2017

CHARACTERIZATION OF GENIC MICROSATELLITE MARKERS (EST-SSRS) IN THE ENDANGERED OAK SPECIES QUERCUS GEORGIANA M.A.CURTIS

Priyanka Dipak Kadav
Michigan Technological University, pdkadav@mtu.edu

CHARACTERIZATION OF GENIC MICROSATELLITE MARKERS (EST-SSRS) IN THE ENDANGERED OAK SPECIES *QUERCUS GEORGIANA* M.A.CURTIS

By

Priyanka Dipak Kadav

A REPORT

Submitted in partial fulfillment of the requirements for the degree of

MASTERS OF SCIENCE

In Forest Molecular Genetics and Biotechnology

MICHIGAN TECHNOLOGICAL UNIVERSITY

2017

This report has been approved in partial fulfillment of the requirements for the Degree of MASTER OF SCIENCE in Forest Molecular Genetics and Biotechnology.

School of Forest Resources and Environmental Science

Report Advisor: *Dr. Oliver Gailing*

Committee Member: *Dr. David Flaspohler*

Committee Member: *Dr. Hairong Wei*

School Dean: *Dr. Terry Sharik*

Table of Contents

List of Tables and Figures

Acknowledgement

Heartily, I express my sincere gratitude to the Almighty for showering blessing upon me and giving me the strength to do the research work.

First and foremost, I am grateful to my guide Prof. Oliver Gailing for his constant support and guidance. His guidance aided me to learn about research in its true sense. He paid invaluable time and painstaking effort throughout my project work. It was my privilege to work with him. I am extremely thankful to him who gave me this opportunity to carry out research and meeting with the highly skilled scientists in the laboratory.

I am grateful to The Morton Arboretum and Chicago Botanical Garden for offering this project and The Morton Arboretum and Ecosystem Science Centre for funding me.

My grateful thanks to our lab manager Dr. Jennifer Sanders for her espouse during the project work.

Sincere thanks to my committee members Dr. David Flaspohler and Dr. Hairong Wei.

I would like to acknowledge the timely help provided by Sudhir Khodwekar for his continuous encouragement and right direction towards the goal. He stood beside me as a pillar throughout my research journey. Things would have been difficult without his constant support. I would like to thank Munkaila Mussah for encouraging me to do my research and who took care of me considering as his little sister.

My family has given me constant and unequivocal support throughout this journey. My mother has always been my strength and inspiration and her belief in me made this journey successful. My uncle Prashant Deshpande back in Abu Dhabi has always been my support in this journey.

Last but not the least; I would like to thank my friends Monica Harmon, Ruhua Zhang, Sirikorn Khumwan, Mehmet Acet, Roba Bdeir and Robert Richard for accompanying me during the journey.

Practical experience is a prerequisite in today's generation. The laboratory work provided me with immense knowledge. It was a memorable journey that will always remain evergreen in my mind.

Abstract

Quercus is important ecologically and economically because it provides food and habitat for wildlife, wood and paper products for humans. Oaks are endangered due to various factors like shifting climates, habitat loss, drought, pathogens and genetic swamping. *Quercus georgiana* (M.A. Curtis) is an endangered and restricted oak species which is remaining only in the southeastern part of the US. Efforts are required to conserve this endangered species from extinction. Conservation of this species can be done through these methods: *ex-situ* conservation (arboretum and botanical garden) and *in-situ* conservation strategies which protect the species in its natural habitat. For this conservation strategy, it is important to ensure that sample collections capture as much of the gene pool as possible so that the biodiversity is maintained. A variety of molecular markers are available for *Quercus.* These markers which are highly polymorphic, co-dominant and multiallelic loci will be useful in the study of population genetics of *Q. georgiana.* Genetic variations in both, among and within the populations, have to be considered if sampling and conservation strategies are developed for this rare and endangered species. These analyses are important in the future for sample collection trips so that the conservation goal is obtained.

Introduction

Deforestation, pollution, and climate change are threatening the forest diversity all over the world (Ledig 1988). Since forests are the habitats for diverse organisms, the threat to forest diversity is extended to these flora and fauna that are associated with forests and not only the forest trees. When a species becomes extinct, along with that much of the genetic diversity is lost too upon which long-term survival and evolution depend. Genetic variation is essential for a population in order to adapt to the environment and survive (Reed, Frankham 2003). Once a species or a population with a unique genetic information has been lost, it cannot be restored. The protection of endangered species and its genes for future generations is important to maintain the genetic diversity in a species (Ledig 1988). Therefore, conservation biology plays an important role in the preservation of these genes. The field of conservation biology is dealing with scientific knowledge useful to preserve biodiversity. A great deal of effort in conservation biology has focused on conservation genetics. Conservation efforts are usually classified as *in-situ* and *ex-situ*, meaning in the natural position (or in place) and out of the natural position, respectively.

Quercus georgiana

According to Botanic Gardens Conservation International (BGCI US), 17 species of oaks in the US are endangered like *Quercus acerifolia* (Palmer), *Quercus alba* (U. S. P), *Quercus hinckleyi* (C. H. Muller), *Quercus georgiana* (M. A. Curtis). *Quercus georgiana* is an endangered US endemic oak species that is left in only a few populations of the Piedmont region in the southeastern US (GA, AL, and NC). There are 14 counties of this species in Georgia and 3 counties in Alabama. There are only five individual trees of the entire population that are remaining in North Carolina and the remnant

populations in the South Carolina which are considered to be the part of the historic range of this species are now eliminated (Oldfield, S. and A. Eastwood 2007). *Quercus georgiana* is important from both ecological and economic perspectives. *Quercus georgiana* has ornamental value because of its glossy green leaves, attractive autumn colors of purples and reds and its resilience to drought and heat. Its leaves and acorns are food sources for larvae and woodpeckers, deers and small mammals (Toppila 2012). Oaks are anemophilous i.e. gene flow occurs by wind pollination.

Threats

Tourism and recreation are major threats to *Q. georgiana*. This species is present at popular hiking trails within state parks and nature reserves at Stone Mountain. Erosion, poor regeneration, climate change, drought conditions and compacted soils that result in foot and vehicle traffic on granite outcrops are also major reasons of threats. Genetic swamping and introgression from surrounding red oak species such as *Quercus* (section Lobatae) results in the threatening of the genetic identity of *Quercus georgiana* (Wenzell 2015).

