*Syst. Biol.* 68(2):187–203, 2019

DOI:10.1093/sysbio/syy084
Advance Access publication December 6, 2018

# Restriction-Site-Associated DNA Sequencing Reveals a Cryptic *Viburnum* Species on the North American Coastal Plain

ELIZABETH L. SPRIGGS[1,*], DEREN A. R. EATON[1,2], PATRICK W. SWEENEY[3], CAROLINE SCHLUTIUS[1], ERIKA J. EDWARDS[1,2], AND MICHAEL J. DONOGHUE[1,3]

[1]*Department of Ecology and Evolutionary Biology, Yale University, PO Box 208106, New Haven, CT 06520-8106, USA;* [2]*Department of Ecology, Evolution, and Environmental Biology, Columbia University, New York, NY 10027, USA; and* [3]*Division of Botany, Peabody Museum of Natural History, Yale University, PO Box 208118, New Haven CT 06529, USA*

**Correspondence to be sent to: Department of Ecology and Evolutionary Biology, Yale University, PO Box 208106, New Haven, CT 06520-8106, USA; E-mail: espriggs@fas.harvard.edu*

*Received 10 October 2017; reviews returned 20 November 2018; accepted 3 December 2018*
*Associate Editor: Mark Fishbein*

*Abstract*.—Species are the starting point for most studies of ecology and evolution, but the proper circumscription of species can be extremely difficult in morphologically variable lineages, and there are still few convincing examples of molecularly informed species delimitation in plants. Here, we focus on the *Viburnum nudum* complex, a highly variable clade that is widely distributed in eastern North America. Taxonomic treatments have mostly divided this complex into northern (*V. nudum* var. *cassinoides*) and southern (*V. nudum* var. *nudum*) entities, but additional names have been proposed. We used multiple lines of evidence, including RADseq, morphological, and geographic data, to test how many independently evolving lineages exist within the *V. nudum* complex. Genetic clustering and phylogenetic methods revealed three distinct groups—one lineage that is highly divergent, and two others that are recently diverged and morphologically similar. A combination of evidence that includes reciprocal monophyly, lack of introgression, and discrete rather than continuous patterns of variation supports the recognition of all three lineages as separate species. These results identify a surprising case of cryptic diversity in which two broadly sympatric species have consistently been lumped in taxonomic treatments. The clarity of our findings is directly related to the dense sampling and high-quality genetic data in this study. We argue that there is a critical need for carefully sampled and integrative species delimitation studies to clarify species boundaries even in well-known plant lineages. Studies following the model that we have developed here are likely to identify many more cryptic lineages and will fundamentally improve our understanding of plant speciation and patterns of species richness. [North America; North American Coastal Plain; plants; RADseq; morphology; species delimitation; *Viburnum*.]

Advances in sequencing technologies (e.g., Baird et al. 2008; Peterson et al. 2012; Etter et al. 2011) and analytical methods (e.g., Yang and Rannala 2010) are revolutionizing the study of phylogeography and species delimitation. This is especially the case for plants, where traditional Sanger loci have typically been insufficiently variable to resolve recent evolutionary history (Schaal et al. 1998). Although these new methods have the potential to resolve long-standing taxonomic problems, their application is still under development and there are few carefully worked examples in plants that integrate morphology, geography, and ecology. Here, our primary aim is to provide just such an example, which we hope will serve as a point of reference for future studies of species and speciation.

Many of the most common woody plant species of eastern North American deciduous forests have exceptionally large geographic ranges (e.g., *Acer rubrum* L., *Quercus alba* L., *Tilia americana* L., *Prunus serotina* Ehrh., *Carya glabra* (Mill.) Sweet, *Cornus alternifolia* L.f., *Platanus occidentalis* Hook. & Arn., *Fraxinus americana* L., *Carpinus caroliniana* Walter, and *Betula nigra* L.; Little 1971; Kartez 2015). For most of these species, there is a history of subspecies and varieties being described based on morphological variation, but few of these segregates are widely recognized, and many have been synonymized in major floristic treatments of the 20th century (Sorrie and Weakley 2001). Phylogeographic studies of North American plants have uncovered significant genetic structure within most species (McLachlan et al. 2005; Morris et al. 2008, 2010; Tanaka et al. 2012; Barnard-Kubow et al. 2015; Laricchia et al. 2015; Nadeau et al. 2015; Thomson et al. 2015; Zinck and Rajora 2016), but in many cases it is unclear whether population structure within these species matches previous taxonomic hypotheses. The application of next-generation sequencing and molecular species delimitation methods to such problems will significantly improve our understanding of eastern North American deciduous forests, and will help resolve whether the flora is dominated by widespread and broadly panmictic species, or by species complexes comprised of multiple independent evolutionary lineages.

Here, we focus on the *Viburnum nudum* species complex (Adoxaceae) and test several alternative hypotheses based on the taxonomic literature. This complex is consistently supported as monophyletic and appears to be sister to the remainder of the *Lentago* clade, a group that contains five other North American species (Clement et al. 2011, 2014; Spriggs et al. 2015; Eaton et al. 2017). The *V. nudum* complex occurs across eastern North America, from Nova Scotia, west to Michigan and south into Texas and Florida, and it exhibits considerable phenotypic variation, particularly in leaf characters. In nearly all taxonomic treatments, two entities are recognized as either species, subspecies, or varieties—"cassinoides" in the North and "nudum" in the South. We consider three alternative hypotheses: 1)

variation occurs along a South to North cline, which is reflected in genetic structure by an isolation by distance pattern. In this case, "cassinoides" and "nudum" would be extreme forms that occur at opposite ends of a more or less continuous spectrum. This thinking is reflected in many recent treatments that recognize two varieties or subspecies (Clark 1971; Jones 1983; Haines 2011). 2) There are two distinct species—*V. cassinoides* in the North and *V. nudum* in the South (Small 1933; Gleason and Cronquist 1991; Ferguson 1966; Radford et al. 1968; Strausbaugh and Core 1978; Weakley 2012). This scenario would be distinguished from the previous one by clear boundaries between the populations and evidence that each lineage is evolving independently (de Queiroz 2007). 3) Finally, we evaluate whether there are more than two species in the clade as suggested in some of the oldest taxonomic treatments of the group that recognized now widely disregarded segregates like "*nitidum*" (Aiton 1789), "*angustifolium*" (Torrey and Gray 1843), or "*harbinsonii*" (McAtee 1956).

To understand the evolutionary history and species limits of the *V. nudum* complex, we assessed both morphological and genetic variation across the geographic range of the complex. Individuals were genotyped using restriction-site-associated DNA (RAD) sequencing, and over 60,000 loci were recovered with more than 250,000 parsimony informative Single nucleotide polymorphisms (SNPs). Genetic clustering methods and phylogenetic inference consistently revealed three distinct lineages, two of which are sympatric across large areas of the southeastern United States. We argue that all three groups are independently evolving lineages and should be recognized as species. Our findings provide insight into the postglacial assembly of eastern North American forests, and are consistent with the argument that the North American Coastal Plain is an under-appreciated biodiversity hotspot of global significance (Noss et al. 2015).

## Materials and Methods

### *Sample Collection and RAD Sequencing*

In 2013–2016, 139 individuals of the *V. nudum* complex were collected from 45 populations spanning the eastern United States, with an extra collecting effort focused in Georgia (Fig. 1). A voucher specimen for each individual is deposited in the Yale University Herbarium (YU) and leaf material was dried in silica gel. One to two individuals from each population (66 total) were selected for RAD sequencing. Four recently collected herbarium specimens were also sampled to add material from under-represented regions, and eight closely related species were included as outgroups. Genomic DNA was extracted for these 78 individuals using either a Qiagen DNEasy kit (Valencia, CA) or following the Beck et al. (2012) modification of the CTAB protocol (Doyle and Doyle 1987).

Each sample for this study was included on one of three separate but identical RAD library preparations.

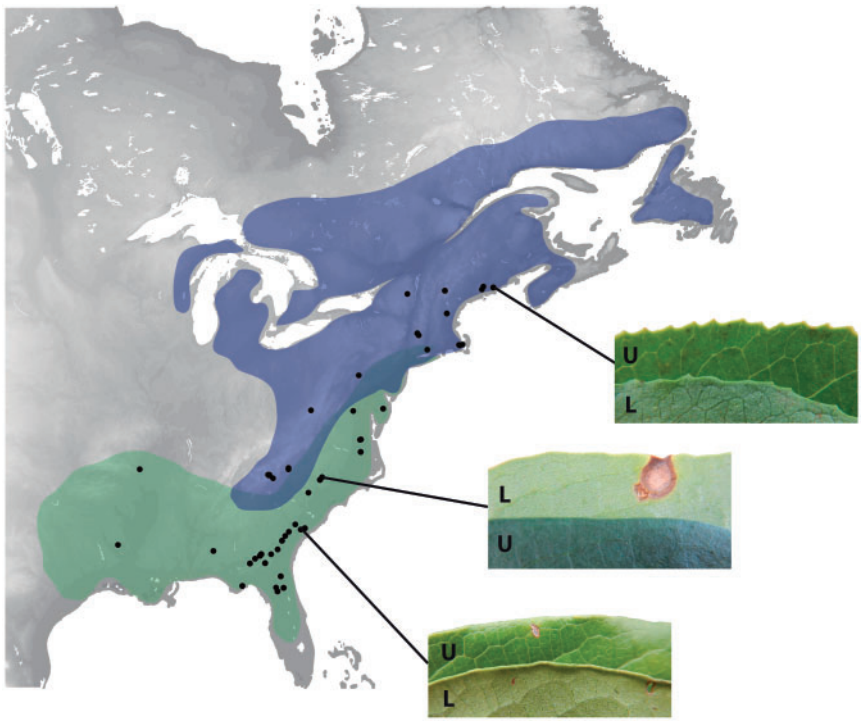


Figure 1. Distribution of the *Viburnum nudum* species complex re-drawn from Jones (1983). Northern "*V. nudum* var. *cassinoides*" range is in blue and southern "*V. nudum* var. *nudum*" range is in green. Points indicate collecting localities for the current study. Photographs show leaf margins: L = lower (abaxial) surface; U = upper (adaxial) surface. "*V. nudum* var. *nudum*" leaves tend to be entire (smooth), whereas "*V. nudum* var. *cassinoides*" leaves tend to be minutely toothed.

All RAD libraries were prepared by Floragenex Inc. (http://floragenex.com) using digestion with the PstI restriction enzyme followed by sonication and size selection for a mean fragment length of 400 bp. Each library was sequenced at the University of Oregon GC3F facility (http://gc3f.uoregon.edu) on an Illumina HiSeq 2000, Hiseq 2500, or Hiseq 4000. Although the total number of reads per sample varied, the identity of the loci recovered across plates was highly repeatable (i.e., there was no evidence of a plate bias) as 93.6% of the loci contained at least one individual from all three plates.

