THE GENETIC DIVERSITY OF EASTERN NORTH AMERICAN TREE SPECIES AS RELATED TO THE EXTENT OF THEIR SPATIAL DISTRIBUTION DURING THE LAST GLACIAL MAXIMUM

by

Robin J. P. Timms

(Carson 2015)

FACULTY OF NATURAL RESOURCES MANAGEMENT
LAKEHEAD UNIVERSITY
THUNDER BAY, ONTARIO

April 2020

THE GENETIC DIVERSITY OF EASTERN NORTH AMERICAN TREE SPECIES
AS RELATED TO THE EXTENT OF THEIR SPATIAL DISTRIBUTION DURING
THE LAST GLACIAL MAXIMUM

by

Robin J. P. Timms

An Undergraduate Thesis Submitted in Partial Fulfillment of the Requirements for the
Degree of Honours Bachelor of Science in Forestry

Faculty of Natural Resources Management

Lakehead University

April 2020

Dr. Ashley Thomson	Dr. Jian Wang
Major Advisor	Second Reader

LIBRARY RIGHTS STATEMENT

In presenting this thesis in partial fulfillment of the requirements for the HBScF degree at Lakehead University in Thunder Bay, I agree that the University will make it freely available for inspection.

This thesis is made available by my authority solely for the purpose of private study and may not be copied or reproduced in whole or in part (except as permitted by the Copyright Laws) without my written authority.

Signature: _____

Date: _____

A CAUTION TO THE READER

This HBScF thesis has been through a semi-formal process of review and comment by at least two faculty members. It is made available for loan by the Faculty of Natural Resources Management for the purpose of advancing the practice of professional and scientific forestry.

The reader should be aware that opinions and conclusions expressed in this document are those of the student and do not necessarily reflect the opinions of the thesis supervisor, the faculty or of Lakehead University.

MAJOR ADVISOR COMMENTS

ABSTRACT

Timms, R. J. P. 2020. The genetic diversity of eastern North American tree species as related to the extent of their spatial distribution during the Last Glacial Maximum. 63 pp.

Key Words: allozyme, deciduous, glacial refugia last glacial maximum, paleoecology, phylogeography, species distribution models

The distribution of boreal and temperate species during the last glacial maximum (LGM) in Eastern North America can have a significant effect on the genetic diversity of present day populations. Here, I modelled the LGM distribution of several keystone deciduous tree species native to northeastern North America. The ecological niche modelling software, Maxent, was used to generate a relative rate of occurrence that was used to infer the probable locations of glacial refugia. These predicted refugia were then validated by paleo-ecological records where available. The genetic variation of each species was assessed by a literature review of previous studies that used allozymes to determine the expected heterozygosity. The area of the refugia were then compared to the expected heterozygosity through a linear regression. The findings of this study indicated no significant relationship ($r^2 = 0.024$) between refugia size and expected heterozygosity. The lack of data on expected heterozygosity likely skewed the results and further studies need to be conducted to correct for this.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

INTRODUCTION

North America has seen very significant glacial events throughout geological history. The most recent maximum ended roughly 18000 years ago, with the retreat of the glaciers leaving permanent marks on the land clearly seen today (Yokoyama et al. 2001). Less clearly seen are the effects on the genetic diversity of surviving tree species that had to recolonize the post glacial tundra-like environments. Temperate tree species persisted during the LGM in pockets of suitable habitat, known as refugia, that helped enable their survival. The size and extent of these refugia can be analyzed through allozymic variation today, along with habitat modelling from paleoclimate data.

Studies that examine fossilized pollen data have been used to reconstruct previous landscapes and species distributions along with the speed of post glacial migration and will aid in the accuracy of the models to be constructed (Gaillard et al. 2008). The western half of North America has had significant studies of glacial refugia and modern genetic diversity (Roberts and Hamann 2015), while the eastern half of North America is lacking in a study to this degree.

The relationship between size of refugia and genetic diversity was proven to be significant in the study by Roberts and Hamann (2015). Moreover, the extent of this relationship was seen in a study by Nei et al. (1975) which found that genetic diversity declines rapidly when a population bottleneck occurs. Expected heterozygosity (He), a measure of genetic diversity, is related to the frequency of common alleles, which are unlikely to be lost due to small population sizes (Nei et al. 1975; as cited by Roberts and Hamann 2015). The genetic differentiation of past populations is favoured in refugia that

are of greater size and distinctly separated, which manifests as higher levels of He in modern populations (Roberts and Hamann 2015).

The objective of this study is to reconstruct the LGM distribution of ten deciduous tree species populations that underwent significant range shifts during the last glacial maximum (LGM) to examine the relationship between present day genetic diversity and LGM range size. The species that will be examined for this study are; *Acer saccharum, Quercus rubra, Quercus macrocarpa, Quercus alba, Quercus velutina, Populus tremuloides, Populus grandidentata, Populus balsamifera, Fagus grandifolia, Castanea dentata.*

SPECIES SELECTION

The selection of species for this study was conducted first on the genus level to compare as many closely related species as possible. The next selection requirement was current habitat relationships, and cohabitation of similar areas. The lack, or presence, of allozyme studies was also a guiding factor in determining species.

One of the goals of species selection was including a diverse set of present-day ranges, which would allow for comparisons to be made. The majority of species selected had similar range sizes that are restricted to Eastern North America. *Populus tremuloides* and *Populus balsamifera* were selected because they had the largest present-day ranges and are more common in the boreal forest. *Populus balsamifera* has a more limited range than *P. tremuloides*, and *P. balsamifera* and was selected as a comparison. *Acer saccharum* and *Fagus grandifolia* were selected because they share a common forest

type today. The *Quercus* genus was selected due to the availability of literature and their relationship to other selected species.

Castanea dentata was selected due to its widespread distribution in Eastern North America before populations were decimated by the chestnut blight in the late 19th century (Dalgleish et al. 2015). The current population of C. dentata is estimated to be 10% of its pre-blight size (Dalgleish et al. 2015). This reduction of population size made *C. dentata* the focus of several studies aiming to aid in conservation and breeding programs (Huang et al. 1999). Furthermore, the historical relationship between chestnut-oak forests is well documented (Huang et al. 1999).

HYPOTHESIS:

Due to the complex nature of this thesis, and the number of species, a simple hypothesis will be hard to predict. A null hypothesis can be assumed to be that there is no significant relationship between present day genetic diversity and the number or size of the genetic refugia that species occupied during the LGM. However, I expect that tree species that are more widely distributed today, and that have a tolerance for more northerly ranges likely faired better in the past. The species that favoured northern ranges more, possibly held large populations close to the glaciers and recolonized more quickly. Specifically, species that are confined to small widely separated refugia will likely have less genetic diversity and share more similar alleles. This is compared to species that had larger more continuous refugia, and consequently retained a higher genetic diversity, and a higher number of different alleles in the population.

LITERATURE REVIEW

PHYLOGEOGRAPHIC EFFECTS ON GENETIC DIVERSITY

A central aspect of this study is the phylogeographic effects on genetic diversity, which will be focused on the last glacial maximum or LGM. In North America, as glaciation moved from north to south, many tree species were pushed out of much of their normal ranges and were forced to survive in glacial refugia in the south (Hollingsworth 1999). A refugium can be defined as "an area, of any size, in which a taxon persisted at any population density, during a cold stage" (Bennet et al. as cited by Hollingsworth et al. 1999). Being restricted to glacial refugia can have a drastic effect on genetic diversity, even more so if the refugia is isolated, with little to no gene flow (Hollingsworth et al. 1999). If isolation, and glaciation, continues for a significant amount of time, then population allele frequencies begin to diverge.

A book by Hollingsworth et al. (1999) focuses on *Quercus* and *Fagus* species of Europe but provides a proof of concept that refugia can be measured through genetic diversity of varying populations to determine where the refugia were and their relative size. Species in Europe with a more northerly range, such as *Betula*, were shown to have had a much larger and more continuous glacial refugium (Huntley and Birks as cited by Hollingsworth et al. 1999). The overall result of this isolation and glacial refugium is speciation in the long term and genetic divergence in the short term (Hollingsworth 1999).

One of the only ways to map a species refugia is by the use of fossil pollen, and macro fossil data. Many studies have used this method, and one example is a study by

Martínez-Meyer and Peterson (2006), which investigated the modelled distribution of tree species niches over the last 20,000 years. This is done by tracking distribution and migration of multiple pollen taxa relative to climate types. The use of the Global Pollen Database was essential to the ecological niche models used. The pollen data was then incorporated into ecological niches summaries, of present day, and paleo, climatology summaries to produce likely geographically distributions. Finally, the outputs from each time period were projected to the "other" time period. The species that this study has in common with this paper is only *Acer saccharum*. However, the approaches used in that study can be applied to the other species used in this study. The problems of these models are that they can only show the predicted ecological niches after the LGM and they do not take into account tree migration. Moreover, the use of fossil pollen data relies on environmental factors that are able to collect and retain pollen over the 20,000-year time period. Many fossil pollen survey locations do not indicate the occurrence of species that may actually have occurred there, because pollen and macrostructures are not always successfully preserved in the fossil record.

SPECIES GENETIC DIVERSITY AND GLACIAL REFUGIA

A similar study to this thesis was conducted on the relationship between genetic diversity and geographic refugia following the LGM for 22 species located in Western North America (Roberts and Hamann 2015). This study found a surprisingly strong relationship between the size of proposed genetic refugia and modern genetic diversity. This is a good indicator that the hypothesis of this thesis is on the right track.

Acer saccharum (Sugar Maple)

  Acer saccharum is a shade tolerant tree which has an average life span of 200 –
300 years. The species is wind and insect pollinated, along with the seeds that are wind
dispersed. These two attributes combine to allow the species to have a medium dispersal
ability A study by Perry and Knowles (1989) looked into the allozyme variation of
multiple populations of sugar maple from two distinct areas near the northern limit of
their range, Wawa and Thunder Bay. This study found that the F_{st} estimates (average
pairwise population differentiation, as measured by allele frequencies) indicated only
3% genetic variability among populations, with no relation to geographic distance
suggested. The study then goes on to suggest that the small populations of sugar maple
might have been remnants of a larger more continuous population from the hypsithermal
interval. This is valuable information when looking at the post glacial state of the
population, suggesting that the species was able to colonize more northerly limits. The
average expected heterozygosity was moderate, with a value of He = 0.110. This can be
compared to two other northern hardwood species, being greater than *Populus
balsamifera* (He=0.037, Bousquet et al.), while lower than *Populus tremuloides* (He =
0.42, Cheliak and Dancik; He = 0.235, Hyun et al.) (as cited by Perry and Knowles
1989).

  Another study found that *Acer saccharum* exhibits slightly lower genetic
diversity than the average of other North American angiosperms, for example *Populus
tremuloides* (Hyun et al. cited by Young et al. 1993) had a distinct genetic structure
among regional populations across their range (Young et al. 1993). This study goes into
further detail by interpreting two broad geographic groups of populations, as

descendants of two refugia from the after the last glacial maximum, when tree migration was occurring (Young et al. 1993). These groups likely originated from: Appalachian refugia, which migrated north into the east coast (Dyke and Prest cited by Young et al. 1993), and from a refugium in the Driftless region of Iowa which populates much of Ontario and Quebec now (Braun 1950; as cited by Young et al. 1993).

While the previous studies looked at genetic diversity, none of them related it to the extent of refugia during the LGM. A study by Graignic et al. (2018) suggests that there is a genetic difference between eastern and western populations, which may possibly represent two distinct migration routes that either separated a significant time ago or different glacial refugia. These populations were in eastern and western Quebec, with the western population arriving 3000 years later than the eastern population. Multiple migration routs could indicate two separate refugia.

Fagus grandifolia (American Beech)

American beech is a widely spread, highly shade tolerant species that inhabits many of the same forest types as sugar maple. This relationship today could potentially indicate a similar relationship as far back as the LGM. Knowing this relationship, a study by Mclachlan et al. (2005) focused on two the genus; maple (*Acer*) and beech (*Fagus*), with no specific species in either genus quantified. This is because of the studies of fossil pollen data not differentiating between species, due to the difficulty (Mclachlan et al. 2005). *Fagus grandifolia* is the only species of the *Fagus* found in North America (Bennet 1985). The Gulf Coastal Plain is identified as the refuge of beech during the last glacial maximum (Davis, Delcourt and Delcourt as cited by Mclachlan et al. 2005). The migration of beech from this refugia proceeded north

following a path seen in the appendices (Mclachlan et al. 2005). While sugar maple (*Acer saccharum*) is not mentioned in the study by Mclachlan et al. (2005), Young et al. (1993) proposed two distinct refugia.

