

University of Alabama in Huntsville

LOUIS

Honors Capstone Projects and Theses

Honors College

5-1-2022

Through the Leaves: Understanding Population Genetic Structure of *Clematis morefieldii*

Keith Andrew Greenway

Follow this and additional works at: <https://louis.uah.edu/honors-capstones>



Part of the [Biology Commons](#), [Genetics Commons](#), and the [Plant Sciences Commons](#)

Recommended Citation

Greenway, Keith Andrew, "Through the Leaves: Understanding Population Genetic Structure of *Clematis morefieldii*" (2022). *Honors Capstone Projects and Theses*. 703.
<https://louis.uah.edu/honors-capstones/703>

This Thesis is brought to you for free and open access by the Honors College at LOUIS. It has been accepted for inclusion in Honors Capstone Projects and Theses by an authorized administrator of LOUIS.

Through the Leaves: Understanding Population Genetic Structure of *Clematis morefieldii*

by

Keith Andrew Greenway

An Honors Capstone

submitted in partial fulfillment of the requirements

for the Honors Diploma

to

The Honors College

of

The University of Alabama in Huntsville

May 1, 2022

Honors Capstone Director: Dr. Paul G. Wolf

Professor and Chair, Department of Biological Sciences

 5/1/2022

Student (signature)

Date



05/01/2022

Director (signature)

Date



05/01/2022

Department Chair (signature)

Date

Honors College Dean (signature)

Date



Honors College
Frank Franz Hall
+1 (256) 824-6450 (voice)
+1 (256) 824-7339 (fax)
honors@uah.edu

Honors Thesis Copyright Permission

This form must be signed by the student and submitted as a bound part of the thesis.

In presenting this thesis in partial fulfillment of the requirements for Honors Diploma or Certificate from The University of Alabama in Huntsville, I agree that the Library of this University shall make it freely available for inspection. I further agree that permission for extensive copying for scholarly purposes may be granted by my advisor or, in his/her absence, by the Chair of the Department, Director of the Program, or the Dean of the Honors College. It is also understood that due recognition shall be given to me and to The University of Alabama in Huntsville in any scholarly use which may be made of any material in this thesis.

Keith Greenway

Student Name (printed)

Keith Greenway

Student Signature

5/1/2022

Date

Table of Contents

Dedication	4
Abstract	5
Introduction	6
Morefield and Krall	6
Taxonomy and Morphology	7
Growing Habit	9
Analysis	9
Materials and Methods	10
Sampling	10
Sample Preparation	11
Data Analysis	12
Results	13
Discussion	15
References	16

Dedication

I would like to thank Tracy Cook with the Huntsville Botanical Garden for first introducing me to the beautiful and enigmatic *C. morefieldii* during my internship at HBG, as well as for providing an endless wealth of knowledge and enthusiasm for Alabama native flora that I still cherish.

Additionally, I would like to thank my project director, Dr. Paul Wolf, for his guidance throughout every stage of this project. I would also like to thank Jake Carter for his integral role in the data processing, analysis, and development of this manuscript, as well as everyone in the Wolf lab for their support over the duration of this project.

I would like to thank my family and friends for their patience and support as I have often clumsily and reclusively juggled the pressures of two majors and a minor. Lastly, I would like to thank my service animal, Rosemary, for providing the 4:30 A.M. support needed to emotionally navigate and complete this season of life.

Abstract

Clematis morefieldii, commonly known as Morefield's Leather Flower, is an herbaceous perennial species of scandent vine endemic to northern Alabama, southern Tennessee, and isolated portions of Georgia. Limited to a small handful of populations, little is known about the population genetic structure of *C. morefieldii*. We collected samples of *C. morefieldii* and related species for analysis via ddRADseq. DNA sequence data were processed and genotypes inferred. This study aims to more clearly define the population genetic structure as well as possible relatives of this curious endemic species. Although the current results are inconclusive, the presence of clustering as well as the potential to add samples from a new population prompt further analysis into the population genetic structure of *C. morefieldii*.