### *Data Assembly*

The software ipyrad (http://github.com/dereneaton/ipyrad) was used to demultiplex sequence data and generate *de novo* assemblies. Reads with phred scores $<20$ for more than five bases were excluded, and a minimum depth of six reads was required for statistical base calling. The clustering threshold of sequence similarity for building loci within individuals was set to 0.88. As individuals in the *V. nudum* complex are diploid ($2N=18$) (Egolf 1962), we removed putative loci that had more than two alleles on the basis that they were likely derived from paralogous regions of the genome. We focused particular attention on the potential impact of missing data. To test these effects, we constructed four RAD data sets: 1) a data set that included all loci shared across at least 10 individuals with 86,174 loci and 61% missing data (min10); 2) a data set with all loci shared across at least 20 individuals with 61,890 loci and 53% missing data (min20); 3) a data set with all loci shared across at least 40 individuals with 26,804 loci and 41% missing data (min40); and 4) a data set with all loci shared across at least 50 individuals with 4990 loci and 32% missing data (min50) (Supplementary Table S1 available on Dryad at http://dx.doi.org/10.5061/dryad.m1vc0). We also assessed the number of loci shared across individuals in each of our data sets to check for systematic biases (Supplementary Fig. S1 available on Dryad). All four data sets were used in the phylogenetic analyses and the min40 was selected as the primary data set for all other analyses.

### *Genetic Clustering*

To estimate the number of genetic clusters within the *V. nudum* complex without a priori assumptions about the number of clusters or individual assignments, we used the Bayesian clustering algorithm STRUCTURE (Pritchard et al. 2000) and discriminant analysis of principal components (DAPC) in the R package ADEGENT (Jombart et al. 2010). For both analyses, we used the min40 data set with outgroups removed and considered a range of clustering scenarios with up to four groups ($K=1$–4). For STRUCTURE, we ran 10 replicates at each value of $K$ for 500,000 generations with a burn-in of 50,000. Replicate runs were grouped using CLUMPP (Jakobsson and Rosenberg 2007) and convergence was assessed by comparing the alpha parameter, P(X|K), and the variance of P(X|K). We used Structure Harvester to compare alternative values of K based on the log probability of the data (log P(X|K)) and the $\Delta$K statistic (Evanno et al. 2005).

Previous studies have demonstrated that when population structure deviates from the standard island model (e.g., when there is hierarchical tree-like structure or when there are strong isolation-by-distance patterns), STRUCTURE frequently identifies only the highest-order pattern (Jombart et al. 2010; Kalinowski 2011). In these cases, running structure within the identified groups can reveal finer patterns of population structure. We, therefore, ran a series of nested analyses, such that when $K=2$ identified a clean split into two nonadmixed groups we ran STRUCTURE separately within each. For each subsampled data set, we retained all loci that were shared across at least 50% of the individuals.

DAPC generates discriminant functions that maximize the differentiation between groups while minimizing the variation within them. For this analysis, we used our min40 SNP data set that included a single SNP sampled randomly from each locus. We selected the number of clusters by running k-means clustering for successive numbers of clusters and compared different numbers of clusters based on the Bayesian information criterion (BIC). We then performed DAPC using only the first two principal components and calculated the membership probabilities for all individuals.

### *Species Delimitation and Tree Inference*

We used BPP v.3.3 (Yang and Rannala 2010) to test species delimitations using a fixed species tree and three putative species identified as clusters in our STRUCTURE analysis. For each run, we randomly sampled 500 loci that were shared across at least five individuals from each of the putative species. We assigned equal prior probabilities to all species delimitation models (1, 2, or 3 species) and conducted two replicate runs under each of a series of different parameters. We used a range of fine-tuning parameters for BPP's reversible-jump MCMC algorithm following the recommendations of Yang (2015): algorithm 0 with $\varepsilon=2$, 5, 10; algorithm 1 with $\alpha=1$, 1.5, 2, and $m=1$, 1.5, 2. With each algorithm, we used a gamma prior for theta (ancestral population size) and tau (root age) that correspond to two alternative demographic scenarios: relatively deep divergences and large population sizes (theta G[1,10] tau G[1,10]) or relatively shallow divergences and small ancestral population sizes (theta G[2,2000] tau G[2,2000]). All runs were sampled every 50 generations for 10,000 samples with a burn-in of 2,000. We evaluated convergence by comparing replicate runs and checked to make sure that the starting tree varied across runs, from a one-parameter model with a single species to a seven-parameter model with three species. We then ran two replicates of a simplified analysis with

fixed species delimitation and a fixed species tree to estimate tau and theta parameters for all supported species. We also conducted a BPP analysis to jointly infer the species tree and delimit species. The selection of loci followed the same procedure as the previous analysis and the same range of priors was considered.

To reconstruct relationships among individuals we took several approaches. Trees were inferred from concatenated data sets using both RAxML version 8.2.10 (Stamatakis 2014) and IQ-TREE version 1.5.5 (Nguyen et al. 2015). We also inferred a tree for each data set using the coalescent-based method tetrad (ipyrad v 0.6.18; Eaton et al. 2017) on an alignment of unlinked SNPs. We included the tetrad analysis because another RADseq study that used the similar quartet-based method, SVDquartets, found SVD-quartets to be more robust to data assembly parameters (e.g., clustering threshold) as compared to concatenation methods (Fernández-Mazuecos et al. 2018). For RAxML, maximum likelihood phylogenies were inferred with 1000 rapid bootstraps (100 for the larger min10 and min20 data sets) using the GTRGAMMA substitution model. Each tree was rooted with *V. rhytidophyllum* + *V. carlesii*, members of the *Euviburnum* clade (which is closely related to the *Lentago* clade within the more inclusive *Valvatotinus*; Clement et al. 2014). In IQ-TREE, the best nucleotide substitution model was selected based on the BIC using the ModelFinder (Kalyaanamoorthy et al. 2017) option, and a maximum likelihood tree was then inferred under the best model. Clade support was calculated with 1000 ultrafast bootstrap replicates (Minh et al. 2013; Hoang et al. 2018) and with the Sh-like approximate likelihood ratio test with 1000 replicates (Guindon et al. 2010). For tetrad, we used SNP alignments sampled such that in each replicate run, a single SNP was randomly sampled from each locus for every quartet comparison. This approach maximizes the number of SNPs in each test while minimizing the linkage among SNPs. Tetrad uses an algorithmic approach similar to SVDQuartets to infer quartet topologies based on phylogenetic invariants (Chifman and Kubatko 2014) and then joins the resulting quartet trees into a supertree using the maxcut algorithm of QMC (Snir and Rao 2012). All possible quartets were sampled and confidence was assessed using 1000 nonparametric bootstrap replicates (100 replicates for the min10 and min20 data sets). Finally, we used SplitsTree4 (Hudson and Bryant 2006) to generate a phylogenetic network for our min40 unlinked SNP data set using the neighbor-net algorithm (Bryant and Moulton 2004).

Of the individuals sequenced, there are 14 pairs of individuals that were sampled within 50 m of one another. To further compare the phylogenies and assess how reliable our data sets are with different amounts of missing data, we counted the number of these pairs that appeared as monophyletic within each tree.

### *Isolation by Distance*

We used custom python scripts (available on Dryad) to calculate the uncorrected pairwise genetic distance between all taxa in the *V. nudum* complex. We took this approach rather than using more common measures of genetic distance to avoid grouping individuals into discrete populations and to eliminate assumptions about the processes generating divergence among individuals. Using the min40 data set, we considered all loci shared between each pair of individuals and calculated the percent sequence divergence with missing data and heterozygous sites removed. Heterozygous sites were removed because we were primarily interested in fixed divergences between populations, and to avoid any bias that could be created by differing levels of genetic diversity among populations. Geographic distance among individuals was calculated using the distm function in the R package geosphere (Hijmans 2016). To test whether the genetic divergence between *V. nitidum* and *V. cassinoides* was significant beyond what would be expected from geographic isolation alone, we compared linear models with geographic distance and species identity as fixed effects. Mantel tests were used to evaluate the correlation between genetic and geographic distance within each species.

### *Tests of Introgression*

We used the D-statistic to test for gene flow between the red and blue clades in the piedmont of North Carolina where members of the two lineages are present. The D-statistic compares the relative frequency of two discordant site patterns (ABBA and BABA) in a four-taxon tree. For each set of tests, we compared one of the five North Carolina piedmont individuals (ELS322, ELS054, ELS567, ELS340, and ELS564) to one of the remaining red clade individuals and one of the remaining blue clade individuals. These individuals were randomly sampled for each replicate and each test was performed 100 times. For the outgroup, three samples from the green clade (ELS586, ELS66, and ELS645) were pooled and SNP frequencies were measured from this pooled group as in Eaton et al. (2015). All tests were performed in ipyrad v 7.21 (http://github.com/dereneaton/ipyrad), and the standard deviation of D was measured with 1000 bootstrap replicates sampled as in Eaton and Ree (2013).

### *Morphological Analysis*

Leaves were collected from all individuals and scanned in the field with a Cannon LiDE110 Color Image Scanner. Leaves with minor insect damage were repaired in Photoshop (http://www.adobe.com) by filling in eaten areas. As in other *Viburnum* species, individuals of the *V. nudum* complex display seasonal heteroblasty, such that leaf form varies consistently along branches (Spriggs et al. 2018). To reduce potential noise introduced by this intra-individual variation, each scanned leaf was classified as "early," "intermediate," or "late" following the scheme of Spriggs et al. (2018). Only early leaves, the most abundant and consistent leaf class, were used. In all, 437 leaves from 118 individuals were included. We used

built-in ImageJ (http://rsbweb.nih.gov/ij) functions to measure leaf area, leaf perimeter, petiole length, and leaf shape. ImageJ shape parameters are based on a fitted ellipse, and for each leaf we included the length of the major and minor axes of this ellipse as well as the aspect ratio (length of the major axis/length of the minor axis). In addition to leaf traits, we measured the length of the inflorescence stalk (peduncle length) and the length of the first order inflorescence branches (cyme ray length) on all collected specimens of reproductive individuals. The number of measurements per individual varied according to the number of inflorescences present, but as many as five measurements of each trait were included for each individual.

We calculated the mean value of all traits for each individual and then log transformed all measurements. We conducted a principal components analysis on the trait data and retained all principal components for clustering analyses. Similar to the approach outlined in Cadena et al. (2018), we used the R package mclust (Fraley et al. 2012) to fit normal mixture models without *a priori* information about the number of species. We conducted a principal component analysis on log-transformed trait measurements and included all PC axes in the clustering analysis. We considered models with up to five different clusters and evaluated them based on the BIC.

### *Range Maps and Distribution Modeling*

Based on the findings of the previous sections, we constructed range maps for each putative species in the *V. nudum* complex. We compiled 969 herbarium specimens from 12 North American herbaria, using either specimen images available in online databases (http://sernecportal.org/portal/) or photos taken in person during herbaria visits (Supplementary Table S2 available on Dryad). We also included our own collections in these maps. Specimens were georeferenced to county and classified as belonging to *V. nudum*, *V. cassinoides*, or *V. nitidum* based on a visual assessment of specimen morphology. We discarded any specimens that were ambiguous in either locality or in species assignment. We chose to georeference the specimens to county because that information is consistently available for most of the specimens. Furthermore, counties in the eastern United States are relatively small so we expect climatic variation within them to be relatively minor.

