RESEARCH ARTICLE

AMERICAN JOURNAL OF BOTANY

Multiple glacial refugia lead to genetic structuring and the potential for reproductive isolation in a herbaceous plant[1]

Karen B. Barnard-Kubow[2], Catherine L. Debban, and Laura F. Galloway

PREMISE OF THE STUDY: Glacial cycles have influenced the genetic structure of many species. In addition to facilitating genetic divergence, isolation in multiple glacial refugia may have contributed to the development of genetic incompatibility and reproductive isolation. We examined the phylogeography of *Campanulastrum americanum*, a monocarpic herbaceous plant that exhibits strong intraspecific reproductive isolation, to determine whether the current genetic structure reflects a history of multiple glacial refugia.

METHODS: Chloroplast loci and nuclear RAD sequencing were used to characterize the range-wide phylogeography of *C. americanum*, in order to determine locations of potential glacial refugia and recolonization routes. Potential locations of refugia during the Last Glacial Maximum were also identified using ecological niche modeling.

KEY RESULTS: Together, the chloroplast and nuclear phylogenies found support for three geographically structured, genetically divergent lineages, among which gene flow appears to be restricted. The distribution of these lineages indicates that *C. americanum* survived the Last Glacial Maximum in at least three refugia located in the Appalachians and on the Atlantic and Gulf coasts. The ecological niche model also supported the existence of multiple refugia.

CONCLUSIONS: The isolation of populations of *C. americanum* in multiple refugia has led to a degree of phylogeographic structure greater than that found in most previously studied plants in eastern North America, which may be attributable to its short generation time. Reproductively isolated populations of *C. americanum* belong to divergent lineages, which suggests that survival in multiple glacial refugia contributed to the development of reproductive isolation in this species.

KEY WORDS *Campanula*; *Campanulastrum*; chloroplast DNA; eastern North America; ecological niche modeling; glacial refugia; phylogeography; RAD-seq; reproductive isolation; spatial distribution modeling

The genetic structure of a species is a manifestation of both historical processes and contemporary patterns of gene flow. One of the primary historical processes that influence genetic structure is climate change, in particular glacial cycles, which have caused cyclic expansions and contractions of many species' ranges (Hewitt, 2000, 2004; Tollefsrud et al., 2008; Gugger et al., 2013; Roberts and Hamann, 2015). Phylogeography (Avise et al., 1987) is often used to understand the effects of glacial cycles on current genetic structure, including the number and location of glacial refugia as well as the recolonization routes from those refugia to current locations (Hewitt, 1996; Taberlet et al., 1998; Hewitt, 2000).

Several themes have emerged from phylogeographic studies. These include the frequent location of glacial refugia in the southern reaches of the Northern Hemisphere, with refugia found in the southern peninsulas of Europe and along the Gulf Coast in the eastern United States (Taberlet et al., 1998; Avise, 2000; Hewitt, 2000; Soltis et al., 2006). There is also now general acceptance of additional "cryptic" refugia, where species survived close to the ice sheet, both in Europe and North America (Stewart and Lister, 2001; Tribsch and Schonswetter, 2003; McLachlan et al., 2005; Provan and Bennett, 2008; Parducci et al., 2012; Suarez-Gonzalez et al., 2015). In Europe, distinct species exhibit congruence with respect to the locations of glacial refugia, subsequent colonization routes,

[1] Manuscript received 5 June 2015; revision accepted 22 September 2015.

Department of Biology, University of Virginia, Charlottesville, Virginia 22904-4328, USA

[2] Author for correspondence (e-mail: kbb7sh@virginia.edu)

doi:10.3732/ajb.1500267

and areas of secondary contact (Taberlet et al., 1998; Hewitt, 2000). While some congruence has also been seen in eastern North America (Avise, 2000; Soltis et al., 2006), species' responses to glacial cycles generally appear to be individualistic and complex (Soltis et al., 2006). This lack of congruence is likely a result of the north–south orientation of the Appalachian Mountains having acted as less of a migration barrier to recolonization from southern refugia than the east–west orientation of the major mountain ranges in Europe. This complexity raises the need for phylogeographic studies covering a diverse array of organisms. However, studies in eastern North America have historically been biased toward animals (reviewed in Soltis et al., 2006), leaving the phylogeography of plant species relatively understudied (but see Griffin and Barrett, 2004; McLachlan et al., 2005; Gonzales et al., 2008; Morris et al., 2008; Li et al., 2013; Nadeau et al., 2015). Particularly lacking are studies on shorter-lived taxa, as the majority of plant phylogeographic studies have focused on trees and perennial herbs.

Isolation in multiple glacial refugia may have led to genetic divergence among populations due to genetic drift and natural selection (Hewitt, 1996). Glacial periods have been longer than interglacials, which means that during the 18–20 glacial cycles in the past 2 million years, species spent more time in glacial refugia than in their current widespread distributions (reviewed in Davis, 1983). Recolonization would have further contributed to drift and genetic divergence as dispersal events led to recurrent bottlenecks (Excoffier et al., 2009). These processes may also have facilitated the development of genetic incompatibility and the potential for reproductive isolation and speciation (Hewitt, 1996; Avise et al., 1998; Carstens and Knowles, 2007; April et al., 2013). If so, then clades identified in phylogeographic studies may correspond to incipient species. However, few studies have directly addressed the interplay between phylogeography and intraspecific reproductive isolation, especially in plants (though see Pinheiro et al., 2013).

To determine whether isolation in multiple glacial refugia can help explain the existence of intraspecific reproductive isolation, we examined the phylogeography of *Campanulastrum americanum*, a herbaceous plant that has strong reproductive isolation between populations (≤90% reduction in cumulative fitness; Galloway and Etterson, 2005; Etterson et al., 2007). Specifically, we address the following questions: Does the phylogeography of *C. americanum* based on chloroplast and nuclear markers support survival in multiple allopatric glacial refugia, and are the patterns concordant between the two types of markers? Does niche modeling also support multiple glacial refugia, and do the proposed locations match those suggested by phylogeography?

MATERIALS AND METHODS

Study system*—Campanulastrum americanum* (L.) Small (= *Campanula americana* L.; Campanulaceae) is an autotetraploid, monocarpic herb found in the eastern half of the United States, with populations from the Appalachians to just west of the Mississippi River (Fig. 1A). Populations are not evenly distributed across the range; the species is less common in the South (southern Mississippi, Alabama, and northern Florida). Individuals can be annual or biennial, are insect-pollinated, and are primarily outcrossing (Galloway et al., 2003). The plant typically grows in disturbed habitats in the understory or along the forest edge, and seeds are dispersed passively—traits that likely contribute to its patchy population structure.

***Chloroplast sequencing and phylogeny*—**To determine the chloroplast phylogeny of *C. americanum*, leaf tissue samples were taken from individuals from 49 populations across the species range (Table 1). Leaf tissue was primarily taken from individuals grown in a greenhouse from field-collected seed, though a small number of field-collected leaf tissue samples were also used. Leaf tissue from *Triodanis biflora* was used as an outgroup for the phylogeny construction because the genus *Triodanis* is closely related to *C. americanum* according to both chloroplast and ITS sequence (Eddie et al., 2003; Wendling et al., 2011). DNA was extracted for all samples using a modified CTAB procedure.

Screening of preexisting chloroplast primer sets (Taberlet et al., 1991; Hamilton, 1999) gave limited success. To obtain more loci, we screened for areas of high polymorphism in a 454 dataset that consisted of full chloroplast genome sequence of four pooled *C. americanum* populations (AL29, MN12, OH64, and VA73; Barnard-Kubow et al., 2014). We obtained four potential loci, which we sequenced in conjunction with a locus that included the trnL intron and trnL–trnF intergenic spacer (Taberlet et al., 1991). Because of difficulties with amplification, *C. americanum*–specific external primers were designed for the trnL intron and trnL–trnF spacer. For primer sequences used in amplification and sequencing and their locations in the chloroplast genome, see Supplemental Data with the online version of this article (Appendix S1).

Initially, a single individual from each population was sequenced at all five loci to determine the overall phylogeographic pattern (geographic distribution of clades). To determine the robustness of this pattern and check for additional polymorphism, we then sequenced 5–13 additional individuals from each of 13 populations (see Supplemental Data with the online version of this article; Appendix S2). These 13 populations were chosen on the basis of proximity to expected southern glacial refugia (from other species) or proximity to populations of divergent chloroplast clades. Most of the additional individuals were genotyped at the subset of loci that distinguished them from nearby haplotypes, because we were primarily concerned with detecting shared haplotypes among populations. If populations were fixed for particular chloroplast haplotypes or contained only closely related haplotypes, we assumed that sampling additional individuals from the remaining populations would be unlikely to alter the overall phylogeographic pattern.

Loci were amplified using 5 PRIME HotMasterMix (5 Prime, Gaithersburg, Maryland, USA). The PCR products were purified using Exonuclease I and Shrimp Alkaline Phosphatase (Affymetrix, Santa Clara, California, USA). Cycle sequencing reactions were carried out using BigDye Terminator v3.1 Cycle Sequencing Kit (Thermo Fisher Scientific, Waltham, Massachusetts, USA) and were cleaned using Sephadex G-50 (GE-Healthcare, Little Chalfont, United Kingdom). The resulting products were analyzed on an ABI 3130xl sequencer (Thermo Fisher Scientific), and sequences were checked and manually aligned using Biodit (Hall, 1999) or Codon Code Aligner version 3.5 (CodonCode Corporation, Centerville, Maine, USA).

