University of North Carolina Asheville
Journal of Undergraduate Research
Asheville, North Carolina
May 2021

Characterizing Population Genetic Markers for Sochan (*Rudbeckia laciniata* L., Asteraceae) to Ensure Sustainable Harvest within the Great Smoky Mountains National Park

Noah J. Rice
Biology
The University of North Carolina Asheville
One University Heights
Asheville, North Carolina, 28804, USA

Faculty Advisor: Dr. Jennifer Rhode Ward

Abstract

The Great Smoky Mountains National Park (GSMNP) is home to over 1,600 species of plants. Among these species is the cutleaf coneflower (*Rudbeckia laciniata* L. Asteraceae). The Eastern Band of Cherokee Indians (EBCI) have harvested cutleaf coneflower from the GSMNP for thousands of years. The EBCI refer to this native plant as sochan, a species which holds both nutritional and cultural significance to the EBCI and is typically harvested in spring. Throughout the almost 100-year history of the GSMNP, multiple rules and regulations pertaining to native plant collection have been administered. In 2016, a regulation was passed which allows the gathering and removal of plants or plant parts by members of Native American tribes for traditional purposes at specific locations within the park. Concerns about genetic diversity of GSMNP sochan populations have been raised, and it is not yet clear which populations and sites are best suited for harvest by the EBCI. The genetic status of these populations has yet to be characterized, and in fact, no population genetic markers have been developed for this species. The purpose of this research was to determine the best markers to use for measuring genetic diversity in *R. laciniata* populations and to use these markers to inform sustainable harvest practices within the GSMNP. Samples of sochan leaves were collected from sites SF, FC, and TP; within the GSMNP. Thirty leaves were taken per population and DNA was extracted from a total of 29 plants across the three populations. A total of 15 microsatellite loci were screened using polymerase chain reaction (PCR) and then gel electrophoresis. Five of these loci were used to assess the 29 samples and successfully amplified PCR products were sent for fragment analysis. Allelic diversity (A) and Lynch genetic distance was calculated, and the site with the greatest diversity was SF. The site with the least diversity based on A and Lynch genetic distance was TP. A principal component analysis comparing SF and FC showed much overlap between these two populations, indicating genetic similarity. However, the F_{ST} value between these two populations fell in the moderate genetic differentiation range. It is recommended that in order to conserve as much genetic diversity as possible, the SF population be protected from future harvests. The population at TP shows the least genetic diversity, and it is recommended to continue harvests at this site. Future research to inform sustainable harvesting of sochan can expand on this work by focusing on geospatial analysis of the data to clearly display patterns of genetic diversity within the GSMNP.

1. Introduction

Humans have been harvesting non-timber forest products such as seeds, flowers, fruits, leaves, and roots for thousands of years[1]. These products contribute to the cultural and social traditions of many indigenous communities and are essential to their livelihood and wellbeing[2]. In recent decades, plant harvests have increased in volume due to a variety of cultural and commercial reasons which is raising concerns about overexploitation of many plant species[1]. Harvesting techniques vary widely, but harvest studies indicate some methods are more sustainable than others. For

example, the Eastern Band of Cherokee Indians (EBCI) harvest ramps (wild leeks) by cutting the petiole and leaving a part of the bulb and rhizomes in the ground which has been shown to have no effect on plant mortality[2]. Other less-sustainable techniques involve complete removal of adult plants or roots, which could curtail reproduction or be fatal; it may take populations years to recover from this[2].

Plant over-harvesting causes alterations in their survival, growth, and reproductive rates, which affects the structure and dynamics of entire populations[1]. Declines in reproductively mature individuals, loss of genetic diversity, and overall population decline are major concerns stemming from uncontrolled harvests[2]. Without management, over-harvesting could result in the extinction of localized gene pools and affect the natural genetic makeup of a region by selective removal of genotypes[3]. The loss of genetic diversity increases the frequency of deleterious alleles due to inbreeding and genetic drift, which in turn causes a loss of adaptive potential and a greater risk of extinction[4]. In contrast, with greater genetic diversity and increasing genetic variation among individuals, the likelihood of inbreeding decreases, and the adaptive potential of plant species is increased[4]. Thus, monitoring genetic diversity is crucial to understanding the population health of a plant species, and can be used to inform sustainable harvest practices.