Introduction

Morefield and Kral

Clinging to a limestone outcrop nestled within a lot on Round Top Mountain in Madison County, Alabama, in 1982, James Morefield, a then 21-year-old botany student, serendipitously discovered a glade of sprawling vines bearing rust-colored stems with pink and green bell-shaped flowers. Suspecting that the plant could be a newfound endemic species, samples were collected and sent to Vanderbilt University's biology department. It wasn't until 1987 that Dr. Robert Kral officially described the plant as a new species within the *Viornae* subsection of *Clematis*, which he named *Clematis morefieldii*, in honor of Morefield's contribution (Kral 1987). Since that time, *C. morefieldii* has remained a scantily-studied subject whose exact population genetic structure is still unknown. Although *C. morefieldii* was listed as federally endangered in 1992 by the U.S. Fish and Wildlife Services, recent findings suggest that the species could meet the criteria for downlisting, thus emphasizing the need for a formal analysis (ECOS 1992, 2019).

Taxonomy and Morphology

Clematis L., a curious genus within the Ranunculaceae family, contains over 300 members and can be found across a wide range of habitats. Most members of this genus exist as climbing lianas, although some can be found as perennial herbs or subshrubs (Xie 2011). The *Viornae* subsection is largely confined to North America and contains several endemics. Members of this subsection can be characterized by leafy bracts, urceolate flowers with beveled-edged sepals, and an arrangement of single or few-flowered cymes situated within leaf axils (Kral 1987). *Clematis morefieldii*, also commonly known as Morefield's Leather Flower and Huntsville Vasevine, is unique from other members of *Viornae* in that it exhibits tomentose to villous pubescence of the stems, has bracts near or at the base of the peduncle instead of further up, and usually contains shorter peduncles with higher numbers of pale pink and green flowers clustered in the leaf axils, as displayed in **Figure 1** (Estes 2006). An herbaceous perennial, *C. morefieldii*, grows up to 5 m long and bears leaves consisting of 9 to 11 paired leaflets that are reduced distally on the rachis (Kral 1987). The aggregate fruits of *C. morefieldii* are roughly 8 mm long and consist of several ovate-acuminate achenes that gain a trichome-covered style when the seed approaches maturity, as shown in **Figure 2** (Cook 2018). According to Paris, the achenes exhibit double dormancy, successfully germinating in the highest amounts in the second year after planting or dispersal (Paris 2016).



Figure 1: *C. morefieldii* flower



Figure 2: *C. morefieldii* aggregate fruit

Growing Habit

C. morefieldii grows in rocky soil near limestone outcrops, being found largely within the Plateau Escarpment ecoregion of the Cumberland Plateau across Alabama, Georgia, and Tennessee (Cook 2018, ECOS 2019). The literature suggests that *Juniperus* and most notably Smoketree, *Cotinus obovatus*, could be indicator species (Kral 1987). Additionally, the U.S. Fish and Wildlife Services lists a newly-opened canopy as highly conducive to new *C. morefieldii* growth, pointing to its potential role following disturbance (ECOS 2019).

Analysis

In light of its ambiguous population genetic structure, *C. morefieldii* demands a formal analysis be done in order to understand both it and its relatives. Studies have shown strong results in analyzing population genetic structure utilizing ddRAD-seq via ipyrad and STRUCTURE (Rowe et al 2019, Lemon and Wolf 2018). Unlike whole genome sequencing, which can be incredibly expensive, ddRAD-seq utilizes restriction site enzymes to precisely cut genomic DNA at homologous sites across up to several hundred samples, thus saving the researcher time and money (Rowe et al 2019). With such large genomic datasets from ddRADseq, ipyrad serves as a scalable method by which to assemble and analyze the Single Nucleotide Polymorphisms (SNPs) (Eaton 2019).

