A PRELIMINARY STUDY OF POPULATION STRUCTURE OF *SIBBALDIOPSIS TRIDENTATA* (AIT.) RYDG. (ROSACEAE) IN THE EASTERN UNITED STATES USING AFLP MARKERS

A Thesis
By
Gerald Edward Bresowar

Submitted to the Graduate School
Appalachian State University
In partial fulfillment of the requirements of the degree of
Masters of Sciences

December 2008
Department of Biology

A PRELIMINARY STUDY OF POPULATION STRUCTURE OF *SIBBALDIOPSIS TRIDENTATA* (AIT.) RYDG. (ROSACEAE) IN THE EASTERN UNITED STATES USING AFLP MARKERS

A Thesis

By

Gerald Edward Bresowar

APPROVED BY:

Gary L. Walker
Chairperson, Thesis Committee

Mary U. Connell
Member, Thesis Committee

Eva Gonzales
Member, Thesis Committee

Randall L. Small
Member, Thesis Committee

Steven W. Seagle
Chairperson, Department of Biology

Edelma D. Huntley
Dean, Research and Graduate Studies

ABSTRACT

A PRELIMINARY STUDY OF POPULATION STRUCTURE OF

SIBBALDIOPSIS TRIDENTATA (AIT.) RYDG. (ROSACEAE) IN THE

EASTERN UNITED STATES USING AFLP MARKERS

(December 2008)

Gerald Edward Bresowar, B.S., University of Tennessee Knoxville

M.S., Appalachian State University

Thesis Chairperson: Gary L. Walker

High-elevation habitats in the Appalachian Mountains of the United States have been shown to harbor many northern disjunct species. Prevailing theory is that populations in these habitats typically represent remnant-isolate populations from previously existing Quaternary communities. In this study I sought to develop a preliminary understanding of the basic populational and regional structuring in one of these northern disjunct species, in hopes of better understanding the relationships between population interactions and histories. Amplified fragment length polymorphisms were used to analyze inter-and intrapopulational genetic variation in nine populations of *Sibbaldiopsis tridentata* (Ait.) Rydb. (Rosaceae); a sub-shrub plant species with a main range in northeastern North America, and with disjunct populations at high elevations in the southern Appalachian Mountains. This study sought to determine patterns of genetic variation related to habitat type, glaciation history, present geographical features, and distances between populations. Nine populations in total were sampled from within the

eastern United States, representing various habitat types and physiographic provinces. Between 10 and 18 individuals were sampled in each population. I found that variation within and among these populations did not correlate strongly to glaciation history of individual sites, and only slightly correlated to habitat type of the sites. Population structural analysis identified regional affiliations between populations in the northeast and a single southern rock-outcrop population, an affiliation between the populations west of the Valley and Ridge Provence, as well as an affiliation between the two southern Appalachian grassy bald populations. The patterns of population structure observed in the Appalachians suggest postglacial population movements, though I was unable to identify population lineages. Further patterns of discontinuity between Appalachian populations and populations west of the Valley and Ridge Provence may be explainable by colonization northward post glaciation, coupled with the Valley and Ridge acting as a geographic or edaphic barrier to migration. I stress that this is a preliminary study, and further research is required to better clarify the population patterns within *S. tridentata* in the eastern United States.

ACKNOWLEDGMENTS

I'd like to thank the people who helped me in the development and implementation of this study. Firstly, to my committee chair, Dr. Gary Walker, I have immense appreciation for all the instruction and assistance he's gladly given me during this process. I'd like to thank Dr. Randall Small, without whom this study likely wouldn't have been possible. I thank Dr. Eva Gonzales and Dr. Mary Connell for their instruction and advice. Furthermore, I'd like to thank Dr. Zack Murrell for helping me grasp more conceptual elements of this field. I'm appreciative of James Sobieraj for laboratory assistance and instruction, Matt Valente for analysis advice and collection help, Steve Furches for laboratory assistance and instruction, and all the graduate students and staff of Appalachian State University Biology, who were more than willing to let me bounce ideas off them. I would also like to thank the ASU Biology Department and The Southern Appalachian Botanical Society for financial support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Climatic oscillations of the late Pleistocene left many present-day populations of plants, animals, and other organisms, with genetic structures indicative of historical-glacial and present-interglacial refugial states. Delcourt and Delcourt (1981) have uncovered evidence suggesting north-south population structurings among refugial and disjunct populations in eastern North America. These patterns presumably result from the north-south ordination of major geographic features in the southeastern United States, such as the Appalachian Mountain Range, along with migrations of species during periods of climate change. The importance of climatic factors on population movements is suggested to be paramount in our understanding species genesis, at least from a phylogenetic viewpoint (Avise 2000, Hewitt 1994). Patterns indicative of population dynamics resulting from glacial histories have been uncovered in studies from Europe, North America, Asia, and elsewhere (Bennett et al. 1991, Ikeda et al. 2006, Soltis et al. 2006). In the unglaciated portion of eastern North America several reoccurring discontinuities have been identified within the structural dynamics of populations. Most of these observed discontinuities have shown an east-west pattern, suggesting north-south migrations, typically with geologic features (rivers, valleys and mountains) at the interface of the distributional types (Parker et al. 1997, Al-Rabab'ah and Williams 2002, Griffin and Barrett 2004, Jody and Bruneau 2004, McLachlan et al. 2005).

GRASSY BALDS AND ROCK OUTCROPS

The grassy balds and rock outcrops of the southern Appalachians harbor several species with more northerly main ranges. Multiple studies using various methodologies have recognized high-elevation outcrops as interglacial refugia for these disjunct populations (Baskin and Baskin 1988, Clebsch and Walker 1988, Wiser et al. 1998, Kennedy and Walker 2007). Questions have persisted for many years about the origins of the grassy-bald habitats of this region, with explanations ranging from natural factors, such as ancient megafauna herbivory, fire regimes, ecotonal location, and recent climatic change (Mark 1958, Weigl and Knowles 1999); to anthropogenic factors such as Native American burns and post-European agricultural practices (Wells 1956, Lindsey and Bratton 1979, Sullivan and Pitillo 1988). Questions of population relationships in the region have had bearings on population management techniques and preservation priorities. Populations of more ancient lineages, or of greater molecular diversity, are often prioritized by conservationists, while more genetically depauperate populations may warrant preservation efforts if resources can be adequately allocated.

Though many scientists have speculated on the origin and maintenance of balds (Wells 1956, Sullivan and Pitillo 1988, Weigl and Knowles 1999), none so far have looked into the problem using a molecular approach. It is the intent of this study to sample and analyze populations of *Sibbaldiopsis tridentata* (Ait.) Rydb., as a southern disjunct in the southern Appalachians, for intra- and interpopulational variation which may increase our understanding of post-glacial movements of this species, as well as better our understanding of relationships between bald and outcrop populations. Differing

levels of variation between populations of balds and outcrops may suggest some barrier to gene flow between the balds and the outcrops. Similar levels of variation between bald and outcrop populations may suggest either ancient refugial status of the balds, or high-levels of gene flow between bald and outcrop communities. If the grassy balds and rock outcrops communities are both similar in approximate age as relictual communities, then both having derived from the same ancient paleoflora, I expect the levels of variation between the communities to be similar. If there is significant gene flow occurring between rock outcrop and bald communities, I expect a higher degree of relatedness between these populations.

STUDY SPECIES

Sibbaldiopsis tridentata (2n = 28), "three-toothed cinquefoil," is a member of the Rosaceae formerly placed within the genus *Potentilla*. Presently it is the only representative of its genus. It is a woody sub-shrub, bearing trifoliate leaves with apically tridentate leaflets, which persist with red coloring in the winter. The flowers (5-10 mm across) are arranged in a compact cyme, and typically have five sepals, five white petals, multiple stamens, and a single-jointed style leading to several single-seeded carpels. Though very little study has been done regarding the life-history of *S. tridentata*, similarities, both morphologically and ecologically, with the better-studied sister species *Sibbaldia procumbens* allows us to speculate on pollination regimes and other reproductive strategies in this species (Coker 1966). Like other cinquefoils, it is suspected to outcross as well as self-pollinate. Small flying insects are suspected to be the outcrossing pollinators, while ants are thought to be the inbreeding pollinators. The

author has documented the presence of minute flies on the flowers. The seeds of *S. tridentata* are small achenes, and likely germinate where they fall, though movement through animal transfer is possible.

Sibbaldiopsis tridentata has a northeastern North American main-range, which extends from New York State northward into eastern Canada and western Greenland. In the southeastern United States, *S. tridentata* is found in disjunct habitats, specifically high-elevation rock outcrops and grassy balds (Weakley 2007). Its occurrence in rock-outcrop habitats continues northward into the higher latitudes of the Appalachian Mountains, though the elevations of occurrence are lower with more northerly populations (Wiser 1998). Coastal populations are known in Maine northward, and populations are found on the rocky shores of the Great Lakes. Low-elevation populations are found in sandy-pine habitats of post-glacial Wisconsin and Michigan, and likely in similar communities in Canada. Population sizes range from a few individuals (<50) to very large numbers of individuals. In this study, larger populations were found in large-area rock fields and grassy habitats, while smaller populations were found in more restricted habitats, such as high-elevation outcrops or small, open patches in pine forests. Populations often occur on granite and amphibolite bedrocks, and are not reported from limestone. This species may be a calcifuge. A high-light level is thought to be a requirement for this species, and though the habitat types change somewhat with latitude, all populations we visited were characterized by a large degree of insolation. In grassy areas the species is often found on rocky microfeatures, exposed earth, and moss beds, and is largely excluded from thick patches of grass, though sometimes occurring

interspersed in less-dense patches of grass. In the forests of eastern Canada *S. tridentata* has been shown to prefer dry piney sites with higher light exposures, and is believed to be completely excluded from shady areas (Rowe 1956).

I sampled populations in the southeast and multiple populations from throughout the range of *S. tridentata* in the eastern United States, in order to expand the geographic range of this study, and to incorporate varying population-glacial histories and habitats in the analysis. Patterns of relationship discontinuity between populations and regions, coupled with our present understandings of genetic patterns resulting from the glacial periods of the Quaternary, were analyzed to help elucidate past population movements between habitats and regions.