Over the past 20 years, one of the most common markers used to genotype plants has been microsatellites[5]. Microsatellites, also called single sequence repeats (SSRs), are 1-6 base pair repeats found in the genomes of all prokaryotes and eukaryotes[6]. Microsatellites are abundant in the euchromatin of eukaryotes, especially in intergenic regions, and are very useful for studying diversity in wild plant species due to high mutation rates (10^{-7} to 10^{-3} mutations per locus per generation). This allows for distinctions to be made among related populations and individuals[5,6]. Differences in the microsatellite regions of DNA, polymorphisms, can be observed using polymerase chain reaction (PCR) and then running the product through gel electrophoresis. Microsatellite loci are useful tools for estimating genetic variability and can be used to analyze natural populations of both plants and animals[11, 13]. Modern methods of characterizing microsatellites provide an opportunity to monitor the dynamics of diversity in populations over time and can be employed to monitor genetic diversity and ultimately inform conservation strategies[12,13].

The Great Smoky Mountains National Park (GSMNP) is known for its biodiversity with an estimated 100,000 plants and animals living in the park[7]. The cutleaf coneflower (*Rudbeckia laciniata* L.), a native plant species found in the GSMNP, is a member of the Asteraceae family. *R. laciniata* is a perennial herb that is typically found in thickets or along streams and can grow up to three meters high. *R. laciniata* is endemic to North America and features yellow daisy flowers with light green leaves. The EBCI refer to *R. laciniata* as sochan and have been using this plant for food and ceremonial purposes for thousands of years[8]. The EBCI typically harvest the young greens of the sochan plant for food in early spring, and the flowers for ceremonial purposes in the summer. The shoots are one of the earliest available spring wild foods and they are a valuable source of nutrients. In recent years, concerns have been raised about the possibility of over-harvesting wild sochan populations within the GSMNP. In the 1980s, the National Park Service (NPS) created regulations on harvesting and otherwise disturbing plants within the National Park System, in an effort to sustain native plant species[9]. Then, in 2016, the NPS issued a ruling that allows the gathering and removal of plants or plant parts by members of Native American tribes for traditional purposes at specific locations[9]. Sochan is of interest to both the GSMNP and the EBCI, due to the shared goal of preserving this species and maintaining sustainable harvests. Sustainable harvests will preserve the plant's genetic diversity, while preserving populations in the park. It is also extremely important that the EBCI can continue their traditions and ceremonies which have taken place for thousands of years.

The purpose of this research was to determine the best genetic markers to use for *R. laciniata* populations and to use these markers to characterize genetic diversity in three populations. The best genetic markers are loci that amplify consistently and are polymorphic. This information can ultimately be used to provide recommendations on how to sustainably harvest sochan, and which sites are most appropriate for harvest. Providing these data to the GSMNP staff will hopefully allow the EBCI to continue their traditional harvest of sochan, and also maintain as much genetic diversity as possible.

2. Methods

Potential microsatellites were developed for *Rudbeckia laciniata* in 2018. Deoxyribonucleic acid (DNA) from a representative *R. laciniata* individual was extracted using the modified CTAB method (Doyle and Doyle 1987) and mailed to West Virginia University (WVU) for Illumina sequencing (Illumina 2018). Electronic data from the WVU core facility were sent to Dr. Jennifer Rhode Ward and Dr. Matt Estep (Appalachian States University). Microsatellite regions were identified by Dr. Estep and research students using msatcommander[10] and low-quality data points were removed. Using sequences generated by msatcommander, primer pairs were ordered from Eurofins Genomics.

Samples of *R. laciniata* leaves were obtained from SF, FC, and TP in the GSMNP. DNA was extracted for 10 plants from TP and FC populations, and nine plants for the SF population. Tissue samples were ground with a mortar and pestle and autoclaved sand in a 1.5 ml microfuge tube. The DNA was extracted using a modified CTAB method developed by Doyle and Doyle (1987) and the quality and concentration of each extract was determined using a Nano Drop ND-1000™.