Sequences were concatenated across loci for each individual. Maximum likelihood trees were constructed in Mega version 5.2.2 (Tamura et al., 2011) for each locus as well as for the concatenated sequences. A GTR model of evolution was used for C11 and rps4, while HKY was used for ycf1 and CLF. Finally, GTR + I was used for rps2 and the concatenated sequences. Models of evolution were chosen on the basis of results from jModeltest version 2.1.5 (Guindon and Gascuel, 2003; Darriba et al., 2012). Bayesian analysis was also

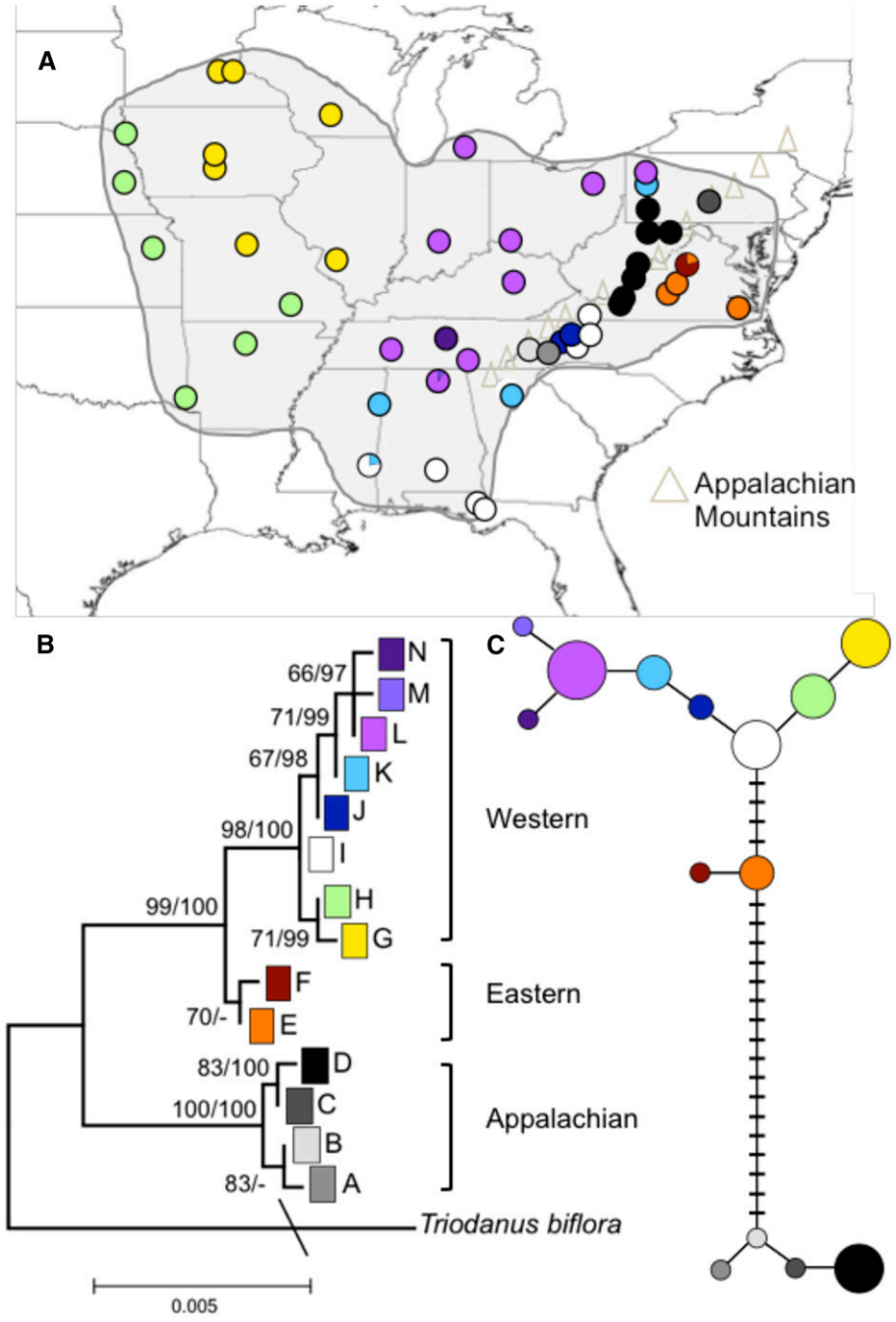

FIGURE 1 (A) Map showing locations of *Campanulastrum americanum* populations (colored circles) within an approximation of the species range (shaded area) in the eastern United States. Color indicates the population's chloroplast haplotype; pie charts represent polymorphic populations. (B) Maximum likelihood tree and (C) statistical parsimony network for concatenated chloroplast sequence data from *C. americanum*. Numbers in B show bootstrap support and Bayesian posterior probabilities (minus sign indicates Bayesian posterior probability <50). Colored circles in C represent individual haplotypes, connecting lines represent single mutations, and the horizontal tick marks represent unobserved haplotypes; size of circles is scaled according to how frequently that haplotype was observed (1–9 times).

run on the concatenated sequences using MrBayes version 3.2.2 (Ronquist et al., 2012), again using a GTR + I model of evolution. Two independent analyses were run starting with random trees and using one cold and three heated chains. Each analysis ran for 1 million generations, sampling one tree every 1000 generations. At the end of the run the standard deviation of split frequencies was <0.01, indicating convergence. The first 25% of samples were discarded from the cold chain, and posterior probabilities for supported clades were determined by majority-rule consensus of trees retained after burn-in. A statistical parsimony network was constructed for the concatenated sequences using TCS version 1.21 (Clement et al., 2000) with a 95% parsimony connection limit.

***Nuclear sequencing and phylogeny*—**To determine the nuclear phylogeny of *C. americanum*, 25 populations were chosen (Table 1) for RAD-seq (Davey et al., 2011). This technique enables the random sampling of single-nucleotide polymorphisms (SNPs) from across the genome. Populations were chosen to represent all chloroplast haplotypes and to achieve a broad geographic sampling, with a focus on the Appalachian region where the chloroplast clades co-occur. For each population, leaf tissue samples were taken from six individuals grown from field-collected seed, except for VA93, where first-generation greenhouse seed was used. DNA was extracted using a modified CTAB procedure, and concentration was determined using a Qubit Assay Kit (Thermo Fisher Scientific). The six DNA extractions from each population were evenly pooled according to concentration and then barcoded using custom-designed adapters with barcode sequences taken from Illumina's TruSeq list of published barcodes.

The barcoded populations were digested with SacI and then pooled for library construction following standard protocols for sequencing on an Illumina Hiseq. Due to low sequence complexity resulting from the restriction enzyme cut site, a 50% PhiX control was spiked into the RAD library before sequencing on two lanes of an Illumina HiSeq with 100-bp reads. Library construction was performed at the Genomics Core Facility in the University of Virginia's Department of Biology, and sequencing was performed at Beckman Coulter Genomics (Danvers, Massachusetts, USA).

The fastq files were run through 'process_radtags' in Stacks (Catchen et al., 2011; 2013) to demultiplex the data and filter out low quality reads. The processed fastq files were individually mapped using Bowtie 2 (Langmead and Salzberg, 2012) to a partial reference genome for *C. americanum* obtained from earlier sequencing of a whole-genome shotgun library (K. Barnard-Kubow, University of Virginia,

TABLE 1. Information on *Campanulastrum americanum* populations sampled in this study, including ID number, location within the United States, chloroplast haplotype, and NCBI Short Read Archive accession numbers for populations included in nuclear RAD sequencing.

