Journal of Biogeography (*J. Biogeogr.*) (2008) **35**, 844–852

ORIGINAL ARTICLE

Identification of glacial refugia in south-eastern North America by phylogeographical analyses of a forest understorey plant, *Trillium cuneatum*

Eva Gonzales[1,2]*, James L. Hamrick[2] and Shu-Mei Chang[2]

[1]*Department of Ecology, Evolution and Natural Resources, School of Biological and Environmental Science, Rutgers University, New Brunswick, NJ 08901, USA and* [2]*Plant Biology Department, University of Georgia, Athens, GA 30602, USA*

*Correspondence: Eva Gonzales, Department of Ecology, Evolution and Natural Resources, Room 152, ENRS Building, Rutgers University, 14 College Farm Road, New Brunswick, NJ 08901-8551, USA.
E-mail: gonzales@aesop.rutgers.edu

ABSTRACT

Aim We examine several hypotheses emerging from biogeographical and fossil records regarding glacial refugia of a southern thermophilic plant species. Specifically, we investigated the glacial history and post-glacial colonization of a forest understorey species, *Trillium cuneatum*. We focused on the following questions: (1) Did *T. cuneatum* survive the Last Glacial Maximum (LGM) in multiple refugia, and (if so) where were they located, and is the modern genetic structure congruent with the fossil record-based reconstruction of refugia for mesic deciduous forests? (2) What are the post-glacial colonization patterns in the present geographical range?

Location South-eastern North America.

Methods We sampled 45 populations of *T. cuneatum* throughout its current range. We conducted phylogeographical analyses based on maternally inherited chloroplast DNA (cpDNA haplotypes) and used TCS software to reconstruct intraspecific phylogeny.

Results We detected six cpDNA haplotypes, geographically highly structured into non-overlapping areas. With one exception, none of the populations had mixed haplotype composition. TCS analysis resulted in two intraspecific cpDNA lineages, with one clade subdivided further by shallower diversification.

Main conclusions Our investigation revealed that *T. cuneatum* survived the LGM in multiple refugia, belonging to two (western, eastern) genealogical lineages geographically structured across south-eastern North America. The western clade is confined to the south-western corner of *T. cuneatum*'s modern range along the Lower Mississippi Valley, where fossil records document a major refugium of mesic deciduous forest. For the eastern clade, modern patterns of cpDNA haplotype distribution suggest cryptic vicariance, in the form of forest contractions and subsequent expansions associated with Pleistocene glacial cycles, rather than simple southern survival and subsequent northward colonization. The north–south partitioning of cpDNA haplotypes was unexpected, suggesting that populations of this rather southern thermophilic species may have survived in more northern locations than initially expected based on LGM climate reconstruction, and that the Appalachian Mountains functioned as a barrier to the dispersal of propagules originating in more southern refugia. Furthermore, our results reveal south-west to north-east directionality in historical migration through the Valley and Ridge region of north-west Georgia.

Keywords
Chloroplast DNA, conservation, biogeography, post-glacial colonization, genetic structure, glacial refugia, phylogeography, *Trillium*.

www.blackwellpublishing.com/jbi
doi:10.1111/j.1365-2699.2007.01834.x

INTRODUCTION

Severe climatic oscillations have led to a series of Pleistocene ice ages that have produced significant changes in species distributions (Hewitt, 2000). In the now classic debate, Deevey (1949) argued that the Pleistocene displaced eastern North American temperate flora much further south than Braun's (1950, 1955) proposed refugia of mixed mesophytic flora in the Alleghany and Cumberland plateaus of Kentucky and Tennessee. Discovery of new fossil sites and improved palaeoecological methods led to an apparent resolution of this debate in Deevey's favour (Jackson *et al.*, 2000). However, the incomplete fossil record and recently emerging molecular evidence indicate that this debate is far from being resolved. More importantly, current rapid climate changes contribute heightened urgency to the study of past species distributions and their dispersal potentials. Arguments regarding past species distributions and post-glacial expansions into their modern ranges remain in the spotlight because the risk of species extinctions from climate change is partially dependent on the ability of species to extend their ranges (Thomas *et al.*, 2004) and/or adapt to new climatic regimes. Our ability to predict a species' future geographical range is, in part, determined by knowledge of past distributions and inferred migration routes and rates.

Several palaeoecological studies provided comprehensive syntheses of the available data (i.e. macrofossils and fossil pollen sequences) in eastern North America during the Last Glacial Maximum (LGM) *c.* 20,000 yr BP (Delcourt & Delcourt, 1981, 1991, 1993; Davis, 1983; Jackson *et al.*, 2000). However, from the available palaeoecological data, we cannot resolve the eastern extent of the mixed *Picea*–temperate deciduous forest that dominated the Lower Mississippi Valley (Jackson & Weng, 1999). The absence of well-dated LGM sites between 30°N and 33°N and east of 91°W prevents determination of the northernmost extent of warm-temperate tree species, associated forest understorey shrubs and herbaceous species except for the Lower Mississippi Valley (Jackson *et al.*, 2000). Temperate hardwood taxa may have occurred from Florida to Mississippi in small widely scattered pockets confined to mesic microsites such as stream courses (Delcourt & Delcourt, 1993; Jackson & Overpeck, 2000). Interpretations of fossil pollen records may be misleading since trace amounts of pollen resulting from long-distance dispersal may lead to overestimates of a species' range during the LGM. Similarly, the absence of a fossil pollen record does not automatically indicate the absence of these species in a sampled range (McLachlan & Clark, 2004). Patterns of range expansions from refugia are difficult to determine from fossil records alone; they may lead to unrealistic estimates of the migration potential of species (McLachlan *et al.*, 2005), and, therefore, other evidence is needed to reconstruct the biogeographical history of modern species (Tremblay & Schoen, 1999).