Molecular methods

Molecular tools such as microsatellites, RAPDs, AFLPs and chloroplast DNA are used for population genetic studies. Microsatellite markers provide information in genetic analysis studies because of their high polymorphism and co-dominant nature of inheritance (Sertse et al. 2013; Keiper et al. 2003; Stefenon et al. 2008). Microsatellites are multiple repeats of short sequences of DNA which are used to assess genetic variation. The populations of threatened or endangered organisms are analyzed using molecular markers that need to be saved before they go extinct. Expressed Sequence Tags (ESTs) are

rich sources of SSRs which are often not only polymorphic within the source taxon, but in related taxa, as well (Ellis & Burke 2007).

Ex-situ and *In-situ* conservation of exceptional species

Quercus georgiana is an 'exceptional species'. The seeds of this species cannot be banked because acorns will not survive long-term dry storage. Research is going on for the acorns of the species which cannot be stored in seed banks using *in vitro* propagation and cryopreservation as a conservation method (Toppila 2012). *Ex-situ* collections must target the entire range of individual samples to seize the maximum genetic diversity possible from them. The second method of conservation is the *in-situ* conservation strategy where the species are protected in its natural habitat.

The Morton Arboretum and Chicago Botanic Garden aims at studying the genetic diversity of *Q. georgiana* from both natural and cultivated collections for future conservation efforts. The main aim of the project is to characterize genic microsatellite markers (EST-SSRs) in this endangered oak species in nine populations in Georgia and Alabama. The sampling from the other populations was not made because *Q. georgiana* was infrequent and the trees appeared to be hybrids in those populations. Hence, these nine populations are the collection representatives that cover the distribution range of the species. The overall goal is to perform genetic variation analyses using *Q. georgiana* as a case study. Decisions can be made regarding which populations in a species and how many should be conserved which is important in the future, for sample collection trips so that the *ex-situ* conservation prime mission is obtained that can lead to future reintroduction and restoration efforts (Whitlock 2016).

Materials and Methods

Plant Material

A total of 215 samples consisting of the nine populations of *Quercus georgiana* were sampled in Alabama and Georgia (Table 1), and sampled plants were at least five meters apart. Also, the GPS coordinates were recorded for each plant. The nine populations consisted of 24 samples for most populations and 23 samples for one population (Stone Mountain).

Marker analyses

A total of 27 genic microsatellites (Expressed Sequence Tag- Simple Sequence Repeats, EST-SSRs) originally developed and genetically mapped in *Quercus robur* (Durand et al. 2010; Bodénès et al. 2012) were tested in a total of eight samples from four *Q. georgiana* populations. Markers were selected based on successful transferability to North American oak species of section *Lobatae* (Lind and Gailing 2013, Sullivan et al. 2013, Lind-Riehl et al. 2014). Twelve of these markers were successfully adopted for use in *Q. georgiana* and all samples were analyzed at these markers. Functional annotations were assigned to EST-SSRs using the Blast2GO software (Conesa et al. 2005).

PCR amplification followed the protocol described in Lind and Gailing (2013). Specifically, samples were amplified in a GeneAmp PCR system 2700 (Applied Biosystem) with the following program: initial denaturation at 95 °C for 15 min followed by 35 cycles of denaturation at 94 °C for 45 sec, annealing at T_a (Table 2) for 45 sec and extension at 72 °C for 45 sec. The final extension step was for 20 min at 72 °C. Each PCR was performed in a 15 µl reaction mix composed of 5 µl HotFIREPol (Oak Biotechnologies, containing 10 mM $MgCl_2$, 0.6 units of HOT FIREPol® Taq polymerase, and 2 mM of each dNTP), 2 µl fluorescently labeled forward primers (5 µM), 2 µl reverse primers (5 µM), 2 µl DNA (~2 ng), and 4 µl molecular grade ultra-pure water (Phenix Research Labs). The PCR amplicons

were checked on 2% agarose gels in 1x TAE buffer at 200 V for 20 minutes. Exact fragment sizes of the amplified DNA products were determined on an ABI3730 Genetic Analyzer with the internal size standard GS-LIZ-500 (Applied Biosystems).

Data Analysis

The scoring of the data was done with the software GeneMarker® V2.6.3 (SoftGenetics) and fragments were assigned to bins after careful visual inspection.

Genetic diversity analyses

Genetic variation parameters expected heterozygosity (H_e) (Nei 1973), observed heterozygosity (H_o) and number of alleles per locus (N_a) were calculated in GENAlEx v. 6 (Peakall and Smouse 2006). Also, the number of private alleles and of locally common alleles (found in ≤ 25% of all populations) was calculated in GENEALEx. For individual markers, the inbreeding coefficient F (Wright 1965) was calculated for all populations. Significant deviation from Hardy-Weinberg proportions was tested in GENEPOP 4.2 (Raymond and Rousset 1995, Rousset 2008) using probability tests for each marker and population. Pairwise genetic differentiation between populations (F_{ST}) was calculated and tested for significance in GENEPOP 4.1with the exact G test using default settings (Raymond and Rousset 1995, Rousset 2008). A Mantel test (Mantel 1967) was performed in GENAlEx v. 6 which compares a genetic distance matrix with a geographical distance matrix to test for correlation between genetic distance and geographical location. The Mantel test performs permutations on one matrix while holding the other constant.

Genetic assignment analysis

STRUCTURE v. 2.3.4 was used to infer population genetic structure using the multi-locus genotype data (Pritchard, Stephens and Donnelly 2000.) The Structure software implements a model-based

genetic admixture analysis using Bayesian methods to assign individuals to populations. To determine the number of clusters K that best fit our data, we performed five independent runs with a burn-in period of 30,000 iterations followed by 10^6 iterations for each value of K (K=1-9) under the admixture model with correlated allele frequencies without any prior information regarding species identification. A dendrogram (Neighbor joining method) was created based on Cavalli-Sforza and Edward's (1967) pairwise genetic distances at EST-SSRs using Populations 2.0 (Langella 1999). Statistical support of clusters was determined with 1,000 bootstrap replicates. TreeViewX (Page 1996) was used to visualize the dendrogram.