A total of 15 microsatellite loci were screened and five were selected to test against the GSMNP populations so that there was a mix of different repeat patterns (dinucleotide, trinucleotide, and tetranucleotide) (Table 1). The locus ID numbers of these five were 2742773, 6256, 6988, 2506889, and 2028163 (Table 2). Annealing temperatures (T_A) for forward and reverse primers were similar and were determined previously based on GC content predictions (Table 2). The forward and reverse sequence and expected product size for each primer is listed below (Table 2).

Polymerase chain reaction (PCR) was performed; each reaction contained 12.5 µL of 2X GoTaq® Green Master Mix (DNA polymerase, dNTPs, cofactors, and reaction buffer), 0.5 µL each of 10 µM forward and reverse primers, and 5.5 µL of PCR water. Six µL of DNA was added, bringing the total reaction volume to 25 µL. DNA was amplified using a T100™ Thermocycler with a 2 minute denature at 94ºC followed by 35 cycles of: 40 s at 94 ºC, 40 s at T_A, and 1 minute at 72 ºC. Samples were held at 12ºC until being removed from the thermocycler. The success of DNA amplification was determined by running PCR products through a 1% agarose gel in TAE buffer (40 mM Tris, 20 mM acetic acid, 1 mM EDTA).

Gels were stained in ethidium bromide and imaged using a BioDoc-It©. Successfully amplified PCR products were sent in a 96 well plate to the North Carolina State University Genomic Sciences Laboratory for fragment analysis on an Applied Biosystems™ 3730xl DNA Analyzer. Each well contained a mixture of 4 µL of PCR product, 0.5 µL of GeneScan™ 500 Liz ladder, and 5.5 µL of Hi-Di™, for a total volume of 10 µL. The fragment files were analyzed using Geneious® 2020 bioinformatics software. The microsatellite plug-in (version 1.4.7, 2021) was used in Geneious® to analyze peaks. Peaks below 1000 fluorescent units were ignored to eliminate background noise, and bins for each locus were predicted based on the type of nucleotide repeat that was being analyzed (dinucleotide, trinucleotide, or tetranucleotide) and expected size. The locus information was set in Geneious® and the software predicted the allele patterns for each locus. Alleles tables were then downloaded and exported for analysis in R (R version 3.6.2, 2019). The R package *polysat* (Clark et al., 2019) was used to calculate allelic diversity (A), fixation index (F_{ST}), Lynch genetic distance, and also to build a principal component analysis (PCA) plot.

Table 1. Characteristics of loci screened.

Repeat Pattern	Locus ID	Microsatellite
Dinucleotide	6256, 6988, 712442, 7389, 494083	AT, AT, AT, AT, AT
Trinucleotide	11765, 2381880, 914246, 1157684, 1373052, 819404, 2506889	AAG, AAG, ATC, AAT, AAC, ATC, AAC
Tetranucleotide	226501, 2742773, 2028163	AAAG, ACAT, ACAT

Table 2. Locus ID, forward sequence of primer, reverse sequence of primer, annealing temperature (ºC) of primer, and expected product size in base pairs (bp).

Locus ID	Forward Sequence	Reverse Sequence	Annealing Temperature (ºC)	Expected Product Size (bp)
6256	AGGGATGATACGATGATTCCC	GTTTCCCGTACAATCCTTATCCCG	59	164
6988	GTTTGTTTGTGGGTTCGACAAGG	TCGGCCAACATCCCTACATC	58	154
2028163	GTTTGGTTGGCAAGCTCATGG	ACCACAACGATGGAGAAGGG	60	217
2506889	TCCTGGTTACACGGTACTGC	GTTTTCCTACACGACGCTCTTC	60	175
2742773	GTTTCGTTCGATTCAGTGAGTGTC	ATTGCTTGACGGTCGATTCG	58	187