To estimate the current and historical ranges of each species, we created species distribution models using the maximum-entropy method MAXENT (Phillips et al. 2006) as implemented in the R package dismo (Hijmans et al. 2017). Current and past climatic data (19 bioclimatic variables) were downloaded from Worldclim (http://www.worldclim.org) at a 2.5 arc-min resolution. We generated 1000 random points within our study region (15° −55° Latitude, −110° to −50° Longitude) and used these points to assess the degree of correlation among the 19 bioclimatic variables. We took two approaches to reducing the number of variables: first, we removed one variable in each pair with a Pearson correlation coefficient $\geq$0.4; second, we removed one variable in each pair with a Pearson correlation coefficient $\geq$0.8. With the first, more stringent, approach we retained three variables: mean annual temperature, annual precipitation, and precipitation seasonality. With the second we retained six variables: the previous three plus mean temperature of the wettest quarter, mean diurnal temperature, and precipitation of the wettest quarter.

Species locality data were derived from our field collections and range maps (see above). Because the current taxonomy of the *V. nudum* complex does not match our species delimitations (see below), we were unable to use taxonomic assignments from publically available occurrence data in most cases (https://www.gbif.org); however, as only the northernmost clade (blue in Figs. 2 and 3) occurs in Canada, we included Canadian GBIF records of "*V. nudum*" with that clade in our analyses. For all specimens used to construct the range maps, we identified the county and calculated the mean of all environmental variables across that county. We used 500 random background points from the study region to fit the models. In each case, species distribution models were fit with 80% of the data used for training and the remaining 20% for testing. We conducted ten replicates for each model and evaluated model fit based on the area under the ROC curve (AUC).

## Results

### *Genetic Clustering*

In our STRUCTURE analysis that considered the entire *V. nudum* complex, $K=2$ was preferred with the $\Delta K$ method and had the highest log P(X|K). With $K=2$, the taxa are split into two distinct clusters with no evidence of significant admixture (Fig. 2). When the larger of these two clusters (the red + blue in Fig. 2) was analyzed alone, again $K=2$ had the highest log P(X|K) and the highest $\Delta K$. Within this group also, $K=2$ split the individuals into two distinct groups (red and blue in Fig. 2). Subsequent clustering within these groups revealed geographic structure in both the red and blue groups (Supplementary Fig. S2 available on Dryad). When the green cluster was considered alone, the preferred model had two groups, one that is dominant in the western portion of the range and a second in the northeast portion. Admixed populations were found throughout Georgia (Supplementary Fig. S2 available on Dryad). All individuals were assigned to one of three discrete clusters (blue, red, and green) with high confidence, and although substructure was identified within all three of these (red, blue, and green), there is no evidence of admixture among the major groups (Fig. 2). Surprisingly, the red and green clusters have overlapping geographic ranges, and individuals collected in close

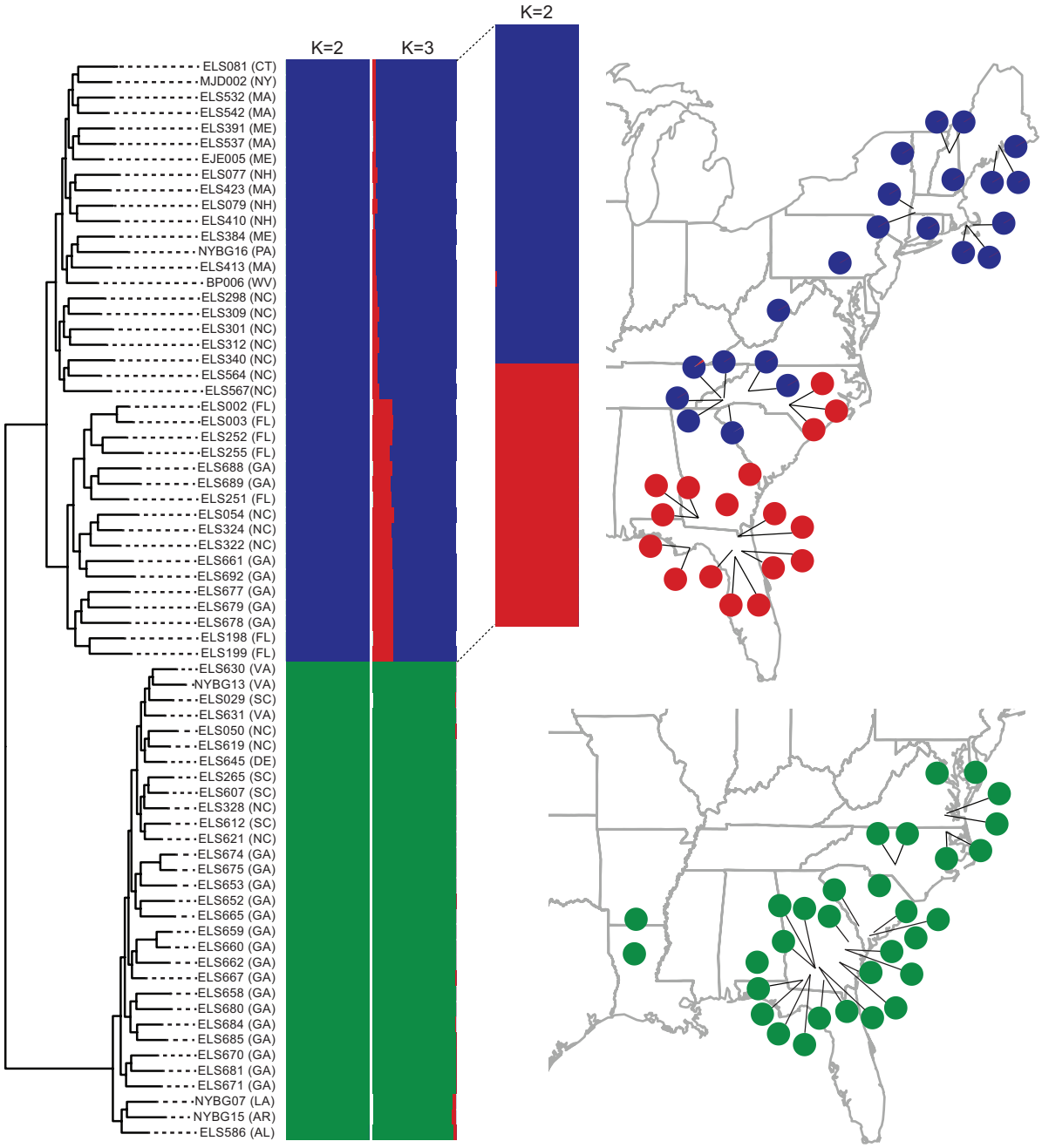


FIGURE 2. Hierarchical structure analyses identified three main clusters that correspond to the phylogeny (Fig. 4). Colors indicate posterior probability of assignment of each individual to a particular cluster based on the combination of 10 replicate runs. The number of clusters (K) for each analysis is displayed at the top of the bar graph. The location of each individual is displayed on the map.

proximity to one another in Georgia or North Carolina were assigned to different groups (Fig. 2).

Using the k-means clustering algorithm in ADEGENT, two clusters had the lowest BIC score (494.5897), but it was not significantly different from that of three clusters (494.8430). Under the three-cluster scenario, we retained two principal components, and all individuals were assigned to the same cluster as in STRUCTURE with a posterior probability of over 0.98. The only exception to this was the sample NYBG16, an individual from Pennsylvania; this grouped with the blue cluster in STRUCTURE but was assigned to the red cluster (posterior probability of 0.967) here (Fig. 3). Despite this assignment, NYBG16 is clearly distinguishable from the red cluster along the second discriminant axis, and it seems likely here that it was misclassified because of a high level of missing data. All individuals with less than 1 million reads passing the initial filtering steps

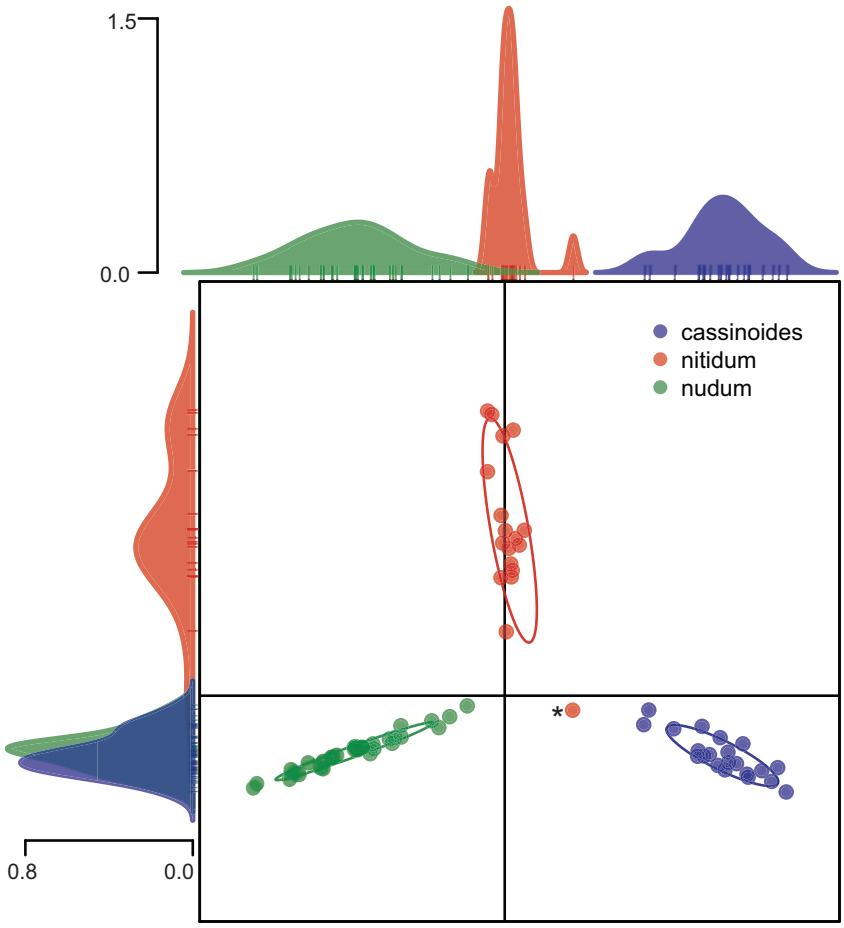


FIGURE 3. DAPC identified three genetic clusters that correspond to the groupings in Fig. 2. Inertia ellipses are drawn for each group and NYBG16 is marked with an asterisk.

(600,000 for NYBG16) had discriminant axis scores with low absolute values (are placed near the origin). This clustering approach seems more sensitive to missing data than STRUCTURE and may be less reliable for classifying individuals with few loci.

### *Species Delimitation and Tree Inference*

BPP strongly supported a three species model regardless of the prior settings used (posterior probabilities >.95). Estimates of divergence times and ancestral population size were also relatively consistent across different runs. Most notably, the ancestral population size for the blue lineage was estimated to be about twice as large as the other two. The species tree inferred with BPP was also strongly supported (posterior probability >0.99) and matched the topology recovered with other phylogenetic methods (below).