Population genetic data offer complementary and independent information to reveal cryptic intraspecific lineages and aid in reconstruction of historical distributions of species. Shifts in geographical ranges have had inevitable genetic consequences; the present distribution of genetic variation over a geographical scale originated in part during Quaternary ice ages and subsequent post-glacial expansions. In spite of our ability to detect single base changes in DNA sequences in plant genomes, the flora of major geographical regions, e.g. the south-eastern United States, remains relatively unexplored phylogeographically (Soltis *et al.*, 2006). Seed plants disperse their genes during two life-cycle phases: by pollen prior to fertilization, and subsequently by seeds. Maternally inherited organelle genomes (i.e. chloroplast and mitochondrial DNA) are dispersed only via seeds in most angiosperms. Maternal inheritance of chloroplast DNA (cpDNA) allows one to uncover the genetic footprint of maternal lineages, and to infer migration and colonization processes that gave rise to the observed pattern. Because the chloroplast genome is non-recombining and its mutation rates are low, the majority of current cpDNA haplotypes can be assumed to pre-date the LGM (Wolfe *et al.*, 1987; McLachlan *et al.*, 2005). Consequently, the modern geographical distribution of cpDNA haplotypes and intraspecific phylogenies should reflect the migration routes of populations expanding from glacial refugia.

In this study we investigated the refugial and post-glacial history of an herbaceous spring ephemeral species, *Trillium cuneatum*. This monocot, an understorey species of mesic deciduous forests, is a member of the Trilliaceae (*sensu* Dahlgren *et al.*, 1985) or the Melanthiaceae (*sensu* APG, 1998). It currently ranges from Kentucky, through Tennessee to central Mississippi and Alabama, eastward into Georgia and the Carolinas (Case & Case, 1997). *Trillium cuneatum* is an especially sedentary species; its seeds, distributed primarily by ants and gravity, move very short distances. Such characteristics and its modern geographical distribution make this species an excellent subject for phylogeographical analyses. The northern margin of its present distribution approaches the ice-sheet margin during the height of the last ice age. At the LGM this treeless tundra-covered area was occupied by arctic vegetation, and open boreal woodlands with northern pines spreading south. With winter temperatures falling as low as −25°C (Jackson *et al.*, 2000), it is unlikely that temperate woodland wildflowers such as *T. cuneatum* could have survived the harsh cold and arid conditions. Based on fossil evidence of associated deciduous tree species, the primary candidate for a LGM refugium is the Lower Mississippi Valley (LMV). However, other scattered refugial sites may have occurred in Florida and southern Georgia and Alabama. Presumably, the southern range of *T. cuneatum* contracted as the climate warmed after the LGM, and populations from southern refugia expanded northward. Although it is unclear where the northernmost populations survived the LGM, isolated refugial populations may also have been in protected cove forests of the southern Appalachian Mountains and expanded north and north-east as the climate moderated.

Several hypotheses have emerged from palaeoecological records regarding the evolutionary history of plant species. However, in the south-eastern USA, until recently, these predictions remained largely untested. The principal goal of this research was to examine hypotheses regarding the biogeographical history of *T. cuneatum* using maternally transmitted cpDNA sequences combined with genetic structure analyses of biparentally inherited nuclear genetic markers (Gonzales & Hamrick, in prep.). In this paper we address the following questions: (1) Did *T. cuneatum* survive the LGM in multiple refugia, and, if so, where were they located? Three scenarios are investigated: a single 'western' refugium located in the LMV; primary refugium located in the LMV, and additional scattered sites in Alabama and Georgia; additional refugia in cove forests of the Southern Appalachian Mountains. (2) What are the post-glacial colonization patterns into the present geographical range, did all refugia contribute equally to post-glacial colonization and are routes from the refugia parallel in a south–north direction?

MATERIALS AND METHODS

Sampling

We sampled 45 locations of *T. cuneatum* throughout its modern geographical range, and recorded geographical coordinates for each site (see Table 1 & Fig. 1). From each population, we collected 10 plants at the peak of flowering (at least 10 m apart to avoid collecting clonal individuals). We preserved fresh leaf tissue in liquid nitrogen and stored the samples at −70°C until DNA extraction. Voucher specimens are deposited in the University of Georgia Herbarium (GA).

Laboratory analyses

We carried out the extraction of total genomic DNA and polymerase chain reaction (PCR) amplification using the REDExtract-N-Amp plant PCR kit (Sigma, Poole, UK) according to the manufacturer's instructions. We amplified and sequenced two regions of the chloroplast, the *trnL* intron and the *trnL-trnF* intergenic spacer (total length: 812 bp), using universal primer pairs (Taberlet *et al.*, 1991). We carried out the PCR using the following procedure: 1 cycle (3 min/94°C), 38 cycles (1 min/94°C, 1 min/48°C, 1.5 min/72°C), and 1 cycle (6 min/72°C). Resulting amplified cpDNA fragments were sequenced on an ABI3700 DNA sequencer using a cycle sequencing BigDye Kit (ABI Applied Biosystems, Inc., Foster City, CA, USA) and original amplification primers on both strands. We initially edited and aligned both sequenced strands using Sequencher 4.2.2 (Gene Codes Corporation, Ann Arbor, MI, USA), and then in ClustalX (Jeanmougin *et al.*, 1998). As cpDNA does not recombine, character states for each fragment were combined to yield haploid genotypes (haplotypes).