3. Results

All loci selected exhibited polymorphism and the loci with the greatest number of alleles per locus were 6256 and 6988 (Table 3). The population showing the greatest allelic diversity was SF and the population showing the least allelic diversity was TP (Table 4). F_{ST} between the SF and FC populations was 0.186. F_{ST} could not be calculated with TP population because there were too many missing data points. The PCA plot did not include TP for the same reason, and the plot showed almost no clustering, well spread-out data points, and the populations were not separated by either axis (Figure 1). Lynch genetic distance data for the SF population showed six pairs of plants with no genetic difference, and 17 pairs with genetic distance of 0.5 or greater within that population (Table 5). The FC population had only one pair of plants with no genetic difference, and seven pairs with a genetic distance of 0.5 or greater (Table 6). The TP population had 10 pairs of plants with no genetic distance, and 14 pairs with values of 0.5 or greater (Table 7).

Table 3. Description of selected loci.

Locus ID	Polymorphic (Y/N)	Alleles Per locus	Successful (Y/N)
6256	Y	7	Y
6988	Y	7	Y
2506889	Y	2	Y
2742773	Y	3	Y
2028163	Y	2	Y

Table 4. Population name, number of samples screened (N), number of microsatellite loci screened, and overall allelic diversity (A) in *R. laciniata.*

Population	N	Loci	A
SF	9	5	19
FC	10	5	14
TP	10	5	8

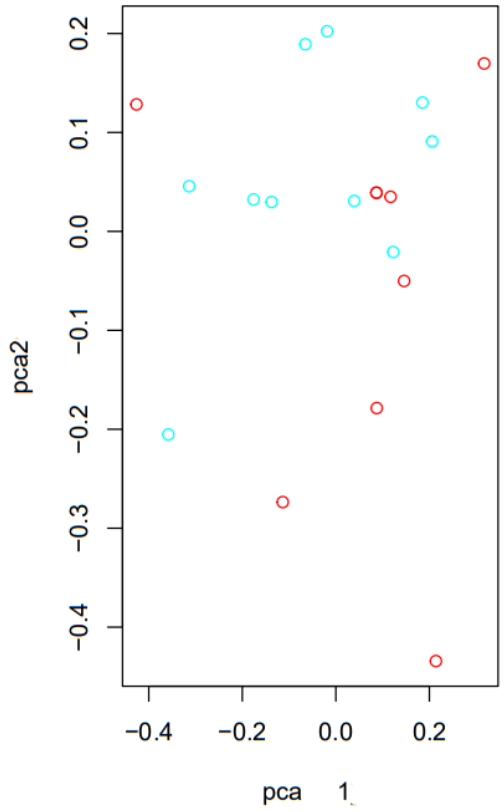

Figure 1. Principal component analysis of SF (red) and FC (light blue) populations.
Note that the populations are not separated by either axis, suggesting overlap (and thus genetic similarity).

Table 5. Lynch genetic distance values for plants in the SF population.

Population	SF 06	SF 09	SF 19	SF 20	SF 21	SF 23	SF 24	SF 28	SF 32
SF 06	0.0000								
SF 09	0.0000	0.0000							
SF 19	0.3333	0.3333	0.0000						
SF 20	0.6000	0.6000	0.6667	0.0000					
SF 21	0.5000	0.5000	0.3333	0.8000	0.0000				
SF 23	0.1667	0.1667	0.3333	0.6000	0.5000	0.0000			
SF 24	0.1667	0.1667	0.3333	0.6000	0.5000	0.5000	0.0000		
SF 28	0.0000	0.0000	0.3333	0.6000	0.0000	0.5000	0.5000	0.0000	
SF 32	0.0000	0.0000	0.3333	0.8000	0.6667	0.4444	0.2222	0.5000	0.0000

Table 6. Lynch genetic distance values for plants in the FC population.

