12-2004

Levels and Patterns of Genetic Diversity in the Rare and Endangered Cumberland Stitchwort, *Minuartia cumberlandensis* (Caryophyllaceae)

Charles Thomas Winder
University of Tennessee - Knoxville

To the Graduate Council:

I am submitting herewith a thesis written by Charles Thomas Winder entitled "Levels and Patterns of Genetic Diversity in the Rare and Endangered Cumberland Stitchwort, *Minuartia cumberlandensis* (Caryophyllaceae)." I have examined the final electronic copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Science, with a major in Botany.

<div align="right">Randall L. Small, Major Professor</div>

We have read this thesis and recommend its acceptance:

Karen W. Huges, B. Eugene Wofford

<div align="right">Accepted for the Council:</div>
<div align="right">Carolyn R. Hodges</div>

<div align="right">Vice Provost and Dean of the Graduate School</div>

(Original signatures are on file with official student records.)

To the Graduate Council:

I am submitting herewith a thesis written by Charles Thomas Winder entitled "Levels and patterns of genetic diversity in the rare and endangered Cumberland Stitchwort, *Minuartia cumberlandensis* (Caryophyllaceae)." I have examined the final electronic copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Science, with a major in Botany.

Randall L. Small

_____

Major Professor

We have read this thesis
and recommend its acceptance:

Karen W. Hughes

_____

B. Eugene Wofford

_____

Acceptance for the Council:

Anne Mayhew

_____

Vice Chancellor and Dean of
Graduate Studies

(Original signatures are on file with official student records.)

LEVELS AND PATTERNS OF GENETIC DIVERSITY IN THE RARE AND ENDANGERED CUMBERLAND STITCHWORT, *MINUARTIA CUMBERLANDENSIS* (CARYOPHYLLACEAE)

A Thesis

Presented for the

Master of Science Degree

The University of Tennessee, Knoxville

Charles Thomas Winder

December 2004

Acknowledgements

The completion of this work would not have been possible without the unwavering support of Nadia Ayoub, my wife. Throughout my struggles, she has been an invaluable advisor on matters both intellectual and personal, and she continues to enlighten and enrich my life with every passing day. I cannot thank her enough.

And I must thank Randy Small, who is undoubtedly the "brains" behind this work. By taking me into his lab, having faith in my abilities, and providing the occasional motivational support, Randy has somehow managed to send me down a strange and winding trail of discovery, only to have me end up where he wanted me all along. For his patience, and his willingness to give freely of his time and energy, I owe Randy a great debt, one I can only repay by offering the same support to those who may one day come under my wing.

I wish to express sincerest thanks to Gene Wofford for introducing me to the fascinating plants that are the subject of this work, and that have captivated my interest and stimulated my desire to learn more. I must also thank Karen Hughes, for her patience and understanding, and for her critical advice, which I value greatly.

Last, but not least, I must thank my lab-mates and fellow students, who have done their best to drag me into the community of botanists, and have offered support far beyond that of mere academic obligation. Kindest thanks to Ed Lickey, Joey Shaw, John "Badass" Beck, Wusheng Liu, Dwayne Estes, Pum Grubbs, Keith Bowman, and Chris Fleming,

who were not only my fellow students, but my teachers as well.

Abstract

Sequences of a highly variable nuclear gene (*G3pdh*) were used to characterize genetic

diversity within and among populations of the endangered rockhouse endemic, *Minuartia*

cumberlandensis (Wofford and Kral) McNeill (Caryophyllaceae), and compared to a

widespread and abundant related species, *M. glabra* (Michaux) McNeill. By

reconstructing phylogenetic relationships among *G3pdh* variants (haplotypes) and

observing the geographical distribution of those ordered variants, an attempt was made to

gauge the effects of historical and contemporary population processes acting within the

species, particularly those with potential implications for long-term conservation. Both

M. cumberlandensis and *M. glabra* were found to have high overall genetic (haplotype)

diversity. In *M. cumberlandensis*, most of this variation was distributed among

populations rather than within them ($F_{ST} = 0.63$), while in *M. glabra* the inverse was true,

with variation largely within populations ($F_{ST} = 0.22$). This pattern suggests significant

divergence among populations of *M. cumberlandensis*, likely caused by a reduced

influence of gene flow relative to genetic drift. Observed heterozygosity in populations

of *M. cumberlandensis* was significantly reduced relative to *M. glabra*, suggesting the

effects of inbreeding within small populations. *Minuartia. cumberlandensis* maintains

broad genetic polymorphism among populations, with genetic similarities between many

populations likely resulting from persistent ancestral alleles rather than ongoing gene

flow. Lineages of haplotypes with significantly restricted geographical ranges provide

further evidence for restricted gene flow among populations of *M. cumberlandensis*. The

bulk of genetic diversity in *M. cumberlandensis* is maintained within the largest, densest

cluster of populations, which occurs primarily in Pickett County, with outlying

populations having relatively low, though still significant, portions of the overall

diversity in the species. Because of the small overall number of outlying populations,

and the unique genetic makeup of each, all occurrences should be given the strongest

possible protection from further human habitat disturbance. It is likely that *M.*

cumberlandensis has persisted in its current location through cyclical changes in climate

during the Pleistocene, and if protected from habitat destruction and population

decimation, it may fare well in the face of future climate change.

Table of Contents

List of Tables

List of Figures

Chapter I

Introduction

Rockshelter endemism. The northern portion of the Cumberland Plateau of Tennessee

and Kentucky in eastern North America is composed of massive deposits of

Pennsylvanian sediment, forming a highly heterogeneous assemblage of horizontally

layered sandstones, shales, and conglomerates (Wilson and Stearns 1958). The Big

South Fork of the Cumberland River and its tributaries define a region of highly complex

topography, formed by the sharp incision and downcutting of the river system through

strata of highly variable composition and durability. This complex topography creates a

mosaic of contrasting physical environments, with dry forested uplands, xeric bluffs and

cliffs, mesic slopes and ravines, and moist bottomlands.

In the gorges and steep-walled valleys of this region, differential weathering of exposed

vertical outcrops forms rockshelters, or cave-like recesses beneath large sandstone

overhangs (Donahue and Adovasio 1990). These abundant formations create an unusual

set of environmental conditions, and provide an ecological niche that supports a unique

endemic flora (reviewed in Walck et al. 1996).

Light intensity within rockshelters is generally low, and other factors such as

temperature, humidity, soil moisture, and substrate type can vary depending on parent

material, aspect, elevation, and surrounding vegetation. The highest diversity of endemic

taxa is found in larger shelters (rockhouses) in close proximity to running water, with

substrates of perpetually moist sand. Temperatures within large rockshelters are moderated by the surrounding rock, remaining relatively cool in summer and warm in winter compared to surrounding forests. Walck et al. (1996) documented a large number of vascular plant taxa residing within shelters, including 11 species occurring exclusively behind the dripline of sandstone rockshelters.

Minuartia cumberlandensis. An intriguing element of the endemic rockshelter flora is the Cumberland stitchwort, *Minuartia cumberlandensis* (Wofford and Kral) McNeill (Caryophyllaceae). Described by Wofford and Kral (1979) as *Arenaria cumberlandensis*, but later transferred to *Minuartia* by McNeill (1983), this small, leafy herb is unusual among members of its genus in dwelling in perpetual shade underneath rock overhangs, whereas all other species of *Minuartia* in the eastern United States characteristically inhabit fully exposed sites such as rock outcrops, barrens, and high-elevation balds (Maguire 1951; Weaver 1970; McCormick et al. 1971; Wofford and Kral 1979). In addition to its unusual ecological niche, the extremely small range of *M. cumberlandensis* distinguishes it from many members of its genus.

The phylogenetic context and origins of *M. cumberlandensis* are largely matters of conjecture, but similarities of morphology and cytology (*n*=10) place the rockshelter endemic in the *M. groenlandica* complex (Fig. 3; All figures and tables are located in the appendices) (Wofford and Kral 1979; Wofford 1981; also see Fernald 1919). This complex consists of *M. groenlandica* (Retz.) Ostenf., an arctic/alpine perennial ranging primarily from Greenland and eastern Canada into the alpine areas of New England with

putative relict populations at high elevations in the southern Appalachians, and *M. glabra* (Michaux) Mattfeld, an annual of dry sandstone or granite outcrops at low elevations in the southeastern United States (Baskin and Baskin 1982). The range of *M. cumberlandensis* is entirely disjunct from any extant populations of *M. groenlandica*, but overlaps the range of *M. glabra* on the Cumberland Plateau (Fig. 3). *Minuartia glabra* is similar to *M. cumberlandensis* in its island-like distributional pattern on the northern Cumberland Plateau, but is more widespread, more frequent, and generally occurs in larger populations.

Wofford and Smith (1980), in assessing the conservation status of *M. cumberlandensis*, observed several scattered populations, many concentrated within Pickett State Park, Tennessee, and concluded that its highly restricted distribution, narrow habitat requirements, and potential for human disturbance warranted special protection under the Endangered Species Act (1973). The species was added to the Federal list of endangered species in 1988 (U. S. Fish and Wildlife Service 1988; listed as *Arenaria cumberlandensis*), and ultimately a Recovery Plan was drafted (U.S. Fish and Wildlife Service 1996), outlining the steps necessary to ensure the continued stability of these populations (although no current deterioration was evident). Characterization of overall genetic variation and structuring among populations of *M. cumberlandensis* was identified as a useful component of a long-term conservation strategy for the species, and is a major goal of the current study.

Populations of *M. cumberlandensis* occur as discrete associations of plants restricted to single shelter caves, each consisting of dozens to thousands of perennial individuals. Populations are sporadically distributed throughout the species' range, and may occur in locally dense clusters or in relative isolation from one another (personal observation). Clusters of populations are frequently arranged in linear fashion, following suitable habitat occurring along the bases of extensive bluffs. In steeper gorges, step-like terraces of massive sandstone may support populations (or clusters of populations) with extensive vertical dimension. All known populations of the species occur within five contiguous counties, and are contained within an area less than 45 kilometers in diameter, although most occurrences are clustered within 10 kilometers of one another in the vicinity of Pickett State Park, Tennessee (Figs. 1, 2).

The major ecological constraints upon the distribution of *M. cumberlandensis* appear to be abundant soil moisture, high humidity, cool temperatures, and deep shade (which certainly limits the number of species capable of competing for space and resources within rockshelters) (Wofford and Smith 1980). Many areas within the overall range of the species, while having plentiful rockshelters with apparently suitable growing conditions, are conspicuously lacking in *M. cumberlandensis* populations (personal observation). This absence may be due to unknown ecological constraints, historical factors (such as widespread extinction of populations through human disturbance), limited seed dispersal into these areas, or some combination of those factors. Ecological processes, particularly mechanisms of seed and pollen dispersal among sites, are little studied in this species, but are of critical importance in understanding its current range,

population structure, and genetic diversity. Extant patterns of genetic diversity in *M. cumberlandensis*, however, undoubtedly hold the telltale signature of the past activities of pollen and seed movement, and may prove informative in assessing the ecological significance of these processes.

In shelters where the plants occur, mild conditions allow the distinctive leafy rosettes to thrive throughout the winter months and send up abundant, leafy flowering stems in spring and summer. Peak flowering occurs from May until July, though flowering plants were observed as late as November in 2002 and 2003, suggesting some degree of flexibility with regard to flowering schedule. This flowering period is entirely distinct from that of the spring-flowering winter annual *M. glabra*, the only other species of *Minuartia* occurring in this area, which makes pollen flow between the two species highly unlikely (Wofford and Kral 1979).

The small white flowers of *M. cumberlandensis* appear well suited to a strategy of generalist insect pollination (as in related species; Levesque and Burger 1982), however no pollination activity was observed during the course of extensive field surveys. Various aspects of floral morphology and development (slight protandry and anthers distant from mature stigmas) suggest predominant outcrossing, though pollination biology studies need to be done to determine what degree of selfing actually occurs. Field observations during the course of this study indicate that copious viable seed was produced every year, and young seedlings occurred frequently in most populations.

Dispersal of the abundant tiny seeds appears to occur primarily through simple dropping of seeds directly into sand around the parent plant. Movement of seeds by the activity of water is also likely, given their small size and the typical proximity of plants to running water. Animal activity could very easily play a role in longer distance dispersal of seeds, given the intensive use of shelters by mammals and birds.

Larger rockshelters in this region were heavily used by pre-historic humans (Ferguson et al. 1986), and have been greatly disturbed within recent history by settlers (and their cattle), moonshiners, loggers, artifact hunters, and hikers. These pre-historic and historic human activities could potentially have affected both the overall distribution of *M. cumberlandensis* and patterns of genetic diversity within the species, but the extent of these possible effects is unknown.

Probably the most significant and immediate threats to *M. cumberlandensis* involve wholesale decimation of populations through activities such as alteration of site hydrology, excavation of rockshelters, and clearing of trees that shade rockshelters (Wofford and Smith 1980, U. S. Fish and Wildlife Service 1996). Such direct threats to the existence of this vulnerable species are of highest priority to conservationists, and many of these issues are being successfully dealt with through sound management practices. Fortunately, the majority of extant populations of *M. cumberlandensis* occur on remote public lands, and are thus buffered from the most critical of the threats, which allows us to address more subtle problems that may affect the long-term viability of the species, particularly in the face of shifting environmental parameters caused by climate

change. Threats to genetic integrity, ecological flexibility, and adaptive potential are of particular importance in *M. cumberlandensis*, given its small population sizes, limited number of populations, and the geographical isolation of these populations from one another. A variety of approaches are available for assessing genetic threats and elucidating ongoing and historical population genetic processes acting within endangered species.

Conservation genetics. Typical genetic threats to endangered species result from the deleterious effects of small population size, including loss of accumulated genetic diversity due to random genetic drift, and reduced heterozygosity caused by inbreeding (Barrett and Kohn 1991). In species with fragmented ranges, reduced gene flow among populations can intensify these effects, leading to greater loss of genetic variation within populations and genetic divergence among isolated populations. Over many generations, loss of genetic variation within populations and individuals may negatively impact reproductive ability and ecological flexibility, and may decrease adaptive evolutionary potential by narrowing the range of heritable variation available to selective processes.

Traditional conservation genetic approaches make use of selectively neutral, highly variable unordered molecular markers (e. g. allozymes and microsatellites) for which genealogical relationships among variants (alleles) are unknown, and focus mainly on measuring amounts of variation within populations and assessing how the overall variation within a species is partitioned among populations. The observed distribution of variation among populations is assumed to represent the effects of ongoing gene flow and

genetic drift, from which can be extrapolated potential hazards to genetic integrity. One

drawback of this approach is that it may downplay the contribution of historical factors to

present-day genetic structuring, which may be significant in species that have undergone

recent demographic shifts (Schaal et al. 1998).

Phylogeography. Recent advances in population genetic analysis incorporating

genealogically ordered variation in the form of DNA sequences have expanded the scope

of conservation genetics by providing an effective means of examining population

genetic structure from a historical perspective.   By observing the geographical

distribution of genealogical lineages (the phylogeographic approach), we can infer

patterns of genetic structure resulting from historical processes such as migration, range

expansion, and habitat fragmentation (Avise 2000).

Most phylogeographic studies to date have made use of haploid, uniparentally inherited,

non-recombining, cytoplasmic DNA variation (for example, see Demesure et al. 1996,

Soltis et al. 1997), but in plants this approach is often unfruitful because of insufficient

intraspecific cpDNA variation (Schaal and Olsen 2000). Recent work has increasingly

utilized diploid nuclear loci for phylogeographic analysis in both plants and animals, with

introns of single-copy nuclear genes showing sufficient fine-scale variation for studying

allelic relationships at the intraspecific level (Olsen 2002).

Under optimal conditions, nuclear phylogeographic analysis of plant species and

populations can provide knowledge of the historical population processes responsible for

current distributions of genetic variation, particularly where data from multiple nuclear

loci are available (Hare 2001; Posada and Crandall 2001; Schaal and Olsen 2000; Cruzan

and Templeton 2000; Zhang and Hewitt 2003; Mort and Crawford 2003; García-Gil et al.

2003; Wright et al. 2003; Järvinen et al. 2003; Olsen and Purugganan 2002; Oh and

Potter 2003).

One important application of this knowledge is to further inform conservation

management decisions, with the goal of mitigating the effects of present and future

anthropogenic environmental impacts on the continued existence and viability of plant

lineages. The phylogeographic approach may be particularly applicable to the study and

conservation of endemic (narrowly restricted) plant species such as *M. cumberlandensis*,

whose histories are obscure, and whose futures seem equally uncertain.