PURPOSE AND HYPOTHESES OF STUDY

The purpose of this study is to elucidate the origins and histories of populations of *S. tridentata* in the eastern United States, attempting to better understand the histories of southeastern populations. This study will estimate levels of inter- and intrapopulational variation using amplified fragment-length polymorphisms (AFLPs). This technique (Vos et al. 1995) for analyzing whole-genome non-random restriction fragments has been used in multiple studies to detect levels of variation in a wide range of taxa (Boucias et al. 2000, Wang et al. 2003, Andrade et al. 2007), and has been used as a sensitive and statistically viable tool for detection of basic population structuring (Bonin et al. 2007). I sampled populations from varying habitat types, both in the glaciated and unglaciated ranges of this species. Several hypotheses were tested:

H1: Levels of genetic variation will differ by habitat type and history. Evidence for refugial populations, genetic drift possibly through founder effect (forming recent populations) or disturbance (such as trampling), along with possible gene flow into a single population, will be elucidated by incorporating knowledge of data, habitat and history.

H1a: Southern Appalachian rock outcrop and grassy bald populations indicate similar levels of variation. This suggests that the balds and outcrops are either similar in age, or are sharing genetic information via gene flow.

H1b: Balds and outcrops have significantly different levels of genetic variation. This indicates that the less variable population has been subjected to some form of population bottlenecking or founder effect.

H2: Levels of variation will differ by region. Inter-populational genetic movement within regions might have led to intra-regional, but not inter-regional similarity. Patterns of relationships between regions would be sought to clarify historical movements of the species into particular regions pre- and post-glaciation. Any significant discontinuities of distribution will be identified, as well as any corresponding features which might explain observed discontinuities.

MATERIALS AND METHODS

COLLECTION AND MOLECULAR PROCEDURES

I sampled nine populations of *Sibbaldiopsis tridentata* (Fig. 1). Four populations were located in the southeast, two bald (RM & BB) and two rock outcrop (BSM & SM). One population was collected within Wisconsin (BC), growing in a sandy low-elevation habitat. A population was sampled in the Dolly Sods wilderness of the Monongahela National Forest (DS) in West Virginia, in a heath and grassy high-elevation plateau habitat. In New York State one population each was sampled from the Delaware Valley escarpment of the Catskills (CE) on a pine-forested rock exposure, and from a rocky-alpine community on the side of Whiteface Mountain (WM) in the Adirondacks. Finally, I sampled a population on the side of Mt. Washington (MW), New Hampshire, in a rocky-alpine area. *S. tridentata* was found to be abundant in the RM, BB, DS, CE, WM and MW populations, with greater than 500 individuals estimated for these populations, and with RM, BB, WM and MW being particularly large (>2000 inds.). The SM, BSM and BC populations were estimated to be smaller, containing less than 500 individuals. Habitat type, GPS coordinates, elevation, and approximate population size were recorded and are given in Table 1.

Sampling involved removal of approximately one leaf from the plant (three leaflets), and drying the sample in silicate gel. In each population between 10 and 18

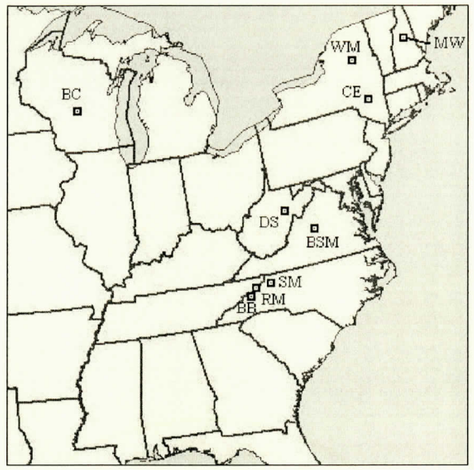


Fig 1. Locations and abbreviated names of *S. tridentata* populations sampled in this study. See (Table 1) for location, size and habitat type data.

Table 1. Physiographic information for populations included in this study. Location, name, abbreviation, no. samples, habitat type, and elevation are given. Population size: small = 1-100 inds., medium = 100-500 inds., large = 500-2000 inds., immense = >2000 inds.. Habitat type is based on the immediate area where the populations were collected.

Population	Abrev.	# ind.	Pop. size	Habitat type	Elev. (M)	Coordinates (Lat)	(Long)
Roan Mt., Chreokee and Pisgah NF, NC/TN	RM	11	immense	Grassy bald	1750	N36.10572	E82.49013
Snake Mt., Pisgah NF, NC	SM	12	small	Rock outcrop	1675	N36.33310	E81.70758
Big Bald, Cherokee NF, TN	BB	18	immense	Grassy bald	1670	N35.99009	E82.49013
Big Stoney Man, Shenadoah NP, VA	BSM	12	medium	Rock outcrop	1220	N38.59750	E78.37319
Dolly Sods Wilderness, Monogahela NF, WV	DS	10	large	Grassy clearings	1215	N39.03814	E79.31319
Catskills Escarpment, Catskills NF, NY	CE	10	medium	Piney outcrop	715	N42.20033	E74.03142
Whiteface Mt., Adirondack Mountains Park, NY	WM	10	immense	Rocky slope	1425	N44.36717	E73.90606
Mt. Washington, White Mountain NF, NH	MW	10	immense	Rock outcrop	1815	N44.28233	E71.27689
Beaver Camp, Warren, WI	BC	11	small	Sandy clearing	290	N44.16289	E90.27689

individuals were sampled. Collections were made every couple of paces. If the population had a patchy distribution pattern, sampling was spread about randomly. In the very large, less patchy populations, such as alpine rock fields, samples were taken several paces apart. I recorded population location using GPS [Garmin ETREX], and population size was classified into one of four classifications: small (1-100 individuals), medium (100-500 individuals), large (500-2000 individuals), and immense (>2000). Voucher specimens were collected from all populations where it was permitted, and are presently stored at Appalachian State University's herbarium, Boone, NC (BOON).

MOLECULAR TECHNIQUE

Genomic DNA of the samples was isolated using a modified CTAB extraction (Doyle and Doyle 1987). The resulting extracts had a gelatinous consistency, and would not amplify easily. I resolved this problem by re-running the samples through the suspension and washing steps of the CTAB procedure. The resulting extracts were much less viscous, with nucleotide concentrations between 50 and 300 ng/ul.

After scanning for genetic variation in five cpDNA regions of reputed variability, either in *Potentilla* (Ikeda et al. 2006), or across varied taxa (Shaw et al. 2007), as well as ITS 4&5 (Urbatsch et al. 2000), no discernable variation was found between two representative individuals from geographically distant populations (RM & MW). Direct sequencing was abandoned, and an Amplified Fragment-Length Polymorphism (AFLP) protocol was adopted.

AFLP data were obtained for 105 individuals using a slightly modified version of the original Vos et al. (1995) protocol. EcoRI and MseI were used to digest genomic

DNA, and custom adapters were ligated to the ends of the fragments. Three pairs of selective primers were used for the nested AFLP protocol (See Appendix A). LIV500 [Applied Biosystems] was used as an internal standard. Amplification of the ligated fragments was done using ExTaq [Takara], on a GeneAmp PCR 9700 thermocycler, and fragment analysis was done on a 3730xl DNA analyzer [both Applied Biosystems]. The resulting data files were loaded onto Peakscanner [Applied Biosystems], and peak tables were loaded onto Excel [Microsoft]. A subset of a peak-height data table is presented in Table 2. A full description and example of the transformation of raw AFLP data into a binary dataset is given in Appendix B.

Peaks which had heights lower than 100 across all samples were excluded from analysis, as were peaks of questionable quality and separation. All viable peaks between 50 and 200 bp in length were included in analysis. Binning was done by identifying specific fragment sizes, and creating bin boundaries 0.5 bp to either side of the fragment marker, the resulting bins being 1.0 bp in size. A binary dataset of scored peaks was created in an Excel [Microsoft] spreadsheet. Multiple runs of three individuals were made to determine reproducibility, and the resulting duplicates were determined to have a mean 97% reproducibility. From this I decided to treat individuals with greater than 97% similarity as analytically identical.

GENETIC ANALYSES

Analysis of molecular variance (AMOVA) is a statistical technique used to determine degrees of genetic variation within and between populations and regions. Populations are sorted into regional treatments so that questions about the significance

Table 2. An example of an AFLP peak/height data table. Population and primer-pair information is given at the top of the figure. Each individual is represented by two columns, one with given peak sizes and the second with the corresponding peak heights.

Beaver Camp Rd. WI Population									
Primers A & I (6FAM) Blue									
Size	**Height**								
GB145		GB148		GB149		GB150		GB151	
35.4132	78	35.6141	178	35.807	132	35.916	4260	35.6125	105
38.8955	162	38.4594	65	38.412	84	38.388	99	38.2506	57
41.5228	2015	42.5667	5385	41.681	2233	42.903	1733	41.462	1939
42.5252	2260	43.4591	1442	42.662	2488	43.71	2014	42.4576	2251
43.4236	716	45.5285	2473	43.543	831	44.522	741	43.4495	846
45.4071	1645	46.8007	4168	45.488	2357	46.242	1308	44.6348	840
46.687	1971	48.5511	557	46.841	3006	47.401	2065	45.5203	1285
48.6418	323	50.5593	388	48.76	372	48.653	347	46.7941	2048
50.4667	203	51.4902	2332	50.552	216	50.439	243	48.3534	453
51.3989	1013	52.4196	4858	51.471	988	51.319	971	50.3751	260
52.4224	2318	53.3474	2816	52.48	2345	52.203	2400	51.3118	1016
53.3511	1211	54.4587	342	53.396	1319	53.089	1452	52.2471	2284
54.3708	171	55.4755	1007	54.493	157	54.067	219	53.2742	1389
55.4809	575	56.4905	1313	55.497	609	55.138	456	54.2996	178
56.4043	560	57.5957	236	56.5	644	56.123	442	55.4162	418

of the regional treatment can be addressed. In this study AMOVA was done in GenAlEX (Peakall and Smouse 2006), and was conducted for several configurations of populations (regions), in attempts to identify the most likely sources of variation (Table 3). The regional treatments tested for in AMOVA were chosen to address questions involving variation between of populations of similar habitat type, glacial history, and geographic location. The east and west geographical regions were delineated by the Valley and Ridge Provence, which separated the higher-elevation regions of the Appalachian Mountains and the Allegheny/Cumberland highlands. The southeastern and northeastern delineation was based on the Virginia/Pennsylvania state line, which is in a low-elevation region, and thus considered a significant barrier to present-day interpopulational gene flow for northern-disjunct taxa. In the southeastern/northeastern treatment, BC and DS were excluded, as they represent geographical outliers.