Initially, we obtained sequences from four individuals in each population and mapped the geographical distribution of the haplotypes. We analysed an additional six individuals from populations that were either polymorphic for cpDNA

Table 1 *Trillium cuneatum* populations, their geographical locations, sample size (*n*), and affiliation with each of the six haplotype groups (A–F). Populations are grouped by states, and the last two letters in each population acronym designate the state of origin (GA, Georgia; AL, Alabama; NC, North Carolina; SC, South Carolina; TN, Tennessee; KY, Kentucky; MS, Mississippi).

Population	Latitude	Longitude	*n*	Haplotype
MACGA	32°54′	83°47′	4	B
TLBGA	32°43′	84°20′	4	B
HRSGA	32°43′	84°52′	4	B
MEAGA	32°40′	84°57′	4	B
FBCGA	32°33′	84°43′	4	B
FBRGA	32°33′	84°47′	4	B
CHATGA	34°53′	84°27′	4	B
ABGGA	33°48′	84°22′	4	B
TMFGA	34°05′	83°51′	4	B
WLKGA	34°39′	85°04′	4	C
FLDGA	34°13′	85°13′	4	B
P°CGA	34°43′	85°23′	10	C
GRSGA	34°51′	84°39′	10	E
BRKGA	34°54′	83°24′	4	E
DALAL	32°19′	86°54′	4	C
LEEAL	32°32′	85°14′	10	B
I20AL	33°42′	86°04′	4	C
BLNAL	34°03′	86°35′	4	C
LAWAL	34°23′	87°14′	4	C
WINAL	34°17′	87°24′	4	C
F°NNC	35°27′	83°49′	4	E
CHE°NC	35°28′	83°54′	4	E
JKNC	35°21′	83°56′	4	E
RCNC	35°21′	83°55′	4	E
SALNC	35°13′	82°21′	4	E
UNGNC	36°04′	79°48′	4	E
UHWNC	35°26′	80°01′	4	E
SCVSC	34°51′	83°05′	4	E
TBRSC	34°52′	83°03′	4	E
JNGSC	35°08′	82°34′	4	F
JGRSC	35°08′	82°30′	10	E
EKRTN	35°17′	85°55′	4	C
DKRTN	35°28′	86°07′	4	C
PRBTN	35°30′	83°56′	4	E
NSHTN	36°04′	86°53′	4	D
NTZTN	35°25′	87°31′	4	C
WHTKY	36°42′	84°14′	4	E
ALLKY	36°51′	86°04′	4	D
LEEMS	34°14′	88°49′	10	C
NEWMS	32°29′	89°04′	10	E
WARMS	32°21′	90°47′	4	A
YAZMS	32°44′	90°26′	10	A(1),C(9)
GRYMS	32°56′	89°28′	4	A
JACKMS	32°18′	90°10′	4	A
LEFMS	32°19′	90°09′	4	A

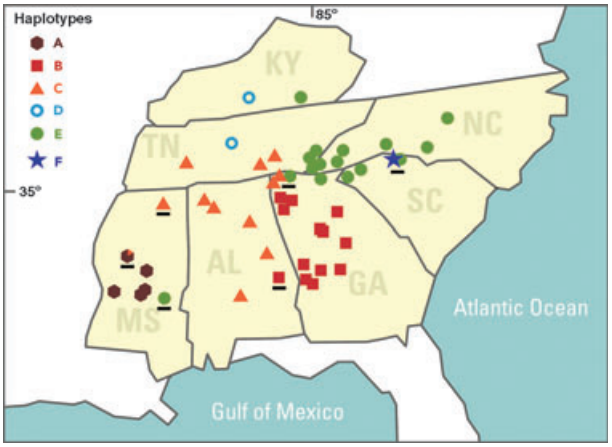

Figure 1 Geographical distribution of cpDNA haplotypes across the range of *Trillium cuneatum*. The haplotypes are indicated as A–F (see Fig. 2 for phylogenetic relationships among haplotypes). Populations with a subtending bar had a sample size of $n = 10$; all others had $n = 4$.

haplotypes or were located in close proximity to another cpDNA lineage (i.e. populations where haplotype polymorphism might occur due to the potential overlap of maternal lineages), resulting in a total of $n = 225$ sequenced individuals (Table 1).

Statistical analyses

We calculated within-population diversity (h_S), i.e. the probability that two randomly chosen haplotypes in a population are different, and total diversity (h_T), i.e. the probability that any two randomly chosen haplotypes are different (Pons & Petit, 1995), and measured genetic differentiation among localities by G_{ST} (Nei, 1987) and N_{ST} (Pons & Petit, 1995, 1996) using the software PERMUT (available at http://www.pierroton.inra.fr/genetics/labo/Software). The N_{ST} parameter takes similarities among haplotypes into account, contrary to G_{ST}. Measures of subdivision that account for the degree of similarity among haplotypes make better use of the information inherent in haplotype data than standard measures (i.e. G_{ST}) based only on frequencies (Petit *et al.*, 2005).