G3pdh. Despite various technical hurdles, a number of highly variable nuclear genes

have proven useful in reconstructing low-level phylogenies, including ITS, *Adh*, waxy,

MADS, phytochrome, PGI, and Vac (Small et al. 2004, Caicedo and Schaal 2004, Mort

and Crawford 2004). Sequences of the *G3pdh* gene have previously been used to study

phylogenetic relationships among recently diverged lineages of *Manihot* (Euphorbiaceae;

Olsen and Schaal 1999, Olsen 2002), the moss *Mitthyridium* (Calymperaceae; Wall

2002), the fungus *Cladonia* (Euascomycetes; Myllys et al. 2003), and intraspecific

variation in *Cryptomeria japonica* (Cupressaceae; Tani et al. 2003) and *Hordeum vulgare*

ssp. *spontaneum* (Poaceae; Morrell et al. 2003).

<u>Objectives</u>. The goals of this study are: 1) to quantify genetic variation within and among populations of *M. cumberlandensis*, and compare observed variation to that in a more widespread and abundant relative, *M. glabra*, and 2) to use the phylogenetic information inherent in DNA sequence polymorphism to gain insight into the historical processes contributing to current population structure of *M. cumberlandensis*.

Chapter II

Materials and Methods

Plant materials. Since *M. cumberlandensis* is a federally listed endangered species,

collection permits were obtained from the U. S. Fish and Wildlife Service, the Tennessee

Department of Environment and Conservation, and the National Park Service. Ten

populations of *M. cumberlandensis* were sampled from across the entire range of the

species (Fig. 2), with most populations containing hundreds to thousands of individuals.

[The population at Big Island was later divided into two populations (A and B)

corresponding to geographically distinct clusters of plants in two separate shelters].

Tissue samples consisting of a single leafy stem were taken from five arbitrarily chosen

plants at widely spaced intervals throughout each population and stored on ice until DNA

extraction. A single whole plant from each population was taken as a voucher specimen

to be deposited in the herbarium at the University of Tennessee (TENN). Four

populations of *M. glabra* were sampled in the same manner, all of which occur on dry

sandstone bluffs and outcrops in the region immediately surrounding the range of *M.*

cumberlandensis. Table 1 lists the sampled populations and their counties of occurrence.

Molecular methods. Whole genomic DNA was isolated from fresh leaves of all

individuals using a miniprep modification of the CTAB procedure of Permingeat et al.

(1998). Initial PCR amplification of the glyceraldehyde 3-phosphate dehydrogenase

(*G3pdh*) region was performed using universal primers designed by Strand et al. (1997;

GPDX7F and GPDX9R), which consistently amplified multiple paralogous *G3pdh* loci in

several species of *Minuartia*. These fragments were cloned and sequenced from several

species, and orthologous loci were identified through phylogenetic analysis. A primer

pair was designed to specifically amplify a 700 bp section of one ortholog: MIN*G3PDH*-

F, ACCCAAAAGACTGTTGATGGC, and MIN*G3PDH*-R,

GGACACRACATCATCCTCTGTGTAG. All PCR amplifications were performed in 25

µL reaction volumes, each containing the following components: 1 µL template DNA

(~10-50 ng), 1X *ExTaq* buffer (TaKaRa/PanVera), 200 µM each dNTP, 3.0 mM MgCl$_2$,

0.1 µM each primer, 0.625 units *ExTaq* (TaKaRa/PanVera), and bovine serum albumin at

a final concentration of 0.2 µg/µL. Cycling parameters were: 30 cycles of denaturation at

95°C for 1 minute, annealing at 63°C for 30 seconds, extension at 72°C for 1 minute;

then final extension at 72°C for 5 minutes. PCR products were cleaned with ExoSap-IT

(Amersham), sequenced directly with the ABI Prism BigDye Terminator Cycle

Sequencing Ready Reaction Kit, v. 2.0 or v. 3.1 (Applied Biosystems Inc.), and

electrophoresed and detected on an ABI Prism 3100 automated sequencer (University of

Tennessee Molecular Biology Resource Facility). Resulting sequences were edited using

Sequencher v. 4.1 (Gene Codes Corp.). Individuals homozygous at the nuclear *G3pdh*

locus were sequenced with a single primer, MIN*G3PDH*-F, though sequencing with this

primer required use of a two-step PCR sequencing cycle to eliminate sequence noise

resulting from promiscuous annealing of the primer at a secondary site within the

fragment. [1] PCR products from heterozygous individuals were sequenced from both ends

---

[1] This promiscuous annealing produced chromatograms with double peaks for the first 50 bp or so of sequence, which then abruptly ended, giving clean sequence for the remainder of the sequence. By noting the base sequence of the underlying "contaminant" fragment and comparing that sequence to all possible positions and complementary positions on the amplified product, it was determined that the primer was also binding at a site about 50 bp from the intended binding site, and sequencing the complementary strand.

(using both amplification primers), allowing reconstruction of sequences interrupted by indel polymorphism. Resolution of haplotypes in heterozygotes was accomplished by haplotype subtraction (Clark 1990), in which known haplotypes from homozygotes are "subtracted" from heterozygous sequences with multiple polymorphic sites, allowing inference of the allelic content of the second haplotype. Sequences of all resulting haplotypes will be deposited in GenBank.

<u>Data analysis</u>. Initial alignment of sequences was performed using ClustalX 1.81 (Thompson et al. 1997), with subsequent manual refinement in MacClade 4.0 (Maddison and Maddison 2000). Insertion/deletion (indel) events inferred during alignment were scored as additional binary characters and added to the data matrix for use in phylogenetetic analyses, though they were excluded from population genetic analyses.

For both *M. cumberlandensis* and *M. glabra*, analyses of genetic variation within and among populations were performed using Arlequin v. 2.000 (Schneider et al. 2000). Standard genetic parameters were calculated for each population, including estimates of haplotype diversity, heterozygosity, and nucleotide diversity. Analysis of Molecular Variance (AMOVA; Excoffier et al. 1992), as implemented in Arlequin v. 2.000, was used to test for significant genetic structuring among populations of each species relative to a null distribution of random assortment of haplotypes within and among populations. To gain further understanding of the relative influences of gene flow and genetic drift at different spatial scales across the range of *M. cumberlandensis*, a test for isolation-by-distance was performed in Arlequin to detect significant correlation of population

13

pairwise genetic distance (F_{ST}) and population pairwise geographic distance (see

Hutchison and Templeton 1999).

Phylogenetic analyses of all haplotypes using the optimality criterion of maximum

parsimony were performed with PAUP* 4.0b10 (Swofford 2002), and support for

obtained phylogenetic relationships among haplotypes was assessed with bootstrap

analysis (1000 replicates, full heuristic search). Relationships defined by indel

polymorphisms were found to be congruent with those defined by nucleotide

substitutions, justifying their inclusion in the matrix for parsimony analysis. Phylogenetic

analysis using the optimality criterion of maximum likelihood was also performed in

PAUP* using the HKY85+G model of nucleotide substitution as indicated by Modeltest

3.06 (Posada and Crandall 1998) and excluding indel characters. The program DnaSP

4.0 (Rozas et al. 2003) was used to detect possible recombination events within the data

set, since such events could potentially affect phylogenetic analyses. A single rogue

haplotype, J, was the obvious product of recombination between distant clades, and was

thus omitted from phylogenetic analysis, though it was included in genetic diversity

measures.

The standard phylogenetic analyses discussed above make certain assumptions about

relationships among taxa (haplotypes) that may not be warranted for allelic relationships

within species (Clement et al. 2000). First, standard trees represent taxa as terminal

points on branches, assuming that no ancestral taxa persist. In addition, standard trees

assume that taxa are related through bifurcating divergence. These assumptions are

14

generally safe for depicting relationships among species or higher taxa, but at the intraspecific level, it is quite likely that ancestral alleles will persist, and that alleles will show reticulating patterns of relationship due to recombination. Thus a more appropriate representation of haplotypic relationships within species is a network, consisting of nodes (haplotypes) connected to one another by lines (mutational steps), and which allows haplotypes to be placed as direct ancestors to other haplotypes, and allows haplotypes to not only diverge from one another, but also converge.

To detect ancestral (interior) haplotypes and reticulating relationships among haplotypes, statistical parsimony analysis was performed with the program TCS (Clement and Posada 2000), which was used to generate a haplotype network (cladogram) for each species.

Chapter III

Results

Diversity of *G3pdh*. The sequenced *G3pdh* region in *M. cumberlandensis* and *M. glabra*

ranged from 664 to 691 base pairs in length, including 247 bp of coding sequence from 3

exons (2 partial and 1 entire) (Fig. 4).

In the 55 diploid individuals (110 alleles) of *M. cumberlandensis* sequenced, 25 unique

haplotypes[2] (alleles) were observed, defined by 35 nucleotide substitutions and 13 indels

(ranging from 1 bp to 18 bp in length). Three substitutions occur in the coding region,

two synonymous and one resulting in an amino acid substitution (glutamic acid to lysine).

Thirteen of the 55 *M. cumberlandensis* individuals (23.6 percent) were heterozygous at

the *G3pdh* locus, having two distinct haplotypes present.

Among 20 individuals (40 alleles) of *M. glabra* sequenced, 18 haplotypes were observed,

defined by 23 nucleotide substitutions and 3 indels (1-6 bp in length). A single

synonymous substitution was observed in the coding region. Eighteen of the 40 *M.

glabr*a individuals (45 percent) were heterozygous at the *G3pdh* locus.

---

[2] In this paper, the term *haplotype* is used interchangeably with *allele*, and refers to a *DNA sequence variant*. Thus, if any two homologous sequences differ at one or more base positions, then by definition they constitute different haplotypes. Since the sequences used in this study were generated from a diploid nuclear locus (*G3pdh*), then each individual plant will possess two homologous sequences (alleles) that may or may not constitute different haplotypes.

Figure 5 shows a matrix of polymorphic sites observed in all haplotypes of *M. cumberlandensis* and *M. glabra*, and Table 2 summarizes patterns of nucleotide polymorphism in populations of both species. Figure 6 shows the number of haplotypes observed in each population for both species. Haplotype diversity and observed heterozygosity were compared for each population in Figure 7, and the nucleotide diversity (π) for each population is shown in Figure 8.

Population genetic structure. Analysis of molecular variance (AMOVA: Table 3) yielded a higher among population variance component ($F_{ST} = 0.63$) for *M. cumberlandensis* than for *M. glabra* ($F_{ST} = 0.22$). Thus, *M. cumberlandensis* shows high among-population variation and low within-population variation, in contrast to *M. glabra*, which shows high variation within populations and low variation among populations. In addition, a significant positive correlation was found between population pairwise F_{ST} and pairwise geographic distance in *M. cumberlandensis*, but with a large range of pairwise F_{ST}s at any given geographic distance (Fig. 10).

Genealogy of *G3pdh* haplotypes. Figure 11 shows the strict consensus of 3608 most parsimonious trees (of length 126) for all haplotypes of *M. cumberlandensis* and *M. glabra*, with level of bootstrap support for the resolved clades indicated at the nodes. The populations of occurrence and number of observed occurrences for each haplotype are noted in the tables above each species. Figure 12 shows the single most likely tree of relationships among all haplotypes, with branch lengths indicating number of substitutions per site. The likelihood tree clearly indicates that the *M. cumberlandensis*

17

and *M. glabra* haplotypes form distinct clusters separated by a significant mutational distance, and when rooted at the midpoint, or with an outgroup (*M. groenlandica*), the alleles of both species form monophyletic groups.

Statistical parsimony analysis was performed with the program TCS in order to reconstruct haplotype networks for both *M. cumberlandensis* and *M. glabra* (Fig. 13). The arrow on the cladogram for *M. cumberlandensis* (Fig. 13a) indicates the location of the most parsimonious and most likely root, as determined in PAUP* using *M. glabra* as the outgroup. The *M. glabra* cladogram is unrooted because of weak support for the root inferred in PAUP*. Reticulations (loops) in the *M. cumberlandensis* network indicate potential recombination events or homoplasious character state changes, which is not surprising given that DnaSP 4.0 detected 3 possible recombination events within those haplotypes. The position of rogue haplotype 'J' was roughly inferred by connecting it to the most similar haplotypes to each of its two recombined segments.

Chapter IV

Discussion

Haplotype diversity and heterozygosity. *Minuartia cumberlandensis* and *M. glabra*

maintain a fairly high amount of genetic diversity at the *G3pdh* locus as measured by

both total number of unique haplotypes and haplotype diversity, a measure that

incorporates the relative frequencies of haplotypes (Table 2). Haplotype diversity, also

known as gene diversity, is defined as the probability that two randomly chosen

haplotypes from a sample will be different, and is equivalent to the expected

heterozygosity in a population at equilibrium (Nei 1987).

Observed heterozygosity is higher in *M. glabra* than in *M. cumberlandensis* (0.45 versus

0.24), and global heterozygosity (over all populations) in both species measures well

below levels expected under equilibrium conditions, which is often the case in species

with a significant amount of population substructuring (Hartl and Clark 1997).

Individual populations of both species also show heterozygosity lower than predictions

based on haplotype diversity, with some populations of *M. cumberlandensis* having little

or no heterozygosity despite having relatively high haplotype diversities (e. g. Big Island

A, Slave Falls, Jamestown Reservoir) (Fig. 7). The reduced heterozygosity observed in

this case is consistent with the effects of frequent mating among closely related plants, or

inbreeding, which characteristically occurs more frequently in smaller populations. A

related factor likely to affect breeding patterns (and thus heterozygosity) in *M.*

cumberlandensis is the potential for restricted movement of pollen and seeds, even within

a single rockshelter population. Limited gene flow within populations (and reduced

19

heterozygosity) could also result from a high occurrence of self-fertilization, but it hasn't

yet been established that *M. cumberlandensis* is even self-compatible, though there is no

apparent physical barrier preventing pollen from contacting mature stigmas on the same

plant. Self-compatibility is thought to occur often in species that colonize new areas of

isolated habitat through rare dispersal events, which makes the ability to self-fertilize a

necessary solution to the lack of sufficient potential mates and/or pollinators (Fishman

and Wyatt 1999; Baker 1955). Certainly, further investigation into the breeding system

of *M. cumberlandensis*, in addition to finer-scale sampling of genetic variation within

populations, could add greatly to our understanding of mechanisms of gene flow within

the species.

Though *M. cumberlandensis* and *M. glabra* have equivalent levels of global *G3pdh*

haplotype diversity, individual populations of *M. cumberlandensis* contain fewer

haplotypes on average (3 versus 4.75) and have lower haplotype diversities on average

(0.52 versus 0.76) than populations of *M. glabra*. This is consistent with expectations of

higher rates of genetic drift due to smaller population sizes in *M. cumberlandensis*.

Populations of *M. cumberlandensis* vary considerably in *G3pdh* haplotype diversity, with

some populations being fixed for a single haplotype, and others having as many as six

distinct haplotypes present (Fig. 6). There is no apparent geographical pattern in the

distribution of populations with high or low allelic diversity, though the population with

the highest number of haplotypes and highest haplotype diversity (Ladder Trail) is a large

population located within the dense cluster of populations in proximity to Pickett State

Park (Fig. 2). However, other populations within this cluster (Hazard Cave, Pickett Dam) have much lower haplotype diversities than the Ladder Trail population, indicating that even within the most densely occupied area of the species' range, rockshelter populations can be sufficiently isolated from one another to allow for significant divergence. Two populations are fixed for single haplotypes (Middle Creek and Puncheoncamp Fork). The Middle Creek population, though in close proximity to other populations, consists of only a few dozen individuals in the corner of a large shelter, which suggests that genetic diversity may be low because of bottlenecking due to recent colonization and/or rapid fixation due to low population size. The Puncheoncamp Fork population contains hundreds of individuals, but is significantly isolated from any other known population, suggesting its low diversity results from insufficient immigration to counter the long-term effects of genetic drift.

Nucleotide diversity. As markers for studying population genetic processes, DNA sequences can be extremely valuable, allowing the reconstruction of evolutionary relationships among variant alleles, and thus providing a phylogenetic context for observed patterns of allelic distribution. Various measures of population genetic diversity incorporate information about relationships (similarities) among sequence haplotypes (alleles) and add a historical dimension to observed patterns of population diversity. One such diversity measure that incorporates phylogenetic distance among alleles is nucleotide diversity (π).