Phylogenetic results were highly consistent across different data sets and inference methods. Here, we consider the results of 12 different analyses: RAxML, IQ-TREE, and tetrad inferences based on the min10, min20, min40, and min50 data sets. In all cases, three major clades that correspond to the red, blue, and green clusters identified by STRUCTURE were recovered with high support (bootstrap support ≥99; SH-aLRT = 100; Fig. 4 and Supplementary Fig. S3 available on Dryad). For the most part, relationships within each of these major clades were not supported and the position of several accessions varied significantly within the major clades. For instance, one individual, ELS567, formed a clade with ELS340 and ELS564 in all IQ-TREE and RAxML analyses except for the RAxML analysis of the min40 data set in which this individual was found to be sister to the rest of the blue clade (Supplementary Fig. S3 available on Dryad). With tetrad, ELS567, ELS340, and ELS564, all from the North Carolina piedmont, are intermixed with individuals collected in the Smoky Mountains (ELS298, ELS301, ELS309, and ELS312). The same general topology and all three major clades were also recovered in a phylogenetic network inferred with SplitsTree (Supplementary Fig. S4 available on Dryad). To further compare trees, we counted the number of times that individuals that were collected within

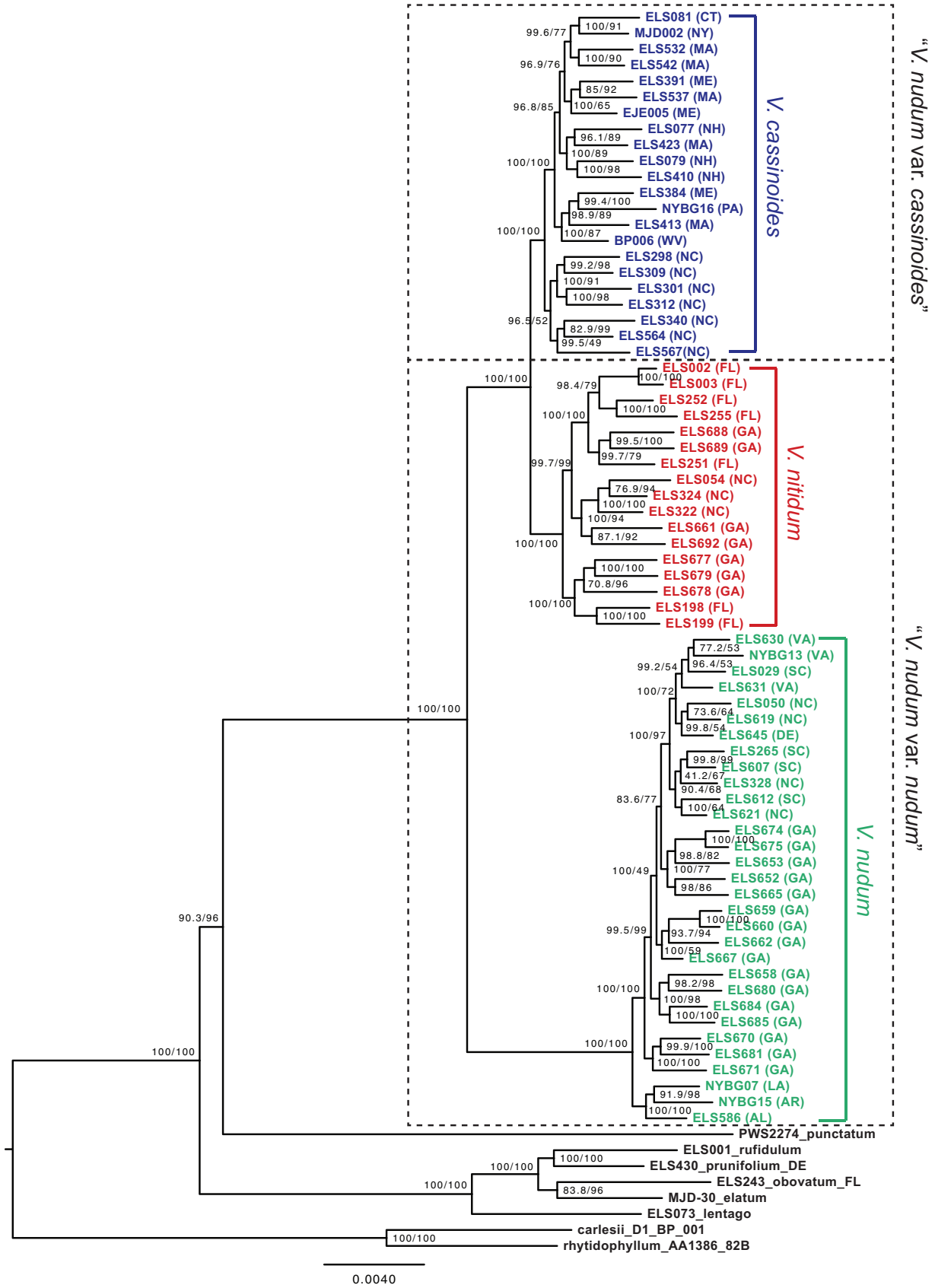


FIGURE 4. Maximum likelihood (IQ-TREE) phylogeny based on the min40 data set. Each node is labeled with the SH-aLRT support/ultrafast bootstrap support reported as percentages. Dotted lines outline the traditional (geographic) delimitations of "*Viburnum nudum* var. *cassinoides*" and "*V. nudum* var. *nudum*".

50 m of one another formed a clade. There are 14 sets of such individuals, and in all trees 10–12 were monophyletic. The existence of three major clades is robust to various levels of missing data and different methods of inference. As noted above, we found that the southern populations that represent the traditional "var. *nudum*" are paraphyletic. There are two clades in the southeastern US, one of which (red in Figs. 2–7) is more closely related to a clade consisting of the northern "cassinoides" populations (blue in Figs. 2–7).

### *Introgression Tests and Isolation by Distance*

For the most part, our D-statistic tests showed no evidence of introgression between the red and blue clades (Supplementary Fig. S5 available on Dryad). A small number of tests were marginally significant with ELS567, and it is possible that this individual is slightly admixed.

In all three putative species, intraspecific genetic distances are lower than interspecific genetic distances (mean of 0.4% and 1.1%, respectively). The greatest genetic distances are between the green lineage and the other two lineages (Fig. 5), with sequences for individuals of the red and the blue clades differing at only 0.4–0.7% of sites. Despite the relative similarity of the red and the blue clades, it is clear that they do not form a continuous cline. There is a "jump" in divergence between the two species and, when linear models are compared, a factor describing whether each comparison is inter- or intraspecific (i.e., species identity) explains more of the variance than geographic distance ($R^2 = 0.617$ and $R^2 = 0.182$, respectively). A model that includes species identity alone is not significantly improved by adding geographic distance. Considering each species separately, we detected a significant pattern of isolation by distance within the blue and green clades ($P = 0.001$ for blue, $P = 0.039$ for green, $P = 0.256$ for red, Supplementary Fig. S6 available on Dryad).

### *Morphological Clustering*

Support values for models with one or two morphological clusters were not significantly different from one another (one group BIC = 915.997 and two groups BIC = 915.288), but both were significantly better than models with more clusters ($\Delta$BIC > 19). Under the two-cluster scenario, one group was primarily composed of the red clade and the blue clade individuals, whereas another contained mostly individuals from the green clade, but 25.4% of individuals were misclassified. Phenograms based on our morphological measurements indicate that the green clade differs from the other two in leaf area, leaf perimeter, the length of the major leaf axis, the inflorescence stalk (peduncle) length, and the length of inflorescence branches (cyme rays) (Fig. 6). The red and blue clades overlap in most of the measured traits, but they are differentiated somewhat by aspect ratio, peduncle length, and leaf margin (Fig. 6).

### *Range Maps and Species Distribution Modeling*

Georeferenced herbarium specimens indicate that the red and the green clades are sympatric across a large portion of the southeastern US (Supplementary Fig. S7 available on Dryad). We identified both species in 54 counties of Alabama, Georgia, South Carolina, and North Carolina. The range of the blue clade may overlap with the green clade in western North Carolina and Virginia, but it seems that areas of sympatry for these species are limited.

Our species distribution models performed well regardless of whether three or six environmental variables were used as predictors (AUC $\geq$0.89). With these models, we found that the northern blue clade may have been relatively widespread at the last glacial maximum (LGM 18,000–23,000 BP), with available area stretching from Texas to Connecticut along the coastal plain (Fig. 7 and Supplementary Fig. S5 available on Dryad). Suitable environments for the other two species appear to have been much more limited. For both, we identify only peninsular Florida, islands in the Caribbean, and parts of Mexico and Central America as potentially suitable.

## Discussion

With extensive geographic sampling and analyses of both RADseq data and morphology, we have identified three distinct lineages in the *V. nudum* complex. We conclusively reject the hypothesis that the *V. nudum* complex is a single species and instead find evidence for three independently evolving species (sensu the unified species concept of de Queiroz 2007). We infer significant range movements in all three in response to LGM climate changes, but we argue that ecological differentiation before the LGM, rather than isolation in LGM refugia, drove the initial diversification of this clade.

Based on a combination of evidence, we propose the recognition of all three lineages in the *V. nudum* complex as species; that is, we argue that all three are separately evolving metapopulation lineages (de Queiroz 2007). The argument for recognizing the green lineage is clear cut. In addition to being morphologically distinctive and genetically divergent, it is broadly sympatric with the red clade without evidence of gene flow. The case for recognizing the red and the blue clades as separate species is more difficult. These lineages are allopatric, recently diverged, and morphologically similar to one another, a pattern at the heart of many debates around species concepts and delimitation methods (Sukumaran and Knowles 2017; Leaché et al. 2019). However, there are six lines of evidence supporting the recognition of these two lineages as separate species: 1) they are consistently identified as discrete groups with genetic clustering methods (Figs. 2 and 3). 2) They are reciprocally monophyletic in trees inferred with both concatenated and coalescent-based phylogenetic methods (Fig. 4 and Supplementary Fig. S3 available on Dryad). 3) There is no evidence of introgression at