Population ID	State	Latitude	Longitude	Haplotype	RAD accession no.
29	Alabama	34.6514	−86.5017	M/L	SRX1032954
69	Alabama	31.8485	−86.6402	I	
58	Arkansas	35.9106	−92.6316	H	
16	Florida	30.7690	−85.2419	I	
83	Florida	30.5647	−84.9598	I	SRX1032955
22	Georgia	34.4681	−84.4298	K	SRX1032956
20	Illinois	38.6074	−89.8992	G	SRX1032957
68	Indiana	39.1643	−86.5269	L	SRX1032958
10	Iowa	42.0728	−93.6725	G	SRX1032959
25	Iowa	41.6884	−93.7121	G	
45	Iowa	42.7554	−96.6069	H	
60	Kansas	39.0474	−95.6815	H	
51	Kentucky	37.9340	−84.2595	L	
5	Maryland	39.6137	−79.1158	D	SRX1032960
44	Michigan	42.2909	−85.5942	L	SRX1032961
12	Minnesota	44.8178	−93.0075	G	
38	Minnesota	44.8212	−93.3187	G	
55	Mississippi	31.7433	−88.5239	I/J	SRX1032962
70	Mississippi	33.9842	−88.4882	K	
7	Missouri	37.1328	−91.2772	H	
50	Missouri	39.1465	−92.6842	G	
59	Nebraska	41.1609	−96.5393	H	SRX1032963
36	North Carolina	36.1850	−81.6683	I	SRX1032964
90	North Carolina	35.7669	−82.1636	I	SRX1032968
91	North Carolina	35.5862	−83.0663	A	SRX1032979
11	Ohio	39.2718	−84.2841	L	
64	Ohio	41.1147	−81.5181	L	
61	Oklahoma	33.9464	−94.5669	H	SRX1032986
13	Pennsylvania	40.2721	−79.8850	D	
27	Pennsylvania	41.0079	−80.0833	K	SRX1032987
94	Pennsylvania	41.4670	−80.0113	L	
95	Pennsylvania	40.4752	−78.2808	C	
102	Pennsylvania	40.3225	−80.1109	D	
87	Tennessee	36.1211	−82.3178	J	
88	Tennessee	35.9833	−82.4989	J	SRX1032988
92	Tennessee	35.6758	−83.5259	B	SRX1032989
19	Tennessee	35.7583	−88.0687	L	
32	Tennessee	35.0748	−85.6250	L	
34	Tennessee	36.0822	−86.2961	N	SRX1032990
71	Virginia	38.3305	−78.4901	E/F	SRX1032991
73	Virginia	37.3534	−80.5522	D	SRX1032992
74	Virginia	37.9500	−78.8900	E	SRX1032993
85	Virginia	37.7576	−79.1876	E	SRX1032994
86	Virginia	36.6344	−81.5881	I	SRX1032995
93	Virginia	37.2067	−76.9476	E	SRX1033036
100	Virginia	37.2774	−80.6126	D	
72	West Virginia	37.9931	−80.3618	D	SRX1033037
84	West Virginia	38.2266	−80.2231	D	
14	Wisconsin	43.3362	−89.9467	G	

unpublished data). The resulting alignments were used as input for Stacks (Catchen et al., 2011, 2013) to identify and genotype loci using the following programs: 'pstacks', 'cstacks', 'sstacks', and 'populations'. A minimum read depth of 12 was used when constructing stacks to increase the chances that each individual within a pool was represented by at least a single read. Because *C. americanum* is an autotetraploid, and to ensure that SNPs were between homologous sequences instead of paralogous ones, we used only SNPs that were fixed within populations but variable between. Therefore, we used the fixed SNP model in 'pstacks' (Catchen et al., 2013). This approach resulted in some potential loss of information, in that population-divergence-based phylogenetic analyses could not be used, and phylogenetic analysis was restricted to SNP relationships. The barcode error frequency was estimated using the percentage of reads that were dropped in 'process_radtags' because of ambiguous barcodes. To ensure that SNPs were from homologous sequence, only one mismatch was allowed between tags when generating the catalog in *cstacks*. Analyses were also run allowing two mismatches, which gave qualitatively similar results. 'Populations' was run using loci for which sequence data were available for at least eight populations. Analyses were also run using loci present in a minimum of four, six, or 12 populations with qualitatively similar results.

A Perl script was used to concatenate the sequences from the Stacks 'FASTA' output for each population (the FASTA output contained the full sequence of each allele from each sample locus). The concatenated sequences were then used to construct a maximum likelihood tree in RAxML version 8.0.6 (Stamatakis, 2014) using 100 bootstrap replicates and a GTR + gamma model of evolution. The tree was visualized in Mega and rooted according to the chloroplast phylogeny. For comparison, the processed fastq files were also run through Stacks denovo, using 'ustacks' instead of 'pstacks' with similar parameter options.

***Isolation-by-distance*—**Isolation-by-distance analyses were carried out using the Isolation By Distance Web Service (Jensen et al., 2005) to calculate Mantel's *r* and examine the extent to which the genetic structuring of *C. americanum* can be explained by geographic isolation. Maximum likelihood distances were calculated for the chloroplast sequences using Mega and for the nuclear RAD data using RAxML. Isolation-by-distance analyses were run separately for both the chloroplast and nuclear data using all populations and all populations minus the divergent Appalachian clade. Separate analyses were also run for each of the three genetic lineages and the Smoky populations (see below).

***Ecological niche modeling*—**To identify possible glacial refugia for *C. americanum*, ecological niche modeling was performed using MaxEnt version 3.3.3k (Phillips et al., 2006). This form of modeling uses known occurrence data and current environmental data to generate a model that predicts the location of suitable habitat for a species. This model can then be projected onto past climates to predict how the location of suitable habitat has changed over time. For the occurrence data, 963 locations were obtained from the Galloway laboratory, the Global Biodiversity Information Facility (http://www.gbif.org), the Wisconsin Vascular Plants Wisflora database (http://wisflora.herbarium.wisc.edu/), the Illinois Natural History Survey (http://www.inhs.illinois.edu/), the Miami University herbarium (http://herbarium.muohio.edu/herbariummu/), the Ohio State University herbarium (https://herbarium.osu.edu/), and the University of North Carolina herbarium (http://www.herbarium.unc.edu/). Any locations collected before 1950 were removed, to match the time range of the current climate data.

Because herbaria are most likely to have samples from the surrounding areas and because some herbaria have many more samples than others, a high frequency of points within a given area is more likely to represent spatial autocorrelation from sampling bias than natural abundance. The original occurrence data were statistically significantly clustered ($z = -26.287$, $P < 0.0001$), indicating spatial autocorrelation between samples. To reduce spatial autocorrelation among these points, data were rarefied using SDMtoolbox (Brown, 2014) so that no occurrence was within 40 km of another, leaving 273 occurrence points for analysis. We used the Average Nearest Neighbor tool in ArcGIS version 10.2 (ESRI, Redlands, California, USA) to determine the significance of the spatial clustering of occurrence data. Increasing the minimum distance between samples to 40 km reduced spatial autocorrelation enough that the data were significantly dispersed ($z = 5.880$, $P < 0.0001$).

Five climatic variables were chosen from 19 bioclimatic variables (http://www.worldclim.org) of present (1950–2000; Hijmans et al., 2005) and Last Glacial Maximum (LGM, ~20c000 ybp) climates (Braconnot et al., 2007) with a resolution of 2.5 arc-minutes: maximum temperature of the warmest month, minimum temperature of the coldest month, precipitation of the wettest month, precipitation of the driest month, and precipitation of the warmest quarter. These variables were selected to encompass those likely to be most biologically important to *C. americanum* given its growing season and susceptibility to drought (Galloway, 2002; Prendeville et al., 2015), followed by removal of variables with a correlation coefficient >0.8 (Sheppard, 2013).

Climate layers were projected using an Albers equal-area projection to train the model to prevent latitudinal bias in pseudoabsence point selection (Brown, 2014). To prevent overfitting due to selection of pseudoabsence points outside the range of equilibrium (Anderson and Raza, 2010), the area where the model trained was restricted to a minimum convex polygon with a buffer of 300 km outside the outermost points. A regularization multiplier of 2 was selected because it created a model with the highest AUC (area under the curve) and lowest omission rate (Warren and Seifert, 2011). Maxent uses regularization multipliers to determine how closely to fit the model to the data; values too low may lead to an overfitted model, wheras higher values give smoother models. Tenfold cross-validation was used to create the model by reserving a random 10% of the occurrence points for testing in turn for 10 models, then averaging those 10 models together. The AUC for this model was in the low range of acceptable values (Swets, 1988; Araujo et al., 2005). However, low AUC is not uncommon for widespread species (Lobo et al., 2008). In addition, AUC may not be the best way of identifying a good model (Lobo et al., 2008), so we also used binomial omission tests to analyze our model. Binomial omission tests first use a threshold for suitability to convert a continuous niche model into a binomial model that identifies an area as "present" or "absent." Omission rates are then calculated as how many known presences would be predicted as absences in a model. Finally, a binomial test is used to evaluate the omission rates and determine whether the model significantly predicts the range of the species.

Finally, a second model was created to verify the main model using spatial jackknifing to reduce spatial bias (Boria et al., 2014; Radosavljevic and Anderson, 2014). In this method, the study area was divided into five spatially independent folds, and five models were created by deleting each fold in turn, then averaging the models together (Brown, 2014).

***Climate analysis*—**To examine whether genetic lineages of *C. americanum* occur in different ecological niches, data for the five bioclimatic variables were extracted for the 49 populations used for chloroplast sequencing. These bioclimatic variables were run through principal component analysis (PROC PRINCOMP in SAS version 9.4; SAS Institute, Cary, North Carolina, USA), and the populations were graphed according to the first and second principal component. Because the second principal component (PC2) appeared to separate Appalachian and non-Appalachian populations, analysis of variance was used to determine whether the chloroplast and/or nuclear clade explained significant variation in PC2 (PROC GLM in SAS).

RESULTS

***Chloroplast sequencing and phylogeny*—**Individual chloroplast loci distinguished between three and nine haplotypes (Appendix S1). Concatenation of the five loci resulted in 2437 bp of sequence with 33 SNPs distinguishing 14 haplotypes. There were no indels

between haplotypes. The maximum likelihood trees generated for the individual loci and the concatenated sequence did not contradict each other. Therefore, only the concatenated tree is presented. Sequences for each locus were deposited in GenBank (accession nos. KP053958–KP054027).