Nucleotide diversity (π) is the mean number of pairwise differences among all sequences in a sample (per nucleotide site), and is a rough measure of the amount of mutational divergence within a group of sequences. A low value of π indicates that all sequences in the group tend to be very similar to one another (little divergence), while a high value indicates that sequences in the group tend to differ by many mutations (high divergence).

Perhaps surprisingly, global nucleotide diversity is higher in the rare endemic *M. cumberlandensis* than in *M. glabra* (0.013 versus 0.007) (Table 2). This indicates that *M. cumberlandensis* as a species maintains a much broader spectrum of allelic variation than does *M. glabra* (at least within the sampled populations). So, even though the two species maintain an equivalent amount of haplotype diversity, in *M. cumberlandensis* haplotypes tend to be distantly related, while the sampled haplotypes in *M. glabra* tend to be more closely related.

Individual populations of *M. cumberlandensis* have a wide range of nucleotide diversities (from zero to 0.013), with an average value of 0.005, while populations of *M. glabra* have a relatively narrow range of nucleotide diversities (from 0.004 to 0.008), with an average value of 0.006 (Fig. 8). Populations that are fixed for a single haplotype (Puncheoncamp Fork, Middle Creek) have nucleotide diversity of zero by definition, while populations with greater haplotype diversities have higher values of π, though populations with similar haplotype diversities may have very different levels of nucleotide diversity (depending on relationships among their constituent haplotypes). For example, the sampled populations at Laurel Fork, Jamestown Reservoir, and Peters

Bridge all have roughly equivalent haplotype diversities (~0.7), but have nucleotide diversities of 0.008, 0.003, and 0.001 respectively. The decreasing π values seen in these populations represent increasing levels of phylogenetic similarity among haplotypes in each population, with the Peters Bridge population having three very closely related alleles. The population with the highest haplotype diversity (Ladder Trail) also has relatively high nucleotide diversity, suggesting broad phylogenetic variation among the six alleles possessed by the sampled individuals. The highest value of π is found in the Slave Falls population, which contains a rogue haplotype (haplotype J) that is the apparent result of recombination between two distantly related haplotypes. This recombination event has the result of increasing the number of pairwise differences between haplotype J and any other sequence, thus raising overall nucleotide diversity in the population (and, to a lesser degree, the species as a whole).

Population genetic structure, gene flow, and genetic drift. It's interesting to note that even though *M. cumberlandensis* has high global nucleotide diversity relative to *M. glabra*, the nucleotide diversity within populations of *M. cumberlandensis* is lower on average than nucleotide diversity within populations of *M. glabra* (Table 2). What this pattern suggests is that the average population of *M. cumberlandensis* contains only a small, non-representative subset of the overall phylogenetic variation in the species, whereas the average population of *M. glabra* contains practically the entire breadth of variation possessed by the aggregated populations sampled.

The analysis of molecular variance (AMOVA: Table 3) confirms the above patterns of genetic structuring, revealing significant structure in both species, but with 63 percent of the total variation in *M. cumberlandensis* being distributed among populations rather than contained within them, and only 21 percent of the total variation in *M. glabra* distributed among populations.

The partitioning of variation seen in *M. cumberlandensis* is consistent with long-term limited gene flow among isolated populations and/or recent establishment of populations from a heterogeneous source population. The relative genetic homogeneity across populations of *M. glabra* is consistent with patterns expected for either high rates of gene flow among populations or, alternatively, a relatively recent establishment of all populations from some homogenous source population. Further evidence, discussed later, may help to elucidate which scenario is more likely.

Relationships among populations. Studies of genetic structuring can fruitfully incorporate information on spatial relationships among populations to gain further insight into population genetic processes. An approach used by Hutchison and Templeton (1999) to elucidate the relative roles of gene flow and genetic drift in contributing to population genetic structure involves testing for correlation of geographic distance between populations with genetic distance between those populations. Assuming a stepping-stone model of regional population structure, in which adjacent populations are most likely to interact through gene flow, they propose that a group of populations at regional equilibrium between loss and gain of alleles (through drift and gene flow) will

show a predictable pattern of increasing genetic distance between populations with increasing physical distance. In fact, a scatterplot of all pairwise comparisons of population genetic distance and geographic distance should show a positive and monotonic relationship across all geographic distances, with the scatter of the pairwise points tending to increase as spatial separation of populations increases (Fig. 9a). This equilibrium pattern is referred to as "isolation-by-distance," and results from the changing relative roles of gene flow and drift as population separation increases. At smaller geographic distances, populations will tend be more genetically similar, and the homogenizing effects of gene flow will reduce the variability of genetic distances between pairs of populations. However, populations separated by greater geographic distances will have reduced influence of gene flow between them, increasing genetic distance between them, and proportionally increasing the influence of genetic drift within populations, leading to higher variability of pairwise genetic distances due to random fluctuations in population genetic makeup.

Theoretically, the equilibrium pattern of isolation-by-distance discussed above will form in any group of populations maintaining stable dispersal conditions for a sufficient amount of time (Hutchison and Templeton 1999). In many cases, particularly in species that have undergone recent demographic shifts, equilibrium conditions will not have been reached, and relationships between genetic and geographic distances between populations will differ significantly from the example in Figure 9a. By evaluating scatterplots of genetic distance (F_{ST}) versus geographic distance between all pairs of populations in a

25

region, the relative roles of gene flow and genetic drift may be extrapolated based on degree of correlation and amount of scatter at various spatial scales.

First, consider a hypothetical group of populations that have recently arisen from some homogenous source population, either through migration or fragmentation of a large continuous population. In this scenario, genetic distances between populations will be low regardless of the physical distance separating them, and there will be little variation in genetic distances among pairwise comparisons (Fig. 9b). If gene flow among populations remains high over time relative to drift, then this pattern (panmixia) will persist. Alternatively, if gene flow among populations remains relatively insignificant, then genetic drift within populations will be the dominant force, leading many populations to diverge genetically and others to remain more similar merely through random genetic fluctuation. With this scenario of extreme fragmentation and isolation, as in the case of panmixia, the genetic distance between a pair of populations will have little or no relationship to the geographic distance between them, but in contrast to panmixia, the amount of variance (scatter) in pairwise genetic distances will be high at any given geographic distance (Fig. 9c).

A final non-equilibrium alternative discussed by Hutchison and Templeton (1999) involves incipient localized gene flow among nearby populations, with genetic drift still dominating population relationships at greater geographical scales (Fig. 9d). Over time, given stable dispersal conditions, this pattern will come to resemble the equilibrium pattern in Figure 9a, but if conditions change so that populations become smaller or more

isolated, then the pattern will shift to look more like the drift-dominated scenario of Figure 9c. Likewise, if conditions change so that dispersal among populations increases or population sizes increase, then the pattern in Figure 9d will begin to look more like the gene flow-dominated scenario of Figure 9b.

The very small range of *M. cumberlandensis*, and the accompanying stochastic effects related to the small geographic scale of its "regional" population distribution make it unlikely that a clear pattern of isolation-by-distance can be resolved, though the scatterplot of all pairwise population comparisons of genetic distance (F_{ST}) and geographic distance (simple linear distance in kilometers) is still informative when viewed in the context of the model scenarios discussed above (Fig. 10). Although there is a significant positive correlation overall between population genetic distance and geographic distance in populations of *M. cumberlandensis* (r=0.165, p=0.015), the high degree of scatter in F_{ST} at most geographic distances supports the conclusion that the species is far from being at regional equilibrium. A reasonable interpretation of this pattern is that populations of *M. cumberlandensis* are significantly isolated from one another at all geographic scales, though populations within about 4 kilometers tend to be more similar to one another. At distances greater than 4 kilometers, there is no discernible relationship between physical separation of populations and the amount of genetic divergence between them, although populations at extreme ends of the species' range (thus greater than 25 km apart) are consistently genetically dissimilar.

These results support the conclusion that populations of *M. cumberlandensis* in general are, at present, essentially independent of one another genetically, and have been for a significant period of time, with the possible exception of densely clustered populations with little geographic separation (as in the Pickett County populations). Nonetheless, it's important to note that certain widely separated populations show a striking degree of genetic similarity. For example, the Ladder Trail and Laurel Fork populations are 11 kilometers apart, but have an extremely low pairwise F_{ST} of only 0.02. It is implausible that high levels of contemporary gene flow between these two populations account for their similarity, given the large distance between them, the lack of apparent intermediate populations, and the observed physical isolation of the Laurel Fork population. There are two plausible explanations for their high degree of similarity. First, assuming that at some point in the distant past the two populations shared a common gene pool that was subsequently fragmented, it is conceivable that long-term maintenance of large, stable populations at each site could have reduced the effects of random genetic drift, thus maintaining the ancestral similarity between the two populations. Alternatively, it is possible that a relatively recent dispersal event from some source population (e. g. Ladder Trail) resulted in the establishment of a distant population (e. g. Laurel Fork) with nearly identical genetic makeup, and drift within the two populations has not had sufficient time to cause significant divergence between them.

Perhaps easier to explain are the occurrence of highly genetically divergent populations with little geographic separation between them. The best examples of this involve populations that have very low variation or are fixed for a single haplotype, and as a

result yield very high F_{ST} values when compared with other populations. The Middle Creek population for example, despite the fact that it has several nearby populations with which it might interact genetically, probably remains genetically depauperate and dissimilar from neighboring populations in spite of occasional genetic exchange due to the overpowering effect of genetic drift due to small population size.

Genealogy of *M. cumberlandensis* and *M. glabra* haplotypes. To gain further insight into the spatial distribution of genetic variation in *M. cumberlandensis*, and the ongoing and historical ecological and genetic processes that have created extant genetic patterns, it is helpful to reconstruct genealogical relationships among *G3pdh* variants. By mapping the geographical distribution of genealogical lineages (haplotypes and clades), we may uncover patterns that reveal the impact of historical processes on overall structuring of genetic variation in the species, which we can then attempt to separate from patterns due to ongoing processes. There are many novel approaches to using phylogeographic variation for inference of historical demographic factors such as range expansion, migration, habitat fragmentation, and population size fluctuation (Templeton 1998, 2004; Templeton et al. 1995; Knowles and Maddison 2002), but this paper takes a more traditional approach, rather than applying rigorous statistical analyses. In fact, given the unique distribution and population structuring in *M. cumberlandensis*, and the small number of populations and individuals sampled for this study, it is not certain that contemporary statistical phylogeography would provide any greater confidence in our conclusions, or any novel insights.

Reconstruction of phylogenetic relationships among *G3pdh* haplotypes using standard

parsimony and likelihood approaches gives us a good approximation of the scale of

evolutionary divergence within *M. cumberlandensis* and *M. glabra*, as well as revealing

relationships between the two species. The cladogram in Figure 11 represents the

consensus of 3608 most parsimonious trees, with the bootstrap values at each node

representing the level of confidence that all haplotypes above that node form a

monophyletic group, or clade. The tree is rooted at the midpoint, which essentially

places the root (ancestral node) at the longest "branch" of the tree, in this case the branch

separating *M. cumberlandensis* haplotypes from *M. glabra* haplotypes.

Note that (with midpoint rooting) the haplotypes of each species form strictly

monophyletic groups (clades), meaning that all haplotypes within each species are more

closely related to one another than they are to any haplotype in the other species. In

terms of evolutionary relationships, we can infer that the haplotypes within each species

are descended from a single common ancestor *within that species*. While it may seem

self-evident that the alleles in each species should form monophyletic groups, it is by no

means uncommon for different species to show paraphyly or polyphyly with respect to

allelic relationships, especially in species that have diverged recently, and especially for

diploid nuclear genes (with large effective population sizes) (Hare 2001). In many cases,

hybridization events will result in alleles from one species appearing to be embedded in

the clade of another species' alleles. Based on this data, there is no evidence that

hybridization occurs between *M. cumberlandensis* and *M. glabra*, and based on the

observed reciprocal monophyly, it is safe to say that the two species diverged in the distant past.

All of the above phylogenetic patterns are also evident in the maximum likelihood tree (Fig. 12), which represents the single most likely inference of relationships among the haplotypes of both species (based on an assumed model of evolutionary change). The tree is drawn as a phylogram to make clear the amount of actual genetic change that has occurred within lineages, with branch lengths corresponding to the amount of mutation separating the various clades and haplotypes. Comparison of phylogenetic patterns in the two species can inform our interpretation of the varying measures of genetic diversity discussed earlier.

Recall that global nucleotide diversity (π) was significantly greater in *M. cumberlandensis* than in the sampled populations of *M. glabra*, despite their equivalent measures of haplotype diversity, which indicated that *M. cumberlandensis* maintains a much broader phylogenetic diversity among its haplotypes than does *M. glabra*. Looking at the maximum likelihood tree, we can see why this is the case: the alleles of *M. cumberlandensis* have a "deeper" divergence on the tree than do the sampled alleles of *M. glabra*, which makes average sequence divergence among *M. cumberlandensis* haplotypes relatively large. This suggests that either the mutation rate is higher within *G3pdh* lineages in *M. cumberlandensis*, or that the most recent common ancestor of the sampled alleles in *M. cumberlandensis* existed further back in time than the most recent common ancestor of the sampled *M. glabra* alleles. Given the relatively low level of

divergence within the two species, we should be cautious in drawing conclusions, but the noticeable grouping of *M. cumberlandensis G3pdh* haplotypes into two distinct clades supported by multiple mutations in the coding region, it's likely that the deep divergence observed in *M. cumberlandensis* represents maintenance of older polymorphism. But why would this be the case?

It's possible that historical fragmentation of *M. cumberlandensis* populations has resulted in the maintenance of isolated lineages whose most recent common ancestor (coalescent) dates back to an ancient common ancestral population, while the shallow divergence (recent coalescence) observed in *M. glabra* is a result of its lack of such fragmentation, and is purely a product of its effective population size. If such extreme fragmentation had occurred, and had isolated the diverse lineages observed in *M. cumberlandensis* for an extended period of time, then certainly those long-isolated lineages are at present fully integrated, occurring frequently in the same populations, and even in the same individuals. It would indeed be ironic if the very fragmentation thought to endanger the existence of *M. cumberlandensis* were in fact responsible for its unusual breadth of genetic diversity, though it's far from certain that this is the case. It's also entirely possible that the deeper coalescence observed in *M. cumberlandensis* could arise if the species had maintained a large effective metapopulation size over a long period of time, while *M. glabra* had undergone a significant bottleneck event that reduced genetic diversity at some point in the "recent" past. Based on the current habitats of both species, and what we know of vegetation changes during the course of the Pleistocene (Delcourt 1979), it's not unreasonable to speculate that *M. cumberlandensis* has persisted in its

current range throughout periods of climate fluctuation, while *M. glabra* is likely to have

recently "invaded" the region of the northern Cumberland Plateau as a single

homogenous migration front, which also may explain, in large part, the genetic

homogeneity observed among populations of *M. glabra* from throughout this region.

Haplotype networks for both *M. cumberlandensis* and *M. glabra* (Fig. 13) largely reflect

the relationships revealed by the parsimony and likelihood trees, with the additional

benefit of showing ancestral alleles as embedded within lineages, rather than terminating

them. Also the *M. cumberlandensis* network has "loops," representing either homoplasy

or the effects of recombination among alleles. Note that the network for *M.*

cumberlandensis is rooted (with *M. glabra* as the outgroup, indicated by the arrow),

while the network for *M. glabra* is not rooted due to lack of strong support for any single

root position. Also note that haplotype J, a recombinant allele that was left out of the

earlier phylogenetic trees, was connected to the two haplotypes most closely related to its

putative "parent" alleles. The black dots on the network represent haplotypes whose

existence is inferred, but were not sampled, and the lines connecting dots represent

mutational steps between haplotypes or inferred haplotypes. Thus, for example, in the *M.*

cumberlandensis network (Fig. 13a), haplotype D is the direct ancestor of haplotype H,

and the two differ by a single mutation. In the case of haplotypes Q, S, and R, the loop

indicates that the ancestry of haplotype R is ambiguous, either because haplotypes Q and

S recombined at some point, or because the direct (inferred) ancestor of haplotype R

converged through homoplasy to be equally distant (one mutation) from both haplotype

Q and haplotype S. A casual comparison of the two networks reaffirms the observation

33

that *M. cumberlandensis* contains broader mutational diversity and displays greater

phylogenetic structure at the *G3pdh* locus, and that *M. glabra*'s alleles tend to be more

closely related to one another.

