MULTI-SCALE GENETICS ANALYSES OF TWO *TRILLIUM* SPECIES (*TRILLIUM RELIQUUM* AND *TRILLIUM CUNEATUM*) IN SPACE AND TIME

by

EVA B. GONZALES

(Under the Direction of J. L. HAMRICK)

ABSTRACT

This dissertation was motivated by an interest to integrate spatial-temporal considerations into understanding the evolutionary history of two forest herbaceous species, *Trillium reliquum* and *T. cuneatum*, and those mechanisms governing their population genetic processes.

The investigation of *T. reliquum* addressed a hypothesis that this endangered species, surviving today in relict, disjunct populations, was previously widespread, and that it became rare due to European settlements and subsequent habitat fragmentation. Comparisons of the distribution of genetic diversity among populations of *T. reliquum* and *T. cuneatum*, its more common. albeit also fragmented, congener, revealed strong genetic structure among populations of both species. However, the disjunct *T. reliquum* populations are much more divergent than those of *T. cuneatum*, in spite of their shared recent history, suggesting that rarity in *T. reliquum* is more ancient, possibly predating the last glacial episode, rather than a consequence of post-European colonization.

Examination of hypotheses emerging from biogeographical and fossil records regarding glacial refugia of *T. cuneatum* in the southeastern US revealed multiple refugia. Surprisingly, the Lower Mississippi Valley refugium, considered by paleoecologists as the main refuge for

deciduous forest species, did not participate in postglacial expansion. Rather, scattered refugial populations in Alabama, Georgia and the southern Appalachian Mountains contributed to the current geographic distribution. Even more unexpected is the conclusion that *T. cuneatum* must have survived at more northern latitudes than the fossil record indicates. Furthermore, this study identified the Ridge and Valley as a corridor for species migration in their response to post-glacial climatic changes.

Investigation of within population mechanisms revealed slow acting consequences of anthropogenic activities on genetic diversity. Although we found evidence for highly localized gene dispersal in all populations, comparisons among continuous mountain populations with fragmented and environmentally stressed Piedmont populations revealed that fragmentation appears to promote pollen-mediated gene movement and to spatially homogenize genetic variation. I attributed these differences largely to edge effects and concluded that human disturbance has increased environmental stress in the Piedmont populations, but has caused only subtle genetic and demographic shifts to date.

INDEX WORDS: Population genetics, phylogeography, ecological genetics, genetic structure, gene flow, conservation, rarity, *Trillium,* forest understory herbs

MULTI-SCALE GENETICS ANALYSES OF TWO *TRILLIUM* SPECIES (*TRILLIUM RELIQUUM* AND *TRILLIUM CUNEATUM*) IN SPACE AND TIME

by

EVA B. GONZALES

Ing., University of Economics, Prague, Czechoslovakia, 1986

B. S., Emory University, 1995

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2005

MULTI-SCALE GENETICS ANALYSES OF TWO *TRILLIUM* SPECIES (*TRILLIUM RELIQUUM* AND *TRILLIUM CUNEATUM*) IN SPACE AND TIME

by

EVA B. GONZALES

Major Professor: James L. Hamrick

Committee: Michael Arnold
John Avise
ShuMei Chang
Ron Pulliam

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
May 2005

DEDICATION

To Daniele

ACKNOWLEDGEMENTS

I would like to thank my advisor, Jim Hamrick for his guidance, encouragement, and extraordinary patience. Jim gave me the freedom and confidence to explore areas of my own interests and to develop my creativity, while continuously challenging me to be a better scientist and a better writer. His extraordinary dedication to research and to his students will always be a great inspiration.

I have benefited from the advice and guidance of my truly excellent committee members, Mike Arnold, John Avise, Shumei Chang and Ron Pulliam. Thanks to their stimulating discussions, they helped to shape my dissertation and provided thorough critiques of my manuscripts. Additionally, Shumei Chang trusted me in her new lab where I conducted a considerable part of my phylogeography project. Besides my formal committee, I am also grateful to Lee Pratt and his lab for their assistance with DNA sequencing.

I particularly treasure the mentorship and friendship freely given by Peter Smouse from Rutgers University and by Jeff Mitten from University of Colorado. I am grateful for their continuous sage advice and assistance, especially in the pollen mediated gene dispersal phylogeography studies. Many thanks go to Rodney Dyer (Virginia Commonwealth University) for his support with data analyses, and to Joe Williams (University of Tennessee) for deciphering the secrets of bisporic gametogenesis in *Trillium*.

I am grateful to numerous UGA faculty, graduate, undergraduate students and lab technicians for their friendship, advice and hard work. In particular, I would like to thank Mary Jo Godt, Kathy Parker, Chris Peterson, Wendy Zomlefer, Susan Watkins, Andy Tull, Brian

Perkins, Jenny Cruse-Sanders (who was always there to celebrate the high points and encourage me during the low ones), Denise Hardesty (my dissertation buddy during the last months), Rebecca Pappert, Cecile Deen, Jenny Moore, JP Schmidt, my office-mates Eleanor Pardini and Eric Fuchs, and other members of the Hamrick and Hubbell labs.

I owe my thanks to numerous people who aided me in the field with continuous enthusiasm as well as logistical support. In particular, DNR state botanist, Tom Patrick, shared his secret *Trillium* places; I would be still looking for my first Relict *Trillium* without his assistance. If it were not for the help of other fellow botanists and *Trillium* enthusiasts, I would never complete the phylogeography investigation. I especially appreciate help of people from the Natural Heritage Programs, Great Smoky Mountains National Park, Chattahoochee and Nantahala National Forests, P. Warner Park in Tennessee and herbarium curators in University of Mississippi and Auburn for their assistance with *Trillium* collections.

This dissertation was made possible by financial support provided by the following organizations: the Palfrey Grant from Plant Biology Department, UGA, Dissertation Improvement Grant from National Science Foundation, Joshua Laerm Award from Georgia Museum of Natural History, Native Plant Society, and Highlands Biological Station, NC.

Finally, I would like to thank my family for their unconditional love and support. I thank my young nieces and nephew, Tereza, Katherine and Martin, for all their hard work in the field and in the lab; I will always treasure my memories of our times spent together in the field, times of enthusiasm, curiosity and discovery. Above all, I thank my husband for his enduring support, wit, endless sense of humour, and for putting up with hearing about my projects at too many dinners.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

> *In the study of dispersal and distribution of animals, it is important to see that the physical conditions lead, and that in a more or less definite succession the flora and fauna follow; thus the fauna comes to fit the habitat as a flexible material does a mold. The time is passed when faunal lists should be the aim of faunal studies. The study must not only be comparative, but* genetic*, and much stress must be laid on the study of the habitat, not in a static, rigid sense, but as a fluctuating or periodical medium.*
>
> Charles Adams, 1901

This dissertation was motivated by an interest to integrate considerations of both space and time into understanding the evolutionary history of a species, and those mechanisms governing population genetic processes. Thus, its central focus is the connection of a *spatial* gradient varying from local within-population microhabitats to the geographic distribution of the species, with a *temporal* gradient extending from the present into the deep history of the Quaternary period.

The integral role of the spatial structure of genetic diversity in evolutionary processes was recognized in the formation of the neo-Darwinian synthesis; founding concepts in evolutionary theory (e.g. shifting balance theory or isolation by distance) were established in the first half of the 20th century (Wright 1946, 1951). In contrast, the integration of temporal aspects of population genetic processes (e.g. direct measurements of contemporary gene flow) were introduced into empirical evolutionary investigations more recently, owing its existence primarily to recent advances in the molecular analyses of populations. Spatial patterns capture the temporal changes; they capture the cumulative effects of evolutionary forces over many generations. We can study spatial patterns, using traditional and new molecular and analytical tools and comparative approaches, to reveal the presence of micro-evolutionary forces and mechanisms responsible for

generating spatial genetic structure. Any valid, comprehensive investigation of micro-evolutionary processes must include a spatial and temporal context; this is true within populations as well as in the context of a system of populations on a widespread, geographic scale.

In this dissertation, I investigated mechanisms generating genetic diversity patterns and inferred micro-evolutionary history of two related *Trillium* species, *T. reliquum* and *T. cuneatum*. These monocot species are members of the Trilliaceae (sensu Dahlgren et al., 1985) or the Melanthiaceae (sensu APG, 1998). Several life-history traits of these perennial spring ephemerals may affect the genetic architecture of their populations. *Trillium* species typically require more than 10 years to reach the reproductive stage (Ohara 1989). Both species have a leaky self-incompatible system; they are polycarpic, and reproduce infrequently, both by seeds and clonal spread. Individual genets may persist for decades. Weak-flying insects pollinate *Trillium* (pers observ), and like many other forest understory herbs, it is myrmecochorous; seeds are primarily dispersed by ants or passively by gravity. Thus gene dispersal mediated by both pollen and seeds is expected to be restricted.

Federally protected *T. reliquum* is a deciduous forest understory spring ephemeral, with a narrow geographic range but broad habitat specificity – an unusual condition for rarity. It has been only recently described (Freeman, 1975), and is thought to have suffered greatly from human activities during the last three centuries. Although its historical geographic distribution is unknown, extant populations are thought to represent remnants of a formerly more widespread species - hence the specific epithet *reliquum* (relict) (Freeman, 1975). Populations are small to moderate in size, isolated from one another, and are primarily restricted to the Fall Line Hills district from the Georgia-South Carolina border to southwestern Georgia and southeastern Alabama. In contrast, *T. cuneatum*, partially sympatric with *T. reliquum*, is more widespread; it often carpets floors of mature mesic deciduous forests, both in large continuous, as well as smaller, disturbed remnant habitats. Its geographic range extends from central Kentucky, through Tennessee to central Mississippi and Alabama, and east into Georgia and the Carolinas.

The investigation of *T. reliquum* focused largely on a hypothesis that this rare and federally endangered species, currently surviving in relict disjunct populations, was previously widespread, and that it became rare due to European settlements and the subsequent fragmentation of its once contiguous habitats. I contrasted the genetic diversity and its spatial distribution among *T. reliquum* populations and sympatric populations of its more common congener *T. cuneatum* to gain insights into the history of rarity of this endangered species. This comparison was based on the assumption that populations of both species were subjected to the same land use history. Consequently, the genetic composition of both species should reflect its common history in a similar fashion, providing these two species were both formerly common. In contrast, if populations of *T. reliquum* were disjunct prior to European arrival, we should observe higher genetic structure among its populations than among the *T. cuneatum* populations.

The remaining objectives of my dissertation focus on *T. cuneatum*. I have extended investigation of this species beyond its sympatric region with *T. reliquum*, and conducted a phylogeography study of its entire geographic range. I employed a combination of molecular markers that differ in their mode of inheritance: maternally inherited chloroplast markers (cpDNA sequences) and biparentally inherited nuclear markers (allozymes). Furthermore, I combine these molecular markers with previously published fossil records of temperate deciduous trees, and traditional, non-molecular biogeographical patterns of southeastern biota to reveal the history of *T. cuneatum* since the end of the last glacial episode approximately 20,000 years ago. Specifically, I used molecular evidence to examine hypotheses (e.g. Adams 1901, Delcourt and Delcourt 1993) which emerged from previously published fossil records in the unglaciated SE North America regarding potential refugia of temperate species associated with mesic deciduous forests. Additionally, I examined the hypothesis that emerged from the geology and topography of the SE US regarding potential postglacial migration corridors via the Tennessee Continental Divide extending from Alabama in a southwest-to-northeast direction, leading to the southern Appalachian Mountains. In spite of numerous studies of southeastern animal taxa, in the plant literature, these hypotheses remain untested. Additionally, I examined whether there was molecular evidence to support the so

called “Out of Appalachia” hypothesis which reflects the fact that the southern Appalachian Mountains house an unusual number of species, many endemic. This hypothesis implies presumed isolated glacial refugia within protected cove forests of the southern Appalachians (otherwise dominated by boreal forests during the glacial maxima) in the process of speciation. Finally, I compared these results with previous phylogeographic studies and looked for concordance with other taxa.

For the remaining objectives, I narrowed the geographic spectrum as well as the temporal scale of my investigation. I focused on within population genetic diversity, its distribution and underlying mechanisms responsible for shaping genetic structure within five populations representing two contrasting environments (Georgia Piedmont and southern Appalachians). Using a comparative approach, I addressed questions of whether different ecological conditions (both natural and human induced) in these two environments affect within population genetic structure, clonal reproduction and contemporary pollen-mediated gene flow. Investigation of fine-scale genetic structure is based on the Isolation-by-Distance model and genetic neighborhood concept first proposed by S. Wright (1951). In continuous populations with restricted gene dispersal, genetic structure develops and relatedness among individuals increases with spatial proximity. Understanding fine-scale genetic structure provides insights into the effective dispersal of propagules and mechanisms involved in maintaining genetic variation operating at the population level. Combined with analyses of clonal reproduction and contemporary measurements of pollen movement, these approaches allow us to evaluate relative contributions of vegetative reproduction, and seed and pollen mediated gene movement to the development of genetic neighborhoods within large populations, and ultimately aids in interpretations of patterns on larger scales. In this study, I combined well established spatial autocorrelation techniques with novel approaches of contemporary pollen-mediated investigations to address the within population dynamics of gene movement. Finally, I attempted to reconcile the observed contemporary and relatively short-term patterns of genetic structure with the species’ recent history (i.e. since European colonization) and with the patterns accumulated since the Last Glacial Maximum.

REFERENCES

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

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

Dahlgren RMT, Clifford HT, Yeo PF (1985). The families of the Monocotyledons: Structure, evolution and taxonomy. Springer-Verlag, Berlin pp. 123-126, 550.

Delcourt PA, Delcourt RH (1993). Paleoclimates, paleovegetation, and paleofloras during the Late Quaternary. In: Flora of North America. Editorial committee (Eds.), Flora of North America. Oxford University Press, New York, pp.71-94.

Freeman JD (1975). Revision of a *Trillium* subgenus Phyllantherum (Liliaceae). *Brittonia* 27, 1-26.

Ohara M (1989) Life history evolution in the genus *Trillium.* Plant Species Biology 4,1-28.

Wright S (1946) Isolation by distance under diverse systems of mating. *Genetics* 31, 39-59.

Wright S (1951) The genetical structure of populations. *Annals of Eugenics* 15, 323-354.

CHAPTER 2

DISTRIBUTION OF GENETIC DIVERSITY AMONG DISJUNCT POPULATIONS OF THE RARE FOREST UNDERSTORY HERB, *TRILLIUM RELIQUUM*[1]

[1] Gonzales E and JL Hamrick. Submitted to *Heredity*.

ABSTRACT

We assessed genetic diversity and its distribution in the rare southeastern US forest understory species, *Trillium reliquum*. Twenty-one loci were polymorphic (P_S = 95.5%) and the mean number of alleles per polymorphic locus was 3.05. However, heterozygosity was relatively low (H_{es} = 0.120) considering the level of polymorphism observed for this outcrossing species. A relatively high portion of the genetic diversity (29.7%) was distributed among populations. There was no relationship between population size and genetic diversity, and we did not detect significant inbreeding. These results are best explained by the apparent self-incompatibility of this species, its longevity and clonal reproduction. Our results suggest that not only is there little gene flow among extant populations, but that rarity and population isolation in this species is of ancient origins, rather than due to more recent anthropogenic fragmentation. Furthermore, the Chattahoochee River was identified as a major barrier to gene exchange.

Introduction

Natural populations often decline as a result of the deterioration of their habitat caused by anthropogenic disturbance. A number of studies document that species become rare and endangered due to habitat loss, small population size and increased isolation of disjunct populations, or due to the detrimental impact of non-native animals and plants (Gemmill et al., 1998). Attempts to generalize these studies have been made repeatedly, but there are numerous exceptions and confounding factors that impede such endeavours (Gitzendanner and Soltis, 2000; Godt and Hamrick, 2001; Hamrick and Godt, 1996; Karron, 1987).

Dissecting the causes and consequences of rarity is often difficult. Species become rare by several pathways, and rarity is associated with a variety of evolutionary and ecological factors, including habitat specificity, local population size and geographic range (Rabinowitz, 1981). We address the genetic consequences of rarity in disjunct populations of *Trillium reliquum* Freeman (Relict trillium). This monocot species is a member of the Trilliaceae (sensu Dahlgren et al., 1985) or the Melanthiaceae (sensu APG, 1998). It is a forest understory spring ephemeral, with a narrow geographic range but broad habitat specificity – an unusual condition for rarity. *Trillium reliquum* has been only recently described (Freeman, 1975), and is thought to have suffered greatly from human activities during the last three centuries. Although its historical geographic distribution is unknown, extant populations are thought to represent remnants of a formerly more widespread species - hence the specific epithet *reliquum* (relict) (Freeman, 1975). Since its description and listing as endangered by the US Fish and Wildlife Service, about 30 additional populations have been identified. Most are small to moderate in size, isolated from one another, and are primarily restricted to the Fall Line Hills district from the Georgia-South Carolina border to south-western Georgia and south-eastern Alabama (Figure 1).

Several life-history traits of *T. reliquum* may affect the genetic response of its populations to habitat disturbance and loss. *Trillium* species typically require more than 10 years to reach the reproductive stage (Ohara 1989, Jules 1996, pers obs). It is self-incompatible in some populations or has a leaky self-incompatible system (M. Brooks, Ch. Heckel, pers. com.). It is polycarpic and

reproduces infrequently, both by seeds and clonal spread. Individual genets may persist for decades. *T. reliquum* is pollinated by weak-flying insects (Calliphoridae flies and beetles) (M. Brooks, pers. com.). *Trillium*, like many other forest understory herbs, is myrmecochorous; seeds are primarily dispersed by ants or passively by gravity. We have observed yellow jackets (*Vespula vulgaris*) foraging inside mature fruits, and they may act as less common dispersers. Similarly, Jules (1996) reported *V. vulgaris* dispersing seeds of the related species, *T. ovatum*. Populations of *T. reliquum* are typically located along rivers with individual plants scattered on flood plains and bluffs; giving rise to the possibility that seeds may, in rare instances, be moved along watersheds during seasonal floods.

In this study, we examine the distribution of genetic variation within and among populations of *T. reliquum* at several spatial scales. Additionally, we surveyed populations of its common congener, *T. cuneatum,* in their sympatric geographic range to yield insights into the history of population distribution and the genetic consequences of fragmentation. Such comparisons are useful because populations of both species were subjected to the same human pressures in areas of sympatry, and they have similar life history characteristics, habitats, pollinators and seed dispersers. Comparisons are based on the expectation that if *T. reliquum* populations were disjunct prior to anthropogenic fragmentation, the genetic structure of its populations would be greater than that of the more common (albeit also fragmented) *T. cuneatum* in the same region. In addition, we analyzed relationships between genetic diversity and population size, geographic distance and watershed association. Finally, we ask whether small, isolated populations experience more inbreeding.

Populations of *T. reliquum* are widely scattered, disjunct, presumably remnant sites of a previously more common, continuously distributed species (Freeman 1975). Population genetics theory predicts that such populations, if isolated by recent habitat fragmentation due to anthropogenic development, would retain the genetic "footprint" of this history and we should detect an isolation-by-distance pattern. Alternatively, if populations were historically disjunct (i.e. before European settlement), they should be significantly differentiated with no discernable pattern of

genetic structure. The overall distribution of genetic diversity should be haphazard due to long-term isolation without any apparent gene flow among populations. Additionally, if *T. reliquum* populations were historically isolated, we might expect to find evidence of unique or otherwise rare alleles at relatively high frequencies in isolated populations. Combined, species level statistics estimating genetic diversity within this species are expected to be considerably higher than mean population values. We use genetic diversity comparisons with its more common congener, *T. cuneatum*, to gain further insights into the history of rarity of *T. reliquum*. If *T. reliquum* was *common* prior to European colonization and agricultural spread, we should see similar genetic differentiation among populations of both species since both were exposed to similar anthropogenic pressures. Alternatively, if *T. reliquum's* rarity is more ancient, we would expect stronger differentiation among remnant populations of *T. reliquum* relative to that for *T. cuneatum.*

Finally, we also address the question of associations between the density of populations and genetic diversity. The north-western portion of *T. reliquum's* geographic distribution has a higher density of populations than the rest of its range (Figure 1). As a result, we expect to find a relationship between inter-population distance and the ability to retain genetic diversity in this region, particularly when allele frequencies and heterozygosity are considered. Closely distributed populations may better counteract genetic drift than widely separated, disjunct populations. In addition, smaller populations should experience greater risk of genetic diversity loss, both in terms of allelic richness and heterozygosity as well as higher levels of inbreeding.

The close proximity of *T. reliquum* population*s* to rivers may facilitate rare seed dispersal events and may provide corridors for pollinators to follow. Consequently, populations within the same river basin should be more genetically similar than nearby populations belonging to different watersheds. A few investigations of watershed influence on genetic relationships of populations have been conducted. While compelling evidence for such affinities was found in *Mimulus caespitotus,* (Ritland 1989) and *Primula sieboldii* (Kitamoto et al 2005), Barrett et al (2004) found less clear evidence supporting the hypothesis of genetic similarities among populations within river basins of a rare endemic monocot, *Narcissus longispathus*.

Material and Methods

Sampling

We sampled 48 plants (at least 10 m apart to avoid collecting clonal individuals) from each of 22 *T. reliquum* sites, representing two-thirds of all known populations throughout its geographic range (Figure 1). Samples were collected at the peak of flowering. Although we did not quantify precise population sizes, we grouped populations according to their estimated number of flowering individuals at the time of sampling into three relative size classes: small (less than 50 flowering individuals), moderate (50-200 flowering individuals), and large (more than 200 flowering individuals). Additionally, for comparative purposes, we collected samples from nine *T. cuneatum* populations from approximately the same geographic area; in five cases (KCNCH, LLCH, MCH, MCO & RCNCH) individuals of both species occur intermingled within the same site; the remaining four *T. cuneatum* populations are monospecific; two of these (TGA, HGA) are located near *T. reliquum* populations, and the final two (CHAT, SAL) are more distant to ensure comparable spatial sampling for both species. Precise geographic coordinates were recorded for each site, which allowed us to determine distances between all pairs of populations using *ArcView 3.3* software (ESRI 2002). Distances between *T. reliquum* populations ranged from (1.2 - 362.5 km) with a mean of 130 km (SD = 77.8 km). Similarly, distance between *T. cuneatum* sites ranged from 6.2 to 306 km, with a mean distance of 115.2 km (SD = 88.9 km). Detailed location information may be requested from the Georgia Department of Natural Resources. Voucher specimens are deposited in the University of Georgia Herbarium (GA).

Genetic analyses

We transported leaf samples on ice to the laboratory and crushed them within 24 hours using a mortar and pestle. An extraction buffer (Wendel and Parks, 1982) was added to solubilize and stabilize enzymes. The extract was absorbed onto chromatography paper wicks and stored at –70^{o}C until electrophoretic analyses. We used starch gel electrophoresis to determine allozyme diversity. We resolved a subset of the following enzymes for each species: amino acid transferase (AAT),

diaphorase (DIA), fluorescent esterase (FE), glutamate dehydrogenase (GDH), isocitrate dehydrogenase (IDH), leucine- amino peptidase (LAP); malate dehydrogenase (MDH), menadione reductase (MNR), 6-phosphogluconate dehydrogenase (6-PGDH), peroxidase (PER), phosphoglucoisomerase (PGI), phosphogluco-mutase (PGM), shikimate dehydrogenase (SKDH), triose-phosphate isomerase (TPI), UTP-glucose-1-phosphate (UGPP). For *T. reliquum*, we employed the following four gel-electrode buffer combinations to resolve 22 loci on 11% starch gels: *Buffer 7:* (AAT); *Buffer 8-:* (DIA, FE-1, FE-2, FE-3, GDH, LAP, MNR , TPI-1, TPI-2), *Buffer 4:* (IDH, MDH-1, MDH-2, MDH-3, 6-PGDH-1, 6-PGDH-3, SKDH-1, SKDH-2, UGPP); and *Buffer 6: (*PER, PGI-1, PGI-2). For *T. cuneatum*, we employed five gel-electrode combinations and resolved the following 20 loci: *Buffer 8*-: (AAT-2, AAT-3, GDH); *Buffer 4*: (IDH, SKDH-1, SKDH-2, UGPP-1, UGPP-2); *Buffer 6:* (DIA, FE-1, FE-2, FE-3, FE-4, PER, TPI-1, TPI-2); *Buffer 11:* (MDH-1, MDH-2, 6-PGDH, PGM). Stain recipes for AAT, DIA and MNR are given in Cheliak and Pitel (1984); UGPP is given in Manchenko (1994). All other stain and buffer recipes were taken from Soltis *et al.* (1983). For enzymes with more than one locus, loci were numbered sequentially with the number one assigned to the most anodal locus.

Statistical analyses

We calculated genetic diversity statistics for both species (as described in Hamrick and Godt, 1989) and for each population (as described in Hedrick, 1985) using a statistical program developed by M. D. Loveless and A. Schnabel. These measures included the percentage of polymorphic loci (*P*), the mean number of alleles per locus (*A*), and per polymorphic locus (*AP*), the effective number of alleles (A_e), observed (H_o) and expected (H_e) heterozygosity. We also calculated a coefficient of variation (CV_{He}) for expected population heterozygosity values (as in Schoen and Brown, 1991). Subscript *s* indicates species values, whereas *p* indicates population values. Deviations from Hardy-Weinberg expectations were examined for each polymorphic locus within each population by calculating Wright's fixation index (Wright, 1922). Fixation indices were tested for significance using Chi-square (Li and Horvitz, 1953).

We estimated population divergence using Nei's gene diversity statistics (Nei, 1973; 1977). This statistic (G_{ST}) estimates the proportion of the total genetic diversity (H_T) found among populations for each polymorphic locus; G_{ST} values were averaged across loci to obtain an overall estimate of population divergence. In addition, we calculated mean G_{ST} values for loci with H_T greater than 0.10 because loci with one common allele and remaining rare alleles are often not informative concerning genetic structure. Each G_{ST} value was tested for significance by Chi-square (Workman and Niswander, 1970). Nei (1977) demonstrated that G_{ST} is equivalent to a multi-allelic F_{ST} (Wright, 1951). Chakraborty and Danker-Hopfe (1991) have also shown that these two indices are empirically equivalent to Weir and Cockerham's (1983) θ when sample sizes are equal and a large number of populations are analyzed as is the case for this study. Genetic identity (*I*) and distance (*D*) measures were also calculated for each pairwise combination of populations (Nei, 1972). To graphically portray genetic relationships between *T. reliquum* populations, we used genetic identities to construct a UPGMA phenogram as well as a Neighbour Joining tree using NTSYS-pc 2.1 software (Rohlf, 1992).

Finally, to gain insights into patterns of genetic diversity across the species range, we subdivided *T. reliquum* populations based on three factors. We investigated the effects of inter-population distance, population size and watershed association on the populations' ability to retain genetic variation. In addition, we used Rousset's (1997) measure of genetic distance $F_{ST}/(1-F_{ST})$ to analyze associations between geographic distance and genetic distance, using a reduced major axis regression (Sokal and Rohlf, 1981); this test assesses whether the pairwise population genetic differentiation matrix is correlated with the pairwise geographic distance matrix. Significance of the correlation between genetic and geographic distance was tested with a Mantel test. Analyses were executed using Isolation by Distance software (Bohonak, 2002). We also subdivided the range of *T. reliquum* according to the density of populations per geographic area (Figure 1) to investigate the effect of inter-population distances on the distribution of genetic variation. The northwestern portion of *T. reliquum's* range has a higher density of known populations with a mean inter-population distance of 21.1 km (additional, unsampled populations exist in this

region), while the remaining populations are more scattered with a mean inter-population distance of 138.3 km.

RESULTS

Genetic diversity within the species

Twenty-one of the 22 loci resolved (95.5%) were polymorphic in at least one *T. reliquum* population (Table 1); a slightly lower number, 17 of the 20 loci (85 %) were polymorphic in *T. cuneatum* populations (Table 2). At the species level, we detected a lower number of alleles per polymorphic locus (AP_s) and alleles per locus (A_s) in *T. reliquum* than in *T. cuneatum* (*T. reliquum:* AP_s = 3.05, A_s = 2.95*; T. cuneatum:* AP_s = 3.71, A_s =3.30*).* For *T. reliquum*, the mean effective number of alleles (A_{es}= 1.16) was low considering its P_s, AP_s and A_s values, and again, it was lower than the values for *T. cuneatum* (A_e = 1.39). Similarly, *T. reliquum* genetic diversity was also low (H_{es} = 0.120) relative to its level of polymorphism and the number of alleles detected (Table 1) and relative to *T. cuneatum* (H_{es} = 0.217).

Genetic diversity within populations

We detected a more striking genetic diversity differences between the two species at the population level. The mean percentage of polymorphic loci (P_p) averaged 33.9% across all *T. reliquum* populations, while *T. cuneatum* populations average 58.3% polymorphic loci. For *T. reliquum*, all of the within population statistics had lower values than *T. cuneatum*. The mean number of alleles per locus (A_p) was 1.41 and 1.89 respectively with a mean of 2.19 and 2.54 alleles per polymorphic locus (AP_p), and 1.11 and 1.36 effective alleles per locus (A_{ep}). In *T. reliquum*, mean genetic diversity (H_{ep}) and observed heterozygosity (H_{op}) were 0.069 and 0.070, respectively (Table 1), while genetic diversity in *T. cuneatum* populations was much higher (H_{ep} = 0.192, H_{op} = 0.183) (Table 2). Observed genotype frequencies conformed to Hardy-Weinberg expectations for 88.4% of the loci in all *T. reliquum* populations. We detected 19 instances (11.6%) of F_{IS} values that differ significantly from zero ($p < 0.05$); three cases with heterozygote excesses and 16 cases with heterozygote deficiencies. Significant F_{IS} values were mostly attributable to loci with

H_S < 0.10 (17 of 19 cases), i.e. loci that are not very informative concerning genotype equilibrium distributions. The mean F_{IS} value across all loci and populations was 0.007 and did not differ significantly from zero.