Both the maximum likelihood and Bayesian phylogenetic analyses, as well as the statistical parsimony network, found support for three chloroplast clades (Fig. 1B, C). Percent sequence divergence ranged from 0.04 to 0.25 within clades, from 0.74 to 0.86 between the Appalachian and Eastern clades, from 0.21 to 0.41 between the Eastern and Western clades, and from 0.95 to 1.15 between the Appalachian and Western clades. An estimate of chloroplast synonymous substitution rate (Wolfe et al., 1987) places the divergence of the Appalachian clade to 2.3–7.0 mya, the Eastern and Western clade to 0.7–2.3 mya, and the divergence of the Western clade branches to 0.5–1.6 mya. These dates place the divergence of the Appalachian clade as likely occurring pre-Pleistocene, while the other divergences appear to have occurred during the Pleistocene.

The clades are geographically structured, with the highly divergent Appalachian clade (haplotypes A–D) restricted to the Appalachian Mountains. Within the Appalachian clade, the populations in the southern Appalachians (haplotypes A and B) are differentiated from the more northern populations (haplotypes C and D). The Eastern clade (haplotypes E and F) contains populations in the eastern Appalachians along with a coastal disjunct population in southeastern Virginia, whereas the Western clade (haplotypes G–N) shows a much wider distribution with populations throughout the range, including the southern Appalachians. The Western clade also shows some substructuring, with two branches occurring either east or proximate to and west of the Mississippi. The remaining haplotype (I) has a disjunct distribution, occurring in the most southern part of the range (Mississippi, Alabama, and Florida) and in the southern Appalachians.

Of the 13 populations where additional individuals were sequenced, 10 were monomorphic (Appendix S2). Two of the polymorphic populations had private haplotypes. Only one population had haplotypes shared by multiple populations (MS55, haplotypes I and J), and both of these haplotypes fall in the Western chloroplast clade. Therefore, sampling additional individuals from populations does not seem likely to change the overall phylogeographic pattern.

***Nuclear sequencing and phylogeny*—**RAD sequencing produced 5–9 million reads per population. Reads were deposited in NCBI's Short Read Archive (see Table 1 for accessions). Approximately 40% of reads from each population mapped to the reference genome, thereby providing an average depth of coverage of 30× across loci. A total of 20 938 SNPs were used to construct the maximum likelihood tree.

The nuclear tree produced similar results to the chloroplast tree in that it also found support for three clades, which are similar in pattern (Fig. 2). The Appalachian chloroplast clade that was restricted to the Appalachians is recovered in the nuclear tree (haplotypes A–D), though the deep divergence between this clade and the remaining clades is no longer apparent. In addition, similar substructuring is seen between populations in the southern (haplotypes A and B) and northern Appalachians (haplotype D). However, bootstrap support for the southern Appalachian populations grouping with the northern Appalachian populations is relatively low compared with the chloroplast tree, which suggests that some of the nuclear data supports an alternative grouping for these southern populations.

The Eastern chloroplast clade (haplotypes E and F) found in the eastern Appalachians and the Atlantic coast was also recovered, with stronger support, in the nuclear tree. Within this clade, there is support for the coastal disjunct population, VA93, being basal to the remaining populations. However, the Eastern clade now groups with the Western chloroplast clade populations that occur in the southern Appalachians (haplotypes I and J; hereafter called "the Smoky populations," after the Great Smoky Mountains in which several of these populations occur).

The third clade in the nuclear tree contains all populations west of the Appalachians and corresponds well with the Western chloroplast clade, with the exception of the Smoky populations discussed above. The nuclear tree again shows support for the populations west of the Mississippi grouping together within a larger, well-supported clade that includes several populations east of the Mississippi. The remaining populations, which are primarily southern, are difficult to resolve in terms of their exact placement on the tree (as reflected by low bootstrap values), though there was support for AL29, TN34, and PA27 grouping together. Bootstrap values can be inflated in large-scale datasets. However, the alternative de novo analysis of the RAD-seq data produced a similar topology (see Supplemental Data with the online version of this article; Appendix S3), providing further confirmation of these patterns.

Taken together, the chloroplast and nuclear phylogenies indicate the existence of three genetically divergent lineages within *C. americanum*. In addition, there are the Smoky populations, which cluster geographically and do not fall clearly into any one lineage because their placement differs in the chloroplast and nuclear trees. Along with the Smoky populations, two of the three lineages, including the highly divergent Appalachian lineage, are restricted primarily to the Appalachians, whereas the Western lineage occurs throughout the range outside of the Appalachians. This distribution results in populations located within the Appalachian Mountains being both genetically divergent and more genetically diverse than the rest of the species range.

***Isolation-by-distance*—**When including all populations, significant isolation-by-distance was found for the nuclear RAD data (Mantel's $r = 0.35$, $P = 0.002$), but not for the chloroplast data (Mantel's $r = 0.03$, $P = 0.286$). However, once the Appalachian lineage was removed from the analysis, the chloroplast data exhibited significant isolation-by-distance (Mantel's $r = 0.27$, $P < 0.001$). Stronger isolation-by-distance was found when analyses were conducted separately for the Appalachian (chloroplast: Mantel's $r = 0.67$, $P = 0.015$; nuclear: Mantel's $r = 0.60$, $P = 0.018$) and Western clades (chloroplast: Mantel's $r = 0.37$, $P < 0.001$; nuclear: Mantel's $r = 0.45$, $P = 0.007$). No significant isolation-by-distance was found within the Eastern clade or the Smoky populations, possibly because of small sample sizes (four and five populations, respectively). These results suggest that isolation-by-distance is influencing the genetic structure of *C. americanum*.

***Ecological niche modeling*—**The suitable range predicted by our ecological niche model corresponds well to the known current range of *C. americanum* (Fig. 3A), except for the northeast, where *C. americanum* does not occur. This discrepancy may be due to ongoing recolonization in this area or to anthropological effects not taken into account by a model built from climate data. The predicted

highly suitable areas in the western part of the range overlap with areas of apparent recent recolonization by haplotypes G, H, and L, whereas the unsuitable area in the Deep South corresponds to a part of the range where *C. americanum* is less common, with only sparsely distributed populations. The average testing AUC of the tenfold cross-validated model was 0.733. The binomial omission test found that our model's predictions were significant ($P < 0.005$) using multiple thresholds, including thresholds that minimize the difference between sensitivity and specificity, thresholds that maximize the sum of sensitivity and specificity, and a 10th-percentile training presence threshold. These thresholds have been shown to be useful measures of model performance (Jiménez-Valverde and Lobo, 2007).

When the model was projected onto LGM climate data (Fig. 3B), potential glacial refugia were apparent in the designation of suitable habitat along the present-day Gulf and Atlantic coasts, as well as in the southern Appalachians. Spatial jackknifing created a model with AUC of 0.747 that is qualitatively similar to the tenfold cross-validated model, although with less of the southernmost part of the range predicted to be suitable during the LGM (see Supplemental Data with the online version of this article; Appendix S4).