Results

All twelve microsatellite primers amplified and were polymorphic (Table 3). For individual markers, observed heterozygosity (H_o) across populations was highest at marker FIR048 and it was found to be 0.838 and expected heterozygosity (H_e) across populations was highest at marker FIR043 and it was found to be 0.802. Marker VIT081 has the lowest observed heterozygosity (H_o) and expected heterozygosity (H_e) values. The mean F values (inbreeding coefficient) was highest at locus GOT037 and it was found to be 0.1924 and the lowest at the locus PIE039 which was found to be -0.1763 (no indication of inbreeding at this marker). The number of alleles per locus ranged from 2 to 9 (Table 3). There was no indication of null alleles in any of the populations. At the population level, CB and AM populations had the highest observed heterozygosity (H_o), i.e. 0.567, and expected heterozygosity (H_e), i.e. 0.574, values whereas PN population had the lowest values of 0.447 and 0.471, respectively. The mean inbreeding coefficient was highest in population AM (0.116) and lowest in population CR (-0.114). It was seen that the F values were not consistently positive at most markers per population. Thus, there was no indication of inbreeding in any of the populations. The number of private alleles and locally common alleles (≤ 25%) were also calculated. CB population had the highest number of private and locally common alleles. The number of alleles per population ranged from 3 to 5 (Table 4).

Hardy- Weinberg Exact tests (HWE) were performed by locus in each population. Values in boldface type are significantly different from Hardy-Weinberg proportions (α=0.05) after Bonferroni correction ($p<0.05/12=0.00416$) (Table 5). It was seen that most of the markers did not show significant deviations from Hardy- Weinberg proportions except for marker FIR028 in population ED (positive F value), GOT037 in population AM (positive F value), FIR043 and PIE200 in population CR (negative F values) which were significantly different from Hardy-Weinberg proportions. ANOVA (Analysis of

Variance) and Posthoc tests were performed which indicate no significant differences in mean H_e, mean H_o and mean N_a between populations.

Pairwise genetic distances (F_{ST}) between populations were found to be significant after exact G test ($p<0.001$). The pairwise genetic distance (F_{ST}) values between the populations PN and DK was highest (0.129) (Table 6) and it was seen in the STRUCTURE results that they are not clustered together. Pairwise geographic distances between populations in kilometers were also calculated. But according to the geographic distance, the distance between PN and DK is 57.22 kilometers. The lowest F_{ST} values were between ED and DK and between WG and DK populations indicating that they are genetically related to each other. But according to the geographic distance, the distance between ED and DK is 176.3 kilometers and between WG and DK is 162.98 kilometers. The highest pairwise geographic distance was between WG and MR and it was found to be 420.3 kilometers (Table 7). A Mantel test, based on the correlation between a genetic distance matrix and a geographical distance matrix was performed. The regression coefficient R^2 is comparatively low (0.2692) and p value is 0.100 (not significant), indicating that there is no correlation between genetic and geographic distance (Fig. 1).

Genetic assignment analysis performed in STRUCTURE indicated K=2 as the most likely number of genetic clusters (Fig. 2). In each population most individuals were assigned genetically to one of the two clusters such that the first four populations (DK, CB, ED, WG) were grouped in cluster 1 (red) and the remaining five populations (PN, AM, SM, MR, CR) were grouped in cluster 2 (green). The dendrogram (Neighbor joining method) based on Cavalli-Sforza and Edwards (1967) pairwise genetic distances at EST-SSRs revealed low and non-significant bootstrap values for most clusters (bootstrap values < 50). The grouping of populations was not related to their geographic location (Fig. 3). The analysis of population structure identified two distinct genetic clusters, but no association between

ancestry in one of the clusters and geographic location in one of the three sampled forest fragments was detected (Fig. 4).

Discussion

Quercus georgiana was used to study genetic diversity through the use of microsatellites. In this study, genetic variation, measured by mean expected heterozygosity, was analyzed in the endangered species *Q. georgiana*.

Pairwise genetic distances (F_{ST}) between populations were found to be significant but no correspondence with geographic distances (Fig. 4). There was no association between genetic and geographic distances. Mantel tests also indicated that there is no correlation between genetic and geographic distance (Fig. 1). Though *Q. georgiana* is discontinuously distributed throughout the Piedmont region because of habitat loss and fragmentation, it might be due to historic gene flow between populations through pollen movement by wind.

Genetic variation of *Quercus georgiana* was low when compared to another endangered oak species *Quercus hinckleyi* (C.H. Muller). *Quercus hinckleyi* is distributed in the Brewster and Presidio counties in Texas (Backs et al. 2016). The genetic variation measured by expected heterozygosity was found to be 0.853 which is high than the genetic variation of *Q. georgiana* species. The comparison of population genetic diversity between a rare, narrowly distributed species *Quercus georgiana* and a common, widespread species of *Quercus rubra* was studied. Northern red oak (*Quercus rubra*) is widely distributed throughout the eastern part of US. It extends from the Atlantic coast in the east to the Mississippi River in the west and from southern Ontario in the north to central Georgia in the south (Sork et al. 1993; Lind and Gailing 2013). For *Quercus rubra*, the expected heterozygosity (mean) was 0.71 at EST-SSRs and that of *Quercus georgiana* was 0.510 (only EST-SSRs were compared for both the species). This indicates the impact of small population size and high isolation on genetic diversity in *Quercus georgiana* and this data can be used in conservation planning of this endangered species.

Decisions should be made on the sampling strategy depending upon the level of genetic variation of this endangered species. Since the pairwise genetic differentiation between the populations is significant, sampling from all the populations should be done to incorporate all the genetic diversity for conservation. But considering the limited resources for *ex-situ* conservation strategy, the results of the pairwise genetic differentiation between the populations with higher values can be given priority for conservation purpose. For example, the populations PN, DK, MR, ED and WG should be given priority for conservation since they have high pairwise genetic differentiation values. The STRUCTURE results identify two genetic clusters (Fig. 2) that can be seen in the dendrogram result as well (Fig. 3). For example, the populations DK, CB, ED and WG are clustered together on both the dendrogram and the STRUCTURE diagram. These findings suggest that when building the *ex-situ* collections at least two populations should be sampled that will represent the two genetic clusters identified in STRUCTURE and dendrogram. Also, in the dendrogram, the populations showing bootstrap values higher than 50% (statistical support) should be prioritize for sampling i.e. populations DK, ED, WG and CB. AM population has the highest genetic distance (Fig. 3) to all populations and therefore future collections should target AM populations since it is genetically differentiated from the other populations.