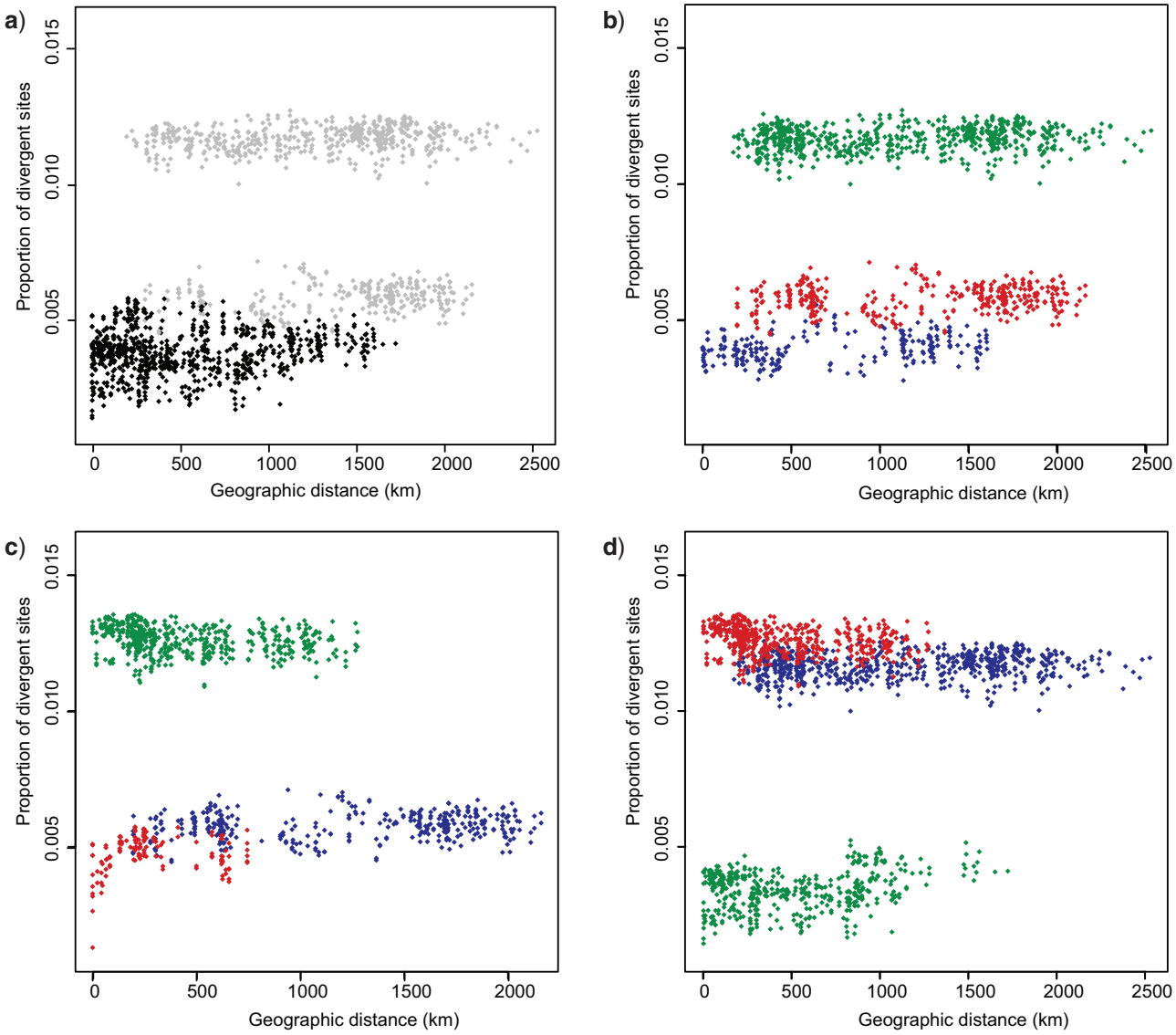


FIGURE 5. Pairwise geographic and genetic distances between all individuals. a) Comparisons across all individuals. Gray dots indicate among group comparisons (interspecific) and black dots indicate within group comparisons (intraspecific). b–d) Pairwise geographic and genetic distances separated by group (species). b) Distances between *Viburnum cassinoides* individuals and all others. *Viburnum cassinoides* to *V. cassinoides* distances are in blue, *V. cassinoides* to *V. nitidum* distances are in red, and *V. cassinoides* to *V. nudum* are in green. c) Distances between *V. nitidum* individuals and all others. *Viburnum nitidum* to *V. cassinoides* distances are in blue, *V. nitidum* to *V. nitidum* are in red, and *V. nitidum* to *V. nudum* are in green. (d) Distances between *V. nudum* individuals and all others. *Viburnum nudum* to *V. cassinoides* are blue, *V. nudum* to *V. nitidum* are red, and *V. nudum* to *V. nudum* are green. Note that each panel in (b–d) contains a different subset of the data in (a). The *V. nudum* to *V. nudum* comparisons in panel (d), for instance, are not included in either (b) or (c). Similarly, the *V. nudum* to *V. nitidum* comparisons in both (c) and (d) are not present in (b).

the boundary between them (Supplementary Fig. S5 available on Dryad). 4) Genetic distances indicate a discrete boundary rather than clinal variation; that is, clade identity is more informative for predicting genetic distances than geographic proximity (Fig. 5). 5) Although there is morphological overlap between the two groups, several traits (aspect ratio, peduncle length, leaf margin) do generally separate them (Fig. 6), and individuals near the range boundary are not intermediate in phenotype. 6) Our estimates suggest that the red and blue clades diverged prior to the LGM, but maintained their isolation throughout significant range movements that likely provided opportunities for secondary contact. Together these data indicate that the red and blue lineages are on independent evolutionary trajectories, and we expect their independence to persist into the future, leading to further divergence between them.

The genetic structure that we have discovered in the *V. nudum* complex is at odds with nearly all taxonomic

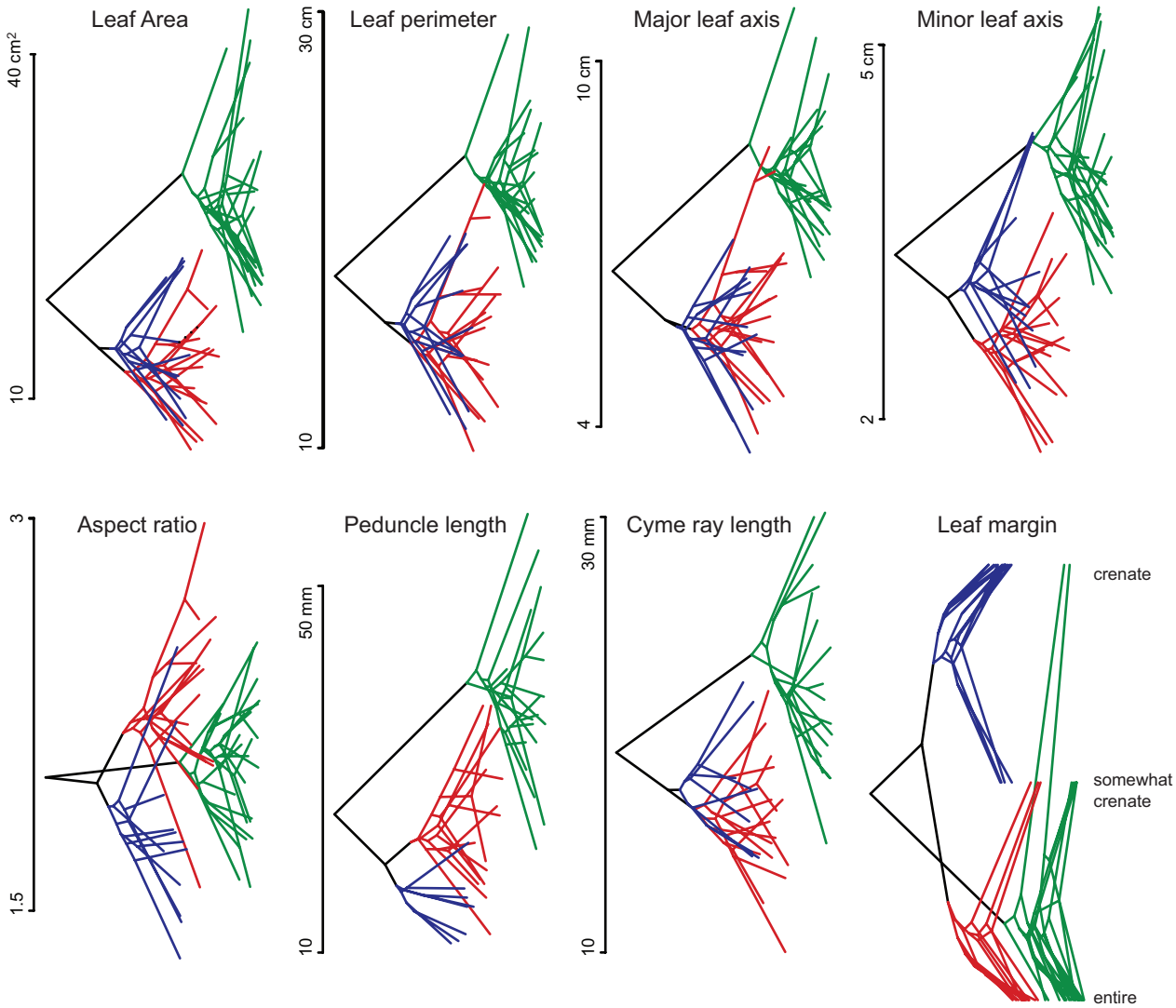


FIGURE 6. Phenograms for each trait constructed in Phytools (Revell 2012). Branch lengths are identical to those of in Fig. 2, and colors match previous figures with *Viburnum cassinoides* in blue, *V. nitidum* in red, and *V. nudum* in green. The vertical placement of each tip corresponds to the mean trait value for that individual. All collected individuals were measured for leaf area, leaf perimeter, major leaf axis, minor leaf axis, aspect ratio (major axis/minor axis), petiole (joining leaf to stem) length, peduncle (inflorescence stalk) length, cyme ray (inflorescence branch) length, and leaf margin (entire, somewhat crenate or regularly crenate).

treatments of the clade. Although most authors have recognized two entities in the complex based on geography ("cassinoides" in the north, "nudum" in the south) and leaf margin (crenate in "cassinoides, entire in "nudum") (Small 1933; Ferguson 1966; Radford et al. 1968; Clark 1971; Strausbaugh and Core 1978; Jones 1983; Gleason and Cronquist 1991; Haines 2011; Weakley 2012), nearly all treatments have placed the red and green clades together in *V. nudum*. Here, we have shown that *V. nudum* is paraphyletic under this circumscription, as the southern red clade is more closely related to the northern blue clade.