Geographic distribution of haplotypes and clades. To begin to understand the importance

of phylogeographic patterns in elucidating factors that structure genetic variation, it is

helpful to both visualize the phylogenetic composition of individual populations, and to

map the distributions of significant haplotypes and clades throughout the range of the

species (using data from the tables on Fig. 11). The diagrams in Figure 14 (a-k) depict

the allelic composition of each population of *M. cumberlandensis*, with the observed

alleles for that population highlighted in green on the overall network for the species.

The number of occurrences of each observed haplotype in that population is noted in

green. Thus the Ladder Trail sample consists of two copies of haplotype A, four copies

of haplotype B, one copy of C, and so on. You may recall that Ladder Trail had one of

the highest measures of nucleotide diversity (π) of all the populations, and it's clear from

this diagram (Fig. 14a) why this is the case: the Ladder Trail population alone carries

nearly the entire breadth of variation that the species carries as a whole. The low value of

pairwise F_{ST} (genetic distance) for Ladder Trail and Laurel Fork (Fig. 14e) is undoubtedly

due to their possession of roughly equivalent frequencies of haplotypes B and C, and the

close relationship of haplotype D in Ladder Trail to haplotype H in Laurel Fork. Other

populations, such as Jamestown Reservoir (Fig. 14g) and Peters Bridge (Fig. 14h) draw

their allelic content from only a small portion of the overall network, which explains their

lower nucleotide diversities, and strongly suggests some historically derived pattern of

genetic partitioning. An extreme example of genetic partitioning is the Puncheoncamp

Fork population (Fig. 14k), which is fixed for haplotype U, resulting in high genetic

distance between it and other populations, though Puncheoncamp Fork apparently has

some genetic relationship to Big Island B (Fig. 14j), which possesses the closely related

haplotype T.

At first glance, the diagrams of population allelic composition for *M. glabra* (Fig. 15)

seem to show that the alleles in each population are randomly scattered across the

haplotype network for the species, rather than being partitioned based on phylogenetic

relationships. Upon closer inspection, however, it's clear that while there is abundant

phylogenetic overlap among the populations, there is little sharing of specific alleles

among them, with the exception of haplotype h, which occurs in both of the Fentress

County populations (Jamestown Reservoir and Darrow Ridge). This pattern indicates

some degree of restricted gene flow among these populations, but it remains to be

determined whether population similarities are primarily due to significant long-term

regional gene flow, or to retention of ancestral polymorphism after a relatively recent

establishment of populations, or both. Further sampling of populations of *M. glabra*

would be a first step in addressing this question, including sampling the entire range of

the species in order to look for the telltale signs of Pleistocene migration and range

expansion.

M. cumberlandensis, in contrast to *M. glabra*, has a significant number of shared

haplotypes among its populations. Mapping the geographical distributions of these

shared haplotypes allows us to visualize the combined effects of contemporary gene flow

and historical population structuring. If we map the geographic ranges of not just

individual haplotypes, but of groups of closely related haplotypes (clades *sensu*

Templeton), then we can visualize patterns that represent the effects of gene flow and

historical contingency over longer spans of time. For example, Figure 16 shows the

distribution of haplotypes W, X, and Y, which are all closely related to one another and

form a monophyletic group that is relatively derived (distant from the root). All three

haplotypes were observed only at the Peters Bridge population and nowhere else,

although they appear to be descended from haplotype A, which is common elsewhere in

the species' range. The fact that this lineage is restricted to a single population at the

extreme end of the range of *M. cumberlandensis*, and has had time to accumulate a

number of mutations, indicates relatively long-term isolation of this population from the

remainder of the range, with no emigration from its gene pool. Another monophyletic

group made up of haplotypes Q, R, and S, is restricted to the two Big Island populations

(Fig. 17), and suggests a similar degree of isolation of these populations from the

remainder of the range. A clade consisting of haplotypes U and T occurs only at

Puncheoncamp Fork and Big Island B, and may represent evidence of limited gene flow

between the two populations (Fig. 18). An interesting example of a small clade with

widespread distribution consists of haplotypes G and N, and occurs at Laurel Fork,

Hazard Cave, and Big Island A (Fig. 19). One might assume that this distribution

provides evidence of long distance gene flow among the three widely spaced populations,

but in fact the haplotypes are relatively old (based on their proximity to the root), and are

therefore likely to occur across the range of *M. cumberlandensis* because of shared ancestral polymorphism among populations.

A final noteworthy pattern of clade distributions involves the two major subclades at the highest level of structure in the haplotype network (Fig. 20). The clade highlighted in purple occurs across the entire range of *M. cumberlandensis*, while the large group of haplotypes highlighted in yellow is restricted to the Pickett County populations and Laurel Fork. What caused this distributional pattern is not immediately evident, but it makes very clear that the majority of the broad genetic variation in the species resides in the central cluster of populations highlighted in yellow. This makes sense given that the highest densities of populations occur in this area, thus making it likely that broad ancestral polymorphism would be maintained because of the larger effective population sizes and more frequent gene flow resulting from the dense clustering of populations.

Conclusions and implications for conservation. The above discussion can be distilled into a few basic points that are directly relevant to the conservation of *M. cumberlandensis*:

- Haplotype diversity within most populations of *M. cumberlandensis* is high, but observed heterozygosity is low, suggesting the effects of inbreeding and the risk of inbreeding depression. However, inbreeding may not be a significant threat to genetic integrity in the species, particularly if purging of deleterious recessive alleles has already taken place.

- *Minuartia cumberlandensis* shows significant structuring of overall genetic variation, with populations tending to carry only a small non-representative subset of the overall variation found in the species. This is due to extremely reduced gene flow among populations, particularly outliers, and the influence of genetic drift within small populations.

- The relatively high nucleotide diversity observed in *M. cumberlandensis* is due to deep coalescence and persistence of broad ancestral polymorphism, particularly in largest contiguous cluster of populations in Pickett County.

- Outlying populations contain a significant amount of the species' overall variation, but are currently genetically isolated from the main cluster, with shared haplotypes likely representing maintenance of persistent ancestral alleles in larger populations.

- It is likely that *M. cumberlandensis* has persisted in its current location through cyclical changes in climate during the Pleistocene, and if protected from habitat destruction and population decimation, it may fare well in the face of future climate change.

Literature Cited

Avise, J. C. 2000. Phylogeography: The History and Formation of Species. Harvard University Press. Cambridge.

Baker, H. G. 1955. Self-compatibility and establishment after "long-distance" dispersal. *Evolution* 9: 347-349.

Barrett, S. C. H., and Kohn, J. R. 1991. Genetic and evolutionary consequences of small population size. *In* Genetics and Conservation of Rare Plants, eds. D. A. Falk and K. E. Holsinger, pp. 3-30. Oxford University Press, New York. 283 pp.

Baskin, J. M., and Baskin, C. C. 1982. Germination ecophysiology of *Arenaria glabra*, a winter annual of sandstone and granite outcrops of the southeastern United States. *American Journal of Botany* 69(6): 973-978.

Clark, A. G. 1990. Inference of haplotypes from PCR-amplified samples of diploid populations. *Molecular Biology and Evolution* 7(2): 111-122.

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

Delcourt, H. R. 1979. Late quaternary vegetation history of the Eastern Highland Rim and adjacent Cumberland Plateau of Tennessee. *Ecological Monographs* 49(3): 255-280.

Demesure, B., Comps, B., and Petit, R. J. 1996. Chloroplast phylogeography of the common beech (Fagus sylvatica L.) in Europe. *Evolution* 50(6): 2515-2520.

Donahue, J., and Adovasio, J. M. 1990. Evolution of sandstone rockshelters in Eastern North America; A geoarchaeological perspective. Pages 231-251 in Lasca, N. P., and Donahue, J., eds., Archaeological Geology of North America. Boulder, Colorado, Geological Society of America, Centennial Special Volume 4.

Excoffier, L., Smouse, P. E., and Quattro, J. M. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial-DNA restriction data. *Genetics* 131: 479-491.

Ferguson, Terry A., Pace, R. A., Gardner, J. W., and Hoffman, R. W. 1986. Final Report of the Big South Fork Archaeological Project: Survey, Testing and Recommendation. Submitted to the U.S. Army Corps of Engineers, Nashville District.

Fernald, M. L. 1919. V. The specific identity of *Arenaria groenlandica* and *A. glabra*. *Rhodora* 21: 17-20.

Fishman, L., and Wyatt, R. 1999. Pollinator-mediated competition, reproductive character displacement, and the evolution of selfing in *Arenaria uniflora* (Caryophyllaceae). *Evolution* 53(6): 1723-1733.

García-Gil, M. R., Mikkonen, M., and Savolainen, O. 2003. Nucleotide diversity at two phytochrome loci along a latitudinal cline in *Pinus sylvestris*. *Molecular Ecology* 12: 1195-1206.

Hare, Matthew P. 2001. Prospects for nuclear gene phylogeography. *Trends in Ecology and Evolution* 16(12): 700-706.

Hutchison, D. W., and Templeton, A. R. 1999. Correlation of pairwise genetic and geographic distance measures: inferring the relative influences of gene flow and drift on the distribution of genetic variability. *Evolution* 53(6): 1898-1914.

Järvinen, P., Lemmetyinen, J., Savolainen, O., and Sopanen, T. 2003. DNA sequence variation in BpMADS2 gene in two populations of *Betula pendula*. *Molecular Ecology* 12: 369-384.

Knowles, L. L., and Maddison, W. P. 2002. Statistical phylogeography. *Molecular Ecology* 11: 2623-2635.

Levesque, C. M., and Burger, J. F. 1982. Insects (Diptera, Hymenoptera) associated with *Minuartia groenlandica* (Caryophyllaceae) on Mount Washington, New Hampshire, U. S. A., and their possible role as pollinators. *Arctic and Alpine Research* 14(2): 117-124.

McCormick, J. F., Bozeman, J. R., and Spongberg, S. 1971. A taxonomic revision of granite outcrop species of *Minuartia* (*Arenaria*). *Brittonia* 23: 149-160.

McNeill, J. 1981. The delimitation of *Arenaria* (Caryophyllacea) and related genera in North America, with 11 new combinations in *Minuartia*. *Rhodora* 82: 495-502.

Maddison, D. R. and W. P. Maddison, 2000. MacClade 4: Analysis of phylogeny and character evolution. Version 4.0. Sinauer Associates, Sunderland, Massachusetts.

Maguire, B. 1951. Studies in the Caryophyllaceae – V. *Arenaria* in America north of Mexico: A conspectus. *American Midland Naturalist* 46(2): 493-511.

Morrell, P. L., Lundy, K. E., and Clegg, M. T. 2003. Distinct geographic patterns of genetic diversity are maintained in wild barley (*Hordeum vulgare* ssp. *spontaneum*) despite migration. *Proceedings of the National Academy of Science* 100(19): 10812-10817)

Mort, M. E., and Crawford, D. J. 2004. The continuing search: low-copy nuclear sequences for lower-level plant molecular phylogenetic studies. *Taxon* 53(2): 257-261.

Myllys, L., Stenroos, S, Thell, A., and Ahti, T. 2003. Phylogeny of bipolar *Cladonia arbuscula* and *Cladonia mitis* (Lecanorales, Euascomycetes). *Molecular Phylogenetics and Evolution* 27: 58-69.

Nei, M. 1987. Molecular Evolutionary Genetics. Columbia University Press, New York.

Oh, S., and Potter, D. 2003. Phylogenetic utility of the second intron of LEAFY in *Neillia* and *Stephanandra* (Rosaceae) and implications for the origin of *Stephanandra*. *Molecular Phylogenetics and Evolution* 29: 203-215.

Olsen, K. M. 2002. Population history of *Manihot esculenta* (Euphorbiaceae) inferred from nuclear DNA sequences. *Molecular Ecology* 11: 901-911.

Olsen, K. M., and Purugganan, M. D. 2002. Molecular evidence on the origin and evolution of glutinous rice. *Genetics* 162: 941-950.

Olsen, K. M., and Schaal, B. A. 1999. Evidence on the origin of cassava: phylogeography of *Manihot esculenta*. *Proceedings of the National Academy of Sciences* 96: 5586-5591.

Permingeat, H. R., Romagnoli, M. V, and Vallejos, R. H. 1998. A simple method for isolating high yield and quality DNA from cotton (Gossypium hirsutum L.) leaves. *Plant Molecular Biology Reporter* 16: 1-6.

Posada, D., and Crandall, K. A. 1998. Modeltest: testing the model of DNA substitution. *Bioinformatics* 14(9): 817-818.

Radford, A. E., Ahles, H. E., and Bell, C. R. 1968. Manual of the Vascular Flora of the Carolinas. University of North Carolina Press, Chapel Hill.

Schaal, B. A., Hayworth, D. A., Olsen, K. M., Rauscher, J. T., and Smith, W. A. 1998 Phylogeographic studies in plants: problems and prospects. *Molecular* Ecology 7: 465-474.

Schneider, S., Roessli, D., and Excoffier, L. 2000. Arlequin Ver. 2.000: A software for population genetics data analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland.

Soltis, D. E., Gitzendanner, M. A., Strenge, D. D., and Soltis, P. S. 1997. Chloroplast DNA intraspecific phylogeography of plants from the Pacific Northwest of North America. *Plant Systematics and Evolution* 206: 353-373.

Strand, A. E., Leebens-Mack, J., and Milligan, B. G. 1997. Nuclear DNA-based markers for plant evolutionary biology. *Molecular Ecology* 6: 113-118.

Swofford, D. L. 2002. PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

Tani, N., Tsumura, Y., and Sato, H. 2003. Nuclear gene sequences and DNA variation of *Cryptomeria japonica* samples from the postglacial period. *Molecular Ecology* 12: 859-868.

Templeton, A. R. 2004. Statistical phylogeography: methods of evaluating and minimizing inference errors. Molecular Ecology 13: 789-809.

Templeton, A. R. 1998. Nested clade analyses of phylogeographic data: testing hypotheses about gene flow and population history. *Molecular Ecology* 7: 381-397.

Templeton, A. R., Routman, E., and Phillips, C. A. 1995. Separating population structure from population history: a cladistic analysis of the geographical distribution of mitochondrial DNA haplotypes in the tiger salamander, *Ambystoma tigrinum*. *Genetics* 140: 767-782.

Thomson, J. D., Gibson, T. J., Plewniak, F. P., Jeamougin, F., and Higgins, D. G. 1997. The Clustal X Windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* 24: 4876-4882.

U. S. Fish and Wildlife Service. 1996. Cumberland Sandwort Recovery Plan. Atlanta, GA. 28pp.

U. S. Fish and Wildlife Service. 1988. Endangered and Threatened Wildlife and Plants; Determination of Endangered Status for *Arenaria cumberlandensis*. *Federal Register* 53(121): 23745-23748.

Walck, J. L., Baskin, J. M., Baskin C. C., and Francis, S. W. 1996. Sandstone rockhouses of the eastern United States, with particular reference to the ecology and evolution of the endemic plant taxa. *The Botanical Review* 62(4): 311-362.

Wall, D. P. 2002. Use of the nuclear gene *glyceraldehyde 3-phosphate dehydrogenase* for phylogeny reconstruction of recently diverged lineages in *Mitthyridium* (Musci: Calymperaceae). *Molecular Phylogenetics and Evolution* 25: 10-26.

Weaver, R. E. 1970. The *Arenarias* of the southeastern granitic flatrocks. *Bulletin of the Torrey Botanical Club* 97(1): 40-52.

Wilson, C. W., and Stearns, R. G. 1958. Structure of the Cumberland Plateau, Tennessee. *Bulletin of the Geological Society of America* 69: 1283-1296.

Wofford, B. E., and Kral, R. 1979. A new *Arenaria* from the Cumberlands of Tennessee. *Brittonia* 31(2): 257-260.

Wofford, B. E., and Smith, D. K. 1980. *Arenaria cumberlandensis* Status Survey Report. Report for the U. S. Fish and Wildlife Service. 20pp.

Wofford, B. E. 1981. External seed morphology of *Arenaria* (Caryophyllaceae) of the southeastern United States. *Systematic Botany* 6(2): 126-135.

Wright, S. I., Lauga, B., and Charlesworth, D. 2003. Subdivision and haplotype structure in natural populations of *Arabidopsis lyrata*. *Molecular Ecology* 12: 1247-1263.

Wyatt, R. and Fowler, N. 1977. The vascular flora and vegetation of the North Carolina granite outcrops. *Bulletin of the Torrey Botanical Club* 104(3): 245-253.

Zhang, D. X., and Hewitt, G. M. 2003. Nuclear DNA analyses in genetic studies of populations: practice, problems and prospects. *Molecular Ecology* 12: 563-584.