Population structuring was analyzed with Bayesian cluster analysis in the program STRUCTURE v2.2 (Falush et al. 2007). This analysis assumes Hardy-Weinberg equilibrium, and has been used recently in several AFLP analyses (Valente 2007, Van Ee et al. 2006). STRUCTURE relies on defined numbers of clusters (K) provided by the user, rather than on established population boundaries. This analysis allowed me to better determine the number of actual populations in the sample set, and to elucidate how these populations associate with each other. Establishing the best-fit K-value (number of clusters) for a given dataset was a process of comparing the estimated natural log probability of data (Ln(N|k) for an array of possible K values. I processed the data

Table 3. Treatments tested for, and associated regions (A&B) in AMOVA analysis. Populations included in regions are given in parentheses. Note that not all populations are included in every parameter tested.

Treatment	Region A	Region B
Habitat	Rock (BSM, CE, MW, SM, WM)	Grassy (BB, DS, RM)
Habitat	Rock outcrop, S. App. (BSM, SM)	Grassy balds, S. App. (BB, RM)
Habitat	Rock outcrop (BSM, MW, SM, WM)	Non-rock outcrop (BB, CE, DS, RM)
Geographic	East of Valley and Ridge (BB, BSM, CE, MW, RM, SM, WM)	West of Valley and Ridge (BC, DS)
Geographic	Southeastern (BB, BSM, RM, SM)	Northeastern (CE, MW, WM)
Historical	Glaciated (BC, CE, MW, WM)	Unglaciated (BB, BSM, DS, RM, SM)

using the admixture model with 10,000 simulated generations for K-values 2-9, and I used the protocol of Evanno et al. (2005) to establish the best-fit K value for the dataset.

The software AFLP_SURV ver. 1.0 (Vekemans et al. 2002) was used to develop inter-population genetic diversity matrices. I chose default parameters which use Bayesian analysis with non-uniform distribution of allele frequencies, and which assume a Hardy-Weinberg distribution. The analysis was run with 500 permutations and 1000 bootstraps for genetic distance. Wright's fixation index F_{ST} (Wright 1951; Hartl and Clark 1989) was calculated, along with Nei's genetic distance (Nei and Li 1979) adapted for binary dominant markers by Lynch and Milligan (1994), and Reynold's et al. (1983) genetic distances. The resulting matrices were analyzed in the PHYLIP (Felsenstein 2007) module NEIGHBOR to construct corresponding dendrograms. From the resulting trees a majority-rule consensus tree was constructed using the PHYLIP module CONSENSE. Though the utility of trees in determining population structuring using AFLPs is perhaps questionable by some authors (Hollingsworth and Ennos 2004), I included them in this study to help determine which patterns seemed to reoccur across analysis techniques.

A pairwise population matrix was created, and population-level Principle Coordinate Analysis (PCA) was done, both in GenAlEx. PCA analysis allows for comparison of similarities between populations or individuals. This is accomplished through statistical analysis of eigenvalues and eigenvectors between samples and the resulting chart represents a three-dimensional (three axes) space in which populations are plotted according to relative similarity. Correlation between geographical and molecular

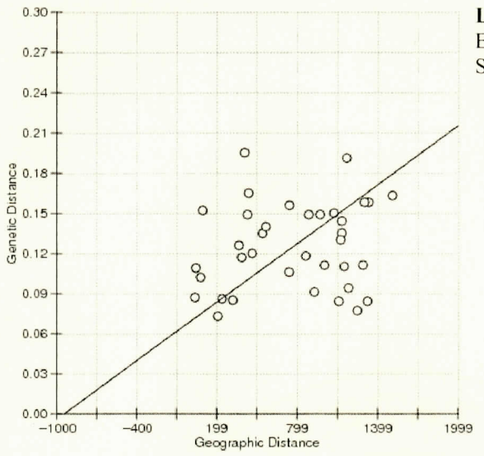
distance was estimated in the program ISOLATION BY DISTANCE (IBDWS) (Bohonak 2002). The program estimates F_{st} between pairs of populations (Weir 1990) and incorporates Mantel Testing to determine statistical significance of observed patterns. This analysis allowed me to estimate correlations between geographical and genetic distance between populations. IBDWS is reported with a p-value of 0.05. In order to determine levels of intrapopulational variation, I calculated percentage of polymorphic loci per population in GenAlEx.

RESULTS

Nine populations were sampled. Habitat type, GPS coordinates, elevation, and relative population size were recorded (Table 1). A total of 105 individuals were sampled, producing an average of 165 AFLPs between 50 and 200 bp in size. A total of 120 unambiguous and informative characters were scored within the fragment-size parameters. Of the 105 individuals sampled, no two were shown to be identical using this analysis (>97% similarity).

Analysis of Isolation by Distance (Fig. 2) showed little relationship between geographic and genetic distance (r^2=0.026; p=0.05). Principle-coordinate analysis (PCA) results for the nine sample populations are given in Figure 3.

Several regional treatments were shown to be significant contributors to variation across sampled populations. AMOVA analysis of all samples as a single region showed that variation within populations accounted for 81% of all variation observed. This left 19% of observed variation being among populations (P=0.001). Of the regional treatments tested for, the most significant treatment was the east/west of the Valley and Ridge Provence, with 7% of observed variation being found from among regions. The regional treatment for glacial history and the treatment for rock outcrop vs. non-rock



Linear model:	Intercept	Slope	r^2	n
Estimate	0.06880	7.304e-05	0.0260	36
SE	0.01097	1.236e-05		

Fig. 2. Linear regression of geographic-by-genetic-distance between pairs of populations (p = 0.05). Analysis was done in the program ISOLATION BY DISTANCE (IBD).

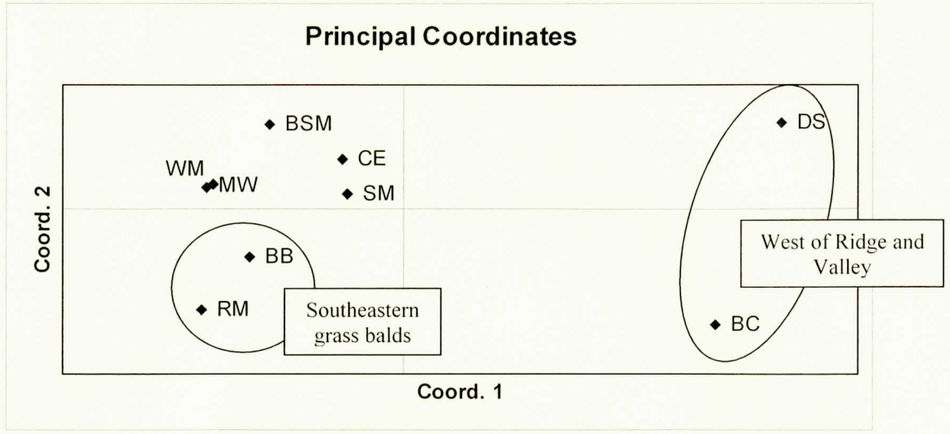


Fig. 3. Principle Coordinate Analysis of sampled populations executed in the program GenAlEx. Nine sampled populations are shown, and clusters representing the non-Appalachian populations (west of the Valley and Ridge) as well as the southeastern grass bald populations are designated.

outcrop populations, though showing zero variation due to these regions, had insignificant p-values (Table 4).

The pairwise population matrix of Nei's Genetic Distance (Table 5) showed WM and MW sharing the greatest genetic similarity, while RM and DS were the least similar. DS and BC displayed the highest mean genetic distance from the other populations (0.154), while SM had the lowest (0.105). The mean-across-populations genetic distance from other populations was 0.125, and the median was 0.119.

Comparison of dendrograms of Nei's genetic distance, Reynold's genetic distance, and pairwise populations F_{st} (Fig. 4) indicated multiple disagreeing clades across analyses. The F_{st} genetic distance tree presented the greatest degree of congruence compared to other analytical techniques applied in this study. The strict-consensus tree gave support for the monophyly of the populations west of the Valley and Ridge Provence, as well as the southern Appalachian grassy bald populations. The remaining populations were unresolved in the consensus treatment.

Structural analysis indicated several prominent and geographically-associated population clusters (Table 6; Fig. 5). Of the nine populations sampled, the northeastern populations (CE, MW and WM) clustered with SM throughout treatments K=2-9 ("northeastern affiliation"). At treatment K=7 the two southern Appalachian grassy bald populations (RM and BB) grouped together as did the populations west of the Valley and Ridge Province (BC and DS). The greatest level of clustering was seen at treatment K=7 and K=5, where three clusters were observed. These were the northeastern affiliation

Table 4. Results from AMOVA for various treatments. Refer to Table 2 for populations included in specific treatments.