There have been two notable exceptions to the standard taxonomic treatment of the complex. First, Torrey and Gray's (1843) Flora of North America recognized three varieties within *V. nudum*, and these appear to correspond to our three lineages. Second, McAtee (1956) recognized a wide-ranging *V. cassinoides* that would appear to include both the red and blue clades in our analysis. It is only in these two treatments that there is a recognition of what turns out to be the deepest genetic and morphological split within the complex (namely the green clade vs. the red + blue clades). Based on examination of the type specimens, we are confident that our green clade (Figs. 2 and 3) corresponds to *V. nudum* L. and that our blue clade corresponds to *V. cassinoides* L. The southern red clade corresponds most closely to what Aiton described in 1789 as *V. nitidum*. Aiton's description was based on a cultivated plant and from intensive searches of the herbaria that housed Aiton's specimens (BM, K, G, and HUH), it seems highly likely that type material does

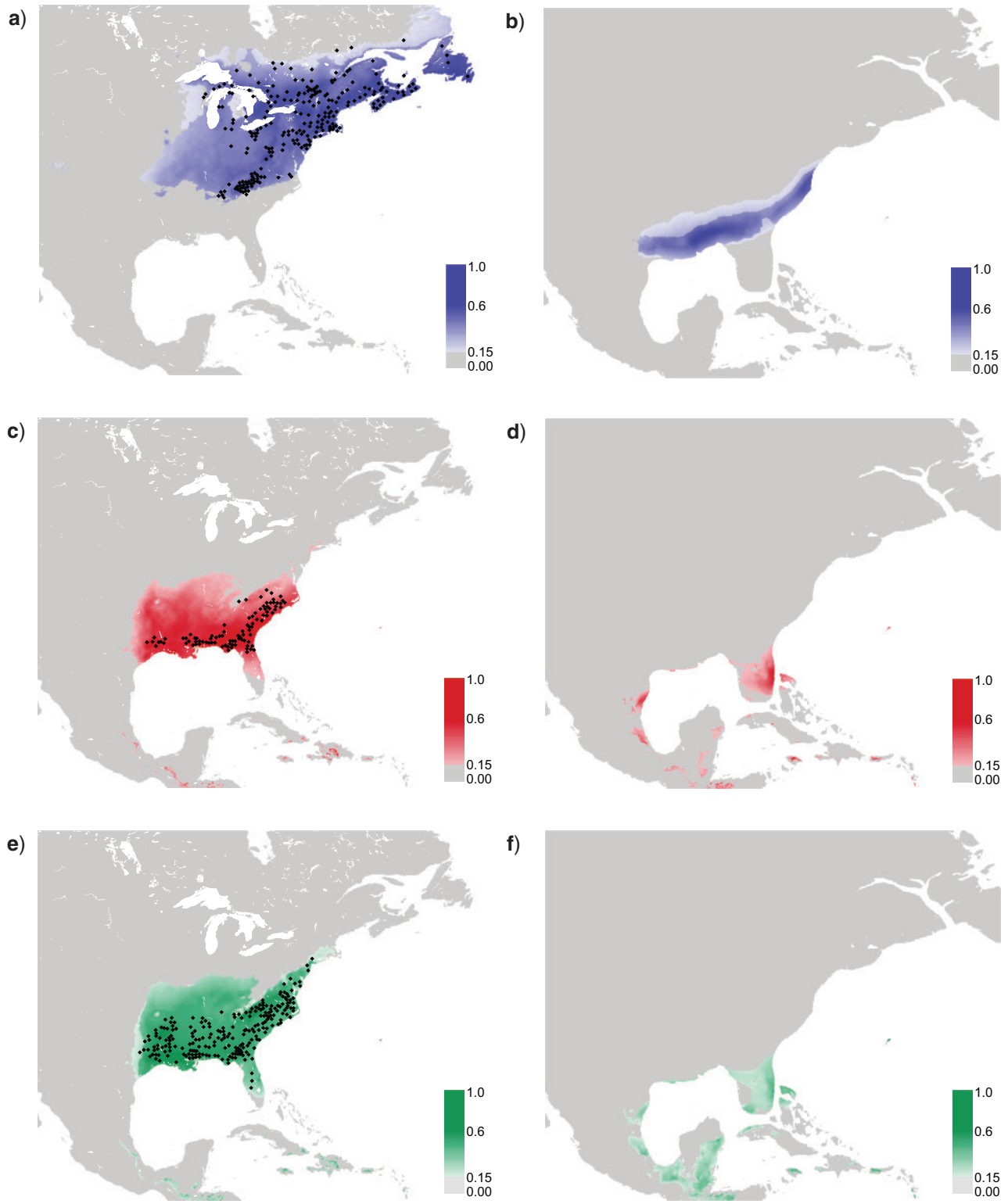


FIGURE 7. Species distribution models using Maxent with three predictor variables downloaded from Worldclim (http://www.worldclim.org) at a 2.5 arc-min resolution (mean annual temperature, annual precipitation, and precipitation seasonality) to infer the current (a, c, e) and Last Glacial Maximum (b, d, f) distributions of each species. Occurrence data for each species are plotted as points on the current map. Relative probability of occurrence at each grid cell is indicated by shading.

not exist for this name, and we will need to designate a neotype elsewhere. However, we note that Torrey and Gray (1843) included *V. nitidum* under their *V. nudum* var. *angustifolium*, and McAtee (1956) included *V. nudum* var. *angustifolium* within his *V. cassinoides* var. *nitidum*. Therefore, the name *V. nitidum* corresponds to the concepts of both authors who previously recognized the existence of such a lineage in the *V. nudum* complex.

### *Morphological Differentiation*

The failure by most botanists to recognize the third species in the *V. nudum* complex (the red clade) suggests that it differs only in cryptic or inconsistent morphological traits. And, indeed, we found overlap among the three lineages in most of the morphological traits that we measured (Fig. 6). Without prior information about species identity, clustering methods identified only two groups. For the most part these groups corresponded to *V. nudum* (green) and to the *V. cassinoides* (blue) + *V. nitidum* (red) clade, although the classification was not perfect. The fact that we did not recover all three lineages with the morphological clustering analysis indicates that there are not clear breaks in the distributions of the measured traits and that these species are somewhat cryptic (i.e., difficult to distinguish morphologically), at least along these specific axes of variation.

Although morphological differences among the three species are subtle and many herbarium specimens can be difficult to assign with certainty, in our experience identifications in the field are relatively straightforward. *Viburnum cassinoides* is the easiest to identify, in part because it is the only member of the complex that occurs in much of the Northeast and Canada. It often has wide, ovate leaves with crenate margins and displays of pink (immature) and black (mature) fruits. In the South, *V. nudum* and *V. nitidum* are broadly sympatric, and can co-occur in close proximity. *Viburnum nudum* leaves tend to be larger (20–30 $cm^2$), thicker, rounder, and more acute or obtuse (as opposed to acuminate) than they are in *V. nitidum* (blade size 10–18 $cm^2$). In addition, *V. nudum* often has in-rolled leaf margins and its leaves frequently dry darker on herbarium specimens. The main stalk of the flat-topped umbel-like inflorescence tends to be longer in *V. nudum* (33–44 mm) than it is in *V. nitidum* (16–29 mm). Finally, in late July and August, *V. nudum* fruits are often pale green to white, whereas the fruits are pink or black in sympatric *V. nitidum* populations. *Viburnum nudum* fruits ultimately mature to a dark blue or black color, but we have never observed the prolonged phase of bright pink that is, typical in both *V. nitidum* and *V. cassinoides*. Based on these characters, the two sympatric species are relatively easy to tell apart. We note that we focused our measurements on traits that have been used previously to distinguish *V. nudum*, in the broad traditional sense, from *V. cassinoides* (Jones 1983), and we suspect that additional, subtler characters (e.g., the shapes and sizes of peltate glands on the leaf surfaces) may vary consistently among the three species and may prove to be particularly useful in differentiating *V. nitidum* from *V. cassinoides*.

### *The Maintenance of Species Boundaries*

Based on our preliminary geographic range maps, the northern *V. cassinoides* may rarely overlap with populations of the two southern species along its southern and eastern boundary. Meanwhile, *V. nitidum* appears to exist entirely within the range of *V. nudum*. Given these distributions, how are the differences among the species maintained? Although closely related plant species are sometimes isolated by having different chromosome numbers (e.g., Wood et al. 2009), chromosome counts for *V. cassinoides* and *V. nudum sensu lato* have yielded 2N = 18 in all cases (Egolf 1962). Furthermore, extensive crossing studies have shown that viable hybrids can be produced when members of the same major clade of *Viburnum*, having the same chromosome number, are crossed with one another, and this holds for species of the *Lentago* clade. Several such hybrids have been cultivated, for example, a cross between *V. lentago* and *V. prunifolium* (*V.* × *jackii*) (Egolf 1956; Dirr 2007). Jones (1983) conducted several crossing experiments between the two entities he recognized in the *V. nudum* complex, but these are difficult to interpret because it is not clear whether the "*V. nudum* var. *nudum*" in his study corresponds to our *V. nudum* or to our *V. nitidum*. Despite this uncertainty, his results are intriguing. In both experiments, cross-species pollination resulted in a lower fruit set (a reduction to <1% in one experiment). Based on these studies we expect that crosses between any members of the *V. nudum* complex could yield offspring, although fruit set might be dramatically reduced.

Our experience in the field suggests that while *V. nudum* and *V. nitidum* can often be found in close proximity, they appear to be differentiated with respect to soil and vegetation types. *Viburnum nudum* is often found on low, poorly drained, and frequently inundated bald cypress (*Taxodium distichum*) and tupelo (*Nyssa biflora*) swamps. In contrast, *V. nitidum* occurs in adjacent and interdigitating, slightly higher and better-drained sandy soils. *Viburnum nudum* is the more common of the two species in most of the southeastern US, but in some regions, such as in northern Florida, *V. nitidum* may be encountered more frequently.

Their adaptation to different habitats may inhibit the success of hybrid offspring, but as we have found no evidence of hybrids at any life stage, we suspect that species boundaries are probably also maintained by prezygotic mechanisms. We have found evidence in other *Viburnum* lineages that consistent differences in flowering time contribute to the isolation of sympatric species (Donoghue 1980; Spriggs et al. *in review*), and temporal isolation might be a factor in this case as well. We note that Jones (1983) described differences in

flowering time in the sympatric populations he observed in Iredell County, North Carolina.

Our analysis highlights several regions of potential contact between the species that warrant more comprehensive sampling. Primary among these is the southern Appalachians, where our range maps indicate that all three species may occur. McAtee (1956) recognized *V. cassinoides* var. *harbisoni*, which he distinguished based on short, narrow leaves that form tufts and smaller inflorescences. This requires further investigation, though we suspect that it is not distinct from our *V. nitidum*.

### *Divergence Times, Population Structure, and Movements since the LGM*

Based on dates inferred across *Viburnum* (Spriggs et al. 2015), the *Lentago* clade probably started to diversify in North America more than 20 million years ago (Ma). Two individuals included in Spriggs et al. (2015) correspond to our *V. cassinoides* and *V. nudum*, and therefore, represent the oldest speciation event within the complex. The age of this split was estimated to be 4.8 Ma (95% posterior probability confidence interval of 1.19–11.34 Ma). Scaling the tau estimates from our BPP analysis to this age, the divergence of *V. cassinoides* from *V. nitidum* would have been some 1.1 Ma. If we consider the youngest reasonable date for these splits (the low end of the confidence interval), the divergence between *V. nudum* and *V. nitidum* + *V. cassinoides* would have occurred approximately 1.19 Ma and the split between *V. nitidum* and *V. cassinoides* at approximately 276,000 years ago. It is possible, therefore, that speciation within the *V. nudum* complex was influenced by earlier rounds of Quaternary glaciation. However, whether we accept the older or the younger age estimates, these speciation events occurred well before the most recent glaciation (~110,000–11,700 years ago).

How, then, did the last glaciation impact the three species? Our LGM distribution models predict nearly complete displacement of the ranges of the three species of the *V. nudum* complex (Fig. 7). The LGM ranges that we infer are broadly consistent with fossil evidence of the communities that occurred in eastern North America at that time. Pollen-based biome reconstructions for the Southeast have classified central to northern Florida as supporting warm mixed deciduous-coniferous forest, and the southern coastal plain (the region identified as suitable for *V. cassinoides*) as either cold mixed forest or boreal forest (Jackson et al. 2000; Williams et al. 2004). Although *Viburnum* pollen is generally rare in these fossil pollen assemblages, it is distinctive enough to be reported on a regular basis. However, few studies have attempted to assign pollen grains to specific *Viburnum* clades and it is often impossible to attribute them to particular species (Donoghue 1985). Therefore, paleopalynological studies can tell us that *Viburnum* existed at a certain time and place, but not which subclades or species were present.

