

RELATIONSHIP BETWEEN GLACIAL REFUGIAL RANGE
AND GENETIC DIVERSITY IN EASTERN NORTH
AMERICAN CONIFER SPECIES

by

DANIEL-JOSEPH GAREAU



FACULTY OF NATURAL RESOURCES MANAGEMENT
LAKEHEAD UNIVERSITY
THUNDER BAY, ONTARIO

April 2020

RELATIONSHIP BETWEEN GLACIAL REFUGIAL RANGE AND GENETIC
DIVERSITY IN EASTERN NORTH AMERICAN CONIFER SPECIES

by

Daniel-Joseph Gareau

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
Major Advisor

Mr. Tomislav Sapic
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: April 12, 2020

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

Gareau, D.J. Relationship between glacial refugial range and genetic diversity in eastern North American conifer species. 63 pp.

Keywords: conifer, distribution, eastern North America, forest genetics, genetic diversity, glacial maximum, paleodistribution modeling, refugia

The aim of this thesis is to relate genetic diversity to the extent of glacial refugia in several important North American conifer species. Refugial locations were hindcasted using MaxEnt software with occurrence points retrieved from the Global Biodiversity Information Facility (GBIF 2019) and Biodiversity Information Serving Our Nation (BISON 2019), and bioclimatic variables of current conditions and conditions during the last glacial maximum according to the MPI-ESM-P global circulation model retrieved from WorldClim version 1.4 (Hijmans *et al.* 2005). Bioclimatic variables were removed according to correlation to the highest contributing variable in multiple iterations until all highly correlated variables were removed to arrive at a final model. The size and number of refugia was compared against heterozygosity values gathered from allozyme studies. Species distribution models performed well with and were able to adequately predict current ranges of the ten selected conifer species and predicted Pleistocene distribution that can largely be corroborated with paleoecological and phylogeographical studies. A strong relationship between expected heterozygosity measured by allozyme analysis and the number and size of modelled refugia (adjusted $r^2 = 0.71$) suggests that population size reductions and reduced gene flow during the last glacial maximum had pronounced effect on the genetic diversity of Eastern North American conifer species. Of these two variables, the number of refugia was more closely correlated to expected heterozygosity than size of refugia (adjusted $r^2 = 0.67$ versus adjusted $r^2 = 0.58$) which could suggest that these multiple refugia preserved different novel alleles that resulted in higher genetic diversity when glaciers receded and population admixture occurred.

CONTENTS

ABSTRACT	v
TABLES	viii
FIGURES	x
ACKNOWLEDGEMENTS	xii
INTRODUCTION	1
LITERATURE REVIEW	3
GENETIC DIVERSITY	3
INFERRING GLACIAL REFUGIA	7
Fossil Records	8
Phylogeography	10
Species Distribution Modelling	14
METHODS	18
RESULTS	22
SPECIES DISTRIBUTION MODEL PERFORMANCE	22
<i>Abies balsamea</i>	24
<i>Larix laricina</i>	25
<i>Picea glauca</i>	26
<i>Picea mariana</i>	27
<i>Picea rubens</i>	28
<i>Pinus banksiana</i>	29
<i>Pinus resinosa</i>	30
<i>Pinus strobus</i>	31
<i>Thuja occidentalis</i>	32
<i>Tsuga canadensis</i>	33

RECONSTRUCTION OF GLACIAL REFUGIA.....	34
DISCUSSION.....	42
PLEISTOCENE SPECIES DISTRIBUTION.....	42
DETERMINANTS OF GENETIC DIVERSITY.....	48
CONCLUSION.....	49
LITERATURE CITED.....	51
APPENDICES.....	58
APPENDIX I CREATED GEOPROCESSING TOOLS.....	59
APPENDIX II AREA UNDER RECIEVER CURVE VALUES AND PERMUTATION IMPORTANCE OF BIOCLIMATIC VARIABLES	61
APPENDIX III PREDICTED REFUGIA LOCATION AND AREA IN SQUARE KILOMETRES OF TEN EASTERN NORTH AMERICAN CONIFER SPECIES	62

TABLES

Table	Page
1. Predicted current distribution of <i>Abies balsamea</i>	24
2. Predicted current distribution of <i>Larix laricina</i>	25
3. Predicted current distribution of <i>Picea glauca</i>	26
4. Predicted current distribution of <i>Picea mariana</i>	27
5. Predicted current distribution of <i>Picea rubens</i>	28
6. Predicted current distribution of <i>Pinus banksiana</i>	29
7. Predicted current distribution of <i>Pinus resinosa</i>	30
8. Predicted current distribution of <i>Pinus strobus</i>	31
9. Predicted current distribution of <i>Thuja occidentalis</i>	32
10. Predicted current distribution of <i>Tsuga canadensis</i>	33
11. Predicted distribution of <i>Abies balsamea</i> during the LGM.....	35
12. Predicted distribution of <i>Larix laricina</i> during the LGM.....	35
13. Predicted distribution of <i>Picea glauca</i> during the LGM.....	36
14. Predicted distribution of <i>Picea mariana</i> during the LGM.	36
15. Predicted distribution of <i>Picea rubens</i> during the LGM.	37
16. Predicted distribution of <i>Pinus banksiana</i> during the LGM.....	37
17. Predicted distribution of <i>Pinus resinosa</i> during the LGM.....	38

18. Predicted distribution of <i>Pinus strobus</i> during the LGM.	38
19. Predicted distribution of <i>Thuja occidentalis</i> during the LGM.....	39
20. Predicted distribution of <i>Tsuga canadensis</i> during the LGM.	39

FIGURES

Figure	Page
1. Predicted current distribution of <i>Abies balsamea</i> .	24
2. Predicted current distribution of <i>Larix laricina</i> .	25
3. Predicted current distribution of <i>Picea glauca</i> .	26
4. Predicted current distribution of <i>Picea mariana</i> .	27
5. Predicted current distribution of <i>Picea rubens</i> .	28
6. Predicted current distribution of <i>Pinus banksiana</i> .	29
7. Predicted current distribution of <i>Pinus resinosa</i> .	30
8. Predicted current distribution of <i>Pinus strobus</i> .	31
9. Predicted current distribution of <i>Thuja occidentalis</i> .	32
10. Predicted current distribution of <i>Tsuga canadensis</i> .	33
11. Predicted distribution of <i>Abies balsamea</i> during the LGM.	35
12. Predicted distribution of <i>Larix laricina</i> during the LGM.	35
13. Predicted distribution of <i>Picea glauca</i> during the LGM.	36
14. Predicted distribution of <i>Picea mariana</i> during the LGM.	36
15. Predicted distribution of <i>Picea rubens</i> during the LGM.	37
16. Predicted distribution of <i>Pinus banksiana</i> during the LGM.	37
17. Predicted distribution of <i>Pinus resinosa</i> during the LGM.	38

18. Predicted distribution of <i>Pinus strobus</i> during the LGM.	38
19. Predicted distribution of <i>Thuja occidentalis</i> during the LGM.	39
20. Predicted distribution of <i>Tsuga canadensis</i> during the LGM.	39

ACKNOWLEDGEMENTS

I would like to acknowledge the immense amount of her time that Dr. Ashley Thomson afforded me to discuss this thesis and her continued mentorship throughout this process. I was privileged to have such an accommodating and invested thesis supervisor.

I would also like to acknowledge my late grandfather, Dick Davidson of Sioux Lookout. Without his support, I would have likely had not been able to finish this degree.

INTRODUCTION

Significant changes in tree species distributions associated with glacial-interglacial cycles have had major consequences on the genetics of most organisms through the mechanics of repeated colonization, genetic bottlenecks, and admixture (Jaramillo *et al.* 2009). Throughout the Pleistocene Epoch, temperate tree species underwent repeated cycles of range expansion and contraction (Hewitt 2000). A relatively recent series of major shifts during this epoch was the last glacial maximum (LGM), the most recent occurrence of continental glaciers reaching their southernmost extent, and the subsequent retreat of these glaciers. This event was approximately 25,000 to 15,000 calendar years before present, when an area roughly one third of North America was under the coalesced Laurentide and Cordilleran ice sheets (Ray and Adams 2001). Along with up to two-kilometre ice sheets covering this landmass, conditions affecting tree habitat included sea levels being up to 130 meters lower than today (Yokoyama *et al.* 2018) and an average temperature of approximately 9° Celsius (Schneider von Deimling *et al.* 2006). Due to these factors, tree species would have existed within completely different ranges that were able to support them during this time.

The ranges likely existed as refugia; small areas separated allopatrically. *Ex situ*, they could have existed south of the glacial extent due to species following favourable climatic conditions during the advance of ice sheets (Rowe *et al.* 2004). Lower sea levels exposed continental and insular shelves, of which some areas remained unglaciated north of the southern glacial extent (Ray and Adams 2001), which could have harboured *ex situ* and *in situ* fragmented populations. *In situ* sites with locally hospitable climates within larger unfavourable areas, known as microrefugia, could have also existed for these species because of the effect of physiographic factors on meteorological conditions (Dobrowski 2011). Following glacial retreat, colonisation has occurred from these refugia into their current ranges (Rowe *et al.* 2004).

With species distribution modelling, it is possible to predict the historical distribution of these refugia by working backwards from contemporary occurrence points and historical climate grids (Philips and Dudik 2008). The objective of this study is to correlate the size and number of these refugia using this information along with genetic diversity measured by expected heterozygosity as provided by various allozyme studies of the trees in focus. Allozyme variation will be used due to the relative availability and their reliability in estimating gene flow (Ouborg *et al.* 1999). I anticipate softwood tree species within multiple and/or large refugia during the LGM have led to higher genetic diversity in modern day populations than those with fewer and/or smaller refugia. Further, species with multiple smaller refugia have experienced less of a genetic bottleneck than ones with a single or few large refugia.

LITERATURE REVIEW

GENETIC DIVERSITY

Conifers are genetically diverse compared to other plants and generally exhibit very low population genetic differentiation (Hamrick *et al.* 1992). Part of the reason for this is that extensive gene flow, even across fragmented landscapes, prevents population genetic divergence due to drift and associated loss of alleles associated with finite population size (O'Connell *et al.* 2007). Hamrick *et al.* (1992) found that a significant proportion of genetic variation within conifer species can be explained by five life history traits: geographic range, distribution within the USA, successional stage, habitat type, and cone type. Other life history traits important for most plant species are common in conifer species. These findings could explain part of the genetic variation seen in the species included in this thesis.

Ten conifer species from eastern North America are included in this thesis: *Abies balsamea*, *Larix laricina*, *Picea glauca*, *Picea mariana*, *Picea rubens*, *Pinus banksiana*, *Pinus resinosa*, *Pinus strobus*, *Thuja occidentalis*, and

Tsuga canadensis. These species have an average expected heterozygosity of 0.128 according to published allozyme analyses (Table 1). According to Little (1971), *Larix laricina*, *Picea glauca*, and *Picea mariana* have large ranges extending from the eastern coast of Canada and New England to Alaska. These three species have the highest published expected heterozygosity values (Cheliak *et al.* 1988; Furnier *et al.* 1990; Perry and Bousquet 2000). *Abies balsamea* and *Pinus banksiana* have large ranges that do not cross the Rocky Mountains (Little 1971). Both species have lower expected heterozygosity than the calculated average of the selected species, with *Abies balsamea* being considerably lower (Godt *et al.* 2001; Shea and Furnier 2002). *Pinus resinosa*, *Pinus strobus*, *Thuja occidentalis*, and *Tsuga canadensis* have ranges restricted to eastern North America (Little 1971) and have lower than average genetic diversity apart from *Pinus strobus*. *Pinus strobus* has the largest proportion of its range within the USA (Little 1971) and the largest expected heterozygosity (Beaulieu and Simon 1994). *Tsuga canadensis* has a likewise large proportion of distribution within the USA (Little 1971) yet has a much lower expected heterozygosity than average (Zabinski 1992).