A study by Bennet (1985) investigates the spread of *Fagus grandifolia* by mapping and interpreting the data from pollen and macrofossils. Between the years 18,000 and 14,000 years before present (bp), the pollen data suggests *F. grandifolia* was present over a wide area in the south (Tennesee and Louisiana), although it was generally very scarce (Bennet 1985). There was believed to be a full glacial refugium in the Tunica Hills and the bluff lands of the Mississippi valley from the Gulf Coast to as far north as Tennessee (Delcourt and Delcourt as cited by Bennett 1985). This would indicate one, albeit widely spread, refugia in an area that is also shown by Mclachlan et al. (2005).

Given the previous studies into glacial refugia, the genetic diversity of *F. grandifolia* needs to be quantified for the purpose of this thesis. A study by Houston and Houston (1994) surveys the genetic structure of two stands located in Massachusetts and West Virginia, by examining the isozyme variation at nine loci. The specific genetic variance data is located in the material and methods section. The study by Houston and Houston (1994) found that the local genetic diversity in each respective population was high, with an average observed heterozygosity of 0.382, and an average number of alleles per locus of 2.9. The mean Fst value of 0.064 indicated a moderate differentiation between the two populations (Houston and Houston 1994). Due to the relative closeness of the sample areas these populations could potentially be from the same refugia during

the LGM. Other studies will have to be used to prove or disprove this point. Also, this expected heterozygosity is quite high compared to the other species in this study.

Populus spp.

In the *Populus* genus, the multiple species that will be investigated are, *P. tremuloides, P. baslamifera* and *P. grandidentata*. Studies of genetic diversity based on allozymes will be compiled along with studies that indicate likely glacial refugia. Not all species had studies that indicated phylogeography and models will have to be made to find likely glacial refugia.

Trembling aspen (*Populus tremuloides*) is the most widely distributed tree species in North America (Perala as cited by Liu and Furnier 1993). While big tooth aspen (*Populus grandidentata*) range is located to Eastern North America, with most populations being found in Canada (southwestern Quebec and southeastern Ontario), but extending down to northwestern Tennessee (Laidly 1990). A study by Liu and Furnier (1993) examines the genetic variation in allozyme loci, RFLPs and RAPDs, in 130 trembling aspen, and 105 big tooth aspen trees. With regards to the allozyme loci, the results were, P=77%; A=2.8; He = 0.25, for trembling aspen, and P=29%; A=1.4; He = 0.08, for bigtooth aspen (Liu and Furnier 1993). The sample populations included 130 trembling aspen and 105 big tooth aspen, located in Michigan, Minnesota and Wisconsin (Liu and Furnier 1993). The results indicate that trembling aspen is more genetically variable than bigtooth aspen at both allozymic and RAPD markers. This could be linked to the difference in ranges of the species as stated above.

A study by Callahan et al. (2013), found two major genetic clusters for *Populus tremuloides*; a south western cluster and a northern cluster. The south western cluster

had two sub clusters but is limited in its current distribution to the Rocky Mountains area (Callahan et al. 2013). The northern cluster had no sub clusters determined, but this could be a failing of the broad scope of the study, with Callahan et al. (2013) recommending a finer scale study to determine if more distinct genetic refugia existed during the LGM. This could potentially indicate that the northern cluster was a series of refugium that were closely situated and had a continuous genetic exchange. Another study, by Ding et al. (2017) constructed a historical biogeography of *Populus tremuloides*, through the use of species distribution modelling and data on genetic structure and within-population variance. The findings suggest three potential glacial refugia; the eastern United States refugia, the south western United States refugia, and the Alaska refugium (Ding et al. 2017). The Alaska refugium is indicated to have been small with a low probability of presence and will not be considered as significant in this thesis (Ding et al. 2017). The eastern refugium was the largest and is credited as the most likely source of migration north into the current species distribution (Ding et al. 2017). This idea can be related to the study Callahan et al. (2013), which also suggests a distinct genetic cluster responsible for the current northern distribution of trembling aspen.

Populus balsamifera offers a problem with this thesis, being that only one study measures the genetic variation. This was a study by Farmer et al. (1988), which measured the isozyme variation of *Populus balsamifera* by sampling five 50 tree populations along a latitudinal transect in Northern Ontario. The overall findings indicate that of the populations surveyed, there was a lack of genetic differentiation (Farmer et al. 1988). Farmer et al. (1988) suggests that this could be the result of a

common ancestral gene pool, and substantial gene flow between populations due to recent migration. The idea of a common ancestral gene pool is corroborated by the multiple genetic clusters found by Keller et al. (2011). This is a small population that, in no way, represents the large range of *P. balsamifera*.

The phylogenetic findings of a study by Keller et al. (2011) are three regional clusters of genetic diversity: a northern cluster found in Alaska and Northern Canada; a central cluster distributed from the Great Lakes to the prairie provinces; and an eastern cluster found in Quebec and the maritime provinces. These findings can be used with modelling to be done on this thesis to help determine if the three genetic clusters can be attributed to three glacial refugia.

Quercus spp.

The *Quercus* genus is a diverse set of species located across the world, but this thesis will focus on the following species; *Quercus rubra*, *Quercus macrocarpa*, *Quercus alba*, and *Quercus velutina*. Of this species list *Q. rubra* and *Q. macrocarpa* had the largest quantity of genetic data, while *Q. alba*, and *Q. velutina* had very little.

A study by Sork et al. (1993), investigates two aspects of genetic variation in *Quercus rubra*; macrogeographic and fine-scale. This was done by sampling multiple locations for the former, and intensely sampling four sub-populations within a 4 ha plot in Missouri, for the latter. The macrogeographic study of genetic variation is of key importance to this thesis. The findings of this aspect of the study indicate a relatively high level of genetic variation within individual populations and a higher than average genetic differentiation between populations (Sork et al, 1993). This genetic differentiation is linked loosely to the idea of glacial history, including bottlenecks,

11

genetic drift, and uneven migration patterns (Schlarbaum et al. as cited by Sork et al. 1993). While this study does not go into depth into the cause of differentiation, future ecological niche modelling, coupled with the genetic variation findings could be used to determine the likelihood of genetic refugia for the red oak species.

Due to the lack of allozyme studies, other literature needed to be gathered to aid in the description of genetic diversity. A study by Magni et al. (2005) investigates the DNA variation of *Q. rubra* across the species range and compares this to other *Fagaceae*. The genetic data of this study is different compared to the allozymic variation of other studies used in this thesis. However, the information can be useful when compared with other data gathered. The general findings indicate a relatively low population differentiation pointing to a weak phylogeographical structure (Magni et al. 2005). This could be a result of a limited number of genetic refugia, but this study does not elaborate on this point, and future modelling will have to be done to prove the validity of this idea.

The only allozymic genetic findings of *Q. velutina* and *Q. alba*, were carried out in a study by Manos & Fairbrothers (1987). The notable species of this study that can be related to this thesis are; *Q. velutina, Q. alba*, and *Q. rubra*. Due to the limited range of the study these values might not represent the entire range accurately and should be supported with findings of other studies, if they are available, to prove their validity.

The other species with a solid literary backing into their expected heterozygosity was *Q. macrocarpa*. A study by Geburek and Tripp-Knowles (1994) examined the genetic architecture of *Q. macrocarpa*, through the use of allozyme variation. The study focused on sampling an isolated population of bur oak near Rainy River and the fine-

scale genetic structure of this population (Geburek and Tripp-Knowles 1994). Due to the limited nature of this, sampling this study will be used as a point of data to compare with and add to data gathered by other studies. One such comparison is a study by Schnable and Hamrick (1990), which examines the population genetic structure of 21 populations of *Quercus macrocarpa*, along with four populations of *Quercus gambelii*.

Overall the literature contains few broad-scale studies of the genetic diversity of eastern North American *Quercus* spp.. However, studies were available for a few species, though a number of them relied on populations that likely are not representative of the entire species range. There was little to no information about likely refugia of any specific *Quercus* species, and only a general look at the *Quercus* genus likely paleohabitat in a study by Prentice et al. (1991).

Castanea dentata (American Chestnut)

The American chestnut has still not recovered from chestnut blight and is in a continually dire state. Overlooked by many, Huang et al. (1999) conducted a study into the allozymic variation of 12 populations. Historically oak-chestnut forests played an important part in eastern North American forests (Davis as cited by Huang et al. 1999). This link between the two genera, *Quercus* and *Castanea*, makes a good contrast and comparison point between potential ancient forest compositions and current compositions. The general findings of this study indicate that American chestnut has an average to low level of genetic diversity, which could have contributed to its demise (Huang et al. 1999). Going further, this study indicates the Appalachian populations originated from multi refugia during the last glacial maximum (Huang et al. 1999). This information will be compared to the models to be created, and to the genetic diversity to

the *Quercus* genus to see if links do exist and if a location for the refugia can be determined.

ALLOZYMES

This thesis relies heavily on the genetic diversity of the species in question and is determined by gathering and evaluating previous studies on the variation of allozymes. The allelic variations of enzymes, or allozymes, are encoded by structural genes and are a reliable metric for the overall genetic diversity of the species in question (Richardson et. al. 1986). The studies that have obtained allozyme data do so by the process of allozyme electrophoresis. This process involves using electrophoresis gels to imprint a banding pattern, which is further analyzed (Richardson et. al. 1986). The selection of allozymes as genetic markers for the tree species in this thesis was made due to the historic use of allozymes in other studies. They are useful for evaluating species boundaries, comparisons among taxa and a quantitative estimate of genetic distinction and distribution (Strauss, 1992; Miller & Libby, 1991, as cited by Millar & Westfall 1992).

EXPECTED HETEROZYGOSITY

The previous sections of the literature review looked at the validity of various studies and at the known expected heterozygosity. Table 1 has been made to summarise the findings of literature and display the available data in a concise and efficient manner. The amount of data on each species varies greatly, with the greatest number of studies on *A. saccharum* for a total of 6, and the lowest number on, *Q. alba*, *P. balsamifera*, and *C. dentata*, with only a single study for each.

Table 1. Summary of the expected heterozygosity by species

Species	Expected Heterozygosity	Source
Acer saccharum	0.113	Perry & Knowles 1989.
	0.171	Fore et al. 1992b.
	0.148	Fore et al. 1992a.
	0.148	Simon et al. 1995.
	0.115	Young et al. 1993a.
	0.112	Young et al. 1993b.
Quercus rubra	0.186	Sork et al. 1993.
	0.115	Manos & Fairbrothers. 1987.
	0.172	Guttman & Weigt. 1989.
Quercus macrocarpa	0.187	Schnabel & Hamrick. 1990.
	0.196	Geburek & Tripp-Knowles. 1994.
Quercus velutina	0.102	Manos & Fairbrothers. 1987.
	0.203	Guttman & Weigt. 1989.
Quercus alba	0.097	Manos & Fairbrothers. 1987.
Populus tremuloides	0.235	Hyun et al. 1987.
	0.220	Lund et al. 1992.
	0.290	Jelinski & Cheliak. 1992.
	0.250	Liu & Furnier. 1993.
Populus grandentata	0.080	Liu & Furnier. 1993.
	0.130	Liu & Furnier. 1993.
Populus balsamifera	0.037	Farmer et al. 1988.
Fagus grandifolia	0.395	Houston & Houston. 1994.
	0.165	Houston & Houston. 2000.
Castanea dentata	0.151	Huang et al. 1999.

METHODS AND MATERIALS

MODELLING

Maximum entropy modelling will be used in this thesis, as implemented in the software Maxent (Phillips, Dudík & Schapire n.d.). Maxent uses presence-only data modelling, which relies on species location of individuals as an input into the system, along with a set of environmental predictors. The data gathered from these models will be statistically analyzed to show its validity and accuracy in the results section. These methods will be updated as models are created to ensure a complete list of steps can be followed. Models using current climate data will be made and compared to species distribution mas from Little's "Atlas of United States trees" (1971).

Occurrence data

Each species that is being modelled is based on occurrence data that will be gathered using GBIF or global biodiversity information facility, an online database for species occurrence (GBIF 2020). The amount of occurrence data varies by species, but the species chosen for this thesis all have several thousand occurrence points, which adds accuracy to the model being created. To remove error, species occurrences with no lat/long were removed from the files, along with occurrences that are outside the range of the study area of North America. Duplicates occurrences will also be removed.

Environmental Variables

The paleoclimate data will be gathered from WorldClim.com, with environment variables factored out during correlation with the ecological niche modelling tools built into Maxent (Hijmans et al. 2005). The data that will be gathered from WorldClim is in

a format that is incompatible with Maxent. This data will be formatted in ArcGIS to change the file type and to include the relevant environmental variables. The study area will be focused on North America, and the data will be altered to fit. This area can be seen in the results and appendices but will include the entirety of North America. The paleoclimate data that will be selected has a spatial resolution of 2.5 arc minutes at the equator or roughly 5 by 5 kilometer blocks as this was the most accurate data available. The three Global Climate Models available for use were; CCSM4, MIROC-ESM, and MPI-ESM-P. These will be referred to CCLGM, MRLGM and MELGM, respectively, for this thesis.