Treatment	Source	df	SS	MS	Est. Var.	%
No regions	Among Pops	8	473.655	59.207	3.705	19%
(P=0.001)	Within Pops	96	1556.116	16.210	16.210	81%
	Total	104	2029.771		19.915	100 %
Southern grass	Among Regions	1	72.754	72.754	0.755	4%
balds/rock outcrops	Among Pops	2	102.273	51.137	2.620	13%
(P=0.005)	Within Pops	49	859.199	17.535	17.535	84%
	Total	52	1034.226		20.909	100%
Rock/grass habitats	Among Regions	1	69.261	69.261	0.268	1%
(P=0.011)	Among Pops	6	317.249	52.875	3.195	16%
	Within Pops	85	1410.199	16.591	16.591	83%
	Total	92	1796.710		20.054	100%
Rock outcrop/	Among Regions	1	58.601	58.601	0.041	0%
non-rock outcrop	Among Pops	6	327.909	54.652	3.322	17%
(P=0.312)	Within Pops	85	1410.199	16.591	16.591	83%
	Total	92	1796.710		19.954	100%
East/west of Valley						
and Ridge	Among Regions	1	102.685	102.685	1.453	7%
(P=0.001)	Among Pops	7	370.970	52.996	3.161	15%
	Within Pops	96	1556.116	16.210	16.210	78%
	Total	104	2029.771		20.823	100%
Southeastern/	Among Regions	1	62.554	62.554	0.388	2%
northeastern	Among Pops	5	244.360	48.872	2.669	13%
(P=0.005)	Within Pops	76	1313.399	17.282	17.282	85%
	Total	82	1620.313		20.339	100%
Glaciated/	Among Regions	1	60.843	60.843	0.035	0%
unglaciated	Among Pops	7	412.812	58.973	3.686	18%
(P=0.287)	Within Pops	96	1556.116	16.210	16.210	81%
	Total	104	2029.771		19.931	100%

Table 5. Population genetic-distance matrices computed in the program AFLP-SURV v.1.0. Hardy-Weinberg equilibrium is assumed, with 500 permutations applied. These data were subsequently used to create relationship trees (with consensus), see Fig 4.

Nei's D after Lynch and Milligan (1994)

	RM	SM	BC	BSM	CE	WM	MW	BB	DS
RM	0.0000								
SM	0.0273	0.0000							
BC	0.0816	0.0738	0.0000						
BSM	0.0781	0.0566	0.1228	0.0000					
CE	0.0744	0.0357	0.0912	0.0840	0.0000				
WM	0.0367	0.0165	0.0879	0.0494	0.0280	0.0000			
MW	0.0325	0.0125	0.0952	0.0560	0.0365	0.0109	0.0000		
BB	0.0341	0.0458	0.0805	0.0753	0.0644	0.0414	0.0379	0.0000	
DS	0.1226	0.0731	0.0933	0.1003	0.0889	0.0950	0.0967	0.1125	0.0000

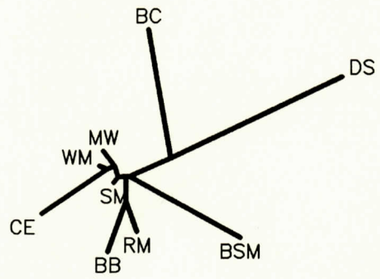
Pairwise F_{st} between populations

	RM	SM	BC	BSM	CE	WM	MW	BB	DS
RM	0.0000								
SM	0.0367	0.0000							
BC	0.1249	0.1185	0.0000						
BSM	0.1164	0.0916	0.2031	0.0000					
CE	0.1150	0.0636	0.1644	0.1498	0.0000				
WM	0.0542	0.0265	0.1449	0.0861	0.0532	0.0000			
MW	0.0481	0.0202	0.1537	0.0957	0.0672	0.0186	0.0000		
BB	0.0524	0.0697	0.1371	0.1259	0.1129	0.0679	0.0623	0.0000	
DS	0.1994	0.1407	0.1945	0.2007	0.1863	0.1801	0.1816	0.2071	0.0000

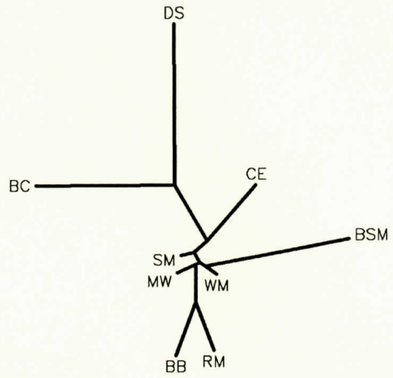
Genetic distance between populations after Reynolds et al. (1983)

	RM	SM	BC	BSM	CE	WM	MW	BB	DS
RM	0.0000								
SM	0.0374	0.0000							
BC	0.1334	0.1262	0.0000						
BSM	0.1238	0.0961	0.2271	0.0000					
CE	0.1222	0.0658	0.1796	0.1623	0.0000				
WM	0.0557	0.0269	0.1566	0.0900	0.0547	0.0000			
MW	0.0493	0.0204	0.1669	0.1006	0.0696	0.0188	0.0000		
BB	0.0538	0.0722	0.1474	0.1346	0.1198	0.0703	0.0643	0.0000	
DS	0.2224	0.1517	0.2163	0.2240	0.2062	0.1985	0.2005	0.2321	0.0000

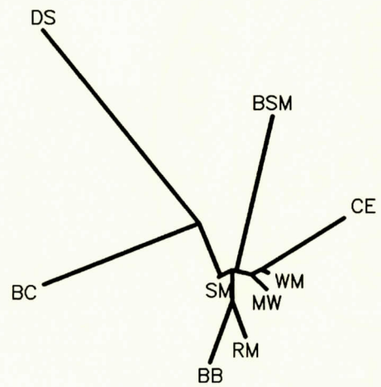
a. Pairwise F_{st}



b. Nei's D after Lynch and Milligan (1994)



c. Reynold et al. (1983) gen. dist.



d. Strict-consensus

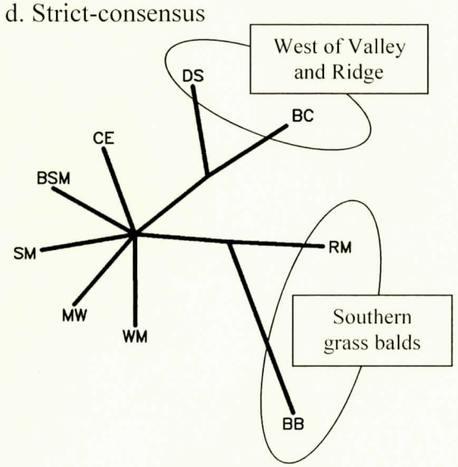


Fig. 4. Unrooted dendrograms produced from binary dominate data analyses using the program AFLP-SURV v.1.0 (a-c), and a strict consensus tree of trees a-c produced in the CONSENSE module of the program PHYLIP v. 3.68 (d). 500 permutations were used to construct a-c. See Table X for distance matrices.

Table 6. Population structure analysis from the program STRUCTURE. Population clustering tables are given for set clusters (K) of K = 9-5. The proportion of association (0-1.0) each population has with a given cluster (PPA) is given. Only populations which sorted into a cluster are shown here. The LnP(N|K) values of each analysis is also given. See Fig. 5 for cluster maps.

K=9	**Ln P(N|k)= -7670.5**	
Pop	**Cluster**	**PPA**
CE	**1**	0.796
WM	**1**	0.509
MW	**1**	0.453
SM	**1**	0.344

K=8	**Ln P(N|k)= -7746.5**	
Pop	**Cluster**	**PPA**
CE	**1**	0.765
WM	**1**	0.489
MW	**1**	0.484
SM	**1**	0.326

K=7	**Ln P(N|k)= -7823.4**	
Pop	**Cluster**	**PPA**
CE	**1**	0.821
WM	**1**	0.537
MW	**1**	0.478
SM	**1**	0.387
BB	**2**	0.595
RM	**2**	0.214
DS	3	0.788
BC	3	0.704

K=6	**Ln P(N|k)= -7946.7**	
Pop	**Cluster**	**PPA**
CE	**1**	0.797
WM	**1**	0.5
MW	**1**	0.462
SM	**1**	0.398
DS	3	0.764
BC	3	0.704

K=5	**Ln P(N|k)= -8137.7**	
Pop	**Cluster**	**PPA**
CE	**1**	0.677
BSM	**1**	0.638
WM	**1**	0.555
MW	**1**	0.542
SM	**1**	0.361
BB	**2**	0.59
RM	**2**	0.274
DS	3	0.806
BC	3	0.694

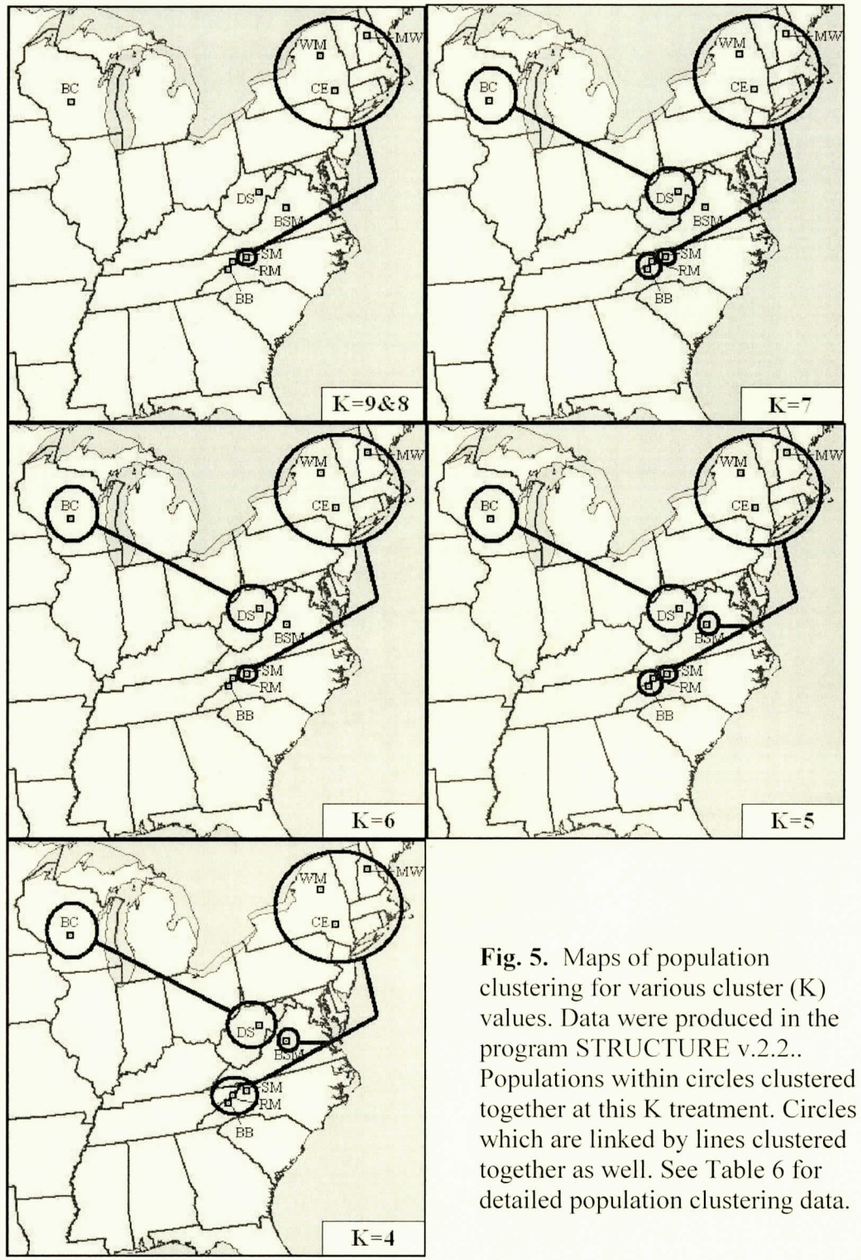


Fig. 5. Maps of population clustering for various cluster (K) values. Data were produced in the program STRUCTURE v.2.2.. Populations within circles clustered together at this K treatment. Circles which are linked by lines clustered together as well. See Table 6 for detailed population clustering data.