Genetic diversity among populations

Genetic diversity varied substantially more among *T. reliquum* populations than among *T. cuneatum* populations. Allele frequencies were significantly different among *T. reliquum* populations for 20 of 21 polymorphic loci, and for 16 of 17 polymorphic loci among *T. cuneatum* populations ($p < 0.001$). For *T. reliquum*, values of P_p ranged from 4.6% to a high of 59.1%, and H_{ep} values ranged from 0.020 to 0.126 ($CV = 0.468$), while we observed much less variation among *T. cuneatum* populations (range of P_p = 55% - 75%) and higher H_{ep} (0.152 - 0.221, $CV = 0.140$). *T. reliquum* populations at either margin of the geographic range (i.e. eastern-most and western-most) had the highest proportion of polymorphic loci. A statistically significant trend of decreasing heterozygosity from east to west ($r = 0.436$, $p < 0.001$) was detected with the notable exception of the three western-most populations (LCCH, WGRCH & FLCH). These three populations are located in Alabama, along the western banks of the Chattahoochee River. A similar trend was observed for the effective number of alleles (A_{ep}) while P_P, AP_p and A_p varied haphazardly and without obvious trends. We did not observe such a pattern in *T. cuneatum*. We identified a relatively high, but comparable number of private alleles (12 for *T. reliquum,* mean frequency = 0.089; 14 for *T. cuneatum,* mean frequency = 0.067). A disproportionate number (six) of the *T. reliquum* private alleles were found in the three Alabama populations (LCCH, WGRCH and FLCH) west of the Chattahoochee River. Similarly, five of the *T. cuneatum* private alleles were detected in the two Alabama populations (LCCH, SAL).

The proportion of total genetic variation attributable to differentiation among *T. reliquum* populations (G_{ST}), was 0.297; values for individual loci ranged from 0.010 to 0.797. Grouping populations by watersheds is responsible for 45% of the differentiation among populations (i.e. G_{ST} among watersheds = 0.133). We also calculated G_{ST} values for the 10 loci with $H_T > 0.10$

resulting in an unusually high G_{ST} value (0.436) for an outcrossing species. Mean genetic identity was moderate (I = 0.942) with a fairly broad range of values (0.841-0.999). In sharp contrast, only 9.2 % of total genetic variation was distributed among populations of *T. cuneatum*; when loci with $H_T > 0.10$ are considered, G_{ST} values increased slightly to 0.11.

Geographic patterns of genetic variation in *T. reliquum*

Geographically close populations exhibited a weak trend of higher genetic similarities. Although the isolation by distance analysis resulted in a statistically significant correlation between log transformed geographic distance and pairwise $F_{ST}/(1-F_{ST})$ measures, only a very small portion of the among population differentiation was explained by geographic distance (r^2 = 0.083, p < 0.004). In spite of the overall "isolation by distance" trend, some populations separated by short geographic distances had relatively large genetic distance values. Such population pairs invariably belonged to different watersheds (e.g. JDOCM & MCH).

We used Nei's genetic distances for UPGMA and Neighbor Joining classification analyses; only the UPGMA tree is presented (Figure 2). Both phenograms depicted distinct groups of populations belonging to the same river basin (e.g. the North Chattahoochee watershed in Georgia), while other populations (e.g. MZF & EPOCM) clustered in the phenogram despite no apparent watershed or other geographical association. The three Alabama populations (LCCH, FLCH and WGRCH) located along the western bank of the Chattahoochee River did not cluster with nearby populations on the Georgia side of the river. Moreover, the Georgia FGCH site is included in the clade of populations belonging to the northern portion of this watershed, rather than with nearby populations WGRCH and FLCH just to the west across the river.

To further investigate patterns of genetic variation, we divided the geographic range of *T. reliquum* into two sub-regions: The north-western portion of the distribution (Figure 1) contains seven populations with relatively small geographic distances among them; and the east-central portion with 15 populations more widely scattered across *T. reliquum's* range. The mean percentage of polymorphic loci per population (P_p) detected in the high density area was 36.4%, while

the mean value for more widely scattered populations was 30.4%. The mean number of alleles per locus (A_p) in the high density area was 1.42 (Table 2), slightly higher than in the low density region (A = 1.38). Although populations in close proximity retained a higher percentage of polymorphic loci and more alleles per locus, not all measures in this region indicated higher genetic diversity. The mean number of alleles per polymorphic locus (AP_p) was 2.14 in the high density region, while populations separated by larger distances average 2.21 alleles per polymorphic locus. Populations from areas of high density had on average 1.07 and the more scattered populations 1.13 effective number of alleles, and genetic diversity (H_e) of 0.05 and 0.08 respectively (Table 3).

We found an insignificant inverse relationship between the mean percentage of polymorphic loci and population size. Small populations (< 50 flowering individuals) had the highest proportion of polymorphic loci (P = 34.9%), moderate populations (50-200 flowering individuals) 33.3% of polymorphic loci, and large populations (> 200 flowering individuals) contained the lowest percentage of polymorphic loci (33.0%). There were no significant differences among the population size categories for any of the other genetic diversity parameters (Table 3).

Discussion

Genetic diversity within the species

Trillium reliquum maintains high levels of genetic polymorphism despite its rarity and disjunct distribution (Table 1, Table 4). The strikingly incongruent characteristic, however, is the low genetic diversity observed for *T. reliquum* (H_{es}=0.120) considering that 95% of the loci are polymorphic with 3.05 alleles per polymorphic locus (Table 1). This discordance between genetic diversity (H_e) and polymorphism is best explained by the relatively low effective number of alleles per locus (A_e = 1.16). This low value results from the high number of low frequency alleles detected for many loci. Approximately 29% of the polymorphic loci have overall heterozygosity (H_T) values below 0.05 and another 24 % have H_T values less than 0.10.

This conclusion is further corroborated by the observation that *T. reliquum* has a slightly higher proportion of polymorphic loci than its more common congener, *T. cuneatum*, while the remain-

ing population genetic statistics (*AP, A,* A_e, H_e) were appreciably higher for *T. cuneatum* (Table 4). Although rare species usually maintain less polymorphism, this generalization does not always hold (Godt and Hamrick 2001, Karron 1987). In fact, Gitzendanner and Soltis (2000) found that approximately 20% of the rare species reviewed contain equal or higher polymorphism than their more common congeners. Our results, however, are consistent with the majority of the congeneric comparisons reviewed by Gitzendanner and Soltis (2000) since they demonstrate a large discrepancy in genetic diversity between *T. reliquum* (H_{es} = 0.120) and *T. cuneatum* (H_{es} = 0.217).

Genetic diversity within populations

In contrast to the high polymorphism within species, *T. reliquum* populations maintain less genetic diversity, relative to its widespread congener. In this regard, *T. reliquum's* population values are more typical of other rare herbaceous perennials (Table 4). However, while rare species often exhibit large discrepancies between mean observed and expected heterozygosities (Ellstrand and Elam, 1993), we observed few such differences. We detected little inbreeding in any of the populations regardless of population size or isolation from nearest neighbouring populations (F_{IS} = 0.007). The most likely explanation for the low inbreeding observed is the apparent self-incompatible or leaky SI breeding system of the species (M. Brooks, pers. com.).

Although *T. reliquum* is geographically restricted, a relatively large portion of the total genetic diversity was partitioned among its populations, (e.g. G_{ST} = 0.279 for all loci and 0.436 for loci with H_T>0.1). This strong genetic differentiation is in sharp contrast to its more widespread congener sampled from a comparable geographic range. Most of the total genetic variation in *T. cuneatum*, is contained within populations as indicated by the low G_{ST} value (0.092; Table 4) over all polymorphic loci, which increased only slightly (to 0.110) for loci with H_T>0.10. Several studies of allopatric *Trillium* species reported variable levels of genetic diversity and its distribution (e.g. Bayer et al 1986, Griffin and Barrett 2004, Tomimatsu and Ohara 2002, Whitkus et al 1987). G_{ST} values (or its analogs F_{ST} and θ) reported in these studies ranged from 0.095 to 0.35. However, populations of these *Trillium* species have been subjected to vastly different histories;

some studies sampled from partially or completely glaciated regions in North America, some contain multiple glacial-refugial lineages and for some monomorphic loci were not analyzed. As a result comparisons of these species with our results for *T. reliquum* are not appropriate to address our question of history of rarity in this species.

Genetic diversity (H_{ep}) among the *T. reliquum* populations ranged from 0.020 to 0.120; (CV = 0.468). Such variation in genetic diversity and proportion of polymorphic loci (P_p) among populations is characteristic either of species with naturally disjunct ranges (Hamrick, 2004; Hamrick and Godt, 1996), or self-pollinated species (Schoen and Brown, 1991). However, even in perennial herbs with widely disjunct populations, (e.g. *Tradescantia hirsuticaulis*, (Godt and Hamrick 1993) a granite outcrop endemic with a similar geographic range to *T. reliquum*, or *Sarracenia leucophyla*, (Wang et al 2004), another relatively rare southeastern perennial herb, such extreme population variability in genetic diversity is uncommon. In contrast, Tomimatsu and Ohara (2004) reported comparable levels of both species and population genetic diversity in *Trillium camschatcense* in eastern Hokaido, Japan. In their study area, previously large, continuous populations had been fragmented into smaller remnants. Their investigation revealed that 91 % of its loci were polymorphic (range 18 - 82 %); and H_{ep} was relatively low (0.079; range 0.035 - 0.133). The relatively low level of differentiation among populations of *T. camschatcense* (F_{ST} = 0.13) was similar to that of *T. cuneatum*.

Contrary to the *T. camschatcense* investigation (Tomimatsu and Ohara 2004) and to our expectations, our study did not reveal a significant relationship between the demographic characteristics of *T. reliquum* populations and the maintenance of genetic variation. We expected such relationships based on population genetics theory, which predicts that larger populations are more likely to have higher heterozygosity and greater allelic richness. In addition, we expected heterozygosity within populations and genetic similarity measures to be positively correlated with the number and density of populations within a region and with geographic distance, presumably reflecting a greater opportunity for historical gene flow and the greater susceptibility of isolated

populations to genetic erosion and inbreeding. While populations from the densely populated region maintained a slightly higher percentage of polymorphic loci and more alleles per locus, the mean effective number of alleles and heterozygosity were higher in the more disjunct populations. Even more surprising is the weak trend of increasing polymorphism with decreasing population size. We also failed to detect significant associations between population size, allelic richness, and heterozygosity. Moreover, we did not identify differences between expected and observed heterozygosity regardless of population size or isolation. Even in a predominantly outcrossing species, one would expect inbreeding to increase by $1/2N_e$ per generation. This supports the argument that present-day population sizes may not be predictive of the genetic composition of populations for a long-lived species.

In the absence of detailed historical information on each population's demographic history (such as recent population bottlenecks), our results suggest that current population size and interpopulation distance may be largely irrelevant to the maintenance and partitioning of *T. reliquum* genetic diversity. Instead, other factors (e.g. clonal reproduction, longevity) might have more important influences on the genetic structure that we see among present-day populations. This notion has been further supported by the isolation by distance analysis. Although we detected a statistically significant relationship between log transformed genetic and geographic distances ($p < 0.004$), the isolation by distance analysis explains only 8.3% of genetic differentiation observed between populations. Culley and Grubb (2003) also detected no relationship between genetic and geographic distance for historically fragmented populations of the perennial forest herb, *Viola pubescens*.

The UPGMA analysis (Figure 2) suggests that associations within watersheds may be an important factor shaping genetic structure in *Trillium reliquum*. In particular, the UPGMA phenogram displays a distinct clade of populations belonging mostly to the western Georgia region. This result at first seems surprising and contradictory. This clade also includes one population (FGCH) separated by a large geographic distance, but located within the same river basin

(eastern banks of the Chattahoochee River in Georgia). In contrast, population LCCH, which is located on the western bank of the Chattahoochee River in Alabama in close geographic proximity to the western Georgia populations, is not included in this clade. Furthermore, the LCCH population is more similar to the two other Alabama sites within the same watershed, albeit separated by larger geographic distances. Other watersheds are represented by two to three populations each and do not provide such a persuasive pattern of watershed alliances. Although there are some exceptions (such as population ATTC and TCTC in the Oconee River basin), some populations seem haphazardously placed within a clade.

The Chattahoochee River may constitute a barrier for pollen and seed dispersal between populations on its eastern and western banks. The magnitude of genetic divergence between populations located east and west of the Chattahoochee River make it possible that these two groups of populations are from separate glacial refugia; a hypothesis we are currently investigating using cpDNA sequences. Such a pattern is in concordance with studies of numerous animal taxa (e.g. *Geomys pinetis, Lepomis punctatus, Amia calva* and *Sternotherus minor*) (Avise, 2000) and a more recently documented fungal (*Septobasidium*) study (T. Turner, unpublished). Geographic analyses of mitochondrial DNA gene trees of these taxa document consistent agreement between divergent branches of gene trees and two geographic regions: the Atlantic and Gulf zones, divided by the Chattahoochee River basin. Such integrated molecular phylogeographic studies suggest that shared biogeographic factors may have shaped the distributional boundaries and contemporary genetic architecture of multiple co-distributed species.

The proposition that *T. reliquum* populations originated from separate glacial refugia on opposite sides of the Chattahoochee River is further supported by the disproportionate number (50%) of private alleles in the Alabama populations, and by the unusual and abrupt change in the trend of decreasing genetic diversity (H_{ep}) from east to west across Georgia. This trend of declining heterozygosity was highly significant ($r^2 = 0.190$, $p < 0.001$), and is even stronger when the three Alabama populations were excluded from the regression analysis ($r^2 = 0.497$). The overall trend of declining heterozygosity was unexpected, especially considering the inconsistent patterns for

other population genetics statistics. Furthermore, contrary to population genetics theory and the majority of empirical evidence, populations with the highest genetic diversity are located in the eastern and western extremes of the geographic range, rather than in the more centrally located populations.

History of rarity – inference from genetic data

High levels of genetic structure indicate that extant *T. reliquum* populations have historically experienced little genetic interchange and that there is no appreciable contemporary gene flow among these remnant populations. Populations are mostly isolated by large geographic distances and inhospitable habitat (due, in part, to human development). The high population divergence, the large number of private and rare alleles at relatively high frequencies (0.089), and significant among population heterogeneity in common allele frequencies suggest that genetic drift has historically been a major influence in shaping the genetic divergence among *T. reliquum* populations. This does not appear to be the case for the more continuously distributed *T. cuneatum*.

While we do not dispute that *T. reliquum* populations have suffered greatly due to anthropogenic disturbances, our results are not consistent with the apparent recency of the severe impact of habitat destruction. Although sympatric populations of the more common *T. cuneatum* were presumably exposed to similar anthropogenic pressures, they exhibit higher intra-population genetic variation and considerably lower inter-population divergence. Ancient isolation affecting genetic diversity levels and patterns is the most likely explanation for the population divergence observed among contemporary *T. reliquum* populations. Our data suggest that the rarity of *T. reliquum* is, in all probability, of more ancient origin than previously proposed, and that this species historically existed as isolated populations long before European settlement in the southeastern US.

For some species, there appears to be a connection between levels of genetic diversity and ecological conditions. For example, in both California and Spain, inland *Avena barbata* populations maintain less genetic diversity than coastal populations (Allard et al 1978*).* In *Lycopersicon pimp-*

inallifolium, the smallest populations correspond to the lowest diversity (Rick et al, 1977). For species with comparatively less variation among populations in gene diversity and N_e, surveys of genetic variation may be less important. Nonetheless, there are species, *T. reliquum* among them, for which such ecological predictors of genetic diversity are lacking. No clear correlation has emerged from our study for a relationship between population size, isolation, or marginal versus centrally-located populations and genetic diversity within populations. In such cases, information about the genetic variation of individual populations, critical in guiding conservation efforts, may only be derived directly from genetic surveys. Rabinowitz (1981) proposed that while natural selection cannot select for rarity, it may favour traits which offset the disadvantage of being rare. Clonal reproduction and considerable longevity of individual *T. reliquum's* ramets (many decades), combined with an outcrossing mating system, promote maintenance of genetic variation within populations and their viability, even if the effective size of the populations declines.

Our results emphasize the importance of genetic surveys for sound management practices and raise several issues concerning conservation strategies. One implication of our study is that *T. reliquum* might be viewed as a species composed of a number of ancient and genetically diverse populations. These populations represent units with a limited subset of genetic diversity and evolutionary potential. Such a viewpoint has relevance for the conservation of genetic resources and is important for the design of sampling strategies intended to conserve the species' genetic diversity. Our results suggest that the Alabama and nearby Georgia populations may represent different historical lineages, further reinforcing the need to protect a larger number of populations to retain genetic diversity in this species.

ACKNOWLEDGMENTS

Numerous people at the Georgia, Alabama and South Carolina Departments of Natural Resources and US Fish and Wildlife Service (especially Al Scholtz, Johnny Stowe and Steve Paris) were instrumental in the successful completion of this study. We particularly wish to thank Tom Patrick, who helped with the location of all populations as well as field sampling. We are also grateful to private land-owners for their collaboration and sampling assistance. We thank Mary Jo

Godt, ShuMei Chang, Jennifer Cruse-Sanders, John Paul Schmidt and two anonymous reviewers for valuable comments on data analyses and the manuscript. Financial support was provided by the Georgia Museum of Natural History (Joshua Laerm Award), the Georgia Native Plant Society and the Palfrey Fund, Plant Biology Department, University of Georgia.

References

Allard RW, Miller RD, Kahler AL (1978). The relationship between degree of environmental heterogeneity and genetic polymorphism. *Verhandelingen der Koninklijke Nederlandse Akademie van Wetenschappen, Afdeling Naturokunde Tweede Reeks* **70:** 49-69.

Angiosperm Phylogeny Group [APG] (1998). An ordinal classification for the families of flowering plants. *Annls Mo Bot Gard* **85:** 531-553.

Avise JC (2000). *Phylogeography* Harvard University Press, Cambridge, Massachusetts.

Barrett SCH (2004). Mating patterns and genetic diversity in the wild Daffodil *Narcissus longispathus* (Amaryllidaceae). *Heredity* **92**: 459-465.

Bohonak AJ (2002). Computer note: IBD (Isolation by distance) A program for analyses of isolation by distance. *J Hered* **93**: 153-154.

Chakraborty R, Danker-Hopfe, H (1991). Analysis of population structure: a comparative study of different estimators of Wright's fixation indices. In: *Handbook of Statistics*. Vol. 8. Edited by CR Rao and R Chakraborty. Elsevier Science Publishers BV, New York, pp 203-254.

Cheliak WM, Pitel JA (1984). Techniques for starch gel electrophoresis of enzymes from forest tree species. Information report P1-X-42. Petawawa National Forestry Institute. Canadian Forestry Service, Agriculture, Chalk River, Ontario.

Culley TM, Grubb Jr. TC (2003). Genetic effects of habitat fragmentation in *Viola pubescens* (Violaceae), a perennial herb with chasmogamous and cleistogamous flowers. *Mol Ecol* **12:** 2919-2930

Dahlgren RMT, Clifford HT, Yeo PF (1985). The families of the Monocotyledons: Structure, evolution and taxonomy. Springer-Verlag, Berlin pp. 123-126, 550.

Environmental Systems Research Institute [ESRI] (2002) *ArcView 3.3*

Ellstrand NC, Elam DR (1993). Population genetic consequences of small population size. Implications for plant conservation. *Annu Rev Ecol Syst* **24**: 217-242

Freeman JD (1975). Revision of a *Trillium* subgenus Phyllantherum (Liliaceae). *Brittonia* **27**: 1-26.

Gemmill CEC, Ranker TA, Ragone D, Perlman SP, Wood KR (1998). Conservation genetics of

the endangered endemic Hawaiian genus *Brighamia* (Campanulaceae). *Am J Bot* **85**: 528-539.

Gitzendanner M, Soltis PS (2000). Patterns of genetic variation in rare and widespread plant congeners. *Am J Bot* **87**: 783-792.

Godt MJW, Hamrick JL (1993). Genetic diversity and population structure in *Tradescantia hirsucalis* (Commelinaceae) *Am J Bot* **80**: 959-966.

Godt MJW, Hamrick JL (1998). Allozyme diversity in the grasses. In: *Population Biology of the Grasses*. (ed. GP Cheplic), pp. 11-29. Cambridge Univ Press, Cambridge, UK

Godt MJW, Hamrick JL (2001). Genetic diversity in rare Southeastern plants. *Natl Areas J* **21**: 61-70.

Griffin SR, Barrett SCH (2004). Post-glacial history of *Trillium grandiflorum* (Melanthiaceae) in Eastern North America. *Am J Bot* **91**:465-473.

Hamrick JL (2004). Response of forest trees to global environmental changes. *Forest Ecol and Mang* **197:**323-335.

Hamrick JL, Godt MJW (1989). Allozyme diversity in plant species. In: *Plant population genetics, breeding and genetic resources* (eds. Brown AHD, Clegg MT, Kahler AL, Weir BS), pp. 43-63. Sinauer, Sunderland, Massachusetts.

Hamrick JL, Godt MJW (1996). Conservation genetics of endemic plant species. In: *Conservation Genetics: Case histories from Nature* (eds. Avise JC, Hamrick JL), pp. 281-304. Chapman and Hall.

Hedrick PW (1985). *Genetics of populations.* Jones and Bartlett, Boston.

Jules E (1996) Yellow jackets (*Vespula vulgaris*) as a second seed dipsperser for the myrmecochorous plant, *Trillium ovatum*. *Am M Nal* **135**:367-369.

Karron JD (1987). A comparison of levels of genetic polymorphism and self-incompatibility in geographically restricted and widespread plant congeners. *Evol Ecol* **1**: 47-58.

Kitamoto N, Honjo M, Ueno S, Takenaka A, Tsumura Y, Washitani I, Ohsawa R (2005). Spatial genetic structure among and within populations of *Primula sieboldii* growing beside separate streams. *Mol Ecol* **14:** 149-157.

Li CC, Horvitz DG (1953). Some methods of estimating the inbreeding coefficient. *Am J Human Genet* **5**: 107-117.

Manchenko GP (1994). *Handbook of detection of enzymes on electrophoretic gels. CRC Press,* Ann Harbour, MI. CRC Press, Ann Harbour, MI.

Nei M (1972). Genetic distance between populations. *Am Nat* **106**: 283-292.

Nei M (1973). Analysis of gene diversity in subdivided populations. *Proc Natl Acad Sci USA* **70**: 3321-3323.

Nei M (1977). F-statistics and analysis of gene diversity in subdivided populations. *Annls Human Genet* **41**.

Ohara M (1989). Life history evolution in the genus *Trillium*. *Plant Spec Biol* **4**: 1-28

Rabinowitz D (1981). Seven forms of rarity. In: *Biological Aspects of Rare Plant Conservation* (ed. Synge H), pp. 205-217. John Wiley & Sons Ltd.

Rick CM, Fobes JF, Holle M (1977) Genetic variation in *Lycopersicum pimpinellifolium*: evidence of evolutionary change in mating systems. *Pl Syst Evol* **127:** 139-170.

Ritland K (1989). Genetic differentiation, diversity, and inbreeding in the mountain monkeyflower (*Mimulus caespitosus*) of the Washington Cascades. *Can J Bot* **67**: 2017-2024.

Rohlf FJ (1992). NTSYS-pc. Numerical taxonomy and multivariate analysis system. Exeter Software, Setauket, NY.

Rousset F (1997). Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics* **145**: 1219-1228.

Schoen DJ, Brown AHD (1991). Intraspecific variation in population gene diversity and effective population size correlates with the mating system in plants. *Proc Natl Acad Sci USA* **88**: 4494-4497.

Sokal RR, Rohlf FJ (1981). *Biometry*, 2nd edn. Freeman, New York.

Soltis DE, Haufler CH, Darrow DC, Gastony GJ (1983). Starch gel electrophoresis of ferns: a compilation of grinding buffers, gel and electrode buffers, and staining schedules. *Am Fern J* **73**: 9-27.

Tomimatsu H, Ohara M (2003). Genetic diversity and local population structure of fragmented populations of *Trillium camschatcense* (Trilliaceae). *Biol Conserv* **109:** 249-258.

Wang ZF, Hamrick JL, Godt MJW (2004) High genetic diversity in *Sarracenia leucophylla* (Sarraceniaceae), a carnivorous wetland herb. *J Hered* **95**: 234-243.

Wendel JF, Parks CR (1982). Genetic control of isozyme variation in *Camillia japonica* L. *J Hered* **73**: 197-204.

Workman PL, Niswander JD (1970). Population studies on southwestern Indian tribes. II. Local genetic differentiation in the Papago. *Am J Human Genet* **22**: 24-49.

Wright S (1922). Coefficient of inbreeding and relationships. *Am Nat* **56**: 330-338.

Table 2.1: Genetic diversity statistics for *Trillium reliquum* populations. Populations are ordered in East-West direction and grouped according to their association with watersheds. The last two letters in each population's name designate a state (SC=South Carolina, GA=Georgia, AL=Alabama). East-west division is indicated by a line between GA and AL populations.

Population	*Drainage*	*Pop Size*	P_p (%)	Total A_p	AP_p	A_{ep}	H_{ep}	H_{op}	*Mean I*
NAS-SC	Savannah	L	45.5	35	2.30	1.21	0.120	0.119	0.957
ALDS-GA	Savannah	L	36.4	34	2.50	1.23	0.126	0.119	0.927
OMTC-GA	Oconee	S	45.5	34	2.20	1.14	0.088	0.113	0.893
ATTC-GA	Oconee	S	45.5	36	2.40	1.21	0.126	0.123	0.937
TCTC-GA	Oconee	S	31.8	30	2.14	1.15	0.089	0.080	0.928
EPOCM-GA	Ocmulgee	M	31.8	31	2.29	1.16	0.092	0.093	0.962
MCOM-GA	Ocmulgee	L	13.6	25	2.00	1.07	0.043	0.045	0.926
JDOCM-GA	Ocmulgee	S	22.7	27	2.00	1.09	0.048	0.045	0.893
FRNF-GA	Flint	S	50.0	34	2.09	1.16	0.100	0.111	0.950
MZF-GA	Flint	M	22.7	28	2.20	1.07	0.041	0.041	0.953
MDF-GA	Flint	M	4.6	23	2.00	1.04	0.020	0.016	0.922
MCH-GA	Chattahoochee	M	40.9	34	2.33	1.09	0.071	0.073	0.965
THCH-GA	Chattahoochee	M	40.9	33	2.22	1.06	0.051	0.045	0.965
JACH-GA	Chattahoochee	M	40.9	33	2.22	1.05	0.040	0.038	0.965
RCNCH-GA	Chattahoochee	L	36.4	31	2.13	1.07	0.050	0.040	0.963
RCSCH-GA	Chattahoochee	S	31.8	29	2.00	1.08	0.058	0.043	0.963
KCNCH-GA	Chattahoochee	S	40.9	32	2.11	1.08	0.057	0.052	0.961
KCSCH-GA	Chattahoochee	S	22.7	27	2.00	1.03	0.026	0.026	0.966
FGCH-GA	Chattahoochee	S	22.7	28	2.20	1.04	0.028	0.030	0.964
LCCH-AL	Chattahoochee	M	59.1	38	2.23	1.13	0.097	0.099	0.918
WGRCH-AL	Chattahoochee	M	22.7	29	2.40	1.14	0.076	0.097	0.923
FLCH-AL	Chattahoochee	M	36.4	31	2.13	1.12	0.067	0.084	0.926

P_p=Percent of polymorphic loci, Total A_p=Total number of alleles per population (including monomorphic loci), AP_p=Mean number of alleles per polymorphic locus, A_e=Mean effective number of allele per polymorphic locus, H_{ep}=Genetic diversity (expected heterozygosity), H_{op}=Observed heterozygosity, I=Genetic identity.

Table 2.2: Genetic diversity statistics for *Trillium cuneatum* populations. Populations were sampled approximately in geographic regions sympatric with *T. reliquum*. The last two letters in each population's name designate a state (GA-Georgia, AL=Alabama).

Population	P_p (%)	Total A_p	AP_p	A_{ep}	H_{ep}	H_{op}	*Mean I*
MCO-GA	55.0	39	2.73	1.38	0.193	0.161	0.968
MCH-GA	50.0	36	2.60	1.80	0.201	0.201	0.958
TGA-GA	65.0	38	2.38	1.33	0.187	0.183	0.967
HGA-GA	75.0	42	2.47	1.41	0.226	0.211	0.954
RCNCH-GA	50.0	34	2.40	1.3	0.152	0.158	0.967
KCNCH-GA	55.0	38	2.82	1.36	0.186	0.177	0.963
CHAT-GA	55.0	33	2.18	1.25	0.154	0.147	0.969
LCCH-AL	60.0	40	2.75	1.39	0.203	0.197	0.946
SAL-AL	60.0	37	2.50	1.42	0.229	0.211	0.962

P_p=Percent of polymorphic loci, Total A_p=Total number of alleles per population (including monomorphic loci), AP_p=Mean number of alleles per polymorphic locus, A_e=Mean effective number of allele per polymorphic locus, H_{ep}=Genetic diversity (expected heterozygosity), H_{op}=Observed heterozygosity, I=Genetic identity.

Table 2.3: Comparison of mean genetic diversity statistics for *T. reliquum* populations based on their inter-population distance, and population size: Area with a high density of populations (mean distance among populations = 21.1 km); and area with low density (mean distance = 138.3 km). Populations were also grouped into relative size categories based on the number of flowering plants.

	N	P_p	AP_p	A_p	A_e	H_{ep}	H_{op}
High density subregion	8	36.4	2.14	1.42	1.07	0.070	0.052
Low density subregion	14	30.4	2.21	1.38	1.13	0.140	0.081
Large populations (>200)	4	33.0	2.23	1.42	1.15	0.085	0.081
Moderate populations (50-200)	9	33.3	2.22	1.42	1.10	0.062	0.065
Small populations (< 50)	9	34.9	2.13	1.40	1.11	0.069	0.069

N= number of sampled populations

Table 2.4: Species level genetic diversity comparisons for *T. reliquum*, its more common congener *T. cuneatum*, south-eastern rare & endemic species, outcrossing perennials, monocots & all seed plants. (Standard errors are in parentheses where available.)