Conclusion

The present studies utilized wild populations of *Quercus georgiana.* Genetic variation in wild populations can be compared to that of *ex-situ* collections so that effective conservation plans can be developed of this threatened species to support the *in-situ* conservation. Results from this study can be used to make decisions regarding the *ex-situ* conservation strategies in this species such that the entire gene pool is captured to maintain high genetic diversity. The *ex-situ* conservation can support the reintroduction and restoration strategies to enlarge wild plants in their current locations. Conversely, this study can also be a baseline for building and improving the collection of *Quercus georgiana* in arboretums and botanical gardens.

Tables

Table 1. Sampled *Quercus georgiana* populations.

Locality	County	State	Pop. Code	Sample size (n)	Latitude	Longitude
Davidson-Arabia Mountain Nature Preserve	DeKalb	GA	AM	24	33.667	-84.125
Chattahoochee Bend State Park	Coweta	GA	CB	24	33.421	-84.962
Concord Road	Pike	GA	CR	24	33.147	-84.464
Dowdell's Knob, F.F. Roosevelt State Park	Harris	GA	DK	24	32.84	-84.746
Eden	St. Claire	AL	ED	24	33.64	-86.369
Moss Rock Preserve	Jefferson	AL	MR	24	33.384	-86.841
Penton	Chambers	AL	PN	24	33.026	-85.483
Stone Mountain	DeKalb	GA	SM	23	33.809	-84.152
Walnut Grove	Walton	GA	WG	24	33.751	-83.832

Table 2. Description of genic EST-SSRs.

Locus	Repeat motif	Forward primer sequence (5'-3')	Reverse primer sequence (5'-3')	T_a (C°)	Size range (in base pairs)	Species	Linkage group[g]	Functional annotation [b]
FIR013	$(CAG)_5$	6-FAM CGGGGAGGTTGATGAGTATT	AACACTGTCACCCCCATAGC	56	133-144	*Q. robur*	2	constans-like protein COL (flowering time)
FIR039	$(CT)_7$	PET-GAGCCTCTTTCATCGCTCAC	TCAACACCCCAAAACTCCAT	59	111-132	*Q. robur*	1	histone deacetylase (drought stress)
FIR043	$(TC)_9$	PET-TTCTCCATTTCACACGCTTC	ACGACATCGTTTTGGAGCTT	56	114-146	*Q. robur*	7	EBP1 (regulates cell growth)
FIR048	$(CT)_9$	PET-TGCACCAAAATTGGAGGATG	TTGATGCAAGGTGCAGTTTC	56	187-219	*Q. robur*	2	Cell division protein ftsH, putative
GOT037	$(CT)_{11}$	PET-CCATCCTTTTCATTCTTTCCA	TGTTGTTGTTGCTGTTGTCG	57	239-265	*Q. robur*	5	Acid-phosphatase, putative
PIE039	$(CTT)_8$	6-FAM GTAAAACGACGGCCAGTGTCCTCACCCTCTGCGGTCT	CAGAAAGGGCTGCAAAGC	59	157-178	*Q. robur*	12	Unknown protein

Locus	Repeat motif	Forward primer sequence (5'-3')	Reverse primer sequence (5'-3')	T_a (C°)	Size range (in base pairs)	Species	Linkage group[g]	Functional annotation [b]
PIE200	$(CAA)_5$	6-FAM ACAACATGTGCC AAAACTGC	TCGATGATGTGGTTG TTGATG	56	107-119	*Q. robur*	Not assigned	Zinc finger a20 an1 domain-containing stress-associated protein 5-like
PIE125	$(GGAAGC)_3$	PET-AATACAAATCGC AGGAGGTG	CTAACCCATCGTTCA TGGAG	57	146-162	*Q. robur*	6	chaperone protein dna J, chloroplast, putative (heat shock protein, stress)
FIR035	$(AT)_6$	NED-GCTAAGGTTCCGT GTTCCAA	GGCCAGCAACTAAA CCAAGA	56	146-152	*Q. robur*	5	chaperone protein dna J, chloroplast, putative (heat shock protein, stress)
FIR028	$(TC)_8$	VIC-GGAAGAGTGTTC GGAAAGCA	CCAGCTCCTCCACAA TAGCA	56	201-237	*Q. robur*	1	tropinone reductase, putative (drought stress)
VIT081	$(CAT)_3$	6-FAM AATTCAAACCCA GCCAACTG	TCCTCTGGATGCTCC ATCA	56	112-136	*Q. robur*	Not assigned	Proline- rich protein
VIT086	$(CAG)_5$	VIC-AAGAACACCCAT TTCCACCA	TAAAATCCATTTGCC GGTTC	56	184-207	*Q. robur*	Not assigned	Chaperone protein dnaj chloroplast

Note. [b] Putative function determined through a BLASTx search following the method in (Luro et al., 2008), [g] on *Q. robur* linkage map in either Durand et al. (2010) or Bodénès et al. (2012). All EST-SSRs were originally developed for *Quercus robur* (Durand et al. 2010) and adapted for North American red oak species (section *Lobatae*) (Sullivan et al. 2013, Lind-Riehl et al. 2014).

Table 3. Mean genetic variation values for each marker across all samples.

	FIR039	FIR013	FIR048	FIR043	PIE039	GOT037	PIE200	FIR035	VIT081	VIT086	PIE125	FIR028
N_a	4.778	3.000	9.111	8.667	3.000	6.333	2.889	2.444	2.000	2.556	4.889	6.000
H_o	0.636	0.535	0.838	0.808	0.407	0.575	0.545	0.345	0.183	0.226	0.550	0.449
H_e	0.629	0.526	0.801	0.802	0.346	0.712	0.505	0.343	0.186	0.260	0.507	0.499
F	0.0111	-0.0171	-0.0461	0.0074	-0.1763	0.1924	-0.0792	0.0058	0.0161	0.1307	-0.0848	0.1002

N_a: number of different alleles, H_o: observed heterozygosity, H_e: expected heterozygosity, F (fixation index) = $(H_e-H_o)/H_e$.

Table 4. Mean genetic variation estimates for each population.

Population	N_a	H_o	H_e	F	No. of private alleles	No. of locally common alleles (≤25%)
DK	4.583	0.497	0.480	-0.035	0.000	0.417
CB	5.000	0.567	0.568	0.001	0.333	0.500
ED	5.000	0.525	0.516	-0.017	0.167	0.250
WG	4.333	0.499	0.506	0.013	0.083	0.333
PN	4.167	0.447	0.471	0.050	0.333	0.167
AM	5.333	0.507	0.574	0.116	0.167	0.250
SM	4.750	0.493	0.501	0.015	0.083	0.250
MR	4.667	0.490	0.479	0.023	0.083	0.167
CR	3.917	0.547	0.491	-0.114	0.083	0.250

Private alleles are alleles that are found only in a single population among a broader collection of populations whereas locally common alleles (≤ 25%) are present in ≤ 25% of populations.