Sampling Bias

To ensure a statistically accurate model, the spatial bias of species occurrence records needs to be accounted for (Fourcade, Engler, Rodder D, & Secondi 2014). This bias is due to the fact that occurrence points are likely to be more clustered towards populated or roaded areas because of an increased sampling intensity of these areas compared to rural or remote areas (Fourcade, Engler, Rodder D, & Secondi 2014). Using Arcmap version 10.7, the select random points tool will be used to remove the bias inherent in the occurrence data. Given that the data was open sourced, there were large clusters of points around populated areas in the species ranges. By using the select random points tool, a minimum of 30 kilometers separation between points was implemented to remove these large clusters. While this does not altogether remove the bias, as there is less likely to be occurrence points in areas that are unpopulated and have limited access, it does allow the model to not over sample specific areas of a species ranges. To further reduce bias, Little's Range Maps will be used to eliminate species

occurrence points outside there accepted ranges. While these species likely exist in areas other than their current ranges, they likely do not have large genetically diverse populations in these areas (Fryer 2018).

<u>Running the Models</u>

Maxent will be used on the default settings, with the outputs being Cloglog. The bioclimatic variables that will be entered are based on present-day climate models assembled by Worldclim (Hijmans et al. 2005). This will allow for the Maxent model to interpret the occurrence points and build a relative rate of occurrence for each species. A random test percentage of 25% will also be used to assess the accuracy of the models, with five iterations conducted. The jackknife output of the bioclimatic variables will be used to remove variables that did not contribute to the predictive power of the model. The software EMNtools (Warren, Glor, & Turelli 2010) will be used to evaluate the similarity (through correlation analysis) of the individual climatic variables. Next, with the jackknife output, the best performing variables will be selected, and the variables with a correlation higher than 0.8 will be sequentially removed. This will be repeated multiple times until the model was performing optimally. After optimization, the paleoclimate data will be projected onto the models to build a relative rate of occurrence map for each species.

Of the three types of bioclimatic data that will be used to project the models, MELGM, MRLGM and CCLGM, one will be selected as the main source of refugia size. Each type of bioclimatic data comes from separate models that are were used to predict environmental conditions during the LGM. The CCLGM data comes from the CCSM4 model, the MRLGM data comes from the MIROC-ESM model, and the

MELGM data comes from the MPI-ESM-P model (Hijmans et al. 2005). All three are a Global Climate Model (GCM) and are based on a variety of different inputs, creating different paleoclimate conditions (PUFC n.d.).

Habitat Suitability

To estimate the size of refugia, the relative rate of occurrence values from the Maxent results will be interpreted and categorized into ranges of habitat suitability. These ranges are determined on an overall basis for the entire ten species and are broken down into four classes:

- Class 1 suitability indicates a low likelihood of any major species populations and little chance of gene flow over large distances. The maxent output of $0 - 0.3$ was included in this class.

- Class 2 suitability indicates low to moderate species occurrence with small isolated populations and the possibility of gene flow through these areas. The maxent output of $0.3 - 0.5$ was included in this class.

- Class 3 suitability indicates moderate to high species occurrence, with populations of varying size on the landscape. The maxent output of $0.5 - 0.7$ was included in this class.

- Class 4 suitability indicates high likelihood species occurrence, with the potential for large populations on the landscape. The maxent output of $0.7 - 1$ was included in this class.

For the purpose of this study, refugia size will only be measured at suitability classes 2 and above. The class level is not indicative of definite species presence, but an indicator of likelihood.

<u>Refugia Size</u>

After the classification of the suitable habitat for likely refugia, the data will be transferred to Arcmap 10.7 (ESRI 2011), where data conversion will be done to allow the total area of refugia to be calculated. This area will be calculated in squared kilometers, and then transferred to Excel for a regression analysis (Microsoft Corporation 2018).

<u>Regression Analysis</u>

The software that will be used is the built-in data analysis tools of Excel (Microsoft Corporation 2018). The refugia size and expected heterozygosity we will import into Microsoft Excel (2018) where a regression analysis will be conducted. These results will then put into a graph that can be found in the results.

LITERATURE SEARCH FOR ALLOZYME DATA

To complete the literature review and to find references to add validity to the discussion section, extensive searches will be completed using Google Scholar. This will involve the use of the keywords; allozyme, glacial refugia, last glacial maximum, palynology, and phylogeography. However, to complete a comprehensive search, papers that were sourced will have their references researched also. A significant amount of studies for other areas on the planet had to be sifted through to find the ones pertaining to eastern North America.

RESULTS

GLACIAL REFUGIA MODELLING

During the process of running the models, the output results seem promising, and the performance of the models, measured by the AUC value, was high. An AUC value of 0.5 would indicate that the model is no better than random (Phillips 2005). The AUC graphs for each species can be found in Appendix B. Of the three sets of bioclimatic variables modelled, only the MELGM models showed glacial refugia for all species, with the MRLGM and CCLGM models predicting no refugia for multiple species.

For the majority of these deciduous species, the refugia were located well below the glaciated areas, with only *P. tremuloides* and *P. balsamifera* extending closer, as seen in Figures 3 and 4. This corresponds with the current day species distributions and the prevalence of *P. tremuloides* and *P. balsamifera* in the boreal forest regions. The most common refugial location was a wide distribution ranging from present day Texas to the East Coast. This included; *Q. rubra, Q. velutina, Q. alba, P. grandidentata*, and *A. saccharum*. Many of these species occupy similar but distinct ecosites within their respective current ranges. The species that had the smallest refugia were; *Q. macrocarpa, F. grandifolia* and *C. dentata*.

The predicted MELGM models can be seen below in Figures 1 to 5, and the MELGM/CCLGM models are located in Appendix A.
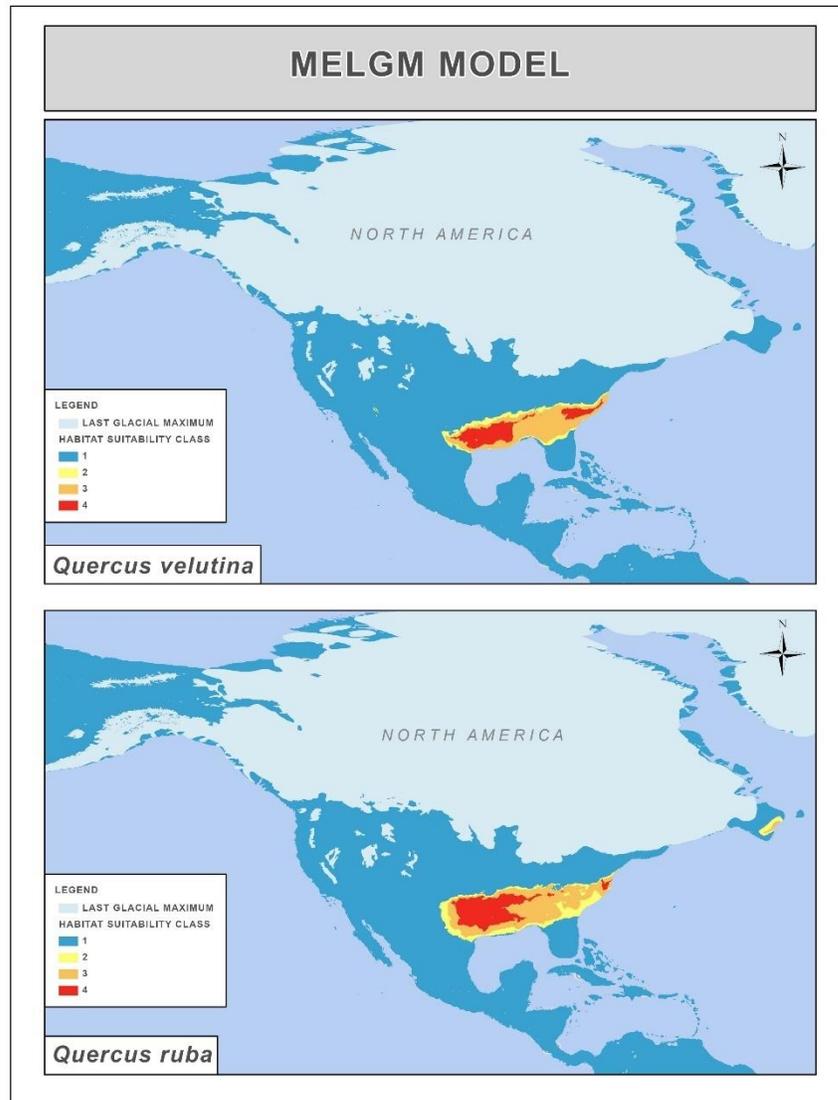
Figure 1. Predicted glacial refugia of *Q. rubra* and *Q. velutina*.

Both *Quercus* species in Figure 1 above have similar refugia that were connected
and likely had adequate gene flow. The separate class 4 suitability areas of *Q. velutina*
might indicate divergent populations with smaller sub populations in-between. *Quercus
rubra* has a more distinct high suitability area, where a single large population was
located, and smaller sub-population further east.
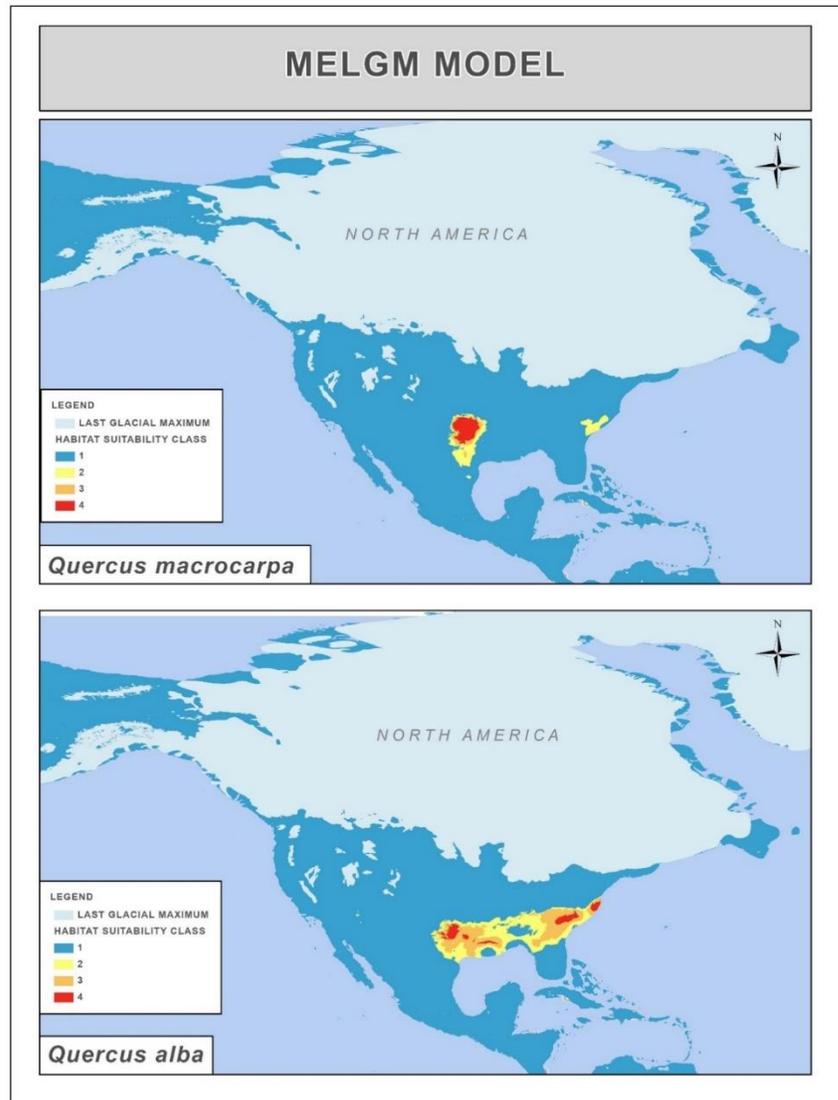
Figure 2. Predicted glacial refugia of *Q. macrocarpa* and *Q. alba.*