Taxonomic group	***P (%)***	***A***	A_e	H_e	G_{ST}
T. reliquum	95.5	2.95	1.16	0.120	0.279
mean population values (SD)	33.9 (12.8)	1.41 (0.17)	1.11 (0.06)	0.069 (0.032)	-
range	4.6-59.1	1.05-1.73	1.03-1.23	0.126-0.20	-
T. cuneatum	85	3.3	1.36	0.217	0.092
mean population values (SD)	58.33 (3.63)	1.89 (0.15)	1.36 (0.06)	0.183 (0.016)	-
range	50.0-75	1.65-2.10	1.25-1.42	0.152-0.229	-
Rare southeastern plants [a]	46.7 (4.5)	1.87 (0.13)	-	0.123 (0.017)	-
SE endemics [b]	40.0 (3.2)	1.8 (0.08)	-	0.096 (0.01)	-
Perennials (outcrossing) [d]	43.7	-	-	0.18	0.218
Monocots (outcrossing) [d]	52.5	-	-	0.165	0.157
All plants [c]	52.2 (1.0)	1.99 (0.03)	-	0.153 (0.004)	0.225 (0.009)

a Godt and Hamrick 2001, b Hamrick and Godt 1989

c Hamrick and Godt 1998, d Hamrick and Godt 1996

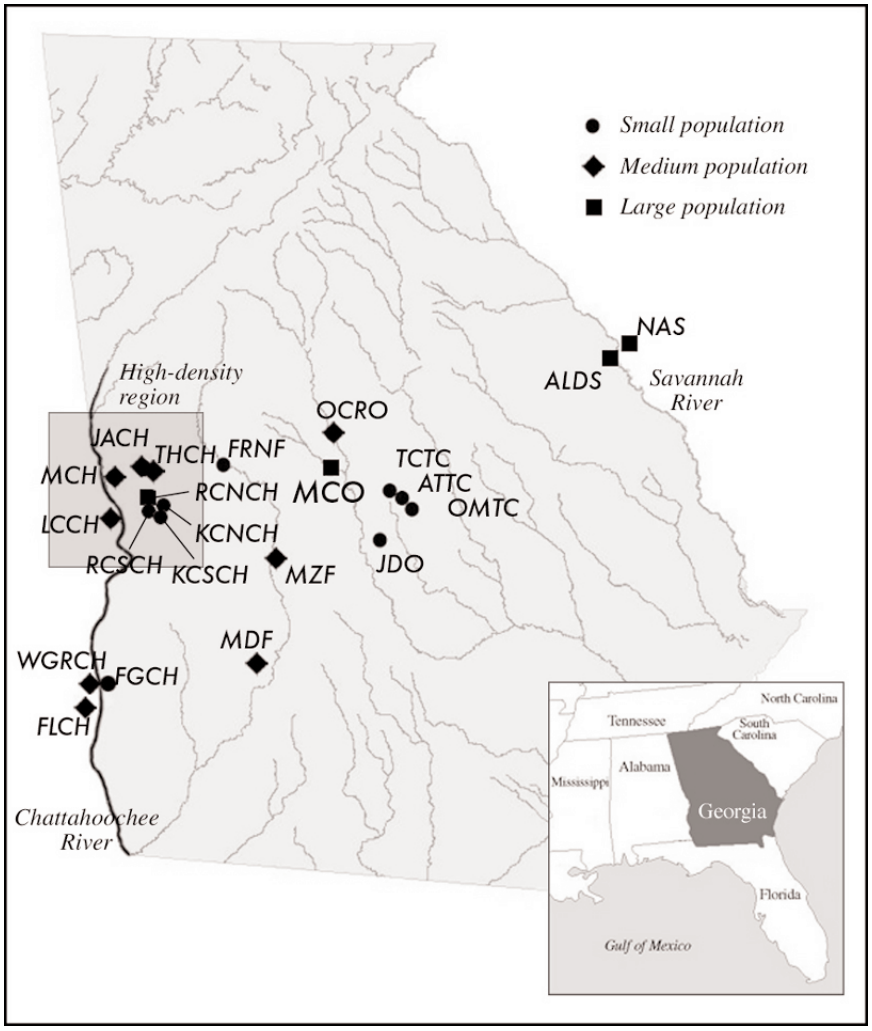

Figure 2.1: Distribution of sampled *Trillium reliquum* populations. The Chattahoochee River is indicated in bold and the high-density region by the shaded area in the NW portion of *T. reliquum* range.

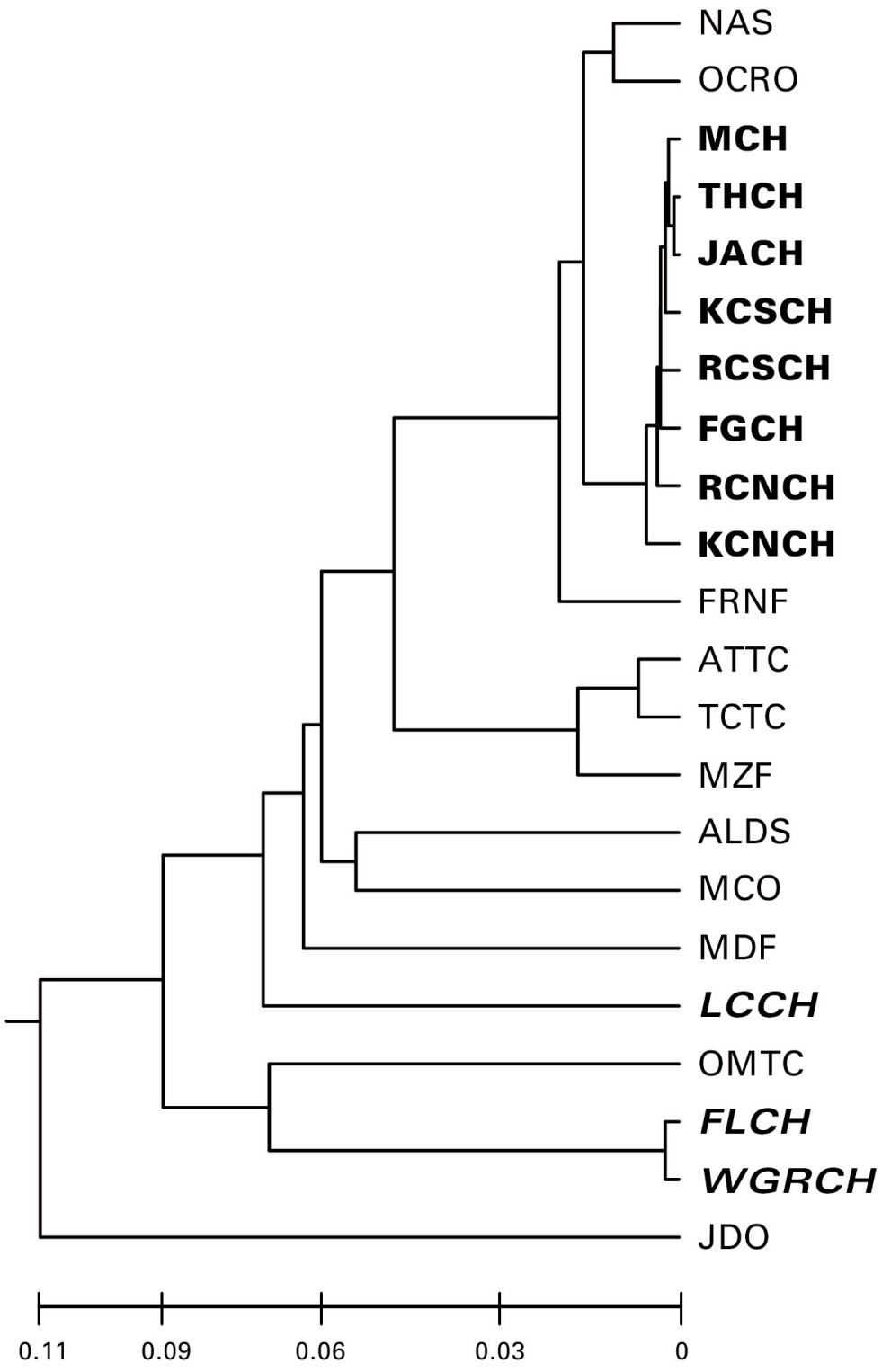

Figure 2.2: UPGMA phenogram based on Nei's (1972) genetic distance values for 22 sampled populations

CHAPTER 3

IDENTIFICATION OF GLACIAL REGUGIA BY PHYLOGEOGRAPHICAL ANALYSES OF A FOREST UNDERSTORY PLANT SPECIES, *TRILLIUM CUNEATUM* [1]

[1] Gonzales E, JL Hamrick and SM Chang. To be submitted to *Molecular Ecology*.

INTRODUCTION

The Earth's climate has experienced severe climatic oscillations which led to a series of Quaternary ice ages that have produced great changes in the species' distribution (Hewitt 2000). Such changes in the geographic distribution of species have had inevitable genetic consequences. The present distribution of genetic variation over a geographic scale was determined in part during the Quaternary ice ages and subsequent post-glacial expansions. The combination of fossil, physiographic, traditional biogeographic and genetic evidence facilitates our understanding of historical events as well as the present distribution of genetic diversity. In spite of the modern technology that allows the detection of single base changes in DNA sequences in maternally inherited plant genomes, the flora of major geographic regions remains nearly unstudied phylogeographically. One such region is the southeastern United States.

The biogeography, as well as the climatic conditions of unglaciated eastern North America during the Last Glacial Maximum (LGM), have been the subject of considerable interest and debate. Several studies provide comprehensive syntheses of the available paleoecological data (i.e. macrofossils and fossil pollen sequences) in North America during the LGM (Delcourt & Delcourt 1981, 1991, 1993, Davis 1983, Jackson et al. 2000). These studies reveal the extent of glaciers on the North American continent and vegetation patterns south of the ice sheet. In eastern North America south of the glacial boundary, there was a relatively narrow zone of tundra expanding westward. Open boreal woodlands and cool-temperate conifers dominated most of the area immediately south of the tundra as far south as 34^{o}N (Jackson et al. 2000).

During this period, the Lower Mississippi Valley was dominated by cool-temperate and temperate deciduous forest species that co-existed with a now extinct species of spruce (*Picea critchfieldii)* as far north as 35^{o}N (Jackson & Weng, 1999). From the available paleoecological data, we cannot resolve the eastern extent of this mixed *Picea*-temperate deciduous forest. The absence of well-dated LGM sites between 30^{o}N and 33^{o}N and east of 91^{o}W prevents determination of the northernmost extent of warm-temperate tree species and associated forest understory shrubs and

herbaceous species except for the Lower Mississippi Valley (Jackson et al. 2000). However, 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). Owing to the abundant representation of northern pines, hardwood pollen (as well as the pollen of forest herbs) must have been below the detection threshold of the pollen sensing system. Furthermore, interpretations of fossil pollen records may be misleading since trace amounts of pollen that has been dispersed long distances may lead to overestimates of the range of species during the LGM. Similarly, the absence of fossil pollen of particular species does not automatically indicate the absence of these species in a sampled range (MacLachlan and Clark 2004). Consequently, fossil records have limited power in the historical reconstruction of vegetation. Population genetic data can offer complementary and independent information to reveal the historical distributions of species.

Furthermore, even in situations where fossil evidence reliably identifies the locations of glacial refugia, we cannot imply that a particular refugium was the principal source of migrants for present day recolonization. Patterns of range expansion from refugia are difficult to determine from fossil records alone and, therefore, other evidence is needed to reconstruct the biogegraphical history of contemporary species (Tremblay and Schoen 1999).

Plants have many unique features that are of great utility to researchers attempting to unravel the spatial and temporal dynamics of populations and their consequences for evolution. Despite their immobility, seed plants move their genes during two life cycle phases: prior to fertilization, genes move in pollen, and subsequently, during dispersal, genes move in seeds. Pollen mediated gene flow plays a major role in connecting extant populations, but only seeds establish new populations (Levin 1981, Petit et al 2005). Consequently, maternally inherited genes are only dispersed by seeds and are especially valuable for clarifying the geographic distribution and spatiotemporal dynamics of plant populations. Chloroplast sequence markers are becoming increasingly popular in phylogeographic reconstructions. Because they are transmitted via seeds, migration

patterns can be inferred and distinguished from the confounding effects of pollen mediated gene flow (measured by nuclear markers).

We investigated the glacial history of a herbaceous temperate forest species, *Trillium cuneatum*. This monocot is a member of the Trilliaceae (sensu Dahlgren et al., 1985) or the Melanthiaceae (sensu APG, 1998). We analyzed contemporary biogeographical patterns of genetic variation, and from its spatial distribution we inferred historical patterns of seed dispersal and pollen mediated gene flow. *Trillium cuneatum* is a long-lived spring ephemeral, growing in mesic temperate deciduous forests. It presently ranges from Kentucky, through Tennessee to central Mississippi and Alabama, eastward into Georgia and the Carolinas (Case & Case, 1997). *Trillium cuneatum* is a relatively sessile species, its seeds (distributed primarily by ants and gravity), and pollen (distributed by weak flying insects) move very short distances. Seeds are produced infrequently, and more than ten years are required to reach the reproductive stage (EG pers obs). Such characteristics make this species a convenient subject for phylogeographic analyses. The northern margin of its present geographic distribution approaches the ice-sheet margin during the height of the last ice age. At the LGM this treeless, tundra covered area was covered with arctic vegetation. Open boreal woodlands and northern pines spread south. With winter temperatures falling as low as –25 ^{O}C (Jackson et al. 2000), it would likely not have been possible for temperate woodland wildflowers such as *Trillium* to survive the harsh arid and cold climatic conditions. Based on fossil evidence, the primary candidate for a LGM refugium is the Lower Mississippi Valley (LMV). However, under Jackson's (2000) scenario of temperate deciduous trees surviving in scattered sites south of 34^{O}N, other refugial sites could have occured 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 the southern refugia expanded northward. In addition, isolated refugial populations may have survived in protected cove forests of the southern Appalachian Mountains.

Several hypotheses emerged from paleoecological records regarding the evolutionary history of plant species. However, in the southeastern US, these predictions remain largely untested. The principal goal of this research was to reveal the biogeographical history of *T. cuneatum* popula-

tions using traditional genetic structure analyses of biparentally inherited genetic markers (allozymes) combined with phylogeographical analyses of maternally transmitted chloroplast DNA sequences (cpDNA). Specifically, we address the following questions: (*i*) Did *T. cuneatum* survive the last glacial maximum in multiple refugia, and, if so, where were these refugia located? Three possible scenarios were investigated: a) a single "western" refugium located in the LMV; b) a primary refugium located in the LMV, and additional scattered sites located in favorable mesic microsites in Alabama and Georgia; c) additional refugia in the cove forests of the Southern Appalachian Mountains. (*ii*) What are the post-glacial colonization patterns into the present geographic range? Did all refugia contribute equally to post-glacial colonization? Are the routes from the refugia parallel in a South-North direction? (*iii*) What are the population dynamics and patterns of genetic diversity in areas where refugial lineages meet? Do maternally inherited cpDNA and biparentaly transmitted nuclear genes leave the same genetic footprint in parapatric areas?

Material and methods

Sampling

We sampled 48 plants at the peak of flowering (at least 10 m apart to avoid collecting clonal individuals) from each of 42 *T. cuneatum* sites throughout its geographic range (Figure 1). We recorded geographic coordinates for each site (Table 1). Voucher specimens are deposited in the University of Georgia Herbarium (GA).

Laboratory analyses

Allozymes: We transported leaf samples on ice to the laboratory and crushed them within 24 hours using a mortar and pestle. To solubilize and stabilize enzymes, we added an extraction buffer (Wendel and Parks, 1982). The extract was absorbed onto chromatography paper wicks and stored at –70^{O}C until electrophoretic analyses. We used starch gel electrophoresis to determine allozyme diversity. We employed four gel-electrode buffer combinations to resolve 20 polymorphic loci on 11% starch gels: *Buffer 8-:* amino acid transferase (AAT-1 and AAT-2), glutamate dehydrogenase

(GDH), *Buffer 4:* isocitrate dehydrogenase (IDH) and shikimate dehydrogenase (SKDH-1 and SKDH-2); *Buffer 6:* peroxidase (PER), diaphorase (DIA), fluorescent esterase (FE-1, FE-2, FE-3 and FE-4), triose-phosphate isomerase (TPI-1 and TPI-2);, *Buffer MC:* malate dehydrogenase (MDH-1 and MDH-2); *Buffer 11*: 6-phosphogluconate dehydrogenase (6-PGDH), phosphoglucomutase (PGM) and UTP-glucose-1-phosphate (UGPP-1 and UGPP-2). The MC buffer was modified following the recipe of Wendel and Weeden (1991). Stain recipes for DIA are given in Cheliak and Pitel (1984); UGPP is given in Manchenko (1994). All other stain and buffer recipes were taken from Soltis *et al.* (1983). For enzymes with more than one locus, loci were numbered sequentially with the number one assigned to the most anodal locus.

Chloroplast DNA: Chloroplast DNA (cpDNA) in *Trillium* as in most angiosperms, is inherited maternally (Griffin and Barrett 2004). We preserved leaf tissue from ten individuals (a subset of plants used for allozyme analyses) in liquid nitrogen and stored the samples at –70^{O}C until DNA extraction. We carried out the extraction of total genomic DNA and 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 trnL intron and trnL-trnF intergenic spacer using universal primer pairs "c & d" and "e & f" (Taberlet et al 1991). We carried out the PCR using the following procedure: 1 cycle (3 min/94^{O}C), 38 cycles (1 min/94^{O}C, 1 min/48^{O}C, 1.5 min/72^{O}C), and 1 cycle (6 min/72^{O}C). Resulting amplified cpDNA fragments were sequenced on ABI370XL DNA sequencer using a Cycle sequencing BigDye Kit (ABI Applied Biosystems, Inc.) and the original amplification primers on both strands. We initially edited and aligned both sequenced strands using Sequencher 4.2.2 (Gene Codes Corporation, Ann Arbor, Michigan, 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).

Statistical analyses

Allozymes: Genetic diversity statistics were calculated for the species (as described in Hamrick and Godt, 1989) and for each population (as described in Hedrick, 1985) using a statistical program developed by M. D. Loveless and A. Schnabel. These measures included the percentage of

polymorphic loci (P), the mean number of alleles per locus (A) and the total number of alleles per population (A_T), and expected (H_e) heterozygosity. Subscript s indicates species values, whereas p indicates population values.

Population divergence was estimated using Nei's gene diversity statistics (Nei, 1973; 1977). This statistic (G_{ST}) estimates the proportion of the total genetic diversity (H_T) found among populations for each polymorphic locus; G_{ST} values were averaged across loci to obtain an overall estimate of population divergence. Each G_{ST} value was tested for significance by Chi-square test (Workman and Niswander, 1970). Nei (1977) demonstrated that G_{ST} is equivalent to a multi-allelic F_{ST} (Wright, 1951). Chakraborty and Danker-Hopfe (1991) have also shown that these two indices are empirically equivalent to Weir and Cockerham's (1984) θ when sample sizes are equal and a large number of populations are analyzed as is the case for this study. Genetic identity (I) and distance (D) measures were also calculated for each pairwise combination of populations (Nei, 1972). To graphically portray genetic relationships between *T. cuneatum* populations, we used genetic identities to construct a UPGMA phenogram as well as a Neighbour Joining tree using PHYLIP (Felsenstein 1989, 1993).

Chloroplast DNA 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 and Petit 1995). We measured genetic differentiation among localities by G_{ST} (Nei 1987) and N_{ST} (Pons and Petit, 1995, 1996) using software PERMUT (developed by RJ Petit, available at http://www.pierroton.inra.fr/genetics/labo/Software). The N_{ST} parameter takes similarities among haplotypes into account, contrary to G_{ST}. Petit et al (2005) demonstrated that measures of subdivision that take into account the degree of similarity among haplotypes make better use of the information inherent in haplotype data than standard measures based on frequencies only (i.e. G_{ST}). We employed multiple approaches (Neighbor Joining, Maximum Parsimony, Maximum Likelihood) to estimate intraspecific phylogenies from cpDNA sequence data using the software package PHYLIP (Felsenstein 1989, 1993).

RESULTS

Nuclear markers

Allozyme analyses revealed high genetic diversity at both the species and population level. Nineteen of 20 resolved loci were polymorphic (P_s = 95%, P_p = 62.5%, range 45 – 80%) (Table 1). A total 102 alleles were detected for all loci. Mean number of alleles per locus at the species level (A_s) was 5.10 and the mean for populations (A_p) was 2.11. Consistent with the high proportion of polymorphic loci and the number of alleles, genetic diversity (H_{es} and H_{ep}) was high (0.305 and 0.210, respectively) (Table 1). We did not observe reduced genetic diversity in the northern margins of the species' geographic range. On the contrary, the northernmost populations contain the highest genetic diversity. The proportion of nuclear genetic variation attributable to differentiation among populations was moderate (G_{ST} = 0.275).

We used Nei's genetic distances to construct UPGMA and Neighbor Joining trees; since both provided comparable results, only one dendrogram (UPGMA) is presented (Figure 3). The populations clustered into two main clades (I, II) (Figure 3) separated by a large dichotomy. Additionally, clade I (western clade) is geographically restricted to the southwestern portion of the geographic range (i.e. Mississippi) while the remaining populations clustered into clade II (eastern). The eastern clade is subdivided into smaller discrete clusters. Although a few populations were haphazardly scattered throughout the dendrogram within clade II (e.g. TBRSC, SCVSC, JNGSC), most genetically similar populations exhibit geographic affinities (e.g. IIa, IIb, IIc) (Figure 3).

Chloroplast DNA

We detected six distinct cpDNA haplotypes (A - F). DNA sequences of these haplotypes differ from one another in six variable characters (both single nucleotide substitutions and single nucleotide or short indels). With the exception of one population (YAZMS), all populations were monomorphic for a single haplotype (Table 1, Figure 1).

Geographical distribution and genetic structure of cpDNA haplotypes: All haplotypes are geographically structured. However, haplotype E, which is abundant in the northeastern portion of *T.*

cuneatum's range (southern Appalachian Mountains), was also detected in the southernmost population (NEWMS) in Mississippi. Three haplotypes (B, C, and E) were widespread, while the remaining haplotypes were detected in a single population (F) or restricted to a few locations (A and D). Among population genetic diversity was much higher for cpDNA (G_{ST} = 0.988, N_{ST} = 0.998) than for the nuclear markers (G_{ST} = 0.275).

Phylogenetic analysis: Phylogenetic analyses based on Maximum Parsimony, Neighbor Joining and Maximum Likelihood did not result in completely congruent and resolved phylogenetic trees. To resolve ambiguities, we compared our data to five congener sequences (*T. erectum, T. grandiflorum, T. maculatum, T. stamineum* and *T. reliquum*). We sequenced *T. stamineum* and *T. reliquum* in our lab, and obtained the remaining three sequences from Genbank (Zomlefer et al 2001). We used the geographic distribution of haplotypes combined with the most parsimonious criteria (invoked after visual examination of haplotypes) to propose the most likely phylogeographic relationships.

While we can resolve the phylogenetic relationships among haplotypes C, D, E and F without ambiguity, we can infer three potential (mutually exclusive) relationships among the more basal lineages A, B and C (Figure 2). Haplotypes B and C are separated by an indel six nucleotides long. While any of these three haplotypes (A, B, C) could be the most basal, ancestral type, in all probability, it is the centrally located haplotype C that gave rise to haplotype B as well as the northern lineages (D, E, and F). Haplotypes B and C are separated by one indel. This difference could either represent a deletion in the C haplotype, or an insertion of six nucleotides into the sequence of haplotype B. We believe that this mutation was most likely an insertion, and that haplotype B was derived from haplotype C. We base this conclusion not only on comparisons to the other *T. cuneatum* lineages (none of them posses this six base pair sequence), but more importantly, all of the congener species (representing both ancestral and sister taxa (Kato et al 1995, Zomlefer et al 2001)) also lack this six bp region. No ambiguity exists about haplotypes D and E which both independently evolved from C, and haplotype F which was derived from E. Each "ancestral-derived" pair of haplotypes in the north-eastern portion is separated by a single

nucleotide substitution. Three mutations separate haplotype A from its closest counterpart, C. Either A or C could represent the ancestral type; alternatively they might be sister lineages evolving independently from an extict common progenitor (or a haplotype not contained in our sample). Regardless of the basal history of the *T. cuneatum* lineages, we infer that clade I (i.e. populations carrying haplotype A) and clade II (i.e. populations carrying haplotypes B-F) have followed independent evolutionary trajectories for a long time. Although we cannot determine an absolute time of separation, considering an average synonymous substitution rate of 0.1 – 0.3 % per MYR in cpDNA genome (Wolfe et al 1987), the separation certainly predates not only the LGM, but also previous glacial episodes of the Wisconsin glaciation period (Zurawski et al 1984).

Discussion

In eastern North America, temperate species survived repeated Pleistocene glaciations by a series of range shifts consisting of southern contractions and northern expansions (Davis 1983, Jackson et al 2000). Fossil evidence suggests that during the last glacial epoch, temperate mesic deciduous forests, associated with herbaceous forest understory species such as *Trillium*, were continuously distributed along the Lower Mississippi Valley. These sites presumably functioned as refugia for temperate forest species during the glacial maxima. Additionally, other, perhaps smaller sites, east of Mississippi (Alabama, Georgia and Florida), with locally favorable ecological conditions may also have functioned as refugia, although only limited pollen evidence exists to support their existence (Jackson et al. 2000, MacLachlan & Clark 2004). Additionally, early biogeographers (e. g. Adams 1901, 1902, Wilder and Dunn 1920) proposed that isolated refugia existed in protected cove forests of the S. Appalachian Mountains and that these refugia gave rise to the rich present day biodiversity in the SE US. Not all refugia may be identified by fossil evidence, especially habitats in which the dominant pines produced copious amounts of pollen, as was the case in the SE US during the LGM (Delcourt and Delcourt 1983. Jackson et al 2000, MacLachlan and Clark 2004). Furthermore, even if the fossil record identifies potential refugial sites, it cannot provide information about their role in post-glacial range expansion.

Our investigation of the evolutionary history of *Trillium cuneatum* revealed a geographically highly structured intraspecific phylogeny, further corroborated by evidence obtained from biparently inherited nuclear markers (Figure 3). Molecular evidence grouped the *T. cuneatum* populations into two main lineages; one currently occupying Mississippi (clade I, haplotype A), and the other representing the remaining range (clade II, haplotypes B-F) (Figures 2 and 3). We cannot determine from our data which haplotype is ancestral to the other lineages; in fact, neither haplotype may have been the progenitor, and the extant lineages may have been evolving independently as "sister clades" (Figure 2).

In spite of the incompletely resolved basal intraspecific history, we can infer the number of glacial refugia, their approximate locations, and postglacial migration patterns. As expected, we identified a refugium in the Lower Mississippi Valley (clade I, lineage A). Congruent with a number of phylogeographic investigations of animal taxa (e.g. Bermingham and Avise 1986, Nedbal and Philipp 1994, Walker and Avise 1998), we identified an eastern clade of lineages whose phylogeny is resolved with less ambiguity (clade II, lineages B-F). Contrary to expectations, the western refugium did not play a role during the postglacial range expansion of *T. cuneatum* into its present range. Our data suggest that the LMV populations have followed their own evolutionary trajectory, separate from the rest of the species even before the Last Glacial Maximum. Most of the *T. cuneatum* populations in the SE US are derived from refugia belonging to clade II, poorly documented by pollen fossil evidence. It is also clear that these sites have played a major role in the post-glacial expansion of this species.

Although our findings are consistent with the "east-west" split between lineages, first identified for several animal species (Avise 2000), we were surprised by the latitudinal stratification of haplotypes in the northern portion of *T. cuneatum's* geographic range. 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 D-F within the eastern lineage (clade II). Additionally, another refugium (haplotype B, most likely derived from C) existed in Georgia, but this lineage is presently confined to the southeastern portion of the species' range, albeit larg-

er than its southwestern counterpart in Mississippi. We cannot be sure that the extant populations in Georgia represent *T. cuneatum*'s southernmost edge during the LGM; other populations (presently extinct) may have existed in Florida. Existence of a Florida refugium has been documented by Sewell et al (1994) in their study of *Liriodendron tulipifera*, and in a wetland herbaceous plant, *Sagittaria latifolia* (Dorken and Barrett 2004).

Intriguingly, propagules originating in Georgia did not migrate very far to the north. This haplotype may have been prevented from advancing by the Tennessee Continental Divide, (referred to as the Valley and Ridge physiographic region) in northwest Georgia, and by the Blue Ridge Mountains in North Georgia. The Valley and Ridge physiographic region deserves special attention. It derives its name from a series of parallel northeast-to-southwest trending valleys and ridges formed by folded and faulted sedimentary rocks extending southwest into Alabama. Biogeographical investigations gave rise to speculations that these parallel ridges and valleys have played a central role in species migrations during range contractions and expansions (Adams 1901, 1902, Wharten 1999). The distribution of *T. cuneatum* haplotypes provides additional support for his hypothesis. Propagules carrying haplotype C advanced from their southern Alabama refugia in a northeast direction following the Valley and the Ridge corridors, leaving their genetic footprint behind. The Southern Appalachians, dominated by boreal forest during the height of the LGM, were colonized by populations carrying haplotype E which was derived from the more centrally located lineage C. The current distribution of populations with haplotype E suggests that this lineage 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. Alternatively, haplotype E could have survived in a refugium southwest of its current distribution. This scenario is, however, unlikely because such a migration should have left a genetic footprint and we should have detected haplotype E in Alabama. We believe, that the distribution of haplotypes C (extending through the migration corridors of the Tennessee Continental Divide) and haplotype E (present only in the mountains and N. Carolina Piedmont) is more consistent with the conclusion that there were glacial refugia within the mountains.