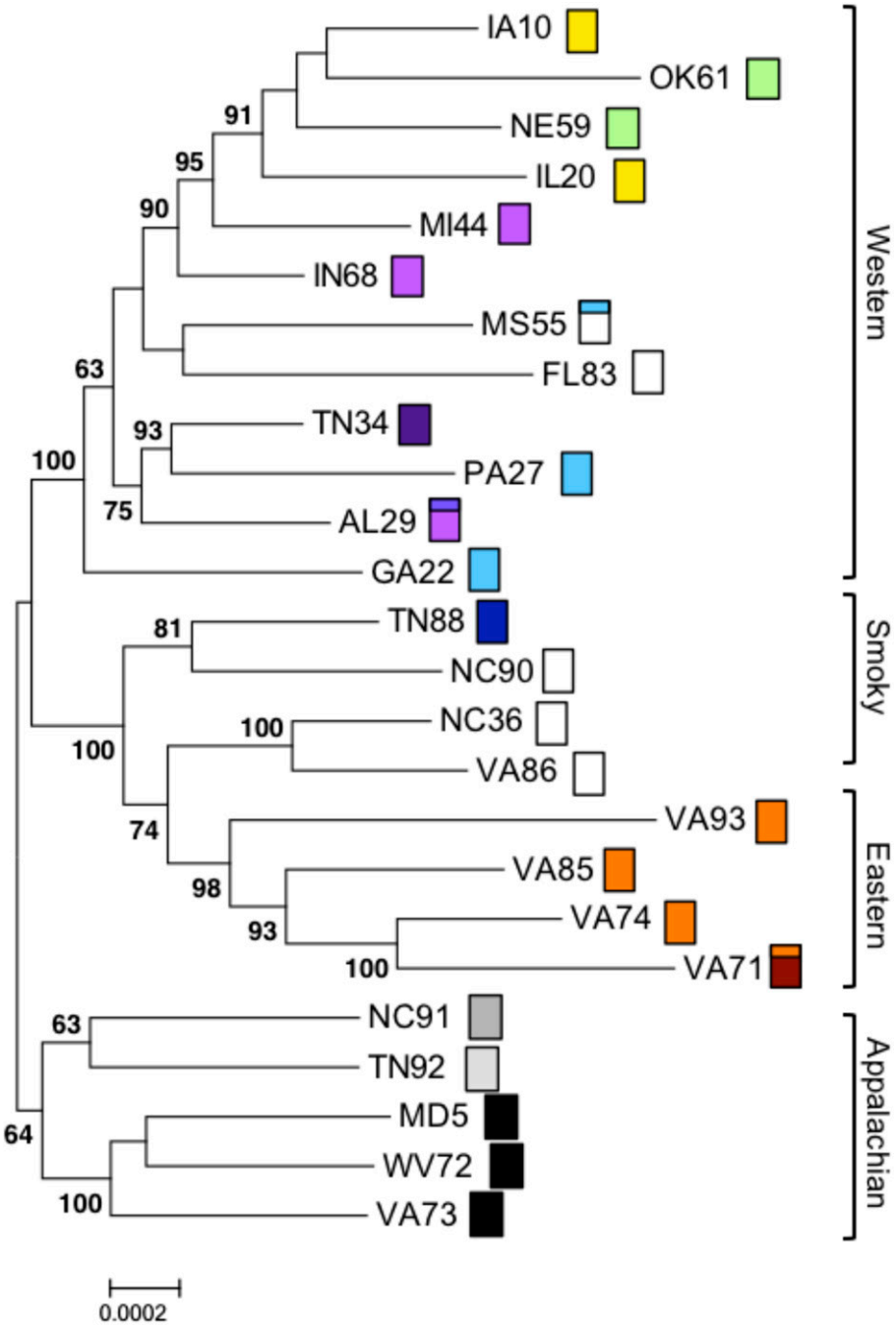

FIGURE 2 Maximum likelihood tree for *Campanulastrum americanum* nuclear data from reference-based Stacks analysis, rooted according to position of root in chloroplast tree. Numbers show bootstrap support values >60. Population IDs (numbers on left) refer to a population's state and ID (as given in Table 1). Colors indicate chloroplast haplotype as in Figure 1.

***Climate analysis*—**Principal component analysis of bioclimatic variables for all 49 populations resulted in first and second principal components that explain 55.06% and 29.65% of the total variance, respectively (for bioclimatic variable loadings, see Supplemental Data with the online version of this article; Appendix S5). Most of the variation in PC1 is found among populations of the Western nuclear clade, while PC2 separates out the Appalachian populations (Appalachian and Eastern nuclear clades), including the Smoky populations (haplotypes I and J) that fall out in the Western chloroplast clade, but Appalachian nuclear clade (see Supplemental Data with the online version of this article; Appendix S6). Both nuclear and chloroplast clades explained variation in the values of PC2 ($P < 0.01$ and $P = 0.04$, respectively). However, more of the variation was explained by nuclear clade ($R^2 = 0.52$ vs. $R^2 = 0.13$), which better delineates Appalachian and non-Appalachian populations. These results suggest that Appalachian populations occupy a different climatic niche than non-Appalachian populations. Altogether, the Smoky I and J populations appear to group closely with the other Appalachian populations in terms of climate and nuclear genetics, though they fall out with the non-Appalachian populations in terms of chloroplast haplotype.

DISCUSSION

Both the phylogeography and ecological niche modeling support the presence of multiple glacial refugia in *C. americanum*, leading to several geographically structured, genetically divergent lineages. The distributions of these lineages—including a Mississippi discontinuity, an east–west Appalachian discontinuity, and an Appalachian refugium—all match

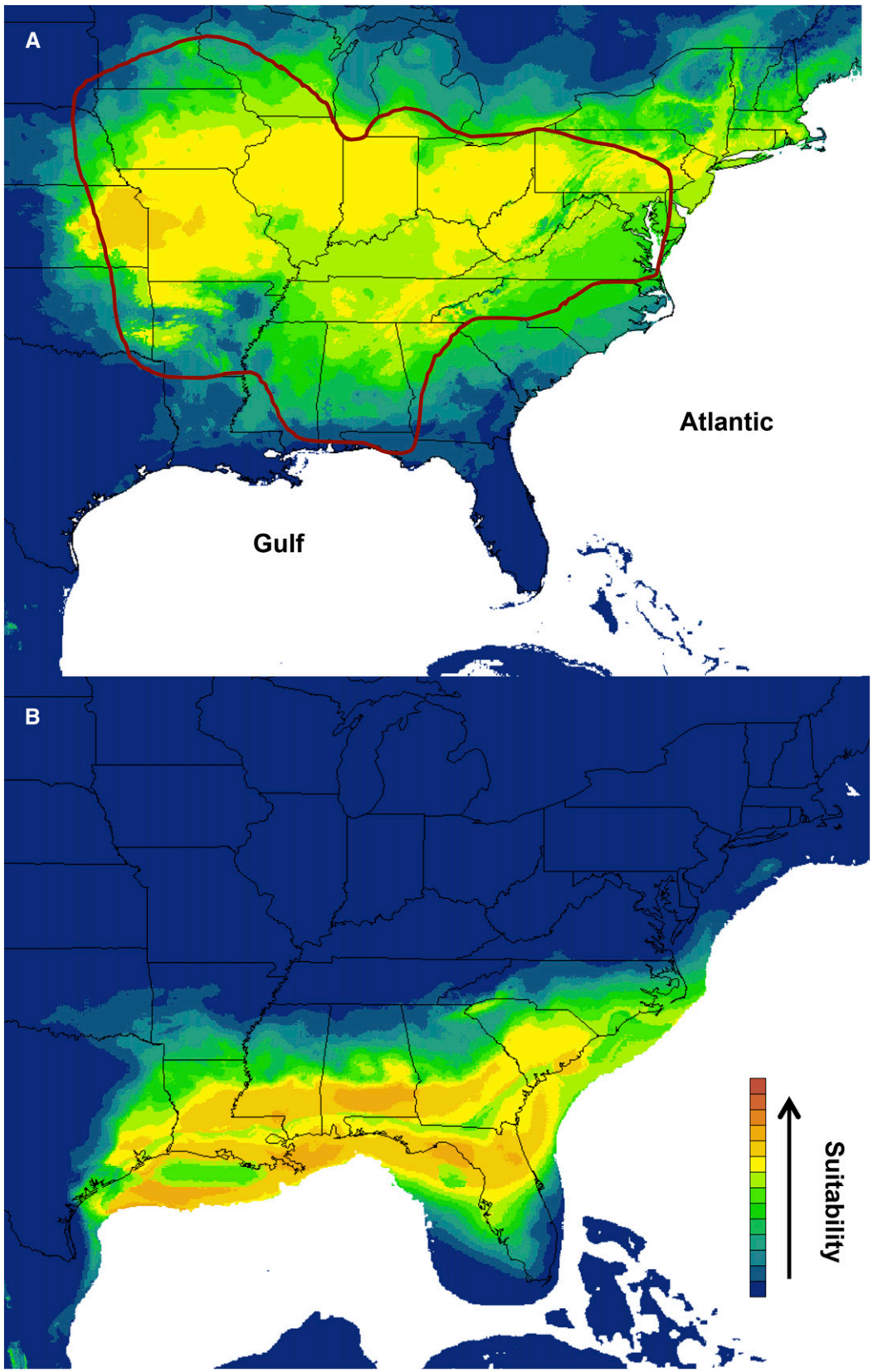

FIGURE 3 Ecological niche model for *Campanulastrum americanum* showing predicted habitat suitability in eastern North America during (A) present day and (B) the Last Glacial Maximum (~20 000 ybp). Log suitability is expressed between 0 and 1, and higher values (represented here by warmer colors) indicate an increased probability of finding *C. americanum*. Dark red line indicates an approximation of the species range.

recurrent patterns found in previous studies of both plants and animals in eastern North America (Soltis et al., 2006). However, rarely are all of these patterns found within a single species. The relatively short generation time (annual–biennial) of *C. americanum* could underlie its greater degree of genetic divergence and geographic structuring compared with the woody and long-lived herbs previously studied in eastern North America (Griffin and Barrett, 2004; McLachlan et al., 2005; Gonzales et al., 2008; Morris et al., 2008; Li et al., 2013). Greater rates of molecular evolution have been found in herbaceous species compared with trees and shrubs (Smith and Donoghue, 2008) and in annuals compared with perennials (Soria-Hernanz et al., 2008; Yue et al., 2010). In addition, *C. americanum*'s patchy population structure and widespread distribution could also contribute to its substantial genetic divergence and geographic structuring. More studies on species with similar life histories are needed to determine whether short generation times and/or patchy population structure generally lead to greater phylogeographic structure.