Table 1. Published expected heterozygosity based on allozyme analysis for ten eastern North American conifer species.

Species	He	Reference
<i>Picea glauca</i>	0.290	(Furnier <i>et al.</i> 1990)
<i>Picea mariana</i>	0.250	(Perry and Bousquet 2000)
<i>Larix laricina</i>	0.220	(Cheliak <i>et al.</i> 1988)
<i>Pinus strobus</i>	0.180	(Beaulieu and Simon 1994)
<i>Pinus banksiana</i>	0.114	(Godt <i>et al.</i> 2001)
<i>Thuja occidentalis</i>	0.094	(Perry <i>et al.</i> 1990)
<i>Picea rubens</i>	0.079	(Hawley and DeHayes 1994)
<i>Tsuga canadensis</i>	0.037	(Zabinski 1992)
<i>Abies balsamea</i>	0.015	(Shea and Furnier 2002)
<i>Pinus resinosa</i>	0.000	(Fowler and Morris 1977)

The geographic distribution traits Hamrick *et al.* (1979) found to be significant could be due to unconsidered factors related to distribution of glacial refugia. Species with distribution on either side of the Rocky Mountains as well as either side of the Appalachian Mountains would by consequence have larger ranges. This distribution could be due to these areas being representative of separated refugia, which has been supported for several of the selected species (Cinget *et al.* 2015; Godbout *et al.* 2010; Jaramillo-Correa *et al.* 2004; Lafontaine *et al.* 2010). The importance found for distribution within the US could be due these populations representing *in situ* refugia that have persisted since the LGM, which could mean they have suffered less severe range contractions and loss of genetic diversity associated with post-glacial founding events. Within western North America, Roberts and Hamann (2015) found larger genetic diversity from tree species that had many and large glacial refugia. Species with restricted refugia had less genetic diversity. This trend emerged regardless of species'

present day distribution. Genetic diversity is also generally higher in boreal species with current day distribution partially or completely within eastern North America. There seems to be little correlation between shade tolerance (Table 2) and expected heterozygosity within the ten selected species in this species.

Table 2. Shade tolerance of ten eastern North American conifer species.

Species	Shade Tolerance
<i>Abies balsamea</i>	Very tolerant
<i>Tsuga canadensis</i>	Very tolerant
<i>Picea mariana</i>	Tolerant
<i>Picea rubens</i>	Tolerant
<i>Picea rubens</i>	Tolerant
<i>Thuja occidentalis</i>	Tolerant
<i>Picea glauca</i>	Intermediate
<i>Pinus strobus</i>	Intermediate
<i>Pinus banksiana</i>	Intolerant
<i>Pinus resinosa</i>	Intolerant
<i>Larix laricina</i>	Very intolerant

Source: (Burns and Honkala 1965).

Genetic variation determined by allozyme analysis has several advantages when collecting genetic diversity of many species. Heterozygosity values from allozyme analysis are generally more available than from microsatellite markers as they are older (Bush and Smouse 1992). They are also reliable in estimating gene flow and genetic diversity (Ouborg *et al.* 1999).

INFERRING GLACIAL REFUGIA

Two research methods have traditionally been used in conjunction to infer glacial refugial locations. The first is the paleoecological approach which examines pollen records and macrofossils (Delcourt and Delcourt 1987). These samples are dated and used to reconstruct migration routes following glacial retreat and glacial refugia (Gavin *et al.* 2014; Roberts and Hamann 2015). The second research approach is phylogeography, which examines modern day geographic distribution of neutral genetic markers (Gavin *et al.* 2014). This distribution reflects the historical barriers between subpopulations of a species. Genetic variation at neutral genetic markers can be used to identify populations that occupied separate glacial refugia based on the variation of allele frequencies among population clusters (Gavin *et al.* 2014; Jaramillo-Correa *et al.* 2009).

A third approach to delineating glacial refugial locations has been developed in contemporary studies. This method is to build Species Distribution Models (SDMs) to correlate species occurrence data with the climatic variables they exist in (Dormann *et al.* 2012; Morin and Thueller 2009). This model is then used to hindcast to determine the distribution of suitable habitat during the LGM. All three approaches have their own sets of assumptions and shortcomings and

are best used in conjunction to address and mitigate these (Gavin *et al.* 2014; Roberts and Hamann 2015).

Fossil Records

Advances in data management have allowed archiving of large curated collections of macro-fossils and pollen fossils from lake sediments (Gavin *et al.* 2014). Pollen records are generally able to prove presence of trees to the genus level, while specific species need macrofossils to provide strong evidence of presence (Delcourt and Delcourt 1987). Macrofossils found *in situ* provide near-definitive evidence for species presence in an area within a given time period. Analysis of multiple sites containing high abundance of microfossils, such as pollen, can suggest presence of their source species with high confidence. Radiocarbon dating can be used to determine the age of micro- and macrofossils with relatively high precision (Gavin *et al.* 2014). However, a major limitation of fossil records is the rarity of samples. Glacial refugia are understood to have been restricted and small, which means they left a likewise small impact on the fossil record. Rare taxa see a similar issue (Roberts and Hamman 2015). Areas suitable for potential glacial refugia are typically south-facing slopes or exposed continental shelves which both are ill-suited for fossil preservation. For these reasons, absence is rarely inferred from lack of fossil samples alone. Refugia not represented in the fossil record are referred to as cryptic refugia (Gavin *et al.* 2014).

Pollen records show the existence of glacial refugia in Eastern North America between the southern edge of the Laurentide continental glacier and modern-day Florida for the genera *Abies*, *Larix*, *Picea*, *Pinus*, and *Tsuga* (Delcourt and Delcourt 1987; Jackson *et al.* 1997). Delcourt and Delcourt (1987) have also suggested presence of the family Cupressaceae, which contains *Thuja occidentalis*. *Picea glauca* and *P. mariana* dominated forests directly south of the Laurentide ice sheet towards the Gulf Coast. *Pinus*, mostly *P. banksiana* with diffuse populations of *P. resinosa* and *P. strobus*, occurred as forest from 34° to 30° north. Macrofossils have also confirmed presence of *Abies balsamea*, *Larix laricina*, and *Tsuga canadensis* within this same general area (Jackson *et al.* 1997).

Pollen microfossils of *Thuja* spp. are not confidently distinguished from those of other members of the Cupressaceae family (Yu 1997). Vegetative macrofossils have not confirmed presence of *Thuja occidentalis* during the LGM, although they have confirmed presence within Southern Ontario after the late glacial period at least as early as 7,500 Years BP. It has been suggested that this species moved along the Niagara Escarpment and south into Southern Ontario following glacial retreat after a considerable lag period (Yu 1997). This is supported by the findings of the oldest macrofossils being found in the Bruce Peninsula dated at 8,200 Year BP (Yu 1997) and Manitoulin Island dated at 10,000 Years BP (Warner 1981). This would suggest a glacial refugia of *Thuja occidentalis* south of Lake Superior, but accurate placement is not possible without macrofossil evidence dated to the LGM.

Beringia, the exposed continental shelf between modern day Alaska and Siberia, remained unglaciated during the LGM (Ray and Adams 2001). Brubaker *et al.* (2005) suggests that according to pollen records, *Larix* spp. was restricted to sites within southwestern Beringia during this time, while *Picea* spp. was present throughout much of eastern Beringia. Though the evidence is unclear, macrofossil evidence of *L. laricina* within eastern Beringia following glacial retreat suggest this was the species of *Larix* that found refugia there. *Picea* was likely represented by multiple species. Presence of *P. glauca* was found to be most widespread in eastern Beringia. *P. mariana* was found to have accompanied *P. glauca* during site colonization in northern British Columbia, which suggest its earlier presence in eastern Beringia as well. Pollen of *Pinus* spp. was also found within Beringia during the LGM, but macrofossil evidence has not suggested presence of the three species included in this thesis.

Phylogeography

A major limitation of phylogeography is that evolutionary events inferred from genetic markers can only be linked to general areas of divergence. They can however provide evidence of refugia that have no supporting evidence within the fossil record, which are known as cryptic refugia (Roberts and Hamman 2015).

Mountains have proven to have had a large effect of the phylogeographic patterns of trees, but this effect is lessened in North America as mountains run

north to south and do not block gene flow as effectively as European mountains which run east to west (Jaramillo-Correa *et al.* 2009). Roberts and Hamann (2015) found larger genetic diversity from tree species that had many and large refugia within western North America. Species with restricted refugia had less genetic diversity. Genetic diversity is generally higher in eastern boreal species which suggests a more disjointed glacial history than western counterparts. Some now submerged Atlantic coastal areas could have contributed to current genetic diversity by harbouring cryptic refugia (Gobout *et al.* 2010).

The phylogeographic structure of *Abies balsamea* supports five distinct genetic lineages resulting from four glacial refugia south of the Laurentide ice sheet, and one northeast of the southern extent. One was inferred to have existed south of the Great Lakes, which was likely to have been the largest. A second refugium was proposed to have been within the Driftless Area west of the Great Lakes. This was likely very limited in size and is considered a cryptic refugium. Another refugium was possibly within the southern Appalachian Mountains, and the last southern refugium was likely on the Atlantic Coast. These refugia could have been completely separated during the LGM or could have represented multiple recolonization routes from fewer combined refugia. The northeastern refugium was likely near modern day Labrador (Cinget *et al.* 2015).

Populations of *Larix laricina* within Alaska show no evidence of gene flow from the larger Canadian population. Due to this and their low genetic diversity, these populations are proposed to be remnants of a glacial refugia within

Beringia (Napier et al 2020). Further phylogeographic evidence suggests a large eastern refugium was located south of the Laurentide ice sheet. Unlike other conifer species, the Appalachian Mountains seem to have not restricted gene flow as this refugium seems to have been on either side. Lastly, a third refugium has been inferred to be within the Driftless Area, an area west of the Great Lakes at the margin of the Laurentide ice sheet (Warren *et al.* 2016).

Anderson *et al.* (2006) reported that most haplotypes of *Picea glauca* within Alaska are unique to that region. This offers strong evidence that this species survived the LGM in or near this region, likely in low densities according to corroboration with pollen records. Lafontaine *et al.* (2010) supported this inference, as well as provided phylogeographic evidence of two eastern refugia. These were defined to distinct areas east and west of the Appalachian Mountains.

Jaramillo-Correa *et al.* (2004) found a distinct geographic variation in mitochondrial DNA within the range wide population of *Picea mariana*. Three southern and one northeastern refugia were inferred from this distribution. *P. mariana* was suggested to be restricted to a glacial refugia south of the Cordilleran ice sheet in Washington or Oregon State within Western North America due to the Rocky Mountains. Within unglaciated Eastern North America, two refugia were likely to have been present, with one on either side of the Appalachian Mountains. A group of mitotypes observed only in Labrador and Eastern Quebec suggests a glacial refugia of this species on exposed Atlantic coastal shelf north of the glacial extent.

Picea rubens is thought to have speciated from a shared progenitor species with *P. mariana* during the LGM (Perron *et al.* 2000). Higher genetic variability than expected within *P. rubens* within the northern half of its range could be due to it being present in two different glacial refugia, one within the southern Appalachian Mountains and another on exposed coastal shelf in the Atlantic Ocean (Hawley and DeHayes 1994).