Traditional phylogenetic methods, typically applied in molecular systematics, are often not well suited to the analysis of intraspecific, recently divergent genetic lineages (Posada & Crandall, 2001). To estimate genealogical relationships among cpDNA haplotypes, we employed an approach of Clement *et al.* (2000) and Templeton *et al.* (1992). Their method built into the TCS software creates a network of haplotypes using the probability of parsimony. It calculates the probabilities of the most parsimonious solutions between haplotype connections. Connections among haplotypes within the network are justified by 95% parsimony criterion. The program also generates hypothetical (extinct or not sampled) intermediate haplotypes. We treated indels as the fifth character. The TCS software package is available at http://darwin.uvigo.es.

RESULTS

Chloroplast DNA diversity

We detected six distinct cpDNA haplotypes (A–F). Their DNA sequences differ from one another in six variable characters (single nucleotide substitutions and single nucleotide or short indels). Sequences were deposited in GenBank (GenBank accession numbers EF613273, EF633033–EF633036, EF633027–EF633032). With the exception of one population (YAZMS), all populations were fixed for a single haplotype (Table 1 & Fig. 1), which resulted in very low average population genetic diversity values ($h_S = 0.006$), and considerably higher total genetic diversity ($h_T = 0.77$). This discrepancy between population and species magnitude of genetic diversity is further reflected in genetic structure: among population cpDNA genetic diversity was very high ($G_{ST} = 0.993$, $N_{ST} = 0.995$).

Geographical distribution and genetic structure of cpDNA haplotypes

All haplotypes are geographically clustered (Fig. 1). Haplotype E, which is characteristic of the north-eastern portion of *T. cuneatum*'s range (Appalachian Mountains), was also detected in the southernmost population (NEWMS) in Mississippi. Three haplotypes (B, C and E) were widespread, while haplotypes A and D were restricted to a few locations or detected only in a single population (F).

Phylogeographical relationships

We used the haplotype geographical distribution combined with the most parsimonious criteria to infer plausible phylogeographical relationships. The resulting TCS parsimony analysis combined haplotypes into a contiguous network with two intermediate missing haplotypes, necessary to link the observed haplotypes (Fig. 2). The TCS analysis clustered populations based on their cpDNA haplotypes into two groups separated by three mutations (i.e. two missing haplotypes). Populations with haplotype A formed clade I, while the remaining populations with haplotypes B–F produced clade II. This separation is congruent with our genetic structure analyses of biparentally inherited nuclear markers (Gonzales & Hamrick, in prep). Although we cannot determine an absolute time of separation of the two lineages, considering an average synonymous substitution rate of $(1.00–3.00) \times 10^{-9}$ per site per year in cpDNA genome (Wolfe *et al.*, 1987), the separation possibly pre-dates not only the LGM but also previous glacial episodes of the Wisconsin glaciation period (Zurawski *et al.*, 1984). The two clades have followed their independent evolutionary trajectories as sister lineages from an extinct common progenitor (or a haplotype not contained in our sample).

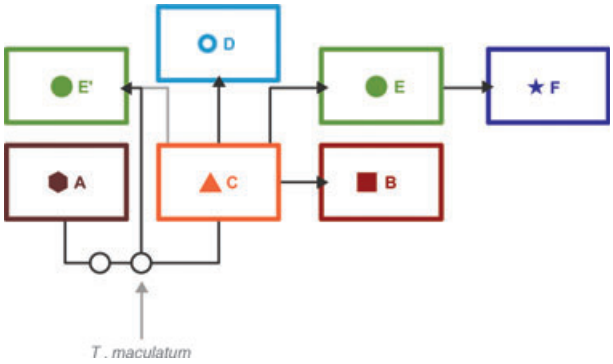

Figure 2 The haplotype network that includes the most parsimonious connections (> 95%) resulting from the TCS analyses. Missing haplotypes are indicated by empty circles (customary in TCS haplotype networks) and each *line* represents *one* mutation; (the 6-bp indel between C and B is considered one character/one mutation). The arrow at the bottom of the network indicates an outgroup (*T. maculatum*; sequences obtained from GenBank). Haplotype E could have evolved directly from C or from an extinct/unsampled haplotype. Both scenarios could have occurred; we refer to a disjunct population with this cpDNA sequence as E′. Alternatively, both E (in the mountains) and E′ (disjunct population) could have evolved independently from different populations with haplotype C (indicated by the grey line).

Phylogeographical relationship within clade II

The C haplotype is the most likely progenitor of the remaining lineages within clade II. Each 'ancestral-derived' pair of haplotypes in the eastern portion is separated by a single nucleotide substitution or indel, with the exception of B which is separated by an indel (most likely an insertion six nucleotides long, treated as one character) from its likely ancestor C. We base this conclusion not only on comparisons with the other *T. cuneatum* lineages (none possess this 6 bp sequence), but additionally, all congener species representing both ancestral and sister taxa (Kato *et al.*, 1995; Zomlefer *et al.*, 2001) also lack this 6-bp region. No ambiguity exists about linkage of haplotype D, which evolved independently from C, or haplotype F, which was derived from E. Haplotype E, characteristic mostly of the north-eastern geographical distribution, may have evolved by two pathways: from populations carrying haplotype C, located south-west of the region with haplotype E, or possibly from another ancestor, a hypothetical haplotype missing or not detected in our samples, indicated by an empty circle in Fig. 2. Interestingly, these two scenarios may not be mutually exclusive. Haplotype E is found both in the mountain populations (north-east portion), and also in a population in the south-west portion of *T. cuneatum*'s distribution. It is plausible that we have detected a case of homoplasy; in such a scenario, NEWMS (the southernmost population in our study) may carry a haplotype that evolved from a missing (extinct) progenitor (henceforth E′), and populations in the mountains may carry a haplotype with identical sequence (E) that evolved independently from haplotype C. On the basis of our data, we cannot distinguish between these possibilities; however, our nuclear data may provide more insight (Gonzales & Hamrick, in prep.).