Given the low mutation rate of the chloroplast genome (Wolfe et al., 1987), the majority of observed chloroplast haplotypes likely predate the most recent postglacial recolonization, allowing inference of recolonization routes from the current geographic distribution of chloroplast haplotypes (McLachlan et al., 2005). Much of the current range of *C. americanum* is occupied by the Western clade, with three haplotypes (G, H, and L) occurring in areas of widespread homogeneity suggestive of recent, rapid recolonization from southern glacial refugia (Hewitt, 1996). Because these areas of widespread homogeneity also overlap with much of the highly suitable habitat designated by the current-climate ecological niche model, rapid recolonization probably occurred as suitable habitat shifted northward after the LGM. The location of the two Western-chloroplast-clade branches east (haplotype L) and west (haplotypes G and H) of the Mississippi may indicate separate recolonization routes, possibly from separate refugial areas (R1 and R2; Fig. 4). The ecological niche model designates suitable habitat during the LGM along the present-day Gulf Coast, providing support for southern refugia as seen in many other species (Avise, 2000; Soltis et al., 2006). A similar east–west Mississippi discontinuity has been observed in other species and is a recurrent pattern in the phylogeography of eastern North America (Soltis et al., 2006; Jaramillo-Correa et al., 2009).

There are also Deep South populations that remain in the area of the hypothesized southern refugia and contain a distinct Western-clade haplotype (haplotype I). Currently, *C. americanum* is only sparsely distributed in the Deep South, and populations in this area may now be relatively isolated. These populations may also be differently adapted, given that they occur in a unique environment (Appendix S6) designated as unsuitable habitat according to the current ecological niche model and are phenotypically different from other populations (Prendeville et al., 2013, Barnard-Kubow personal observation). Other plant species also appear to have genetic lineages mostly restricted to Florida or the Deep South, including tulip poplar (Parks et al., 1994) and red maple (McLachlan et al., 2005).

In addition to southern refugia, there is evidence for *C. americanum* having survived the LGM in a refugium located on the Atlantic coast, as suggested by the location of the Eastern chloroplast clade (haplotypes E and F) in the eastern Appalachians and on the Virginia coast. The ecological niche model and the basal position of the Virginia coastal population (VA93) in the nuclear tree in relation to the other Eastern-clade populations provide support for a coastal refugium for this clade (R3; Fig. 4). Coastal refugia along the Atlantic coast have been proposed for other species, including tiger salamander (Church et al., 2003), American sweetgum (Morris et al., 2008), and eastern white pine (Nadeau et al., 2015). The coastal plain in the Carolinas has also been considered a likely location for glacial refugia because of the high endemism of plant species there (Sorrie and Weakley, 2001). Overall, this clade appears to have contributed much less to recolonization than the Western chloroplast and nuclear clade, likely because of the barrier imposed by the Appalachians.

Restriction of the Appalachian clade to the Appalachian Mountains suggests that these populations survived the LGM in an Appalachian glacial refugium. Furthermore, the deep genetic divergence in the chloroplast tree indicates that this clade has likely remained in the Appalachians through many glacial cycles. The divergence of this clade was estimated to have occurred 2.3–7 mya, which indicates that the divergence of the Appalachian lineage likely predates the Pleistocene (2.5–0.01 mya). Appalachian refugia have been proposed for other taxa, including plants (McLachlan et al., 2005; Gonzales et al., 2008) and animals (Brant and Orti, 2003; Church et al., 2003; Austin et al., 2004; Lee-Yaw et al., 2008). The ecological niche model finds a small amount of moderately suitable habitat in the southern Appalachians during the LGM, providing some support for an Appalachian refugium (R4; Fig. 4). The lack of higher-suitability habitat in the southern Appalachians may be due to the resolution of the climatic datasets generally not being fine enough to detect variation due to microclimates, particularly when modeling past climates (Gavin et al., 2014). Mountain ranges are likely to have fine-scale climatic variation due to topography, raising the possibility of microrefugia, such as sheltered valleys or south-facing slopes, that could be difficult to detect via niche modeling.

Overall there is concordance between the nuclear and chloroplast phylogenies, indicating little gene flow or hybridization between clades. However, there does appear to be less divergence among clades when looking at the nuclear versus chloroplast markers. This difference may be due to the lower effective population size of the chloroplast in relation to the nuclear genes, which can lead to increased drift and stronger population differentiation (Birky et al., 1989; Levy and Neal, 1999). *Campanulastrum americanum* has increased rates of nucleotide substitutions in a subset of chloroplast genes (Barnard-Kubow et al., 2014), which could contribute to greater divergence in the chloroplast tree. However, the deep divergence between the Mountain and Western–Eastern chloroplast clades is found in all five chloroplast loci, including the noncoding regions, indicating that this divergence is likely not driven by increased substitution rates in chloroplast genes. A similar depth of divergence between genetic lineages has not been found in other phylogeographic studies of plant species in eastern North American (Griffin and Barrett, 2004; McLachlan et al., 2005; Gonzales et al., 2008; Morris et al., 2008; Li et al., 2013). As discussed above, this depth of divergence may be due to *C. americanum*'s short generation time.

Although the chloroplast and nuclear phylogenies give similar results overall, there are some differences, particularly in regard to the Western-chloroplast-clade haplotypes I and J. The Smoky I and J populations group with the Eastern-chloroplast-clade populations in the nuclear tree and occur in a similar environment to the other Appalachian populations, distinct from the remaining Western-clade populations (Appendix S6). These Smoky populations

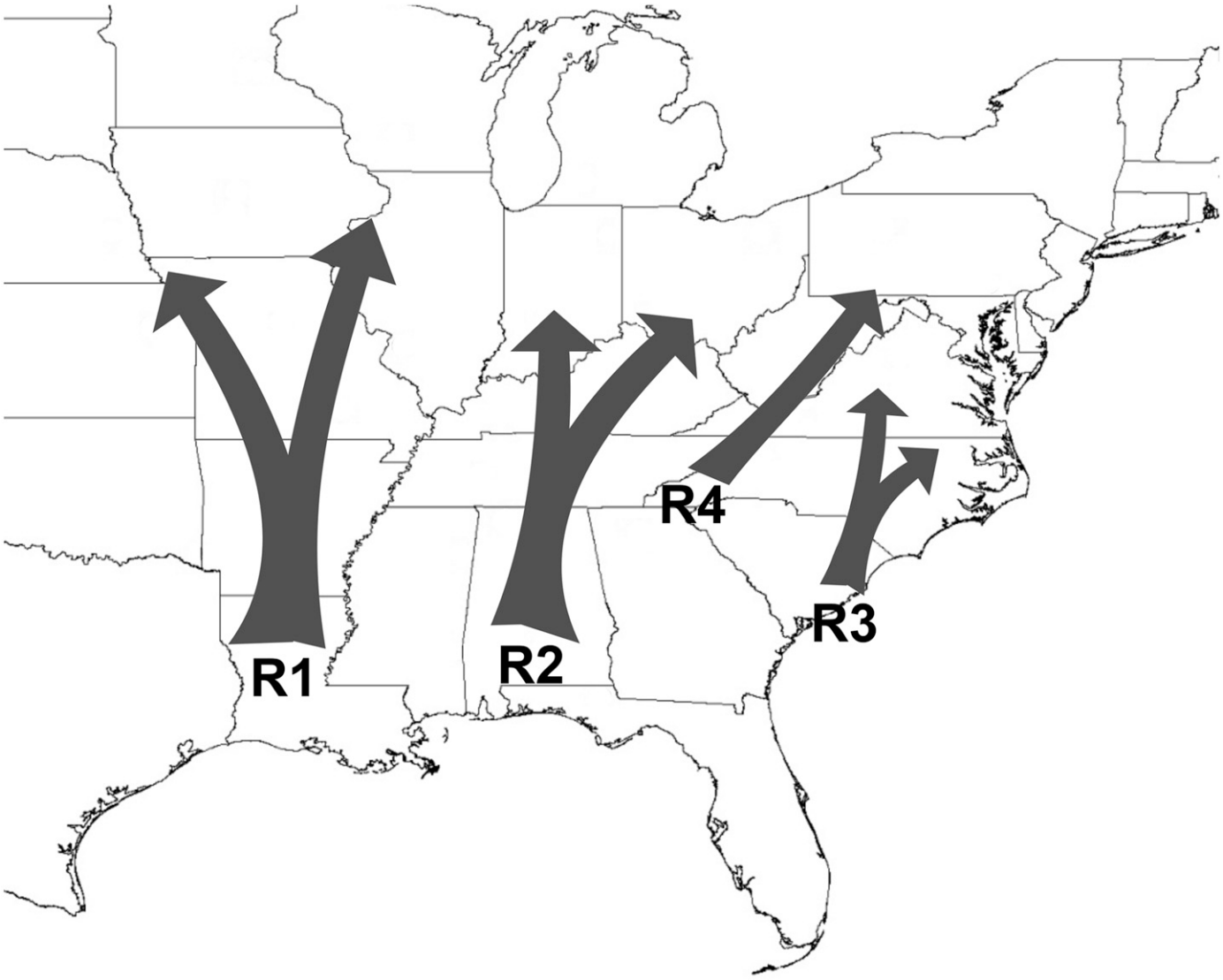

FIGURE 4 Map showing hypothetical locations of glacial refugia in the eastern United States for *Campanulastrum americanum*, as well as potential recolonization routes after the Last Glacial Maximum (for details, see text).

appear to have an evolutionary history distinct from the other Western- and Eastern-clade populations. Perhaps they originated from one of the same glacial refugia as the other Western-clade populations but have now experienced secondary contact and gene flow with other Appalachian populations, or they may have survived and diverged in a separate glacial refugium altogether.