Godbout *et al.* (2010) found that populations of *Pinus banksiana* within the Maritimes region of Canada were genetically distinct from central and western populations. This genetic structure suggests a glacial refugium northeast of the extent of glaciation. Further analysis pointed to two additional glacial refugia: on the Atlantic coast south of the glacial extent, and west of the Appalachian Mountains.

Pinus resinosa is considered unique among North American as it has extremely low genetic diversity, yet it is widespread geographically (Fowler and Morris 1977). Eastern *Pinus* species are generally thought to have survived in the southeastern USA. Despite its low genetic diversity, it is unlikely that this species existed in a single glacial refugium. Instead, it has been inferred that another refugium was present in unglaciated areas in the northeast USA (Walter and Epperson 2005).

Zinck and Rajora (2016) has suggested a single refugium of *Pinus strobus* in central North American south of the Laurentide ice sheet which branched into two recolonization routes, resulting in three genetically different

lineages by the modern day. This refugium was suggested to be in a single large area south of the Laurentide ice sheet.

Potter *et al.* (2012) suggested that the most likely area for glacial refugia of *Tsuga canadensis* was the Southeast USA. This is due to the relatively high genetic diversity found within populations closer to there. More exactly, a glacial refugium was placed southeast of the Appalachian Mountains. There is evidence that this large refugium was partially split into three clusters. It is possible that another area of refugium existed on exposed Atlantic coastal shelf.

Species Distribution Modelling

SDMs use differing approaches to identify the determinants of a species suitable habitat and predict their suitable ranges either over space, time, or both (Dormann *et al.* 2012; Gavin *et al.* 2014; Morin and Thuiller 2009). Applications for SDMs include predicting species' range shifts under projected climate change (Morin and Thuiller 2009; Schneiderman *et al.* 2015), modelling the spread of invasive species (Higgins *et al.* 2000; Roy-Dufresne *et al.* 2019), and hindcasting refugia during glaciation events (Svenning *et al.* 2008). These models exist within two main classes, process-based and correlative, with hybrid models existing between the two extremes (Dormann *et al.* 2012; Morin and Thuiller 2009). Process based SDMs formulate a mathematical function to represent the ecology of a species. This means that processed based SMDs are causal, explicitly relating a species' performance to environmental conditions

(Dormann *et al.* 2012; Gavin *et al.* 2014). These models use a bottom up approach that simulates site level interactions between species and environment, then expanding to landscape and regional scales. This means it is necessary to have a deep knowledge of a species life history, usually derived from observations from multiple individuals in controlled or laboratory environments which is only available for the most well studied species. Species occurrence data is not strictly required (Morin and Thueller 2009; Schneiderman *et al.* 2015). These properties make process based SDMs most applicable for stand or landscape scale distribution models (Lischke *et al.* 2006), outlining habitat outside of a species fundamental niche (Evans *et al.* 2016) and identifying future limitations to current ranges (Morin *et al.* 2008). Along with the relative intensity of study needed for the inputs, these SDMs have been criticized due to their larger parameterization demands (Araújo and Guisan 2006; Gavin *et al.* 2014).

Correlative SDMs directly relate environmental variable to observed species distribution. Instead of causal, these models use statistical relationships to make predictions (Dormann *et al.* 2012; Morin and Thueller 2009). These models employ a top down approach by beginning with statistically relating regional scale climatic variables usually in the form of climate grids to species occurrence, then adding landscape scale variables such as soil to improve the prediction (Morin and Thueller 2009; Schneiderman *et al.* 2015). Correlative models necessitate presence/absence data to draw predictions, but do not strictly require ecological knowledge (Dormann *et al.* 2012). MaxEnt, the

program selected for use in this thesis, is one of many available within this class (Jamevich *et al.* 2015). Several assumptions and shortfalls of correlative SDMs must be considered before implementation, as well as limitations of the occurrence data and climate grids used for input. Firstly, a correlative SDM provides predictions of suitability, not a guarantee of species presence. Additionally, these predictions are of realized niche rather than fundamental niche and this can be affected by factors not covered by input variables such as anthropogenic influence, competition, and dispersal limitations (Gavin *et al.* 2014). These are also static models that do not consider changing relationships in abiotic and biotic conditions (Dormann *et al.* 2012). Due to these considerations, SDMs are to be treated as a hypothesis and require their own validation (Jamevich *et al.* 2015). In the case of paleodistribution modeling of climate refugia, paleoecological evidence, at both micro- and macro-scales, and phylogeographic studies should be used to supplement models produced (Gavin *et al.* 2014).

An assumption of correlative SMDs is that inputs represent a species at equilibrium with its environment which means that using species records from more unstable settings, such as climate changes or invasions, underrepresent current suitable habitat, which would in turn underrepresent future species distribution (Dormann *et al.* 2012; Elith and Leathwick 2009). Violation of this assumption is necessary when modelling species constrained by invasion processes during early stages (Barbet-Massin *et al.* 2018; Václaví and Meentemeyer 2012).

Proper data handling must be used to ensure that highly correlated climate input variables are not included, as this can result in “overfitting” of resultant SDMs. (Warren and Seifert 2011). Causality of correlations is critical to determine as variables are often correlated to one another. An example of this would be relative humidity seeming to be a direct limiting factor of a species distribution when the actual limiting factor is one such as moisture levels that determines relative humidity (Dormann *et al.* 2012).

SDMs are sensitive to both sample biases and sample size. These attributes must be addressed in order to produce accurate habitat predictions (Araújo and Guisan 2006). As frequent sources of input data are natural history collections (NHCs), associated biases should be understood. NHCs often suffer from spatial biases reflecting opportunistic sampling from easily accessible locations such as roadside. High elevations and wet areas can be underrepresented due to access difficulties (Wiz *et al.* 2008). Additional area for bias from occurrence data is collectors bias, where NHCs are biased towards collecting novel and rare specimens (Araújo and Guisan 2006).

MaxEnt is an SDM program that utilizes presence-only records but does not use absence points which are instead inferred (Philips *et al.* 2019; Philips *et al.* 2017). This means MaxEnt is sensitive to under-representation due to unrealized portions of fundamental niches (Elith *et al.* 2010). Portions of habitat niche can also be missed due to the spatial resolution of climate input variables if it is too large to reflect changes in variables that exist within smaller scales (Guisan *et al.* 2008).

METHODS

Ten tree species were selected for study within this thesis. The criteria for selection were as follows; Gymnosperms, generally single stemmed and taller than 20 feet when mature (*ie.* tree species rather than shrubs), modern day range within the area occupied by continental glaciers during the LGM, and presence within Eastern North America. Occurrence points for these species were obtained from the Global Biodiversity Information Facility (GBIF 2019) and Biodiversity Information Serving Our Nation (BISON 2019). The downloads were in .csv format and combined into one file per species retaining the species name, longitude, and latitude fields in that order. These files were added to an ArcMap session, plotted by longitude and latitude, and were constrained to their natural ranges identified by Little (1971) using the clip tool. The clipped out outliers likely represent individuals within botanical gardens or otherwise introduced outside their native ranges. Random points were then created with the clipped species occurrences acting as the constraining feature class. Points were functionally unconstrained in the total number of points but constrained to be a minimum of 50 kilometres apart in order to reduce spatial clustering. Care was taken to retain an appropriate number of points as reducing these too much

can affect the final model (Wisiz *et al.* 2008). These steps were performed using a created model tool (Appendix I). These final shapefiles were converted to a single .csv table using the table to excel tool in ArcMap containing all ten species and coordinates.

Global environmental data were obtained at a 2.5 arcminute resolution from WorldClim version 1.4 (Hijmans *et al.* 2005). These .tiff files contain current conditions representative of 1960-1990 and past conditions based on three general circulation models (GCMs) (CCSM4, MIROC-ESM, and MPI-ESM-P) reflecting conditions approximately 22,000 years ago. These 19 climatic variables (Table 3) were clipped to the extent of North America and converted to ASCII format using the raster to ASCII tool in ArcMap.

Table 3. Bioclimatic variables used for SDM building.

Code	Description of variable	Unit
BIO1	Annual mean temperature	C° *10
BIO2	Mean diurnal Range (Mean of monthly (maximum temperature – minimum))	C° *10
BIO3	Isothermality (BIO2/BIO7) (*100)	-
BIO4	Temperature seasonality (standard deviation *100)	-
BIO5	Maximum temperature of warmest month	C° *10
BIO6	Minimum temperature of coldest month	C° *10
BIO7	Temperature annual range (BIO5-BIO6)	C° *10
BIO8	Mean temperature of wettest quarter	C° *10
BIO9	Mean temperature of driest quarter	C° *10
BIO10	Mean temperature of warmest quarter	C° *10
BIO11	Mean temperature of coldest quarter	C° *10
BIO12	Annual precipitation	mm
BIO13	Precipitation of wettest month	mm
BIO14	Precipitation of driest month	mm
BIO15	Precipitation seasonality (coefficient of variation)	mm
BIO16	Precipitation of wettest quarter	mm
BIO17	Precipitation of driest quarter	mm
BIO18	Precipitation of warmest quarter	mm
BIO19	Precipitation of coldest quarter	mm

Source: (Hijmans *et al.* 2005)

Correlation coefficients were calculated for pairwise comparisons between each bioclimatic variable using ENMTools (Warren *et al.* 2009). SDMs were then constructed for each species using MaxEnt (Phillips *et al.* 2019). First, a model was built with all 19 climatic variables. Five replicates were performed for every model run, and values averaged. This means 20% of the points were used for testing while 80% were used for training in any given replicate. The points used for testing changed between replicates.

Jackknife resampling was used to evaluate the importance of individual variables to the model. Variables were removed in an iterative process according to correlation to the highest contributing variable. In the second model iteration, variables were removed according to contribution to the second highest contributing variable, in the third iteration according to the third highest contributing variable, and so on until only uncorrelated variables were left. Once the models were constructed, they were projected for the LGM using the bioclimatic variables from the three GCMs. This predicted refugia suitable for these species during this time frame. The output rasters were reclassified, converted to polygons, and projected according to a created model tool (Appendix I). Species habitat index values under 0.33 within the output rasters were removed when converting to feature classes in order to remove noise and define the predicted core habitat of the target species.

The resultant multipart polygons were opened in ArcMap and individual refugia selected. Refugia in this sense were, in some instances, multiple polygons occupying different geographic areas but belonging to the same

multipart feature. They were then split into single polygon feature classes and had areas calculated in square kilometres using a third created model tool (Appendix I). These polygon feature classes were merged into polygon feature classes representing predicted species refugia according to each GCM using a fourth created model tool. (Appendix I). These species polygon feature classes were then merged into a polygon for each GCM. These three GCM polygon feature classes were then merged into a final polygon feature class and converted to an excel file using the table to excel tool in ArcMap. The resultant table had the following fields in this order: Global Circulation Model, Species, Square Kilometres.

Using a pivot table in Microsoft Excel, number of identified refugia and total area was calculated for each species and GCM. Retrieved heterozygosity values from allozyme studies were recorded for each species from published literature. The total area and number of refugia were then compared with expected heterozygosity values using simple linear regression and multiple linear regression for each species using the regression tool in the analysis toolpack of Excel. The results for the set of predicted refugia that was best supported by other lines of evidence (*i.e.* paleoecological and phylogeographical) was presented in the results. Current distribution predictions were quantified using the same parameters as those used for delineating glacial refugia. The calculated area was then compared to area of distribution predicted during the LGM.