(including BSM at K=5), the southern Appalachian grassy balds, and populations west of the Valley and Ridge Provence. Throughout all K treatments RM displayed the weakest association with individual clusters. The associated log probability of data (Ln P(N|k) values are plotted in a line graph for K-values 2-10, and is given in Figure 6. The K-values 7 and 8 provided the highest Ln P(N|k)-values.

Percentage of polymorphic loci per population (Table 7) indicated that the highest percentage of polymorphic loci was found in the SM population, while the lowest percent was found in the DS population. The highest percent mean-over-geographic-region was found in the southeastern RM, BB and SM populations (83.33%), followed by the WM and MW populations of the northeast (79.59%). The lowest mean percentage of polymorphic loci by geographic region was found in the west of the Valley and Ridge populations DS and BC (57.5%). Between the grassy-bald populations and the rock-outcrop populations of the southeast, the highest percent polymorphic loci mean was found in the grassy-bald populations (82.09%), while the outcrop mean was lower (76.67%). Between population groupings based on approximated-population size, populations with greater than 2000 individuals had the highest mean percent polymorphic loci (80.84%), followed by small populations with less than 100 individuals (78.33%). The lowest percent-polymorphic loci for population size group was 44.17%, and was found in the single population designated as large (500-2000 individuals), DS.

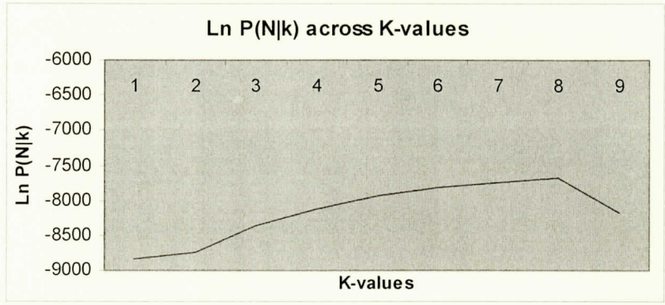


Fig. 6. Graph of Ln P(N|k)-values across the sampled K-values modeled in the software program STRUCTURE v2.2 (Falush et al. 2007). According to the protocol of Evanno et al. (2005), the most likely number of clusters (K) is best deterimined by comparison of Ln P(N|k)-value across sampled K-values. K-values with comparitively high Ln P(N|k)-values are considered more likely to be accurate representations of true population structuring.

Table 7. Percentage of polymorphic loci in each sampled population, with mean and standard error. Number of individuals sampled per population, habitat type of population, and approximate population sizes are given. Population size: small = 1-100 inds., medium = 100-500 inds., large = 500-2000 inds., immense = >2000 inds. The numerical data were calculated in GenAlEx (Peakall and Smouse 2006).

Population	%P	# ind.	Habitat type	Pop size
RM	84.17%	11	Grass bald	immense
BB	80.00%	10	Grass bald	immense
SM	85.83%	12	Rock outcrop	small
BSM	67.50%	12	Rock outcrop	medium
WM	80.00%	10	Rocky slope	immense
MW	79.17%	10	Rocky slope	immense
DS	44.17%	11	Grass heath	large
CE	61.67%	10	Piney outcrop	medium
BC	70.83%	11	Sandy clearing	small
Mean	72.59%			
SE	4.44%			

DISCUSSION

The main objective of this study was to determine the genetic structure of *Sibbaldiopsis tridentata* populations in the eastern United States, in order to gain a better understanding of the natural history of grassy balds in the southern Appalachians. While genetic variation in this species appeared to be relatively low, detectable levels of variation were found within and among populations using AFLP data analysis. The nature of the data did not permit construction of chronologically-arranged lineages and determining relative bald and outcrop ages was not possible. Nonetheless, this analytical technique revealed several discontinuities across the ranges of the populations sampled.

GEOGRAPHIC POPULATION STRUCTURING

The low r^2-value produced from the Isolation by Distance Mantel Tests suggested that the distribution patterns of *S. tridentata* are not strongly geographically correlated. Several analytical techniques employed indicated shallow but detectable levels of geographical structuring of *S. tridentata* in the eastern United States. The major characteristics of these structurings were somewhat congruent across multiple analytical techniques. The most prevalent populational pattern was an east-west of the Valley and Ridge Provence discontinuity. Also prevalent across analyses was a comparatively strong affiliation between northeastern populations, a group which seemed to often associate with the SM-outcrop population in North Carolina. Strong, geographically distant,

associations, like these, possibly account for the failure of the Isolation by Distance analysis to detect strong geographical by genetic correlations.

AMOVA analysis of several regional configurations of the nine sampled populations showed shallow but detectable levels of interregional variation on a whole. The western populations DS and BC were shown to be genetically removed from the eastern populations of the Appalachian Mountains. This discontinuity accounts for the greatest degree of interregional variation detected (6.98% according to AMOVA). Furthermore, both the Principle Coordinate Analysis (PCA) and the Bayesian population structure analysis indicated the existence of this east-west discontinuity, and these populations were treated as monophyletic in the strict-consensus genetic distance tree analysis. The Valley and Ridge discontinuity is illustrated by the relative genetic distance between BSM and DS, while these populations are geographically proximate.

The presence of the northeastern affiliation (associating with SM in North Carolina) was supported in the Bayesian cluster-based population structuring analyses, though the distinction did not resolve in the strict-consensus distance dendrograms. The northeastern-population region (CE, WM and MW) accounted for a small (~2%) but significant degree of interregional variation in AMOVA, when compared to the southern Appalachian population-region.

SM is probably the least-disturbed small population sampled in this study, as there are no established trails or roads approaching the population. Similarity between SM and the northeastern populations, when compared to the relative dissimilarity between SM and more geographically proximate populations, perhaps suggests that SM

represents a more relictual lineage type. The similarities between SM and more northern populations, which are necessarily recent due to glaciation events, might suggest that the paleoflora community was migrating along the lagging end of glacier retreat, and colonizing newly available habitat from this source population. This history of population colonization has left similar patterns in the population structure of other plant species (McLachlan et al. 2005). AFLP data does not permit accurate chronological analysis of population lineages, and determining ancestral and derived populations was not possible. As variation between these populations within the southeastern United States is possibly very low, it is possible that genetic lineages will prove difficult to detect.

HABITAT BASED POPULATION STRUCTURING

Though AMOVA analysis based on habitat distinctions showed no significant inter-populational variation, there were distinctions made with various other analyses. Population structuring showed common affinity between the southern Appalachian grass bald populations (at K=7). Likewise, the genetic distance consensus tree resolved these populations as separate and sister clades. These two balds are geographically proximate (37 km), thus this proximity is possibly responsible for the observed genetic similarity, though rates and distance of gene dispersal in *S. tridentata* are unknown. The two sampled southern Appalachian grass bald populations are located on mountaintops, with lower elevation areas between these populations. Therefore the likelihood of unknown populations existing between the bald populations is considered low.

It is possible that the observed degree of similarity between the southern-bald populations is accounted for through colonization of the balds from nearby rock-outcrop

populations within the region, and subsequent gene flow between the grass bald populations. Gene flow between grass bald populations is considered more likely than gene flow from grass balds to outcrop populations, due to the large size of the grass balds, as well as greater niche availability for colonization and recruitment. Likewise, small size and lack of niche availability in rock outcrop habitats suggests that colonization of rock outcrops from external populations is unlikely, and that intrapopulation recruitment may be low. Percent-polymorphic loci per population data showed that the grass bald populations have high-degrees of variation, statistically identical to that of the SM rock outcrop population, despite the extreme size difference between the outcrop and bald populations. Further sampling of bald and outcrop populations may be required to better clarify the distinctions between populations from these habitat types in the southern Appalachian Mountains.

There was statistical evidence for interpopulational differences between southeastern habitat types using population-structure clustering analyses, with SM (outcrop) being more closely aligned with the northeastern populations, and BSM (outcrop) as a comparatively genetically-isolated population. The genetic isolation of BSM, though not resolved in the consensus tree, was supported by long-branch length in the three component trees (Fig 4). The degree of genetic variation in the BSM-outcrop population (67.50% polymorphic loci) was shown to be significantly lower than that in the SM-outcrop population (Table 6), and statistically identical to the CE and BC populations. This variation discrepancy between CE/BSM- and SM-outcrop populations may be explained by the highly-visited nature of the BSM and CE populations. The BSM

population is in close proximity to the Shenandoah Parkway, and is located on a popular outcrop with short-distance hiking trails leading to the outcrop ledge, while the CE population is located along an escarpment edge and has a popular trail leading directly through the rocky-population habitat. The SM population, on the other hand, is isolated and has no major trails leading to the population. Disturbance resulting from human trampling might account, to some degree, for observed levels of variation in visited outcrops. Lower degrees of variation within populations, if due to disturbance by visitors, poses an interesting question regarding the levels of variation observed throughout the sampled populations in this study. Questions as to what degree human mediated reduction in variation might influence observed patterns have yet to be looked into.