The refugia for *Q. macrocarpa* and *Q. alba* are distinctly different from the species in Figure 1. The extremely limited refugia of *Q. macrocarpa* indicates one population group, where the entirety on species genetics originated from. Conversely, *Q. alba* has two distinct refugia that are linked by an area that had limited populations and little to no gene flow.
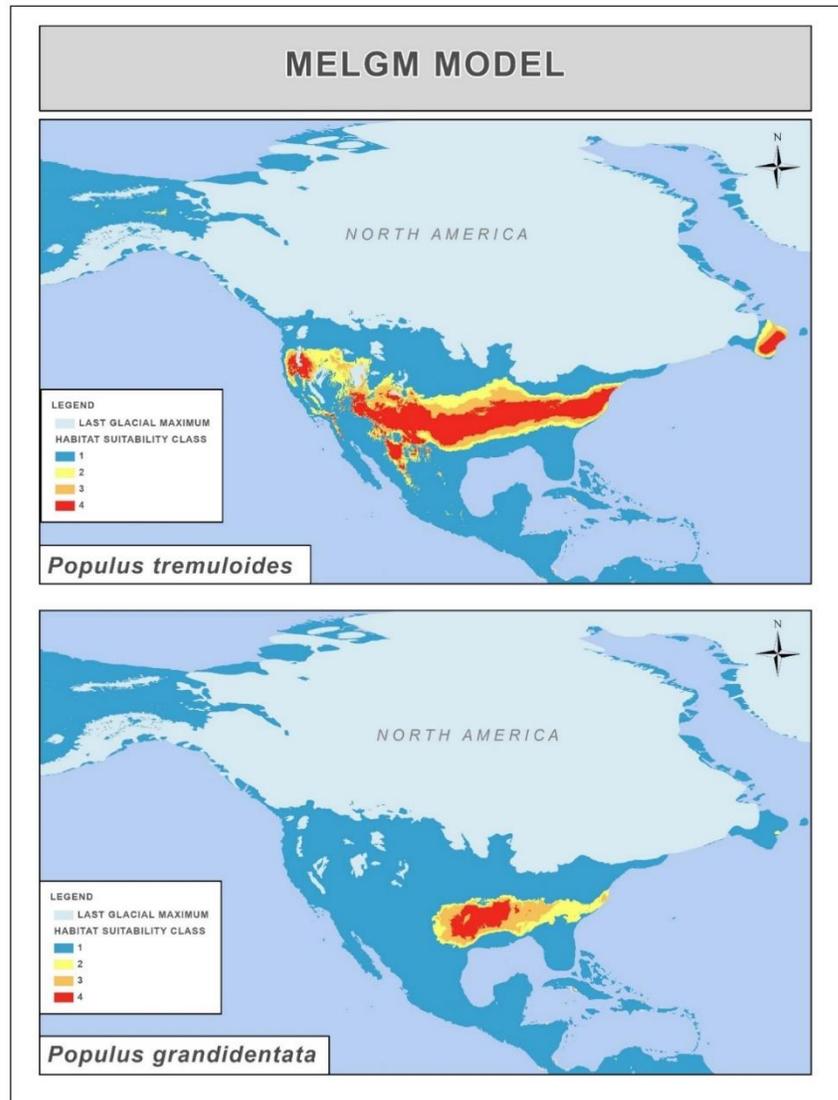
Figure 3. Predicted glacial refugia of *P. tremuloides* and *P. grandidentata*.

The refugia for *P. tremuloides* is wide ranging, spanning the entirety of North America, and approaching the ice sheet. The populations of *P. tremuloides* were well connected with no distinct single refugium. On the contrary, the refugia for *P. grandidentata* is smaller and has a single distinct population with smaller sub-populations near the East Coast.
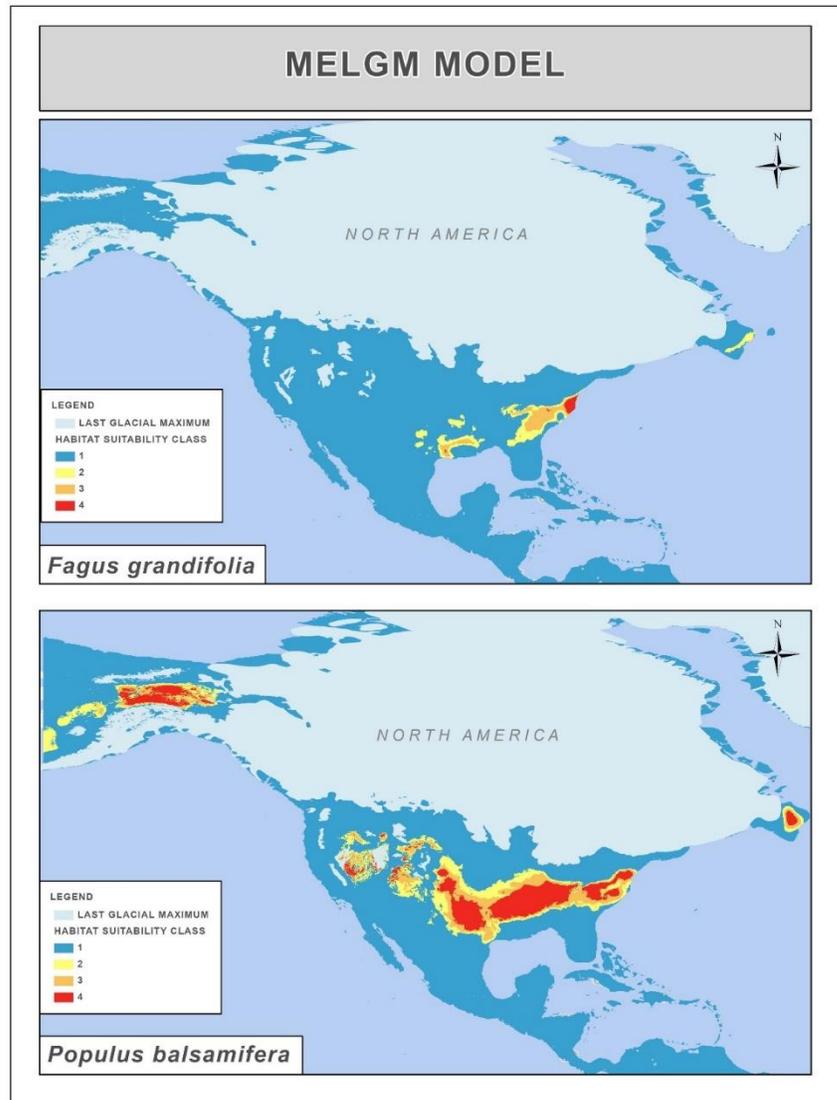
Figure 4. Predicted glacial refugia of *F. grandifolia* and *P. balsamifera.*

Of all the species, only *P. balsamifera* was predicted to have a refugium located in present day Alaska. The southern population of *P. balsamifera* was well established with smaller refugium in the mountains to the west. The limited refugia of *F. grandifolia* are widely separated and have very little class 4 habitat suitability. There was a small core population directly on the East Coast, and scattered subpopulations, with little to no gene flow.
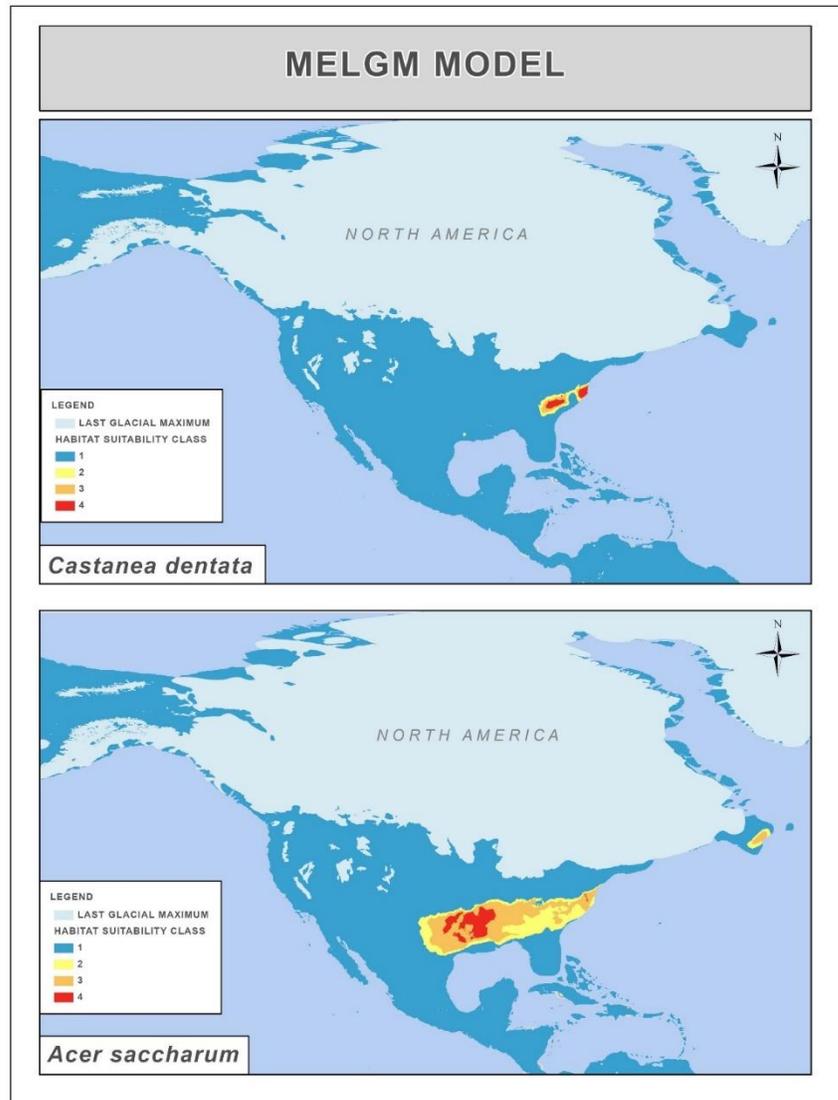
Figure 5. Predicted glacial refugia of *C. dentata* and *A. saccharum*

The two refugia of *C. dentata* are small in overall size but close in proximity. Gene flow may have occurred between them, but its more likely that two distinct populations existed. The large refugia of *A. saccharum* had a core population in class 4 suitable habitat, with smaller populations located near the East Coast. Wide spread populations likely existed in small numbers. A notable exception is the lack of refugium, for any species, located in present day Florida, which was likely a warmer area during

the LGM. The paleoclimate variables were likely the cause of this omission, as numerous studies theorized extensive forested areas in that location.

SPECIES GENETIC DIVERSITY

The genetic variation of the species in this thesis is as diverse as the range of sizes of the modelled glacial refugium. However, no strong relationship between the two can easily be seen by looking at the data. Were this relationship to exist, the two species with the largest refugium would have the highest expected heterozygosity. To illustrate this point, Table 2 was created to show the findings of the modelling and genetic diversity data. This summary table is sorted by descending order of highest to lowest expected heterozygosity for ease of understanding.

Table 2. Summary of predicted refugia size and average expected heterozygosity

Species	Refugia (#)	Refugia Size (10^6 km^2)	Average He
Fagus grandifolia	3	0.7	0.280
Populus tremuloides	2	3.8	0.249
Quercus macrocarpa	2	0.4	0.191
Quercus rubra	2	2.0	0.158
Quercus velutina	1	1.4	0.152
Castanea dentata	1	0.2	0.151
Acer saccharum	2	1.8	0.135
Populus grandidentata	1	1.4	0.105
Quercus alba	1	1.4	0.097
Populus balsamifera	6	3.2	0.037

* He - Expected heterozygosity

A cursory glance at Table 2 shows *P. balsamifera* at the very bottom with a H_{et} value of 0.037, while *F. grandifolia* is at the top with a H_{et} value of 0.280. This is surprising given the refugium size of *P. balsamifera* is 3.2×10^6 km^2, and F. grandifolia is 0.7×10^6 km^2. This runs contrary to the hypothesis and expected results of this thesis.

There may be a relationship, however, as *P. tremuloides* was the second highest H_{et}

value of 0.249, and the largest refugium size of 3.8×10^6 km^2. To fully assess the

relationship between the refugium size and genetic diversity, a linear regression analysis

was conducted.

REGRESSION ANALYSIS

Unsurprisingly the linear regression of all ten species, with refugia size as the

independent variable Y, and the average expected heterozygosity as the dependent

variable X, failed to show a meaningful relationship. In fact, only 2.37% of the variation

of expected heterozygosity can be explained by the size of the refugia, as seen in Table 3

below. Furthermore, the adjusted r square value indicates an overall negative

relationship, which means as refugia size goes up, genetic diversity goes down (Table

3).