The objectives of this study were to utilize ddRAD-seq in order to analyze and better determine the population genetic structure of *Clematis morefieldii*, as well as determine possible relatives.

Materials and Methods

Sampling

Samples of *C. morefieldii* were collected from a selection of populations thought to be implicative of its habitat range. Given the plant's confinement largely within the Cumberland Plateau, six populations across southern Tennessee and northern Alabama were collected from, for a total of 63 samples, as displayed in **Table 1**. Samples were collected from individuals spaced at least one meter apart in order to avoid collecting from potential clones.

For each plant sampled, leaf tissue was carefully removed from the plant and wrapped in a dry tea filter, subsequently being sealed in an airtight container housing silica gel in order to dessicate and preserve the tissue for analysis.

Population	Locality	Lat	Long
Walls of Jericho	Franklin County, TN	35.000329 N	-86.07395 W
Snowbird Hollow	Franklin County, TN	35.1105 N	-85.042 W
Hawkins Cove	Franklin County, TN	35.168956 N	-85.96359 W
Buggytop Mountain	Franklin County, TN	35.122803 N	-85.927949 W
Monte Sano Preserve	Madison County, AL	34.741411 N	-86.7543305 W
Monte Sano State Park	Madison County, AL	34.737611 N	-86.51867 W
Keel Mountain Preserve	Madison County, AL	34.66016 N	-86.41732 W

Table 1: *C. morefieldii* Populations Sampled

Additionally, six samples of *C. viorna* and six unknowns were included to the total as outgroups for the analysis.

Sample Preparation

The Genomic DNA samples were prepared by taking the 57 leaf tissue samples collected and giving them identification numbers corresponding to individual Qiagen Microtubes (Qiagen Inc., Valencia, California, USA). A leaf from each sample was taken and cut into three different sections, putting all three sections into the individual eppendorf tube associated with the sample. The genomic libraries were prepared using standard double digest restriction-site associated DNA sequencing (ddRAD-seq) protocols (Parchman, et al. 2012; Rowe, et al 2019). The genomic DNA of the samples was cut using the restriction enzymes EcoRI and MseI. Illumina index barcodes were also added to track each sample during analysis. The barcoded oligonucleotides were ligated to the EcoRI ends of the DNA fragments, while the MseI ends were ligated with a standard, non-barcoded oligonucleotide. Subsequently, the samples were amplified via PCR using iproof high-fidelity DNA polymerase (New England Biolabs Inc., <https://www.neb.com>), along with primers that overlap their ligated oligonucleotides. To mitigate some unwanted, random variation during PCR amplification, the DNA fragments were only mixed with one individual, being subsequently amplified in duplicate. After PCR amplification, the total library was reduced to fragments falling within the size range of 350-450 bp using BluePippin (Sage Science, Beverly, MA). The library was further verified using TapeStation 220 (Agilent Technologies, Santa Clara, CA). The new, reduced library samples were sequenced with 100-bp single-end sequencing at the University of Colorado in Denver.

Data Analysis

The genomic library was returned with about 500 million DNA sequence reads. These were analyzed with iPyRAD v. 0.9.84, a data assembly software, to assemble the raw DNA sequence data for further analysis (Eaton 2020). iPyRAD first reads the barcode for demultiplexing the raw data, which is the process of separating each sample into files based on Illumina index barcodes. The second step consists of quality control, including filtering and trimming to account for adapter contamination. Subsequently, read copies from the same locus are identified by employing reference mapping or *de novo* clustering, followed by an estimation of both the sequencing error rate and sample heterozygosity. Once this has been accomplished, iPyRAD makes haplotype calls and consensus basecalls for the samples, followed by identification of orthology across the sample set. Lastly, iPyRAD employs a last run, filtering and trimming the assembled loci in order to create assembly statistics and output files for analysis (Eaton 2020).