Populations throughout the southern Appalachians 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 three km away carry haplotype E. Additionally, JNGSC has the lowest allozyme diversity (measured by P and A) and it is not genetically similar to any other population according to the UPGMA analysis (Figure 3). Such an outcome is consistent with a historical bottleneck and long-term isolation from other populations for a long time. Population JNGSC may be a relict population from an isolated refugium in the southern Appalachian mountains. Both, the unique haplotype derived from the common lineage in the mountains as well as the relatively low genetic diversity at the allozyme loci support this conclusion. Alternatively, the unique haplotype in JNGSC could have arisen by more recent, post-expansion mutation. Such an event in combination with a population bottleneck is also consistent with the patterns observed in both the cpDNA and nuclear markers. 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 should reveal similar patterns and provide more support for the "Out of Appalachia" hypothesis.

Additionally, we identified a haplotype D in two northern populations: central Tennessee (NSHTN) and Kentucky (ALLKY). The shared history of these two populations is also apparent from the nuclear markers (Figure 3) despite their geographic separation. This lineage is derived directly from haplotype C. Propagules carrying haplotype D probably migrated north during post-glacial range expansion, although our sampling does not allow us to determine how far south lineage D may have survived. Similarly, Dorken and Barrett (2004) identified rare haplotypes of *Sagittaria latifolia* in the same general area, and Griffin and Barrett (2004) in their study of *Trillium grandiflorum* identified two haplotypes whose current range is near the glacial boundary. However, this is not too surprising since *S. latifolia* currently occupies boreal forest in central Ontario, and *T. grandiflorum's* current range reaches southern Canada. Conversely, *T. cuneatum* has not expanded beyond 37^{O} N latitude.

The majority of the populations were fixed for the same haplotype with a single exception (YAZMS, a member of the western clade I) where nine of the ten sampled individuals had haplo-

type 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). The 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).

Trillium cuneatum haplotypes are typically clustered into a single geographic subregion. Besides the polymorphic YAZMS population, we detected another exception to this general observation. In population NEWMS at the southernmost edge of *T. cuneatum's* distribution, we detected haplotype E, the haplotype that is common in the S. Appalachian Mountains. The distance between NEWMS and the range of haplotype E exceeds 500 km. In theory, the detection of identical sequences in these disjunct locations could be explained by homoplasy. Alternatively, a rare long distance dispersal event could be responsible for the introduction of haplotype E to the NEWMS site. Although we cannot categorically exclude two identical independent mutations, we believe that homoplasy is less likely than long distance dispersal. Furthermore, the allozyme data exclude the possibility of a recent human introduction since NEWMS is included in the same clade with the nearby Mississippi populations that possess haplotype A (Figure 3). Given the short pollen dispersal distances (Gonzales et al in prep), slow sexual reproduction and patchy distribution of populations in the fragmented Mississippi landscape, it is unlikely, that recently introduced propagules from the mountains would have had the time to acquire nuclear genes typical of local Mississippi populations. Such an introduction must have been of older origins, likely predating the arrival of European settlers, and perhaps even predating the LGM.

Conclusions

Our findings demonstrate the phylogeographical concordance of diverse co-distributed regional biota. Comparative molecular assessments of vertebrates in the southeastern United States have revealed repeated patterns of a deep, phylogenetic "breaks" which typically distin-

guish populations in the eastern portion of the species' range from those to the west; an additional substructure is evident within these two phylogeographic lineages, but these differences are typically more shallow relative to the matrilineal separation between regions (Walker and Avise 1998, Avise 2000). Furthermore, our results provide molecular evidence consistent with the role of the Valley and the Ridge region as a natural corridor for the passage of species into the interior of Appalachia. Similar patterns were recorded in several turtle and freshwater fish species (Walker and Avise 1998, Avise 2000); however, they were attributed to watershed affinities of rivers running parallel to the Valley and the Ridge corridors. From our results, it becomes apparent that this region may have played an important role not only for aquatic organisms but also in the migration of plants, and potentially other terrestrial species as well.

While there is continuing interest in where and how the remarkable biota of the Southeast survived during the Quaternary, the number and location of glacial refugia within the southeastern US is still a matter of considerable debate among biogeographers and paleoecologists, primarily due to the paucity of fossil records. Since no distinctively southern species have been found in the boreal *Pinus-Picea* pollen fossil records, it poses a question of where the present southern flora survived during the full-glacial. The penetration of northern species deep into the South suggests that southern species must have been present along the Gulf Coast and the Florida peninsula, although some temperate tree species may have survived in rare more northern disjunct refugia in Alabama and Georgia, and within the Appalachian Mountains. Our results confirm the western refuge area in the LMV and further provide clear evidence of refugia in Alabama and Georgia. Surprisingly, the LMV refugium, considered by paleoecologists as the main refugium for broad leaf deciduous temperate forest, did not participate in the post-glacial expansion of *T. cuneatum*, although the same may not be true for other species surviving the full-glacial there. Furthermore, our data suggest that isolated refugia of southern species existed further north than indicated by the fossil record. Our study highlights the usefulness of molecular studies in understanding the complex vegetation history of the southeastern US. Combined with geological, paleoecological and traditional non-molecular biogeographical data, this investigation deepens our understanding of past biotas and provides novel insights not possible without such a multidisciplinary approach.

References

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

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

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

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

Bermingham E, Avise JC (1986) Molecular zoogeography of freshwater fishes in the southeastern United States, *Genetics*, 113, 939-965.

Case FW, Case RB (1997) *Trilliums*. Timber Press Inc. Portland, Oregon.

Davis MB (1983) Quaternary history of deciduous forests in eastern North America and Europe. *Annals of Missouri Botanical Garden*, 70, 550-63.

Chakraborty R, Danker-Hopfe H (1991). Analysis of population structure: a comparative study of different estimators of Wright's fixation indices. In: *Handbook of Statistics*. Vol. 8. Edited by CR Rao and R Chakraborty. Elsevier Science Publishers BV, New York, pp 203-254.

Cheliak WM, Pitel JA (1984). Techniques for starch gel electrophoresis of enzymes from forest tree species. Information report P1-X-42. Petawawa National Forestry Institute. Canadian Forestry Service, Agriculture, Chalk River, Ontario.

Dahlgren RMT, Clifford HT, Yeo PF (1985). The families of the Monocotyledons: Structure, evolution and taxonomy. Springer-Verlag, Berlin pp. 123-126, 550.

Delcourt PA. Delcourt RH (1981) Vegetation maps for eastern North America: 40,000 yr BP to the present, In: Romans, R. C. (Eds.) "Geobotany II." Plenum, New York, pp.123-126.

Delcourt PA, Delcourt RH (1991) Quaternary ecology. A paleoecological perspective. Chapman and Hall, New York. P. 242.

Delcourt PA, Delcourt RH (1993) Paleoclimates, paleovegetation, and paleofloras during the

Late Quaternary. In: Flora of North America. Editorial committee (Eds.), Flora of North America. Vol. I.Introduction. Oxford University Press, New York, pp.71-94.

Dorken ME, SCH Barrett (2004) Chloroplast haplotype variation among monoecious and dioecious populations of *Sagittaria latifolia* (Alismataceae) in eastern North America. *Molecular Ecology*, 13, 2699-2707.

Felsenstein J (1989) PHYLIP (Phylogeny Inference Package) Version 3.2. *Cladistics,* **5**, 164-166.

Felsenstein J (1993) PHYLIP (Phylogeny Inference Package) Version 3.5c. Distributed by the author. Department of Genetics, University of Washington, Seattle.

Griffin SR, Barrett SCH (2004) Post-glacial history of *Trillium grandiflorum* (Melanthiaceae) in Eastern North America, *American Journal of Botany,* 91, 465-473.

Hamrick JL, Godt MJW (1989) Allozyme diversity in plant species. In: *Plant population genetics, breeding and genetic resources* (eds. Brown AHD, Clegg MT, Kahler AL, Weir BS), pp. 43-63. Sinauer, Sunderland, Massachusetts.

Hedrick PW (1985) *Genetics of populations.* Jones and Bartlett, Boston.

Hewitt GM (2000) The genetic legacy of the Quaternary ice ages. *Nature*, 405, 907-913.

Jackson ST, Weng C (1999) Later Quaternary extinction of a tree species in eastern North America. *Proceedings of the National Academy of Sciences USA*, 96, 13847-13852.

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

Jackson ST, Webb RS, Anderson, KH, Overpeck JT, Webb T, Williams JW, Hansen BCS (2000) Vegetation and environment in Eastern North America during Last Glacial Maximum. *Quaternary Science Reviews,* 19, 489-508.

Jeanmougin F, Thompson JD, Gouy M, Higgins DG, Gibson TJ (1998) Multiple sequence alignment with ClustalX. *Trends in Biochemical Sciences*, 23, 403–405.

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

Levin D (1981) Dispersal versus gene flow in plants. *Annls MO Bot Grdn*, 88, 233-53.

MacLachlan J S, Clark JS (2004) Reconstructing historical ranges with fossil data at continental scales. *Forest Ecology Management,* 197, 139-147.

Manchenko GP (1994). *Handbook of detection of enzymes on electrophoretic gels. CRC Press,* Ann Harbour, MI. CRC Press, Ann Harbour, MI.

Nedbal MA, Philipp DP (1994) Differentiation of mitochondrial DNA in largemouth bass. *Transactions of American Fisheries Society, 123*, 460-468.

Nei M (1972). Genetic distance between populations. *American Nauralist,* 106, 283-292.

Nei M (1973). Analysis of gene diversity in subdivided populations. *Proceedings of National Academy of Science USA*, 70, 3321-3323.

Nei M (1977). F-statistics and analysis of gene diversity in subdivided populations. *Annals of Human Genetics,* 41, 255-233.

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

Petit RJ, Duminil J, Fineschi S, Hampe A, Salvini D, Vendramin GG (2005) Comparative organization of chloroplast, mitochondrial and nuclear diversity in plant populations. *Molecular Ecology*, in press.

Pons O, Petit RJ (1995) Estimation, variance and optimal sampling of gene diversity. I. Halploid locus. *Theoretical and Applied Genetics*, 90, 462-470.

Pons O, Petit RJ (1996) Measuring and testing genetic differentiation with ordered vs. unordered alleles. *Genetics,* 144, 1237-1245.

Sewell MM, Clifford RP, Chase MW (1996) Intraspecific chloroplast DNA variation and biogeography of North American *Liriodendron* L. (Magnoliaceae). *Evolution,* 50, 1147-1154.

Tremblay NO, JD Schoen (1999) Molecular Phylogeography of *Dryas intergrifolia*: glacial refugia and postglacial recolonization. *Molecular Ecology*, 8, 1187-1198.

Soltis DE, Haufler CH, Darrow DC, Gastony GJ (1983) Starch gel electrophoresis of ferns: a compilation of grinding buffers, gel and electrode buffers, and staining schedules. *American Fern Journal,* **73**, 9-27.

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.

Vellend M, Myers JA, Gardescu S, Marks PL (2003). Dispersal of *Trillium* seeds by deer: implications for long-distance migration of forest herbs. *Ecology* 84, 1067-1072.

Walker D, Avise JC (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.

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution,* 38, 1358-1370.

Wendel JF, Parks CR (1982) Genetic control of isozyme variation in *Camelia japonica* L. *Journal of Heredity,* 73, 197-204.

Wendel JF, Weeden NF (1991) Visualization and interpretation of plant isozymes. In Soltis D. E. and P. S. Soltis, eds. Isozymes in plant biology. Portland, USA. Dioscorides Press, 5-45.

Wharten CH (1999) The Cumberland Plateu and the Valley and Ridge. In Highroad guide to the Georgia Mountains Longstreet Press Athens, GA. pp.1-2.

Wolfe KH, Li WH, Sharp PM (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.

Workman PL, Niswander JD (1970) Population studies on southwestern Indian tribes. II. Local genetic differentiation in the Papago. *American Journal of Human Genetics,* 22, 24-49.

Wright S (1951) The genetical structure of populations. *Annals of Eugenics* 15, 323-354.

Zomlefer WB, Norris H, Williams N, Whitten WM, Judd WS (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 MT, Brown AHD (1984) The nature of nucleotide sequence divergence between barley and maize chloroplast DNA. *Genetics*, 106, 735-749.

Table 3.1. List of *Trillium cuneatum* populations, their geographic coordinates, percent polymorphic allozyme loci (*P*), total number of allozyme alleles (A_T) and genetic diversity (H_e). Affiliation to each haplotype group (A-F) is also indicated.

Population	Latitude	Longitude	*P (%)*	A_T	H_e	Haplotype
MACGA	32°53.948’	83°47.225’	60.0	42	0.204	B
TLBGA	32°43.155’	84°20.462’	60.0	39	0.184	B
HRSGA	32°42.571’	84°52.234’	75.0	44	0.219	B
MEAGA	32°39.500’	84°57.250’	50.0	38	0.201	B
FBCGA	32°33.100’	84°42.805’	55.0	40	0.185	B
FBRGA	32°33.250’	84°47.004’	50.0	35	0.151	B
CHATGA	34°52.66’	84°26.522’	55.0	35	0.155	B
WLKGA	34°38.835’	85°03.680’	65.0	47	0.192	C
FLDGA	34°12.921’	85°12.766’	85.0	54	0.272	B
POCGA	34°43.392’	85°23.477’	70.0	53	0.225	C
GRSGA	34°51.293’	84°38.900’	65.0	42	0.222	E
BRKGA	34°54.400’	83°24.000’	65.0	44	0.211	E
DALAL	32°19.428’	86°54.435’	60.0	39	0.228	C
LEEAL	32°31,634”	85°13.813”	60.0	42	0.200	B
I20AL	33°42.385”	86°03.569	50.0	39	0.184	B
BLNAL	34°02.571”	86°34.957”	65.0	44	0.298	C
LAWAL	34°22.803”	87°13.715”	65.0	45	0.167	C
WINAL	34°17.130”	87°23.903	50.0	39	0.128	C
FONNC	35°26.626’	83°49.209’	80.0	56	0.269	E
CHEONC	35°28.067’	83°54.083’	57.9	50	0.219	E
JKNC	35°21.260’	83°56.165’	70.0	54	0.220	E
RCNC	35°20.900’	83°55.000’	52.6	45	0.203	E
SALNC	35°13.150’	82°20.876’	60.0	45	0.266	E
UNGNC	36°04.370’	79°48.325’	65.0	41	0.255	E
UHWNC	35°25.750’	80°01.320’	55.0	38	0.160	E
SCVSC	34°50.500’	83°05.000’	57.9	40	0.206	E
TBRSC	34°51.850’	83°03.200’	65.0	39	0.187	E
JNGSC	35°07.534’	82°34.220’	45.0	34	0.196	F
EKRTN	35°17.338’	85°54.905’	55.0	53	0.266	C
DKRTN	35°28.271’	86°07.388’	70.0	50	0.240	C
PRBTN	35°29.959’	83°56.066’	75.0	55	0.271	E
NSHTN	36°04.251’	86°53.261’	65.0	56	0.288	D
NTZTN	35°24.909’	87°30.733’	70.0	46	0.219	C
WHTKY	36°41.978’	84°13.681’	75.0	54	0.317	E
ALLKY	36 °50.550’	86°04.443’	55.0	43	0.242	D
LEEMS	34°14.430’	88°48.558’	60.0	36	0.177	C
NEWMS	32°28.500’	89°03.612’	70.0	46	0.176	E
WARMS	32°21.237’	90°47.008’	70.0	50	0.272	A
YAZMS	32°44.100’	90°26.200’	65.0	44	0.184	A,C
GREYMS	32°56.205’	89°28.105’	70.0	43	0.172	-
JACKMS	32 °17.960’	90°09.980’	60.0	42	0.150	A
LEFMS	32 °19.458’	90 °09.373’	60.0	40	0.177	A

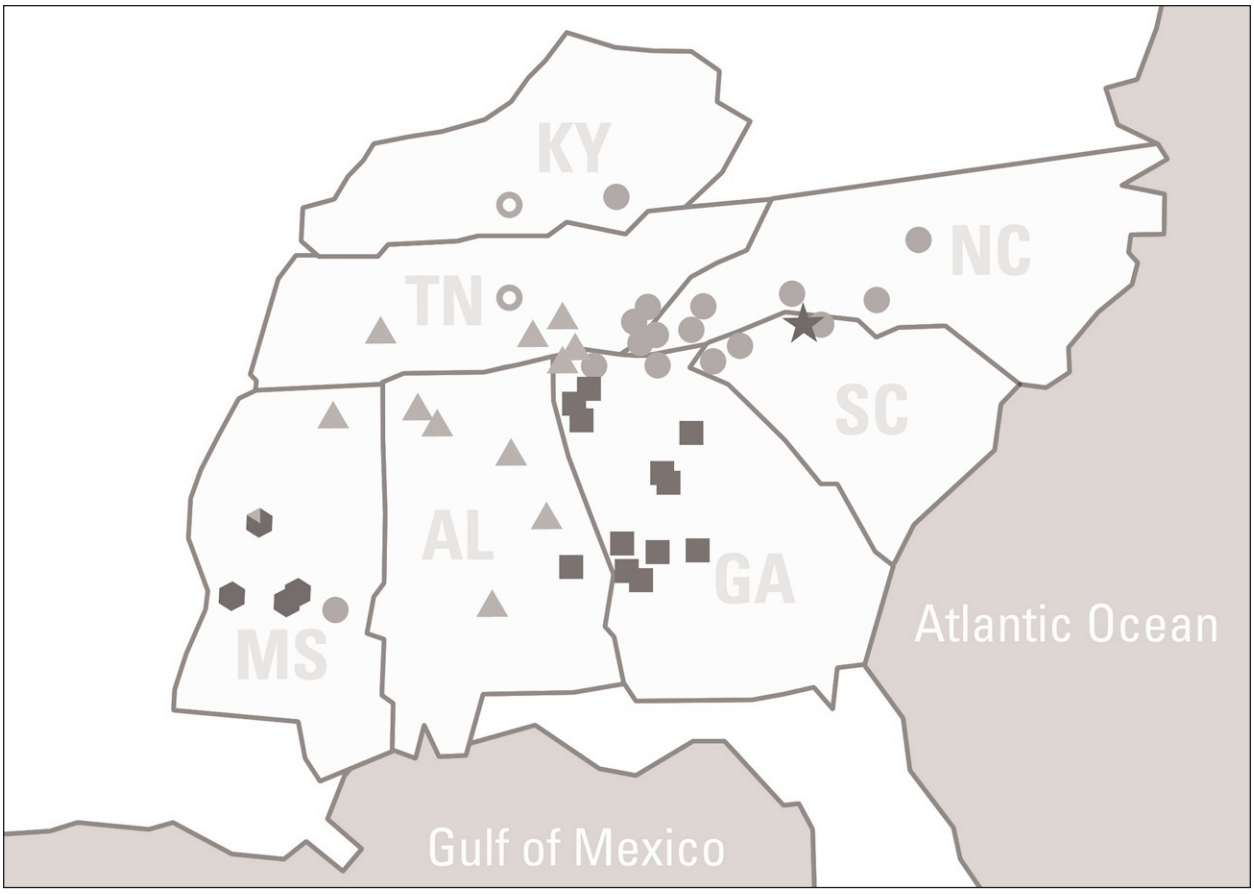

Figure 3.1: Geographic distribution of cpDNA haplotypes among *Trillium cuneatum* populations. The haplotypes are indicated as A – F. (See Figure 3.2 for phylogenetic relationships among haplotypes)

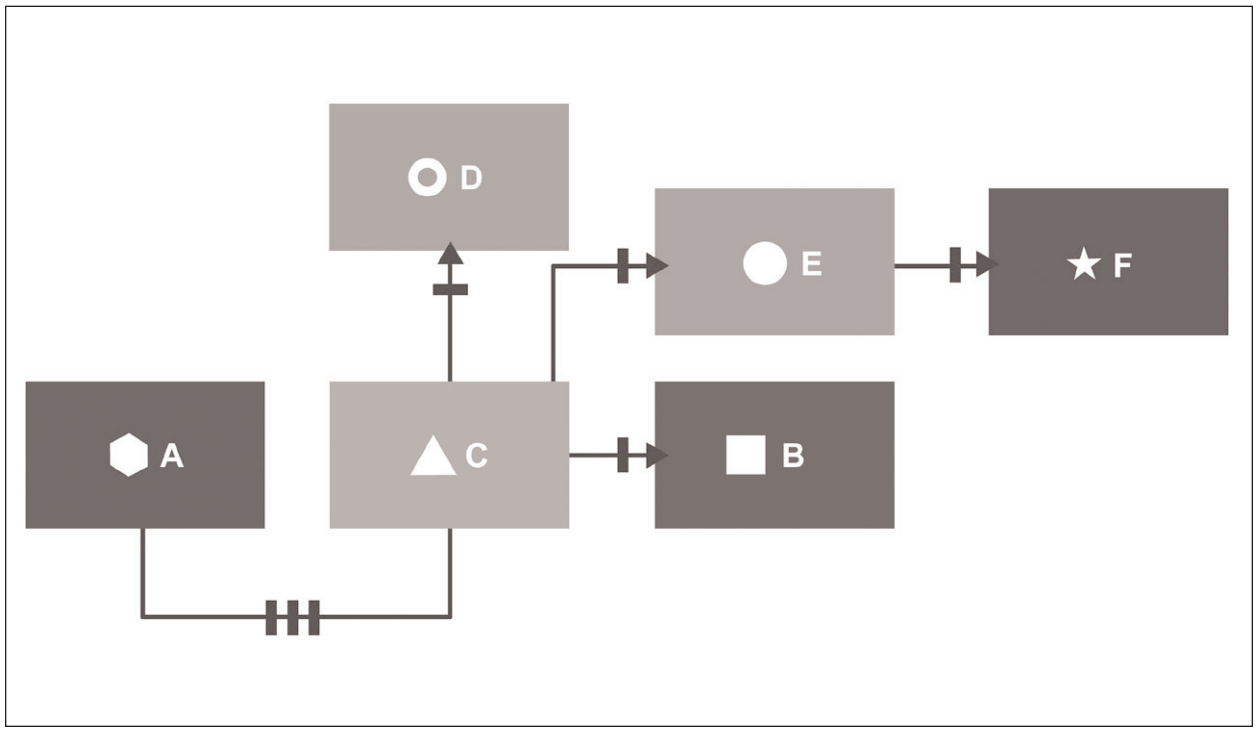

Figure 3.2: Maximum parsimony analysis of phylogenetic relationship among cpDNA haplotypes. Number of mutational steps is indicated by slashes across branches of the parsimony network. Arrows indicate “basal→derived” relationship among haplotypes.

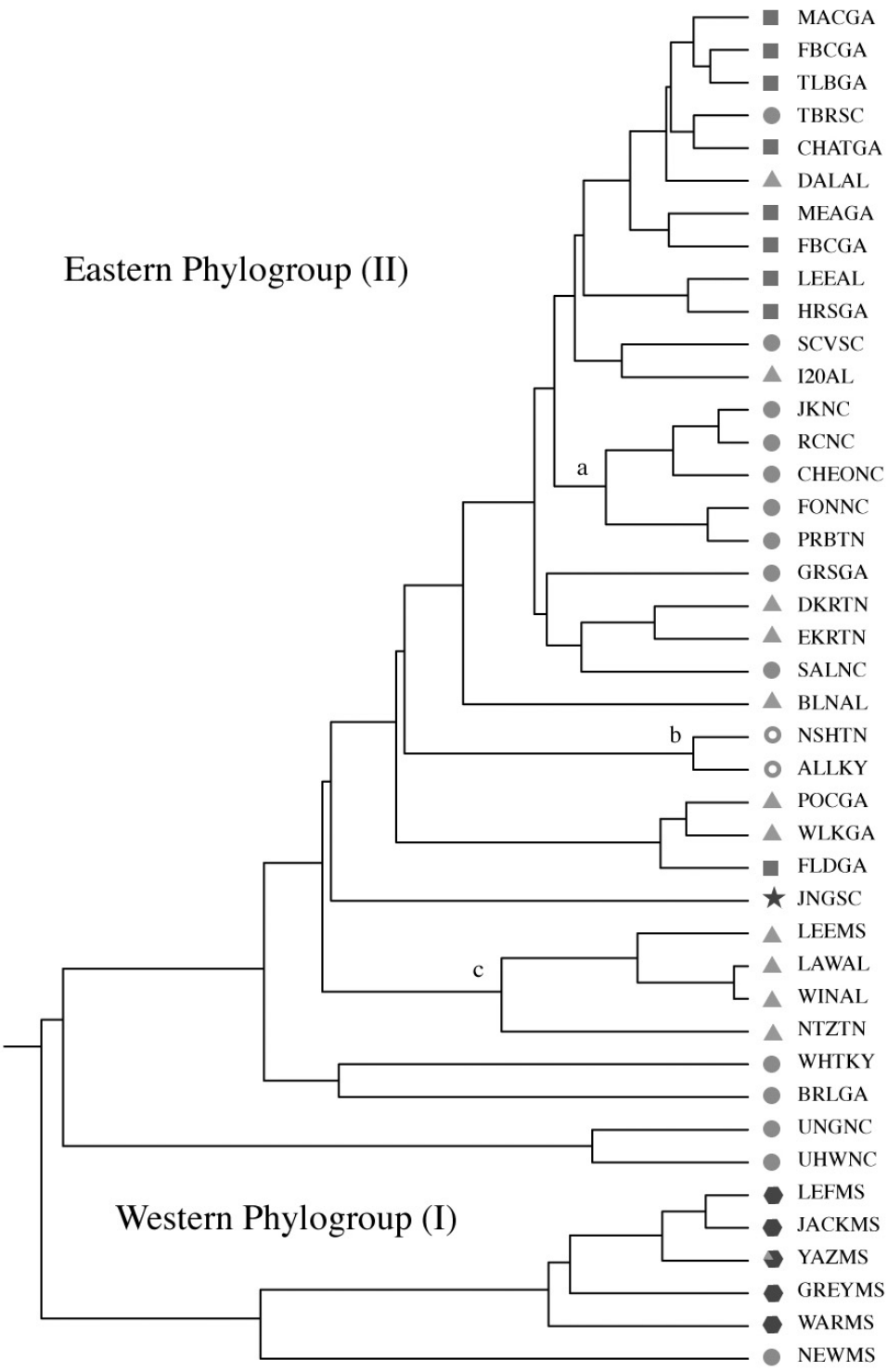

Figure 3.3: Genetic relationships among *Trillium cuneatum* populations based on 20 allozyme loci. UPGMA dendrogram was built according to the Nei's (1972) genetic distances. Haplotypes are designated consistently with Figures 3.1 and 3.2.

CHAPTER 4

VARIATION IN CLONAL REPRODUCTION AND FINE-SCALE GENETIC STRUCTURE IN A FOREST UNDERSTORY HERBACEOUS SPECIES, *TRILLIUM CUNEATUM* [1]

[1] Gonzales E and JL Hamrick, to be submitted to *Heredity*

Introduction

Genetic diversity is required for populations to adapt to continuous environmental changes (Frankel 1970); consequently, the maintenance and distribution of genetic diversity is a central theme in evolutionary biology. Large populations of naturally outbreeding species usually have extensive genetic diversity; however, genetic variation in continuous populations may be subdivided into smaller neighborhoods (Wright 1951). Furthermore, fine scale genetic structure (FSGS), defined as the non-random spatial distribution of genetic variation within a population, is closely associated with effective population size (N_e), a key parameter in population genetics (Chambers 1995). Effective population size has consequences for genetic drift, selection, and the loss of heterozygosity due to inbreeding. Effective population size and the distribution of genetic diversity also affect the establishment and spread of new alleles and the maintenance of polymorphism within populations. In a spatially structured population with limited seed dispersal, a cluster of related individuals carrying a mutant gene may develop, eventually increasing the frequency of a novel allele in a local neighborhood and ultimately in the population. Thus, FSGS can play a central role in evolutionary processes, and investigations of within population spatial genetic structure are of vital interest to population geneticists.

Two major factors have been implicated for the development and maintenance of FSGS within plant populations: *restricted gene dispersal* and *natural selection*. In addition, in clonally reproducing species, vegetative spread can contribute significantly to genetic sub-structuring within populations. In situations with heterogeneous microhabitats, localized genetic differentiation often results for loci affected by selection. Although exceptions exist (e.g. Hamrick and Allard, 1972), allozymes, microsatellites or dominant PCR based markers are assumed to be selectively neutral, and therefore, in outcrossing species selection should not influence their spatial structuring. Consequently, we consider gene dispersal rather than selection as the major factor in the development of genetic neighborhoods within large continuous populations (Wright 1951, Jain and Bradshaw 1966).

Gene flow influences the scale of local adaptation (Fenster et al 2003) and the role of population structure in evolutionary processes. Since established plants are immobile, they provide an ideal opportunity to study the effects of gene flow on the distribution of genetic variation over various spatial scales. Although established plants do not move, their genes are mobile; moreover, gene dispersal occurs in two stages: as haploid pollen, and, subsequent to successful pollination and zygote formation, as diploid seeds. Additionally, in some species, apomictic seeds and clonal propagules may also contribute to gene dispersal. Following gene dispersal, other processes (besides natural selection) contribute to the physical distribution and spatial location of genotypes within populations (e.g. demographic factors such as seedling establishment, thinning, colonization patterns and disturbance frequency).

In continuous populations with restricted gene dispersal, genetic structure can develop in the absence of selection, and the relatedness of individuals increases with spatial proximity (Wright 1951). It is important to make a distinction between the relative roles of seed versus pollen meditated gene dispersal in shaping FSGS. The spatial scale of patches of related genotypes depends primarily on seed dispersal patterns (not pollen), while the magnitude of relatedness within the genetic neighborhood is a function of both seed and pollen movement. The relative role of seed and pollen dispersal in shaping genetic structure is not always clearly distinguished in the plant population genetics literature. Kalisz et al (2001) provided clear and concise general predictions for four possible scenarios resulting from combinations of relative seed and pollen dispersal distances within plant populations: (1) When pollen and seed dispersal within large populations are panmictic, neither genetic structure or inbreeding will be detected. (2) Similarly, the lack of genetic structure will also be typical for populations with restricted pollen movement, but widely dispersed seeds. Additionally, a significant overlap of seed shadows reduces fine scale genetic structure. (3) Contrary to the previous two scenarios, localized seed dispersal relative to pollen movement will lead to the development of strong genetic structure. (4) Furthermore, if pollen and seed dispersal are both restricted, inbreeding may reinforce genetic substructuring and the relatedness of individuals within patches, while the spatial extent of the neighborhood will not be significant-

ly affected. In addition to these four scenarios, the extent and frequency of clonal reproduction can further complicate interpretations of FSGS (Setsuko et al 2004).