Table 3. Regression analysis statistics of refugia size and expected heterozygosity

Regression Statistics	
Multiple R	0.154
R Square	0.024
Adjusted R Square	-0.098
Standard Error	0.075
Observations	10

To visualize this confusing finding, Figure 6 was created from the linear

regression of the two aforementioned variables.
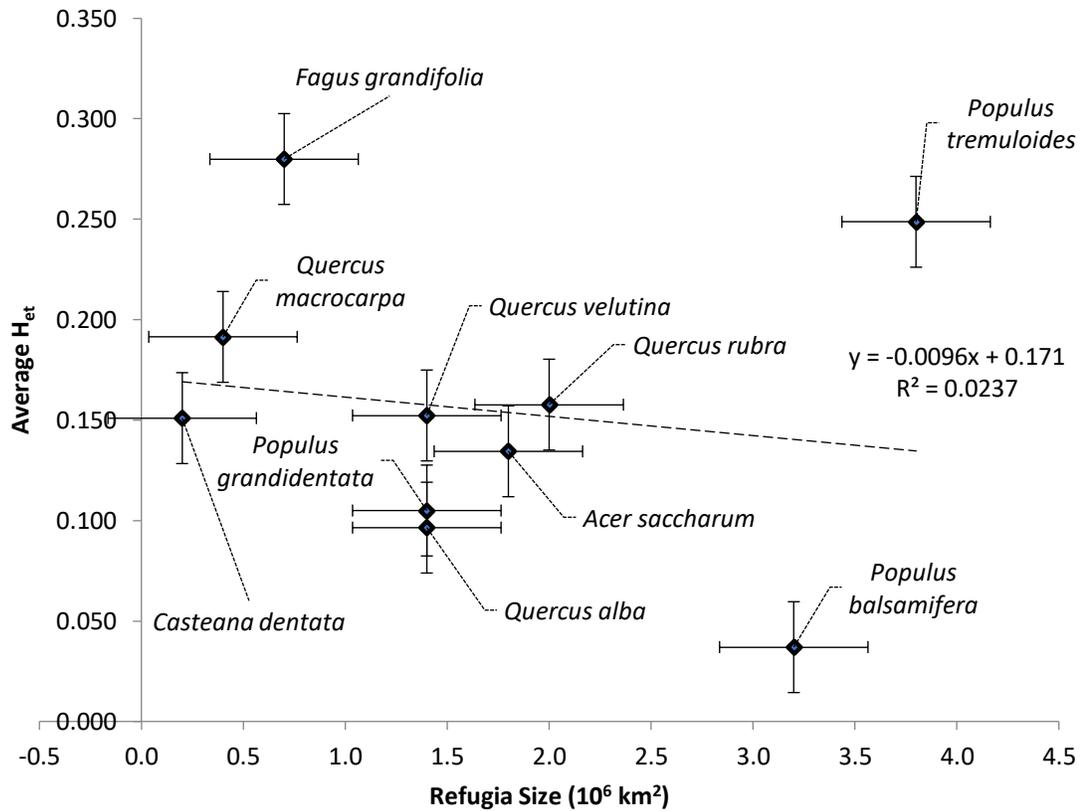
Figure 6. Expected heterozygosity as a function of modelled species refugia size at the last glacial maximum.

The reasons for this weak relationship likely have to do with four species in this study; *C. dentata, Q. macrocarpa, F. grandifolia* and *P. balsamifera.* Removing these four species from the linear regression, produces excellent results, with an r^2 value of 0.874, as seen in Table 4. This reduction of the number of species in the regression analysis does lower the number of points to a level where its unlikely any finding can be significant. Many consider it a rule of thumb that a minimum of 10 points are required for statistical significance (Chernick 2012).

Table 4. Regression analysis for the six species; *P. tremuloides, Q. rubra, Q. velutina, A. saccharum, P. grandidentata, Q. alba*

Regression Statistics	
Multiple R	0.935
R Square	0.874
Adjusted R Square	0.842
Standard Error	0.022
Observations	6

Interestingly, of the species removed for Table 4, *Q. macrocarpa,* and *F. grandifolia*, had relatively small refugia, and high expected heterozygosity. Furthermore, two of the removed species, *C. dentata* and *P. balsamifera*, had only one source for genetic diversity. However, *Q. alba* only had one source also but performed well in the regression in Table 4.

To further this study, the problems with the expected heterozygosity data must be reconciled with current literature. The predicted refugia also must be validated, if possible, to ensure the accuracy of the geographic locations.

DISCUSSION

KNOWN REFUGIA

The literature that pertains to this thesis has a few examples of known refugia for several species. These are proven through fossil pollen data, macro fossil data, and the interpretation of genetic variation. The species in this section can have their refugia models either validated or proven, to be inaccurate.

Castanea dentata

One of the problem species for this thesis was *Castanea dentata*. This is due mainly to the small size of refugia and the relatively high expected heterozygosity. However, a study by Kubisiak and Roberds (2006) looked into the genetic diversity and allele frequencies of *Castanea dentata* and made the proposition of a glacial refugium existing in present day Mississippi and Alabama. This was due to the findings of the largest number of rare alleles located in the current southwestern portion of *C. dentata* range (Kubisiak and Roberds 2006). These findings add validity to the model predictions for *C. dentata*; however, there are notable discrepancies. The predicted model has higher habitat suitability further east than the study by Kubisiak and Roberds (2006). This difference could be explained by the occurrence points of *C. dentata* being biased due to the reduction of the population through the chestnut blight in the last 100 years. The model does predict lower habitat suitability in the area of the theorized refugia; however, this suitability fell under the class 1 habitat suitability designation during the modelling, and thus was not considered to be a refugium in the model. However, this does not eliminate the possibility of populations existing in the area.

The relatively high expected heterozygosity of 0.151, as proposed by Huang et.

(1999) would indicate that populations of *C. dentata* were of a healthy size, with

adequate gene flow, within their refugia. While a known refugium for *C. dentata* exists,

it is based solely on a single study by Kubisiak and Roberds (2006), with little to no

corroborating evidence for other sources.

Acer saccharum

As the species with the most sources for expected heterozygosity and a large

number of studies that theorized LGM refugia, *A. saccharum* should be the easiest to

validate. The *Acer saccharum* model predicted a large continuous refugium in the

Southern USA, as seen in Figure 5. The model can be validated by fossil pollen data that

suggest *A. saccharum* migrated north from a single refugium in a similar area to the

model (Delcourt and Delcourt 1987; as cited by Vargas-Rodriguez et al. 2015).

However, this assumption is thrown into question by Vargas-Rodriguez et al. (2015),

who had to reject their hypothesis of a single refugium and was unable to reject their

hypothesis of multiple refugia. This brings in the question of the likelihood of small

refugia existing outside the predicted models' ranges and the potential inaccuracy of the

model itself. Further doubt is placed on the model by the identification of two maple

genetic refugia, with the use of pollen and macrofossil data (Jackson et al. 2000; as cited

by Graignic et al. 2018).

It is possible that the model prediction of habitat suitability could confuse the

ability for a tree to grow in a particular climate, with the ability to grow on specific

ecosites in that climate. It is unlikely that there were numerous widespread populations

of *A. saccharum*, due to competition with other trees deciduous and coniferous. More

likely, there were small areas of core habitat and smaller fragmented populations within the refugia.

Fagus grandifolia

A unique problem within this thesis, *F. grandifolia*, has one of the highest values for expected heterozygosity, but the model predicted one of the smallest refugia. The models predicted two distinct refugia located, as seen in Figure 4. This is a viable possibility, as a study by Kitamura and Kawano (2001) revealed definite regional differences, with two genetically distinct clusters. These two clusters likely originated from separate refugia, with one in the area of the Gulf-Coastal plain, eastern coastal plain, Piedmont Plateau, and Ozark plateau, and to the northeast, nearer to the east coast. (Kitamura and Kawano 2001). This validates an aspect of the model produced in this thesis, in which the East Coast refugium showed high habitat suitability; however, the refugia on the Gulf Coast, showed much lower, and more scattered habitat suitability.

The predicted western refugia on the gulf-coastal plain is further legitimized by the suggestion of a full-glacial refugium for *Fagus grandifolia*, and other northern deciduous trees, located "in the Tunica Hills and the bluff-lands either side of the Mississippi valley from the Gulf coast at least as far north as Tennessee" (Delcourt & Delcourt 1975; as cited by Bennett 1985). Moreover, Fossil pollen records were used by Delcourt and Delcourt (1987) to create a northern range of *Fagus grandifolia* migration during the LGM, that encompasses both of the predicted refugia.

Fagus – Acer Relationship

The link between *Acer saccharum* and *Fagus grandifolia* is well documented in current forest conditions, where they inhabit similar areas and tend to be found in the

33

same forest types. This link likely existed during the LGM with the migration of maple-beech associations being theorized by Delcourt and Delcourt (1987). Of the two species, *A. saccharum* had a larger refugium, with higher habitat suitability in areas, wherein those same areas, *F. grandifolia* had fragmented habitat suitability. There is a point on the East Coast, where both *A. saccharum* and *F. grandifolia* have high habitat suitability and definite refugial overlap. While the refugia data does not prove a link between the species, there likely was a relationship during the LGM.

Populus tremuloides/balsamifera

Of the three *Populus* species modelled in this study, *P. tremuloides*, and *P. balsamifera* have a more northerly present-day range and are associated with boreal coniferous species. This range indicates that these two species are more adapted to harsher climates and lower mean average temperatures. This is represented in the models by habitat suitability near the North American ice sheet. The predicted models of *P. tremuloides* suggested a wide refugia stretching coast to coast, with the only barrier being the Rocky Mountains to create a division. This can be backed up by a study, which found two major genetic clusters based on microsatellite markers (Callahan et al. 2013). The proposed division is between the current day northern and western distributions. Where the northern distribution has a high level of genetic diversity, which was likely the result of large refugia in the southern USA during the LGM (Callahan et al. 2013). While the western population likely resulted from a smaller refugium that was located in the Rocky Mountains and was unable to have adequate gene flow to the larger population (Callahan et al. 2013). This is a complete validation of the *P. tremuloides* model accuracy and can be associated with the high expected heterozygosity. Another

study proposed a glacial refugium for *P. tremuloides* in present day Alaska; however, the findings disproved this theory (Latutrie, Bergeron & Tremblay 2016). This is more validation of the model's accuracy as it did not predict any refugia in that area.

One of the largest problems with the linear regression was the inclusion of *P. balsamifera,* given its large refugia size and low heterozygosity. A study by Breen, Murray and Olson (2012) was able to map fossil pollen data accurately and predicted a large glacial refugium in the southern USA, and small refugia in present day Alaska. This would indicate that the predicted model for *P. balsamifera* is accurate, given it was the only species to have even a low level of suitable in Alaska. Therefore, the only limitation of *P. balsamifera*, to this thesis, was the lack of allozyme studies to represent the species' real genetic diversity accurately.

UNKNOWN REFUGIA

Due to the relative lack of comprehensive glacial refugia studies within North America, certain species chosen for this study lack any direct literary evidence of known refugia. This will make validation predicted models more complicated.

Quercus Species

There is a notable lack of studies into the presence of glacial refugia for the *Quercus* genus. Moreover, there is no information to differentiate between the separate *Quercus* species likely refugia, in the studies that do exist. One such example is a study by Jackson et al. (2000), that mapped fossil pollen data for the Southeastern USA. This study did not gather any data past the Mississippi River; thus, the scope is limited. The findings showed a significant amount of *Quercus spp*. macro fossils along the southern

gulf coast, ranging from the Mississippi to the East Coast and down into present day Florida (Jackson et al. 2000). Another study, by Davis (1983), theorized a similar location for a *Quercus* spp. glacial refugium, which roughly matches the predicted models.

Three species, *Q. alba, Q. rubra,* and *Q. velutina*, all have relatively similar refugia that were modelled, with large areas located in the Southern USA. The outlier of the four is *Q. macrocarpa,* with a small refugium in the Southwestern USA.

The only method of providing a literary example of known refugia is looking at the relationships of the *Quercus* species with other tree species in this study. While this is not a strong validation of the models, the current day intra-species relationships likely indicate past relationships. Therefore, there is a way to partially validate the models of this study for the *Quercus* genus. One known relationship is the Oak-Pine forest type. While *Pinus spp*. were not used in this thesis, they were a major component of the forest composition of the Southeastern USA during the LGM (Webb 1988; as cited by Abrams 1992). This relationship could indicate small oak populations surviving across the landscape within the predicted refugia of this thesis.

Being unable to differentiate between *Quercus* species in known literature makes validating the individual models for the four *Quercus* species difficult. Only a general validation of the models fit into known refugia for the *Quercus* genus.

Populus grandidentata

This species offered the most problems with validating the predicted glacial refugia in this thesis. One of the only studies that map a glacial refugium for *Populus*

spp. is by Jackson et al (2000), which highlights a vague small area north of Florida. *P. grandidentata* has a current species distribution that is more related to southern deciduous species, inhabiting many of the same forest types as *A. saccharum, F. grandifolia*, and the *Quercus* genus in this study. This relationship between *P. grandidentata* and the other deciduous species in this thesis will have to serve as the only source of likely refugia. No studies exist that theorize likely glacial refugia of *P. grandidentata*. Given the low expected heterozygosity, this could indicate small populations that are widely separated.