DISCUSSION

Our investigation revealed a *geographically* highly structured intraspecific phylogeny (Fig. 1) indicating that *T. cuneatum* survived the LGM in multiple glacial refugia, belonging to two (western and eastern) main genealogical lineages clustered across south-eastern North America. Assuming the low chloroplast DNA mutation rates (Wolfe *et al.*, 1987), modern patterns of cpDNA haplotype distribution suggest cryptic vicariance, in the form of forest contractions and subsequent expansions associated with Pleistocene glacial cycles, rather than simple southern survival and subsequent northward colonization. The latitudinal stratification in cpDNA haplotype distribution was unexpected, suggesting that populations of this rather southern thermophilic species might have survived in more northern locations than we initially expected, based on fossil records and LGM climate reconstruction, and that the Appalachian Mountains functioned as a barrier to dispersal of propagules originating in the more south-eastern refugia. Furthermore, our results reveal south-west to north-east directionality in historical migration through the Valley and Ridge region of north-west Georgia and south-east Tennessee. However, an alternative interpretation of our results could be also proposed considering the uncertainty of the molecular clock for the cpDNA sequences used in our study. If mutation rates were higher than we assumed based on Wolfe *et al.* (1987), then some of the cpDNA haplotypes within the eastern lineage may have differentiated during post-glacial expansion. We view the latter scenario as being less likely because of the high interpopulation divergence and the number of private alleles in nuclear markers among the cpDNA lineages (Gonzales & Hamrick, in prep.). We consider our results, therefore, to be more indicative of long-term separation rather than of founder effects and bottlenecks associated with the colonization of newly available habitat. Pollen dispersal in this species is very limited (Gonzales *et al.*, 2006), and thus it is unlikely that it would completely erase the evidence of post-glacial propagule dispersal.

Our findings reinforce the phylogeographical concordance of a diverse co-distributed regional biota (Soltis *et al.*, 2006). Comparative molecular assessments of vertebrates in the south-eastern United States have revealed repeated phylogenetic 'breaks' which typically distinguish populations in the eastern portion of the species' range from those to the west; additional substructure is evident within these two phylogeographical lineages, but these differences are typically shallow relative to the matrilineal separation between regions (Bermingham & Avise, 1986; Nedbal & Philipp, 1994; Walker & Avise, 1998; Avise, 2000). Although *T. cuneatum* cpDNA haplotypes are not as strongly differentiated as mitochondrial DNA (mtDNA) in animal genomes (due to a slower mutation rate of cpDNA), molecular evidence also grouped *T. cuneatum* populations into two main lineages; one occupying Mississippi

(western clade I, haplotype A), and the other representing the remaining, more easterly range (clade II, haplotypes B–F) (Figs 1 & 2). Our results suggest that these two lineages may have evolved from a now extinct ancestral haplotype, and the extant lineages have evolved independently as 'sister clades'. Assuming the low rate of cpDNA mutations, the lineages probably diverged prior to the LGM (Wolfe *et al.*, 1987; Dorken & Barrett, 2004). The two lineages seem to be separated geographically by a region with few populations. We were not successful in finding any natural populations (or their records in UGA or Auburn University Herbaria) in the area along today's Alabama/Mississippi border (Fig. 1). This could be due to a natural lack of dispersal into the area or due to anthropogenic extirpation; however, we have no reason to believe that the anthropogenic pressures on this species were stronger here than elsewhere across its southern range. Given that this gap in population density occurs between the two clades, the spatial discontinuity between the two lineages appears natural.

The number of cpDNA haplotypes represents the minimum number of glacial refugia, their approximate locations, and provides a basis for inferring post-glacial expansion patterns, assuming that the haplotype sequences differentiated before the LGM. As predicted by the palaeontological and palynological records, we identified a refugium in the Lower Mississippi Valley (western clade I, lineage A). Additionally, we identified a more easterly clade of lineages (clade II, lineages B–F). Contrary to expectations, the Mississippi refugium has not played an important role during post-glacial range expansion of *T. cuneatum* into its present range. Our data suggest that the LMV populations may have followed their own evolutionary trajectory, separate from the rest of the species, probably before the LGM. Most of the populations in the south-eastern USA are derived from at least three refugia (haplotypes B, C and E) belonging to clade II, poorly documented by temperate tree fossil pollen evidence. It is clear that these sites played a major role in the post-glacial expansion of the species.