We infer that *V cassinoides* was the most widespread member of the *V. nudum* complex during the LGM (Fig. 7), occupying the coastal plain from eastern Texas to southern New Jersey (Fig. 7 and Supplementary Fig. S5 available on Dryad). This finding is consistent with our BPP estimates of a much larger ancestral population size (theta) for *V. cassinoides* than for the other two species (Fig. 4). These results are also consistent with reports of *Viburnum* pollen at depths corresponding to the LGM (18,000–23,000 BP) at several sites in North Carolina (Whitehead 1981), Alabama (Delcourt et al. 1980), Louisiana (Jackson and Givens 1994), and Texas (Camper 1991) (Neotoma Paleoecology Database, http://www.neotomadb.org). Although a coastal range for *V. cassinoides* is plausible, it is surprising to us that our distribution models do not also identify the Upper Mississippi Alluvial Valley (western Tennessee, eastern Missouri) as potentially suitable habitat, as this region is considered to be an important refugial area for other cold-temperate forest species (e.g., *Liriodendron, Fagus, Carya, Ulmus, Fraxinus*; Delcourt and Delcourt 1984).

The glacial history of coastal plain species like *V. nudum* and *V. nitidum* has received less attention and is difficult to infer from paleoecological data (Jackson et al. 2000). We reconstruct small ranges for both species in peninsular Florida with tiny strips of available habitat along the Gulf Coast, and patches in Mexico and Central America as well. The climate of southern Florida at the LGM is somewhat uncertain, but evidence suggests that much of the central peninsula was dry, characterized by sand dunes and scrub vegetation (Watts 1980). It seems likely that both *V. nudum* and *V. nitidum* would have occurred in the vicinity of streams or rivers, as they often do today, and therefore, would have occupied only a subset of the LGM range that we reconstruct. We presume that the ecological (and possibly phenological) differences between them evolved well before the LGM, and that they sorted to different soil types and remained reproductively isolated even within their reduced ranges. We note that *Viburnum* pollen has been found in south central Florida at depths corresponding to between 46,200 BP and 18,700 BP (Grimm et al. 2006). Although our distribution models favor a scenario in which all three species were almost entirely displaced at the LGM, climate estimates for that time remain uncertain, and we cannot exclude the possibility that *Viburnum* populations persisted in specialized habitats, for example, steep ravines or river bluffs within their current ranges.

### *Diversity of the North American Coastal Plain*

Two of the three species that we recognize here—*V. nudum* and *V. nitidum*—occur primarily on the North American Coastal Plain (NACP), a region that has recently been proposed as a global biodiversity hotspot (Noss et al. 2015). Due to its correspondence with the geologic coastal plain, the NACP floristic province is one

of the most clearly defined in North America (Takhtajan 1986), and has an exceptionally high proportion of endemic species (Takhtajan 1986; Gleason and Cronquist 1991; Thorne 1993; Sorrie and Weakley 2001; Noss et al. 2015). Given its relatively flat topography and homogenous climate, explanations for the high diversity of the NACP have often focused on glacial refugia. However, there are many other factors that may have contributed to diversification in this region (Sorrie and Weakley 2001). In addition to the isolation caused by the contraction of geographic ranges at glacial maxima, dispersal barriers were created by extensive flooding of river systems as the glaciers receded and by coastal flooding during interglacial periods (Christensen 1988). Chronic disturbance caused by hurricanes, flooding, and, most importantly, fire have maintained a patchwork of diverse habitats across much of the region and created opportunities for species to adapt to specific habitats (Christensen 1988; Sorrie and Weakley 2001; Noss et al. 2015). Perhaps most significantly for *Viburnum*, the NACP contains a wide range of soil types (Buol 1973) and a diversity of freshwater wetland communities (Christensen 1988; Sorrie and Weakley 2001). *Viburnum nudum* and *V. nitidum* provide an example of close relatives differentiating into adjoining habitats, with *V. nitidum* on well-drained sandy soils frequently at the margins of streams, and *V. nudum* in more stagnant swamps. Our resurrection of *V. nitidum*, a NACP endemic, is consistent with the assertion of Sorrie and Weakley (2001) and of Noss et al. (2015) that the plant diversity of the NACP has been under-described. Here, phylogeographic and species delimitation studies of the type presented are very likely to expose additional hidden diversity and will help us to identify the factors that have generated the remarkable diversity of this region.

## Conclusions

The view is widespread that plant species are often difficult to delimit due to hybridization, polyploidy, asexual reproduction, and phenotypic plasticity (e.g., Levin 1979). At the outset, the *V. nudum* complex appeared to be a classic case of such difficulties. However, it turned out to be surprisingly straightforward, and this offers hope that the confluence of molecular, morphological, geographical, and ecological data, along the lines demonstrated here, will prove to be the rule in plants rather than the exception.

Our molecular analyses have clearly identified three distinct lineages, one of which (*V. nudum*) is highly divergent from the other two. The morphological differences among the three clades are minor, but *V. nudum* can be distinguished with considerable confidence based on morphology alone. The more recently diverged *V. nitidum* and *V. cassinoides* are not so clearly separated based on morphological traits, but molecular evidence indicates that these two clades are isolated by more than geographic distance, and they appear to have maintained their separation through significant geographic movements during glacial cycles. Based on all of the available evidence we hypothesize that these three lineages are evolving independently, and we therefore, recognize them as separate species.

Speciation in the *V. nudum* complex, by our estimates, occurred well before the LGM, and although recent glaciation had significant impacts on the geographic distributions of these species, it was not responsible for initiating the divergences among them. The relatively old ages of the divergences makes it difficult to identify the conditions that drove speciation. However, we suspect that ecological differentiation occurred early on, with the split between *V. nudum* and *V. nitidum* + *V. cassinoides*. This differentiation was probably maintained during subsequent glacial cycles and range shifts. Recent glaciation undoubtedly resulted in the shifting of species ranges in eastern North America, but in this case, and perhaps in many others, it did not drive speciation.

## Supplementary Material

Data available from the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.m1vc0.

## Funding

This work was supported by the National Science Foundation [grants IOS-1256706 to M.J.D., IOS-1257262 to E.J.E., DGE 1122492 to E.L.S., and NSF 1501188 to E.L.S. and M.J.D.]. Additional funding was provided by the Torrey Botanical Society, the American Society of Plant Taxonomists, the Yale Peabody Museum of Natural History, and the Yale Institute for Biospheric Studies.

## Acknowledgments

We thank Thomas Near, Felipe Zapata, Bruce Sorrie, Alan Weakley, Brian Park, Yoshihisa Suyama, and Tetsukazu Yahara for helpful discussions. We are also grateful to the Florida State Parks, and the North Carolina State Parks for assistance with collecting, and to the New York Botanical Garden for permission to work with their specimens. Finally, we thank the herbaria of the following institutions: New York Botanical Garden (NY), Louisiana State University (LSU), University of Connecticut (CONN), University of North Carolina, Chapel Hill (NCU), Florida State University (FSU), Ohio State University (OS), the University of Wisconsin Herbarium (WIS), Tulane University (NO), University of Texas (TEX, LL), Delta State University (DSC), Mississippi State University (MISSA), and the Mississippi Museum of Natural Science (MMNS) for providing the specimens or images examined to generate our range maps.

## References


Aiton W. 1789. Hortus Kewensis. London: printed for George Nicol, bookseller to his Majesty.

Baird N.A., Etter P.D., Atwood T.S., Currey M.C., Shiver A.L., Lewis Z.A., Selker E.U., Cresko W.A., Johnson E.A. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS One. 3:e3376.

Barnard-Kubow K.B., Debban C.L., Galloway L.F. 2015. Multiple glacial refugia lead to genetic structuring and the potential for reproductive isolation in a herbaceous plant. Am. J. Bot. 102:1842–1853.

Beck J.B., Alexander P.J., Allphin L., Al-Shehbaz I.A., Rushworth C., Bailey C.D., Windham M.D. 2012. Does hybridization drive the transition to asexuality in diploid *Boechera*? Evolution. 66:985–995.

Bryant D., Moulton V. 2004. Neighbor-net: an agglomerative method for the construction of phylogenetic networks. Mol. Biol. Evol. 21:255–265.

Buol S.W. 1973. Soils of the southern states and Puerto Rico. Southern Cooperative Series Bulletin No. 174.

Cadena C.D., Zapata F., Jiménez I. 2018. Issues and perspectives in species delimitation using phenotypic data: Atlantean evolution in Darwin's Finches. Syst. Biol. 67(2):181–194.

Camper H.A. 1991. Pollen analysis of Patschke Bog. College Station (TX): Texas A and M University.

Chifman J., Kubatko L. 2014. Quartet inference from SNP data under the coalescent model. Bioinformatics. 30:3317–3324.

Christensen N.L. 1988. Vegetation of the southeastern coastal plain. In: Barbour M.G., Billings W.D., editors. North American terrestrial vegetation. Cambridge: Cambridge University Press. p. 317–363

Clark R.C. 1971. The woody plants of Alabama. Ann. Mo. Bot. Gard. 58:99–242.

Clement W.L., Arakaki M., Sweeney P.W., Edwards E.J., Donoghue M.J. 2014. A chloroplast tree for *Viburnum* (Adoxaceae) and its implications for phylogenetic classification and character evolution. Am. J. Bot. 101:1029–1049.

Clement W.L., Donoghue M.J. 2011. Dissolution of *Viburnum* section Megalotinus (Adoxaceae) of Southeast Asia and its implications for morphological evolution and biogeography. Int. J. Plant Sci. 172:559–573.

de Queiroz K. 2007. Species concepts and species delimitation. Syst. Biol. 56(6):879–886.

Delcourt H.R., Delcourt P.A. 1984. Ice age haven for hardwoods. Nat. Hist. 93:22–28.

Delcourt P.A., Delcourt H.R., Brister R.C., Lackey L.E. 1980. Quaternary vegetation history of the Mississippi embayment. Quat. Res. 13:111–132.

Dirr M.A. 2007. Viburnums: flowering shrubs for every season. London: Timber Press.

Donoghue M.J. 1980. Flowering times in *Viburnum*. Arnoldia. 40:2–22.

Donoghue M.J. 1985. Pollen diversity and exine evolution in Viburnum and the Caprifoliaceae sensu lato. J. Arnold Arbor. 66:421–469.

Doyle J., Doyle J.L. 1987. Genomic plant DNA preparation from fresh tissue—CTAB method. Phytochem. Bull. 19:11–15.

Eaton D.A.R., Hipp A.L., González-Rodríguez A., Cavender-Bares J. 2015. Historical introgression among the American live oaks and the comparative nature of tests for introgression. Evolution. 69(10):2587–2601.

Eaton D.A.R., Ree R.H. 2013. Inferring phylogeny and introgression using RADseq data: an example from flowering plants (Pedicularis: Orobanchaceae). Syst. Biol. 62(5):689–706.