Table 5. HWE exact test (Cockerham, 1984) by locus in each population.

DK				
Locus	**H_o**	**H_e**	**F**	**p-value**
FIR039	0.609	0.548	-0.1103	1.0000
FIR013	0.955	0.658	-0.4505	0.0237
FIR048	0.905	0.856	-0.0570	0.6833
FIR043	0.667	0.822	0.1890	0.0681
PIE039	0.273	0.247	-0.1046	1.0000
GOT037	0.636	0.536	-0.1869	0.5628
PIE200	0.333	0.497	0.3287	0.0845
FIR035	0.292	0.249	-0.1707	1.0000
VIT081	0.167	0.153	-0.0909	1.0000
VIT086	0.083	0.081	-0.0323	1.0000
PIE125	0.667	0.590	-0.1294	0.5708
FIR028	0.375	0.523	0.2824	0.1690
CB				
Locus	**H_o**	**H_e**	**F**	**p-value**
FIR039	0.591	0.510	-0.1579	0.7003
FIR013	0.478	0.521	0.0817	0.6770
FIR048	1.000	0.848	-0.1795	0.6867
FIR043	0.824	0.867	0.0499	0.0359
PIE039	0.750	0.735	-0.0201	0.1048
GOT037	0.708	0.767	0.0769	0.6560
PIE200	0.167	0.288	0.4217	0.0553
FIR035	0.500	0.469	-0.0667	1.0000
VIT081	0.208	0.249	0.1638	0.3976
VIT086	0.500	0.452	-0.1056	0.8193
PIE125	0.542	0.596	0.0917	0.0058
FIR028	0.542	0.558	0.0295	0.5152
ED				

Locus	H_o	H_e	F	p-value
FIR039	0.714	0.687	-0.0396	0.5387
FIR013	0.636	0.563	-0.1303	0.8851
FIR048	0.913	0.774	-0.1795	0.6923
FIR043	0.750	0.866	0.1343	0.2234
PIE039	0.708	0.484	-0.4650	0.0602
GOT037	0.458	0.589	0.2212	0.2663
PIE200	0.391	0.446	0.1229	0.4628
FIR035	0.333	0.278	-0.2000	1.0000
VIT081	0.042	0.041	-0.0213	-
VIT086	0.083	0.155	0.4637	0.1276
PIE125	0.750	0.629	-0.1917	0.8439
FIR028	0.522	0.679	**0.2312**	0.0000

WG

Locus	H_o	H_e	F	p-value
FIR039	0.632	0.633	0.0022	0.4249
FIR013	0.609	0.499	-0.2197	0.4212
FIR048	0.864	0.811	-0.0650	0.5446
FIR043	0.833	0.839	0.0072	0.1420
PIE039	0.417	0.359	-0.1622	1.0000
GOT037	0.591	0.707	0.1637	0.2825
PIE200	0.292	0.353	0.1744	0.5533
FIR035	0.208	0.187	-0.1163	1.0000
VIT081*	0.043	0.043	-0.0222	-
VIT086	0.500	0.477	-0.0492	0.2785
PIE125	0.583	0.648	0.0992	0.3547
FIR028	0.417	0.519	0.1973	0.0400

PN

Locus	H_o	H_e	F	p-value
FIR039	0.750	0.722	-0.0385	0.9295
FIR013	0.458	0.580	0.2096	0.2754

FIR048	0.458	0.575	0.2024	0.0818
FIR043	0.708	0.778	0.0893	0.6377
PIE039	0.125	0.119	-0.0511	1.0000
GOT037	0.667	0.732	0.0890	0.0689
PIE200	0.833	0.611	-0.3636	0.0056
FIR035	0.375	0.305	-0.2308	0.5512
VIT081	0.333	0.424	0.2147	0.4963
VIT086	0.238	0.427	0.4430	0.0495
PIE125	0.250	0.226	-0.1077	1.0000
FIR028	0.167	0.156	-0.0667	1.0000

AM

Locus	**H_o**	**H_e**	**F**	**p-value**
FIR039	0.591	0.685	0.1373	0.0932
FIR013	0.417	0.444	0.0625	1.0000
FIR048	0.739	0.839	0.1194	0.0513
FIR043	0.833	0.882	0.0551	0.3658
PIE039	0.500	0.404	-0.2379	0.7764
GOT037	0.318	0.793	**0.5990**	0.0000
PIE200	0.750	0.659	-0.1383	0.0530
FIR035	0.292	0.415	0.2971	0.0383
VIT081	0.292	0.249	-0.1707	1.0000
VIT086	0.125	0.117	-0.0667	1.0000
PIE125	0.667	0.659	-0.0119	0.4492
FIR028	0.565	0.744	0.2402	0.0185

SM

Locus	**H_o**	**H_e**	**F**	**p-value**
FIR039	0.591	0.628	0.0592	0.4391
FIR013	0.348	0.468	0.2566	0.3340
FIR048	0.826	0.845	0.0224	0.3878
FIR043	0.783	0.830	0.0569	0.0552
PIE039	0.348	0.287	-0.2105	1.0000

GOT037	0.524	0.706	0.2584	0.0334
PIE200	0.636	0.533	-0.1938	0.5766
FIR035	0.435	0.386	-0.1275	1.0000
VIT081	0.478	0.364	-0.3143	0.2742
VIT086	0.130	0.198	0.3397	0.2128
PIE125	0.364	0.319	-0.1392	1.0000
FIR028	0.455	0.452	-0.0046	0.2500

MR

Locus	**H_o**	**H_e**	**F**	**p-value**
FIR039	0.583	0.574	-0.0166	0.5846
FIR013	0.375	0.471	0.2044	0.4518
FIR048	0.917	0.839	-0.0932	0.8283
FIR043	0.875	0.779	-0.1237	0.2438
PIE039	0.000	0.000	-	-
GOT037	0.667	0.782	0.1476	0.0435
PIE200	0.667	0.530	-0.2570	0.3371
FIR035	0.500	0.584	0.1441	0.4440
VIT081	0.000	0.000	-	-
VIT086	0.000	0.080	1.0000	0.0210
PIE125	0.667	0.560	-0.1907	0.6089
FIR028	0.625	0.551	-0.1339	1.0000