In this study, we investigate patterns of FSGS within large continuous populations of a myrmecochorous forest understory species, *Trillium cuneatum*, at various spatial scales. We compare FSGS and clonal reproduction in populations of this perennial herb located in higher elevations of the Southern Appalachian Mountains and at the lower elevations of the Piedmont Province of Georgia. Estimates of contemporary pollen mediated gene dispersal (Gonzales & Hamrick, in prep), as well as field observations of pollinators and seed dispersing vectors (ants) revealed that both pollen and seed mediated gene dispersal is limited. These observations, further corroborated by other investigations of myrmecochorous species (e.g. Kalisz et al 1999, 2001), predict highly genetically sub-structured populations. Our pollen dispersal study detected small, but statistically significant regional differences; pollen dispersal distances in the Piedmont are greater than in the mountains (EG et al in prep); however, our study provided only a temporal snapshot of pollen movement within a single season. Such contemporary differences in gene dispersal may not be indicative of gene movement patterns over longer periods of time. Consequently, they reveal little about long-term evolutionary and ecological processes that develop and maintain FSGS in populations of this long-lived perennial species. In long-lived polycarpic species with overlapping generations, further complicated by clonal reproduction and a seed bank, genetic neighborhoods are developed and maintained by processes encompassing many reproductive events. Moreover, examination of contemporary pollen movement does not provide insights into differences among microhabitats within populations. In this study, we investigated FSGS within five populations of *T. cuneatum*, in which we specifically asked whether there are differences in clonal reproduction and spatial genetic structure: a) among geographic regions (i.e. populations from the warmer, drier Piedmont Province versus populations located within the more mesic and cooler Southern Appalachian Mountains); b) among populations within the two geographic regions; c) among plots in each population (i.e. within population variation).

A central theme of this paper is the appropriateness of spatial scale. At what scale do demographic and ecological variables affect evolutionary processes, and what are the consequences for

our interpretations when appropriate scales are ignored or not considered? We illustrate how ecological and evolutionary processes interact at various spatial scales. We demonstrate how inappropriate sampling can mislead interpretations of fine scale genetic structure and inferences of microevolutionary processes. Finally, we propose an appropriate (sensitive) sampling scheme to detect and quantify within population genetic structure that would also render comparisons among studies of population genetic substructuring possible.

Material and methods

Study species

Trillium cuneatum is a long-lived spring ephemeral, carpeting floors of mature mesic deciduous forests, both in large continuous, as well as smaller, disturbed remnant habitats. Its geographic range extends from central Kentucky, through Tennessee to central Mississippi and Alabama, and east into Georgia and the Carolinas. This monocot species is a member of the Trilliaceae (sensu Dahlgren et al., 1985) or the Melanthiaceae (sensu APG, 1998).

The reproductive biology of *Trillium* is poorly understood. It takes ten or more years to reach reproductive maturity (Ohara 1989, Jules 1999, EG pers obs). Sexual reproduction (i.e. successful fruit production) in *T. cuneatum* is infrequent. In March and April, a mature plant may produce a single flower (rarely two) pollinated by weak flying insects such as fruit flies and less frequently Calliphoridae. Preliminary investigation suggests that this species is predominantly outcrossing and that it possesses a weak self-incompatiblity system (EG pers observ). Several studies described comparable mating systems in two congeners, *T. grandiflorum* and *T. erectum* (Broyles et al 1997, Kalisz et al 1999, Sage et al 2001). *T. cuneatum* reproduces sporadically by seeds; once established, however, individuals may persist for many decades (EG pers observ). Fruits mature at the end of June/early July. *T. cuneatum* is polycarpic; each maternal plant produces a single fruit containing, on average, 30-40 (rarely more) seeds, each with a fleshy elaiosome (EG pers observ). Ants, attracted to these elaiosomes, disperse the seeds and, in doing so, reduce the number of seeds germinating close to the maternal plant. Consequently, ant-mediated seed dispersal affects pat-

terns of seedling emergence and their relatedness. Additionally, yellow jackets and deer play a role as rare seed dispersers of *Trillium* (EG pers observ, Jules 1996, Vellend *et al.* 2003).

Ohara (1989) reported that *T. cuneatum* "reproduce exclusively by seeds" although in another section of the same report, he mentions a rare possibility for clonal reproduction. Jacobs and Jacobs (1997) observed vegetative spread in some *T. cuneatum* plants under cultivation. We have made similar observations of rhizomes transplanted into the greenhouse; furthermore, excavated rhizomes from natural populations provide evidence of clonal reproduction. In this paper, we also present quantitative evidence of cloning based on multilocus genetic data.

Study sites and sampling

Study sites. We conducted a comparative study of five populations located in two geographic regions: the Piedmont Province (three sites) and the Southern Appalachian Mountains (two sites) in northern Georgia and western North Carolina, USA (Figure 1, Table l). Two Piedmont populations were located in Atlanta, GA; both sites occur in small, fragmented forests, completely surrounded by urban development and subjected to ongoing human disturbance. One site (ATL1) is located in the Storza Woods, Midtown; the other (ATL2) is adjacent to Emory University. Although these populations occur in small remnants of mature, secondary forests, the *Trillium* populations are not small; each site contains several thousand flowering individuals. The third Piedmont population, the University of Georgia's Thompson Mill Experimental Forest (TMF), is located in a larger area, less affected by anthropogenic disturbances. *Trillium* plants occur mostly along the margins of 130 hectares of this mature, secondary-growth deciduous forest surrounded by pastures. Although this site is not subjected to direct anthropogenic disturbances, we have observed heavy deer browsing of *Trillium*. The remaining two sites are located in the Appalachian Mountains of north Georgia and western North Carolina. The Grassy Mountain (GRM) population is in relatively undisturbed Chattahoochee National Forest, GA in the southernmost Appalachian Mountains. The wilderness area of the Joyce Kilmer Memorial Forest (JKMF), part of the Nantahala National Forest in western North Carolina, provides habitat for the fifth popula-

tion. The two mountain populations consist of tens of thousands of patchily distributed individuals spreading over hundreds of hectares of continuous hardwood forest. Although Grassy Mountain has been deforested, the land was never farmed and it is likely that many forest understory herbaceous species, including *T. cuneatum,* survived as a seed bank (or rhizomes), rather than re-colonized the mature forest after initial timber harvest. JKMF is one of the very few remaining undisturbed old-growth forests in the SE US, and it is the only primary forest with *T. cuneatum* populations; thus it represents the most ancient population of *T. cuneatum* included in our study. The forest herbaceous layer in GRM and JKMF is species rich, while the Piedmont populations maintain lower species diversity.

Sampling design. Our sampling design is not completely balanced, because some plots were originally established for a contemporary pollen dispersal study; in these, only flowering individuals were sampled. In all populations, we established at least two rectangular sampling plots (min size 10 x 16 m). Plots in which we sampled only flowering individuals were larger than plots established for complete sampling. Plot size varied somewhat due to different plant densities. Plots were separated by at least 100 m within the smaller Piedmont populations, and by larger distances (> 300 m) in the Mountain sites. Spatial (X,Y) coordinates were recorded for all plants within each plot, and we classified each plant as either juveniles (plants with one leaf), non-reproductive (plants with three leaves, but no flower) and reproductive (flowering plants) (Ohara 1989). This classification scheme allows approximate age estimates, although some overlap between the two older categories exists since occasionally reproductive individuals do not flower (EG pers observ, Hanzawa and Kalisz 1993). All mapped plants were sampled for allozyme analyses. In addition, we established a third sampling site of the same size; in this plot only 48 haphazardly selected plants were sampled.

Data collected from plots with complete sampling and spatial coordinates were used for analyses of clonal diversity and spatial autocorrelation. Data obtained from the third plot were used to examine genetic structure among plots within populations using standard genetic structure statistics (e.g. analysis of private alleles, G_{ST}).

Laboratory Analyses

Leaf samples were transported on ice to the laboratory and crushed within 24 hours using a mortar and pestle. An extraction buffer (Wendel and Parks, 1982) was added to solubilize and stabilize the enzymes. The extract was absorbed onto chromatography paper wicks and stored at –70^{o}C until electrophoretic analyses. We used starch gel electrophoresis to determine multilocus genotypes for each individual sampled. We employed four gel-electrode buffer combinations to resolve the following 14 polymorphic loci on 11% starch gels: *Buffer 8-:* diaphorase (DIA), fluorescent esterase (FE-1, FE-2 and FE-3); *Buffer 4:* isocitrate dehydrogenase (IDH) and shikimate dehydrogenase (SKDH-2); *Buffer MC:* malate dehydrogenase (MDH-1 and MDH-2); *Buffer 11*: 6-phosphogluconate dehydrogenase (6-PGDH), phosphoglucomutase (PGM-1 and PGM-2) and UTP-glucose-1-phosphate (UGPP-1, UGPP-2 and UGPP-3). The MC buffer was modified following the recipe of Wendel and Weeden (1991). Stain recipes for DIA are given in Cheliak and Pitel (1984); UGPP is given in Manchenko (1994). All other stain and buffer recipes were taken from Soltis *et al.* (1983). For enzymes with more than one locus, loci were numbered sequentially with the number one assigned to the most anodal locus. The number of polymorphic loci varied from population to population.

Statistical Analyses

We used co-dominant allozyme markers to describe clonal, as well as fine-scale genetic structure. Although allozyme loci have fewer alleles and lower genetic diversity than microsatellites, they are ideal markers for measuring genetic structure and making inferences about gene dispersal. In addition to low mutation rate, allozyme analyses typically resolve more loci giving more reliable estimates of relatedness among individuals (Ennos 2001).

Clonal variation. The ability to group individuals into the same genet using genetic markers and to resolve clonal patterns in a population is a function of the number of resolved polymorphic loci and genetic diversity at each locus. The level of resolved genetic variation determines the exclusion probability (i.e. the likelihood of a second individual having a different diploid genotype from the focal individual by chance in a sample of N individuals). In our study, all individuals

were sorted according to multi-locus genotypes and all plants with identical multi-locus genotypes were assumed to be ramets of the same genet. The clonal exclusion probability (Paetkau and Strobeck 1994) calculated from independently determined allele frequencies in each population was greater than 0.99, thus the probability of getting the same multi-locus genotype by chance is less than one percent.

We measured clonal diversity as the number of multilocus genotypes (*G*) divided by the number of samples in each plot (*N*). *G/N*, clonal diversity, is interpreted as the probability that a randomly chosen plant has a unique genotype. The effective number of genotypes, a measure of clonal diversity adjusted for the frequencies of detected multilocus genotypes, was determined according to the formula

$$N_{eg} = 1/\sum g_i^2$$

where g_i represents the frequency of i-th multilocus genotype. We estimated genet size as the number of ramets with identical multilocus genotypes. We measured genotypic diversity at each site by the Complement to Simpson's index corrected for finite sample size (Simpson 1949, Pielou 1969):

$$D = 1 - \sum\{[n_i(n_i-1)] / [(N-(N-1)]\},$$

where n_i equals the number of individuals of genotype *i* and *N* is plot sample size. Values of *D* range from zero to one, with higher values corresponding to greater genotypic diversity. To investigate differences among sites in genet diversity, we compared the values of all Complements to Simpson's index using pairwise t-tests. We tested the null hypothesis that the two samples came from locations having the same genotypic diversity with the following statistic:

$$t = [D_{s1} - D_{s2}]/[s_1^2 + s_2^2]^{1/2},$$

where D_{s1} and D_{s2} are the complements to Simpson indices for plots 1 and 2 respectively; s_1^2 and s_2^2 are the variances of D_{s1} and D_{s2} (Keefe and Bergeson, 1977). Finally, the distribution of genets in each plot was assessed by Fager's (1972) Evenness index

$$E = (D_{obs} - D_{min})/(D_{max} - D_{min})$$

where D_{min} = [(G-1)(2N-G)]/[N(N-1)], D_{max} = [(G-1)N]/[G(N-1)] and G is the number of unique multilocus genotypes (i.e. number of genets) (Fager 1972). Genotypic evenness (*E*) ranges from zero (each sample has the same genotype or one genotype dominates) to one (all genotypes are represented by the same number of ramets) and scales *D* to the level of polymorphism in each plot. This index provides no information about the spatial distribution of genotypes.

Spatial Genetic Structure. We used two approaches to assess within population genetic structure. First, we calculated genetic diversity statistics for each population (as described in Hedrick, 1985) using a statistical program LYNSPROG developed by M. D. Loveless and A. Schnabel. These diversity statistics measure observed (H_o) and expected (H_e) heterozygosity and the number of private alleles occurring in each plot (A_p). Deviations from Hardy-Weinberg expectations were examined for each polymorphic locus within each plot by calculating Wright's fixation index (Wright, 1922). Fixation indices were tested for significance using Chi-square (Li and Horvitz, 1953). We used Nei's gene diversity statistics to estimate genetic structure among plots within each population (Nei, 1973; 1977). This statistic (G_{ST}) estimates the proportion of the total genetic diversity (H_T) found among plots for each polymorphic locus; G_{ST} values were averaged across loci to obtain an overall estimate of plot differentiation. Each G_{ST} value was tested for significance by Chi-square (Workman and Niswander, 1970). However, to detect FSGS using F-statistics (or G_{ST}), one must delineate neighborhoods within a continuous population; without a priori knowledge of population subdivision, this method may not be sensitive enough to detect and quantify genetic structure. Furthermore, it may lead to misinterpretations about inbreeding due to a spatial Wahlund effect.

As an alternative approach, we used spatial autocorrelation to evaluate the spatial distribution of genetic diversity within populations. The utility of spatial autocorrelation analysis for detecting genetic structure has been thoroughly demonstrated since its introduction by Sokal and Oden (1978 a, b, Heywood 1991, Ennos 2001). Unlike classical spatial autocorrelation analysis that is usually executed on one allele at a time, we employed a multivariate approach to autocorrelation analysis applicable to co-dominant, multilocus arrays, based on genetic distance methods follow-

ing the methods of Smouse and Peakall (1999). This procedure reduces the stochastic noise by avoiding the need for allele-by-allele, locus-by-locus analysis.

We generated autocorrelation coefficient *r* between pairwise geographic and pairwise squared genetic distance matrices for all possible pairs of individuals within each plot from their multilocus genotypes. The *r*-values approximate mean pairwise relatedness between individuals. For instance, a *r*-value of 0.5 represents parent-offspring or full-sib relationships; a value higher than 0.5 would indicate a history of inbreeding, while $r = 0.25$ is interpreted as a half-sib relationship. Negative values indicate greater genetic dissimilarity among individuals than expected by chance. We calculated mean *r*-values for each discrete distance interval by averaging over all pairs of individuals located within a distance class, and tested for statistical significance by randomly shuffling the genotypes among geographic locations and recomputing *r*. We used 1000 permutations to define the upper and lower bounds of the 95% confidence interval generated under the null hypothesis of no genetic structure. We superimposed the confidence envelope about the null hypothesis of no spatial genetic structure (random) as determined by permutations on the graph and plotted the autocorrelation coefficient *r* as a function of distance. For a given distance class, *r* is significantly different from zero (at $p < 0.05$) if the observed value falls outside the confidence envelopes. Significant correlation of $r > 0$ indicates that in a given distance class, individuals are more closely related than expected by chance, while for a value of $r < 0$ individuals are less related than expected by chance. This method is further explained in Smouse and Peakall (1999). All spatial autocorrelation statistics and simulations were performed using program GenAlex V5_1.1 (Smouse & Peakall, 1999) freely available at http://www.anu.edu.au/BoZo/GenAlex.

For each plot, we determined a mean *r*-value for nearest neighbors (distance d = 0 – 10 cm) and then for d = 10 – 25 cm, (see Figure 3 for remaining distance classes). The number of distance classes varied according to plot size. A set of 50 individuals scattered across the entire population was included in the analysis of each plot to ensure that data in each plot were compared to the same "*genetic background*" of the population. Individuals comprising this genetic background

were scattered across the landscape at greater distances from each other than the largest distance class in the above analysis to avoid their inclusion in the within plot analyses.

To separate the confounding effect of clonal reproduction from gene dispersal, we assessed FSGS in two ways. The first analysis included all the ramets within each plot. Subsequently, we excluded all replicates of identical multilocus genotypes (i.e. presumed clones) and conducted the spatial autocorrelation analysis for all genets represented only once within each plot. The genetic background was held constant with both analyses (i.e. with and without clonal individuals). Differences between the two correlograms for each plot were attributed to clonal reproduction. Correlograms excluding the clonal individuals were then used to make inferences regarding factors other than vegetative reproduction responsible for shaping FSGS.

Results

Clonal reproduction

Clonal reproduction of *T. cuneatum* differed significantly between Mountain and Piedmont habitats. Although a high proportion of the individuals had a unique multilocus genotype in all populations, and clonal clusters were spatially restricted (maximum diameter 20 cm) we detected vegetative reproduction more frequently in the Piedmont populations. Furthermore, we observed more variation among plots within the Piedmont populations, while the Mountain sites were very similar in clonal reproduction, both between populations within the region and within populations. The mean clonal diversity (*G/N*) in the mountains was 91.7% (range 90.0% – 94.2%, SD = 1.4%), while mean clonal diversity for the Piedmont populations was significantly lower (G/N = 72.3%), ($p < 0.005$) with higher variation among populations and plots (range 60.4% – 85.4%, SD = 10.5%) (Table 2). This difference between the Mountains and the Piedmont was further magnified when we excluded the flowering plots from our calculations (mean G/N_{MT} = 91.5% vs. mean G/N_{PD} = 68.2%, $p < 0.003$). The flowering plots tend to inflate genotype diversity because non-reproductive individuals are not represented in the analysis; from the completely surveyed plots, we observed that a large portion of the asexually produced ramets were non-reproductive.

The two geographic regions also varied greatly in clone size (i.e. number of ramets per genet) and the frequency that clonal spread was observed (Figure 2). Mountain populations rarely contained clones composed of more than two ramets; only in one case, did we detect a genotype with five ramets. Piedmont populations not only clone more frequently, but clonal clusters are significantly larger than in the mountains with a maximum observed size of 23 ramets (Table 2).

Clonal diversity measured by the Complement to Simpson's Index *D* was consistent with the relative values of *G/N*. The mean *D* value in the mountains was 0.999 (SD = 0), while Piedmont populations had a slightly lower, yet statistically significant, mean value of 0.991 (range 0.976 - 0.998, SD = 0.008) ($p < 0.05$). Fager's Evenness measure ranged from 0.77 -0.97; we did not detect any significant differences between regions.

Within population genetic structure

Genetic structure measured by G_{ST} among the three plots within each population was weak in all five populations (range 0.014-0.025) (Table 3). In contrast, the number of private alleles (i.e. alleles occurring exclusively in one plot) in each population was unexpectedly high (range 5 – 9) (Table 3); an additional 11 (range 1 – 4) alleles were detected in two out of the three sampled plots. In each population and within each plot, we detected an excess of homozygotes over Hardy-Weinberg expectations; the inbreeding coefficient (F_{IS}) ranged from 0.048-0.143. The discrepancy between observed and expected H_e values was greater in the mountain populations.

Spatial autocorrelation analyses (excluding clonal individuals). While the G_{ST} statistics did not reveal significant genetic structure differences among plots within populations in either geographic region, the spatial autocorrelation approach proved to be more sensitive in detecting FSGS. Correlograms resulting from the multilocus spatial autocorrelation analyses graphically demonstrate how mean genetic relatedness among pairs of individuals declines as a function of inter pair distances (Table 4, Figures 3a, 3b, 3c). The complete analyses resulted in 24 correlograms; we provide correlograms for three representative plots (Figure 3)

Nearest neighbors (plants occurring within 10 cm) in the mountain populations have relatively high mean genetic similarity ($r = 0.406$, SD = 0.077). The mean *r*-value detected for the near-

est neighbors in the Piedmont populations is slightly lower. However, variation among plots and populations is much greater (r = 0.352, SD = 0.163) in the Piedmont. In the next distance class (10-25 cm), differences in mean r-values between the regions decrease, while variability within the Piedmont region (even among plots within the same populations) remains high (SD is four times greater than in the mountains) (Table 4).

The spatial extent of significant FSGS also differed between the two regions. While the mean size of significantly positive genetic structure in the mountains is 294 cm (SD = 66 cm, CV = 22.5%), genetic sub-structuring in the Piedmont locations was detected on a smaller scale and with considerably greater variation among plots than in the mountains (mean d = 150 cm, SD = 106 cm, CV = 70.8%) (Table 4).

Very little consistency among plots and populations, however, resulted from the cohort comparisons. In three plots, we observed increasing spatial genetic structure and relatedness among near neighbors, while in five plots the opposite pattern emerged.

Spatial autocorrelation analyses (including clonal individuals). Comparisons of analyses based on all the ramets revealed an increase of mean pairwise genetic relatedness for near neighbors; however, it had only a minor effect on the spatial extent of FSGS within the plots (Figure 3). As might be expected from the clonal analyses, r-values for the Piedmont populations were higher than those of the mountains (r = 0.607 and 0.518 respectively). Nevertheless, the Piedmont populations continue to have greater variation among sampled sites (Table 4).

Discussion

Consistent with our expectations, we documented clonal reproduction and significant spatial genetic sub-structuring within all *T. cuneatum* populations. Decoupling clonal reproduction from sexually reproduced individuals allowed us to interpret the observed FSGS as a consequence of restricted seed and pollen dispersal around maternal plants. Relatively high relatedness values for this predominantly outcrossing species most likely result from several mutually re-enforcing factors. This species experiences correlated mating and/or biparental inbreeding; a conclusion that is

consistent with the results of a companion study of contemporary pollen movement conducted in four of these populations (Gonzales et al in prep). Furthermore, some seeds may not be dispersed from their maternal plants, leading to the establishment of seedlings near their parents (Kalisz et al 1999). Such a scenario combined with short pollen movement distances, correlated mating, inbreeding and longevity (i.e. overlapping generations) would account for the development of genetic structure with the high genetic relatedness values we observed.

Our results provide evidence of regional differences in the distribution of within-population genetic structure as a result of both variation in clonal reproduction as well as gene dispersal associated with sexual reproduction. Furthermore, while the Piedmont populations experience considerable heterogeneity in all the parameters measured, the mountain populations were remarkably similar. These results are consistent with expectations of greater stability and more equilibrium conditions in the older, less disturbed mountain populations. The more unstable conditions that characterize the environmentally stressed Piedmont region are probably responsible for the greater heterogeneity observed both within and among populations.

Clonal reproduction

Our data provide evidence that clonal reproduction is influenced by the population's geographic location. Clonal reproduction was relatively common in Piedmont populations, while vegetative spread was rarely detected in mountain populations. Moreover, in the Piedmont the degree of asexual reproduction varies both among populations and among plots separated by 100 m or more within populations. We did not observe significant heterogeneity in clonal growth at either spatial scale in the mountain region.

Trillium cuneatum occupies temperate forests of the southeastern US where temperature, light and moisture vary in both space and time. There is a strong elevation and climatic gradient from cooler and moister conditions in the mountains to drier and warmer conditions in the Piedmont. Periods of droughts are more often encountered in the Piedmont at the southern edge of the *T. cuneatum* distribution. In the Mountains, droughts occur less frequently and their duration is

shorter. Species distributed across this climatic gradient and exposed to temporal abiotic variability have evolved specialized adaptations to their local environments (Etterson 2004).

A number of studies have investigated clonal reproduction in natural populations of perennial herbaceous species (e.g. Cheplick and Gutierrez 1999, Brzosko et al 2002), but little is known about the processes that determine variation in asexual reproduction either within or among sites. In some species vegetative spread was associated with stressful environmental conditions and is interpreted as an important adaptation for survival during unfavourable conditions (e.g. Tybjerg and Vestergaard 1992, Kime et al 1997). Kudoh (1999) attributed a significant role in shaping clonal structure of a forest understory herb *Uvularia perfoliata* to light availability and forest canopy disturbance. Pulliam et al (in review) detected a significant negative relationship between soil moisture and vegetative reproduction for another forest herbaceous perennial species, *Polygonatum biflorum* across this mountain – Piedmont gradient.. *Polygonatum* clone significantly more often in the drier and more stressful Piedmont environments than in the mesic, cooler and stable higher elevation conditions. Additionally, several other studies (e.g. Marshall 1990, Evans and Cain 1995) have documented reduced clonal growth under inter-specific competition, presumably due to resource limitations.

Ohara (1989) proposed that pedicelate *Trillium* species distributed in the Appalachian Mountains (e.g. *T. grandiflorum, T. sulcatum, T. ovatum*) reproduce exclusively by seeds or that vegetative reproduction in such species is very rare. He further noted an exception and reported that *T. cuneatum* and *T. luteum* (although species belonging to subgenus *Phyllantherum* having sessile flowers), are more similar to species found in the mountains: they do not clone. This conclusion was based on limited sampling in one high elevation Georgia site. He further suggested that other sessile *Trillium* species in lower elevations (mostly the Coastal Plain) reproduce by cloning more commonly because of environmental stress and disturbance.

We believe that the same arguments are pertinent to the regional intra-specific differences in cloning that we observed between the Mountains and Piedmont *T. cuneatum* populations.

Piedmont populations experience higher environmental stress (e.g. heat, drought, human disturbance, heavy deer browsing), while the mountain plants are exposed to more stable and less disturbed environmental conditions. Additionally, the forest understory in both Mountain populations is species rich lending support to the argument that inter-specific competition reduces vegetative reproduction.

Such diverse biotic and abiotic conditions may have contributed to ecological differentiation in the reproductive strategies of *T. cuneatum*. Just as related species of *Trillium* are adapted to local, specific environments in deciduous forests of the Eastern US (Ohara 1989), *T. cuneatum* may have evolved differences in reproductive traits in response to different environmental conditions found in the Mountains and the Piedmont. These differences may represent genetic adaptations to warmer, drier, more disturbed and stressful conditions or they may be primarily environmentally induced. Our data provide no direct evidence for either hypotheses, but we have observed that rhizomes replanted from Piedmont populations to the greenhouse clone readily, while transplanted rhizomes from the mountains did not branch.

Genetic structure within populations

Several measures of genetic diversity can be used to quantify the distribution of variation within natural populations. We used Nei's (1973) coefficient of gene diversity (G_{ST}) to partition genetic variation among the three plots within each population. In all populations, the low G_{ST} values (0.014–0.025) do not suggest strong population sub-structuring on a spatial scale of 100 meters or more. The lack of detecting genetic structure at this spatial scale probably results from sampling across several neighborhoods within each plot. This explanation is further corroborated by the observation of a slight heterozygote deficiency within the plots, which can be explained by an internal spatial Wahlund effect. Our pollen movement study also indicates the possibility of biparental inbreeding due to near neighbor mating (Gonzales et al in review).

Since G_{ST} is based on allele frequencies, the common alleles primarily determine its value. In contrast, private alleles, whose frequencies are often low, do not significantly contribute to G_{ST} values. However, the common occurrence of private alleles in our plots indicates restricted gene

movement by both seed and pollen within these populations. We did not anticipate the number of private alleles (range 5 – 9 per population); furthermore, even more surprising was their relatively high mean frequency (0.019 – 0.072) (Table 3). Such a pattern is consistent with restricted gene dispersal within all populations, in spite of low G_{ST} values among plots. This observation is consistent with our companion pollen dispersal study (EG et al in prep) and with presumed predominant ant-mediated and gravity seed dispersal.

Spatial autocorrelation analyses, our second approach, demonstrated regional differences in population substructuring. Plots in the mountain populations revealed stronger FSGS in terms of both genetic similarity (mean r = 0.406) among nearby plants and the spatial extent of FSGS (mean d = 3 m). Piedmont populations are subdivided into spatially smaller units (mean d = 1.5 m) and genetic relatedness among near neighbors is lower (mean r = 0.352). There are two potential explanations for this observation. (1) Smaller genetic neighborhoods may be explained by more restricted seed dispersal in the Piedmont than in the Mountains. (2) Alternatively, seed dispersal distances may be comparable in both environments, but overlapping seed shadows in the Piedmont due to higher densities of fruiting individuals blur the footprint of seed dispersal and, thus, decrease the apparent size of the genetic neighborhoods. Additionally, overlapping seed shadows will lead not only to smaller patches; the magnitude of relatedness within these patches is expected to be lower than in neighborhoods where seeds from different families are less frequently mixed. Our observations are consistent with the later scenario. Genetic neighborhoods in the mountains are larger and near-neighbors have higher genetic relatedness than in the Piedmont, providing little evidence for overlap during seed dispersal. This conclusion is also consistent with considerably lower fruiting density, more restricted pollen movement and lower effective number of pollen donors per fruit observed in the Mountains (Gonzales et al in prep). Fruiting individuals are separated by distances which are several times greater than the observed spatial extent of neighborhoods (approximately 7 m^2). Our estimates of genetic neighborhoods are consistent with other studies directly measuring seed dispersal by ants (e.g. Kalisz et al 1999).

Furthermore, while there is little variation among plots within the Mountains, plots within the Piedmont populations vary considerably for both the r-values and for the spatial scale of FSGS.

This plot-to-plot variability in the Piedmont may be attributed to anthropogenic disturbances that continually disturb both biotic and abiotic conditions. Such instability particularly characterizes the two small Atlanta populations that are exposed to frequent human disturbance and enhanced edge effects. These populations occur in small fragmented forests surrounded by urban development; the entire populations are affected by edge effects (i.e. increased light, temperature, urban heat) since all individuals are within approx. 50 m or less of the forest edge. Interestingly, the TMF site, which is exposed to less human disturbance than the two Atlanta populations has some features of both environments (e.g. high relatedness among near neighbors and low fruiting density as in the Mountains, but high clonal reproduction as in the other Piedmont populations), possibly due to less severe edge effects and absence of urban surroundings resulting in intermediate spatial genetic structures.