We explored the heterozygosity levels of loci across individuals. We used the total data set for total representation of the species that had sequencing significance, samples to be considered the outgroup, and an unknown group blind for analysis (51 individuals of *C. morefieldii*, 6 individuals of *C. viorna*, 6 individuals of unknown). Exclusion of samples for analysis were done due to sequencing error rate parameters (5 individuals of *C. morefieldii* and 1 *C. viorna*) were excluded from iPyRAD analysis. Total read count (488,576,739) for *C. morefieldii* data before demultiplexing the file and average read counts (621) for individual *C. morefieldii* samples.

Assembled data was assessed using the program STRUCTURE (Pritchard et al. 2000; Falush et al. 2003, 2007; Hubisz et al. 2009). We used STRUCTURE for a Bayesian clustering approach to statistically assign individuals to their source populations based on genetic data. We ran 4 replicates for K values (clusters or groups) 3 and 4 with 25,000 burn-in periods and 100 repetitions. Results were assembled and summarized using STRUCTURE's front end software. True value of K was found by comparison of cluster outputs ranging of K values from 2 through 6.

Results

After data filtering, 7 total filtered loci remained. The average read depth was 64 reads per sample, and the average sequencing error rate across samples was 0.0069.

The mean level of heterozygosity of individuals across the 20 total loci STRUCTURE counted was insufficient due to a parameter file error during iPyRAD step1. Figure 3 (K value is 3) is indicative of 3 separate peaks, showing trace amounts of heterozygosity clustering among different samples from different locations. The y-axis in figure 3 are inferred cluster separations among the 3 K values, ranging from 0.00 to 1.00. The x-axis are individuals sampled. Populations that are showing clustering in red, green, and blue regions are from Hawkins Cove(red), Buggytop Mountain(Green), and a mixture of Snowbird Hollow and Hawkins Cove(blue). Figure 4 (K value is 4) shows 4 less precise peaks indicating that the true K value is not higher than K=3. Population clustering while K=4 shows the same clustering as K=3, but also loosely indicates that *C. viorna* (green) has genetic similarity among the *C. morefieldii* sample populations.

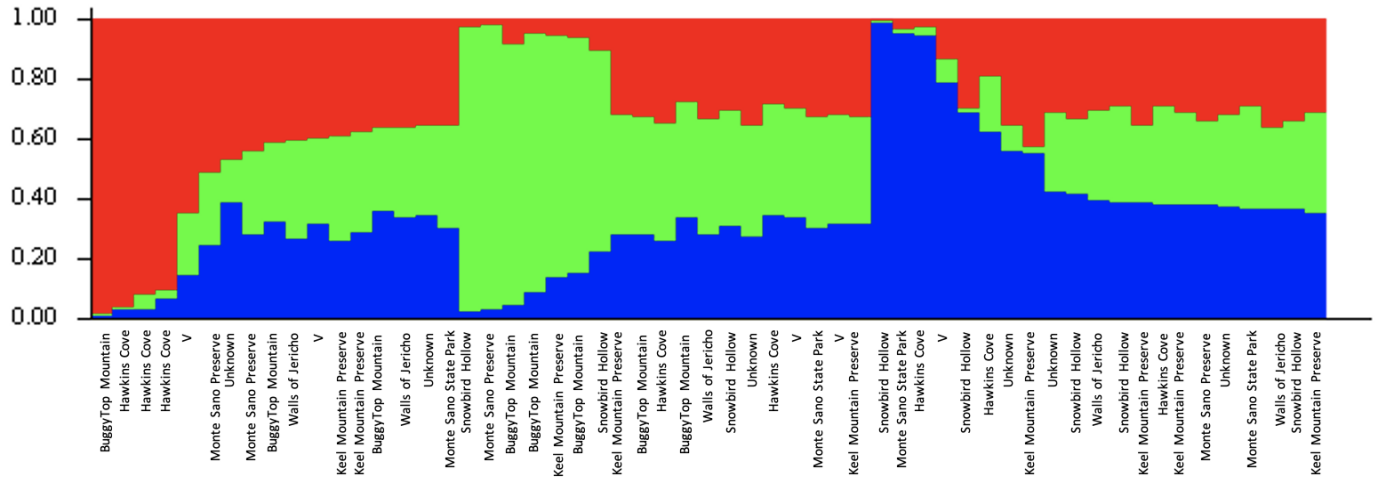


Figure 3: STRUCTURE Analysis of K value at 3.