CR

Locus	**H_o**	**H_e**	**F**	**p-value**
FIR039	0.667	0.669	0.0039	0.2623
FIR013	0.542	0.574	0.0560	0.8198
FIR048	0.917	0.819	-0.1186	0.5243
FIR043	1.000	0.556	**-0.7972**	0.0000
PIE039	0.542	0.478	-0.1325	0.8896
GOT037	0.609	0.796	0.2352	0.0644
PIE200	0.833	0.625	**-0.3333**	0.0017
FIR035	0.167	0.219	0.2381	0.2987

VIT081	0.083	0.153	0.4545	0.1283
VIT086	0.375	0.353	-0.0614	1.0000
PIE125	0.458	0.380	-0.2055	1.0000
FIR028	0.375	0.312	-0.2033	1.0000

Significant values in boldface type are significantly different from Hardy-Weinberg proportions (α=0.05) after Bonferroni correction ($p<0.05/12=0.00416$). *: no p-value was calculated for this marker with very low variation within population WG.

Table 6. Pairwise genetic distances (F_{ST}) between populations.

Population	DK	CB	ED	WG	PN	AM	SM	MR	CR
DK	0.000								
CB	0.054	0.000							
ED	0.028	0.048	0.000						
WG	0.028	0.038	0.035	0.000					
PN	0.129	0.094	0.125	0.111	0.000				
AM	0.075	0.065	0.065	0.079	0.058	0.000			
SM	0.087	0.066	0.077	0.092	0.046	0.025	0.000		
MR	0.115	0.087	0.102	0.115	0.083	0.049	0.049	0.000	
CR	0.083	0.060	0.077	0.062	0.050	0.039	0.034	0.070	0.000

All values are significant ($p<0.001$, exact G test).

Table 7. Pairwise geographic distance in kilometers between populations.

Population	DK	CB	ED	WG	PN	AM	SM	MR	CR
DK	0.000								
CB	82.39	0.000							
ED	176.3	152.07	0.000						
WG	162.98	103.21	245.81	0.000					
PN	57.22	97.5	138.03	200.52	0.000				
AM	134.97	71.74	212.13	38.7	167.92	0.000			
SM	147.35	76.76	203.73	43.36	172.56	13.91	0.000		
MR	285.16	314.88	199.45	420.3	233.94	389.52	388.27	0.000	
CR	61.99	50.84	192.92	101.51	105.01	75.42	91.57	339.27	0.000

Figures

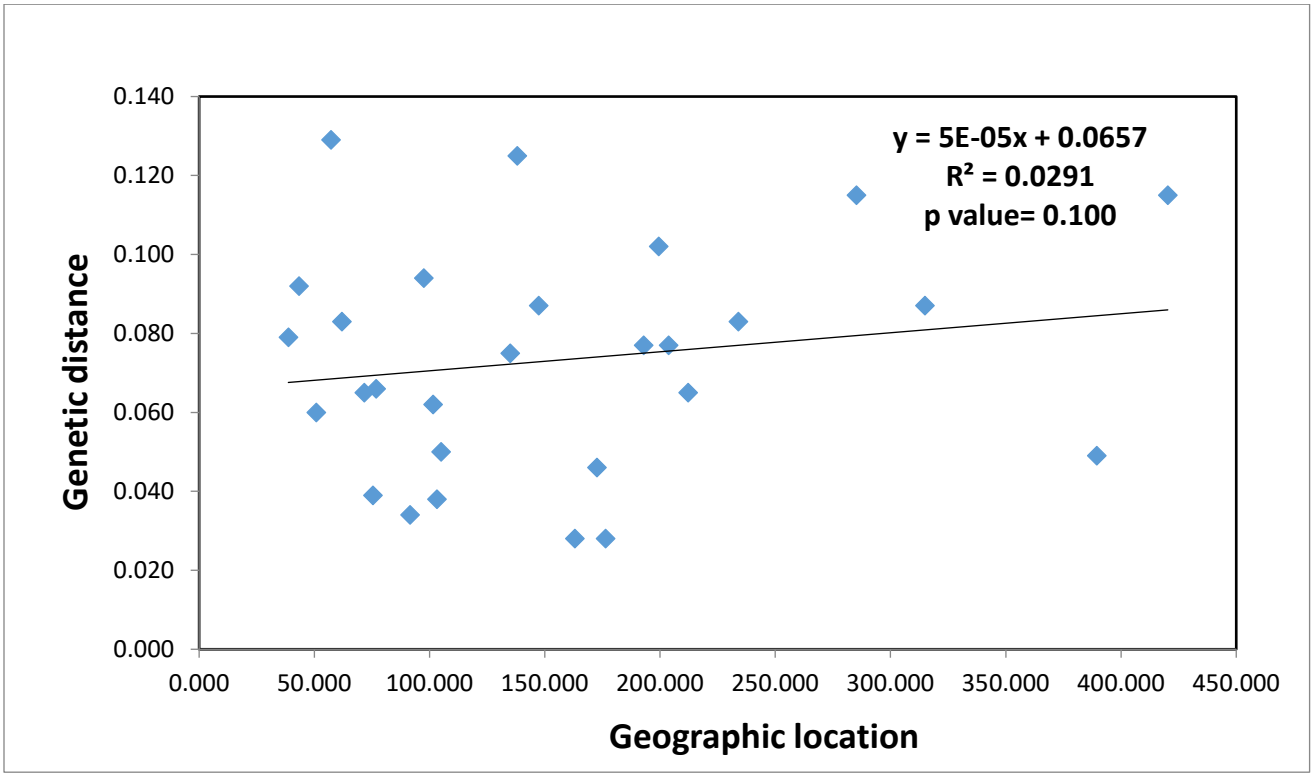

Figure 1. Geographic distance (measured in kilometers) versus genetic distance for pairwise comparisons of nine populations of *Quercus georgiana* using a Mantel test.