The greatest contribution to the current range comes from centrally located haplotype C, placing a refugium in Alabama. This haplotype also appears to be the progenitor of haplotypes B and D–F within the eastern lineage (clade II). Additionally, another refugium (haplotype B) might have existed in Georgia (and possibly in northern Florida), but this lineage is currently confined to the south-eastern portion of the species' range, albeit larger than its south-western counterpart in Mississippi. It is unclear whether the extant Georgia populations represent *T. cuneatum*'s southernmost extent during the LGM; other populations (now extinct) may have existed in Florida. Such Florida refugia have been proposed for other species (e.g. *Liriodendron tulipifera*, Sewell *et al.*, 1996; *Sagittaria latifolia*, Dorken & Barrett, 2004; Mylecraine *et al.*, 2004; *Acer rubrum*, McLachlan *et al.*, 2005). Propagules originating in Georgia did not migrate very far to the north. Spread of haplotype B may have been blocked by the Tennessee Continental Divide (referred to as the Valley and Ridge physiographical region) in north-west Georgia, and by the Blue Ridge Mountains in north Georgia. The Valley and Ridge physiographical region deserves special attention. It derives its name from a series of parallel south-west to north-east trending valleys and ridges formed by folded and faulted sedimentary rocks extending south-west into Alabama. Previous biogeographical investigations speculated that these parallel ridges and valleys facilitated species migrations during range contractions and expansions (Adams, 1901, 1902; Wharten, 1999). The distribution of *T. cuneatum* haplotypes provides additional support for this hypothesis. Our data suggest that propagules carrying haplotype C expanded from their southern Alabama refugia in a north-east direction following the Valley and Ridge corridors, leaving their genetic footprint behind. The southern Appalachians, dominated by boreal forest during the LGM, are occupied by populations with haplotype E, derived from the more centrally located haplotype C. The current distribution of haplotype E indicates that this lineage probably did not enter the mountains during the most recent post-glacial range expansion; rather, a few populations may have survived in the Appalachian Mountains during the LGM and subsequently repopulated the north-eastern portion of the current range. This scenario is congruent with evidence for *Fagus grandifolia* and *A. rubrum* refugia in the southern Appalachians (McLachlan *et al.*, 2005). Alternatively, haplotype E could have survived in a refugium south-west of its current distribution. This scenario is, however, less likely because we did not detect haplotype E in Alabama. The distribution of haplotype C (extending through the migration corridors of the Tennessee Continental Divide) and haplotype E (present only in the mountains and North Carolina Piedmont) is more consistent with the conclusion that there were glacial refugia within the mountains during the LGM. Consistent with studies by Dorken & Barrett (2004) and McLachlan *et al.* (2005), we identified unique haplotypes in more northerly locations than we expected, based on climate reconstruction; for instance, haplotype F was only observed in a single population above 35°N, and D was not found south of 36°N. Such patterns indicate the possible existence of glacial refugia further north than initially anticipated.

The southern Appalachian populations share the same haplotype (E) with the exception of population JNGSC which has a unique haplotype, F, derived from E. Surprisingly, plants sampled in a population approximately 3 km away have haplotype E. Thus, population JNGSC may be a relict from an isolated refugium in the southern Appalachian Mountains; alternatively, the unique haplotype in JNGSC could have arisen more recently, after the LGM. If JNGSC is indeed the actual site of a glacial refugium, rather than a more recent, post-glacial mutation, sampling of other species in this area may reveal similar patterns and provide further support for the multiple Appalachian refugia.

Additionally, propagules carrying haplotype D may have originated in south-central Tennessee or may have migrated/dispersed to their current geographical location during

post-glacial range expansion. Additional sampling in the northern Alabama–southern Tennessee area should provide more details on the location of this refugium. This finding is congruent with patterns detected in *A. rubrum* and *F. grandifolia* (McLachlan *et al.*, 2005). Similarly, Dorken & Barrett (2004) identified rare *S. latifolia* haplotypes in the same general area, and Griffin & Barrett (2004) identified two unique haplotypes of *Trillium grandiflorum* whose current range is near the glacial boundary.

Generally, the populations were fixed for one haplotype, with a single exception (YAZMS) where nine of the ten sampled individuals had haplotype A and one individual had haplotype C. Population YAZMS, located at the zone of contact between lineages A and C, occupies a secondary forest of a recently abandoned cotton field (R. Wieland, personal communication). Deciduous trees in this site are young, and are mixed with pines, a sign of recent secondary succession. It is feasible that seeds were moved into this site from nearby populations belonging to different lineages by white-tailed deer (*Odocoileus virginianus*). In a concurrent study, we observed deer browsing of mature *Trillium* fruits, and a similar observation in *T. grandiflorum* was reported by Vellend *et al.* (2003).

By and large, *T. cuneatum* haplotypes are each clustered into single geographical subregions. We detected a single exception to this general observation. Plants from the NEWMS population at the south-westernmost edge of *T. cuneatum*'s distribution have haplotype E (=E′), the haplotype otherwise characteristic of the southern Appalachian Mountains. The distance between NEWMS and the rest of the range of haplotype E exceeds 500 km. The detection of identical sequences in these disjunct locations could be explained either by homoplasy or, alternatively, by a rare long-distance dispersal event. The TCS analyses yielded two possible pathways (Fig. 2): populations with haplotype E could have evolved from the C haplotype or from a missing haplotype (two mutational events). Both scenarios are more than 95% likely; however, they do not need to be mutually exclusive. If E evolved just once, then NEWMS is an outcome of a long-distance (possibly human-mediated) dispersal. Alternatively, this population could have evolved from a now extinct haplotype and represents a unique relict of a southern refugium. Another possibility of homoplasy, consistent with the TCS analysis outcome (but not directly considered by it) exists: haplotype E′ in NEWMS may have evolved from a different ancestral population within lineage C, possibly nearby in Alabama (Fig. 2). Based on our cpDNA sequence data, we cannot exclude either scenario; however, sequencing of additional cpDNA fragments might provide further information.