The extensive genetic structuring of a species, as observed here, has the potential to contribute to speciation by facilitating the evolution of genetic incompatibility and reproductive isolation (Hewitt, 1996; Carstens and Knowles, 2007). Is there any evidence for this in *C. americanum*? Strong reproductive isolation, manifested as reduced germination and survival, is found between populations of *C. americanum*, in particular when crossing Western-clade populations with populations from either the Appalachian or Eastern clades (Galloway and Etterson, 2005; Etterson et al., 2007; Barnard-Kubow, 2015). Moderate reproductive isolation is also found when crossing between the Appalachian and Eastern clades (Barnard-Kubow, 2015). By contrast, crossing populations within any of the three lineages results in little or no reproductive isolation (Galloway and Etterson, 2005; Etterson et al., 2007; Barnard-Kubow, 2015). This pattern suggests that reproductive isolation in *C. americanum* occurs between genetically divergent lineages that recolonized from different glacial refugia. Accordingly, the historical processes of climate fluctuations and glacial cycles appear to have facilitated the early stages of the speciation process in *C. americanum*, in that phylogeographic lineages appear to be incipient species. Our results are consistent with other studies that have found varying degrees of reproductive incompatibility among phylogeographic lineages (Gómez et al., 2007; Pinheiro et al., 2013; Singhal and Moritz, 2013), suggesting a widespread role for climatic shifts in the speciation process.

ACKNOWLEDGEMENTS

We thank the many people who assisted us by collecting seed from populations of *C. americanum*. We also thank D. Triant for writing the Perl script used in concatenating the FASTA sequence data;

A. Nguyen for constructing the RAD-seq library; M. Sebasky for discussions and advice on ecological niche modeling; J. Antonovics, B. Blackman, E. Brodie III, D. Carr, and D. Taylor for guidance throughout this project; and the reviewers who provided constructive feedback on the manuscript. This work was supported by DEB-0922214 and the ARCS Foundation.

LITERATURE CITED

Anderson, R. P., and A. Raza. 2010. The effect of the extent of the study region on GIS models of species geographic distributions and estimates of niche evolution: Preliminary tests with montane rodents (genus *Nephelomys*) in Venezuela. *Journal of Biogeography* 37: 1378–1393.

April, J., R. H. Hanner, A. M. Dion-Cote, and L. Bernatchez. 2013. Glacial cycles as an allopatric speciation pump in north-eastern American freshwater fishes. *Molecular Ecology* 22: 409–422.

Araujo, M. B., R. G. Pearson, W. Thuiller, and M. Erhard. 2005. Validation of species-climate impact models under climate change. *Global Change Biology* 11: 1504–1513.

Austin, J. D., S. C. Lougheed, and P. T. Boag. 2004. Discordant temporal and geographic patterns in maternal lineages of eastern north American frogs, *Rana catesbeiana* (Ranidae) and *Pseudacris crucifer* (Hylidae). *Molecular Phylogenetics and Evolution* 32: 799–816.

Avise, J. C. 2000. Phylogeography: The history and formation of species. Harvard University Press, Cambridge, Massachusetts, USA.

Avise, J. C., J. Arnold, R. M. Ball, E. Bermingham, T. Lamb, J. E. Neigel, C. A. Reeb, and N. C. Saunders. 1987. Intraspecific phylogeography: The mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Ecology and Systematics* 18: 489–522.

Avise, J. C., D. Walker, and G. C. Johns. 1998. Speciation durations and Pleistocene effects on vertebrate phylogeography. *Proceedings. Biological Sciences* 265: 1707–1712.

Barnard-Kubow, K. B. 2015. Cytonuclear incompatibility contributes to incipient speciation. Master's thesis, University of Virginia, Charlottesville, Virginia, USA.

Barnard-Kubow, K. B., D. B. Sloan, and L. F. Galloway. 2014. Correlation between sequence divergence and polymorphism reveals similar evolutionary mechanisms acting across multiple timescales in a rapidly evolving plastid genome. *BMC Evolutionary Biology* 14: 268.

Birky, C. W., P. Fuerst, and T. Maruyama. 1989. Organelle gene diversity under migration, mutation, and drift—equilibrium expectations, approach to equilibrium, effects of heteroplasmic cells, and comparison to nuclear genes. *Genetics* 121: 613–627.

Boria, R. A., L. E. Olson, S. M. Goodman, and R. P. Anderson. 2014. Spatial filtering to reduce sampling bias can improve the performance of ecological niche models. *Ecological Modelling* 275: 73–77.

Braconnot, P., B. Otto-Bliesner, S. Harrison, S. Joussaume, J. Y. Peterchmitt, A. Abe-Ouchi, M. Crucifix, et al. 2007. Results of PMIP2 coupled simulations of the Mid-Holocene and Last Glacial Maximum—Part 1: Experiments and large-scale features. *Climate of the Past* 3: 261–277.

Brant, S. V., and G. Orti. 2003. Phylogeography of the Northern short-tailed shrew, *Blarina brevicauda* (Insectivora: Soricidae): Past fragmentation and postglacial recolonization. *Molecular Ecology* 12: 1435–1449.

Brown, J. L. 2014. SDMtoolbox: A python-based GIS toolkit for landscape genetic, biogeographic and species distribution model analyses. *Methods in Ecology and Evolution* 5: 694–700.

Carstens, B. C., and L. L. Knowles. 2007. Shifting distributions and speciation: Species divergence during rapid climate change. *Molecular Ecology* 16: 619–627.

Catchen, J., P. A. Hohenlohe, S. Bassham, A. Amores, and W. A. Cresko. 2013. Stacks: An analysis tool set for population genomics. *Molecular Ecology* 22: 3124–3140.

Catchen, J. M., A. Amores, P. Hohenlohe, W. Cresko, and J. H. Postlethwait. 2011. Stacks: Building and genotyping loci de novo from short-read sequences. *G3* 1: 171–182.

Church, S. A., J. M. Kraus, J. C. Mitchell, D. R. Church, and D. R. Taylor. 2003. Evidence for multiple pleistocene refugia in the postglacial expansion of the eastern tiger salamander, *Ambystoma tigrinum tigrinum*. *Evolution* 57: 372–383.

Clement, M., D. Posada, and K. A. Crandall. 2000. TCS: A computer program to estimate gene genealogies. *Molecular Ecology* 9: 1657–1659.

Darriba, D., G. L. Taboada, R. Doallo, and D. Posada. 2012. jModelTest 2: More models, new heuristics and parallel computing. *Nature Methods* 9: 772.

Davey, J. W., P. A. Hohenlohe, P. D. Etter, J. Q. Boone, J. M. Catchen, and M. L. Blaxter. 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics* 12: 499–510.

Davis, M. B. 1983. Quaternary history of deciduous forests of eastern North-America and Europe. *Annals of the Missouri Botanical Garden* 70: 550–563.

Eddie, W. M. M., T. Shulkina, J. Gaskin, R. C. Haberle, and R. K. Jansen. 2003. Phylogeny of Campanulaceae s. str. inferred from its sequences of nuclear ribosomal DNA. *Annals of the Missouri Botanical Garden* 90: 554–575.

Etterson, J. R., S. R. Keller, and L. F. Galloway. 2007. Epistatic and cytonuclear interactions govern outbreeding depression in the autotetraploid *Campanulastrum americanum*. *Evolution* 61: 2671–2683.

Excoffier, L., M. Foll, and R. J. Petit. 2009. Genetic consequences of range expansions. *Annual Review of Ecology Evolution and Systematics* 40: 481–501.

Galloway, L. F. 2002. The effect of maternal phenology on offspring characters in the herbaceous plant *Campanula americana*. *Journal of Ecology* 90: 851–858.

Galloway, L. F., and J. R. Etterson. 2005. Population differentiation and hybrid success in *Campanula americana*: Geography and genome size. *Journal of Evolutionary Biology* 18: 81–89.

Galloway, L. F., J. R. Etterson, and J. L. Hamrick. 2003. Outcrossing rate and inbreeding depression in the herbaceous autotetraploid, *Campanula americana*. *Heredity* 90: 308–315.

Gavin, D. G., M. C. Fitzpatrick, P. F. Gugger, K. D. Heath, F. Rodriguez-Sanchez, S. Z. Dobrowski, A. Hampe, et al. 2014. Climate refugia: Joint inference from fossil records, species distribution models and phylogeography. *New Phytologist* 204: 37–54.

Gómez, A., R. N. Hughes, P. J. Wright, G. R. Carvalho, and D. H. Lunt. 2007. Mitochondrial DNA phylogeography and mating compatibility reveal marked genetic structuring and speciation in the NE Atlantic bryozoan *Celleporella hyalina*. *Molecular Ecology* 16: 2173–2188.

Gonzales, E., J. L. Hamrick, and S. M. Chang. 2008. Identification of glacial refugia in south-eastern North America by phylogeographical analyses of a forest understorey plant, *Trillium cuneatum*. *Journal of Biogeography* 35: 844–852.