Appendices

Appendix A: Tables

Table 1: Populations sampled and counties of occurrence.

Species	County	Population Sampled
M. cumberlandensis	Pickett, TN	Ladder Trail
		Hazard Cave
		Pickett Dam
	Scott, TN	Slave Falls
		Middle Creek
		Big Island A
		Big Island B
		Puncheoncamp Fork
	Fentress, TN	Laurel Fork
		Jamestown Reservoir
	Morgan, TN	Peters Bridge
M. glabra	Grainger, TN	Clinch Mountain
	Morgan, TN	Lilly Bridge
	Fentress, TN	Darrow Ridge
		Jamestown Flatrock

Table 2: Summary of population variation for *M. cumberlandensis* and *M. glabra*.

Population	n*	No. of haplotypes	Proportion heterozygous	Haplotype Diversity (SD)	Polymorphic sites (S)	Nucleotide diversity pi (SD)
All M. cumberlandensis	**110**	**25**	**0.236**	**0.947 (0.006)**	**34 (3 coding)**	**0.0126 (0.0065)**
Ladder Trail	10	6	0.400	0.844 (0.103)	18	0.0109 (0.0063)
Hazard Cave	10	2	0.000	0.356 (0.159)	11	0.0057 (0.0035)
Pickett Dam	10	3	0.200	0.378 (0.181)	17	0.0054 (0.0036)
Slave Falls	10	3	0.000	0.711 (0.086)	18	0.0133 (0.0076)
Laurel Fork	10	4	0.600	0.711 (0.118)	17	0.0082 (0.0049)
Middle Creek	10	1	0.000	0.000 (0.000)	0	0.0000 (0.0000)
Jamestown Reservoir	10	4	0.200	0.733 (0.101)	6	0.0026 (0.0018)
Peters Bridge	10	3	0.600	0.733 (0.076)	2	0.0014 (0.0012)
Big Island A	10	2	0.000	0.533 (0.095)	7	0.0054 (0.0034)
Big Island B	10	4	0.600	0.733 (0.101)	7	0.0020 (0.0015)
Puncheoncamp Fork	10	1	0.000	0.000 (0.000)	0	0.0000 (0.0000)
Avg. across populations		3	0.236	0.521	9.4	0.0050
All M. glabra	**40**	**18**	**0.450**	**0.939 (0.019)**	**23 (0 coding)**	**0.0071 (0.0039)**
Clinch Mountain	10	3	0.600	0.689 (0.104)	5	0.0040 (0.0026)
Lilly Bridge	10	6	0.600	0.889 (0.075)	15	0.0082 (0.0048)
Darrow Ridge	10	7	0.400	0.933 (0.062)	11	0.0073 (0.0044)
Jamestown Flatrock	10	3	0.200	0.511 (0.164)	9	0.0043 (0.0028)
Avg across populations		4.75	0.450	0.756	10	0.0060

*Number of alleles sampled

Table 3: Analyses of molecular variance (AMOVA).

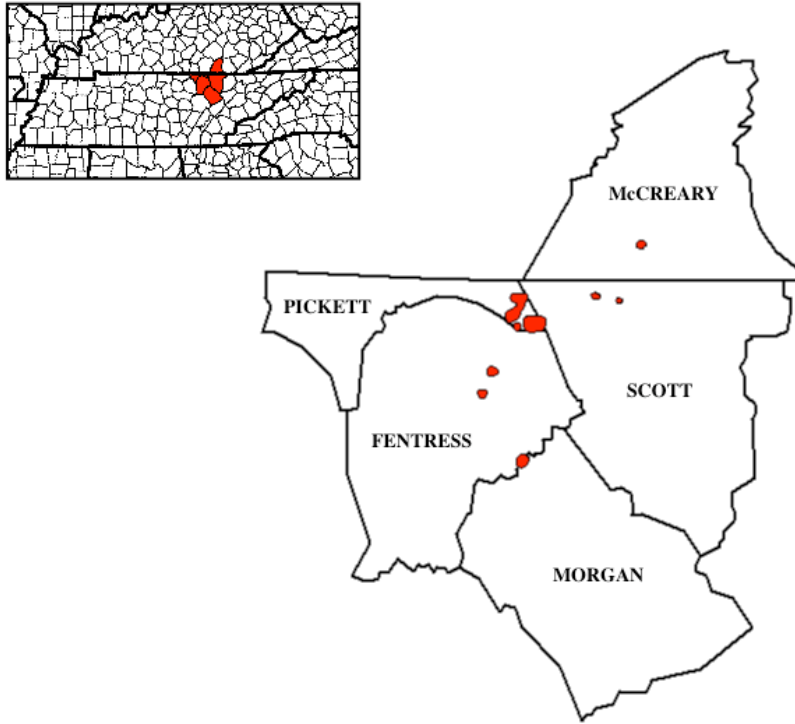| Source of variation	Percentage of Variation	
	Minuartia cumberlandensis	*Minuartia glabra*
Among Populations	63.27	21.73
Within Populations	36.73	78.27
	p<0.00001 for all values	

Appendix B: Figures

Figure 1: Range of *M. cumberlandensis* in Tennessee and Kentucky.
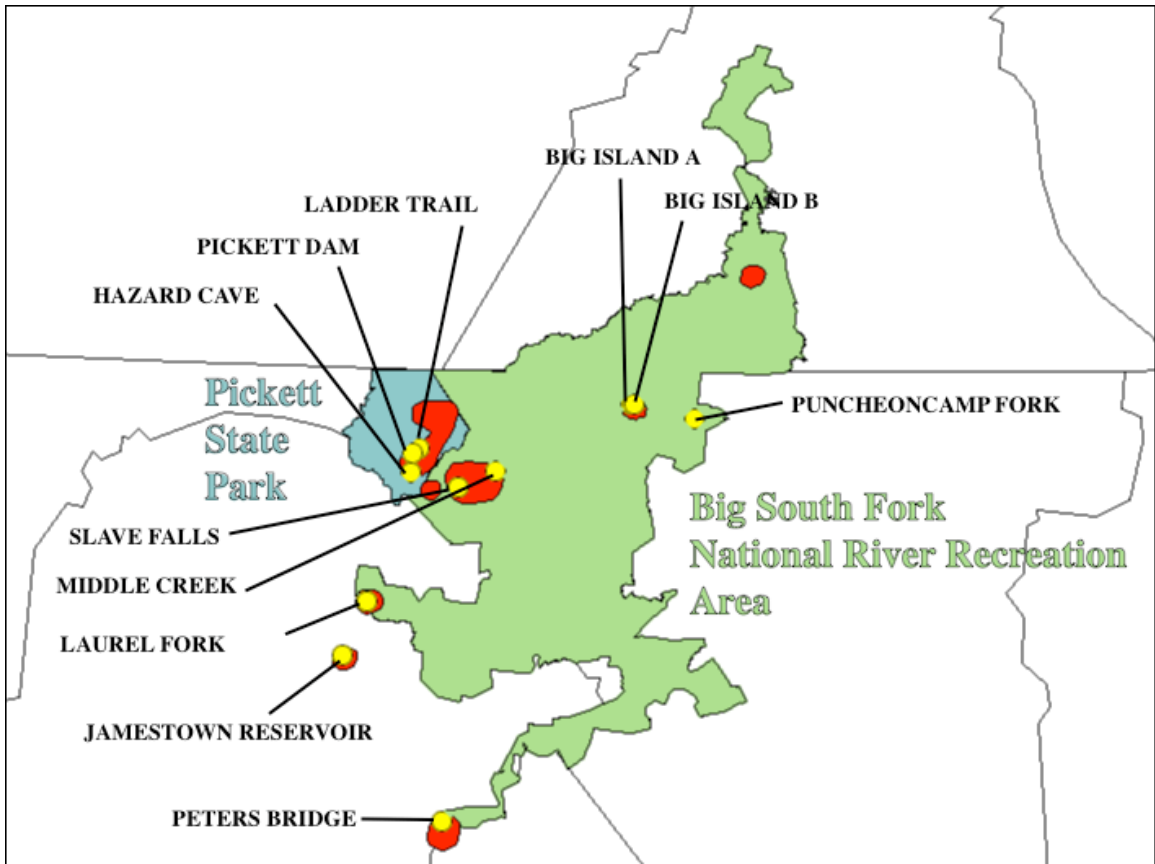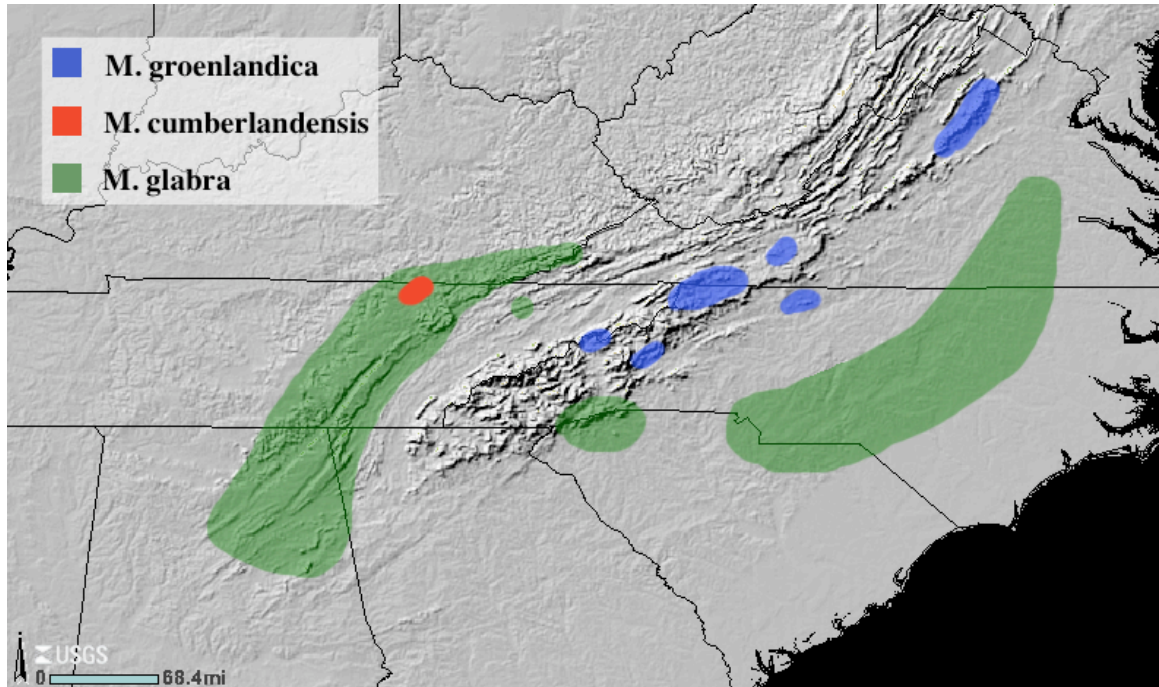
Figure 2: Sampled populations of *M. cumberlandensis*.

Figure 3: Southern ranges of species in the *M. groenlandica* complex (from Weaver 1970, Fishman and Wyatt 2004).
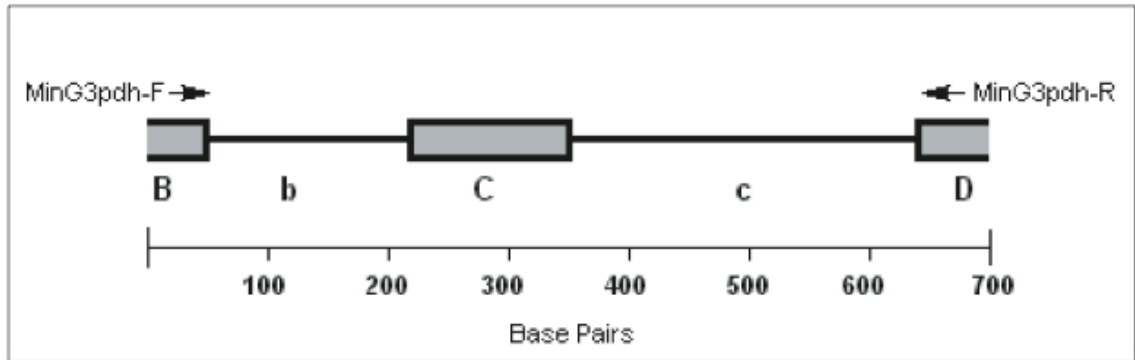
Figure 4: Sequenced region of glyceraldehyde 3-phospate dehydrogenase (*G3pdh*). Primer binding sites are indicated by arrows. Exon regions are B, C, and D, and introns are b and c.
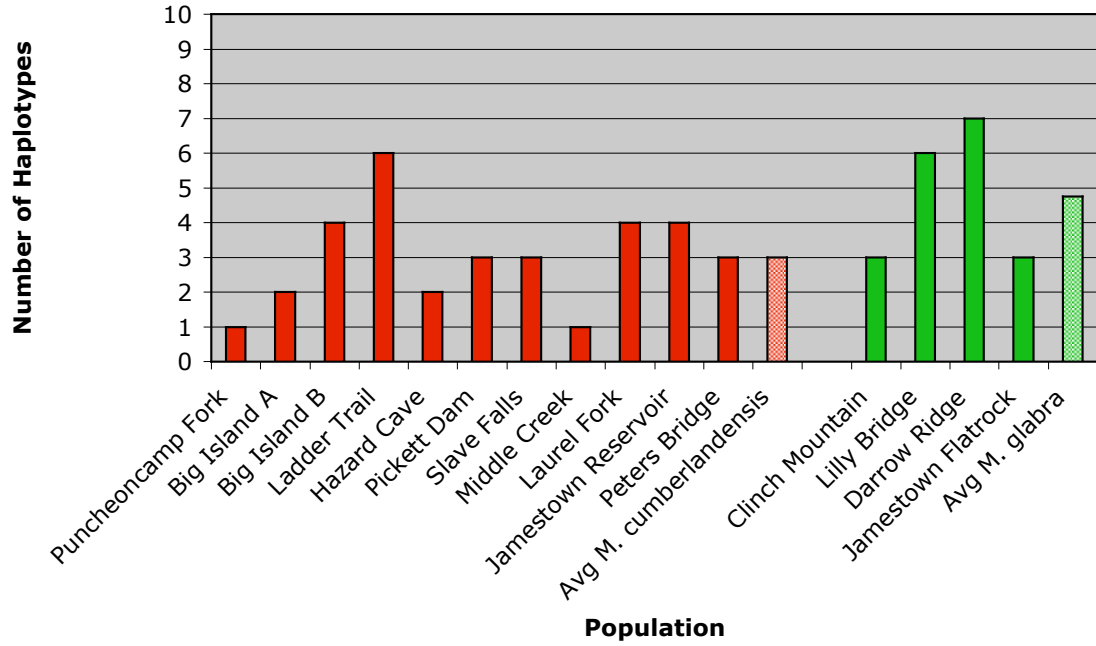
Figure 5: Matrix of polymorphic sites. Haplotypes *A* through *Y* are from *M. cumberlandensis* and haplotypes *a* through *r* are from *M. glabra*. The positions of polymorphic sites in the overall alignment are indicated by the vertically displayed numbers along the top of the figure. Sites in the coding region are underlined. A period (".") indicates that the sequence position is identical to the top sequence (haplotype C), and a dash ("-") represents an insertion or deletion at that position.