Population	FC 02	FC 04	FC 07	FC 13	FC 14	FC 23	FC 24	FC 28	FC 32	FC 33
FC 02	0.0000									
FC 04	0.1111	0.0000								
FC 07	0.3889	0.2778	0.0000							
FC 13	0.2778	0.1667	0.1111	0.0000						
FC 14	0.4444	0.3333	0.4444	0.3333	0.0000					
FC 23	0.4444	0.3333	0.2778	0.1667	0.0833	0.0000				
FC 24	0.1111	0.0000	0.2778	0.1667	0.3333	0.2500	0.0000			
FC 28	0.6111	0.5556	0.3889	0.3333	0.6667	0.4000	0.4167	0.0000		
FC 32	0.4444	0.3333	0.2222	0.1111	0.2500	0.0667	0.2500	0.3333	0.0000	
FC 33	0.6667	0.6667	0.5556	0.4444	0.5000	0.2667	0.5000	0.4667	0.2000	0.0000

Table 7. Lynch genetic distance values for plants in the TP population.

Population	TP 10	TP 13	TP 16	TP 20	TP 23	TP 26	TP 30	TP 31	TP 33	TP 35
TP 10	0.0000									
TP 13	0.1667	0.0000								
TP 16	0.4444	0.6667	0.0000							
TP 20	0.1667	0.3333	0.1667	0.0000						
TP 23	0.1667	0.3333	0.1667	0.0000	0.0000					
TP 26	0.0000	0.0000	1.0000	0.3333	0.3333	0.0000				
TP 30	0.1111	0.0000	0.6667	0.3333	0.3333	0.0000	0.0000			
TP 31	0.0000	0.1667	0.5000	0.1667	0.1667	0.0000	0.1667	0.0000		
TP 33	0.5000	0.6667	0.5000	0.5000	0.5000	1.0000	0.6667	0.5000	0.0000	
TP 35	0.0000	0.1667	0.5000	0.1667	0.1667	0.0000	0.1667	0.0000	0.5000	0.0000

4. Discussion

This research determined the best genetic markers to use for *R. laciniata* are loci 6256 (dinucleotide), 6988 (dinucleotide), 2506889 (trinucleotide), 2742773 (tetranucleotide), and 2028163 (tetranucleotide). These loci were all found to be polymorphic and amplified consistently across samples. Furthermore, allelic diversity data and Lynch genetic distance values indicated that the SF population has the most genetic diversity and may need to be protected from future harvests in order to preserve as much genetic variation as possible. An area of high genetic diversity, such as SF, should be targeted for conservation in order to foster future adaptive responses which may be beneficial to the species, even though it is not at risk of extinction[11]. However, the population density of *R. laciniata* within the GSMNP, is lower than density outside the park. Genetic variation that will be important for survival in the future is not currently known, so it is best to conserve as much variation as possible[11]. The allelic diversity calculations and Lynch genetic distance values also indicate that the population at TP appears to be the least genetically diverse and thus may be the most appropriate site for harvest by the EBCI.

However, it is worth noting that no simple general relationship between genetic diversity and risk of species extinction exist, and a comprehensive understanding of functional genetic diversity and ecological relationships can help inform the most effective genetic conservation strategies[15]. The F_{ST} value between the FC and SF populations indicates that those two populations have moderate genetic differentiation but, F_{ST} values can be tricky to interpret in a biological context and are subject to overestimations and underestimations due to the effects of polymorphisms[16]. The PCA plot did not support the interpretation of the F_{ST} value as there was almost no clustering of data points based on population indicating they are not very different genetically. PCA plots are useful tools in population genetics and widely used for investigating population structure due to the formation of clusters between similar individuals[14].

Future research to inform sustainable harvesting of sochan can expand on this work by focusing on geospatial analysis of the data. Geographic information systems (GIS) should be used to allow for clear graphical presentation of genetic diversity through a map of the GSMNP, which can enhance the use and incorporation of these findings[12]. Using a spatial analysis will be a better way to visualize and understand the geographic patterns of genetic diversity among populations of sochan[12]. Alpha and beta diversity should also be analyzed in order to help with assessment of patterns in genetic variation across local and regional populations. Previous research using GIS to display spatial patterns of genetic diversity has been used to inform conservation strategies in multiple regions by revealing high priority areas of conservation[11,12].

Future population genetics research for sochan can also use these same approaches to focus on screening more loci and testing the allelic diversity of more *R. laciniata* populations. Different populations of sochan within the GSMNP, aside from the 3 used for this research, should be screened with the loci 6256, 6988, 2742773, 2506889, and 2028163. It is possible that sochan populations from locations outside of the GSMNP could be incorporated to compare relative diversities. The intention of this research is to preserve the genetic diversity of sochan, while also upholding the rights of the EBCI to continue their traditional harvests within the GSMNP.