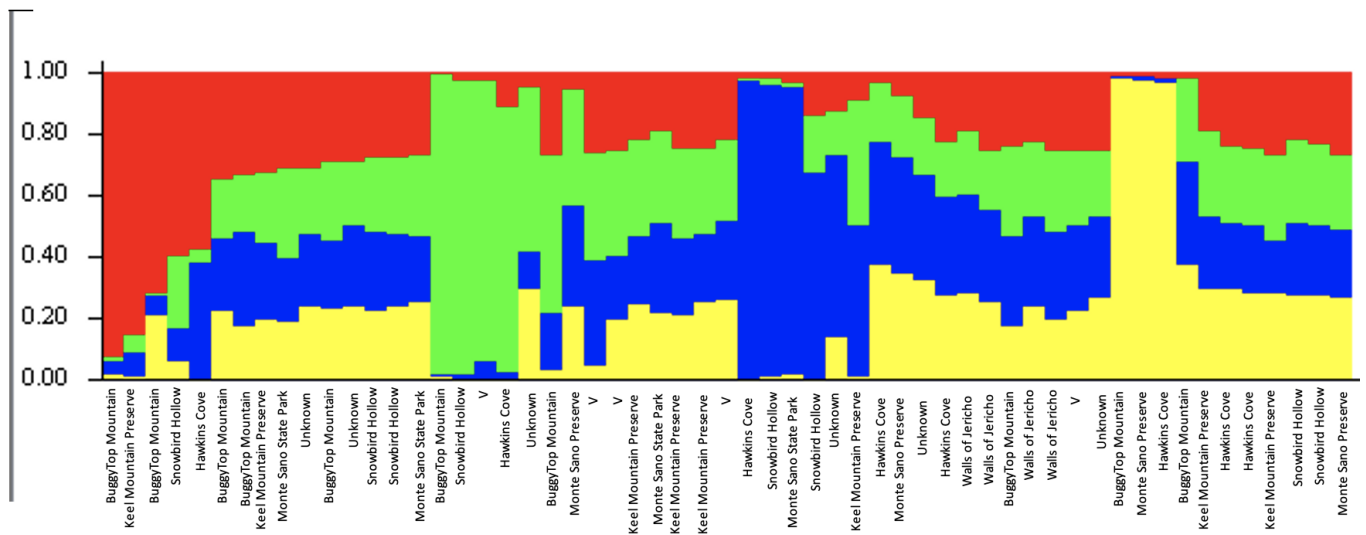


Figure 4: STRUCTURE Analysis of K value at 4

Discussion

Although inconclusive, the presence of clustering across the *C. morefieldii* dataset suggests that further analysis will be required to gain clearer results. This will be accomplished through future runs of iPyRAD that avoid the parameter file error encountered during this project. Further analyses could include Principal component analysis (PCA), where the relationship between genetic distance and geographical distance could be plotted and determined (Bro and Smilde 2014). Additionally, collection trips in the Blackbelt region of Mississippi have yielded samples of what appear to be an additional *C. morefieldii* population. Aside from being unique in its soil conditions, the Blackbelt population, if truly *C. morefieldii*, would extend the current range of the plant and would have potential implications regarding its federal listing status. In a further analysis of *C. morefieldii*, including the Blackbelt population to the current dataset could help to refine and verify the tentative results of this project, thus prompting continued analyses on one of Alabama's enigmatic natives.