CONCLUSION

This study established phylogeographical patterns of *T. cuneatum,* a herbaceous species with limited seed dispersal which is associated with mesic deciduous forests across south-eastern North America. We identified a phylogeographical break between two main lineages associated with a rather weak natural range disjunction between Mississippi populations and the remainder of the range. The eastern lineage contains several closely related cpDNA haplotypes likely predating LGM and representing multiple glacial refugia that persisted as low-density populations in locally favourable conditions and provided propagules for more or less continuous present-day distribution. Our results indicate that the past range limits were similar to its present-day geographical distribution. The persistence of unique haplotypes in higher latitudes suggests that some populations most likely survived closer to the Laurentide Ice Sheet than previously thought. This finding, combined with the fact that haplotypes are very strongly spatially clustered, that cpDNA lineages do not overlap, and that, with one possible exception, there is no evidence for long-dispersal events, has an important consequence for conservation and the future of the species. The migration rate in this species seems very low and it might prove difficult to keep up with the predicted climate changes; a situation further exacerbated by anthropogenic fragmentation of deciduous forests isolated by inhospitable habitats. Additionally, the presence of multiple lineages revealed cryptic evolutionarily significant units, a finding important for designing sound conservation strategies, defining management units and the appropriate geographical scale for management. There are just a few phylogeographical studies of plant species of un-glaciated south-east North America available, thus making generalizations premature. The findings of this study will be better understood once they are combined with nuclear molecular data, and ultimately placed in a broader context of more comprehensive investigations of co-distributed species with analogous seed dispersal strategies and life-history traits. Characterization of conservation units, combined with ecological and adaptation processes will lead to appropriate and effective strategies to maintain the rich south-eastern North American flora.

ACKNOWLEDGEMENTS

We are very grateful to Tom Patrick, Jeff Mitton, Lee Pratt, John Avise, Michael Arnold, Ron Pulliam and Peter Smouse for their continuous support throughout this project and helpful critiques on earlier drafts of this manuscript. We also appreciate the help of herbaria at UGA, Auburn University and University of Mississippi for their assistance in the identification of collection sites. Additionally, we thank the staff from National Heritage Programs in Mississippi, Alabama, Georgia and South Carolina, Great Smoky Mountains National Park, National Forests in Georgia and Alabama, and Warner Park, Nashville, TN. This project was supported by grants from NSF-BCS 0327147, Highlands Biological Station, NC, Georgia Native Plant Society, and the University of Georgia Department of Plant Biology Palfrey fund.

REFERENCES

Adams, C.C. (1901) Baseleveling and its faunal significance, with illustrations from southeastern United States. *The American Naturalist*, **35,** 839–852.

Adams, C.C. (1902) Southeastern United States as a center of geographical distribution of flora and fauna. *Biological Bulletin*, **3,** 115–131.

APG (Angiosperm Phylogeny Group) (1998) An ordinal classification for the families of flowering plants. *Annals of the Missouri Botanical Garden*, **85,** 531–553.

Avise, J.C. (2000) *Phylogeography: the history of formation of species*. Harvard University Press, Cambridge, MA.

Bermingham, E. & Avise, J.C. (1986) Molecular zoogeography of freshwater fishes in the southeastern United States. *Genetics*, **113,** 939–965.

Braun, E.L. (1950) *Deciduous forests of eastern North America*. Hafner Press, New York.

Braun, E.L. (1955) Phytogeography of the unglaciated eastern United States and its interpretation. *Botanical Review*, **21,** 297–375.

Case, F.W. & Case, R.B. (1997) *Trilliums*. Timber Press Inc., Portland, OR.

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

Dahlgren, R.M.T., Clifford, H.T. & Yeo, P.F. (1985) *The families of the monocotyledons: structure, evolution and taxonomy*. Springer-Verlag, Berlin.

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

Deevey, E.S., Jr (1949) Biogeography of the Pleistocene, Part I: Europe and North America. *Bulletin of the Geological Society of America*, **60,** 1315–1416.

Delcourt, P.A. & Delcourt, R.H. (1981) Vegetation maps for eastern North America: 40,000 yr BP to the present. *Geobotany II* (ed. by R.C. Romans), pp. 123–126. Plenum, New York.

Delcourt, P.A. & Delcourt, R.H. (1991) *Quaternary ecology. A palaeoecological perspective*. Chapman & Hall, New York.

Delcourt, P.A. & Delcourt, R.H. (1993) Paleoclimates, paleovegetation, and paleofloras during the Late Quaternary. *Flora of North America, Vol. I. Introduction* (ed. by Morin N.R. & Brouillet L.), pp. 71–94. Oxford University Press, New York.

Dorken, M.E. & Barrett, S.C.H. (2004) Chloroplast haplotype variation among monoecious and dioecious populations of *Sagittaria latifolia* (Alismataceae) in eastern North America. *Molecular Ecology*, **13,** 2699–2707.

Griffin, S.R. & Barrett, S.C.H. (2004) Post-glacial history of *Trillium grandiflorum* (Melanthiaceae) in Eastern North America. *American Journal of Botany*, **91,** 465–473.

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

Jackson, S.T. & Overpeck, J. (2000) Responses of plant populations and communities to environmental changes of the late Quaternary. *Paleobiology*, **26,** 194–220.

