

Evolutionary Divergence and Population Structure of the North American *Castanea* Species

M. Taylor Perkins¹, Alexander M. Sandercock², Paul H. Sisco³, Frederick L. Paillet⁴, Ronald S. Revord⁵, Jared W. Westbrook⁶, Jason A. Holliday⁷, J. Hill Craddock¹

¹Department of Biology, Geology, & Environmental Science, The University of Tennessee at Chattanooga, Chattanooga, TN, USA ²Genetics, Bioinformatics, and Computational Biology, Virginia Tech, Blacksburg, VA, USA ³The American Chestnut Foundation – Carolinas Chapter, Asheville, NC, USA ⁴Department of Geosciences, University of Arkansas, Fayetteville, AR, USA ⁵Center for Agroforestry, University of Missouri, Columbia, MO, USA ⁶The American Chestnut Foundation, Asheville, NC, USA ⁷Department of Forest Resources and Environmental Conservation, Virginia Tech, Blacksburg, VA, USA

Background

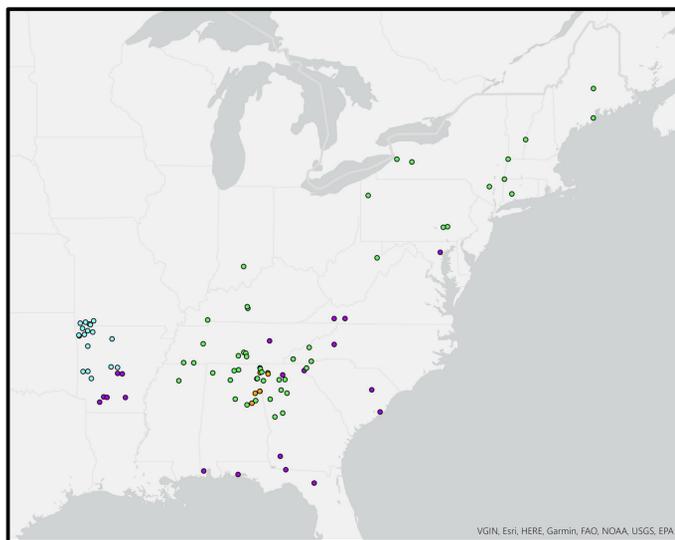
American chestnut (*Castanea dentata*) and the chinquapins (*C. pumila* sensu lato[†]) are an evolutionary sister species pair that represent a promising system for studies of admixture and adaptation in wild plants. Since the 1920s, botanists and geneticists have hypothesized that hybridization between American chestnut and the chinquapins is common and that this process has been involved in the origins of some *Castanea* taxa and populations [1-3]. Rigorously testing these hypotheses, however, has been difficult until the recent availability of high-throughput DNA sequencing technologies and associated computational tools. Here, we present preliminary results of our analyses of whole genome sequencing data from 255 plants representing all North American *Castanea* species and subspecific taxa

Research Questions

- Have American chestnut populations experienced introgression from any of the chinquapin taxa (*pumila*, *ozarkensis*, or *alabamensis*), particularly where they occur in sympatry?
- Has Alabama chinquapin (*C. alabamensis*) experienced introgression from *C. dentata*?
- What is the frequency of admixture between different chinquapin taxa?

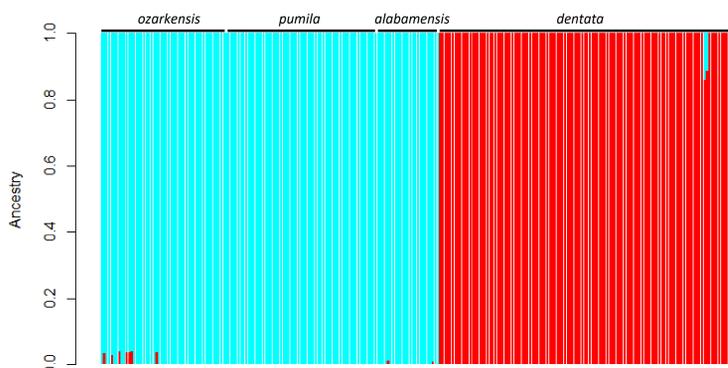
Sampling

Between 2006 and 2021, we collected leaf tissue samples representing all North American *Castanea* taxa. We generated whole-genome sequencing (WGS) data from 255 individual plants using methods described previously [4]. Datasets containing >27,000,000 single-nucleotide polymorphism loci (SNPs) were used for all analyses. Sample sites are shown below (green dot = *C. dentata*, purple = *pumila*, gold = *alabamensis*, light blue = *ozarkensis*).