5. Acknowledgments

I would like to express my gratitude to Dr. Jennifer Rhode Ward for her guidance, advice, and time throughout the completion of this research. This research was funded in part by the Carlos Campbell Memorial Fellowship, from the Great Smoky Mountains Conservation Association. I would also like to thank my committee members Dr. David Clarke and Dr. Graham Reynolds for their feedback, and the UNC Asheville Biology Department for their support. Thank you to Dr. Matt Estep for the work done to identify microsatellite regions.

6. References

1. Ticktin T. 2004. The ecological implications of harvesting non-timber forest products. Journal of Applied Ecology. 41(1):11–21.
2. Chamberlain J, Small C, Baumflek M. 2019. Sustainable Forest Management for Nontimber Products. Sustainability. 11(9).
3. Laikre L, Larsson LC, Palmé A, Charlier J, Josefsson M, Ryman N. 2008. Potentials for monitoring gene level biodiversity: using Sweden as an example. Biodiversity & Conservation. 17(4):893–910.
4. Love Stowell SM, Pinzone CA, Martin AP. 2017. Overcoming barriers to active interventions for genetic diversity. Biodiversity and Conservation. 26(8):1753–1765.
5. Vieira MLC, Santini L, Diniz AL, Munhoz C de F, Vieira MLC, Santini L, Diniz AL, Munhoz C de F. 2016. Microsatellite markers: what they mean and why they are so useful. Genetics and Molecular Biology. 39(3):312–328.
6. Taheri S, Abdullah TL, Yusop MR, Hanafi MM, Sahebi M, Azizi P, Redmond RS. 2018. Mining and Development of Novel SSR Markers Using Next Generation Sequencing (NGS) Data in Plants. Molecules. 23(2):399.
7. Dye AS, Shaw S-L. 2007. A GIS-based spatial decision support system for tourists of Great Smoky Mountains National Park. Journal of Retailing and Consumer Services. 14(4):269–278.
8. Lewis C. 2012. The Case of the Wild Onions: The Impact of Ramps on Cherokee Rights. Southern Cultures. 18(2):104-117,119.
9. Gathering of certain plants or plant parts by federally recognized Indian tribes for traditional purposes. 36 CFR2.

10. Faircloth BC. 2008. msatcommander: detection of microsatellite repeat arrays and automated, locus-specific primer design. Molecular Ecology Resources. 8(1):92–94.
11. Zonneveld M Van, Dawson I, Thomas E, Scheldeman X, van Etten J, Loo J, Hormaza JI. 2014. Application of Molecular Markers in Spatial Analysis to Optimize In Situ Conservation of Plant Genetic Resources. In: Tuberosa R, Graner A, Frison E, editors. Genomics of Plant Genetic Resources: Volume 1. Managing, sequencing and mining genetic resources. Dordrecht: Springer Netherlands. p. 67–91.
12. Zonneveld M Van, Scheldeman X, Escribano P, Viruel MA, Damme PV, Garcia W, Tapia C, Romero J, Sigueñas M, Hormaza JI. 2012. Mapping Genetic Diversity of Cherimoya (Annona cherimola Mill.): Application of Spatial Analysis for Conservation and Use of Plant Genetic Resources. PLOS ONE. 7(1):e29845.
13. Sekino M, Hara M. 2001. Application of Microsatellite Markers to Population Genetics Studies of Japanese Flounder Paralichthys olivaceus. Marine Biotechnology. 3(6):0572–0589.
14. Gaspar HA, Breen G. 2019. Probabilistic ancestry maps: a method to assess and visualize population substructures in genetics. BMC Bioinformatics. 20(1):116.
15. Teixeira JC, Huber CD. 2021. The inflated significance of neutral genetic diversity in conservation genetics. PNAS. 118(10).
16. Balloux F, Lugon-Moulin N. 2001. The estimation of population differentiation with microsatellite markers. Molecular Ecology. p. 155-165.