Jackson, S.T. & Weng, C. (1999) Late Quaternary extinction of a tree species in eastern North America. *Proceedings of the National Academy of Sciences USA*, **96,** 13847–13852.

Jackson, S.T., Webb, R.S., Anderson, K.H., Overpeck, J.T., Webb, T., Williams, J.W. & Hansen, B.C.S. (2000) Vegetation and environment in Eastern North America during Last Glacial Maximum. *Quaternary Science Reviews*, **19,** 489–508.

Jeanmougin, F., Thompson, J.D., Gouy, M., Higgins, D.G. & Gibson, T.J. (1998) Multiple sequence alignment with ClustalX. *Trends in Biochemical Sciences*, **23,** 403–405.

Kato, H., Terauchi, R., Utech, F.H. & Kawano, S. (1995) Molecular systematics of the Trilliaceae sensu lato as inferred from rbcL sequence data. *Molecular Phylogenetics and Evolution*, **4,** 184–193.

McLachlan, J.S. & Clark, J.S. (2004) Reconstructing historical ranges with fossil data at continental scales. *Forest Ecology and Management*, **197,** 139–147.

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

Mylecraine, K.A., Kuser, J.E., Smouse, P.E. & Zimmermann, G.L. (2004) Geographic allozymes variation in Atlantic white-cedar, *Chamaecyparis thyoides* (Cupressaceae). *Canadian Journal of Forest Research*, **34,** 2443–2454.

Nedbal, M.A. & Philipp, D.P. (1994) Differentiation of mitochondrial DNA in largemouth bass. *Transactions of the American Fisheries Society*, **123,** 460–468.

Nei, M. (1987) *Molecular evolutionary genetics*. Columbia University Press, New York.

Petit, R.J., Duminil, J., Fineschi, S., Hampe, A., Salvini, D. & Vendramin, G.G. (2005) Comparative organization of chloroplast, mitochondrial and nuclear diversity in plant populations. *Molecular Ecology*, **14,** 689–701.

Pons, O. & Petit, R.J. (1995) Estimation, variance and optimal sampling of gene diversity. I. Haploid locus. *Theoretical and Applied Genetics*, **90,** 462–470.

Pons, O. & Petit, R.J. (1996) Measuring and testing genetic differentiation with ordered vs. unordered alleles. *Genetics*, **144,** 1237–1245.

Posada, D. & Crandall, K.A. (2001) Intra-specific gene genealogies: trees grafting into networks. *Trends in Ecology & Evolution*, **16,** 37–45.

Sewell, M.M., Clifford, R.P. & Chase, M.W. (1996) Intra-specific chloroplast DNA variation and biogeography of North American *Liriodendron* L. (Magnoliaceae). *Evolution*, **50,** 1147–1154.

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

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

Templeton, A.R., Crandall, K.A. & Sing, C.F. (1992) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics*, **132,** 619–633.

Thomas, C.D., Cameron, A., Green, R.E., Bakkenes, M., Beaumont, L.J., Collingham, Y.C., Erasmus, B.F.N., de Siqueira, M.F., Grainger, A., Hannah, L., Hughes, L., Huntley, B., van Jaarsveld, A.S., Midgley, G.F., Miles, L., Ortega-Huerta, M.A., Peterson, A.T., Phillips, O.L. & Williams, S.E. (2004) Extinction risk from climate change. *Nature*, **427,** 145–148.

Tremblay, N.O. & Schoen, J.D. (1999) Molecular phylogeography of *Dryas integrifolia*: glacial refugia and postglacial recolonization. *Molecular Ecology*, **8,** 1187–1198.

Vellend, M., Myers, J.A., Gardescu, S. & Marks, P.L. (2003) Dispersal of *Trillium* seeds by deer: implications for long-distance migration of forest herbs. *Ecology*, **84,** 1067–1072.

Walker, D. & Avise, J.C. (1998) Principles of phylogeography as illustrated by freshwater and terrestrial turtles in the southeastern United States. *Annual Review of Ecology and Systematics*, **29,** 23–58.

Wharten, C.H. (1999) The Cumberland Plateau and the Valley and Ridge. *Highroad guide to the Georgia Mountains*, pp. 1–2. Longstreet Press, Athens, GA.

Wolfe, K.H., Li, W.H. & Sharp, P.M. (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.

Zomlefer, W.B., Norris, H., Williams, N., Whitten, W.M. & Judd, W.S. (2001) Generic circumscription and relationships in the tribe Melanthieae (Liliales, Melanthiaceae), with emphasis on *Zigadenus*: evidence from ITS and trnL-F sequence data. *American Journal of Botany*, **88,** 1657–1669.

Zurawski, G., Clegg, M.T. & Brown, A.H.D. (1984) The nature of nucleotide sequence divergence between barley and maize chloroplast DNA. *Genetics*, **106,** 735–749.

BIOSKETCHES

This research stems from the authors' shared interest in historical biogeography, natural history and evolutionary biology. This work represents a part of **Eva Gonzales**'s PhD dissertation, conducted under the supervision of J. L. Hamrick. E. Gonzales is currently working as a post-doctoral associate at Rutgers University.

J. L. Hamrick studies the genetics and evolution of natural plant populations with the emphasis on genetic structure and gene dispersal mediated by pollen and seeds.

S. M. Chang is interested in ecological and evolutionary questions concerning plant reproduction and factors that generate and maintain genetic variation in natural populations.

Editor: Brett Riddle