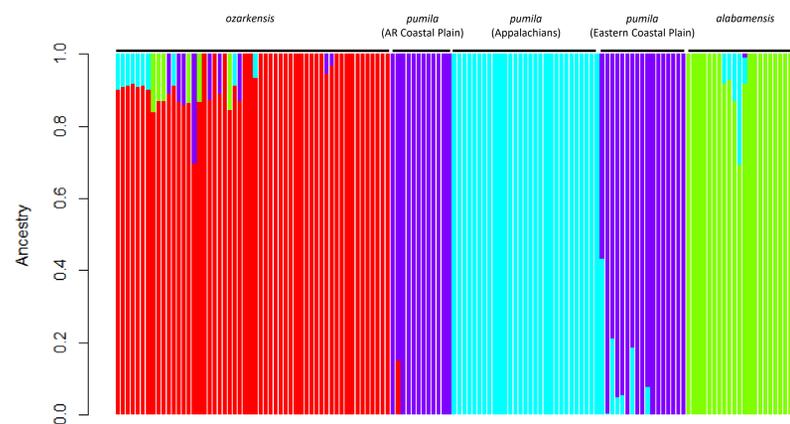
Population structure analysis suggests admixture between American chestnut and the chinquapins has been rare

We inferred population structure of the North American *Castanea* species using ADMIXTURE software [5]. The genetic structure of our entire sample set was best characterized by a model involving two clusters ($K=2$), which correspond to (1) the chinquapins collectively and (2) *C. dentata*.



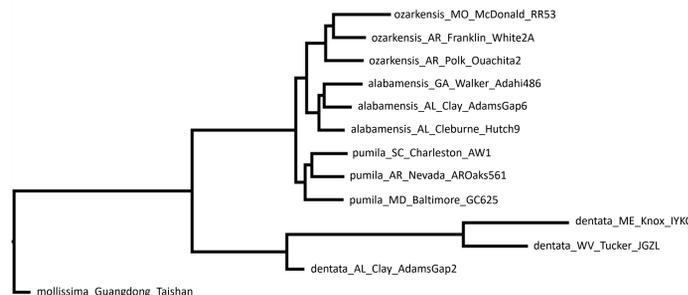
Population structure analysis suggests admixture between different chinquapin taxa has been more common

The genetic structure of the North American chinquapin taxa was best characterized by a model involving four clusters ($K=4$): (1) Ozark chinquapins, (2) Allegheny chinquapins from Coastal Plain sites, (3) Allegheny chinquapins from sites in the Appalachians, and (4) Alabama chinquapins.



The chinquapin taxa comprise a sister group to American chestnut

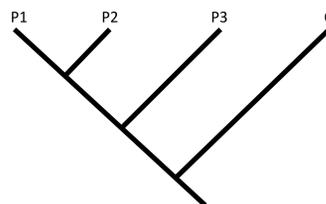
We inferred a maximum-likelihood phylogenetic tree using IQ-TREE software [6]. Bootstrap support for nodes separating taxa was 100% in all cases.



D statistics (i.e., ABBA-BABA) to test for gene flow

We used Patterson's *D* statistic (also known as the ABBA-BABA test) [7] to test for gene flow/admixture between different *Castanea* species and between different taxa within the chinquapin group.

Consider the four populations (labeled P1, P2, P3, and O) and their phylogenetic tree shown below. With *D* statistics, we test the null hypothesis that P1 and P2 descend from a common ancestral population that split from the ancestors of P3 at an earlier time, without any gene flow between P3 and P1 or P2 after they split. The alternative hypothesis is that P3 exchanged genes with P1 or P2 after these two populations diverged. *D* statistics can also be applied to tests for gene flow between species, as we have done in our study.

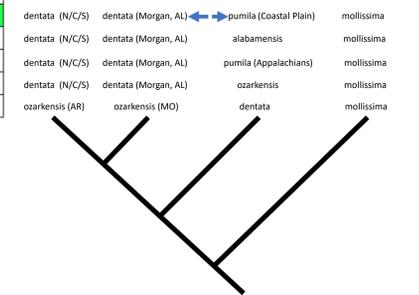


The *D* statistic considers biallelic sites in the genome where we have coverage for P1, P2, P3, and O. The allele in the O population is denoted "A", and any alternative alleles in P1, P2, or P3 are denoted "B". The pattern "ABBA" refers to sites in the genome where P1 has the outgroup allele (A) and P2 and P3 share the derived allele (B). The pattern "BABA" refers to sites where P1 and P3 share the derived allele (B) and P2 has the outgroup allele (A). If P1 and P2 have diverged without experiencing gene flow with P3, the frequency of ABBA sites will be equal to the frequency of BABA sites, and the *D* statistic will not differ significantly from zero.

Hybridization between American chestnut and the chinquapins has been rare

Using the software package 'Dsuite' [8], we calculated Patterson's *D* statistics to test for gene flow between *C. dentata* and chinquapin populations for which ADMIXTURE had produced evidence suggestive of hybridization. Significant test results are highlighted in green.