RESULTS

SPECIES DISTRIBUTION MODEL PERFORMANCE

Two bioclimatic variables were retained in every SDM: temperature seasonality and precipitation seasonality. Isothermality was retained in nine of ten SDMs. Maximum temperature of the warmest month was the variable with the highest permutation importance for six of the ten SDMs, mean temperature of warmest quarter was highest for three models, and mean annual temperature had the highest permutation importance for one species (Table 4). In terms of receiver operating curves, models resulted in area under curve (AUC) values between 0.869 and 0.985. The models were able to generally predict current distribution that matched published range maps (Little 1971). Southern range edges were overpredicted for most species. Current predictions of each selected species are presented within their respective sections, as well as AUC values and permutation importance of retained variables. A summary table of these numbers is within Appendix II.

Table 4. Retention and permutation importance of 19 bioclimatic variables within ten SDMs.

Code	Description of variable	Retention in model (# of models)	Highest permutation importance (# of models)
BIO1	Annual mean temperature	1	1
BIO2	Mean diurnal Range	1	-
BIO3	Isothermality	9	-
BIO4	Temperature seasonality	10	-
BIO5	Maximum temperature of warmest month	6	6
BIO6	Minimum temperature of coldest month	0	-
BIO7	Temperature annual range	0	-
BIO8	Mean temperature of wettest quarter	1	-
BIO9	Mean temperature of driest quarter	0	-
BIO10	Mean temperature of warmest quarter	3	3
BIO11	Mean temperature of coldest quarter	0	-
BIO12	Annual precipitation	1	-
BIO13	Precipitation of wettest month	0	-
BIO14	Precipitation of driest month	1	-
BIO15	Precipitation seasonality	10	-
BIO16	Precipitation of wettest quarter	4	-
BIO17	Precipitation of driest quarter	6	-
BIO18	Precipitation of warmest quarter	4	-
BIO19	Precipitation of coldest quarter	1	-

Abies balsamea

The SDM constructed for *Abies balsamea* was able to predict current distribution well along the northern and eastern range edges according to Little's (1971) range map (Figure 1). It extended predictions slightly past the southern range edge, and poorly predicted the northwestern range edge. Seven uncorrelated bioclimatic variables were retained in the final model according to their permutation importance: annual mean temperature (40.7%), precipitation of warmest quarter (37%), precipitation seasonality (6.7%), temperature seasonality (6.0%), mean temperature of wettest quarter (5.1%), precipitation of driest quarter (3.9%), and mean diurnal range (0.5%). The AUC value for this model was 0.940.

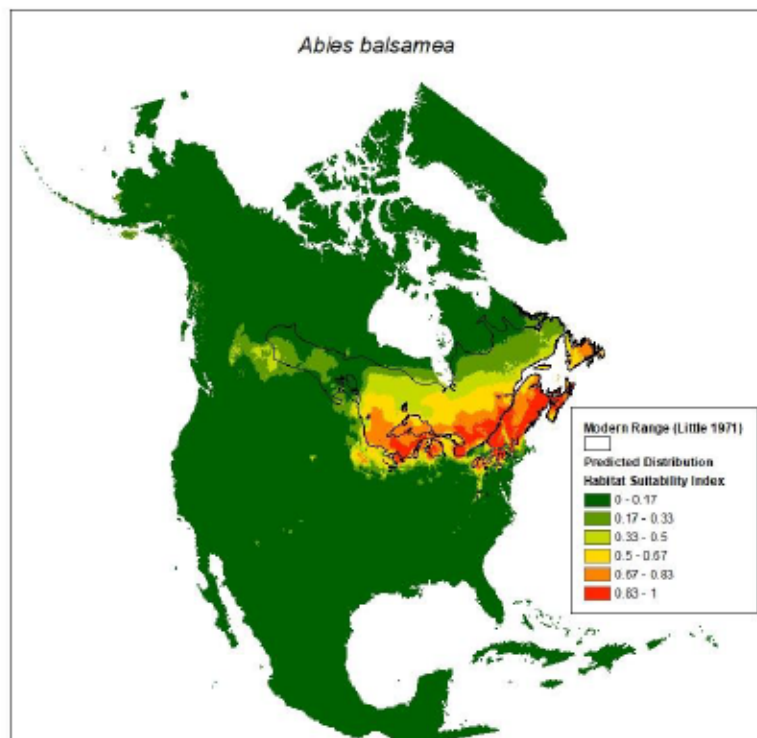


Figure 1. Predicted current distribution of *Abies balsamea*.

Larix laricina

The SDM constructed for *Larix laricina* was able to predict current distribution well along the northern, eastern, and western edges of the main range according to Little's (1971) range map (Figure 2). It extended predictions past the southern range edge and the disjunct Alaskan range. Six uncorrelated bioclimatic variables were retained in the final model according to their permutation importance: maximum temperature of warmest month (50.6%), precipitation of warmest quarter (25.8%), isothermality (12.3%), precipitation seasonality (5.0%), temperature seasonality (4.5%), and precipitation of driest quarter (1.8%). The AUC value for this model was 0.891.

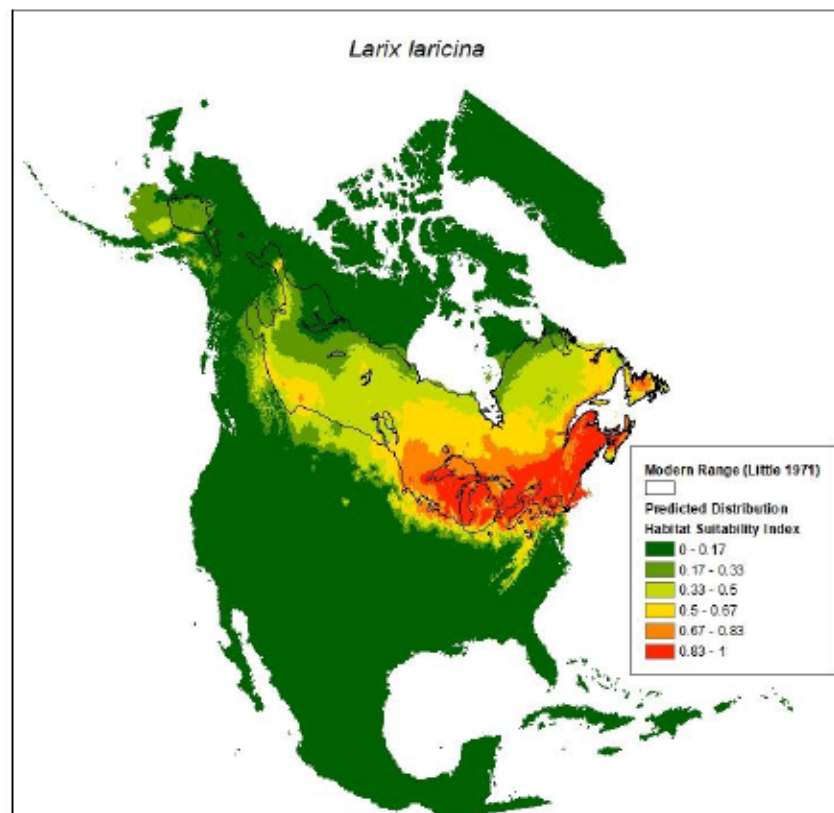


Figure 2. Predicted current distribution of *Larix laricina*.

Picea glauca

The SDM constructed for *Picea glauca* was able to predict current distribution along the northern, eastern, and western range edges according to Little's (1971) range map (Figure 3). It overextended predictions along the southern range edge slightly, as well as predicted as cluster of low habitat suitability southwest of the range. Six uncorrelated bioclimatic variables were retained in the final model according to their permutation importance: maximum temperature of warmest month (69.9%), isothermality (19.3%), temperature seasonality (5.7%), precipitation of coldest quarter (5.7%), precipitation of warmest quarter (1.2%), and precipitation seasonality (1.2%). The AUC value for this model was 0.869.

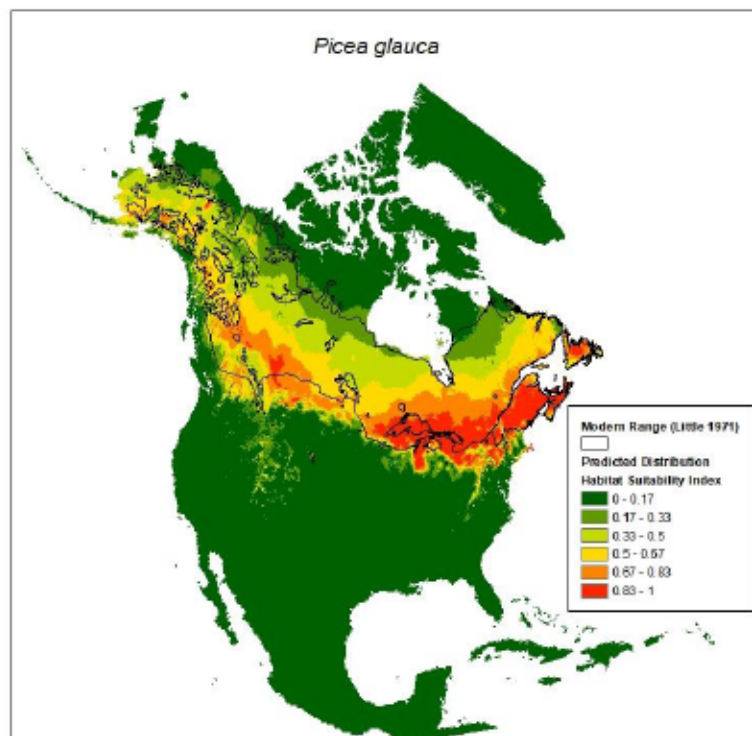


Figure 3. Predicted current distribution of *Picea glauca*.

Picea mariana

The SDM constructed for *Picea mariana* was able predict current distribution along the northern, eastern, and western range edges according to Little's (1971) range map (Figure 4). It overextended predictions along the southern range edge slightly, as well as an over prediction following the Appalachian Mountains south. Five uncorrelated bioclimatic variables were retained in the final model according to their permutation importance: maximum temperature of warmest month (46.2%), precipitation of warmest quarter (18.8%), isothermality (13.8%), precipitation seasonality (12.0%), precipitation of driest quarter (6.5%), and temperature seasonality (2.7%). The AUC value for this model was 0.876.

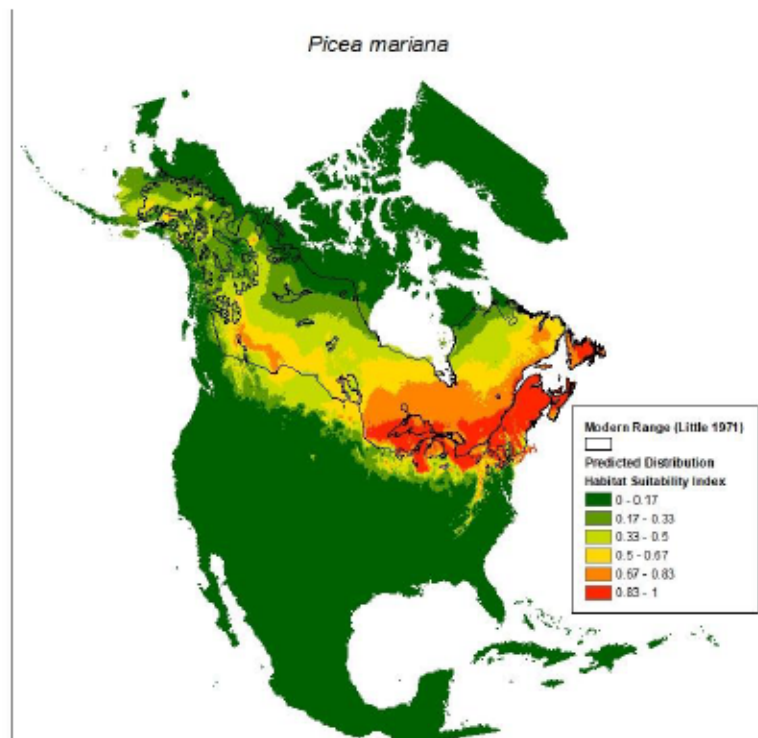


Figure 4. Predicted current distribution of *Picea mariana*.