Figure 6: Number of haplotypes per population. *Minuartia cumberlandensis* is represented by red bars, and *M. glabra* by green bars. The average number of haplotypes across populations of a species is shown as a hatched bar.
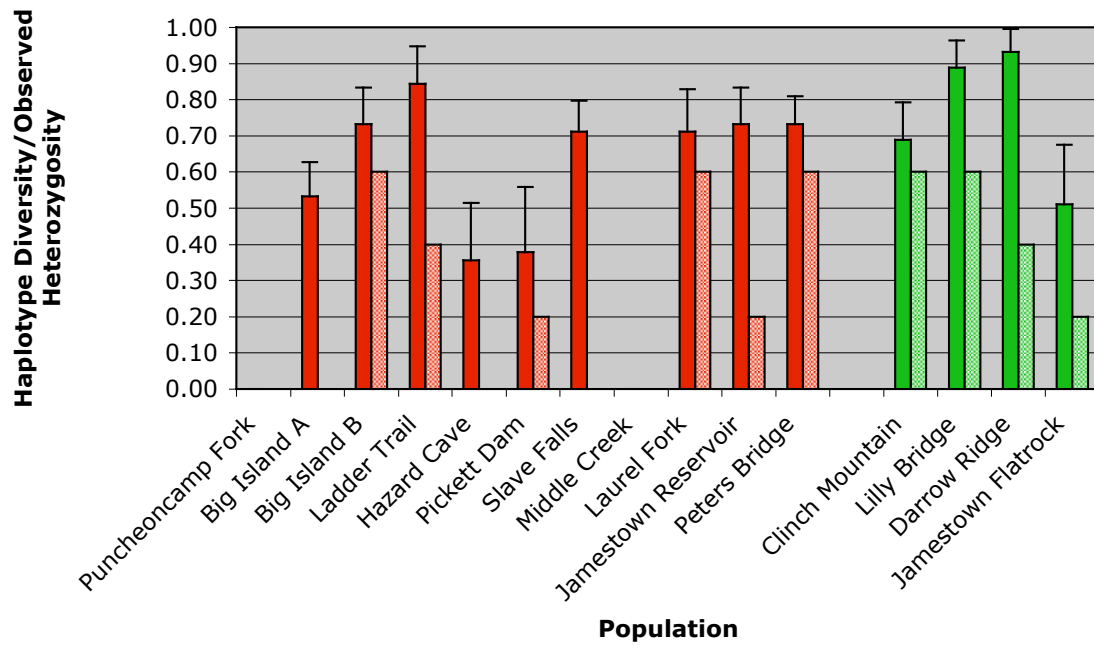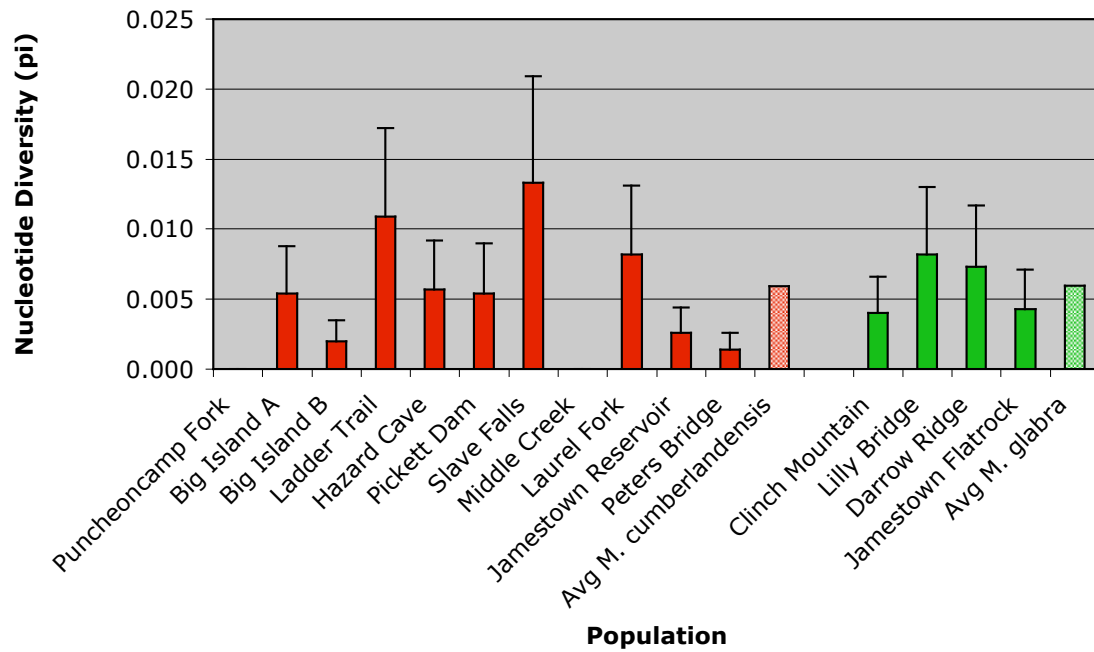
Figure 7: Haplotype diversity and observed heterozygosity within populations.
Populations of *M. cumberlandensis* are represented by red bars, and populations of *M. glabra* are represented by green bars. Solid bars are estimates of haplotype diversity and hatched bars indicate observed heterozygosity.

Figure 8: Nucleotide diversity within populations. Populations of *M. cumberlandensis* are represented by red bars and populations of *M. glabra* by green bars. The average value of nucleotide diversity across populations of a species is shown as a hatched bar.
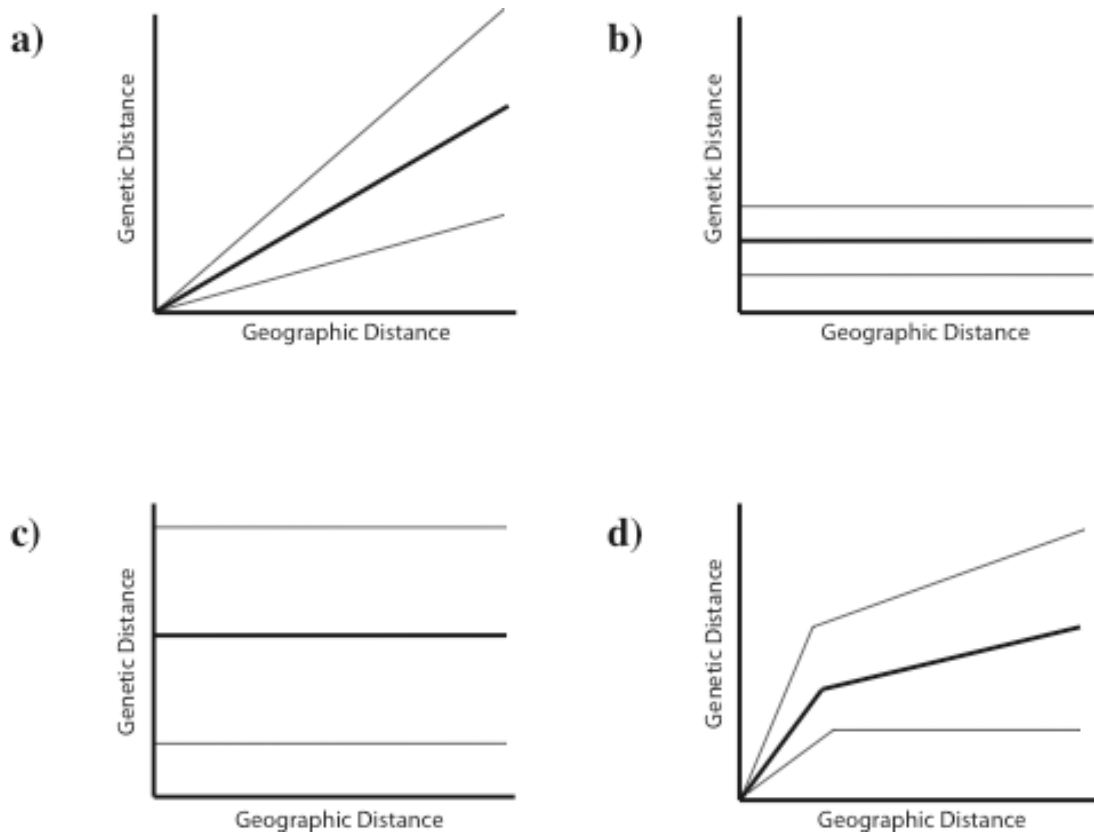
Figure 9: Possible relationships of geographic and genetic distance between populations. Shows hypothetical relationships of population genetic distance and geographic distance based on various scenarios of population structure under a stepping-stone model of gene flow; a) under conditions of equilibrium between loss of alleles due to drift and gain of alleles through gene flow; b) under conditions of panmixia, or dominance of gene flow; c) under conditions of complete fragmentation, or dominance of genetic drift within populations; and d) under conditions of shifting roles of gene flow and drift, with nearby populations experiencing high rates of gene flow, but more distant populations still showing divergence due to drift (adapted from Hutchison and Templeton 1999).
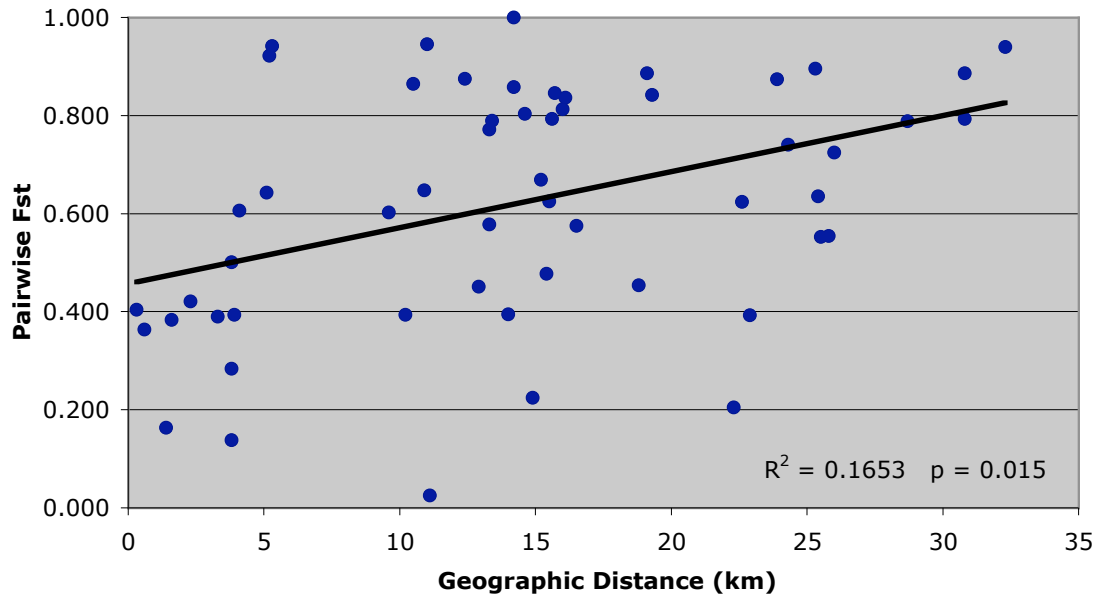
Figure 10: Population pairwise F_{ST} vs. geographic distance for *M. cumberlandensis*.
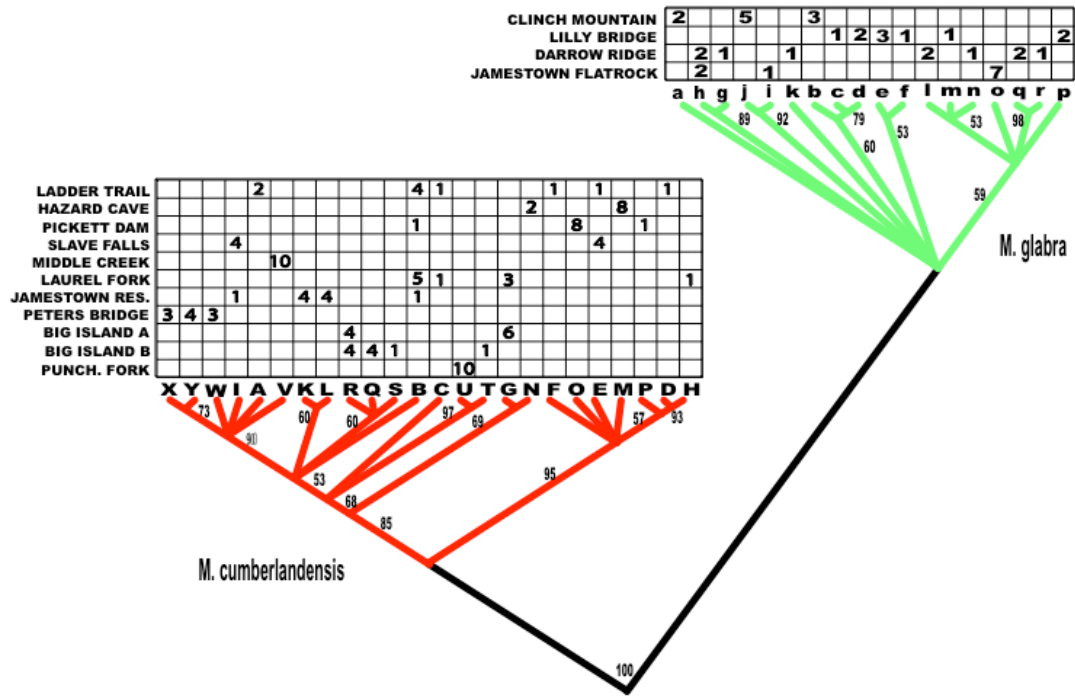
Green clade (M. glabra) haplotype counts:

	a	h	g	j	i	k	b	c	d	e	f	l	m	n	o	q	r	p
CLINCH MOUNTAIN	2			5		3												
LILLY BRIDGE							1	2	3	1		1						2
DARROW RIDGE	2	1				1						2		1		2	1	
JAMESTOWN FLATROCK	2			1											7			

Red clade (M. cumberlandensis) haplotype counts:

	X	Y	W	I	A	V	K	L	R	Q	S	B	C	U	T	G	N	F	O	E	M	P	D	H
LADDER TRAIL					2							4	1				1		1			1		
HAZARD CAVE												2					8							
PICKETT DAM											1					8		1						
SLAVE FALLS				4										4										
MIDDLE CREEK						10																		
LAUREL FORK									5	1			3										1	
JAMESTOWN RES.			1				4	4			5	1												
PETERS BRIDGE	3	4	3																					
BIG ISLAND A								4					6											
BIG ISLAND B							4	4	1				1											
PUNCH. FORK											10													