How do human disturbance and edge effects influence fine scale genetic structure? Patterns of genetic diversity reflect both species biology and abiotic conditions. Furthermore, genetic substructuring likely reflects the history of these populations. In all populations, restricted gene dispersal via both pollen and seeds is apparent. Mountain populations experience less natural environmental fluctuations. Furthermore, JKMF is an ancient population, never disturbed by human activities; consequently, there has been ample time to establish stable FSGS. Similarly, the GRS population, free of repeated human disturbance and edge effects is also stable, although it occurs in a secondary forest. In contrast, both Atlanta populations are subjected to edge effects, increased pollen movement and produce higher fruiting densities than the mountain populations (EG et al in prep). Interestingly, in spite of the greater density of fruiting plants, we detected fewer seedlings and juvenile plants in the Piedmont plots indicating lower germination and/or higher seedling mortality in this environment, observations consistent with more stressful conditions. Higher pollen mediated gene flow, overlapping seed shadows and lower recruitment via sexual reproduction may account for the lower relatedness values and smaller scale genetic subdivisions of the Piedmont populations. In all probability these populations originally maintained stable fine scale genetic structure that was more similar to the mountain populations than it is today. European set-

tlers deforested the Piedmont and converted its previously contiguous ecosystems into an archipelago-like landscape, with small natural habitats embedded in agricultural and urban zones, exposing the forest habitats to edge effects and their consequences. The currently observed regional differences may reflect different land-use history as well as environmental conditions, whether natural or human induced.

Most of within-population genetic structure studies have focused on one or very few populations (usually represented by a single plot). A handful of these studies have taken a demographic genetics approach and investigated differences in spatial structure among age cohorts. These investigations of mostly tree species often report deterioration of spatial genetic structure from the seedling stage to the reproductive adult stage, a pattern consistent with stochastic thinning during recruitment (Parker et al 2001, Hamrick et al 1993, Epperson and Alvarez-Buylla 1997). Trees lend themselves to such studies because we can usually estimate their age; however, in most perennial herbaceous plants, age usually cannot be determined. We know of only a few studies of herbaceous perennials which have taken a cohort approach (e.g. Kalisz et al 1999, Tonsor et al 1993, Cruse-Sanders and Hamrick 2004). Both Kalisz (1999) and Tonsor (1993) reported an increase in genetic structure from seedling or seeds to the adult stage for *Trillium grandiflorum* and *Plantago lanceolata*, while Cruse-Sanders and Hamrick (2004) reported stronger genetic structure among juvenile ginseng plants (*Panax quinquefolius*) than reproductive adults in five out of seven plots. We have taken a similar approach in multiple plots and populations across the landscape, taking clonal reproduction into account. Very little consistency among plots emerges from the cohort comparisons. In three plots, we observed an increase in genetic structure with age; similar to the scenario in the *T. grandiflorum* and *P. lanceolata* studies (Kalisz et al 1999, Tonsor et al 1983). In the remaining five plots, where demographic data were available, we observed the opposite pattern: strongest genetic structure occurred in the juvenile stages, and FSGS decayed with age, although the degree of erosion varied considerably. The development of genetic structure varies from plot to plot among age cohorts. This lack of a consistent pattern may be a consequence of year-to-year variability in seed production, recruitment and/or mortality. Furthermore,

when clonal individuals were included in the cohort analyses, a different pattern of genetic structure development emerged. For instance, in the TMF-2 plot, the adult reproductive life stage showed the strongest FSGS, while the same plot reanalyzed without clonal individuals demonstrated the weakest FSGS for the same cohort. We observed similar confounding effects of cloning in all the other plots.

Two take home messages emerged from our cohort analyses: (*i*) clonal reproduction can lead to misinterpretations of ecological and evolutionary factors shaping patterns of genetic structure; (*ii*) in addition to restricted pollen and seed dispersal, stochastic processes appear to be important during the development of FSGS within all populations. In spite of the overall apparent stability of genetic structure within the large mountain populations, population subdivision remains a dynamic process.

Furthermore, our results strongly point out the need for analyses of multiple populations and plots within populations before we can make generalizations for the species. Spatial autocorrelation can be exploited in full, and result in interpretations within the appropriate ecological context only if comparative analyses are available. Studies based on analyses of one or very few sites render limited interpretations. Investigation of one population may not be enough to reveal microevolutionary forces acting in the population or specific ecological conditions.

Conclusions

This study concentrated on comparative analyses of FSGS in continuous populations of *T. cuneatum* located in two environmentally different geographic regions. Our findings document that restricted gene dispersal, clonal reproduction, and longevity with multiple overlapping generations all contribute to significant spatial genetic structure in this species. From the analyses of FSGS, we infer that vegetative spread is spatially restricted, mating occurs predominantly among near neighbors, and that neighborhood areas may be delineated by restricted seed dispersal on the order of 7 m^2 or less, resulting in fine scale subdivision of the continuous populations into a mosaic of small patches of related individuals.

Furthermore, our study revealed significant geographic differences both in clonal reproduction and fine-scale genetic structure; we attribute these differences to varying environmental conditions, both natural and human induced. Although our study could not test directly for the long-term stability of FSGS, the results indicate that older, relatively undisturbed populations have more homogenous patterns of FSGS, perhaps because their more stable environment allows the FSGS to approach equilibrium conditions. In contrast, populations in harsher, more disturbed environments are heterogeneous, and their FSGS is perturbed. Although we can speculate that naturally more stressful conditions are largely responsible for the differences detected in clonal reproduction among the two regions, and anthropogenic disturbance and fragmentation is the major cause of unstable FSGS, our data cannot clearly discriminate between the two potential causes.

This study not only provides important insights about variation in FSGS among geographic regions, among populations and across microsites, but it also links theses differences to potential causes. Today, the greatest challenge is to reveal cause and effect association between observed patterns and their underlying mechanisms. It is the understanding of the mechanisms that will facilitate generalization to unstudied sites and species with comparable life history traits (Nathan and Muller-Landau 2000). By placing replicate plots within populations located in different environmental regions, we could demonstrate the scale of heterogeneity in FSGS. Our design also allowed us to interpret our results in the context of the underlying causal mechanisms shaping genetic structure within populations. To determine the important factors, we need large-scale integrated studies which can generate crucial insights into microevolutionary processes. Increasing recognition of the importance of large-scale studies represents prospects for further progress. We might hope to develop predictive understanding of how gene dispersal and the development of FSGS is affected by abiotic agents, and ultimately begin applying these insights for a better understanding of the microevolutionary forces that govern processes within large, continuous populations.

References

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

Broyles, S. B., S. L. Sherman-Broyles, and P. Rogati. 1997. Evidence of outcrossing in *Trillium erectum* and *T. grandiflorum* (Liliaceae). *Heredity,* 88: 325-329.

Brzosko, E., A. Wroblewska, and M. Ratkiewicz. 2002. Spatial genetic structure and clonal diversity of island populations of lady's slipper (*Cypripedium calceolus*) from the Biebrza National Park (northeast Poland). *Molecular Ecology* 11: 2499-2509.

Chambers, S. M. 1995. Spatial structure, genetic variation and the neighborhood adjustment to effective population size. *Conservation Biology* 9: 1312-1315.

Cheliak., W. M., and J. A. Pitel. 1984. Techniques for starch gel electrophoresis of enzymes from forest tree species. Information report P1-X-42. Petawawa National Forestry Institute. Canadian Forestry Service, Agriculture, Chalk River, Ontario.

Cheplick, G. P., and C. M. Gutierrez. 2000. Clonal growth and storage in relation to competition in genets of the rhizomatous perennial *Amphibromus scabrivalis*. *Canadian Journal of Botany* 78: 537-546.

Cruse-Sanders, J. M. and J. L. Hamrick. 2004. Spatial and genetic structure within populations of wild American ginseng (*Panax quinquefolius*, L. Araliaceae). *Journal of Heredity* 95: 309-321.

Dahlgren, R. M. T., H. T. Clifford, and P. F. Yeo. 1985. The families of the Monocotyledons: Structure, evolution and taxonomy. Springer-Verlag, Berlin pp. 123-126, 550.

Ennos, R. A. 2001. Inferences about spatial processes in plant populations from the analysis of molecular markers. *In* J. Silvertown, and J. Antonovics [eds.] Integrating Ecology and Evolution in a Spatial Context, 45-72. Blackwell Science, London, UK.

Epperson, B. K., and E. R. Alvarez-Buylla. 1997. Spatial autocorrelation analysis of family structure in multiple life stages of *Cecropia obtusifolia. Evolution* 51: 275-282.

ETTERSON, J. R. 2004. Climate change I. Clinal patterns of selection along an environmental gradient in the Great Plains. *Evolution* 58: 1446-1458.

EVANS, J. P., AND M. L. CAIN. 1995. A spatially explicit test of foraging behavior in a clonal plant. *Ecology* 76: 1147-1155.

FAGER, E. W. 1972. Diversity: a sampling study. *American Naturalist* 106:293-310.

FENSTER, C. B., X. VEKEMANS, AND O. J. HARDY. 2003. Quantifying gene flow from spatial genetic structure data in a metapopulation of *Chamaecrista fasciculata* (Leguminosae) *Evolution* 57: 995-1007.

FRANKEL, O. H. 1970. Variation, the essence of life. Sir William Macleay Memorial Lecture *Proceedings of the Linnean Society* 95: 158-169.

HAMRICK, J. L., AND R. W. ALLARD. 1972. Microgeographical variation in allozymes frequencies in *Avena barbata. Proceedings of National Academy of Science* 68: 2100-2104.

HAMRICK, J. L., D. A. MURAWSKI, AND J. D. NASON. 1993. The influence of seed dispersal mechanisms on the genetic structure of tropical tree populations. *Vegetatio* 107/108: 281-297.

HANZAWA, F. M., AND S. KALISZS. 1993. The relationship between age, size, and reproduction in *Trillium grandiflorum* (Liliaceae). *American Journal of Botany* 80:405-410.

HEDRICK, P. W. 1985. *Genetics of populations.* Jones and Bartlett, Boston.

HEYWOOD, J. S. 1991. Spatial analysis of genetic variation in plant populations. *Annual Review of Ecology and Systematics* 22:235-255.

JACOBS, D. L., AND R. L. JACOBS. 1987. Trilliums in woodland garden: American Treasures. Ecogardens, USA.

JAIN S. K., AND A. D. BRADSHAW. 1966. Evolutionary divergence among adjacent populations I. The evidence and its theoretical analysis. *Heredity* 21: 407-441.

JULES, E. 1996. Yellow jackets (*Vespula vulgaris*) as a second seed disperser for the myrmecochorous plant, *Trillium ovatum*. *American Midland Naturalist* 135: 367-369.

KALISZ, S., J. D. NASON, F. M. HANZAWA, AND S. J. TONSOR. 2001. Spatial population genetic structure in *Trillium grandiflorum*: the roles of dispersal, mating history, and selection. *Evolution* 55: 1560-1568.

KALISZ, S., F. M. HANZAWA, S. J. TONSOR, D. A. THIEDE, AND S. VOIGT. 1999. Ant-mediated seed dispersal alters pattern of relatedness in populations of *Trillium grandiflorum*. *Ecology* 80: 2620-2624.

KEEFE, T. J., AND E. P. BERGESON. 1977. A simple diversity index based on theory runs. Water Res. 11: 689-691.

KIME, L., KLIMESOVÁ J., HENDRIKS R., AND J. VAN GROENENDAEL. 1997. Clonal plant architecture: a comparative analysis of form and function. *In* H. de Kroon and J. van Groenendael [eds.] The ecology and evolution of clonal plants 1-29. Backbuys Published, Leiden, The Netherlands.

KUDOH, H., H. SHIBAIKE, H. TASASU, D.F. WHIGHAM, AND S. KAWANO. 1999. Genet structure and determinants of clonal structure in a temperate deciduous woodland herb, *Uvularia perfoliata*. *Journal of Ecology* 87: 244-257.

LI, C. C., AND D. G. HORVITZ. 1953. Some methods of estimating the inbreeding coefficient. *American Journal of Human Genetics* 5: 107-117.

MARSHALL, C. 1990. Source-sink relations of interconnected ramets. *In* J. van Groenendael and H. de Kroon [eds] Clonal growth in plants: regulation and function, 23-42 The Hague: SPB Academic Publishers, Netherlands.

MANCHENKO, G. P. 1994. *Handbook of detection of enzymes on electrophoretic gels. CRC Press,* Ann Harbour, MI. CRC Press, Ann Harbour, MI.

NATHAN, R., AND H. C. MULLER-LANDAU. 2000. Spatial patterns of seed dispersal, their determinants and consequences for recruitment. *Trends in Evolution and Ecology* 15: 278-285.

NEI, M. 1973. Analysis of gene diversity in subdivided populations. *Proceedings of National Academy of Science* 70: 3321-3323.

NEI, M. 1977. F-statistics and analysis of gene diversity in subdivided populations. *Annals of Human Genetics* 41: 225-233.

OHARA, M. 1989. Life history evolution in the genus *Trillium*. *Plant Species Biology* 4: 1-28.

PARKER, K. C., J. L. HAMRICK, J. D. NASON, AND A. PARKER. 2001. Fine-scale genetic structure in *Pinus clausa* (Pinaceae) populations: effects of disturbance history. *Heredity* 87: 99-113.

PAETKAU, D. AND C. STROBECK. 1994. Microsatellite analysis of genetic variation in black bear populations. *Molecular Ecology* 3:489-495.

PIELOU, E. C. 1969. An Introduction to Mathematical Ecology. Wiley-Interscience, New York.

SAGE, T. L., S. R. GRIFFIN, V. PONTIERI, P. DROBAC, W. W. COLE, AND S. C. H. BARRETT. 2001. Stigmantic self-incompatibility and mating patterns in *Trillium grandiflorum* and *Trillium erectum* (Melanthiaceae). *Annals of Botany* 88: 829-841.

SETSUKO, S., K. ISHIDA, AND N. TOMARU. 2004. Size distribution and genetic structure in relation to clonal growth within a population of *Magnolia tomentosa* Thunb (Magnoliaceae). *Molecular Ecology* 13: 2645-2653.

SIMPSON, E. H. 1949. Measurement of diversity. *Nature* 163:688.

SMOUSE, P. E., AND R. PEAKALL. 1999. Spatial autocorrelation analysis of individual multiallele and multilocus genetic structure. *Heredity* 82: 561-573.

SOKAL, R. R., AND N. L. ODEN. 1978a. Spatial autocorrelation in biology. 1 Methodology. *Biological Journal of the Linnean Society* 10:199-228.

SOKAL, R. R., AND N. L. ODEN. 1978b. Spatial autocorrelation in biology. 2 Methodology. *Biological Journal of the Linnean Society* 10:229-249.

SOLTIS, D. E., C. H. HAUFLER, D. C. DARROW, AND G. J. GASTONY. 1983. Starch gel electrophoresis of ferns: a compilation of grinding buffers, gel and electrode buffers, and staining schedules. *American Fern Journal* 73: 9-27.

TONSOR, S. J., S. KALISZ, J. FISHER, AND T. P. HOLTSFORD. 1993. A life-history based study of population genetic structure: seed bank to adults in *Plantago lanceolata*. *Evolution* 47: 833-843.

TYBJERG, H. AND P. VESTERGAARD. 1992. Growth dynamics in the rhizomatous herb *Polygonatum verticillatum*. *Oikos* 65: 395-408.

VELLEND, M., J. A. MYERS, S. GARDESCU, AND P. L. MARKS. 2003. Dispersal of *Trillium* seeds by deer: implications for long-distance migration of forest herbs. Ecology 84: 1067-1072.

WENDEL, J. F, AND N. F. WEEDEN. 1991. Visualization and interpretation of plant isozymes. In Soltis D. E. and P. S. Soltis, eds. Isozymes in plant biology. Portland, USA. Dioscorides Press, 5-45.

WENDEL, J. F. AND C. R. PARKS. 1982. Genetic control of isozyme variation in *Camillia japonica,* L. *Journal of Heredity* 73: 197-204.

WRIGHT, S. 1922. Coefficient of inbreeding and relationship. *American Naturalist* 56: 330-338.

WRIGHT, S. 1951. The genetical structure of populations. *Annals of Eugenics* 15: 323-354.

Table 4.1: Locations, elevations and spatial co-ordinates of investigated *Trillium cuneatum* populations.

Location (Population code)	Elevation (m)	Spatial coordinates	
		Latitude	Longitude
A. MOUNTAINS			
Joyce Kilmer Memorial Forest, Nantahala National Forest, North Carolina (JKMF)	1,000	N 35^{o}21.26’	W 83^{o}56.16’
Grassy Mountain, Chattahoochee National Forest, Georgia (GRM)	1,200	N 34^{o}50.29’	W 84^{o}39.95’
B. PIEDMONT PROVINCE			
Thompson Mill Experimental Forest, Georgia (TMF)	280	N 34^{o}05.07’	W 83^{o}49.00’
Atlanta, Storza Woods, Georgia (ATL 1)	300	N 34^{o}47.56’	W 84^{o}22.36’
Atlanta, Emory University, Georgia (ATL 2)	300	N 34^{o}49.00’	W 84^{o}19.80’

Table 4.2: Differences in clonal diversity between geographic regions and among populations within geographic regions (Southern Appalachian Mountains and Piedmont Province).

Geographic Region Population-plot #	*N*	*G*	*G/N*	*Max* n_i	N_{eg}/G	*D*	*E*
A. MOUNTAINS							
JKMF-1	231	212	0.918	5	0.846	0.999	0.910
JKMF-2	210	189	0.900	3	0.766	0.999	0.953
GRM-1*	259	244	0.942	4	0.872	0.999	0.773
GRM-2	238	221	0.929	4	0.838	0.999	0.868
Mean			0.917	4	0.831	0.999	0.876
(SD)			(0.014)	(0.82)	(0.045)	(0)	(0.077)
Mean **			0.915	4	0.817	0.999	0.910
(SD)			(0.014)	(1.0)	(0.044)	(0)	(0.043)
B. PIEDMONT							
TMF-1	192	145	0.755	12	0.449	0.994	0.928
TMF-2	240	196	0.817	5	0.652	0.998	0.968
ATL1-1	191	160	0.838	8	0.344	0.996	0.887
ATL1-2	266	165	0.620	13	0.186	0.993	0.967
ATL1-3*	72	49	0.681	8	0.696	0.976	0.828
ATL2-1*	288	246	0.854	5	0.570	0.998	0.934
ATL2-2	240	148	0.854	23	0.345	0.982	0.895
ATL2-3	192	116	0.604	11	0.371	0.990	0.958
Mean			0.723	10.6	0.452	0.991	0.921
(SD)			(0.106)	(5.83)	(0.175)	(0.008)	(0.048)
Mean **			0.682	12	0.391	0.989	0.924
(SD)			(0.087)	(6.1)	(0.154)	(0.008)	(0.055)

* Plots with flowering individuals only (i.e. non-reproductive individuals present, but not sampled)

** Mean values and standard deviation excluding plots with flowering individuals only

N = sample size (# of individuals analyzed/plot)

G = # of unique multilocus genotypes/plot

N_{eg} = effective # of genotypes/plot

G/N = genotype diversity (probability that a randomly chosen plant has a unique genotype)

Max n_i = maximum number of ramets/multilocus genotype (i.e. largest clone in a plot)

D = Complement to Simpson's Index

E = Fager's Evenness Index

Table 4.3: Within population (among plots) genetic structure measured by G_{ST} statistics and number of private alleles (i.e. alleles detected in only one plot). Mean expected and observed heterozygosity values are also reported for each population. Each population is represented by three plots.

		Private alleles				
	G_{ST}	Number	Mean frequency	Rare alleles*	H_e (observed)	H_e (expected)
A. Mountains						
JKMF	0.014	5	0.038	3	0.269	0.317
GRM	0.021	6	0.072	2	0.324	0.348
B. Piedmont						
TMF	0.025	9	0.019	4	0.328	0.346
ATL-1	0.017	6	0.069	1	0.319	0.330
ATL-2	0.018	5	0.053	1	0.344	0.367

G_{ST} : measure of sub-population divergence (Nei, 1973; 1977). This statistic (G_{ST}) estimates the proportion of the total genetic diversity found among plots for each polymorphic locus; G_{ST} values were averaged across loci to obtain an overall estimate of population divergence.

***Rare alleles** are alleles detected in two out of three plots in each population.

Table 4.4: Summary of spatial autocorrelation analyses for the two nearest distance classes. In all plots, results are reported separately for analyses with and without clonally reproduced individuals.

	r		***r***		***d***
	(clones excluded)		**(all plants, including clones)**		**FSGS spatial**
	0-10 cm	**10-25 cm**	**0-10 cm**	**10-25 cm**	**extent (cm)**
A. Mountains					
JKMF 1	0.506	0.254	0.694	0.308	300
JKMF 2	0.343	0.251	0.419	0.265	325
GRS 1	0.425	0.185	0.514	0.248	350
GRS 2	0.349	0.223	0.418	0.235	200
Mean	0.406	0.228	0.518	0.264	294
(SD)	(0.077)	(0.032)	(0.131)	(0.032)	(66.0)
B. Piedmont					
TMF 1	0.455	0.431	0.603	0.485	200
TMF 2	0.479	0.302	0.580	0.479	145
ATL1-1*	0.498	0.338	0.743	0.510	300
ATL 1-2	0.154	0.099	0.229	0.100	110
ATL 1-3	0.093	0.066	0.393	0.117	50
ATL 2-1	0.412	0.141	0.741	0.262	70
ATL 2-2	0.242	0.180	0.870	0.525	85
ATL 2-3*	0.482	0.288	0.693	0.391	300
Mean	0.352	0.231	0.607	0.342	158
(SD)	(0.163)	(0.128)	(0.208)	(0.169)	(99.5)

* Plots with flowering plants only

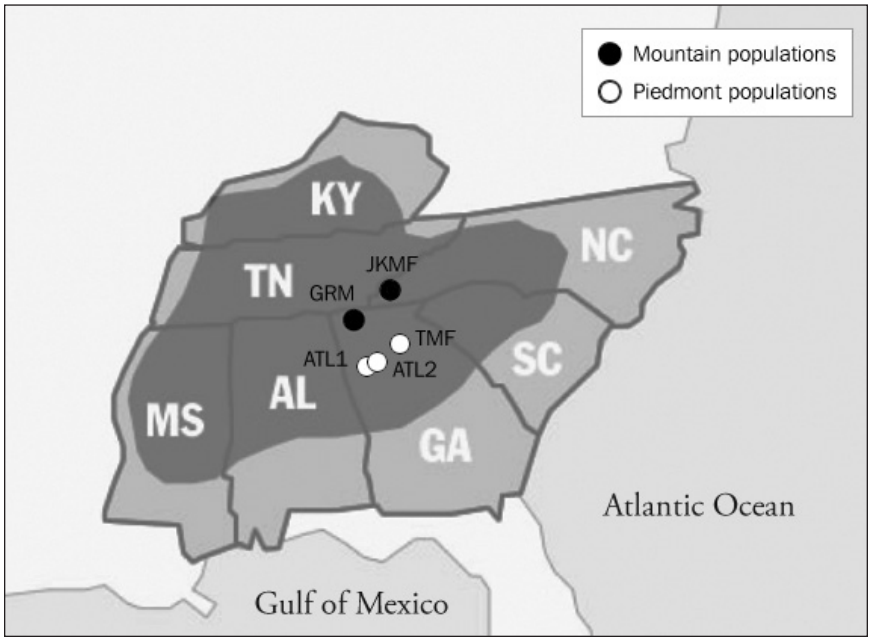

Figure 4.1: Geographic distribution of sampled populations. Populations located in the Southern Appalachian Mountains are indicated by black dots, while the Piedmont Province populations are represented by light dots. The overall *Trillium cuneatum* geographic distribution is indicated by dark grey area within SE US.

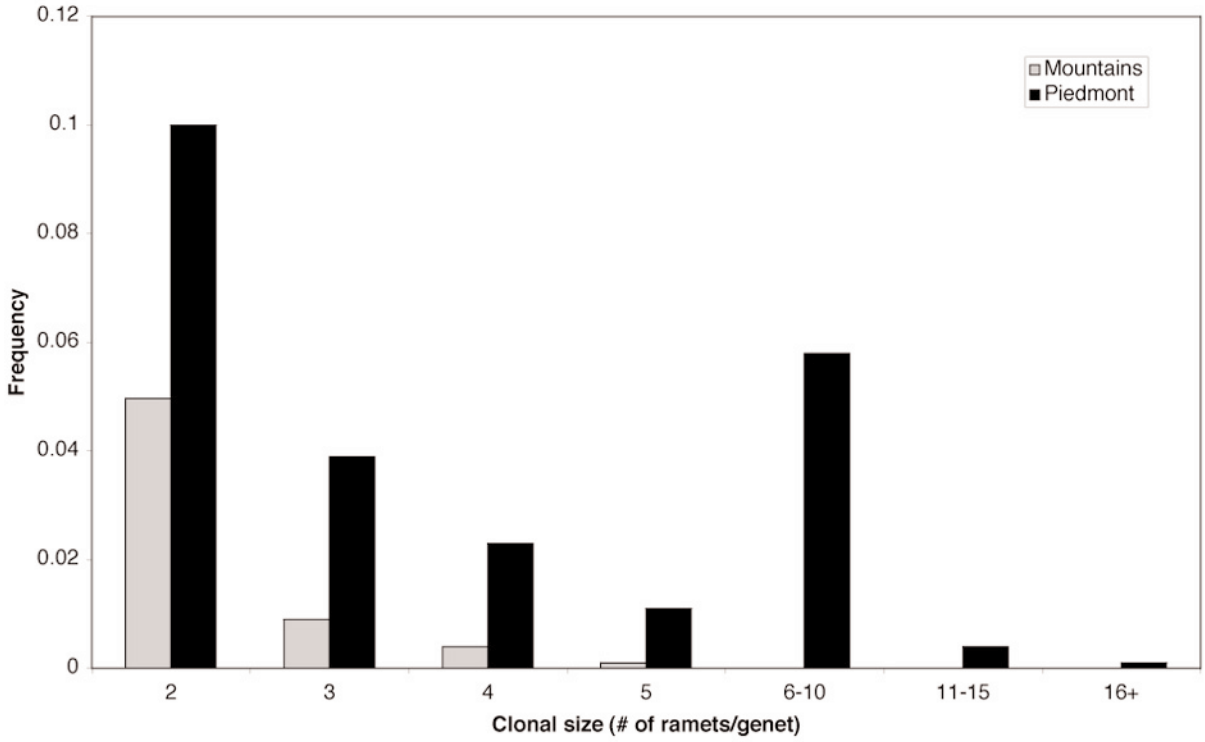

Figure 4.2: Differences in clonal structure of *Trillium cuneatum* in two geographic regions. All samples within each region were pooled for comparison (N_{MT} = 938, N_{PD} = 1681). Plants in the Piedmont form large clumps and clone more frequently; while clonal reproduction in the mountains is rare and large clusters of ramets with identical genotypes are absent.

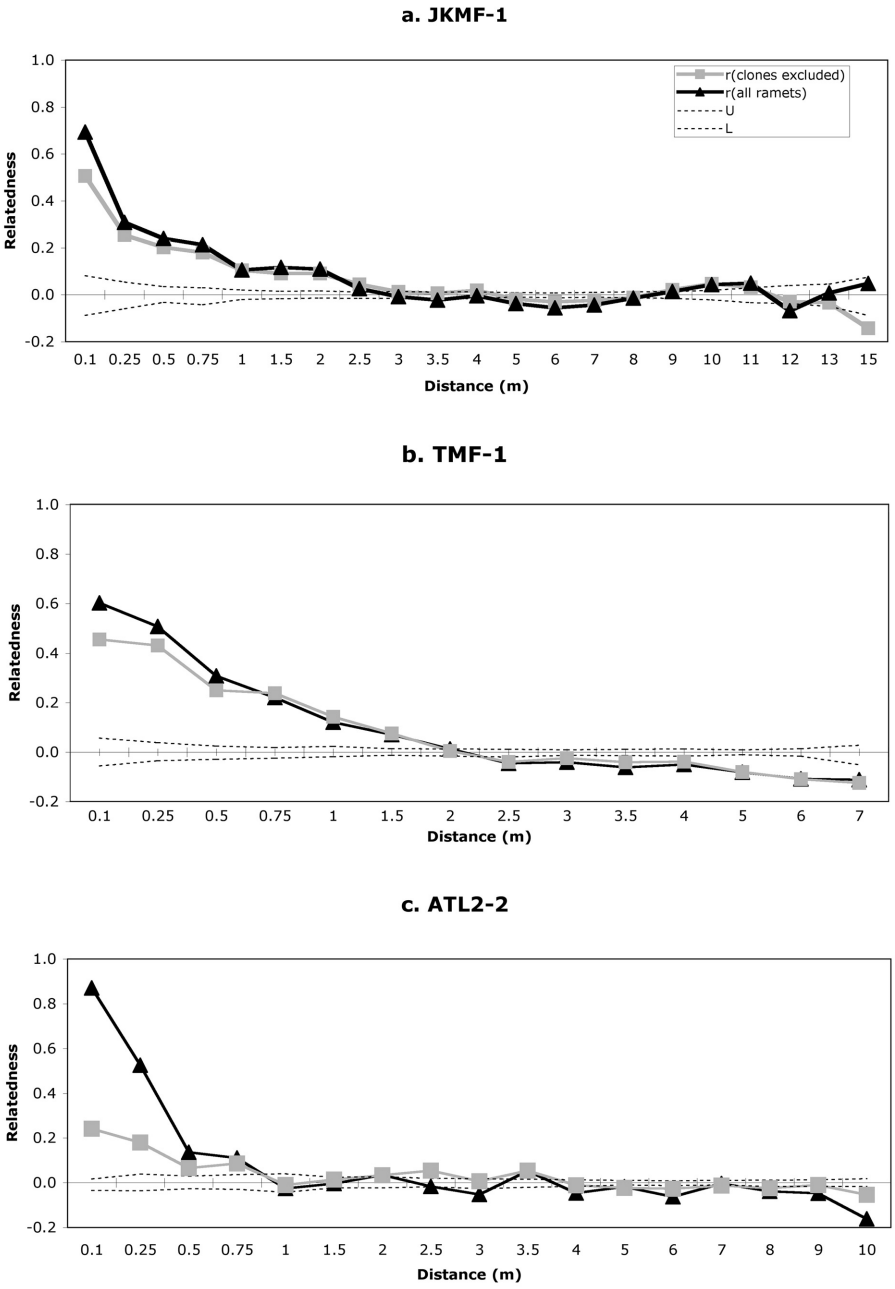

Figure 4.3: Spatial autocorrelograms of estimated relatedness and spatial extent of genetic neighbourhoods for three plots representing differences between the mountain (JKMF-1) and the Piedmont populations (TMF-1 and ATL2-2). Only the first seven meters are displayed. Correlograms of all ramets are represented by black line with triangular symbols; grey lines with squares indicate the relationship between plants after excluding clones. The 95 % confidence envelope around the null hypothesis r = 0 is indicated by dashed lines.