SUMMARY

This preliminary study has sought to determine levels of intra- and interpopulational variation in *S. tridentata* in the eastern United States. The populations located west of the Valley and Ridge Provence (DS and BC) are responsible for the largest degree of total interpopulational variation observed. Throughout the structure and F_{st} analysis, these two populations were shown to be comparatively distant from the eastern (Appalachian) populations. The varying habitats between all sampled populations, as well as the differences in glaciation history of the sites, did not significantly influence observed variation. It is likely that I have not collected from enough populations in order to answer questions of habitat differentiation, as well as effects of glacial history in *S. tridentata*. I recognize that AFLP does not allow for application of a molecular clock to relationship trees, and it is therefore unlikely that further AFLP analysis of more populations in the southeast will elucidate specific historical timelines of species history. Though fragment analysis may be the best option for further population-level studies, I nonetheless recommend efforts into identifying useful molecular regions, as the analytical utility of sequencing techniques is more robust than that of fragment analysis.

Though variation within populations was shown to be significant, patterns of this variation did not correlate with geography, size or habitat. The patterns observed across

populations of *S. tridentata* in the eastern United States were similar to patterns observed in other pant and animal species in the region (Sewell et al. 1996, Broyles 1998, Church et al. 2003, McLachlan et al. 2005). The exact genetic relationships between populations were difficult to resolve, largely due to a high degree of molecular similarity throughout the eastern United States populations. There was greater regional-level resolution than actual inter-populational resolution, and regional analysis has shown two reoccurring patterns across multiple analysis techniques:

1. This preliminary study has uncovered what is possibly a significant discontinuity between *S. tridentata* populations located along the Appalachian Mountain region and those located west of the Valley and Ridge Provence. This east-west discontinuity reflects patterns observed for some tree and other species (Scribner and Avise 1993, Parker et al. 1997, McLachlan et al. 2005), and similarly might be the result of northern glacial refugia, much closer to the glacial maxima than typical for most plant species at that time. Recolonization after ice retreat from populations near the trailing ice edge, along with rapid development of the Valley and Ridge as a geographic barrier, could have led to the development of this discontinuity, though our present understanding of ecological dynamics in the Valley and Ridge during the Quaternary is incomplete. The lower elevation of the Valley and Ridge, though a modern barrier to *S. tridentata*, likely was not restricting during glacial periods, as the climate was much cooler during this time, and there is evidence that the Valley and Ridge was a corridor for other boreal species migration during climatic oscillations. Nonetheless, the prevalence of limestone

in the Valley and Ridge may have prevented movement of *S. tridentata*, as the present populations are not known to inhabit limestone sites, and the species may be a calcifuge.

2. The three northeastern populations did associate together in the majority of analyses, incorporating a single southeastern rock outcrop-population (SM), raising questions of which southeastern populations might represent the older populations. This pattern is possibly the result of an ancient flora type which migrated northward post glaciation, and colonized newly uncovered habitat after glacial retreat from northern regions. Elements of this flora type, presumably including *S. tridentata*, might have persisted during interglacial periods on rock outcrops in the southern Appalachian Mountains, as well as outcrops of the northern Alleghany Plateau, due to reduced competition at those sites. AFLP data does not allow for the development of chronological lineage models for populations, and it is not possible to determine the relative ages of the bald and outcrop populations. Further sampling of populations in the southern Appalachian Mountains, and application of AFLP and other molecular methodologies, is required to clarify our understanding of intra-populational dynamics between bald and outcrop populations. With further study, questions regarding approximate bald age may be answerable.

The questions of this study regarding relative age of the grassy balds and rock outcrop populations of *S. tridentata* were not answerable due to restrictions in the analytical utility of AFLP data. Nonetheless, several prominent discontinuities were detected in the interpopulation structure of the species. The results of this study provided a preliminary model of population structure of *S. tridentata* in the eastern United States,

and further studies now can use this information to better structure molecular projects aimed at either this study species, or its community associate species.

REFERENCES

Al-Rabab'ah, M.A. and C.G. Williams. 2002. Population dynamis of *Pinus taeda* L. based on nuclear microsatellites. Forest Ecol. Manag. 163:263-271.

Andrade, I.M., S.J. Mayo, C. Van den Berg, M.F. Fay, M. Chester, C. Lexer, and D. Kirkup. 2007. A preliminary study of genetic variation in populations of Monstera adansonii var. klotzschiana (Araceae) from north-east Brazil, estimated with AFLP molecular markers. Ann. Bot. 100:1143-1154.

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

Baskin, J.M. and C.C. Baskin. 1988. Endemism in rock outcrop plant communities of unglaciated eastern United States: an evaluation of the roles of the edaphic, genetic and light factors. J. Biogeogr. 15:829-840.

Bennett, K.D., P.Z. Tzedakis, and J.K. Willis. 1991. Quaternary refugia of north European trees. J. Biogeogr. 18:103-115.

Bohonak, A.J. 2002. IBD (Isolation By Distance): A program for analyses of isolation by distance. J. Heredity 93:153-154.

Bonin, A., D. Ehrich, and S. Manel. 2007. Statistical analysis of amplified fragement length polymorphism data: a toolbox for molecular ecologist and evolutionists. Mol. Ecol. 16:3737-3758.

Boucias, D., C. Stokes, A. Suazo, and J. Funderburk. 2000. AFLP analysis of the entomopathogen Nomuraea rileyi. Mycologia 92:638-648.

Broyles, S.B. 1998. Postglacial migration and the loss of allozyme variation in northern populations of *Asclepias exaltata* (Asclepiadaceae). Amer. J. Bot. 85:1091-1097.

Church, S.A., J.M. Kraus, J.C. Mitchell, D.R. Church, and D.R. Taylor. 2003. Evidence formultiple Pleistocene refugia in the postglacial expansion of the Eastern Tiger Salamander *Ambystoma tigrinum tigrinum*. Evolution 57:372-383.

Clebsch, E.C. and G.L. Walker. 1988. Disjunct phanaerogams associated with *Thuja occidentalis* in its southern disjunct range. Castanea 12:54-55.

Coker, P.D. 1966. Sibbaldia procumbens L.. J. Ecol. 54:823-831.

Delcourt, P.A. and H.R. Delcourt. 1981. Vegetation maps for eastern North America: 40,000 yr b.p. to the present. *IN*: Romans, R. (ed.). Geobotany II. New York: Plenum Press. p. 123-165.

Doyle, J.J. and J.L. Doyle. 1987. Isolation of plant DNA from fresh tissue. Focus 12:13-15.

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

Falush, D., M. Stephens, and J.K. Pritchard. 2007. Inference of population structure using multilocus genotype data: dominant markers and null alleles. Mol. Ecol Notes 7:574-578.

Felsenstein, J. 2007. PHYLIP: Phylogenetic inference package, version 3-67. Distributed by the author, Department of Genome Sciences, University of Washington, Seattle Washington.

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

Hartl, D.L. and A.G. Clark. 1989. Principals of population genetics, 2nd ed. Sinauer Associates, Sunderland, Massachussets.

Hewitt, G.M. 1994. Genetic consequences of climatic oscillations in the Quaternary. Philosophical Transactions: Biological Sciences 359:183-195.

Hollingsworth, P.M. and R.A. Ennos. 2004. Neighbor joining trees, dominant markers and population genetic structure. Heredity 92:490-498.

Ikeda, H., K. Senni, N. Fujii, and H. Setoguchi. 2006. Refugia of *Potentilla matsummurae* (Rosaceae) located at high mountains in the Japanese archipelago. Mol. Ecol 15:3731-3740.

Jody, S. and A. Bruneau. 2004. Evolution of triploidy in *Apios americana* (Leguminosae) revealed by genealogical analysis of the histone H#-D gene. Evolution 58:284-295.

Kennedy, A.H. and G.L. Walker. 2007. The population genetic structure of the Showy Lady's Slipper Orchid (*Cyprepedium reginae* Walter) in its glaciated and unglaciated ranges. Castanea 72:248-261.

Lindsey, M.M. and S.P. Bratton. 1979. The vegetation of grassy balds and other high elevation disturbed areas in the Great Smoky Mountains National Park. Bull. Torrey Bot. Club 106:264-275.

Lynch, M. and B.G. Milligan. 1994. Analysis of population genetic structure with RAPD markers. Mol. Ecol 3:91-99.

Mark, A.F. 1958. The ecology of the southern Appalchian grass balds. Ecol. Monogr. 28:294-336.

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

Nei, M. and W.H. Li. 1979. Mathmatical model for studying genetic variation in term of restriction endonucleases. Proc. Natl. Acad. Sci. USA 76:5269-5273.

Parker, K.C., J.L. Hamrick, A.J. Parker, and E.A. Stacy. 1997. Allozyme diversity in *Pinus virginiana* (Pinaceae): intraspecific and interspecific comparisons. Amer. J. Bot. 84:1372-1382.

Peakall, R. and P.E. Smouse. 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol. Ecol. Notes 6:288-295.

Reynolds, J., B.S. Weir, and C.C. Cockerham. 1983. Estimation of the coancestry coefficient: basis for a short-term genetic distance. Genetics 105:767-779.

Rowe, J.S. 1956. Uses of understory plant species in forestry. Ecology 3:461-473.

Scribner, K.T. and J.C. Avise. 1993. Cytonuclear genetic architecture in mosquitofish populations. Mol. Ecol. 2:139-149.

Sewell, M.M., C.R. Parks, and M.W. Chase. 1996. Intraspecific chloroplast DNA variation and biogeography of North American *Liriodendron* L. (Magnoliaceae). Evolution 50:1147-1154.

Shaw, J., E.B. Lickey, E.E. Shilling, and R.L. Small. 2007. Comparison of whole genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. Amer. J. Bot. 94:275-288.

Soltis, D.E., A.B. Morris, J.S. McLachlan, P.S, Manos, and P.S. Soltis. 2006. Comparative phylogeography of unglaciated eastern North America. Mol. Ecol. 15:4261-4293.

Sullivan, J.H. and J.D. Pitillo. 1988. Succession of woody plants into a high elevation grassy bald of the Balsam Mountains. Castanea 53:275-288.

Urbatsch, L.E., B.G. Baldwin, and M.J. Donoghue. 2000. Phylogeny of the coneflowers and relatives (Heliantheae: Asteraceae) based on nuclear rDNA internal transcribed spacer (ITS) sequences and chloroplast DNA restriction site data. Syst. Bot. 25:539-565.