References

- Bro, Rasmus, and Age K. Smilde. "Principal Component Analysis." *Anal. Methods*, vol. 6, 2014, pp. 2812-31.
- Cook, Tracy Lea. *Phytogeographic Analysis of Clematis morefieldii in the Southern Cumberland Plateau*. 2018. Alabama A & M U, MS thesis.
- Eaton, Deren A. R., and Isaac Overcast. "ipyrad: Interactive assembly and analysis of RADseq datasets." *Bioinformatics (Oxford, England)*, vol. 36, no. 8, 2020, pp. 2592-94.
- Environmental Conservation Online System (ECOS). *U.S. Fish and Wildlife Service, Federal Register*. vol. 57, no. 98, U.S. Fish and Wildlife Service , 20 May 1992.
- Environmental Conservation Online System (ECOS). (1992). *U.S. Fish and Wildlife Service, Federal Register*, 57(98).
- Estes, Dwayne, and Chris Fleming. "*Clematis morefieldii* (Ranunculaceae) New to Tennessee." *Contributions to Botany*, vol. 22, no. 1, 2006, pp. 821-24.
- Falush, D., Stephens, M., and Pritchard, J. K. "Inference of population structure using multilocus genotype data: dominant markers and null alleles." *Mol. Ecol. Notes*, vol. 7, 2007, pp. 574–578. doi: 10.1111/j.1471-8286.2007.01758.x.
- Falush, D., Stephens, M., and Pritchard, J. K. "Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies." *Genetics*, vol. 164, 2003, pp. 1567–1587.
- Hubisz, M. J., Falush, D., Stephens, M., and Pritchard, J. K. "Inferring weak population structure with the assistance of sample group information. *Mol. Ecol. Resour.*, vol. 9, 2009, pp. 1322–1332. doi: 10.1111/j.1755-0998.2009.02591.x.

- Kral, Robert. "A New 'Viorna' Clematis from Northern Alabama." *Annals of the Missouri Botanical Garden*, vol. 73, no. 3, 1987, pp. 665-69.
- Lemon, Jenessa B. and Paul G. Wolf. "Genetic Differentiation between Endemic *Eriogonum soredium* and Its Common Relative *E. shockleyi* (Polygonaceae)," *Systematic Botany*, vol. 43, no. 4, November 2018, pp. 901-909.
- Parchman, Thomas L., et al. "Genome-wide association genetics of an adaptive trait in lodgepole pine." *Molecular Ecology*, vol. 21, no. 12, 2012, pp. 2991-3005.
- Paris, Kyle J., et al. "Reproductive Biology of the Federally Endangered Clematis Morefieldii Kral (Ranunculaceae)." *Castanea*, vol. 81, no. 3, 2016, pp. 175–87, <http://www.jstor.org/stable/26353984>.
- Pritchard, J. K., Stephens, M., and Donnelly, P. "Inference of population structure using multilocus genotype data." *Genetics*, vol. 155, 2000, pp. 945–959.
- U.S. Fish & Wildlife Services. *Morefield's Leatherflower Recovery Plan Amendment*. By *Environmental Conservation Online System (ECOS)*. 2019. https://ecos.fws.gov/docs/recovery_plan/Morefield_s%20Leather%20Flower%20Recovery%20Plan%20Amendment.pdf.
- Rowe, Carol A., et al. "Do Genetic Differences Explain the Ability of an Alkaline Shrub to Grow in Both Uplands and Wetlands?," *Western North American Naturalist*, vol. 79, no. 2, July 2019, pp. 260-269.
- Xie, Lei, et al. "Phylogenetic Analyses of Clematis (Ranunculaceae) Based on Sequences of Nuclear Ribosomal ITS and Three Plastid Regions." *Systematic Botany*, vol. 36, no. 4, 2011, pp. 907–21, <http://www.jstor.org/stable/41416907>.