CHAPTER 5

COMPARISONS OF POLLEN MEDIATED GENE FLOW WITHIN LARGE POPULATIONS OF A FOREST UNDERSTORY SPECIES, *TRILLIUM CUNEATUM*[1]

[1] EVA GONZALES, JL HAMRICK, PE SMOUSE and RJ DYER. To be submitted to Molecular Ecology.

ABSTRACT

Pollen movement plays a critical role in the distribution of genetic variation within and among plant populations. Direct measures of pollen movement in the large, continuous populations that characterize many herbaceous plant species are often technically difficult and biologically unreliable. Here, we studied contemporary pollen movement in four large populations of *Trillium cuneatum*. Three populations located in the Georgia Piedmont are exposed to strong anthropogenic disturbances, while the fourth population, located in the Southern Appalachian Mountains is relatively undisturbed. Using the recently developed TWOGENER analysis, we compared estimates of the effective number of pollen donors *(N_{ep})*, effective mating neighbourhood size (A_{ep}) and the average distance of pollen movement (δ) in each of these four populations. Our estimates indicate that maternal plants do not sample pollen randomly from a global pool; rather, pollen movement in all four populations is highly restricted. Although the effective number of pollen donors per maternal plant is low (1.22-1.66) and pollen movement is highly localized in all populations, N_{ep} in the disturbed Piedmont populations is higher and there is more pollen movement than in the mountains. The distance pollen moves is greater in disturbed sites and fragmented populations, possibly due to edge effects in *Trillium* habitats.

Introduction

Gene dispersal is a key factor in the maintenance and distribution of genetic variation. Pollen-mediated gene movement, in particular, has been advocated as a major factor that shapes the distribution of genetic diversity within and among plant populations (Hamrick & Godt 1989, Streiff *et al.* 1999). Much of the recent interest among evolutionary and conservation biologists in studying gene flow stems from two seemingly diverse sources. First, there is considerable concern regarding the impact of habitat and population fragmentation on the ability of species to maintain natural levels of genetic diversity (e.g. Sork *et al.* 1999, Irwin *et al.* 2003, Young *et al.* 1996). Second, with the introduction of genetically modified organisms, there is a need to determine the potential impact of such organisms on natural plant populations. Knowledge of rates and patterns of gene flow are essential if we are to predict the effects of either of these anthropogenic disturbances (e.g. Burke *et al.* 2002, Celis *et al.* 2004, Jarosz *et al.* 2004, Stevens *et al.* 2004).

Historically, two approaches have been employed to quantify gene movement. *Indirect* methods (e.g. Wright 1951, Slatkin 1985) transform measures of genetic structure into the number of migrants per generation (N_em). This approach, based on the contemporary distribution of genetic diversity, reflects the cumulative effects of both pollen and seed mediated gene dispersal and provides insights into historical levels of gene flow (Neigel 1997). *Direct* approaches, based on parentage analyses, permit estimates of contemporary gene flow. They provide details about contemporary gene movement (e.g. Devlin *et al.* 1988, White *et al*. 2002, Burczyk *et al.* 2002) but are usually constrained to one or a few small sites by the need to identify all potential pollen donors and to genotype large progeny arrays. Paternity analyses work best with a limited number of potential pollen donors, but in large populations, the number of potential pollen donors within even a limited area may exceed our ability to delineate and sample them. In addition, even with high exclusion probabilities based on hyper-variable microsatellite markers, considerable paternal ambiguity may remain. These considerations have generally limited parentage analyses to small, discrete populations, and only a limited set of species and populations have been characterized for

their patterns of mating. As a result, it is not usually possible to adequately describe the effects of landscape variables on rates and patterns of gene flow (Sork *et al.* 1999, Smouse & Sork 2004).

Yet, many plant species occur in large, continuous populations, with potential pollen donor pools consisting of thousands of flowering individuals. *Trillium cuneatum*, the subject species of our investigation, is a long-lived understory spring ephemeral, often carpeting floors of mature mesic deciduous forests from southern Kentucky, through Tennessee to central Mississippi and Alabama, and eastward into Georgia and the Carolinas. It occupies both large continuous and smaller disturbed (remnant) habitats. This woodland understory herb often occurs in populations with thousands of reproductive individuals, a situation typical of many herbaceous species occupying a variety of habitats (e.g., grasslands, coastal marshes, mountain meadows, and forest understories). However, although thousands of *T. cuneatum* individuals flower each spring, less than one percent successfully yields a fruit during any given season (EG pers. observ.). Unpredictability and low reproductive success render studies of gene movement based on traditional paternity analyses, unrealistic. Despite their potential evolutionary and ecological impact, patterns of gene movement within such populations remain difficult to quantify reliably. As a result, gene flow in large continuous populations remains poorly studied. Moreover, a critical question that needs to be addressed is the extent to which the landscape context of populations influences gene movement (Sork *et al*. 1999), regardless of whether landscape heterogeneity is natural or has been created by recent anthropogenic disturbances. Clearly, we need new approaches to measure contemporary gene movement in such plant populations to be aware of potential consequences of changing environmental conditions and habitat alteration on gene dispersal.

One such attempt is represented by the analytical method dubbed TwoGener, which combines straightforward analyses of the indirect genetic structure approach with the direct parent-offspring deductive aspect of parentage analyses (Smouse *et al.* 2001). TwoGener uses a statistic, Φ_{ft}, extractable from a molecular analysis of variance (AMOVA) (Excoffier *et al*. 1992) of the male gametes that fertilized several ovules from a number of maternal individuals. The Φ_{ft} statistic (analogous to F_{ST}, Wright's (1951) and Weir & Cockerham's (1984) θ coefficients of differenti-

ation among populations) represents the proportion of the pollen pool variance attributable to differentiation among pollen pools sampled by different maternal plants. Assuming selectively neutral markers, no local genetic structure among reproductive individuals, equally fecund and phenologically synchronized adults and normal distribution of pollen dispersal, the overall effective number of pollen donors (N_{ep}) can be derived for individual maternal plants from the Φ_{ft} estimate. We can further deduce information about the size of the effective pollination/mating neighborhoods and the effective distance of pollen movement (Austerlitz & Smouse 2001a, Smouse *et al.* 2001, Austerlitz *et al*. 2004). Genotypes of maternal plants and their progeny provide data for this analysis, with no need to identify pollen donors.

TwoGener represents a step toward understanding pollen dispersal, because it reflects contemporary dynamics of gene exchange and allows studies of gene movement in a landscape context, regardless of population size. The approach provides a means of assessing contemporary gene movement, and rather than assuming that propagule movement is independent of environmental context, it allows us to determine how environmental heterogeneity and anthropogenic changes in landscapes influence contemporary gene flow, how far propagules move, and how landscape features impact pollen movement patterns. We consider the TwoGener method particularly well suited for analyses of large continuous herbaceous plant populations, because it depends only on the identification, sampling, and location of relatively few of the thousands of potentially reproductive individuals.

In this paper, we address the pollination dynamics of four large populations of *Trillium cuneatum*, using TwoGener methods. Our first objective is to *estimate* and *compare* pollen movements in four populations located in different habitats (Figure 1). In each population, we first ask whether maternal plants sample pollen from the same global pollen pool (i.e., randomly) or maternal plants sample from different pools of male gametes, by virtue of their physical location. Our second objective is to use this information to estimate and compare the effective number of pollen donors per maternal plant, the effective size of the pollination neighborhoods and the effective pollen movement distances for each locality. Finally, we compare the pollen dynamics of the four populations.

Materials and Methods

The reproductive biology of *Trillium* is poorly understood. Plants typically require more than ten years to reach the reproductive stage (Ohara 1989, Jules 1996), and sexual reproduction (i.e., successful fruit production) is infrequent. We know nothing directly of *Trillium cuneatum* mating system, but conflicting information exists in literature. For example, Fukuda (1987) reported that most North American *Trillium* species are autogamous, but Kalisz et al. (1999) reported an outcrossing mating system for *T. grandiflorum*. Furthermore, Broyles et al. (1997) and Sage et al. (1997) reported a mixed mating system for *T. grandiflorum* and *T. erectum*. Similarly, Ohara *et al*. (1995) documented some *T. kamtschaticum* self-incompatible populations, while other populations had mixed mating systems.

Although *Trillium cuneatum* reproduces sporadically, both by seeds and clonal spread (EG pers. obs.), individual ramets and genets persist for many decades. In March and April, a mature plant may produce a single flower (rarely two), pollinated by weakly flying insects such as fruit flies and, less frequently, Calliphoridae. Pollination occurs during warm, moist evenings when plants emit an odor reminiscent of fermented fruits (EG pers. obs.). The flower remains open for several days and can be visited by multiple pollinators. The vast majority of plants (i.e., all non-reproductives and flowering individuals that do not set fruit) wither within a few weeks after flowering; the plants remain dormant as underground rhizomes until the following spring. Fruits of successfully reproducing individuals (less than 1%) mature at the end of June/early July. By then, their leaves are also shriveled, and a single fruit remains attached to a weakened stem, often leafless or with dried leaves (EG pers obs). *Trillium cuneatum* is polycarpic; each maternal plant produces a single fruit containing, on average, 30-40 (rarely more) seeds, each with a fleshy elaiosome. Ants, attracted to these elaiosomes, are the primary seed-dispersers and by dispersing seeds, reduce the number of seeds germinating in the immediate vicinity of the maternal plant. Consequently, ant-mediated seed dispersal affects patterns of seedling emergence and relatedness. Additionally, yellow jackets (*Vespula vulgaris*) and deer (*Odocoileus virginianus*) play roles as less frequent seed dispersers (EG pers observ, Jules 1996, Vellend *et al.* 2003).

Study sites. We studied three sites in the Piedmont Province and one site in the Southern Appalachian Mountains of northern Georgia, USA (Figure 1). Two of the Piedmont populations were located in Atlanta; both occur in small, fragmented forests, completely surrounded by urban development and disturbed by human activities. The first (ATL1) is located in the Storza Woods, Midtown, the other (ATL2) adjacent to Emory University. Although these populations occur in small remnants of a mature, secondary forest (approximately 300 x 400 m in extent), their *Trillium* populations are not small. Each contains several thousand flowering individuals. The two remaining populations are located in more extensive forests. In the University of Georgia's Thompson Mill Experimental Forest (TMF), *T. cuneatum* occurs patchily, mostly along the margins of 130 hectares of a mature, secondary deciduous forest, surrounded by pastures. Grassy Mountain (GRS), the forth site, is located in relatively undisturbed Chattahoochee National Forest in the southernmost Appalachian Mountains. This largest population constitutes tens of thousands of patchily distributed flowering individuals, spreading over hundreds of hectares of continuous secondary deciduous forest.

Field sampling. The frequency of fruiting individuals was low in all populations (< 1%), and maternal individuals were scattered haphazardly across the landscape. The low frequency of fruiting plants, and our inability to predict (at the time of flowering) which flowering plants would produce fruits, precludes the use of traditional paternity analyses, but it does not impinge on the sampling requirements of TwoGener methodology.

In 2001, we sampled 25 fruits, one per plant, from which we extracted 351 seeds for a pilot study at the TMF site. As the original intent was a mating system analysis, the locations of these plants were not mapped. During the following season (2002), we collected fruit from the other three study sites, 49 fruits from GRS (816 seeds), 83 fruits from ATL1 (1456 seeds) and 79 fruits from ATL2 (1378 seeds). We increased the number of progeny per family from 12 to 18 when enough seeds were available. We mapped the plants for these other sites, so that we have a record of spatial relationships between all maternal individuals sampled in 2002.

Laboratory Analyses. Genetic assay presents some difficulties in *Trillium*. The seeds are difficult to germinate, even under permissive conditions in the greenhouse. *Trillium* seeds have "double dormancy", which means they need to be exposed to two winter seasons to germinate (Case and Case 1997), so we must infer offspring genotypes from seeds, rather than from seedling leaf tissue. Further, we lack available adult leaf tissue from fruiting plants due to the early senescence of leaves. Instead, we used elaiosomes (diploid maternal tissue) to genotype maternal individuals. That has two consequences for the analyses that follow: (a) We can genotype only those adults in the population that are fruiting at the time of seed collection, generally less than 1% of the population within any given season. Any adult contributing pollen that does not subsequently fruit will not be included in the sample, which is problematic for paternity analysis. (b) We obtain progeny genotypes from assaying seed tissue. Whole seeds (sans elaiosomes and seed coat) contain both diploid embryo and triploid endosperm tissues, but the size of the diploid embryo is negligible, relative to that of the triploid endosperm. We followed the approach of Broyles et al. (1987) and Sage et al. (2001), who extracted allozymes from seeds (sans elaiosomes) of the closely related congeners, *T. grandiflorum* and *T. erectum*, for their investigations of mating systems and self-incompatibility. In reality, assay of seeds represents analysis of triploid endosperms associated with the diploid embryos, not of the embryos themselves.

In *T. cuneatum*, we detected an unexpectedly high proportion of heterozygous seeds in families with heterozygous maternal plants, and in some cases an individual seed displayed a third allele (i.e., three alleles per individual). Both observing three different alleles in triploid endosperm tissue and the excess heterozygosity for the two maternal alleles are not consistent with the usual (monosporic) form of megagametophytic development. Such observations are, however, compatible with a relatively uncommon bisporic type of female gametogenesis (Schoen 1980) reported in *Trillium* (Dyer 1963; Zomlefer 1996), but it presents an additional challenge for the inference of the paternal gametic contribution to the embryo. We provide the pertinent details on development of the *Trillium* embryo sac in Appendix A, and we also explain our procedure for extracting male gametic inference from the analysis of the bisporic endosperm and the elaiosome.

We haphazardously selected 18 seeds from each fruit; for small fruits, we used all available seeds. We extracted enzymes by grinding seed tissue (without the seed coat or elaiosome), using an extraction buffer in a mortar and pestle (Wendel & Parks 1982), so as to solubilize and stabilize the enzymes. We then absorbed the extract onto chromatography paper wicks, which we stored at -70^{o}C until electrophoretic analyses. We used starch gel electrophoresis to determine genotypes of all seeds for 10-13 polymorphic allozyme loci. The number of polymorphic loci varied among populations. Five gel-electrode buffer combinations were employed to resolve the following loci on 11% starch gels: Buffer 6: diaphorase (DIA), fluorescent esterase (FE-1, FE-2 and FE-3), Buffer 8: leucine-B-naphthyl-amide phosphatase (LAP-1 and LAP-2), Buffer 4: isocitrate dehydrogenase (IDH), Buffer MC: malate dehydrogenase (MDH-1, and MDH-2), UTP-glucose-1-phosphate (UGPP-1 and UGPP-2), Buffer 11: phosphogluco-mutase (PGM) and 6-phosphogluconate dehydrogenase (6-PGDH). Stain recipes for AAT and DIA are given in Cheliak & Pitel (1984); UGPP is described in Manchenko (1994). All other stain and buffer recipes were taken from Soltis et al. (1983). For enzymes with more than one locus, the loci were numbered sequentially, with the number '1' assigned to the most anodal.

Data Analyses. Our principal goal was to use the observed genetic structure of spatially separated pollen pools to draw inferences about mean pollen dispersal distances and effective pollination neighborhood sizes. The TwoGener analysis of pollen pool structure utilizes the AMOVA (Excoffier et al. 1992) model for the analysis of pollen structure; the detailed analytical procedure has been published in Smouse et al. (2001). In general terms, the TwoGener analysis relies upon the multilocus pair-wise squared genetic distance matrix, which serves as input for the AMOVA, with families (maternal plants) as the strata and paternal gametes within them as replication. AMOVA computes components of variation; i.e., it separates variation into within- and among-family components. Instead of the customary Fisher's F-ratio, TwoGener calculates the intra-class correlation Φ_{ft}, which is the fraction of the total variance accounted for by inter-family differences. Significance testing of the Φ_{ft} value within each population is conducted by randomly shuffling male gametes among maternal plants, on the null premise that if there were no differ-

ences among maternal pollen pools, it would not matter which collection of male gametes was associated with which maternal plant.

Estimates of the effective number of pollen donors, effective pollination neighborhood size and effective pollen movement distance were obtained as in Smouse *et al.* (2001) and Austerlitz & Smouse (2001). They provide conversion equations based on a relationship $\Phi_{ft} = (1/8\pi\sigma^2 d_e)$, (where σ^2 = variance in pollen flow distance, d_e = effective reproductive density of potential pollen donors). Inserting our estimates of Φ_{ft}, along with an estimate of adult density, we obtain estimates of σ^2, assuming isotropic pollen flow. Austerlitz and Smouse (2001) show that $N_{ep} = (1 / 2\Phi_{ft}) = 8\pi\sigma^2 d_e$, and that effective neighborhood size $A_{ep} = 4\pi\sigma^2$, a circle of radius σ. The average pollination distance is .

Results

Genetic resolution. The statistical robustness of our estimates is a function of genetic diversity in each population, and it can be expressed in terms of the multilocus exclusion probability (E_L) (Jamieson 1994). The number of polymorphic loci, the number of alleles per locus and their frequencies varied from population to population (Table 1). We used population specific allele frequencies to calculate E_L values for each site (range = 0.910-0.946) (Table 1). Smouse *et al*. (2001) showed that resolution above $E_L = 0.7$ is sufficient for TwoGener analysis, so we have more than adequate statistical power for our estimates of pollen dynamics.

Pollen structure analysis. The results of the TwoGener analyses represent a striking departure from the null hypothesis of random pollen movement in all four populations ($p < 0.001$). The Φ_{ft} values ranged from 0.301 (ATL1) to 0.409 (GRS) (Table 2), indicating that spatially separated maternal plants sample strongly structured pollen pools in all four populations. Our data provide a robust foundation for the Φ_{ft} estimates. Smouse *et al.* (2001) recommended sampling $K = [1 / \Phi_{ft}]$ for within family replication (K = number of seeds sampled per maternal plant). We sampled 12 – 18 seeds/fruit, but in retrospect, our results suggest that $K = 4$ would have been sufficient to estimate the mean level of differentiation among maternal plants in all populations.

Effective pollinators, pollination neighborhood and pollen movement. We converted the Φ_{ft} values to estimates of the effective number of pollinators N_{ep}. The N_{ep} estimates per fruit (per plant) were low in all four populations, ranging from 1.224 in GRS to 1.662 in ATL2 (Table 2). Similarly, the effective pollination neighborhood was quite small (1.654 $m^2 < A_{\mathrm{ep}} < 3.963$ m^2), with the GRS having an A_{ep} value less than half those of the Piedmont populations. Corresponding effective distance of pollen movement ranged from $\delta = 0.455$ m (GRS) to $\delta = 0.704$ m (TMF) (Table 2).

Discussion

Pollen dynamics within populations. We used the TwoGener analysis to describe pollen movement dynamics, and to determine the spatial distribution of pollen donors from identified mothers within extensive *Trillium cuneatum* populations. Our results reveal that the genetic composition of pollen pools was significantly differentiated among individual maternal plants and that pollen-mediated gene movement in all populations was severely limited. The effective pollination neighborhood (A_{ep}) around a maternal plant is evidently less than $4m^2$, suggesting that maternal plants are preferentially pollinated by pollen drawn from near neighbors less than a meter away (Table 2). Small amounts of pollen may have come from considerably longer distances, but with our observed density of flowering individuals, most pollen is drawn from very few flowering individuals and within a short distance of the maternal seed parent. This raises the possibility of matings among related individuals. Furthermore, the low N_{ep} values observed (Table 2) suggest that a majority seeds within each fruit are sired by only one or two pollen donors, with a few progeny having different fathers, due either to pollen carryover or multiple pollinator visits. We cannot differentiate between these two possibilities.

The effective number of pollen donors, mating neighborhood size and pollen movement distances are consistent with our field observations of pollinators (i.e., weakly flying insects, hovering over patches of flowers, infrequent visits, pollinator activity only during warm evenings, no visits on rainy and/or cold days). In addition, since each *Trillium* individual produces a single

flower, the expected number of fathers would be low, regardless of pollen movement distances**.** Although we conclude with confidence that there are very few effective pollen donors for a single maternal plant, two far-away fathers would yield the same estimates for the other two parameter estimates (A_{ep} and δ). Our assertion that maternal plants sample pollen from nearby pollen donors is based on the assumption of normal pollen distribution (assuming an exponential distribution would yield marginally greater estimates, data not presented here). Additionally, relatedness among near neighbors detected in our companion study of fine scale genetic structure (FSGS) suggests that the invocation of a pollen dispersal distribution, for which the likelihood of pollination (looking backward from the maternal plant) declines with distance, is plausible. Genetic similarity among near neighbors, detected in the FSGS study, is too high to be explained by restricted seed dispersal alone and, thus, indicating mating between related individuals (EG and JLH in prep).

Our estimated effective number of pollen donors is also consistent with previous mating system studies of two congeners, *T. grandiflorum* and *T. erectum*. Broyles *et al.* (1997) observed high correlations of outcrossed progeny, indicating that there is a high probability of full-sib progeny within a fruit. Their estimates of effective mates per fruit are even lower than ours (N_{ep} for *T. grandiflorum* = 1.01; N_{ep} for *T. erectum* = 1.12). These results were further corroborated by Kalisz's *et al.* (1999) investigation of *T. grandiflorum's* mating system. Their estimates of correlated matings within individual plants are also high (0.818) and suggest that most seeds within a fruit are likely to share the same pollen parent.

Moreover, this limited number of pollen donors is not unique to single-flowered species. Studies of mating systems in species producing multiple flowers, for which estimates of correlated matings per fruit are available, demonstrate that while seeds sampled from multiple fruits of a maternal individual may be sired by multiple pollen donors, there is a high probability that a single father sired seeds within the same fruit. For instance, James *et al.* (1998) examined mating systems of three tropical dry forest tree species (*Cedrela odorata, Jacaranda copaia,* and *Stemmadenia donnell-smithii*). While correlated mating within fruits of *C. odorata* was low (r = 0.159), the probability of single father per fruit within *J. copaia* and *S. donnell-smithii* trees was

high (0.814 and 0.970, respectively). Similarly, Brown *et al.* (1986) measured the degree of shared paternity between progeny within a fruit of a multi-flowered herbaceous Australian legume *Glycine argyrea*; they estimated that 85% of the outcrossed seeds within a single fruit had the same father.

Pollen dispersal differences among populations/habitats. Although our findings demonstrate that pollen movement within each population is restricted, they also reveal that it varies with landscape context. Comparison of Φ_{ft} indicates differentiation between GRS and the three Piedmont sites. The greatest pollen structure and most restricted pollen dispersal were observed in the population occupying the large, continuous, and relatively undisturbed Appalachian habitat; the three Piedmont populations showed significantly higher pollen movement and less pollen structure. We observed variation in pollination neighborhood size and effective number of pollen donors among the three Piedmont populations, differentiation in pollen dynamics among these three sites was less than between the Piedmont and Mountain populations. Question now arises regarding possible factors that vary among these sites and that could be responsible for the observed differences.

Climatic conditions. Grassy Mountain is located at 1200 meters above sea level in the North Georgia Appalachian Mountains, while the other three populations are situated at lower elevations (around 300 meters) of the Piedmont Province (Table 1). These two regions differ not only in latitude and elevation, but also in the amount of precipitation and temperature, with the mountains receiving more rain and having cooler temperatures during the flowering season. We have observed an absence of pollinator activity on cool, rainy days. Since mountain populations experience such conditions more frequently, more restricted pollinator activity could potentially explain the differences in pollen movement between GRS and the warmer and dryer Piedmont populations. The effect of climatic conditions on pollinator activity could, however, be confounded with another factor: anthropogenic disturbance.

*Anthropogenic disturbance***.** European settlers deforested the Piedmont Province and converted its previously contiguous ecosystems into an archipelago-like landscape, with small natural

habitats embedded in agricultural and urban zones. Currently, all *Trillium* habitats in the Piedmont are restricted to small forest fragments, in which they experience an increased edge effect, (i.e. exposure to new abiotic and biotic factors such as increased light and temperature) compared with the pre-fragmentation state.

Not surprisingly, habitat loss, disturbance and fragmentation are cited as major forces leading to the general disruption of natural pollination systems (James *et al.* 1998, Kremen & Rickett 2000). Over the last few decades, the perception has been that pollinators have declined in numbers, resulting in decreased fruit and seed set, contributing to the idea that habitat fragmentation often leads to reductions in pollinator effectiveness (Allen-Wardell *et al.* 1998, Buchmann & Nabhan 1996, Kearns *et al.* 1998). More recent studies have demonstrated that habitat fragmentation can lead to a mixture of positive and negative effects on pollinators. Research illustrating the complex, often unintuitive, responses of pollination systems to multiple anthropogenic insults documents that the consequences of pollination disruption vary widely in complex, idiosyncratic ways. Very little is known about anthomophilous flies (i.e. *T. cuneatum* pollinators, among other species) and the consequences of human disturbances on their abundance. They tend to be small, non-charismatic, difficult to observe and identify. As a result, their role as pollinators is poorly understood (Kearns 2001).

For the *Trillium cuneatum* populations studied, fragmentation-associated edge effects and human disturbance appear to have promoted greater pollen movement. *Trillium* populations in disturbed habitats are exposed to greater radiation flux, potentially promoting insect activity that may facilitate increased pollen movement and fruit set. For example, in TMF population, *Trillium* fruits were predominantly located along forest margins. Additionally, the two Atlanta populations are surrounded by urban development, potentially promoting early flowering and insect activity, due to higher ambient temperatures. Urban habitats experience human modification of naturally vegetated surfaces, with replacement by buildings and paved streets that absorb solar radiation and storing heat, leading to an increase in local temperature. As a result, the Atlanta populations typically flower a few days earlier than populations located beyond the metropolitan area.

Trails in the two populations are also littered with soft drink cans and beer bottles, banana peels and other discarded fruits. Pollinators are also attracted to compost piles and trash cans in human neighborhoods that surround both *Trillium* populations in Atlanta (EG pers. observ.). While fragmentation has led to declines of pollinator populations in some systems, to the point that they are no longer able to provide adequate pollination, urban disturbance in our system seems to operate in the opposite direction.

Other examples of disturbance-enhanced pollinator abundance exist. For instance, Cane (2001) examined the consequences of growing urbanization for populations of pollinating bees. While the abundance of ground nesting bee species declined, urbanization led to an increase of nesting resources for cavity nesting pollinating bees and actually increased the abundance of such pollinators, relative to intact desert shrub habitat. Similarly Sork *et al*. (2005) found that insect pollinated forest subcanopy tree species, *Cornus florida*, experienced increased pollen mediated gene flow and an increase in N_{ep}, associated with habitat disturbance.

Another factor may play a role in determining the magnitude of Φ_{ft}. TwoGener analysis assumes the absence of spatially distributed genetic structure among reproductive individuals, an assumption violated in *Trillium* populations. We have detected statistically significant within-population genetic structure of flowering/adult individuals (Gonzales and Hamrick in prep). Where such fine-scale genetic structure exists and where pollen sampled by a maternal individual is drawn mostly from nearby (i.e. related) individuals, estimates of Φ_{ft} will be inflated (Austerlitz and Smouse 2001) and as a consequence, estimates of pollination neighborhood and pollination distance will be decreased, but how large the inflation may be remains unclear. The methodology for extracting the impact of observed adult structure (Gonzales and Hamrick in prep) of Φ_{ft} is still somewhat limited (Austerlitz and Smouse, 2001; Dyer et al. 2003), and we are currently developing computer simulations, which should shed some additional light on the subject.

The TwoGener analysis has provided unexpected insights into the breeding patterns of populations of a plant species that are too refractory for traditional paternity analyses. Our study

shows the value of the TwoGener approach for studies of gene flow in a species that occurs in large, dense populations or those that have a low and unpredictable fruit set and predation. Our results have allowed us to demonstrate that there are differences in the pollen movement dynamics and effective number of pollen donors that are consistent with the characteristics of the four sites analyzed. Counter to the usual dogma, our results revealed that disturbed populations in the Piedmont have greater fruit set and greater pollen movement, a fact that we attribute to the edge effects generated by fragmentation and other human disturbance.

Acknowledgments

We are grateful to lab technicians and undergraduate students for their assistance in the field and laboratory work, especially to C Deen, L Kappa, S Conrad, T Bercikova, and K Bercikova. This research was supported by NSF-DEB-0211526, NSF-BCS-0327147 and Department of Plant Biololgy Palfrey Fund grant; USDA/McIntire-Stennis 17309, NSF-DEB-0211430 and NSF-DEB-0089238).

References

Allen-Wardell G, Bernhardt P, Bitner R, Burquez A, Buchmann S, Cane J, Cox PA, Dalton V, Feinsinger P, Ingram M, Inouye D, Jones CE, Kennedy K, Kevan P, Koopowitz H, Medellin R, Medellin-Morales S, Nabhan GB (1998) The potential consequences of pollinator declines on the conservation of biodiversity and stability of food crop yields. *Conservation Biology*, **12**, 8-17.