Eaton D.A.R., Spriggs E.L., Park B., Donoghue M.J. 2017. Misconceptions on missing data in RAD-seq phylogenetics with a deep-scale example from flowering plants. Syst. Biol. 66:399–412.

Egolf D.R. 1956. Cytological and interspecific hybridization studies in the genus *Viburnum*. Ithaca (NY): Cornell University.

Egolf D.R. 1962. A cytological study of the genus *Viburnum*. J. Arnold Arbor. 43:132–172.

Etter P.D., Preston J.L., Bassham S., Cresko W.A., Johnson E.A. 2011. Local *de novo* assembly of RAD paired-end contigs using short sequencing reads. *PLoS One.* 6:e18561.

Evanno G., Regnaut S., Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol. Ecol. 14:2611–2620.

Ferguson I.K. 1966. The genera of Caprifoliaceae in the southeastern United States. J. Arnold Arbor. 47:33–59.

Fernández-Mazuecos M., Mellers G., Vigalondo B., Llorenç S., Vargas P., Glover, B.J. 2018. Resolving recent plant radiations: power and robustness of genotyping-by-sequencing. Syst. Biol. 67(2):250–268.

Fraley C., Raftery A.E., Murphy T.B., Scrucca, L. 2012. Mclust Version 4 for R: Normal Mixture Modeling for Model-Based Clustering, Classification, and Density Estimation. Department of Statistics, University of Washington. Technical Report No. 597.

Gleason H.A., Cronquist A. 1991. Manual of vascular plants of Northeastern United States and adjacent Canada. Bronx (NY): The New York Botanical Garden Press.

Grimm E.C., Watts W.A., Jacobson G.L., Hansen B.C.S., Almquist H.R., Dieffenbacher-Krall A.C. 2006. Evidence for warm wet Heinrich events in Florida. Quat. Sci. Rev. 25:2197–2211.

Guindon S., Dufayard J.-F., Leford V., Anisimova M., Hordijk W., Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. 59(3):307–321.

Haines A. 2011. Flora Novae Angliae. New Haven (CT): New England Wildflower Society & Yale University Press. p. 1–973.

Hijmans R.J. 2016. Geosphere: spherical trigonometry. R package version 1.5-5. https://CRAN.R-project.org/package=geosphere.

Hijmans R.J., Phillips S., Leathwick J., Elith J. 2017. dismo: species distribution modeling. R package version 1.1-4. https://CRAN.R-project.org/package=dismo.

Hoang D.T., Chernomor O., von Haeseler A., Minh B.Q., Vinh L.S. 2018. UFBoot2: improving the ultrafast bootstrap approximation.Mol. Ecol. Evol. 35(2):518–522.

Hudson D.H., Bryant, D. 2006. Application of phylogenetic networks in evolutionary studies. Mol. Ecol. Evol. 23(2):254–267.

Jackson S.T., Givens C.R. 1994. Late Wisconsin vegetation and environment of the Tunica Hills Region, Louisiana/Mississippi. Quat. Int. 41:316–325.

Jackson S.T., Webb R.S., Anderson K.H., Overpeck J.T., Webb T.W., Williams J.W., Hansen B.C.S. 2000. Vegetation and environment in Eastern North America during the last glacial maximum. Quat. Sci. Rev. 19:489–508.

Jakobsson M., Rosenberg N.A. 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics. 23:1801–1806.

Jombart T., Devillard S., Balloux F. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genet. 11:1–15.

Jones T.H. 1983. A revision of the genus *Viburnum* section Lentago (Caprifoliaceae) [Ph.D.]. [Raleigh (NC)]: North Carolina State University.

Kalinowski S.T. 2011. The computer program STRUCTURE does not reliably identify the main genetic clusters within species: simulations and implications for human population structure. Heredity. 106(4):625–632.

Kalyaanamoorthy S., Minh B.Q., Wong T.K.F., von Haeseler A., Jermiin L.S. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods. 14:587–589.

Kartez J.T. 2015. The Biota of North America Program (BONAP). Taxonomic Data Center.

Laricchia K.M., McCleary T.S., Hoban S.M., Borkowski D.S., Romero-Severson J. 2015. Chloroplast haplotypes suggest preglacial differentiation and separate postglacial migration paths for the threatened North American forest tree *Juglans cinerea* L. Tree Genetics & Genomes. 11:1–11.

Leaché A.D., Zhu T., Rannala B., Yang Z. 2019. The Spectre of too many species. Syst. Biol. 68(1):168–181.

Levin D.A. 1979. The nature of plant species. Science. 204:381–384.

Little E.L. 1971. Atlas of the United States trees. Vol. 991. U.S. Washington (DC): Department of Agriculture Miscellaneous Publication. p. 1–97.

McAtee W.L. 1956. A review of the Neartic *Viburnum*. Chapel Hill (NC): Privately printed.
McLachlan J.S., Clark J.S., Manos P.S. 2005. Molecular indicators of tree migration capacity under rapid climate change. Ecology. 86:2088–2098.
Minh B.Q., Nguyen M.A.T., von Haeseler A. 2013. Ultrafast approximation for phylogenetic bootstrap. Mol. Biol. Evol. 30(5):1188–1195.
Morris A.B., Graham C.H., Soltis D.E., Soltis P.S. 2010. Reassessment of phylogeographical structure in an eastern North American tree using Monmonier's algorithm and ecological niche modelling. J. Biogeogr. 37:1657–1667.
Morris A.B., Ickert-Bond S.M., Brunson B., Soltis D.E., Soltis P.S. 2008. Phylogeographical structure and temporal complexity in American sweetgum (*Liquidambar styraciflua*; Altingiaceae). Mol. Ecol. 17:3889–3900.
Nadeau S., Godbout J., Lamothe M., Gros-Louis M-C., Isabel N., Ritland K. 2015. Contrasting patterns of genetic diversity across the ranges of *Pinus monticola* and *P. strobus*: a comparison between eastern and western North American postglacial colonization histories. Am. J. Bot. 102:1342–1355.
Nguyen L.-T., Schmidt H.A., von Haeseler A., Minh B.Q. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. Mol. Biol. Evol. 33:268–274.
Noss R.F., Platt W.J., Sorrie B.A., Weakley A.S., Means D.B., Costanza J., Peet R.K. 2015. How global biodiversity hotspots may go unrecognized: lessons from the North American Coastal Plain. Divers. Distrib. 21:236–244.
Peterson B.K., Weber J.N., Kay E.H., Fisher H.S., Hoekstra H.E. 2012. Double digest RADseq: an inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. *PLoS One.* 7:e37135.
Phillips S., Anderson R.P., Schapire R.E. 2006. Maximum entropy modeling of species geographic distributions. Ecol. Model. 190:231–259.
Pritchard J.K., Stephens M., Donnelly P. 2000. Inference of population structure using multilocus genotype data. Genetics. 155:945–959.
Radford A.E., Ahles H.E., Bell R.C. 1968. Manual of the vascular flora of the Carolinas. Chapel Hill (NC): The University of North Carolina Press.
Revell L.J. 2012. Phytools: an R package for phylogenetic comparative biology (and other things). Methods Ecol. Evol. 3:217–223.
Schaal B.A., Hayworth D.A., Olsen K.M., Rauscher J.T., Smith WA. 1998. Phylogeographic studies in plants: problems and prospects. Mol. Ecol. 7:465–474.
Small J.K. 1933. Manual of the Southeastern Flora. New York (NY): Published by the Author.
Snir S., Rao S. 2012. Quartet MaxCut: a fast algorithm for amalgamating quartet trees. Mol. Phylogenet. Evol. 62:1–8.
Sorrie B.A., Weakley A.S. 2001. Coastal plain vascular plant endemics: phytogeographic patterns. Castanea. 66:50–82.
Spriggs E.L., Clement W.L., Sweeney P.W., Madriñán S., Edwards E.J., Donoghue M.J. 2015. Temperate radiations and dying embers of a tropical past: the diversification of *Viburnum*. New Phytol. 207:340–354.
Spriggs E.L., Schlutius C., Eaton D.A., Park B., Sweeney P.W., Edwards E.J., Donoghue M.J. Differences in flowering time maintain species boundaries in a continental radiation of *Viburnum*. Am. J. Bot. In review.
Spriggs E.L., Schmerler S.B., Edwards E.J., Donoghue M.J. 2018. Leaf form evolution in Viburnum parallels variation within individual plants. Am. Nat. 191(2):235–249.
Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 30:1312–1313.
Strausbaugh P.D., Core E.L. 1978. Flora of West Virginia. 2nd ed. Grantsville (WV): Seneca Books, Inc.
Sukumaran J., Knowles L.L. 2017. Multispecies coalescent delimits structure, not species. Proc. Natl. Acad. Sci. USA 114(7):1607–1612.
Takhtajan A.L. 1986. Essays on the evolutionary morphology of plants. Washington (DC): American Institute of Biological Sciences.
Tanaka A., Ohtani M., Suyama Y., Inomata N., Tsumura Y., Middleton B.A., Tachida H., Kusumi J. 2012. Population genetic structure of a widespread coniferous tree, *Taxodium distichum* [L.] Rich. (Cupressaceae) in the Mississippi River Alluvial Valley and Florida. Tree Genetics & Genomes. 8:1135–1147.
Thomson A.M., Dick C.W., Dayanandan S. 2015. A similar phylogeographical structure among sympatric North American birches (*Betula*) is better explained by introgression than by shared biogeographical history. J. Biogeogr. 42:339–350.
Thorne R.F. 1993. Phytogeography. In: Committee FoNAE, editor. Flora of North America Editorial Committee, Vol. I. New York (NY): Oxford University Press. p.132–153.
Torrey J., Gray A. 1843. Flora of North America. New York (NY): Wiley & Putnam.
Watts W.A. 1980. The late Quaternary vegetation history of the Southeastern United States. Ann. Rev. Ecol. Syst. 11:387–409.
Weakley A.S. 2012. Flora of the Southern and Mid-Atlantic States. Available from: URL http://www.herbarium.unc.edu/flora.htm. Accessed August 2017.
Whitehead D.R. 1981. Late-Pleistocene vegetational changes in northeastern North Carolina. Ecol. Monogr. 52:451–471.
Williams J.W., Shuman B.N., Webb T.I., Bartlein P.J., Leduc P.L. 2004. Late-Quaternary vegetation dynamics in North America: scaling from taxa to biomes. Ecol. Monogr. 74:309–334.
Wood T.E., Takebayashi N., Barker M.S., Mayrose I., Greenspoon P.B., Rieseberg L.H. 2009. The frequency of polyploid speciation in vascular plants. Proc. Natl. Acad. Sci. USA. 106:13875–13879.
Yang Z. 2015. The BPP program for species tree estimation and species delimitation. Curr. Zool. 61(5) 854–865.
Yang Z., Rannala B. 2010. Bayesian species delimitation using multilocus sequence data. Proc. Natl. Acad. Sci. USA. 107:747–761.
Zinck J.W.R., Rajora O.P. 2016. Post-glacial phylogeography and evolution of a wide-ranging highly-exploited keystone forest tree, eastern white pine (*Pinus strobus*) in North America: single refugium, multiple routes. BMC Evol. Biol.16:1–17.