M. glabra

M. cumberlandensis

Figure 11: Bootstrap parsimony tree. Shows relationships of *G3pdh* haplotypes of *M. cumberlandensis* and *M. glabra*, rooted at midpoint. The numbers at the nodes of the tree represent levels of bootstrap support. The number of observed copies of each haplotype for each population are displayed in the tables above the tree.
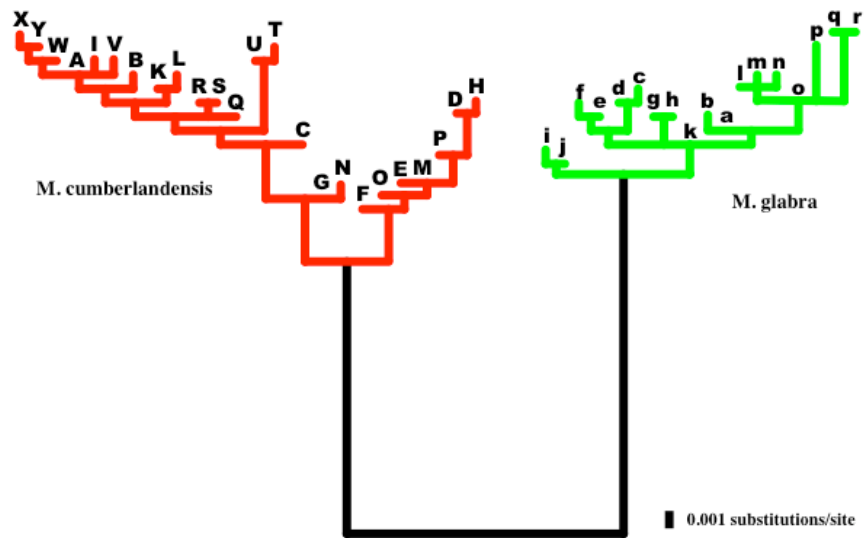
Figure 12: Maximum likelihood tree. Shows relationships among *G3pdh* haplotypes of *M. cumberlandensis* and *M. glabra*.
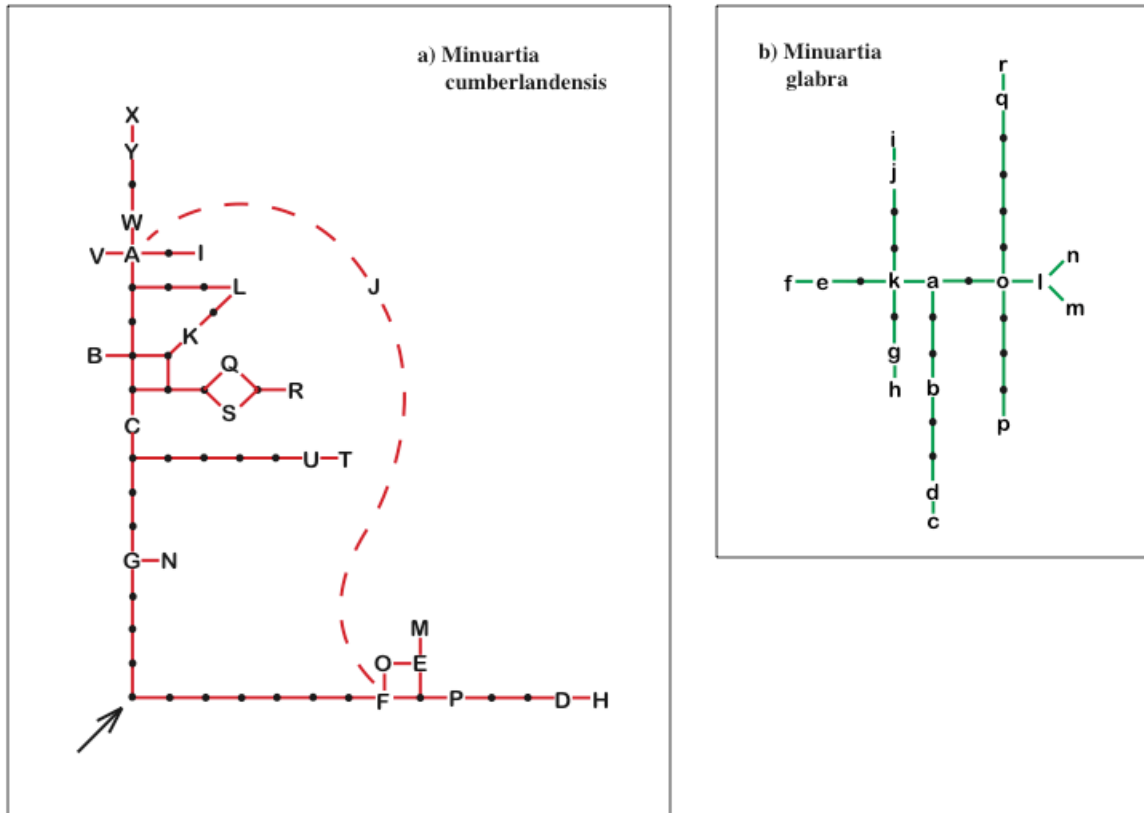
Figure 13: *G3pdh* haplotype networks. Networks for a) *M. cumberlandensis* and b) *M. glabra*, showing reticulations and interior haplotypes. Each dot represents an inferred haplotype that was not observed, and the line segments connecting haplotypes or inferred haplotypes represent mutational steps. The arrow in the *M. cumberlandensis* network represents the root of the tree, or the most recent common ancestor of all sampled haplotypes. The exact placement of haplotype J is indeterminate, but it is the product of recombination between haplotypes closely related to A and F.
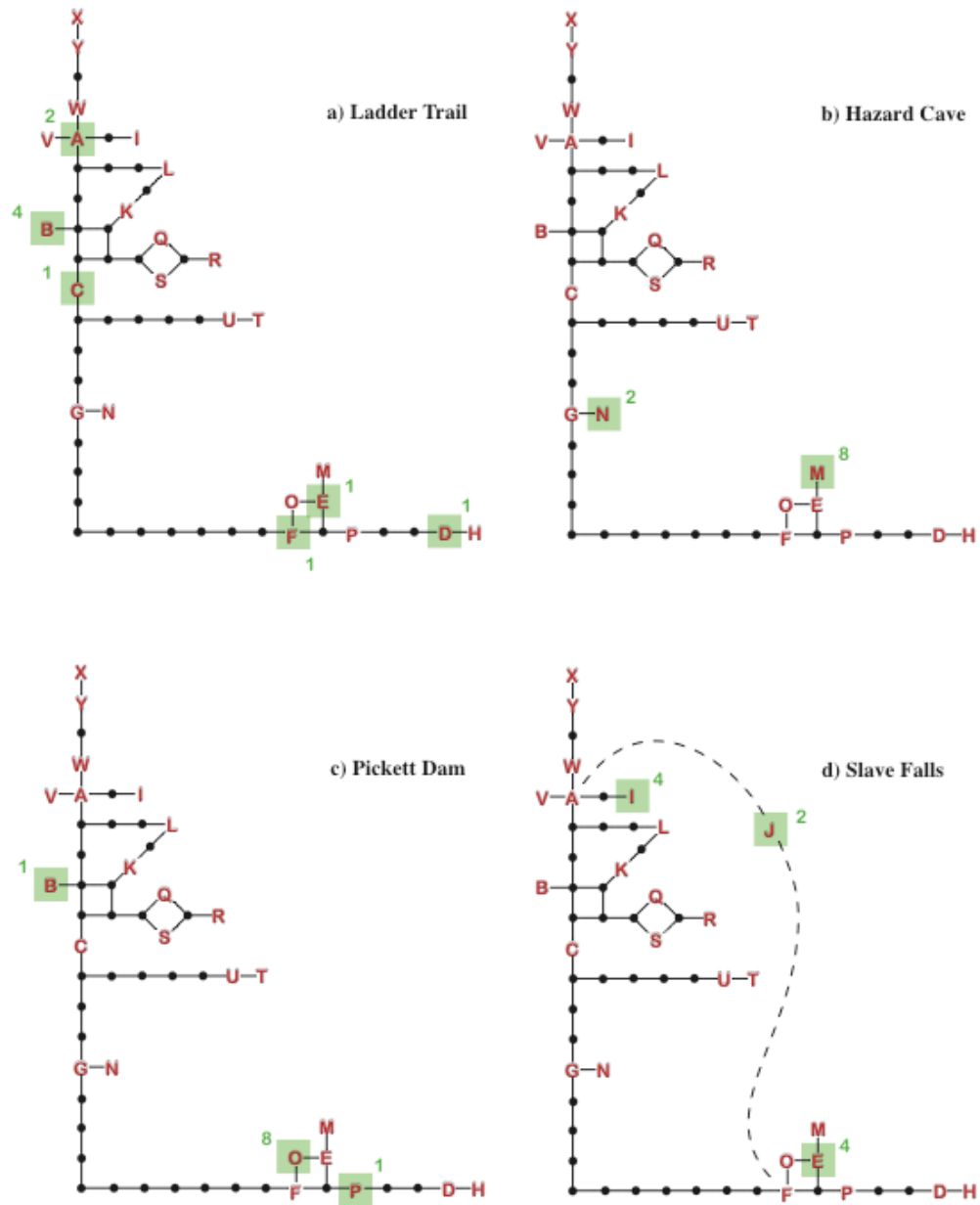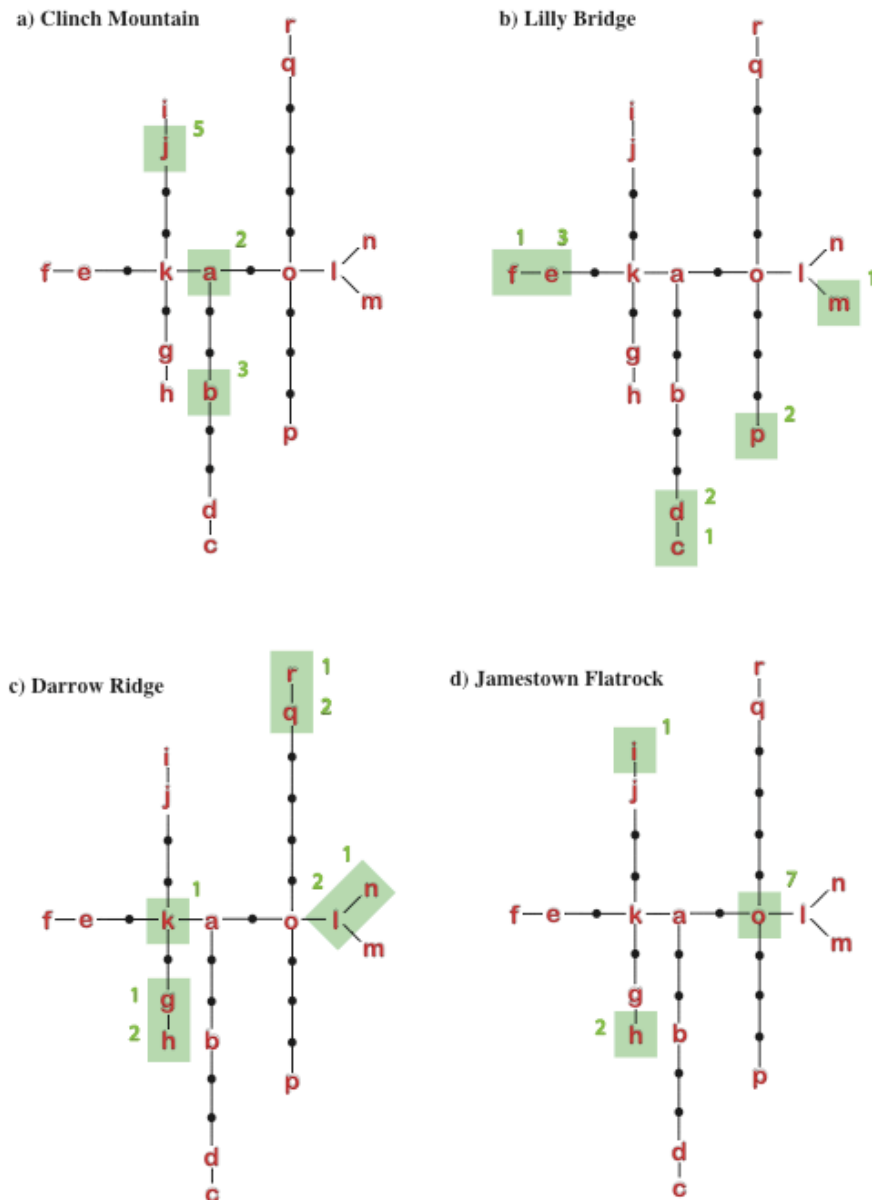
Figure 14: Population composition charts for *M. cumberlandensis*. For each population, observed haplotypes were highlighted in green on the overall species cladogram, with the number of copies of each observed haplotype indicated in green. Includes a) Ladder Trail; b) Hazard Cave; c) Pickett Dam; and d) Slave Falls. In the Slave Falls population (d), haplotype J was observed, which is a recombined form related to the two other observed haplotypes in the population (I and E).
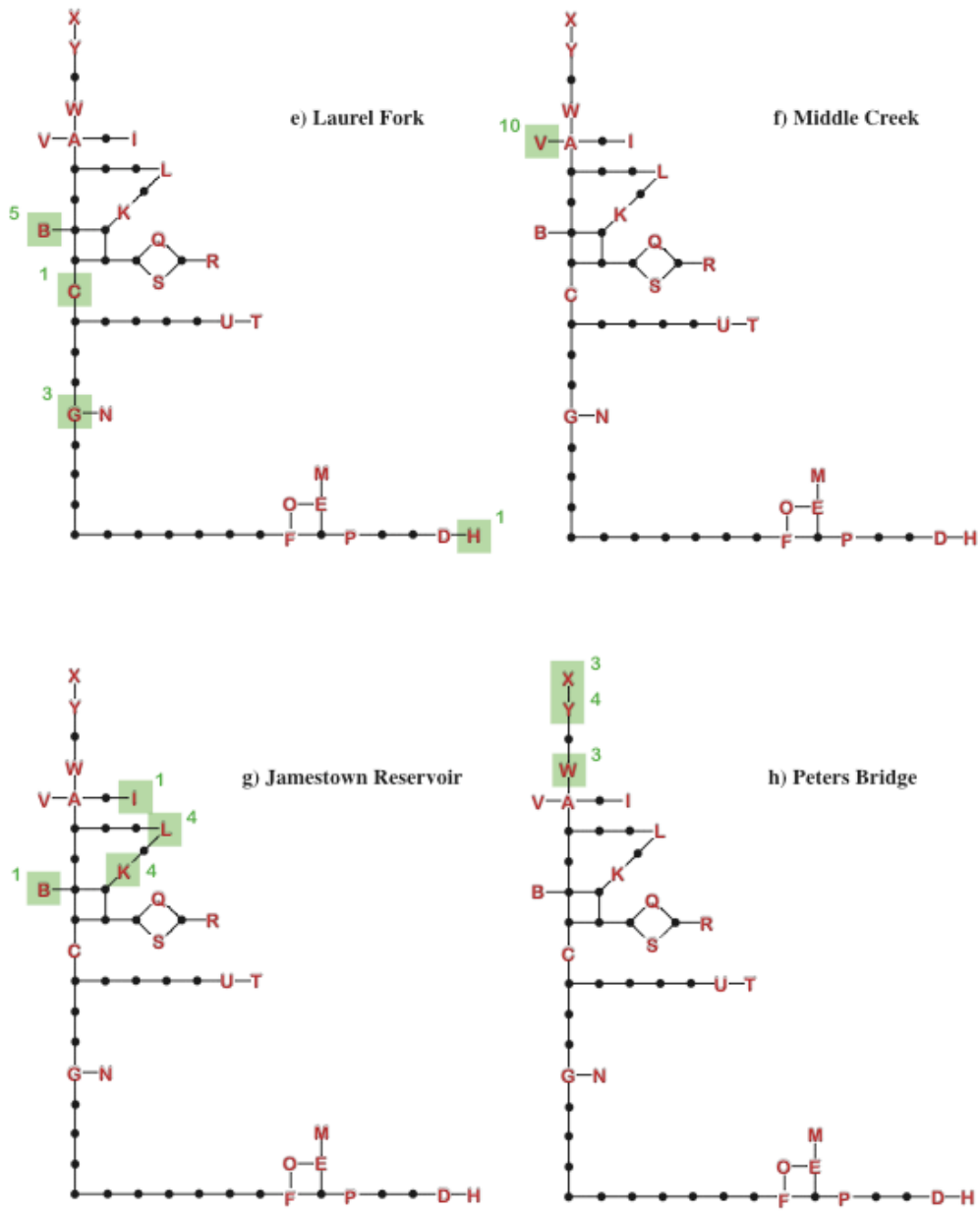
Figure 14 continued: Population composition charts for *M. cumberlandensis*. Continued. Includes e) Laurel Fork; f) Middle Creek; g) Jamestown Reservoir; and h) Peters Bridge.
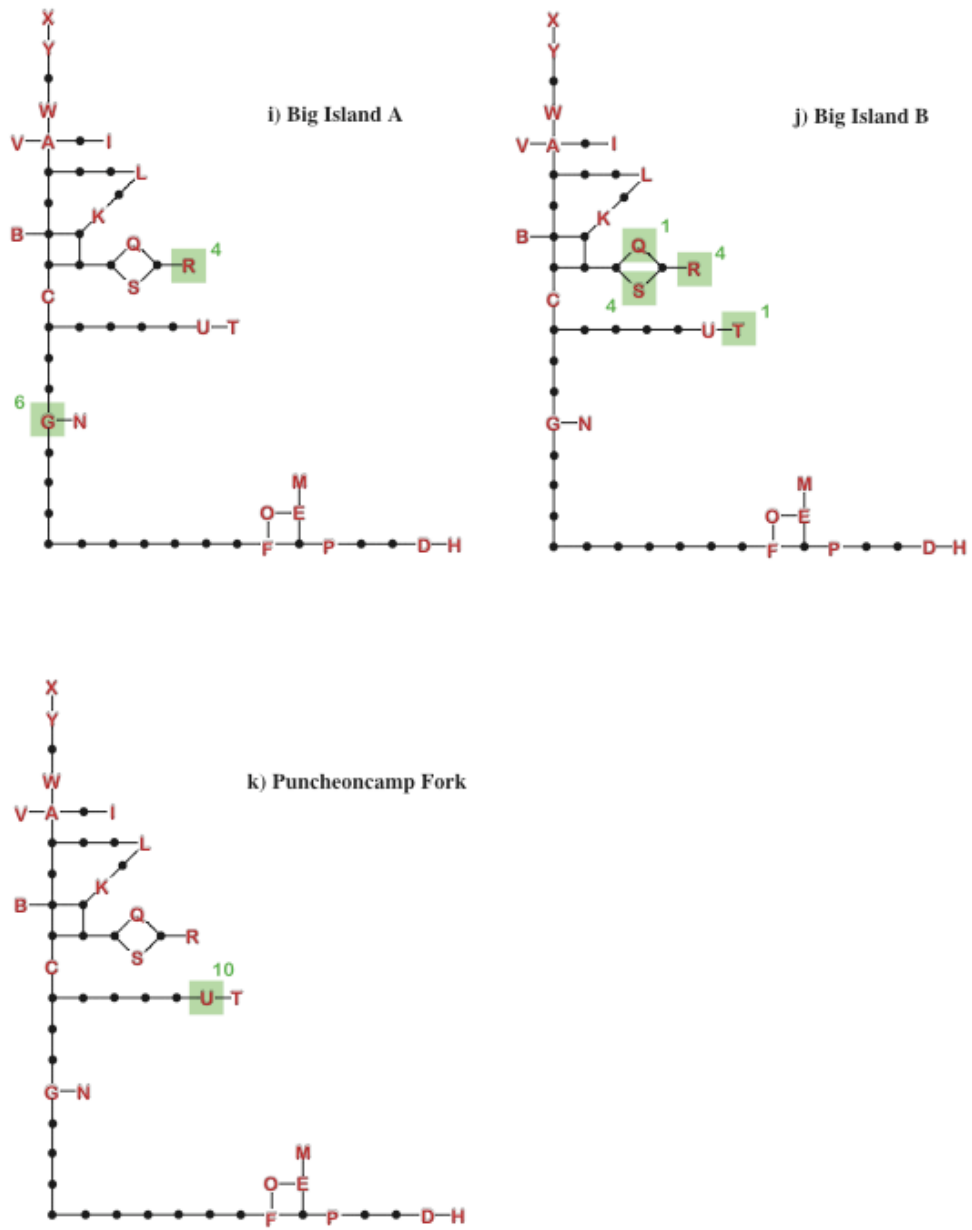
Figure 14 continued: Population composition charts for *M. cumberlandensis*. Continued. Includes i) Big Island A; j) Big Island B; and k) Puncheoncamp Fork

Figure 15: Population composition charts for *M. glabra*. For each population, the observed haplotypes were highlighted in green on the overall species cladogram, with the number of copies of each observed haplotype indicated in green. Includes a) Clinch Mountain; b) Lilly Bridge; c) Darrow Ridge; and d) Jamestown Flatrock.