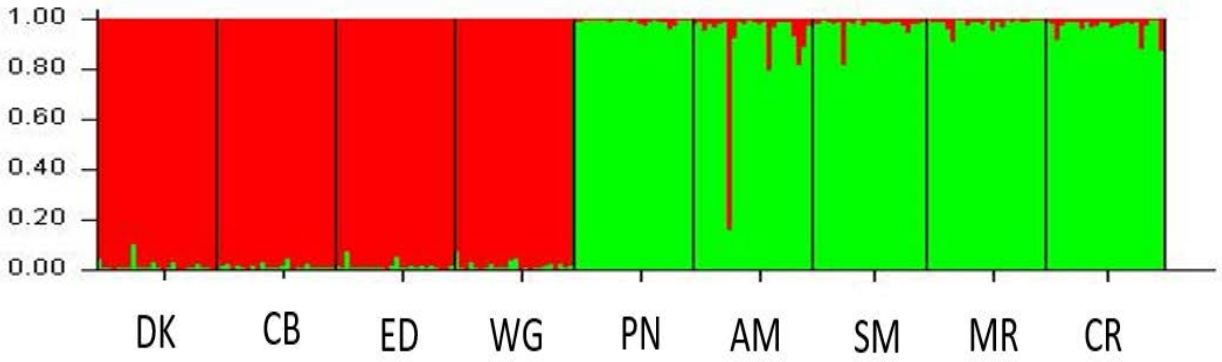

Figure 2. Genetic assignment of individuals using the Bayesian method in STRUCTURE 2.3.4 (Pritchard et al. 2000). Individuals are grouped by population.

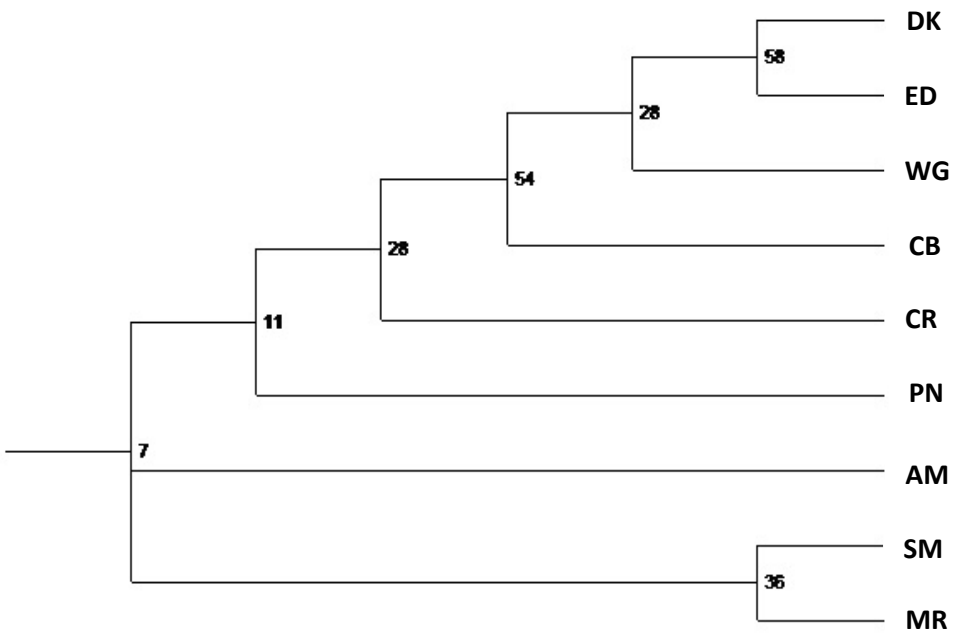

Figure 3. Dendrogram (Neighbor joining method) based on Cavalli-Sforza and Edwards (1967) genetic distances at EST-SSRs; numbers at nodes are percentages over 1,000 bootstrap replicates using Populations 2.0 (Langella 1999).

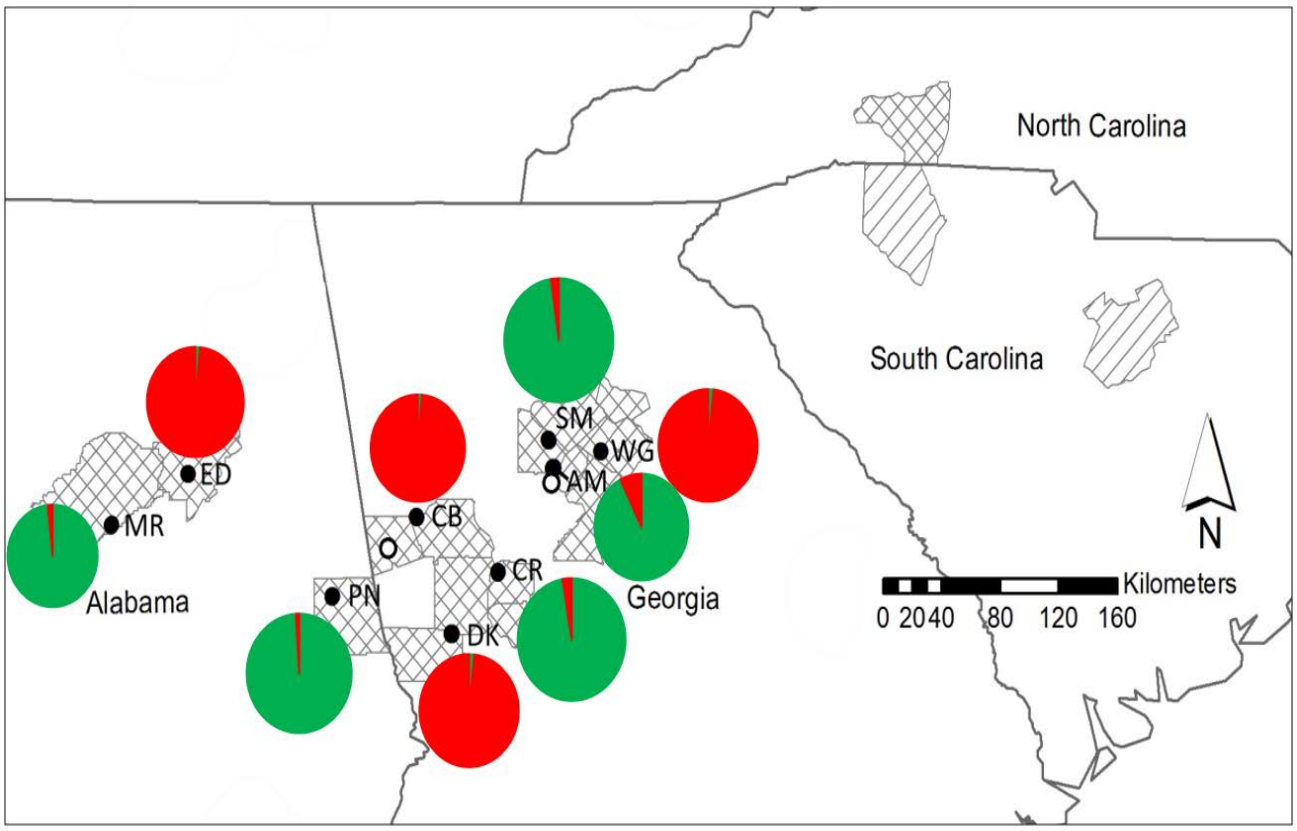

Figure 4. Geographic distribution of percentage of ancestry per population as inferred in STRUCTURE. Geographic distribution of genetic cluster 1 (red) and cluster 2 (green).