Potential Refugia

Four of the models predicted species occurrences located in the far east of North America in current day Newfoundland. The model suggested a high relative rate of occurrence for *P. tremuloides* and *B. balsamifera* in the mid-Atlantic coastal plain. A phylogeographic study by Godbout et al. (2000) found indications of a jack pine refugia in the mid-Atlantic coastal plain. The present-day range of jack pine is similar to *P. tremuloides* and *B. balsamifera* and could validate the models' predictions.

The model suggested a low relative rate of occurrence for both *A. saccharum* and *Q. rubra,* which can be interpreted as the existence of sparse populations in the mid-Atlantic coastal plain. Furthermore, a study by Tomson et al. (2014) linked the chloroplast DNA structure of yellow birch to a refugia mid-Atlantic coastal plain. The similarity of present-day ranges between *A. saccharum, Q. rubra*, and *Betula alleghaniensis* could provide validation of the modelled refugia.

The current literature available confirms the presence of the modelled refugia in the mid-Atlantic coastal plain.

INHERENT BIAS

One of the main problems with this thesis is the inherent bias at all levels. While some bias can be accounted for, such as the occurrence point clusters, others simply can not be corrected for, and it is important to blatantly state what those are. A number of problems seen in the results of this study, mainly the lack of genetic variation data on several species. *Populus balsamifera* is the clearest example, with the only study on the expected heterozygosity relying on a small population of individuals in their northern range in Ontario. This data shows low expected heterozygosity of 0.037. This bias can be compounded even further by the general validation of most refugia models. Believing the size and extent of the models to be correct is, in its way, another form of bias. Moreover, these models likely overestimated the size of the refugia for most species, given the literary examples are usually small in size, and predict very low occurrences overall.

Some of the base assumptions of Maxent also level a certain amount of doubt on the results of the model projections. Maxent is known to have certain problems with presence only data when sampling is not entirely random (Royle et al. 2012). Moreover, the logistic output is incorrectly interpreted as occurrence probability, when it should only be considered a relative rate of occurrence (Yackulic et al. 2012.)

CONCLUSION

This study was able to successfully reconstruct LGM refugia for several Eastern North American tree species. The extent of these refugia were measured and compared

to the He found in current literature. However, no significant relationship between He and glacial refugia size was found. The hypothesis is now considered to be incorrect.

The main reason for this outcome was the lack of genetic diversity data available in the current literature. The majority of species relied on He values from one or two studies (Table 1). Furthermore, many of these studies sampled small populations at the limits of their ranges, and without correlating data, the genetic diversity is inaccurate.

Future studies should incorporate the average number of alleles per locus of each species to better represent genetic diversity. Moreover, if further studies into allozyme variation of the selected species are not conducted, a different genetic marker should be chosen.

There are proven relationships between genetic diversity and refugia size during the last glacial maximum for Western North American species (Roberts and Hamann 2015). However, this study was unable to prove this relationship existed for the selected Eastern North American species, and future studies will need to be conducted.

LITERATURE CITED

Abrams, M. D. 1992. Fire and the Development of Oak Forests. BioScience, 42(5), 346–
353.

Bennett, K. D. 1985. The Spread of Fagus grandifolia Across Eastern North America
During the Last 18 000 years. *Journal of Biogeography*, *12*(2), 147–164.

Breen, A. L., Murray, D. F., & Olson, M. S. 2012. Genetic consequences of glacial
survival: the late Quaternary history of balsam poplar (Populus balsamifera L.) in
North America. *Journal of Biogeography*, *39*(5), 918–928.

Callahan, C. M., Rowe, C. A., Ryel, R. J., Shaw, J. D., Madritch, M. D., & Mock, K. E.
(2013). Continental-scale assessment of genetic diversity and population
structure in quaking aspen (Populus tremuloides). *Journal of
Biogeography*, *40*(9), 1780–1791.

Carson, Justine. 2015. Alaska: Copper River Delta and Prince William Sound. Half
Moon Bay, CA. Justine Carson Photography https://www.justinecarson.com/

Chernick, Michael R. (https://stats.stackexchange.com/users/11032/michael-r-chernick).
2012. Minimal number of points for a linear regression, URL.
https://stats.stackexchange.com/q/37835

Dalgleish H., Nelson C., Scrivani J., Jacobs D. 2015. Consequences of shifts in
abundance and distribution of American chestnut for restoration of a foundation
forest tree. Forests. 7. 4- 9. doi:10.3390/f7010004

Davis, M. B. . 1981. Quaternary history and the stability of forest communities. Pages
132–153 in D. C. West, H. H. Shugart, and D. B. Botkin, editors. Forest
succession: concepts and application. Springer-Verlag, New York, New York,
USA.

Delcourt, P. A. and H. R. Delcourt . 1987. Long term forest dynamics of the temperate
zone: a case study of late-Quaternary forests in eastern North America. Springer-
Verlag, New York, New York, USA.

Ding, C., Schreiber, S. G., Roberts, D. R., Hamann, A., & Brouard, J. S. 2017. Post-
glacial biogeography of trembling aspen inferred from habitat models and
genetic variance in quantitative traits. *Scientific Reports*, *7*(1).

ESRI. 2011. ArcGIS Desktop: Release 10. Redlands, CA: Environmental Systems
Research Institute.

Farmer, R.E., Jr., W.M. Cheliak, D.J. Perry, P. Knowles, J. Barrett and J.A. Pitel. 1988
Isozyme variation in balsam popular along a latitudinal transect in northwestern
Ontario. Can. J. For. Res. 18: 1078-1081.

Fore, S. A., R. J. Hickey, J. L. Vankat, S. I. Guttman, and R. L. Schaefer. 1992b. Genetic structure after forest fragmentation: a landscape ecology perspective on Acer saccharum. Canadian Journal of Botany 70:1659–1668

Fourcade Y, Engler JO, Rodder D, and Secondi J. 2014. Mapping Species Distributions with MAXENT Using a Geographically Biased Sample of Presence Data: A Performance Assessment of Methods for Correcting Sampling Bias. PLoS ONE 9(5): e97122. doi:10.1371/journal.pone.0097122

Fryer, Janet L., comp. 2018. Tree species distribution maps from Little's "Atlas of United States trees" series. In: Fire Effects Information System, [Online]. U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Fire Sciences Laboratory (Producer).

Gaillard, M.-J., Sugita, S., Bunting, M. J., Middleton, R., Broström, A., Caseldine, C., Veski, S. 2008. The use of modelling and simulation approach in reconstructing past landscapes from fossil pollen data: a review and results from the POLLANDCAL network. *Vegetation History and Archaeobotany*, *17*(5), 419–443.

GBIF: The Global Biodiversity Information Facility (2020) *What is GBIF?*. Available from https://www.gbif.org/what-is-gbif

Geburek, T., & Tripp-Knowles, P. 1994. Genetic architecture in bur oak,Quercus macrocarpa (Fagaceae), inferred by means of spatial autocorrelation analysis. *Plant Systematics and Evolution*, *189*(1-2), 63–74.

Godbout, J., Beaulieu, J., & Bousquet, J. 2010. Phylogeographic structure of jack pine (Pinus banksiana; Pinaceae) supports the existence of a coastal glacial refugium in northeastern North America. *American Journal of Botany*. *97*. 11. 1903–1912. doi: 10.3732/ajb.1000148

Graignic, N., Tremblay, F., & Bergeron, Y. 2018. Influence of northern limit range on genetic diversity and structure in a widespread North American tree, sugar maple (Acer saccharumMarshall). *Ecology and Evolution*, *8*(5), 2766–2780.

Guttman, S. I., & Weigt, L. A. 1989. Electrophoretic evidence of relationships among Quercus (oaks) of eastern North America. Canadian Journal of Botany, 67(2), 339–351.

Hijmans et al., 2005. WorldClim 1.4 (historical climate conditions). Int. J. of Clim. 25: 1965-1978.

Hollingsworth, P. M., Bateman, R. M., & Gornall, R. J. 1999. *Molecular systematics and plant evolution*. New York: Taylor & Francis.

Houston, B., & Houston, R. (1994). Variation in American Beech (*Fagus grandifolia* Ehrh.): Isozyme Analysis of Genetic Structure in Selected Stands . *Silvae Genetica*, *43*(5), 277–284.

Houston, D. B., and D. R. Houston. 2000. Allozyme genetic diversity among Fagus grandifolia trees resistant or susceptible to beech bark disease in natural populations. Canadian Journal of Forest Research 30:778–789.

Huang, H., Dane, F., & Kubisiak, T. 1999. Allozyme and RAPD Analysis of the Geographic Variation and Genetic Diversity in Natural Populations of the American Chestnut. *American Journal of Botany*, *85*(7), 1013–1021.

Hyun, J. O., O. P. Rajora, and L. Zsuffa 1987. Genetic variation in trembling aspen in Ontario based on isozyme studies. Canadian Journal of Forest Research 17:1134–1138.

Jackson, S. T., Webb, R. S., Anderson, K. H., Overpeck, J. T., Iii, T. W., Williams, J. W., & Hansen, B. C. (2000). Vegetation and environment in Eastern North America during the Last Glacial Maximum. *Quaternary Science Reviews*, *19*(6), 489–508.

Jannson, S., Bhalerao, R. P., & Groover, A. T. (Eds.). 2010. Genetics and Genomics of Populus. *Plant Genetics/ Genomics*,

Jelinski, D.E., and Cheliak, W.M. 1992. Genetic diversity and spatial subdivision of Populus tremuloides (Salicaceae) in a heterogeneous landscape. Amer. J. Bot. 79: 728-736.

Keller, S. R., Soolanayakanahally, R. Y., Guy, R. D., Silim, S. N., Olson, M. S., & Tiffin, P. (2011). Climate-driven local adaptation of ecophysiology and phenology in balsam poplar, Populus balsamifera L. (Salicaceae). *American Journal of Botany*, *98*(1), 99–108.

Kitamura, K., & Kawano, S. 2001. Regional Differentiation in Genetic Components for the American Beech, Fagus grandifolia Ehrh., in Relation to Geological History and Mode of Reproduction. *Journal of plant research, 114*, 353-368.

Kubisiak, T. L., & Roberds, J. H. 2006. Genetic Structure of American Chestnut Populations Based on Neutral DNA Markers. In Restoration of American Chestnut to Forest Lands (pp. 109–122). U.S Department of the Interior, National Park Service, National Capital Region, Center for Urban Ecology.

Laidly, P. R. 1990. Populus grandidentata Michx. bigtooth aspen. *Silvics of North America*, *2*, 544-550.

Latutrie, M., Bergeron, Y., & Tremblay, F. 2016. Fine-scale assessment of genetic diversity of trembling aspen in northwestern North America. *BMC Evolutionary Biology*, *16*(1).

Li, P., Bousquet, J., & Mackay, J. 1992. Genetic diversity in Canadian Hardwoods: Implications for conservation. *The Forestry Chronicle*, *68*(6), 709–719.

Little, E. L. (1971). *Atlas of United States trees*. Washington: U.S. G.P.O.

Liu, Z., & Furnier, G. R. (1993). Comparison of allozyme, RFLP, and RAPD markers for revealing genetic variation within and between trembling aspen and bigtooth aspen. *Theoretical and Applied Genetics*, *87*(1-2), 97–105.

Lund, S. T., G. R. Furnier, and C. A. Mohn 1992. Isozyme variation in quaking aspen in Minnesota. Canadian Journal of Forest Research 22:521–524

Magni, C. R., Ducousso, A., Caron, H., Petit, R. J., & Kremer, A. 2005. Chloroplast DNA variation of Quercus rubra L. in North America and comparison with other Fagaceae. *Molecular Ecology*, *14*(2), 513–524.

Manos, P. S., & Fairbrothers, D. E. 1987. Allozyme Variation in Populations of Six Northeastern American Red Oaks (Fagaceae: Quercus Subg. Erythrobalanus). *Systematic Botany*, *12*(3), 365.

Martínez-Meyer, E., & Peterson, A. T. 2006. Conservatism of ecological niche characteristics in North American plant species over the Pleistocene-to-Recent transition. *Journal of Biogeography*, *33*(10), 1779–1789.

Mclachlan, J. S., Clark, J. S., & Manos, P. S. 2005. Molecular Indicators Of Tree Migration Capacity Under Rapid Climate Change. *Ecology*, *86*(8), 2088–2098.

Microsoft Corporation. 2018. *Microsoft Excel*. Retrieved from https://office.microsoft.com/excel

Millar, C. I., & Westfall, R. D. 1992. Allozyme markers in forest genetic conservation. *New Forests*, *6*(1-4), 347–371.

Miller, M. D., & Parker, K. C. 2009. The Impacts of the Appalachian Mountains on the Post-Glacial Migration Pathways and Gene Flow of Sugar Maple (Acer Saccharum). *Physical Geography*, *30*(2), 89–104.

Mullin, T.J. and S. Bertrand. Genesis Forest Science Canada Inc. 1999. Forest management impacts on genetics of forest tree species. SCSS Technical Report #113.