Picea rubens

The SDM constructed for *Picea rubens* was able to predict current distribution along all range edges according to Little's (1971) range map (Figure 5). It oversimplified the disjointed range clusters southwest of the main range. Six uncorrelated bioclimatic variables were retained in the final model according to their permutation importance: mean temperature of warmest quarter (54.9%), precipitation seasonality (18.8%), precipitation of warmest quarter (12.2%), precipitation of driest month (8.2%), isothermality (3.2%), and temperature seasonality (2.8%). The AUC value for this model was 0.985.

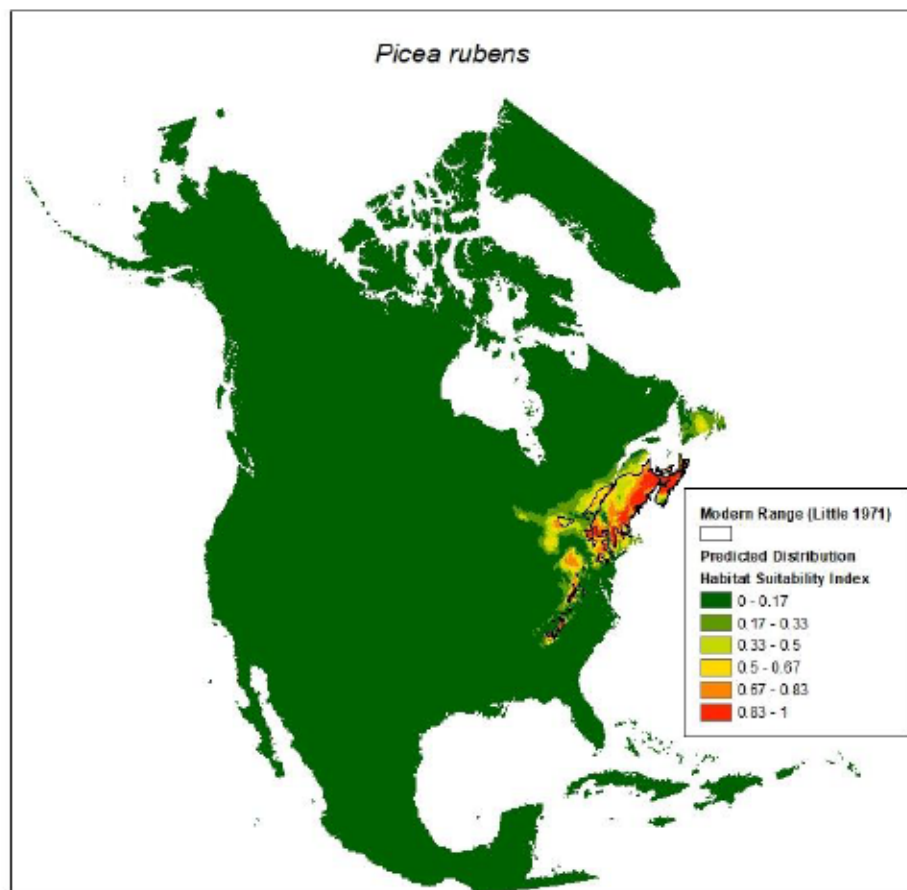


Figure 5. Predicted current distribution of *Picea rubens*.

Pinus banksiana

The SDM constructed for *Pinus banksiana* was able to predict current distribution along the northern, eastern, and western range edges according to Little's (1971) range map (Figure 6). It overextended predictions along the southern range edge and predicted clusters of low habitat suitability west of the range. Six uncorrelated bioclimatic variables were retained in the final model according to their permutation importance: maximum temperature of warmest month (57.9%), precipitation of wettest quarter (16.4%), temperature seasonality (11.1%), isothermality (6.0%), precipitation seasonality (5.3%), and temperature seasonality (2.8%). The AUC value for this model was 0.923.

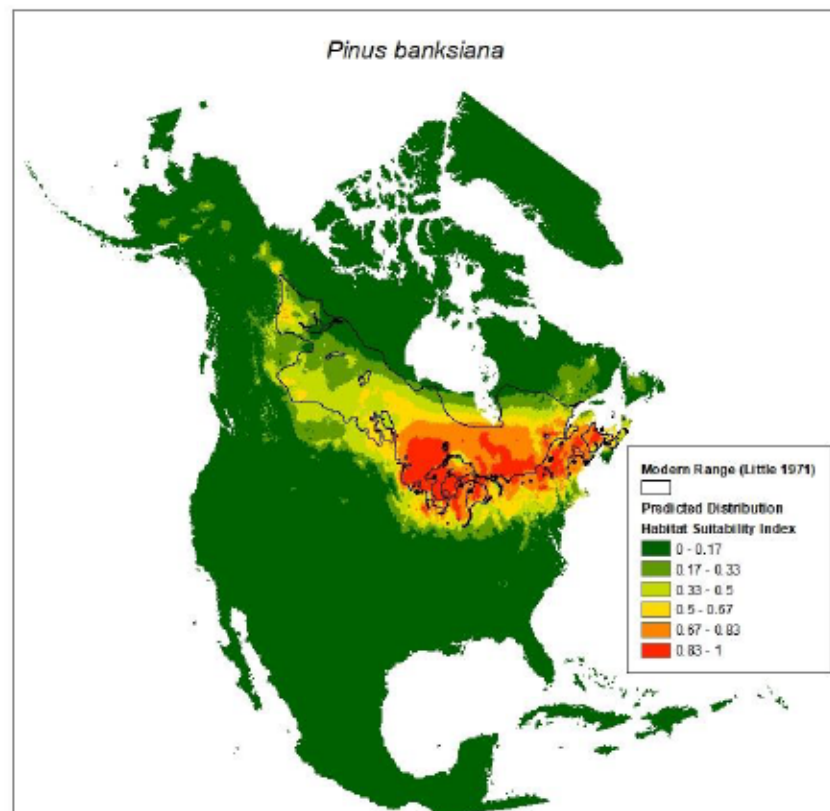


Figure 6. Predicted current distribution of *Pinus banksiana*.

Pinus resinosa

The SDM constructed for *Pinus resinosa* was able to predict current distribution along the northern, eastern, and western range edges well according to Little's (1971) range map (Figure 7). It overextended predictions of the southern range edge slightly, and predicted suitable habitat following the Appalachian Mountains south. southern range edge and predicted clusters of low habitat suitability west of the range. Six uncorrelated bioclimatic variables were retained in the final model according to their permutation importance: mean temperature of warmest quarter (63.5%), precipitation of warmest quarter (25.2%), temperature seasonality (7.1%), precipitation seasonality (2.1%), precipitation of driest quarter (1.8%), and isothermality (0.3%). The AUC value for this model was 0.965.

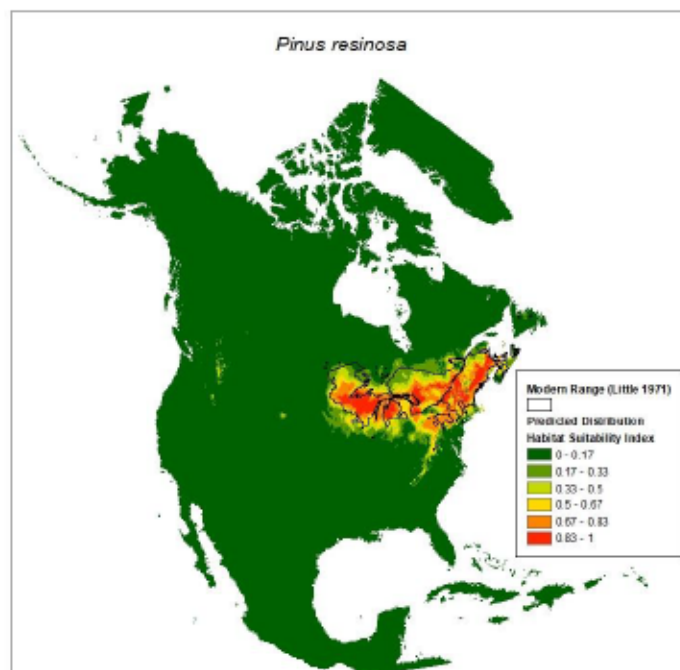


Figure 7. Predicted current distribution of *Pinus resinosa*.

Pinus strobus

The SDM constructed for *Pinus strobus* was able to predict current distribution along the northern, eastern, and southeastern range edges according to Little's (1971) range map (Figure 8). It overextended predictions along the southwestern range edge and predicted clusters of low habitat suitability far west of the range. Five uncorrelated bioclimatic variables were retained in the final model according to their permutation importance: maximum temperature of warmest month (48.7%), annual precipitation (42.2%), temperature seasonality (7.0%), precipitation seasonality (1.6%), and isothermality (0.6%). The AUC value for this model was 0.944.

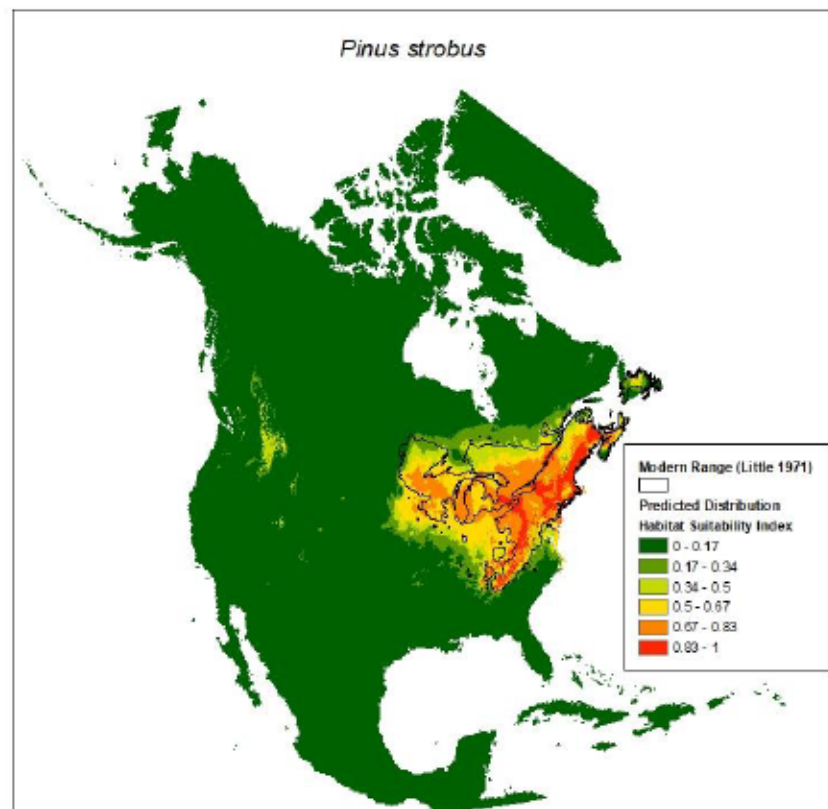


Figure 8. Predicted current distribution of *Pinus strobus*.