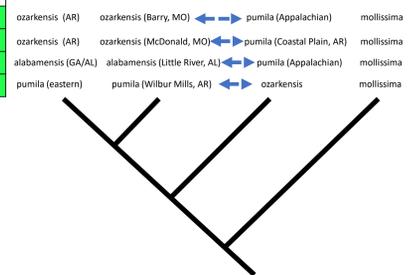
# ABBA sites	# BABA sites	<i>D</i> statistic	Z score	<i>p</i>
199,297	195,668	0.009	1.9	0.047
198,348	198,872	0.001	0.2	0.811
198,796	197,011	0.005	0.9	0.349
197,658	197,885	0.001	0.1	0.903
277,987	278,543	0.001	0.7	0.508



Hybridization between different chinquapin taxa has been more common, but still infrequent

We calculated Patterson's *D* statistics to test for gene flow between chinquapin population pairs for which ADMIXTURE had produced evidence suggestive of gene flow. Significant test results are highlighted in green.

# ABBA sites	# BABA sites	<i>D</i> statistic	Z score	<i>p</i>
400,659	382,624	0.023	8.0	1.5e-15
413,841	344,558	0.091	22.9	2.3e-16
438,159	365,488	0.092	30.7	2.3e-16
424,897	379,010	0.057	14.8	2.3e-16



Conclusions

- Population structure analysis and *D* statistics provided concordant results for tests of gene flow between American chestnut and chinquapin. We found evidence of chinquapin ancestry in only one American chestnut sample site, a location in northern Alabama.
- We found no evidence to support the hypothesis that *C. alabamensis* has introgressed ancestry from *C. dentata*. Interestingly, our phylogenetic analysis indicates that *C. alabamensis* is a sister group to Ozark chinquapin. This is consistent with hypotheses on the relationship between Ozark chinquapin and *C. alabamensis* that date back to the early 20th century [9,10].
- *D* statistics and analysis of population structure indicate that admixture between *C. alabamensis* and Allegheny chinquapin is infrequent, but not rare. The same analyses indicate that admixture between Ozark and Allegheny chinquapin is also infrequent. Discrepancies between results from population structure analysis and *D* statistics are probably due to incomplete lineage sorting.
- We found evidence for the existence of two distinct ecotypes of Allegheny chinquapin – one from the Coastal Plain and a second from the Appalachians. These two ecotypes appear to have experienced substantial gene flow in the Atlantic Coast and eastern Gulf Coastal Plain regions. This is consistent with more limited information obtained in a recent study using genotyping-by-sequencing [11].
- Future work will include analyses of local (chromosome-scale) ancestry and tests for evidence of natural selection at introgressed ancestry tracts.

Notes

[†]The genus *Castanea* in North America is comprised of four taxa: American chestnut (*Castanea dentata*), Allegheny chinquapin (*C. pumila* var. *pumila*), Ozark chinquapin (*C. pumila* var. *ozarkensis*), and Alabama chinquapin (*C. alabamensis*). Controversy exists over whether the chinquapin taxa should be considered distinct species or varieties. Because taxonomic treatment of the chinquapins has been in flux, we follow the naming system of the most recent taxonomic revision of the group [10] by treating Ozark and Allegheny chinquapins as varieties. Because Alabama chinquapin was not included in the most recent taxonomic revision, we treat this taxon using its formally recognized name, *Castanea alabamensis*, consistent with its original description [9]. We do not, however, consider this to be a definitive recommendation, since more study is needed to revise the naming system of the chinquapins.

References

- 1: Dode. 1908. *Bulletin de la Societe Dendrologique de France* 8:140-166
- 2: Camus. 1929. *Encyclopedie Economique de Sylviculture*, vol. 3
- 3: Li and Dane. 2013. *Tree Genetics & Genomes* 9:107-116
- 4: Sandercock et al. 2022. *Molecular Ecology* 31(18):4640-4655
- 5: Alexander et al. 2009. *Genome Research* 19(9):1655-1664
- 6: Nguyen et al. 2015. *Molecular Biology and Evolution* 32:268-274
- 7: Durand et al. 2011. *Molecular Biology and Evolution* 28(8):2239-2252
- 8: Malinsky et al. 2021. *Molecular Ecology Resources* 21:584-595
- 9: Ashe.1925. *The Charleston Museum Quarterly* 1:28-32
- 10: Johnson. 1988. *Journal of the Arnold Arboretum* 69: 25-49
- 11: Perkins et al. 2021. *Systematic Botany* 46(4):973-984

Acknowledgments

MTP's work on this project was supported by a TACF External Grants program award. JHC was supported, in part, by The Summerfield Johnston Endowment for the Restoration of the American Chestnut.