Nei M, Maruyama T, & Chakraborty R. 1975 The bottleneck effect and genetic variability in populations. Evolution 29, 1 – 10. (doi:10.2307/ 2407137)

Perry, D. J., & Knowles, P. 1989. Allozyme variation in sugar maple at the northern limit of its range in Ontario, Canada. *Canadian Journal of Forest Research*, *19*(4), 509–514.

Phillips, S. 2005. A Brief Tutorial on Maxent. *Lessons in Conservation*, *3*, 108–135.

Prentice, I. C., Bartlein, P. J., & Webb, T. 1991. Vegetation and Climate Change In Eastern North America Since The Last Glacial Maximum. *Ecology*, *72*(6), 2038–2056.

Princeton University Forrestal Campus. [PUFC] n.d. Climate Modeling. Retrieved from https://www.gfdl.noaa.gov/climate-modeling/

Ray, N., & Adams, J. 2001. A GIS-based Vegetation Map of the World at the Last Glacial Maximum (25,000-15,000 BP). *Internet Archaeology*, (11).

Richardson, B. J., Baverstock, P. R., & Adams, M. 1986. *Allozyme Electrophoresis A Handbook for Animal Systematics and Population Studies*. Sand Diego: Academic Press. Inc. .

Roberts, D. R., & Hamann, A. 2015. Glacial refugia and modern genetic diversity of 22 western North American tree species. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1804), 2014-2903.

Royle, J. A., Chandler, R. B., Yackulic, C., & Nichols, J. D. 2012. Likelihood analysis of species occurrence probability from presence-only data for modelling species distributions. Methods in Ecology and Evolution, 3(3), 545–554.

Schnabel, A., & Hamrick, J. L. 1990. Comparative Analysis of Population Genetic Structure in Quercus macrocarpa and Q. gambelii (Fagaceae). *Systematic Botany*, *15*(2), 240–251.

Simon, J.-P., Y. Payette, and M.-H. Longpre 1995. Comparative analysis of the genetic composition of canopy and juvenile sugar maple individuals (Acer saccharum) in an old-growth forest in southern Quebec as related to anthropogenic disturbance. Canadian Journal of Forest Research 25:743–752.

Sork, V., Huang, S., & Wiener, E. 1993. Macrogeographic and fine-scale genetic structure in a North American oak species, Quercus rubra L. *Annales Des Sciences Forestières*, *50*(Supplement).

Steven J. Phillips, Miroslav Dudík, Robert E. Schapire. [Internet] Maxent software for modeling species niches and distributions (Version 3.4.1). Available from url: http://biodiversityinformatics.amnh.org/open_source/maxent/.

Thomson, A. M., Dick, C. W., & Dayanandan, S. 2014. A similar phylogeographical structure among sympatric North American birches (Betula) is better explained by introgression than by shared biogeographical history. Journal of Biogeography. 42. 2. 339–350. doi: 10.1111/jbi.12394

Vargas-Rodriguez, Y. L., Platt, W. J., Urbatsch, L. E., & Foltz, D. W. 2015. Large scale patterns of genetic variation and differentiation in sugar maple from tropical Central America to temperate North America. BMC Evolutionary Biology, 15(1), 1–14.

Warren, D. L., Glor, R. E., & Turelli, M. 2010. ENMTools: a toolbox for comparative studies of environmental niche models. *Ecography*, *33*, 607–611. doi: 10.1111/j.1600-0587.2009.06142.x

Yackulic, C. B., Chandler, R., Zipkin, E. F., Royle, J. A., Nichols, J. D., Grant, E. H. C., & Veran, S. 2012. Presence-only modelling using MAXENT: when can we trust the inferences? *Methods in Ecology and Evolution*, *4*(3), 236–243. doi: 10.1111/2041-210x.12004

Yokoyama, Y., Lambeck, K., Deckker, P. D., Johnston, P., & Fifield, L. K. 2001. Erratum: correction: Timing of the Last Glacial Maximum from observed sea-level minima. *Nature*, *412*(6842), 99–99.

Young, A. G., S. I. Warwick, and H. G. Merriam 1993b. Genetic variation and structure at three spatial scales for Acer saccharum (sugar maple) in Canada and the implications for conservation. Canadian Journal of Forest Research 23:2568–2578.

Young, A., H. G. Merriam, and S. I. Warwick 1993a. The effects of forest fragmentation on genetic variation in Acer saccharum Marsh. (sugar maple) populations. Heredity 71:277–289.

APPENDICES

CCLGM MODEL

Populus tremuloides

Populus grandidentata

MRLGM MODEL

Populus tremuloides

Populus grandidentata

MRLGM MODEL

Castanea dentata

Acer saccharum

MRLGM MODEL

Quercus velutina

Quercus ruba

MLRGM MODEL

Fagus grandifolia

Populus balsamifera

MRLGM MODEL

Quercus macrocarpa

Quercus alba

LEGEND
LAST GLACIAL MAXIMUM
HABITAT SUITABILITY CLASS
1
2
3
4
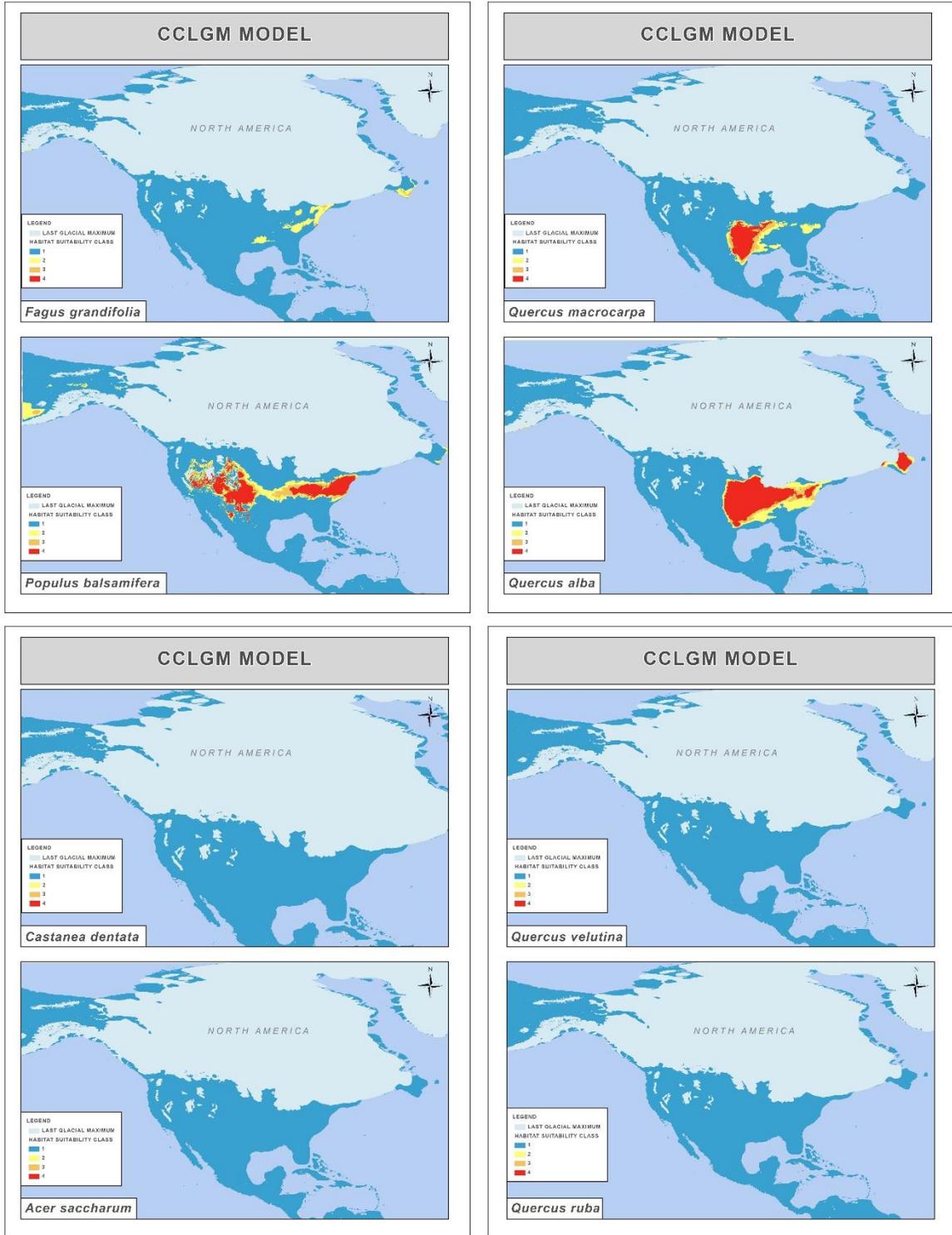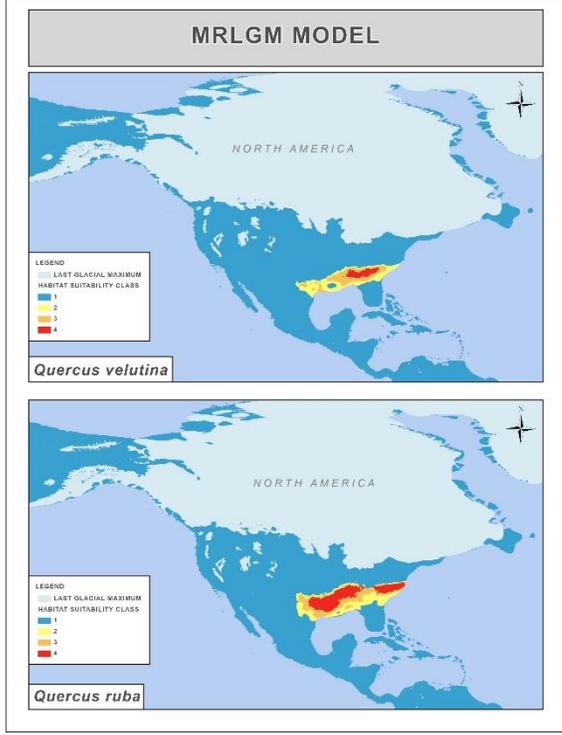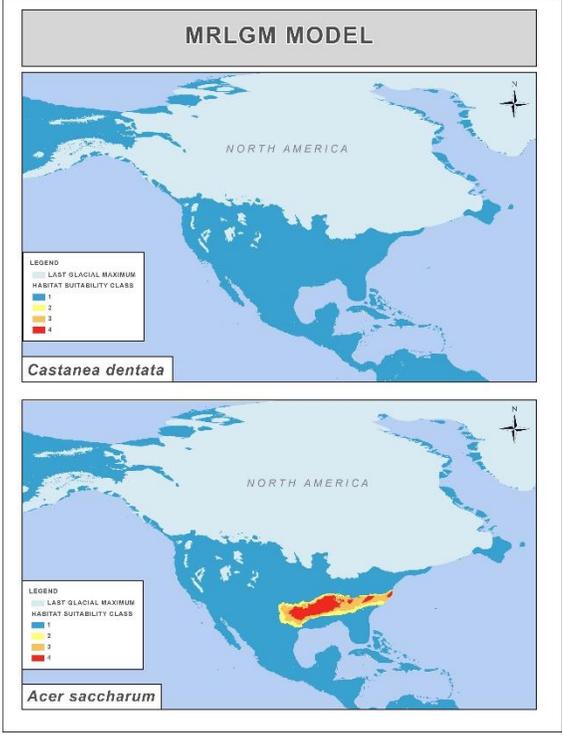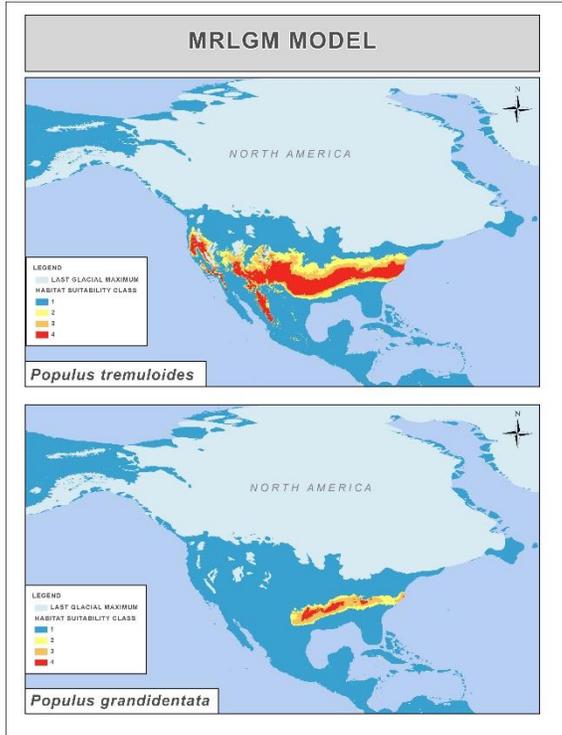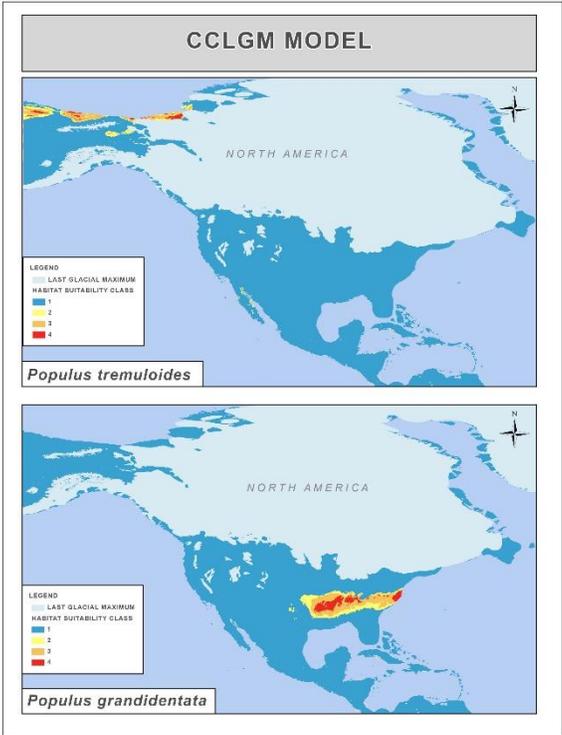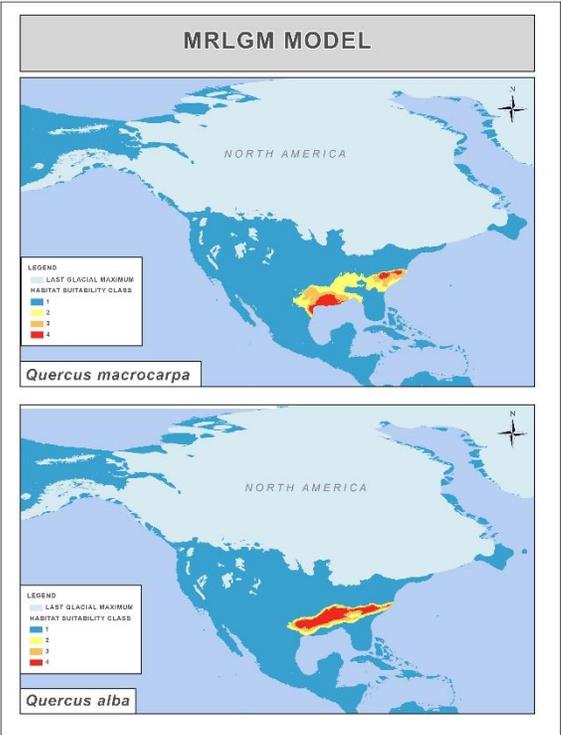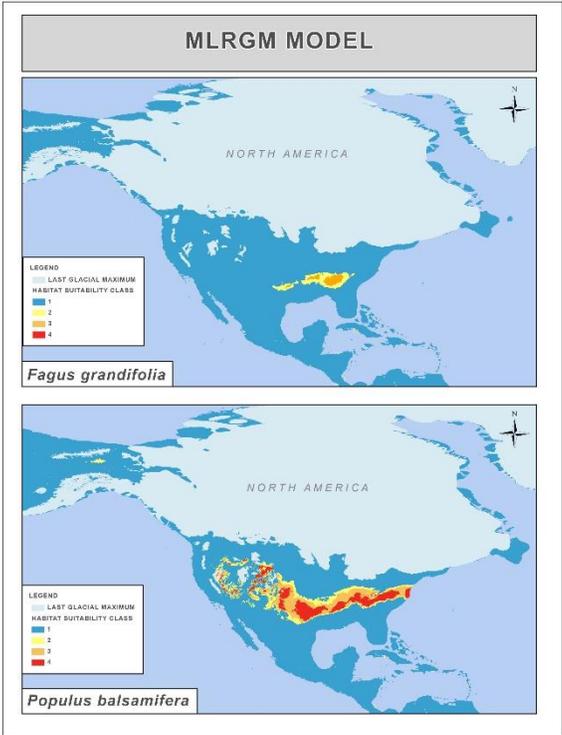
APPENDIX B - Maxent AUC Output Graphs