Thuja occidentalis

The SDM constructed for *Thuja occidentalis* was able to predict current distribution along the northern, eastern, and southeastern range edges according to Little's (1971) range map (Figure 9). It slightly overextended predictions along the southern range edge, oversimplified the disjointed range through the Appalachian Mountains, and predicted clusters of low habitat suitability far west of the range. Six uncorrelated bioclimatic variables were retained in the final model according to their permutation importance: maximum temperature of warmest month (50.1%), precipitation of wettest quarter (30.4%), precipitation seasonality (9.1%), temperature seasonality (8%), precipitation of driest quarter (2.3%), isothermality (0.1%). The AUC value for this model was 0.951.

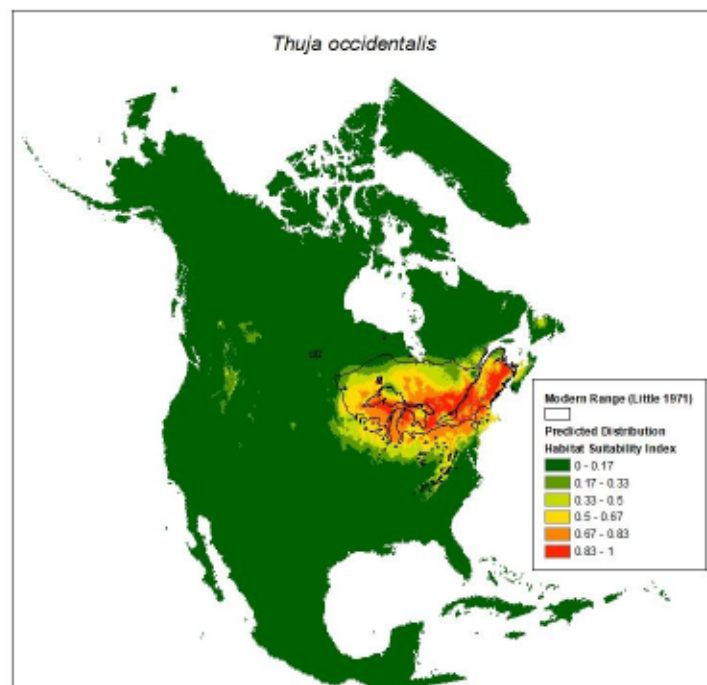


Figure 9. Predicted current distribution of *Thuja occidentalis*.

Tsuga canadensis

The SDM constructed for *Tsuga canadensis* was able to predict current distribution along the northern, eastern, and southeastern range edges according to Little's (1971) range map (Figure 10). It overextended predictions along the southwestern range edge and predicted clusters of low habitat suitability far west of the range. Five uncorrelated bioclimatic variables were retained in the final model according to their permutation importance: mean temperature of warmest quarter (56.3%), annual precipitation (24.7%), temperature seasonality (12.4%), precipitation seasonality (4.3%), and isothermality (2.3%). The AUC value for this model was 0.962.

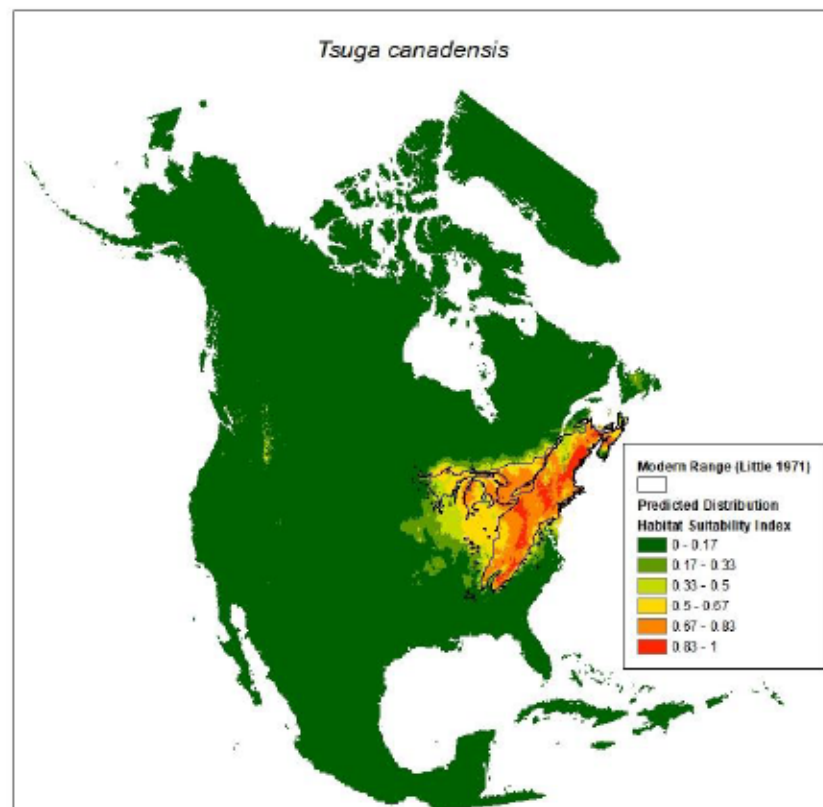


Figure 10. Predicted current distribution of *Tsuga canadensis*.

RECONSTRUCTION OF GLACIAL REFUGIA

As fossil record and phylogeographic evidence discussed within the literature review best support the predicted Pleistocene distributions according to the MPI-ESM-P GCM, they are the predictions presented here. Predicted Pleistocene distributions according to the CCSM4 and MIROC-ESM GCMs are available within the electronic supplementary material. Models predicted Pleistocene distribution for the selected species in four broad areas: Beringia including Alaska, the Grand Banks of Newfoundland, east of the Appalachian Mountains, and west of the Appalachian Mountains (Figure 11 to Figure 20). Some SDMs predicted distribution that was not restricted by the Appalachian Mountains, and one prediction distribution within the northwestern USA. The numbers and total area of predicted refugia found are included in Table 5. Locations and area of individual predicted refugia can be seen in Appendix III.

Table 5. Total area and number of predicted refugia for ten eastern North American conifer species.

Species	Total Area of Refugia (ha)	Number of Refugia
<i>Abies balsamea</i>	873902	3
<i>Larix laricina</i>	1615596	5
<i>Picea glauca</i>	2626508	11
<i>Picea mariana</i>	1595663	8
<i>Picea rubens</i>	804482	2
<i>Pinus banksiana</i>	1309879	3
<i>Pinus resinosa</i>	1011284	1
<i>Pinus strobus</i>	1948221	2
<i>Thuja occidentalis</i>	1111839	3
<i>Tsuga canadensis</i>	1585309	2



Figure 11. Predicted distribution of *Abies balsamea* during the LGM.



Figure 12. Predicted distribution of *Larix laricina* during the LGM.

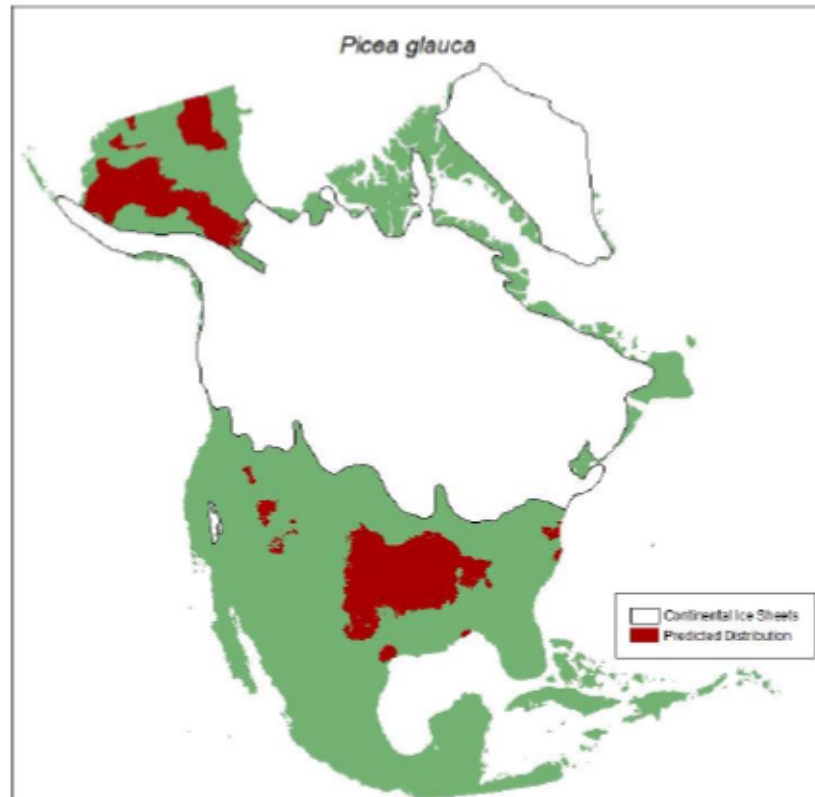


Figure 13. Predicted distribution of *Picea glauca* during the LGM.

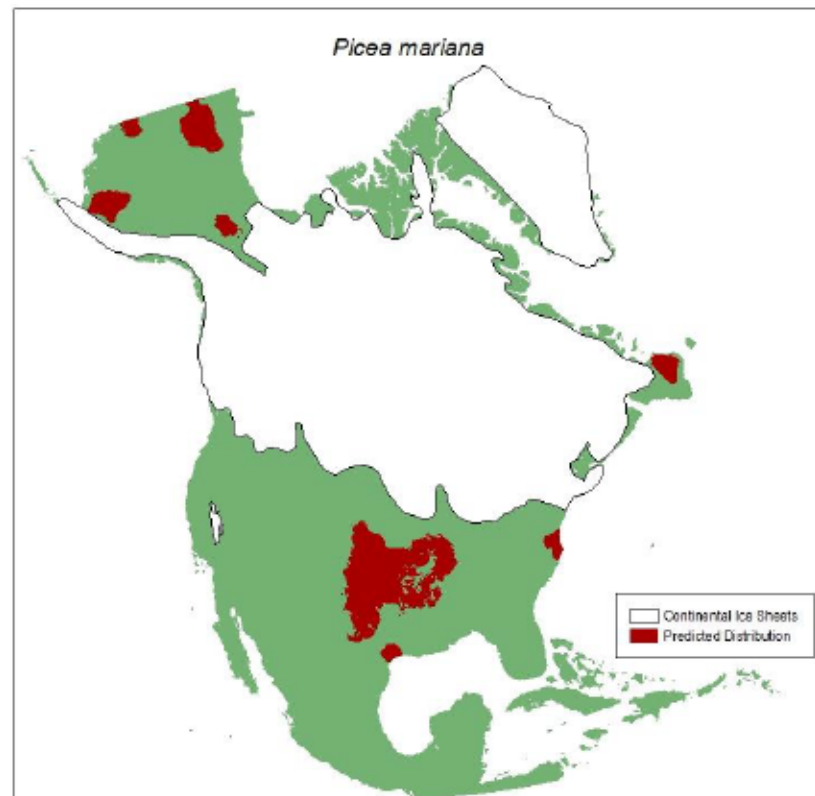


Figure 14. Predicted distribution of *Picea mariana* during the LGM.