Austerlitz F, Smouse PE (2001a) Two-generation analysis of pollen flow across a landscape. II. Relation between ϕ_{ft}, pollen dispersal, and inter-female distance. *Genetics,* **157,** 851-857.

Austerlitz F, Smouse PE (2001b) Two-generation analysis of pollen flow across a landscape. III. Impact of adult population structure. *Genetical Research*, **78,** 271-280.

Austerlitz F, Dick CW, Dutech C, Klein E, Oddou-Muratorio S, Smouse PE and Sork VL (2004) Using genetic markers to estimate the pollen dispersal curve. *Molecular Ecology,* **13**, 937-954.

Brown, AHD, JE Grant, R Pullen (1986) Outcrossing and paternity in *Glycine argyrea* by paired fruit analysis. *Biological Journal of the Linnean Society*, **29**, 283-294.

Broyles, SB, Sherman-Broyles SL, Rogati P (1997) Evidence of outcrossing in *Trillium erectum* and *T. grandiflorum* (Liliaceae). *Heredity,* **88,** 325-329.

Buchmann SL, Nabhan GP (1996) The Forgotten Pollinators. Island Press, Washington, DC.

Burczyk J, Adams WT, Moran GF, Griffin AR (2002) Complex patterns of mating revealed in *Eucalyptus regnans* seed orchard using allozymes markers and the neighbourhood model. *Molecular Ecology*, **11**, 2379-2391.

Burke JM, Gardner KA, Rieseberg LH (2002) The potential for gene flow between cultivated and wild sunflower (*Helianthus annuus*) in the United States. *American Journal of Botany*. **89,** 1550-1552.

Cane JH (2001) Habitat fragmentation and native bees: a premature verdict? *Conservation Ecology*, **5**, Art No. 3.

Case FW, Case RB (1997) Trilliums. Timber Falls Press, Portland, Oregon, USA.

Cheliak WM, Pitel JA (1984). Techniques for starch gel electrophoresis of enzymes from for-

est tree species. Information report P1-X-42. Petawawa National Forestry Institute. Canadian Forestry Service, Agriculture, Chalk River, Ontario.

Devlin B, Roeder K, Ellstrand NC (1988) Fractional paternity assignment: theoretical development and comparison to other methods. *Theoretical and Applied Genetics*, **76**, 369-380.

Dyer AF (1963) Endosperm development after controlled pollination within and between species of *Trillium* and *Paris*. *Chromosoma*, **14,** 549-567.

Dyer RJ, Westfall RD, Sork VL, Smouse PE (2004) Two-generation analysis of pollen flow across a landscape. V. A stepwise approach for extracting factors contributing to pollen structure. *Heredity*, **92**, 204-211.

Excoffier L, Smouse PE, Quattro JM, (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics*, **131**, 479-491.

Fukuda, I (1987). Mode of speciation in *Trillium*-chromosome evolution and differentiation of the reproductive system. *American Journal of Botany*, **74,** 641-642.

Hamrick JL, Godt MJW (1989) Allozyme diversity in plant species. In: *Plant population Genetics: Breeding and germplasm resources*. (eds AHD Brown, MT Clegg, AL Kahler and BS Weir) pp. 430-463 Sinauer, Sunderland, Mass.

Irwin AJ, Hamrick JL, Godt MJ, Smouse PE (2003) A multiyear estimate of the effective pollen donor pool for *Albizia jullibrissin*. *Heredity*, **90**, 187-194.

James T, Vege S, Aldrich P (1998) Mating systems of three tropical dry forest tree species. *Biotropica,* **30,** 587-594.

Jamieson A (1994) The effectiveness of using co-dominant polymorphic allelic series for (1) checking pedigrees and (2) distinguishing full-sib pair members. *Animal Genetics,* **25**, 37-44.

Jules ES (1996) Yellow jackets (*Vespula vulgaris*) as a second seed disperser for the myrmecochorous plant, *Trillium ovatum*. *American Midland Naturalist,* **135**, 367-369.

Kalisz, S, Hanzawa FH, Tonsor SJ, Thiede DT, Voight S (1999) Ant-mediated seed dispersal alters patterns of relatedness in a population of *Trillium grandiflorum. Ecology,* **80,** 2620-2634.

Kearns CA (2001) North American dipteran pollinators: assessing their value and conservation status. *Conservation Ecology*, **5**, 5.

Kearns C, Inouye D, Waser N (1998) Endangered mutualism: the conservation of plant-pollinator interactions. *Annual Review of Ecology and Systematics*, **29,** 83-112.

Kremen, C, Ricketts T (2000) Global perspective on pollination disruptions. *Conservation Biology*, **14**, 1226-1228.

Maheshwari SC (1955) The occurrence of bisporic embryo sac in angiosperms – critical review. *Phytomorphology*, **5**,67-99.

Manchenko GP (1994). *Handbook of detection of enzymes on electrophoretic gels. CRC Press,* Ann Harbour, MI. CRC Press, Ann Harbour, MI.

Neigel JE (1997) A comparison of alternative strategies for estimating gene flow from genetic markers. *Annual Review of Ecology and Systematics*, **28**, 105-128.

Ohara, M (1989) Life history evolution in the genus *Trillium*. *Plant Species Biology,* **4**, 1-28.

Ohara, M, Haruko T, Ohno Y, Shimamoto Y (1995) Variations in the breeding system and the population genetic structure of *Trillium kamtschaticum*. *Heredity*, **76**, 476-484.

Sage, TL, Griffin SR, Pontieri V, Drobac P, Cole WW, Barrett SCH (2001) Stigmantic self-incompatibility and mating patterns in *Trillium grandiflorum* and *Trillium erectum* (Melanthiaceae). *Annals of Botany,* **88**, 829-841.

Slatkin, M (1985) Gene flow in natural populations. *Annual Review of Ecology and Systematics*, **16**, 393-430.

Smouse PE, Dyer RJ, Westfall RD, Sork VL (2001) Two-generation analysis of pollen flow across a landscape. I. Male gamete heterogeneity among females. *Evolution,* **55,** 260-271.

Smouse PE, Sork VL (2004) Measuring pollen flow in forest trees: an exposition of alternative approaches. *Forest Ecology and Management*, **197**, 21-38.

Sork VL, Nason J, Campbell DR, and Fernandez JF (1999) Landscape approaches to historical and contemporary gene flow in plants. *Trends in Ecology and Evolution,* **14**, 219-224.

Sork VL, Smouse PE, Apsit V, Dyer RJ, Westfall RD (2005) A Two-Generation analysis of pollen pool genetic structure in flowering dogwood (*Cornus florida*, Cornaceae) in the Missouri Ozarks. *Americal Journal of Botany*, **92**, 10 pages (in press)

Soltis DE, Haufler CH, Darrow DC, Gastony GJ (1983). Starch gel electrophoresis of ferns: a compilation of grinding buffers, gel and electrode buffers, and staining schedules. *American Fern Journal,* **73**, 9-27.

Sork VL, Nason J, Campbell D, Fernandez J (1999) Landscape approaches to historical and contemporary gene flow in plants. *Trends in Ecology and Evolution*, **14**, 219-224.

Streiff R , Ducousso A, Lexer C, Steinkellner H, ,Gloessl J, Kremer A (1999)
Pollen dispersal inferred from paternity analysis in a mixed oak stand of *Quercus robur* L. and *Q. petraea* (Matt.) Liebl. *Molecular Ecology,* **8**, 831-841.

Vellend M, Myers JA, Gardescu S, Marks PL (2003) Dispersal of *Trillium* seeds by deer: Implications for long-distance migration of forest herbs. *Ecology*, **84**, 1067-1072.

Wendel JF, Parks CR (1982). Genetic control of isozyme variation in *Camillia japonica* L. *Journal of Heredity,* **73**, 197-204.

White, GM, Boshier DH, Powell W (2002) Increased pollen flow counteracts fragmentation in a tropical dry forest: An example from *Swietenia humilis* Zuccarini. *Proceedings of the National Academy of Science of the USA*, **99**, 2038-2042.

Weir, BS, Cockerham CC (1984) Estimating *F-statistics* for the analysis of population structure. *Evolution*, **84**,1358-1370.

Wright S (1951) The genetical structure of populations. *Annals of Eugenics,* **15**, 323-354.

Young A, Boyle T, Brown T (1996) The population genetic consequences of habitat fragmentation for plants. *Trends in Ecology and Evolution,* **11**, 413-418.

Table 5.1: Genetic variation in polymorphic loci examined in four *Trillium cuneatum* populations and resulting multilocus exclusion probabilities. Elevations of the four sites are also given.

Population	Elevation (m)	Number of loci	Mean Number of alleles	Exclusion Probability (E_L)
GRS	1,200	13	3.62	0.956
TMF	300	10	3.40	0.910
ATL1	310	12	3.08	0.934
ATL2	310	12	3.25	0.931

E_L = Exclusion probability (Jamieson 1994)

Table 5.2: Estimates of pollen pool differentiation, effective number of pollen donors, pollination neighborhood, and pollen movement distances in four *Trillium cuneatum* populations based on TwoGener analysis.

Population	J	n	d_e	Φ_{ft}	N_{ep}	A_{ep}	δ
GRS	49	816	0.74	0.409	1.224	1.654	0.455
TMF	22	352	0.40	0.315	1.585	3.963	0.704
ATL1	83	1456	0.42	0.327	1.529	3.639	0.674
ATL2	79	1378	0.52	0.301	1.662	3.196	0.632
Mean (Piedmont populations only)			0.45	0.314	1.592	3.599	0.670

J = total number of maternal plants

n = total number of offspring (seeds) analyzed

d_e = density of flowering individuals/m^2 (density of potential pollen donors)

Φ_{ft} = pollen pool structure (proportion of the pollen pool variance attributable to differentiation among sampled maternal plants)

N_{ep} = effective number of pollen donors

A_{ep} = effective pollination neighborhood (m^2)

δ = average distance of pollen movement (m)

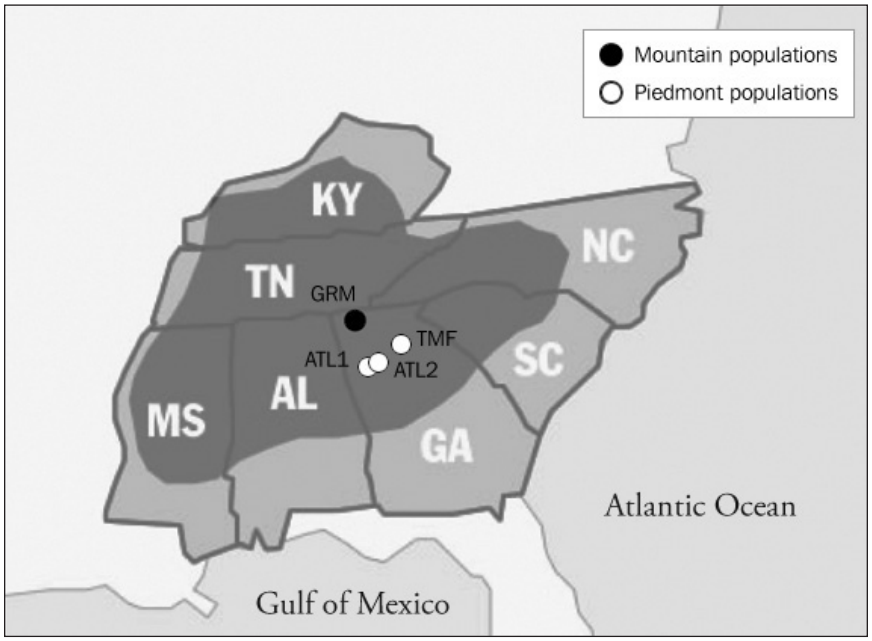

Figure 5.1: Map showing geographic distribution of *Trillium cuneatum* in the SE USA (dark shaded area). Locations of the Piedmont Province populations (ATL1, ATL2, TMF) are indicated by white dots, while the Appalachian Mountains population (GRS) is shown in black.

Appendix A

TRANSLATING DIPLOID ELAIOSOME AND TRIPLOID ENDOSPERM GENETIC ASSAY INTO INFERENCE ABOUT THE PATERNAL HAPLOID CONTRIBUTION

For our allozyme analyses, we assayed elaiosomes (determining maternal genotype exactly) and seeds (consisting mostly of triploid endosperm) to genotype the offspring. From the elaiosomes, we know which mothers are homozygous and which heterozygous. For homozygous mothers, no ambiguity arises concerning paternal contributions to embryonic genotypes (Table A1). For heterozygous mothers, we obtain a combination of categorical and ambiguous pollen assignments. Smouse et al. (2001) have described how to translate diploid maternal and diploid embryo assay into paternal gametic assignments. We will here demonstrate an analogous translation for diploid maternal and triploid endosperm assays, for *bisporic* megagametogenesis.

The most common megagametogenesis developmental pattern in angiosperms is the monosporic type, in which the diploid megaspore mother cell undergoes normal meiosis, resulting in four haploid megaspores; one of them (the germinal spore) will give rise to both the female gametophyte (and the maternal contribution to the endosperm), and the remaining three will abort. In bisporic gametogenesis, meiosis I proceeds normally, giving rise to two daughter cells. One daughter cell aborts, and the surviving daughter cell undergoes meiosis II. However, cytokinesis fails, and a bi-nucleate cell is formed. Both surviving nuclei contribute to the embryo sac, following two mitotic divisions to form an eight-nucleate female gametophyte (Figure A1).

For both types of sporogenesis, recombination occurs in meiosis I. For homozygous mothers, recombination is of no consequence, but for heterozygous mothers it is. For the monosporic form, the maternal contribution to the endosperm is either of the type (A_iA_i) or (A_jA_j). For the bisporic form, the endosperm is derived from two haploid nuclei, so it is possible to obtain a heterozygous (A_iA_j) maternal contribution to the endosperm (Maheswari 1955). Consider the schematic of recombination in meiosis I of a maternal heterozygote (Figure A2).

Heterozygous maternal plants will contribute two identical alleles to the endosperm with probability $(1 - \theta)$, so with probability $(1 - \theta)/2$, we will see a maternal contribution of A_iA_i, and with probability $(1 - \theta)/2$, we will see the alternative maternal contribution of A_jA_j (Table A2). With recombination, with probability (θ), two different alleles (A_iA_j) will be incorporated into the endosperm. The frequencies of various endosperm triploid genotypes will then be a function of recombination rates for each heterozygous locus, as well as the population allelic frequencies. Any paternal allele that is not A_i or A_j (e.g., A_k) will create an obvious paternal allocation to allelic class (A_k), but if the pollen genotype is A_i or A_j (the same as either of the alleles from the maternal individual), potential ambiguity arises. For this last eventuality, there are six different outcomes, detailed in Table A2. The $A_iA_iA_i$ and $A_jA_jA_j$ cells yield categorical paternal designation, A_i and A_j, respectively, but the $A_iA_iA_j$ and $A_iA_jA_j$ cells are ambiguous.

Now, the A_i paternal gamete appears in ambiguous endosperms with probability $(\theta p_i \{A_iA_iA_j\} + (1 - \theta)p_i /2 \{A_iA_jA_j\}) = (\theta p_i + (1 - \theta)p_i/2) = (1 + \theta)p_i/2$ and the A_j paternal gamete appears with probability $(\theta p_j \{A_iA_jA_j\} + (1 - \theta)p_j /2 \{A_iA_iA_j\}) = (\theta p_j + (1 - \theta)p_j/2) = (1 + \theta)p_j /2$. The $(1 + \theta)/2$ term is the same for both, so while we cannot determine whether the father contributed A_i or A_j in such cases, the posterior likelihoods of the endosperm (and hence the embryo) having received A_i or A_j are $\alpha_{ij} = p_i / (p_i + p_j)$ and $\alpha_{ji} = p_j / (p_i + p_j)$, respectively. These are the same results as for the standard diploid embryo case (Smouse et al. 2001).

For the diploid embryo case, half of the embryos from A_iA_j mothers that receive either A_i or A_j male gametes will be homozygotes (A_iA_i or A_jA_j), yielding unambiguous paternal inference, and the other half will be ambiguous heterozygotes (A_iA_j). The monosporic endosperm case (not developed here) yields the same expectation. The bisporic triploid endosperm case (developed above) has a fraction $(1 - \theta)/2$ of unambiguous homozygotes and a fraction $(1 + \theta)/2$ of ambiguous heterozygotes. The frequency of paternally ambiguous progeny from heterozygous mothers increases with the recombinational distance between the centromere and the locus in question for the bisporic endosperm case. We have no knowledge of specific crossover rates (θ) of the loci used for this study, but loci for which the crossover rates are high would be expected to yield a

higher proportion of seeds with the same heterozygous genotype as the mother. Male gametic scoring is identical with that described for the original TwoGener treatment (Smouse et al. 2001), but the frequency of ambiguous male gametes.

Table 5.A1: Categorical assignments of paternal contributions to progeny originating from homozygous maternal families. The paternity inference is based on genotyping elaiosomes (diploid maternal tissue) and seeds (sans seed coat)

Maternal Genotype	Elaiosome Genotype	Endosperm Genotype	Maternal Genotype	Paternal Contribution
A_iA_i	A_iA_i	$A_iA_iA_i$	A_iA_i	A_i
	A_iA_i	$A_iA_iA_j$	A_iA_i	A_j

Table 5.A2: Possible maternal and paternal contributions to triploid endosperm for bisporic gametogenesis, and their frequencies for a maternal heterozygote (A_iA_j). There are six possible triploid outcomes for the endosperm, two that are categorical for paternal assignment (shaded) and four that provide ambiguous paternal assignments.

Maternal Heterozygote	Maternal Contribution	Endosperm genotype given pollen contribution		Maternal Probability
		A_i	A_j	
A_iA_j mother	A_iA_i	$A_iA_iA_i$	$A_iA_iA_j$	$(1-\theta)/2$
	A_iA_j	$A_iA_iA_j$	$A_iA_jA_j$	θ
	A_jA_j	$A_iA_jA_j$	$A_jA_jA_j$	$(1-\theta)/2$
	Pollen pool frequency	p_1	p_2	

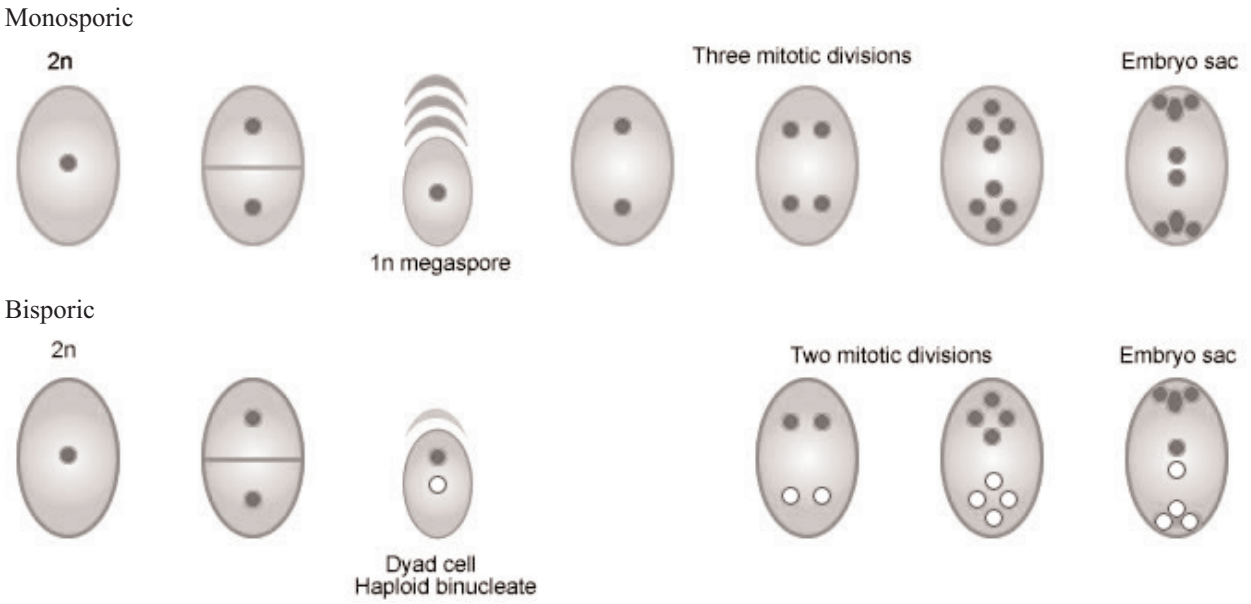

Figure 5.A1: A contrast of monosporic and bisporic female gametophytic development. In the monosporic type, the diploid megaspore mother cell undergoes normal meiosis resulting in four haploid megaspores; one of them (germinal spore) will give rise to the female gametophyte, and the remaining three (somatic megaspores) abort. In the bisporic type, cytokinesis after meiosis I proceeds normally; however, it fails after meiosis II. This results in two haploid binucleate dyad cells. One of them aborts, the other gives rise to the female gametophyte and the endosperm.

Non-recombinant (probability = $1 - \theta$) *Recombinant (probability = θ)*

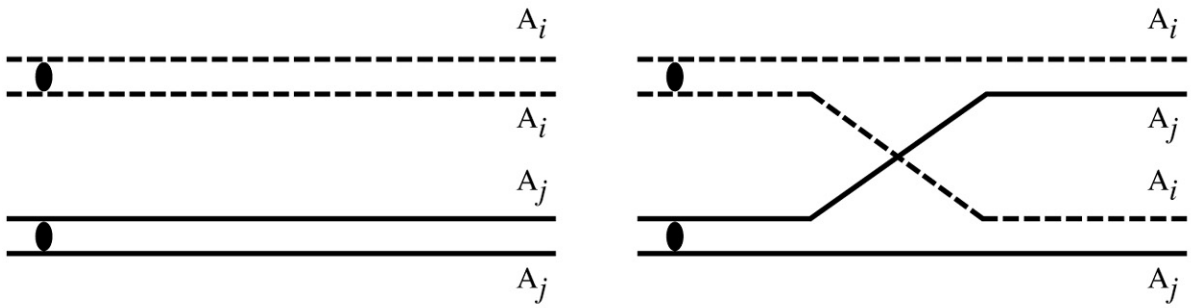

Figure 5.A2: Recombination between chromatids of homologous chromosomes in Meiosis I, between the centromere and the heterozygous A-locus. Recombination, and thus two different alleles (A_iA_j) will be contributed to the endosperm with probability θ.

CHAPTER 6

CONCLUSIONS

This dissertation has focused on micro-evolutionary questions framed in a spatial-temporal context. Such an approach allows the examination of species as dynamic entities, capable of adapting to environmental change, and furthermore to identify those factors that determine evolutionary change. Understanding the evolutionary history of a species requires an integrated approach. Population genetics techniques aimed at such an understanding offer powerful tools for measuring processes occurring within and among populations. They become more effective when combined with other factors, including the plant's life history traits, human pressures on natural populations, physiographic information, and environmental conditions, both contemporary and historical. I used such an integrated approach to investigate the overall objective of my dissertation: to examine the processes that have played major roles in determining patterns of genetic variation in two species of forest understory herbaceous ephemerals, *Trillium reliquum* and *T. cuneatum*, across spatial and temporal gradients.

This broad scale investigation detected strong genetic structure among geographically separated populations of both species of *Trillium*. Disjunct populations of the rare and endangered *T. reliquum* are much more divergent than sympatric populations of its common, albeit also fragmented, congener *T. cuneatum*, in spite of their shared recent history. This observation suggests that rarity in *T. reliquum* is of ancient origins, possibly predating the most recent glacial episode, rather than a consequence of post European colonization of the southeastern United States as previously suggested (Freeman 1975).

Phylogeographic investigation of *T. cuneatum* based on integrated information from molecular markers, fossil evidence, physiographic and traditional biogeographic information revealed several important discoveries. My results provide evidence of more moderate geographic genetic structure with respect to nuclear markers, but remarkably strong genetic differentiation among populations for maternally inherited (i.e. seed dispersed) cpDNA markers. Maternal cpDNA

analyses revealed refugial locations where this southern species survived Quaternary glacial episodes. Remarkably, the southwestern portion of *T. cuneatum's* range, considered by paleoecologists as the main refuge for species associated with temperate deciduous forest, did not participate in the postglacial expansion of the species. Rather, propagules were drawn from refugia in Alabama, Georgia and the southern Appalachian Mountains. We infer pathways of ancient migrations from the geographic patterns of cpDNA haplotype distribution. Propagules originating from different maternal lineages almost never mix, we detected only one polymorphic population at the zone of contact between two lineages. Even more unexpected is our conclusion that *T. cuneatum* must have survived at more northern latitudes than the fossil record and estimates of glacial climates indicate. Furthermore, this study provides molecular evidence for the role of the Tennessee Continental Divide as a corridor for the migration of terrestrial species in their response to postglacial climatic changes.

How can the high genetic structure observed among populations develop and be maintained? Investigation of population dynamics across a smaller spatial scale offers clues into the mechanisms of gene dispersal and the population's ability to maintain relatively high genetic variation and its structure. To address species and population genetic architecture, we have broken these complex questions into contrasting combinations of factors: (*i*) sexual versus vegetative reproduction, (*ii*) seed versus pollen mediated gene dispersal (in addition to clonal spread), (*iii*) contrasting ecological factors: drier and more variable environmental conditions in the Georgia Piedmont versus more mesic and stable conditions in the southern Appalachian Mountains, and finally (*iv*) anthropogenic factors: pristine natural versus fragmented, disturbed habitats. One particularly important result is the association of strong within population genetic structure with genetic structure at a geographic scale. Moreover, this population subdivision is not static; rather it represents a dynamic process responding to a series of natural and human induced factors.

Field observations of pollinators and seed dispersal suggest that gene movement is spatially limited. *Trillium* is a myrmechocorous species; seeds are dispersed primarily by ants over short distances (typically less than three meters). Although recruitment through sexual reproduction is

infrequent, once established, individual plants live for many decades. Additionally, clonal reproduction prolongs their longevity. Application of molecular markers and novel analytical methods to measure pollen-mediated gene movement in large continuous populations (TwoGener analysis) revealed that the effective number of pollen donors per maternal plant is low (less than two) and that the effective distance of pollen movement within populations is even more restricted than is seed dispersal. Correlated matings and spatially restricted gene movement within populations, in addition to spatially limited vegetative spread and longevity, give rise to populations subdivided into a mosaic of small genetic neighborhoods. Such strong fine scale genetic structure within populations is consistent with the relatively high genetic structure observed on a larger geographic scale. However, a question arises as to how this species with such restricted gene dispersal capabilities has become widespread. Although I did not directly address this question, my observation of deer ingestion of mature fruits suggests that deer may be responsible for rare long-distance dispersal events; similar observations have been documented for other forest understory species (Vellend et al 2003), including related *Trillium* species. Additionally, our phylogeographic investigation suggests that *T. cuneatum* survived in multiple refugia, some of them positioned in more northern latitudes than previously believed. Such a spatial distribution of refugial areas would require less long distance seed dispersal to achieve its current geographical range.

Furthermore, by using a comparative approach, I could address questions of how anthropogenic activities (particularly post-European settlement) affected the distribution and maintenance of genetic diversity. Although we found evidence for highly localized gene dispersal in all populations, comparisons among large, continuous and relatively undisturbed populations in the mountains with fragmented and more environmentally stressed Piedmont populations revealed some unexpected results. The most noteworthy difference between the two environments is that fragmentation does not impede pollen mediated gene movement within the fragmented populations; rather it appears to promote it. Pollen in fragmented populations moves over larger distances and more pollen donors contribute to the next generation resulting in more genetically homogeneous populations. Additionally, fragmented populations produce more fruits than their counter-

parts in undisturbed environments. We attributed these differences largely to edge effects: increased light availability and temperature due to urban activities appear to promote greater pollinator activity and extends the photosynthetic period for this forest understory species. However, in spite of higher fruit production in fragmented populations, we found more seedlings in the mountains, suggesting that the more stressful ecological conditions of the Piedmont limit successful recruitment. I concluded that human disturbance increases environmental stress in the Piedmont populations, causing subtle demographic shifts and age structure in these populations.

To what degree are fragmented populations of *T. cuneatum* effectively distressed? Our investigation provides evidence that the consequences of human disturbance and environmental conditions are subtle, although they are statistically significant. So far, populations affected by human activities seem to have coped with the pressures they have experienced during the last few centuries. Comparisons of *T. reliquum* populations presumably isolated for a much longer time than the recently fragmented *T. cuneatum* sites suggests that mechanisms responsible for the loss of genetic variation within populations operates at a very slow pace. A longer time will be needed for *T. cuneatum* populations to suffer the genetic consequences of anthropogenic interference. Most likely, *T. cuneatum's* longevity, its ephemeral habit which narrows the window of opportunity for herbivores, invasive species or pathogens to attack, its clonal reproduction, and its outcrossing mating system has prevented the erosion of genetic diversity in fragmented populations since the European colonization of the southeastern US.

Patterns of genetic variation at the geographic scale are of broad interest to population geneticists, biogeographers, ecologists, and conservation biologists. The population genetic patterns observed through spatial-temporal analyses has increased our understanding of the micro-evolutionary processes acting on *Trillium* populations. A natural extension of this work is to frame additional questions of adaptations in light of the spatial genetic variation revealed across environmental gradients. To make inferences about natural selection and adaptation, themes central to conservation biology, we need to investigate the role the matrilineal history of species plays in the adaptation to changing environmental conditions. Further research is needed to address the relative role

of shared genetic history (i.e. association with maternal lineages) versus shared ecological conditions. For example, phylogeographic studies now allow us to ask questions concerning whether the different maternal lineages have evolved the same coadopted gene complexes in response to similar environmental conditions. Such work is needed to characterize population fitness and the potential for survival in a constantly changing world.

References

Freeman JD (1975). Revision of a *Trillium* subgenus Phyllantherum (Liliaceae). *Brittonia* 27, 1-26.

Vellend M, Myers JA, Gardescu S, Marks PL (2003). Dispersal of *Trillium* seeds by deer: implications for long-distance migration of forest herbs. *Ecology* 84, 1067-1072.