References

Backs, R. J., Terry, M. and Ashley, M. 2016. Using genetic analysis to evaluate hybridization as a conservation concern for the threatened species *Quercus hinckleyi* C. H. Muller (Fagaceae). *Int. J. Plant Sci* 177(2): 122-131.

Bodenes, C., Chancerel, E., Gailing, O., Vendramin, G.G., Bagnoli, F., Durand, J., Goicoechea, P.G., Soliani, C., Villani, F., Mattioni, C., Koelewijn, H.P., Murat, F., Salse, J., Roussel, G., Boury, C., Alberto, F., Kremer, A. and Plomion, C. 2012. Comparative mapping in the Fagaceae and beyond with EST-SSRs. *BMC Plant Biology* 12:153.

Cockerham, B. S. W. a. C. C. 1984. Estimating F-Statistics for the analysis of population structure. *Evolution* 38(6): 1358-1370.

Conesa, A., Gotz, S., Garcia-Gomez, J. M., Terol, J., Talon, M. and Robles, M. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21: 3674-3676.

Durand, J., Bodenes, C., Chancerel, E., Frigerio, J-M., Vendramin, G., Sebastiani, F., Buonamici, A., Gailing, O., Koelewijn, H-P., Villani, F., Mattioni, C., Cherubini, M., Goicoechea, P.G., Herran, A., Ikaran, Z., Cabane, C., Ueno, S., Alberto, F., Dumoulin, P-Y., Guichoux, E., de Daruvar, A., Kremer, A. and Plomion, C. 2010. A fast and cost-effective approach to develop and map EST-SSR markers: oak as a case study. *BMC Genomics* 11:570.

Ellis, J. R. and Burke, J. M. 2007. EST-SSRs as a resource for population genetic analyses. *Heredity (Edinb)*, 99**:** 125-32.

Keiper, F. J., Hayden, M.J., Park, R. F. and Wellings, C. R. 2003. Molecular genetic variability of Australian isolates of five cereal rust pathogens. *Mycological Research* 107: 545-556.

Kramer, A., A. Hird, K. Shaw, M. Dosmann, and R. Mims. 2011. Conserving North America's Threatened Plants: Progress Report on Target 8 of the Global Strategy for Plant Conservation: Botanic Gardens Conservation International U.S.

Langella, O. 1999. Populations 1.2.30: a population genetic software.

Ledig, F. T. 1988. The Conservation of Diversity in Forest Trees. *Bioscience* 38 No 7: 471-473.

Lind, J. and Gailing, O. 2013. Genetic structure of *Quercus rubra L*. and *Quercus ellipsoidalis* E. J. Hill populations at gene-based EST-SSR and nuclear SSR markers. *Tree Genetics & Genomes* 9:707–722.

Lind-Riehl, J., Sullivan, A. and Gailing, O. 2014. Evidence for selection on a CONSTANS-like gene between two red oak species. *Annals of Botany* 113: 967-975.

Luro, F. L., Costantino, G., Terol, J., Argout, X., Allario, T., Wincker, P., Morillon, R. 2008. Transferability of the EST-SSRs developed on Nules clementine (Citrus clementina Hort ex Tan) to other Citrus species and their effectiveness for genetic mapping. *BMC Genomics* 9: 287.

Nei, M. 1973. Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences USA* 70: 3321-3323.

Oldfield, S. and A. Eastwood. 2007. *The Red List of Oaks*. Fauna & Flora International, Cambridge, UK.

Peakall, R. and Smouse, P.E. 2006. GENEALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* 6:288-295.

Pence, V. 2012. The possibilities and challenges of *in vitro* methods for plant conservation. *Kew Bulletin* 64.

Pritchard, J.K., Stephens, M. and Donnelly, P. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155:945-959.

Raymond, M. and Rousset, F. 1995. GENEPOP (Version 1.2): population genetics software for exact tests and ecumenicism. *J Hered* 86:248-249.

Reed, D.H., and R. Frankham. 2003. Correlation between fitness and genetic diversity. *Conservation Biology*. 17(1):230-237.

Rousset, F. 2008. Genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources* 8:103-106.

Sertse, D., Gailing, O., Eliades, N. and Finkeldey, R. 2013. Transferability and application of microsatellites (SSRs) from *Juniperus communis* L. to *Juniperus procera Hochst*. Ex endl. *Open Journal of Genetics* 3: 115-126.

Sork, V., Huang, S. and Wiener, E. 1993. Macrogeographic and fine-scale genetic structure in a North American oak species, *Quercus rubra* L. *Ann For Sci* 50:261–270.

Steferon, V. M., Gailing, O. and Finkeldey, R. 2008. The role of gene flow in shaping genetic structures of the subtropical conifer species *Araucaria angustifolia*. *Plant Biology* 10: 356-364.

Sullivan, A. R., Lind, J. F., McCleary, T. S., Romero-Severson, J., and Gailing, O. 2013. Development and characterization of genomic and gene-based microsatellite markers in North American red oak species. *Plant Molecular Biology Reporter* 31: 231-239.

Toppila, R. 2012. *Ex-situ* conservation of oak (*Quercus L.*) in botanic gardens: A North American perspective. Master of Science thesis, University of Delaware.

Vazquez, D., and Gittleman,J. 1998. Biodiversity conservation: Does phylogeny matter? *Current Biology* 8, no. 11.

Wenzell, K. & Kenny, L. 2015. *Quercus georgiana*. The IUCN Red List of Threatened Species 2015.

Whitlock, R., Hipperson, H., Thompson, D.B.A., Butlin, R.K. and Burke, T. 2016. Consequences of *in-situ* strategies for the conservation of plant. *Biological Conservation* 203: 134-142.

Wright, S. 1965. The interpretation of population structure by F-statistics with special regards to systems of mating. *Evolution* 19: 395-420.

http://globaltrees.org/threatened-trees/trees/quercus-georgiana/