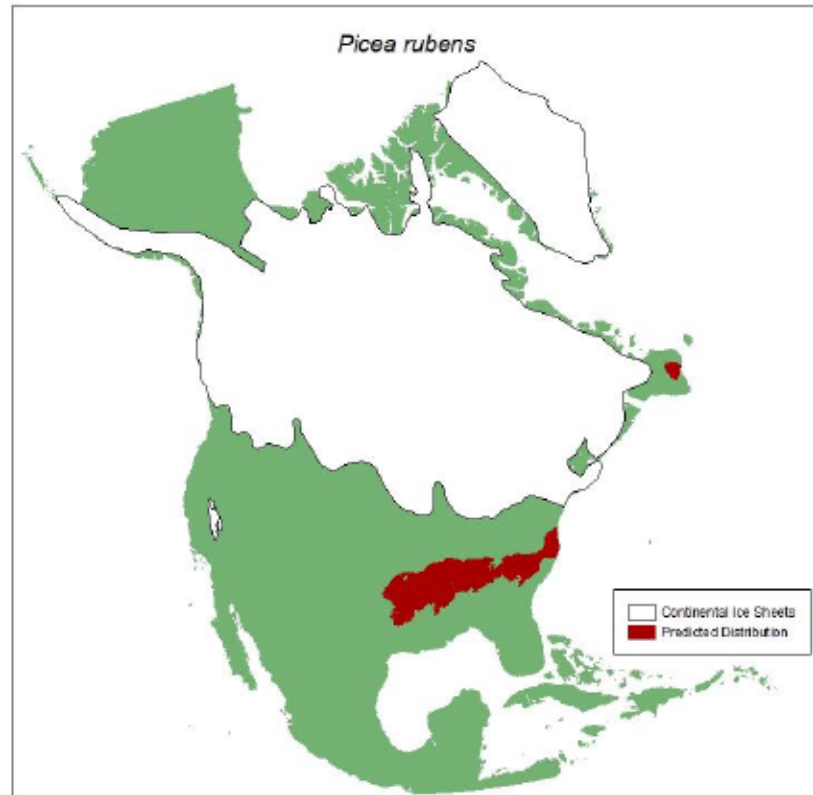


Figure 15. Predicted distribution of *Picea rubens* during the LGM.

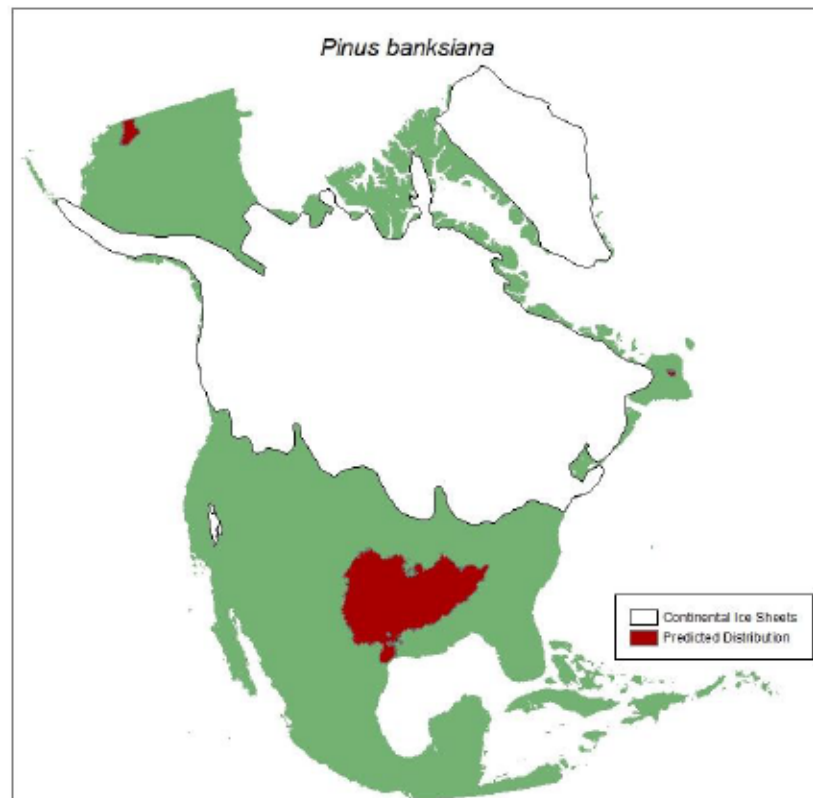


Figure 16. Predicted distribution of *Pinus banksiana* during the LGM.

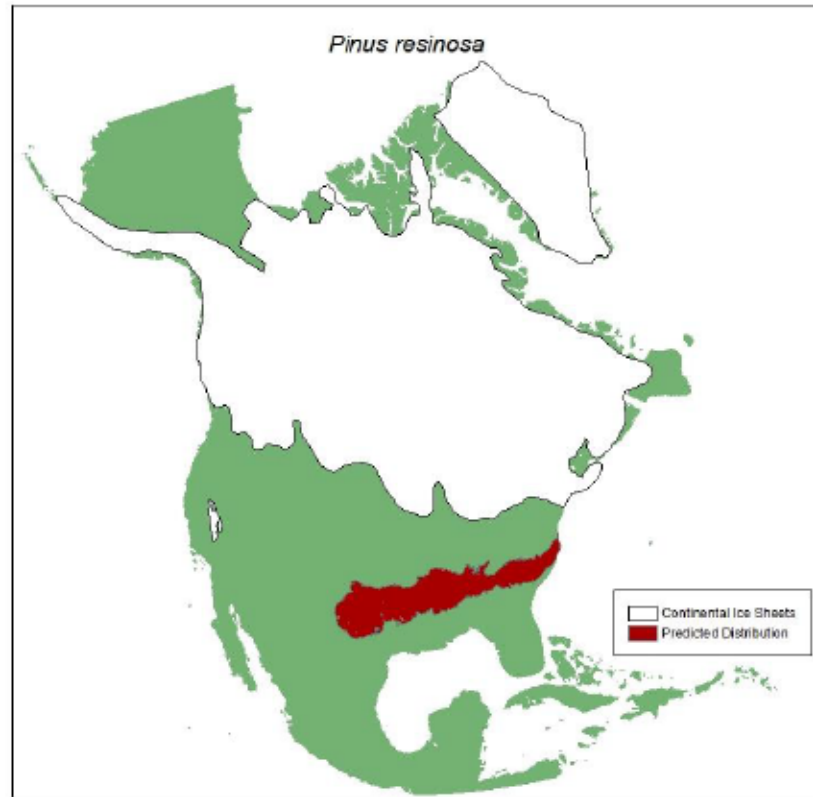


Figure 17. Predicted distribution of *Pinus resinosa* during the LGM.

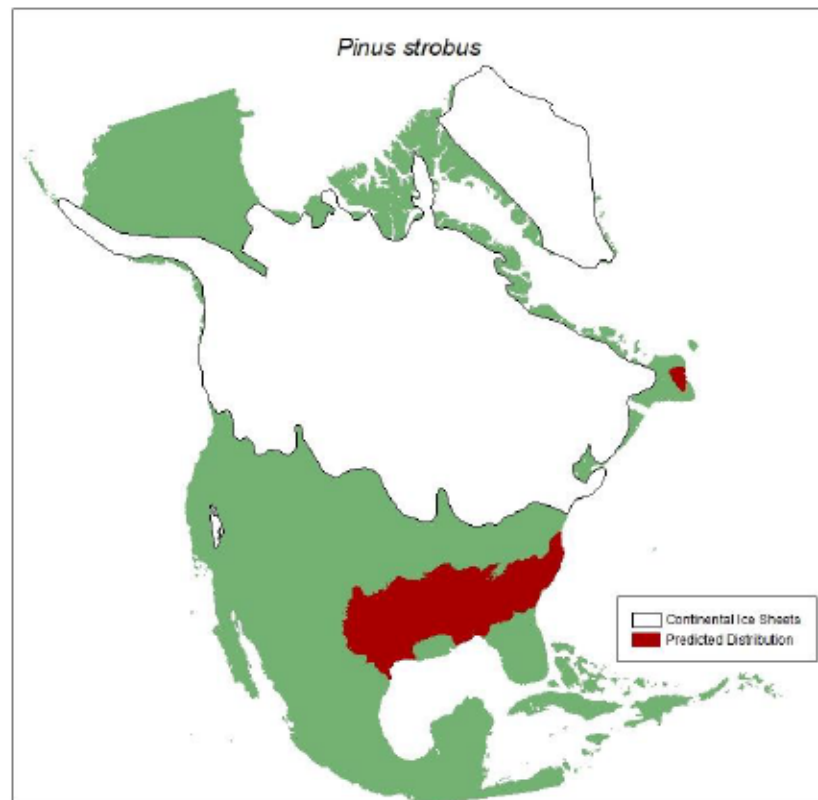


Figure 18. Predicted distribution of *Pinus strobus* during the LGM.

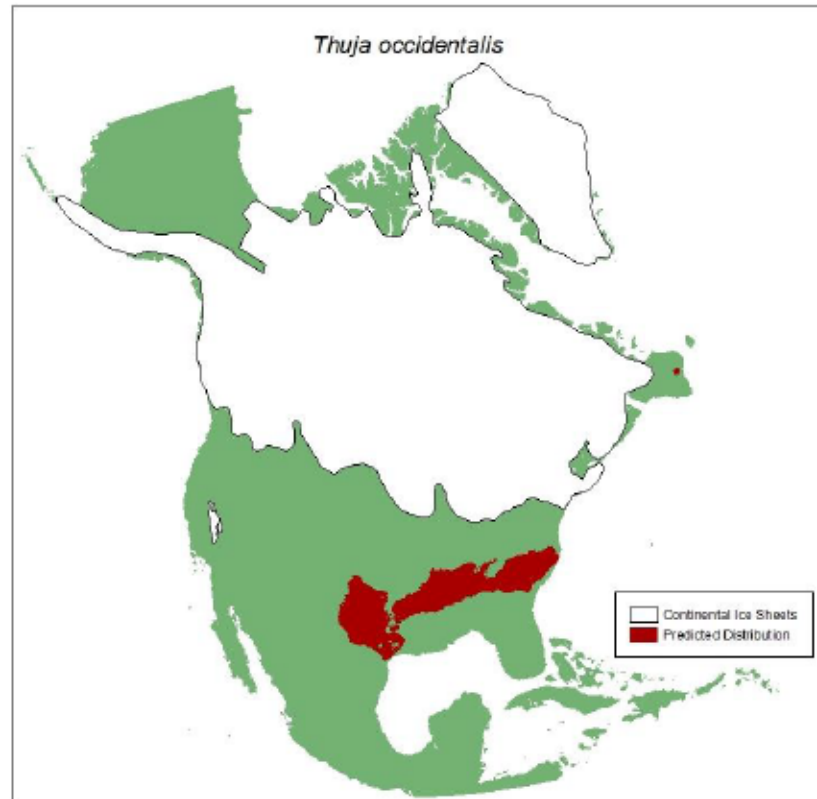


Figure 19. Predicted distribution of *Thuja occidentalis* during the LGM.

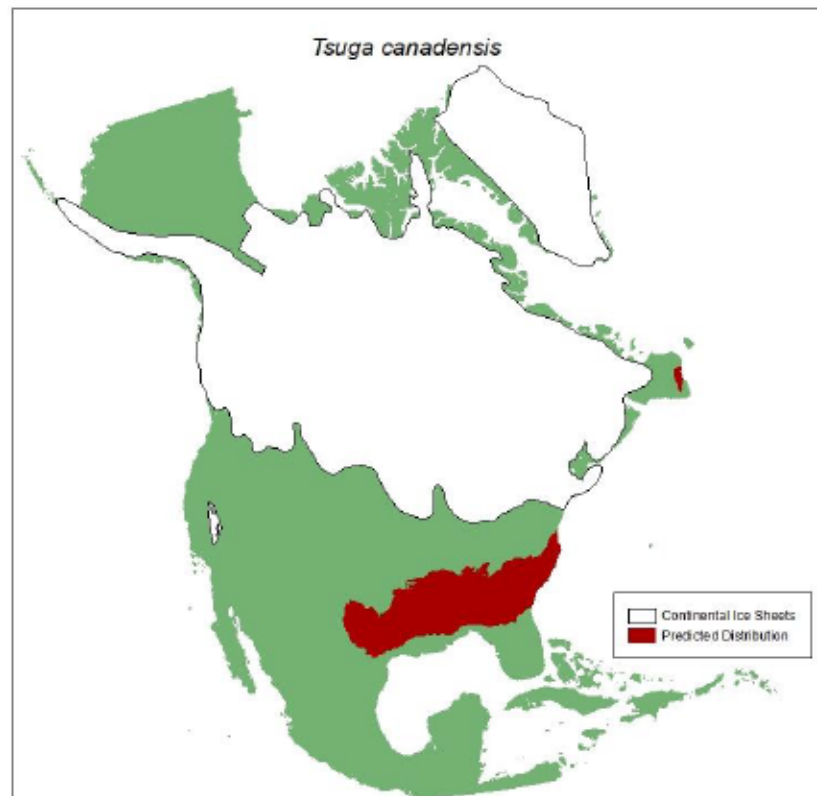


Figure 20. Predicted distribution of *Tsuga canadensis* during the LGM.