67
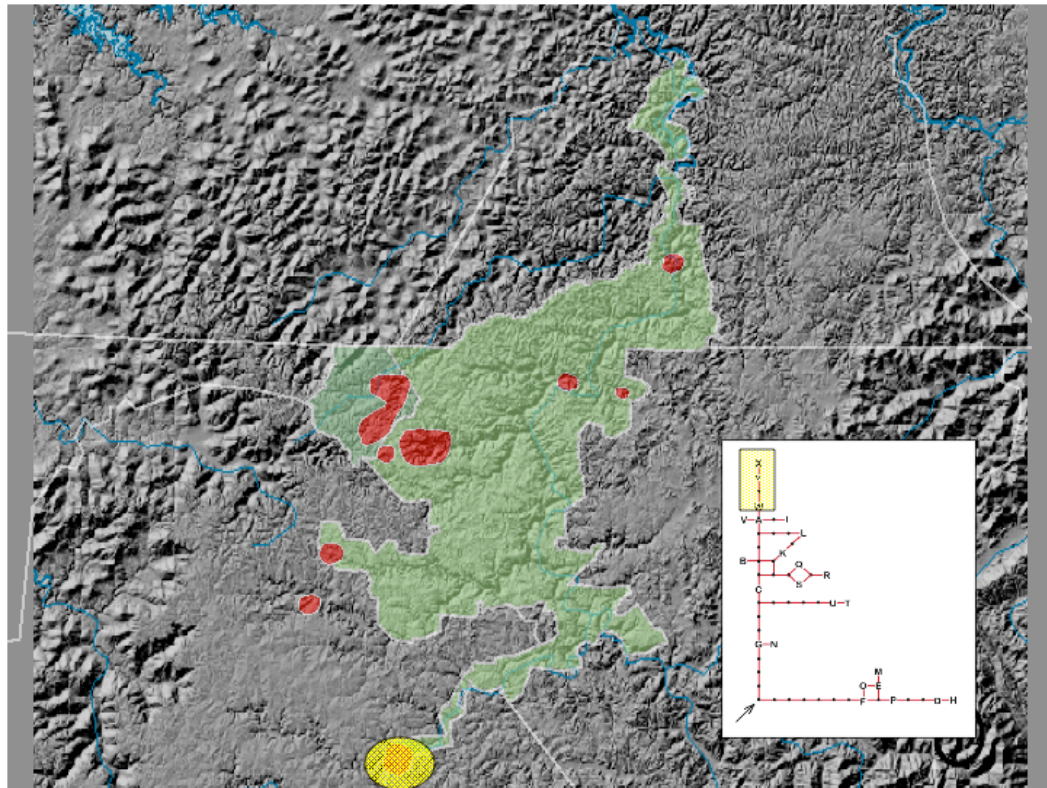
Figure 16: Distribution of *M. cumberlandensis G3pdh* haplotypes in the XYZ clade. Areas in red indicate the known range of the species, and areas in yellow indicate observed occurrences of haplotypes X, Y, or Z. The phylogenetic position of this clade is highlighted on the network in the right corner of the map.
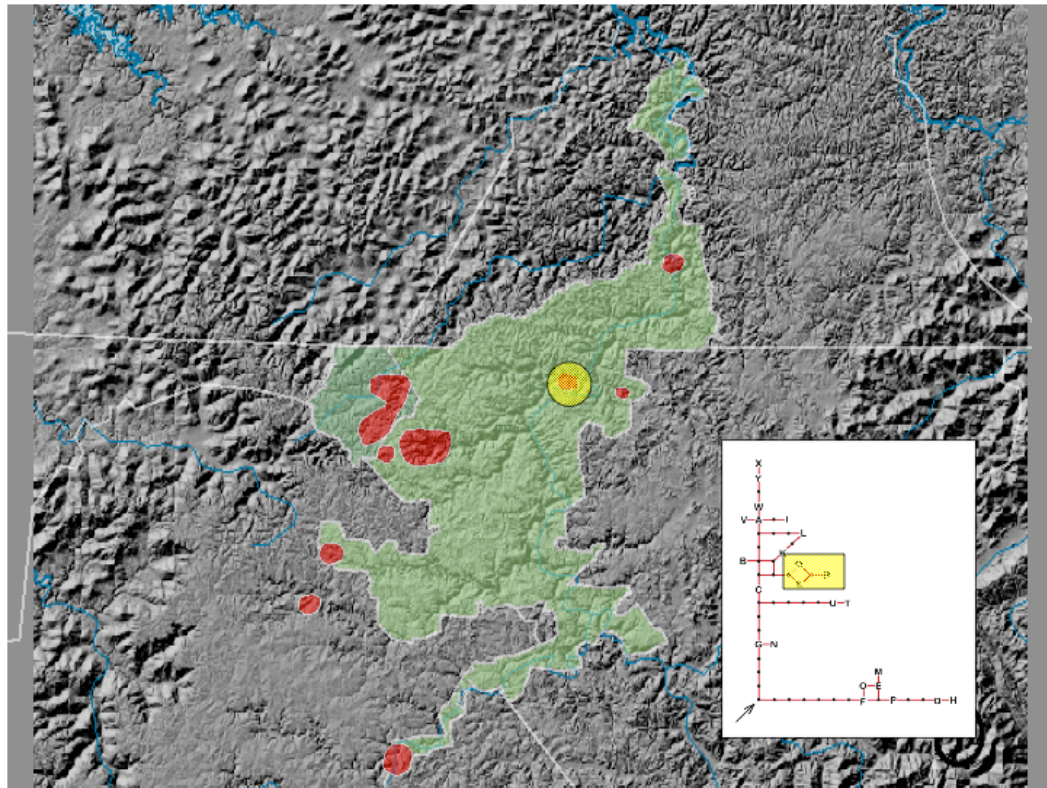
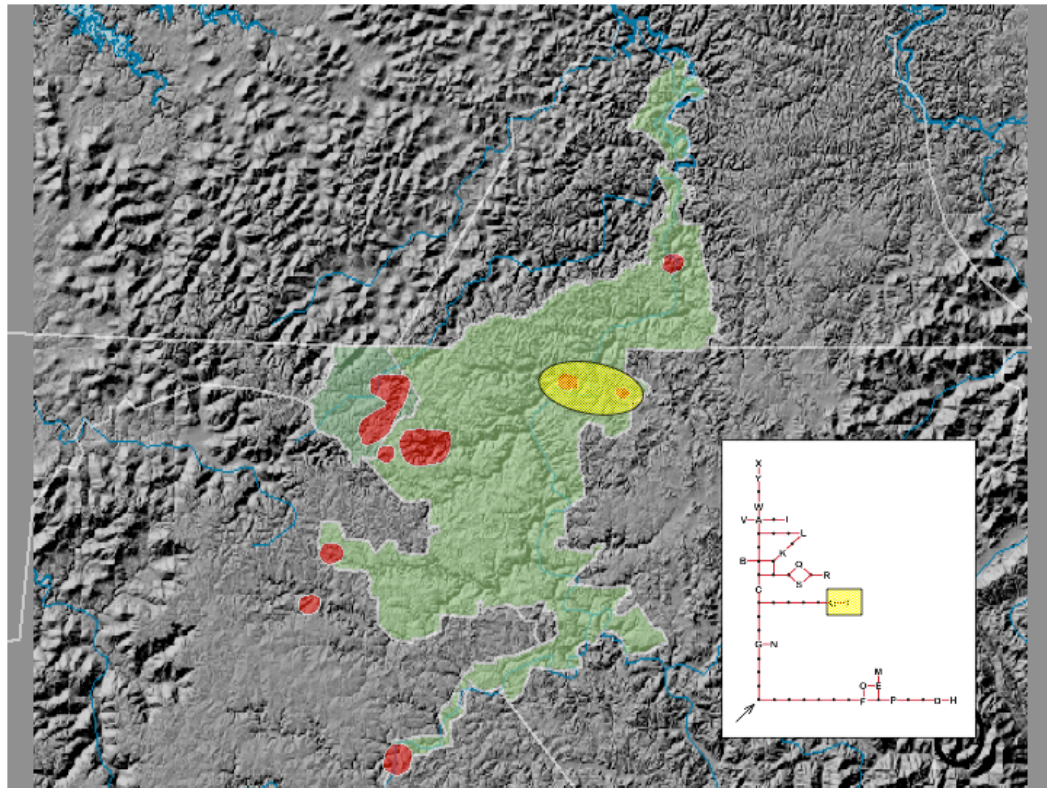**Figure 17:** Distribution of *M. cumberlandensis G3pdh* haplotypes in the QRS clade.

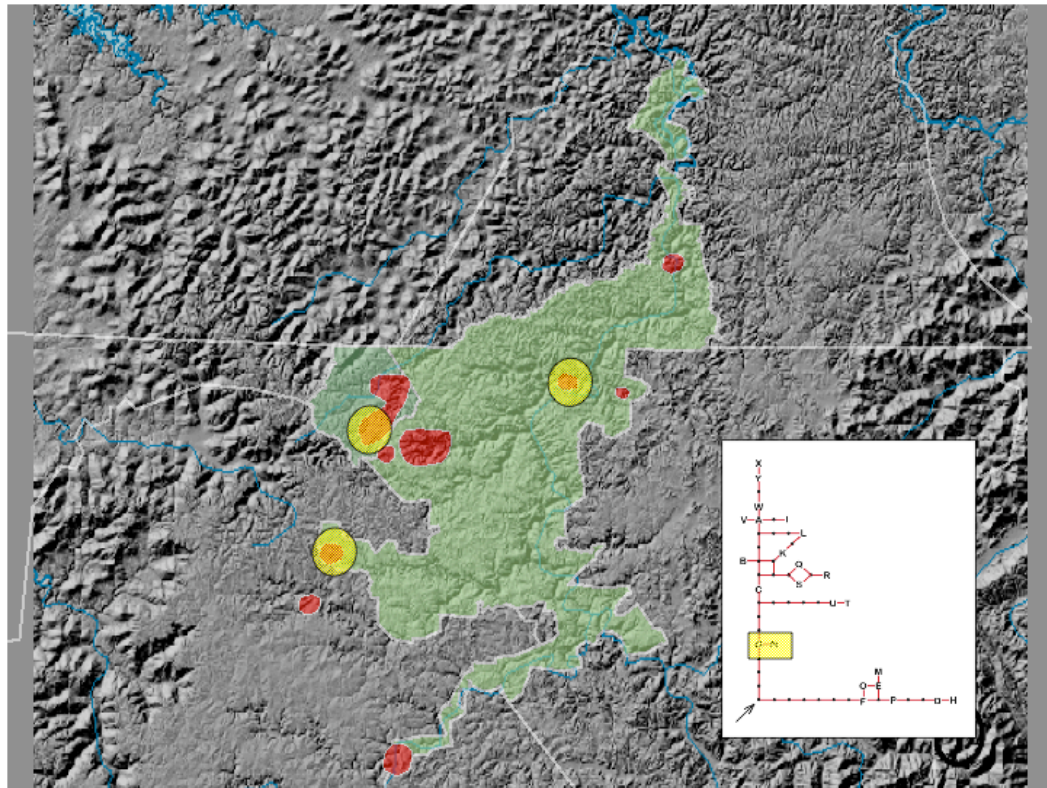**Figure 18:** Distribution of *M. cumberlandensis G3pdh* haplotypes in the UT clade.

Figure 19: Distribution of *M. cumberlandensis G3pdh* haplotypes in the GN clade.
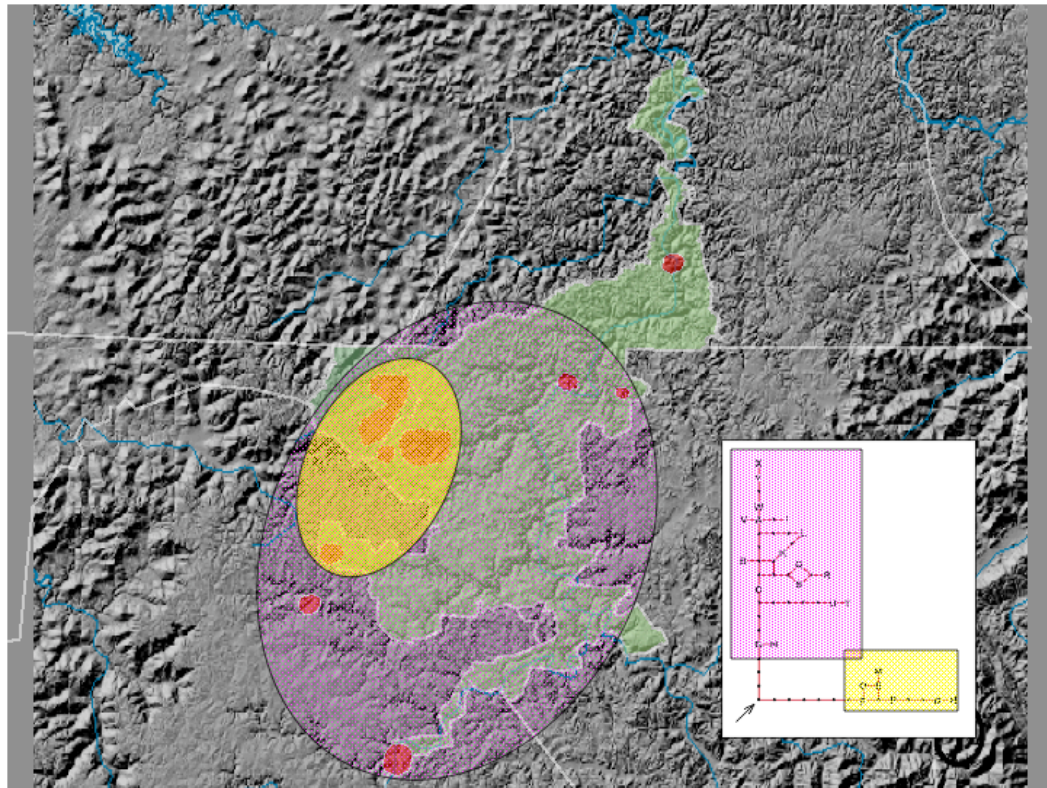
Figure 20: Distribution of *M. cumberlandensis G3pdh* haplotypes in the two major subclades of the network. The clade highlighted in purple occurs across the entire sampled range of the species, while the clade highlighted in yellow occurs only in the Pickett County populations and Laurel Fork.

Vita

Charles Thomas Winder, a native of Michigan, later transplanted to the South, was introduced at an early age to the botanical world when his mother put him to work in the garden pulling weeds. Ever since, he has had a fascination with the wonders of plants and their habits, and of the interaction of plants and humanity. With a strong concern for environmental impacts of human populations on natural systems, Charles is a devout believer in conservation of our natural heritage as a gift to our children and grandchildren, and will continue to work to protect endangered plants so that we may continue to learn from them.