Average Sensitivity vs. 1 - Specificity for Quercus_rubra

Mean (AUC = 0.933) ■
Mean +/- one stddev ■
Random Prediction ■

Sensitivity (1 - Omission Rate)

1 - Specificity (Fractional Predicted Area)

Average Sensitivity vs. 1 - Specificity for Quercus_velutina

Mean (AUC = 0.957) ■
Mean +/- one stddev ■
Random Prediction ■

Sensitivity (1 - Omission Rate)

1 - Specificity (Fractional Predicted Area)
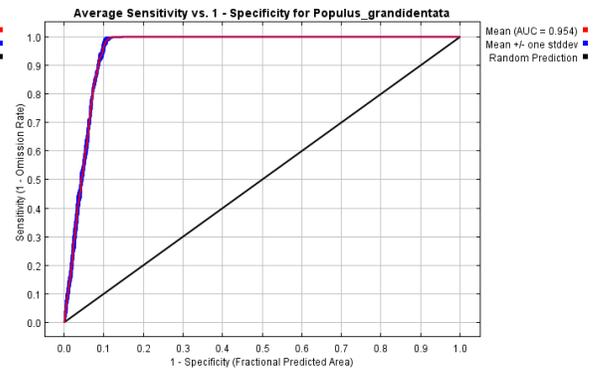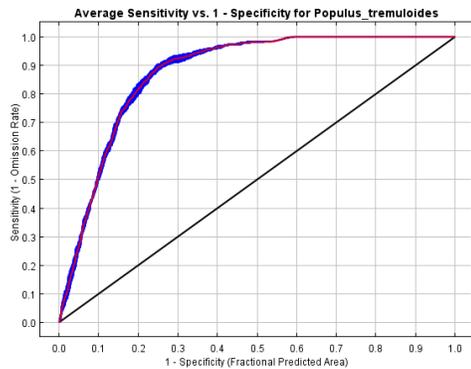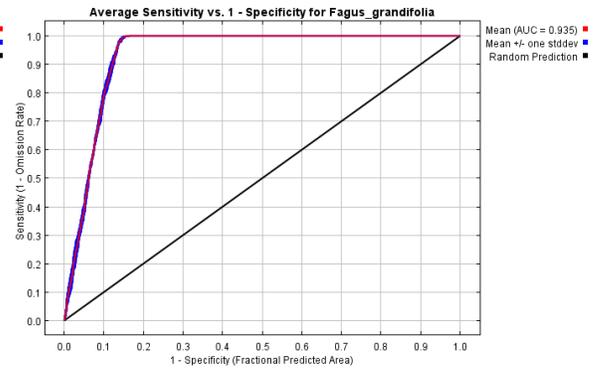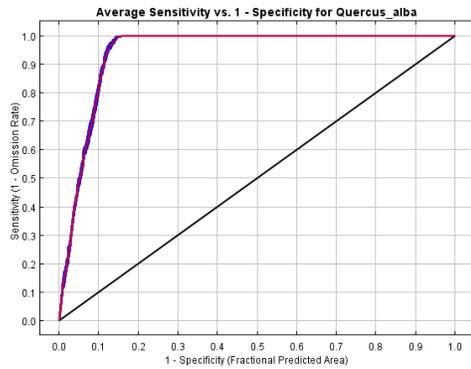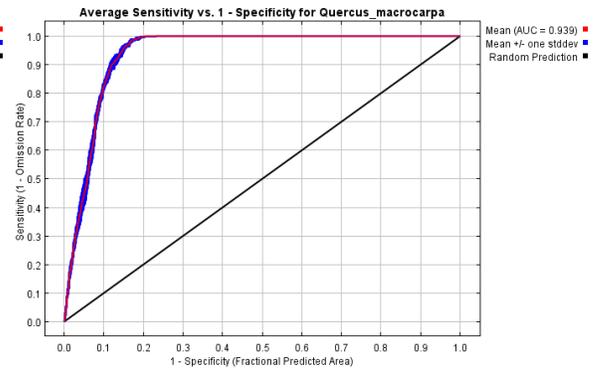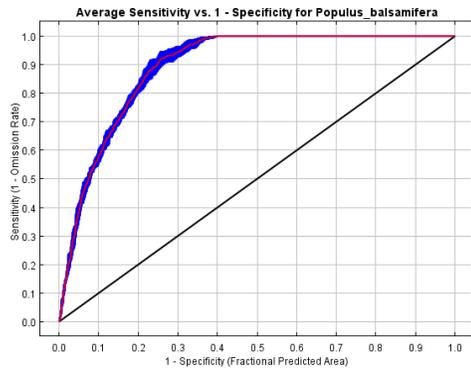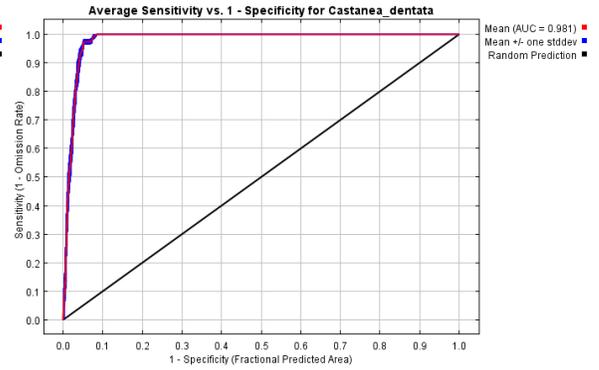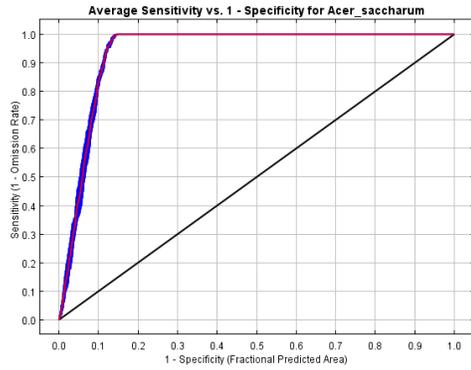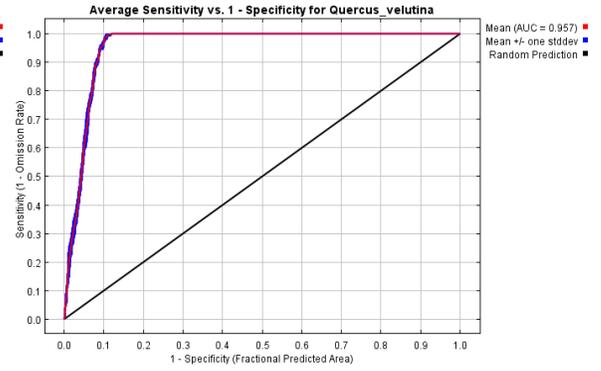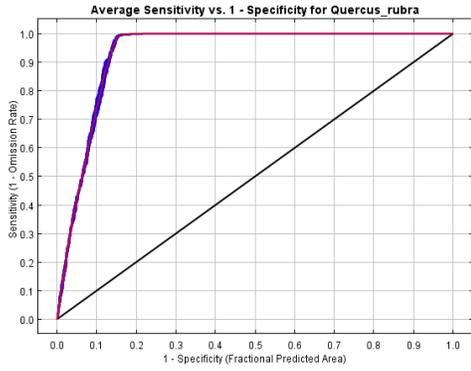
APPENDIX C – Gbif Occurrence Data Citations

GBIF.org (17 January 2020) GBIF Occurrence Download
https://doi.org/10.15468/dl.7yiuso

GBIF.org (17 January 2020) GBIF Occurrence Download
https://doi.org/10.15468/dl.4hc8sr

GBIF.org (17 January 2020) GBIF Occurrence Download
https://doi.org/10.15468/dl.slv4rv

GBIF.org (17 January 2020) GBIF Occurrence Download
https://doi.org/10.15468/dl.so5fnl

GBIF.org (17 January 2020) GBIF Occurrence Download
https://doi.org/10.15468/dl.ciw61o

GBIF.org (17 January 2020) GBIF Occurrence Download
https://doi.org/10.15468/dl.cfxx30

GBIF.org (17 January 2020) GBIF Occurrence Download
https://doi.org/10.15468/dl.zdtvfa

GBIF.org (12 January 2020) GBIF Occurrence Download
https://doi.org/10.15468/dl.1qjuk6

GBIF.org (10 November 2019) GBIF Occurrence Download
https://doi.org/10.15468/dl.6jrwu5

GBIF.org (10 November 2019) GBIF Occurrence Download
https://doi.org/10.15468/dl.axwrju