Griffin, S. R., and S. C. H. Barrett. 2004. Post-glacial history of *Trillium grandiflorum* (Melanthiaceae) in eastern North America: Inferences from phylogeography. *American Journal of Botany* 91: 465–473.

Gugger, P. F., M. Ikegami, and V. L. Sork. 2013. Influence of late Quaternary climate change on present patterns of genetic variation in valley oak, Quercus lobata Née. *Molecular Ecology* 22: 3598–3612.

Guindon, S., and O. Gascuel. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology* 52: 696–704.

Hall, T. A. 1999. BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* 41: 95–98.

Hamilton, M. B. 1999. Four primer pairs for the amplification of chloroplast intergenic regions with intraspecific variation. *Molecular Ecology* 8: 521–523.

Hewitt, G. M. 1996. Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society* 58: 247–276.

Hewitt, G. M. 2000. The genetic legacy of the Quaternary ice ages. *Nature* 405: 907–913.

Hewitt, G. M. 2004. Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society of London, Series B* 359: 183–195.

Hijmans, R. J., S. E. Cameron, J. L. Parra, P. G. Jones, and A. Jarvis. 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* 25: 1965–1978.

Jaramillo-Correa, J. P., J. Beaulieu, D. P. Khasa, and J. Bousquet. 2009. Inferring the past from the present phylogeographic structure of North American forest trees: Seeing the forest for the genes. *Canadian Journal of Forest Research* 39: 286–307.

Jensen, J. L., A. J. Bohonak, and S. T. Kelley. 2005. Isolation by distance, web service. *BMC Genetics* 6: 13.

Jiménez-Valverde, A., and J. M. Lobo. 2007. Threshold criteria for conversion of probability of species presence to either–or presence–absence. *Acta Oecologica* 31: 361–369.

Langmead, B., and S. L. Salzberg. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods* 9: 357–359.

Lee-Yaw, J. A., J. T. Irwin, and D. M. Green. 2008. Postglacial range expansion from northern refugia by the wood frog, *Rana sylvatica. Molecular Ecology* 17: 867–884.

Levy, F., and C. L. Neal. 1999. Spatial and temporal genetic structure in chloroplast and allozyme markers in *Phacelia dubia* implicate genetic drift. *Heredity* 82: 422–431.

Li, P., M. M. Li, Y. Shi, Y. P. Zhao, Y. Wan, C. X. Fu, and K. M. Cameron. 2013. Phylogeography of North American herbaceous *Smilax* (Smilacaceae): Combined AFLP and cpDNA data support a northern refugium in the Driftless Area. *American Journal of Botany* 100: 801–814.

Lobo, J. M., A. Jiménez-Valverde, and R. Real. 2008. AUC: A misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography* 17: 145–151.

McLachlan, J. S., J. S. Clark, and P. S. Manos. 2005. Molecular indicators of tree migration capacity under rapid climate change. *Ecology* 86: 2088–2098.

Morris, A. B., S. M. Ickert-Bond, D. B. Brunson, D. E. Soltis, and P. S. Soltis. 2008. Phylogeographical structure and temporal complexity in American sweetgum (*Liquidambar styraciflua*; Altingiaceae). *Molecular Ecology* 17: 3889–3900.

Nadeau, S., J. Godbout, M. Lamothe, M. C. Gros-Louis, N. Isabel, and K. Ritland. 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. *American Journal of Botany* 102: 1342–1355.

Parducci, L., T. Jorgensen, M. M. Tollefsrud, E. Elverland, T. Alm, S. L. Fontana, K. D. Bennett, et al. 2012. Glacial survival of boreal trees in northern Scandinavia. *Science* 335: 1083–1086.

Parks, C. R., J. F. Wendel, M. M. Sewell, and Y. L. Qiu. 1994. The significance of allozyme variation and introgression in the *Liriodendron tulipifera* complex (Magnoliaceae). *American Journal of Botany* 81: 878–889.

Phillips, S. J., R. P. Anderson, and R. E. Schapire. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling* 190: 231–259.

Pinheiro, F., S. Cozzolino, F. de Barros, T. M. Z. M. Gouveia, R. M. Suzuki, M. F. Fay, and C. Palma-Silva. 2013. Phylogeographic structure and outbreeding depression reveal early stages of reproductive isolation in the Neotropical orchid *Epidendrum denticulatum. Evolution* 67: 2024–2039.

Prendeville, H. R., K. Barnard-Kubow, C. Dai, B. C. Barringer, and L. F. Galloway. 2013. Clinal variation for only some phenological traits across a species range. *Oecologia* 173: 421–430.

Prendeville, H. R., J. C. Steven, and L. F. Galloway. 2015. Spatiotemporal variation in deer browse and tolerance in a woodland herb. *Ecology* 96: 471–478.

Provan, J., and K. D. Bennett. 2008. Phylogeographic insights into cryptic glacial refugia. *Trends in Ecology & Evolution* 23: 564–571.

Radosavljevic, A., and R. P. Anderson. 2014. Making better MAXENT models of species distributions: Complexity, overfitting and evaluation. *Journal of Biogeography* 41: 629–643.

Roberts, D. R., and A. Hamann. 2015. Glacial refugia and modern genetic diversity of 22 western North American tree species. *Proceedings of the Royal Society of London, Series B* 282: 20142903.

Ronquist, F., M. Teslenko, P. van der Mark, D. L. Ayres, A. Darling, S. Höhna, B. Larget, et al. 2012. MrBayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology* 61: 539–542.

Sheppard, C. S. 2013. How does selection of climate variables affect predictions of species distributions? A case study of three new weeds in New Zealand. *Weed Research* 53: 259–268.

Singhal, S., and C. Moritz. 2013. Reproductive isolation between phylogeographic lineages scales with divergence. *Proceedings of the Royal Society of London, Series B* 280: 20132446.

Smith, S. A., and M. J. Donoghue. 2008. Rates of molecular evolution are linked to life history in flowering plants. *Science* 322: 86–89.

Soltis, D. E., A. B. Morris, J. S. McLachlan, P. S. Manos, and P. S. Soltis. 2006. Comparative phylogeography of unglaciated eastern North America. *Molecular Ecology* 15: 4261–4293.

Soria-Hernanz, D. F., O. Fiz-Palacios, J. M. Braverman, and M. B. Hamilton. 2008. Reconsidering the generation time hypothesis based on nuclear ribosomal ITS sequence comparisons in annual and perennial angiosperms. *BMC Evolutionary Biology* 8: 344.

Sorrie, B. A., and A. S. Weakley. 2001. Coastal plain vascular plant endemics: phytogeographic patterns. *Castanea* 66: 50–82.

Stamatakis, A. 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30: 1312–1313.

Stewart, J. R., and A. M. Lister. 2001. Cryptic northern refugia and the origins of the modern biota. *Trends in Ecology & Evolution* 16: 608–613.

Suarez-Gonzalez, A., J. T. Sutton, A. J. Trant, E. Zamlynny, and S. V. Good. 2015. Rethinking refugia: Tree topology, divergence dates, and demographic history trace the distribution of the endangered Plymouth Gentian (*Sabatia kennedyana*) from the Pleistocene glaciation to present day. *American Journal of Botany* 102: 609–620.

Swets, J. A. 1988. Measuring the accuracy of diagnostic systems. *Science* 240: 1285–1293.

Taberlet, P., L. Fumagalli, A.-G. Wust-Saucy, and J.-F. Cosson. 1998. Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology* 7: 453–464.

Taberlet, P., L. Gielly, G. Pautou, and J. Bouvet. 1991. Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Molecular Biology* 17: 1105–1109.

Tamura, K., D. Peterson, N. Peterson, G. Stecher, M. Nei, and S. Kumar. 2011. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution* 28: 2731–2739.

Tollefsrud, M. M., R. Kissling, F. Gugerli, O. Johnsen, T. Skroppa, R. Cheddadi, W. O. van Der Knaap, et al. 2008. Genetic consequences of glacial survival and postglacial colonization in Norway spruce: Combined analysis of mitochondrial DNA and fossil pollen. *Molecular Ecology* 17: 4134–4150.

Tribsch, A., and P. Schonswetter. 2003. Patterns of endemism and comparative phylogeography confirm palaeoenvironmental evidence for Pleistocene refugia in the Eastern Alps. *Taxon* 52: 477–497.

Warren, D. L., and S. N. Seifert. 2011. Ecological niche modeling in Maxent: The importance of model complexity and the performance of model selection criteria. *Ecological Applications* 21: 335–342.

Wendling, B. M., K. E. Galbreath, and E. G. DeChaine. 2011. Resolving the evolutionary history of *Campanula* (Campanulaceae) in western North America. *PLoS ONE* 6: e23559.

Wolfe, K. H., W. H. Li, and P. M. Sharp. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proceedings of the National Academy of Sciences, USA* 84: 9054–9058.

Yue, J. X., J. P. Li, D. Wang, H. Araki, D. C. Tian, and S. H. Yang. 2010. Genome-wide investigation reveals high evolutionary rates in annual model plants. *BMC Plant Biology* 10: 242.