evolutionary aspects of forest health. For example, these analyses could assess whether regions with more phylogenetically clustered forest tree communities are more susceptible to major environmental changes, because of the possibility that a higher proportion of species will be affected by a given stressor. Another important area of research using these methods could quantify the evolutionary impacts of the loss of species, in terms of loss of ecological functionality and evolutionary potential.

Finally, this study establishes that Forest Inventory and Analysis data are suitable for addressing evolutionary biology questions at regional scales. In fact, few if any other available data sets have the extent and resolution necessary to conduct phylogenetic community analyses for forest communities at broad scales. FIA inventory data have considerable potential for use in many such future studies, including several with forest health applications. Future research, for example, could incorporate FIA data to:

1. assess whether nonnative invasive forest tree species are more phylogenetically related than expected by chance (Strauss and others 2006), which could allow for the identification of groups of closely related nonnative species that may be likely to become invasive.

2. determine the phylogenetic community effects of the loss of species or groups of species as a result of forest tree insects or pathogens, such as chestnut blight, hemlock woolly adelgid, and white pine blister rust.
(3) test for phylogenetic signal in forest tree susceptibility to insect and disease infestation, particularly from those with multiple hosts, and to broad environmental changes such as climate change.

(4) quantify the phylogenetic diversity and community structure of entire forest plant communities, and of strata within those communities, using FIA Phase 3 data.

**Acknowledgments**

I wish to thank Frank Koch, Mark Ambrose and Barb Conkling from the Forest Health Monitoring Group at North Carolina State University for their insights about this research and for their technical and administrative support. I further thank Kurt Riitters from the U.S. Forest Service and Cam Webb of the Arnold Arboretum at Harvard University for their technical support. This research was supported in part through Research Joint Venture Agreement 06-JV-11330146-124 between the USDA Forest Service and North Carolina State University.

**Literature Cited**


http://www.soilinfo.psu.edu/index.cgi?soil_data&conus