Valente, M. 2007. Clonality and genetic diversity revealed by AFLPs in *Schisandra glabra* (Brickell) Rehder (Schisandraceae), a rare basal angiosperm. M.S. Thesis, University of Tennessee, Knoxville, Tennessee.

van Ee, B.W., N. Jelinski, P.E. Berry, and A.L. Hipps. 2006. Phylogeny and biogeography of *Croton alabamensis* (Euphorbiaceae), a rare shrub from Texas and Alabama, using DNA sequence and AFLP data. Mol. Ecol. 15:2735-2751.

Vekemans, X., T. Beauwens, M. Lemaire, and I. Roldan-Ruiz. 2002. Data from amplified fragment length polymorphism (AFLP) markers show indication of size homoplasy and of relationships between degree of homoplasy and fragment size. Mol. Ecol. 11:139-151.

Vos, P., R. Hogers, M. Bleeker, M. Reijans, T. Van de Lee, M. Hornes, *et al.* 1995. AFLP: a new technique for DNA fingerprinting. Nucleic Acids Research 23:4407-4414.

Wang, Z., A.J. Baker, G.E. Hill, and S.V. Edwards. 2003. Reconciling actual and infered population histories in the house finch (Carpodacus mexicanus) by AFLP analysis. Evolution 57:2852-2864.

Weakley, A.S. 2007. Flora of the Carolinas, Virginia, Georgia, and surrounding areas (working draft). University of North Carolina, Chapel Hill, North Carolina.

Weigl, P.D. and T.W. Knowles. 1999. Antiquity of southern Appalachian grass balds: the role of keystone megaherbivores. p. 215-223. *In:* Eckerlin, R.P. (ed.). Proceedings of the Appalachian Biogeography Symposium. Virginia Museum of Natural History, Martinsville, Virginia.

Weir, B.S. 1990. Genetic data analysis: methods for discrete population analysis. Sinauer Associates, Sunderland, Massachusetts.

Wells, B.W. 1956. Origin of southern Appalachian grass balds. Ecology 37:592.

Wiser, S.K. 1998. Comparison of southern Appalachian high-elevation outcrop plant communities with their northern Appalachian counterparts. J. of Biogeogr. 25:501-513.

Wiser, S.K., R.K. Peet, and P.S. White. 1998. Prediction of rare plant occurrence: a southern Appalachian example. Ecol. Applications 8:909-920.

Wright, S. 1951. The genetical structure of populations. Ann. Eugenics 15:323-354.

APPENDIX A
AFLP PROTOCOL

APENDIX A
AFLP PROTOCOL

Adapted from Vos et al. (1995).

1. It is important that the DNA extracts be accessible for enzymatic reactions. The presence of polysaccharides will hinder this interaction if concentration is high enough. In the past, PCR has been achieved through a dilution of the extract (1:1000 - 1:3000), along with the use of Ex taq polymerase (Takara Inc.). The resulting PCR products have shown ample amplification. As one of the strengths of AFLP is ability to work with small amounts of DNA, the dilution method might be a suitable option. If the contamination becomes an issue, or if dilution concentrations become too low, there is the second clean-up option of running the extracts through a second round of the CTAB extraction, excluding the actual CTAB itself. This method effectively treats the extract with B-merc and PVP, which will help in the reduction of protein and polysaccharide contaminates. When this second clean up has utilized, the resulting extract is much less viscous, though the concentration of DNA is somewhat reduced.

2. Adapter preparation.

ADAPTERS

Eco-F:	5' - CTC GTA GAC TGC GTA CG - 3'
Eco-R:	5' - AAT TGG TAC GCA GTC TAC - 3'
Mse-F:	5' - GAC GAT GAG TCC TGA G - 3'
Mse-R:	5' - TAC TCA GGA CTC AT - 3'

Bring adapter concentrations to 100µM

a. for EcoRI adapter pair mix:

25 µl	Eco-F
25 µl	Eco-R
450 µl	TE Buffer
Total: 500 µl	(final conc. 5µM)

b. for MseI adapter pair mix:

250 µl	Mse-F
250 µl	Mse-R
Total: 500 µl	(final conc. 50 µM)

After mixing the adapters, heat at 95° C for 5 min to denature. Then allow to cool slowly (in Styrofoam) to room temperature. Make aliquots. Store at -20° C.

3. Primers. *

Two sets of primers will be used. The first round of AFLP PCR will incorporate the +1 primer set, which is pre-selective. The second round will use the +3 primer set, which is selective.

The +1 set will have A's in the +1 location. The +3 sets will have A's in the first spot, and then a combination of nucleotides in the remaining spots, which gibe a combined total of three G's or C's on both primers. **Note: the 3+ primers must be appropriately labeled (florescent)**

Mse+3:	5' - GAT GAG TCC TGA GTA ANN N -3'
Eco+3:	5' - GAC TGC GTA CCA ATT CNN N -3'

* For this projects AFLP analysis of *Sibbaldiopsis tridentata* (Ait.) Rydb., three pairs of selective
(+3) primers were used. These were designed to associate with the above primers:

Pair 1:	5'- TATCTGCGTACCAATTCAGC -3'
	5'- GACGATGAGTCCTGAGTAACTG -3'
Pair 2:	5'- TATCTGCGTACCAATTCACA -3'
	5'- GACGATGAGTCCTGAGTAACGG -3'
Pair 3:	5'- TATCTGCGTACCAATTCACC -3'
	5'- GACGATGAGTCCTGAGTAAC -3'

4. Digestion of genomic DNA.

In accordance with virtually all the literature I've read, I will use EcoRI and MseI restriction enzymes for the digestion. Protocols of Vos et al. 1995 will be followed for reaction concentrations of DNA and enzymes. The universal buffers provided with the enzymes (Promega) will be used instead of the labor intensive reaction buffer of Vos et al 1995. Reaction volumes will be 50µl.

5. Ligation of adapters.

T4 ligase will be used to connect the adapters to the digested DNA. T4 ligase contains ATP (which is unstable), and therefore should be made into aliquots, and kept cold at all times.

This is a two step process where the enzyme master mix is made, and then the ligation mix is made.

Enzyme Master Mix:

T4 Buffer, 10X	0.1 µl
NaCl, 0.5M	0.1 µl
BSA @ 1mg/ml	0.05 µl
MseI, 1U	0.1 µl (usually, but check conc of stock)
EcoRI, 5U	0.12 µl (usually, but check conc of stock)
T4 ligase, 1U	0.2 µl
DI H20	0.33 µl
Total:	1.0 µl

Ligation Reaction:

T4 Buffer, 10X	1.0 µl

NaCl, 0.5M	1.0 µl
BSA @ 1mg/ml	0.5 µl
Enzyme Master Mix	1.0 µl
Mse Adapter	1.0 µl
EcoRI Adapter	1.0 µl
Total:	5.5 µl
Digested DNA	5.5 µl (at about 10 ng/ul)

Mix well. Centrifuge just enough to condense rxn. at bottom of tube. Incubate 37°C for 2 hrs. Store at 4°C.
Add 90 µl $TE_{0.1}$ (EDTA in TE at conc. of 0.1M, make sure).

6. +1 PCR Reactions, Pre-selective AFLP

Because the T4 ligase only ligates one strand of the adapters to the sticky ends of the digested DNA, the other strand is hanging on due only to base pairing between the two strands. Therefore it is necessary to have an initial 72°C hold in order to allow the Taq polymerase to ligate the second strand to the sticky end. If this hold is omitted, the reaction will loose the second strand, and the AFLP PCR will fail. DO NOT DO A HOT START. Do not put the reaction into a hot thermocycler either, you knucklehead.

The Ex taq protocol (Takara) will be used for concentrations of reagents. The reaction volumes will be 25 µl.

The following PCR parameters will be used: (30 cycles)

initial hold	72°C for 2 min
denature	94°C for 30 sec
annealing	56°C for 30 sec
extension	72°C for 2 min
final ext	60°C for 10 min
hold	4°C ∞

Check for reaction success on 1% agarose gel. Should see smear in the 100 to 1000 bp range. Sometimes bands might be visible in the smear. Add 100 µl $TE_{0.1}$ to the remaining reaction

7. +3 PCR Reactions, Selective AFLP

This reaction used the 1+ PCR products as template. The 3+ primers are the same sequence except they have two additional bases, which increase the selectivity. The 3+EcoRI primers are florescent labeled.
Initially multiple primer pairs need to be tested to identify which pairs provide the best and most readable bands. If there are multiple EcoRI primes that work well with a single MseI primer, then it is possible to label the EcoRI primers with separate dyes, and do what is called a multiplex reaction. Then you can verify that they produce the same bands as when they're run individually. But of course, multiplexing is not necessary.
Again, the Ex taq (Takara) protocol will be followed for the mixing of reagents.

The following PCR parameters will be used for the +3 reactions:

Initial:	94°C for 2 min
Denature	94°C for 30 sec
Annealing	65°C for 30 sec (reduce by 0.7°C per cycle)
Extension	72°C for 2 min
(This cycle runs 13 times)	
Denature	94°C for 30 sec
Annealing	56°C for 30 sec
Extension	72°C for 2 min
(This cycle runs 24 times)	
Final ext	72°C for 10 min
Hold	4°C ∞

8. Fragment Analysis

Fragment analysis will be done through the ABI sequencer at Cornell Core Laboratories.

APPENDIX B

Explanation of the process of transforming Amplified Fragment Length Polymorphism (AFLP) raw data into a binary dataset

APPENDIX B

Explanation of the process of transforming Amplified Fragment Length Polymorphism (AFLP) raw data into a binary dataset

After completion of the AFLP laboratory protocol of Vos et. al. (1995), fragment analysis is done on an appropriate analyzer. This requires the inclusion of an internal size standard as well as the soup of various fragments of interest. The output file is then opened in a suitable fragment analysis software package (ex. Genographer, Applied Biosystems).

Once the appropriate interpretation parameters are set (size standard type and sizes, primers present or not) the software will produce a series of tables for each sample run. These tables include the fluorescent dye detected (peak color), the size of the peaks (in bp) and the height of the peaks (unit-less number). Each color represents a separate set of AFLP primers used. This way it is possible to include up to four separate primer pairs in a single AFLP fragment analysis run. This stacking of primer pairs is called "nested" AFLP.

The resulting tables are then copy and pasted into Excel or Quatro. The table below depicts a peak height and size table for three individuals (GB145, GB148 and GB149) from the Beaver Camp Rd. population in Wisconsin. This is a sample from AFLP of the primers labeled "A" and "I," which are tagged with the 6FAM blue fluorescent dye.

Each individual has two columns associated with it. The leftmost column is the size in base pairs of the occurrence of a peak in the fragment analysis, and in the column to the right of that is the associated size of the peak. For example, GB145's first peak is given as being a fragment 35.4132 bp in size, with a height of 78.

Beaver Camp Rd. WI Population
Primers A & I (6FAM) Blue

Size	**Height**				
GB145		GB148		GB149	
35.4132	78	35.6141	178	35.807	132
38.8955	162	38.4594	65	38.412	84
41.5228	2015	42.5667	5385	41.681	2233
42.5252	2260	43.4591	1442	42.662	2488
43.4236	716	45.5285	2473	43.543	831
45.4071	1645	46.8007	4168	45.488	2357
46.687	1971	48.5511	557	46.841	3006
48.6418	323	50.5593	388	48.76	372
50.4667	203	51.4902	2332	50.552	216
51.3989	1013	52.4196	4858	51.471	988
52.4224	2318	53.3474	2816	52.48	2345
53.3511	1211	54.4587	342	53.396	1319

54.3708	171	55.4755	1007	54.493	157
55.4809	575	56.4905	1313	55.497	609
56.4043	560	57.5957	236	56.5	644
57.6022	74	58.515	189	57.592	113
58.43	83	61.9038	443	58.5	96
60.4482	120	62.8162	420	61.851	246
61.8199	240	65.1821	280	62.844	231
62.7325	217	66.6333	72	65.184	135
65.1891	131	67.4481	292	66.531	51
66.5492	59	69.1645	11628	67.427	168
67.4541	242	70.1559	21915	69.127	7824
69.0794	7842	71.5053	184	70.109	11472
70.0704	11870	72.2238	53	72.248	63
72.1369	76	73.2103	319	73.137	194
73.2122	194	78.2687	589	78.255	390
78.1842	343	79.2957	705	79.277	392
79.2143	310	80.6027	198	80.672	129
80.5253	160	82.4695	2935	82.438	1701
82.3045	1842	83.4961	5700	83.461	2675
83.3345	2857	84.8958	523	84.855	204
84.8326	212	87.4147	9093	87.364	7312
87.267	6967	88.4407	18217	88.386	11870
88.297	11384	89.9329	139	92.382	991
92.3229	997	92.3573	1980	94.705	497
94.5698	563	94.5948	874	95.727	792
95.5997	929	95.6201	1580	96.841	63
96.7232	60	96.8317	87	98.421	83
98.2212	85	98.3228	167	99.536	112
99.4383	169	99.441	311	100.61	133
100.527	178	100.523	373	104.44	54
105.991	60	102.883	64	106.1	79
106.877	2956	104.374	128	106.98	3294
107.853	3911	106.925	4985	107.94	4531
109.63	311	107.895	8482	109.79	317
110.699	368	109.663	483	110.76	404
111.77	451	110.638	814	111.81	415
112.753	780	111.792	535	112.87	739
116.702	373	112.77	1180	116.86	341
117.693	590	116.788	416	117.84	558
120.223	54	117.775	894	121.59	138
121.492	153	120.293	101	122.57	128
122.492	142	121.466	321	125.44	110
125.408	127	122.461	341	126.44	186
126.414	189	125.365	161	128.33	95
128.247	89	126.366	287	129.33	108
129.258	105	128.283	144	132.14	81
132.116	89	129.29	194	133.14	86
133.041	86	131.954	108	138.36	148

138.345	172	133.06	141	139.4	202
139.412	229	136.858	245	141.71	1507
141.664	1434	138.347	273	142.81	2424
142.885	2286	139.412	364	144.59	54
146.714	508	141.765	1876	146.67	399
147.814	899	142.884	3887	147.85	775
149.107	140	145.409	220	149.12	129
150.368	84	146.713	921	150.37	85
151.377	2043	147.912	1415	151.37	1993
152.379	2536	149.107	247	152.37	2649
153.285	58	150.277	114	155.96	111
155.97	136	151.469	2646	168.81	79
168.714	90	152.471	4399	169.76	114
169.681	124	156.06	180	179.05	955
178.903	1025	168.758	150	180	1872
179.956	1988	169.72	219	181.39	344
181.271	357	178.974	1400	182.34	441
182.323	441	180.019	3083	187.01	144
186.968	163	181.413	534	188.04	203
188.019	216	182.371	789	189.33	93
189.332	102	186.896	260	192.18	56
193.707	128	188.027	376	193.81	121
196.243	1485	189.33	216	196.31	1357
197.204	1951	192.196	113	197.25	1818
214.49	106	193.064	201	214.52	109
215.474	119	194.626	53	215.58	119
217.53	211	197.227	2765	217.62	188
218.601	410	198.787	91	218.68	367
219.939	73	204.95	71	219.91	55
220.919	58	212.564	73	222.65	310
222.521	374	214.438	178	223.61	547
223.588	684	215.508	188	224.67	274
224.655	312	217.556	314	225.72	384
225.632	415	218.623	639	248.88	65
248.866	83	219.956	130	250	76
249.826	91	220.844	72	252.81	75
252.696	76	222.529	469	253.7	139
253.675	155	223.593	935	257.55	89
257.51	75	224.655	417	258.59	181
258.488	153	225.628	570	263.24	60
263.13	74	243.205	66	264.2	137
264.188	143	248.87	148	281.69	50
281.805	59	249.826	173	282.8	110
282.777	124	252.677	93	289.97	535
289.98	569	253.65	203	290.93	830
290.95	892	257.54	138	296.18	117
296.285	134	258.511	276	297.14	306
297.173	346	263.203	115	315.91	68

315.941	85	264.173	228	340.84	52
340.858	52	267.567	56	356.6	73
356.453	92	276.601	52	357.57	201
357.514	246	281.836	77	358.37	84
358.33	80	282.802	177	417.21	109
417.275	115	289.96	765	460.84	58
460.825	67	290.924	1357	470.82	63
470.848	55	292.612	58		
		296.146	149		
		297.19	425		
		309.858	67		
		315.01	52		
		316.041	97		
		340.844	80		
		356.474	113		
		357.445	334		
		358.255	132		
		384.696	56		
		417.269	122		
		460.845	77		
		470.821	71		
		481.864	54		

Raw peak height and size data such as this must be "scored" into a series of ones and zeros (presences and absences). In order to do this "bins" must be created. Bins are the name for the base pair sizes which you are going to mark present of absent for. Each bin has two components:

1. The actual base pair size which defines the location of the bin. (ex. 55.5 bp)
2. The boundaries of the bin, which define where the "line" is that determines what actual peaks are included in the bin.

So if, for example, I have defined a bin at the size of 55.5 bp, and with boundries 0.5 bp to either side of the bin size, the I've created a bin which is in effect 1.0 bp in size. If the sample I'm scoring has a fragment which is 55.878 bp in size, then I would score a "1" for that sample in that bin. If there was no peak at that size I would score a "0."

Scoring also requires the judicial calling of peaks relative to their height. The primary purpose of this is to eliminate the peaks which are noisy, or to keep from calling background noise in the sample a peak. This is done by picking a height that seems to be clearly out of the noise, but not excluding of too many peaks. In this study I chose a peak cutoff height of 100, and excluded any peaks which were consistently under this height.

When the binning and scoring is done, you have a dataset which looks like this:

	57.5	58.5	60.5	61.5	62.7	63.7	64.7	65.4	66.4	67.5	69.1	74.9	76.6	80.7
GB145	1	1	1	1	1	0	1	0	1	1	1	0	0	1
GB148	1	1	0	1	1	0	1	0	1	1	1	0	0	1
GB149	1	1	0	1	1	0	1	0	1	1	1	0	0	1

	83	84.5	86.4	90.1	96.4	97.4	98.4	99.4	101	103	105	106	114	120
GB145	1	1	0	0	1	0	1	1	1	0	0	1	0	0
GB148	1	1	0	0	1	0	1	1	1	1	1	0	0	0
GB14	1	1	0	0	1	0	1	1	1	0	1	0	0	0

	121	132	133	135	136	137	138	141	145	146	147	148	149	150
GB145	1	1	1	0	0	0	1	0	0	0	1	1	0	1
GB148	1	1	1	0	1	0	1	0	0	1	1	1	0	1
GB149	0	1	1	0	0	0	1	0	1	0	1	1	0	1

	153	157	164	165	167	169	170	172	174	187	188	189	191	193
GB145	1	0	0	0	0	1	1	0	0	1	1	1	0	0
GB148	1	0	0	0	0	1	1	0	0	1	1	1	0	1
GB149	1	0	0	0	0	1	1	0	0	1	1	1	0	1

This table is a complete binary dataset for three individuals (GB145, GB148 and GB149) for the primer pair A & I, along with the base pair bin markers associated with the scoring. For example, the bin marked at 57.5 bp has a peak located within its boundaries (57.0-57.9 bp) for all three individuals scored. The bin marked at 60.5 has a peak located within its boundaries (60.0-60.9 bp) for individual GB145, but the peak is absent from this bin for individuals GB148 and GB149.

The binary dataset can then be arranged into formats suitable for use in statistical programs such as GenAlEx and AFLP_SURV.

BIBLIOGRAHICAL INFORMATION

Gerald Edward Bresowar was born in Baton Rouge, Louisiana, in 1977, one of five children of Gerald and Sylvia Bresowar. He was raised in the Birmingham, Alabama suburb of Homewood, and graduated from Homewood High School in 1995. In December 2005 he graduated with a B.S. in Plant Science from the University of Tennessee, Knoxville. He entered Appalachian State University in January 2006, seeking a M.S. in Biology, which was awarded in December 2008. He is married and, at the time of his thesis submission, is living in Fort Collins, Colorado.