Genetic diversity measured by expected heterozygosity values (H_e) were retrieved from published range wide or central population allozyme analysis studies for the ten selected conifer species (Table 1). The average H_e among these species is 0.128. *Pinus resinosa* had the lowest H_e at 0.000, while the highest was *Picea glauca* with 0.290.

Table 6 contains comparisons of predicted current and LGM distribution of the selected species. The five species with transcontinental modern ranges had the lowest percentage of LGM predicted distribution compared to their current predicted distribution, between 24% and 37%. *Picea rubens* was the only species with a larger predicted distribution in the LGM compared to current at 112%. The highest percentage of LGM predicted distribution compared to current predicted distribution that does not represent a decrease in distribution was for *Tsuga canadensis* at 82%.

Table 6. LGM total predicted distribution compared to current predicted distribution for ten eastern North American conifer species.

Species	Total predicted distribution (ha)		LGM distribution as percent of current distribution
	LGM	Current	
<i>Abies balsamea</i>	873,902	2,825,188	31%
<i>Larix laricina</i>	1,615,596	6,018,690	27%
<i>Picea glauca</i>	2,626,508	7,183,019	37%
<i>Picea mariana</i>	1,595,663	6,584,443	24%
<i>Picea rubens</i>	804,482	717,420	112%
<i>Pinus banksiana</i>	1,309,879	4,156,778	32%
<i>Pinus resinosa</i>	1,011,284	1,625,606	62%
<i>Pinus strobus</i>	1,948,221	2,788,247	70%
<i>Thuja occidentalis</i>	1,111,839	2,414,024	46%
<i>Tsuga canadensis</i>	1,585,309	1,937,132	82%

Simple and multiple linear regressions compared number and total area of refugia as independent variables to He as the dependent variable. The analysis for the predicted Pleistocene distributions according to the MPI-ESM-P GCM is presented here (Table 7), while the analysis for predictions according to the two other GCMs used are included in the electronic supplementary material.

Table 7. Simple and multiple linear regression results comparing number and total area of glacial refugia to expected heterozygosity.

Statistic	Total Area of Refugia	Number of Refugia	Both Dependent Variables
r^2	0.63	0.70	0.78
Adjusted r^2	0.58	0.67	0.71
Standard Error	0.07	0.06	0.05
Significance F	0.01	0.00	0.01

According to the calculated adjusted r^2 , 71% of the variance between He values can be explained by both total area and number of refugia for a given species. Of the two independent variables, number of refugia has a higher adjusted r^2 than total area of refugia at 0.67 versus 0.58. The standard error values for these regression analyses are all quite low, ranging from 0.05 to 0.07. The significance F statistic is also adequately low, ranging from 0.00 to 0.01.

DISCUSSION

PLEISTOCENE SPECIES DISTRIBUTION

The placement of glacial refugia predicted by the SDMs were generally well supported by prior paleoecological and phylogeographic studies for their respective species. Multiple species have phylogeographic studies that suggest a northeastern genetic lineage resulting from a glacial refugia somewhere near Newfoundland and Labrador. The area that many of the models predicted refugia in that fits this suggestion is the Grand Banks, the exposed continental shelf south of Newfoundland. This area is not supported as glacial refugium by fossil record, but this could be due to the ill suitedness of exposed continental shelf to fossil preservation (Gavin *et al.* 2014).

Cinget *et al.* (2015) stated that phylogeographic structuring of *Abies balsamea* suggests that the modern distribution was formed from five genetic lineages. The suggested refugium northeast of the glacial extent is supported well by the predicted distribution on the Grand Banks. The four remaining genetic lines suggested could have resulted from the two predicted refugial

areas south of the glacial extent. The predicted refugium west of Appalachia could have split into two recolonization routes which would explain the suggestion of genetic lineages south and west of the Great Lakes. Similarly, the refugium predicted east of Appalachia could have split into the suggested southern Appalachian and Atlantic coast genetic lineages. Fossil records have placed *Abies balsamea* between the southern edge of the Laurentide glacier and Florida (Delcourt and Delcourt 1987; Jackson *et al.* 1997).

The predicted glacial refugia of *Larix laricina* south of the glacial extent are more disjointed than suggested by Warren *et al.* (2016). The suggestion that the Appalachian Mountains did not restrict gene flow between populations on either side of them could be due to a connecting population not predicted by the SDM, or because of long distance pollen dispersal effectively connecting the populations. Either way, the constructed SDM is likely missing a component of glacial refugia in this area. The model also predicted a refugium on the Grand Banks which is not present in either fossil records or phylogeographic studies. The prediction of *Larix laricina* having refugia within Beringia is well supported by the phylogeographic evidence presented by Napier *et al.* (2020) and fossil record presented by Brubaker *et al.* (2005).

The model for *Picea glauca* predicted a higher number of refugia than supporting studies. Within Beringia, the dispersion of refugia could support Anderson *et al.*'s (2006) suggestion that this species survived within this region in low densities. This placement of *Picea glauca* is well supported in the fossil record according to Brubaker *et al.* (2005), especially within eastern Beringia

due to evidence of early colonization of this species following glacial retreat in Northern British Columbia. The predictions south of the glacial extent are less supported, especially the cluster within western USA. There is a large predicted refugia south of the Great Lakes which is supported by the fossil record presented by Jackson *et al.* (1997), as well as two small refugia on the Gulf Coast which could have been assumed to be the same refugia in the fossil record due to proximity. Lafontaine *et al.* (2010) proposed a refugium east of the Appalachian Mountains which could be the two small refugia predicted on the Atlantic Coast.

The distinct phylogeographic distribution found by Jaramillo-Correa *et al.* (2004) in *Picea mariana* suggested three southern genetic lineages and one northeastern. The suggested northeastern lineage could be explained by the predicted refugia on the Grand Bank, which would have easily recolonized the areas found hosting the unique mitotypes in Labrador and Eastern Quebec. The predicted refugia within Beringia are well supported by the fossil record presented by Brubaker *et al.* (2005), but not supported by Jaramillo-Correa *et al.* (2004), and vice versa for the inferred refugia in Washington or Oregon states. It could be that the phylogeographic pattern seen in this region is in fact a result of recolonization from the Beringian refugia. The two remaining southern genetic lineages are said to be on either side of the Appalachian Mountains, which reflects the predicted Atlantic Coast refugia, and the Gulf Coast and west of Appalachia refugium.

The predicted distribution of *Picea rubens* is closely supported by the phylogeographic evidence presented by Hawley and DeHayes (1994) for both the Appalachia/Atlantic Coast and Grand Banks refugia. No fossil evidence was found for *Picea rubens* specifically between the Laurentide ice sheet and the Gulf Coast, but pollen records confirm that *Picea* spp. was indeed present (Jackson *et al.* 1997).

Considering the genetically distinct population of *Pinus banksiana* within the Maritimes region of Canada according to Godbout *et al.* (2010), the predicted refugium within the Grand Banks is well supported. The refugia suggested to be west of the Appalachian Mountains is also present within the predicted distribution, but it did not predict any refugia on the Atlantic coast as suggested. The placement of this species throughout the eastern USA is supported according to the fossil record presented by Jackson *et al.* (1997). The species distribution predicted in Beringia is not supported by either phylogeography or macro fossil records, but pollen records have been published that place the *Pinus* genus there (Delcourt and Delcourt 1987).

The predicted distribution of *Pinus resinosa* was restricted to only one refugia spanning from the Atlantic Coast to north of the Gulf Coast. Walter and Epperson (2005) suggested that according to their phylogeographic findings, this species should have been within two refugia, with one in northeast USA and the other in the southeast. It is possible that the Appalachian Mountains restricted gene flow more than the predicted distribution seems to suggest,

which would have formed these two refugia. *Pinus resinosa* is accepted to have been within this area according to fossil records (Jackson *et al.* 1997).

Zinck and Rajora (2016) suggested a single refugium of *Pinus strobus* in central North America which split into two recolonization routes and three genetic lineages. The large central refugia is supported by the predicted distribution. *Pinus strobus* is accepted to have been within this area according to fossil records (Jackson *et al.* 1997). The predicted refugia on the Grand Banks could possibly have been the source of the third genetic lineage and could have remained spatially undetected due to large amounts of admixture.

Due to the uncertainty of the fossil record of *Thuja occidentalis* during the LGM (Warner 1981; Yu 1997), and the lack of phylogeographic study, support of the predicted distribution is uncertain. The best estimate according to the fossil record of this species following glacial retreat places a refugia somewhere south of Lake Superior. Though this definition is broad, the two predicted refugia south of the glacial extent could fit it. As for the predicted refugia on the Grand Banks, it is unlikely any fossils would have survived here so it is hard to confidently support or reject it.

The phylogeographic findings of Potter *et al.* (2012) regarding refugia placement of *Tsuga canadensis* is well reflected in the predicted LGM distribution for both the southeast USA refugia and the Grand Banks Refugia. The evidence that the southern refugia was split into three clusters is not reflected in the predicted distribution, however this evidence was inconclusive.

The actual distribution could also have been disjointed due to unrealized variables. Macrofossils of *Tsuga canadensis* have been found within the general area north of the Gulf Coast (Jackson *et al.* 1997).

In general, the predictions of refugia within Beringia and the Grand Banks add to the body of work that explains a phenomenon called “Reid’s Paradox”. This paradox suggests that it would be impossible for temperate tree species to have experienced the high rates of post glacial migration temperate necessary in order to recolonize their current distributions from southern refugia within the relatively short time frame since glacial retreat (Clark *et al.* 1998). These refugia are closer to the current ranges of many temperate trees, which would allow them to recolonize them following glacial retreat before the time lag expected by “Reid’s Paradox”.

The performance of the SDMs in this thesis could be interpreted that modelling software such as MaxEnt is able to adequately predict distribution dynamics at the continental level to a sufficient degree to capture general trends, but at the regional level they do not reflect fossil records well enough. Similar to what Roberts and Hammans (2015) found, this is could be due to the species in question simply not having time to reach the extent of their available niche or be stopped from doing so by factors not included in the input variables. The overprediction of current southern range edges could be because distribution here is determined by soil type and topography before bioclimatic variables (Bemmels and Dick 2018). This shortcoming could have been reflected in predicted LGM distributions as well.

DETERMINANTS OF GENETIC DIVERSITY

Considering the generally accepted effects of founder's effect, genetic bottlenecks, and genetic drift on genetic diversity of a species (Jaramillo et al. 2009), the hypothesis that multiple and large refugia during the LGM would have led to higher genetic diversity today is a sound one. Supporting evidence suggested fewer refugia for species with the highest genetic diversity values than predicted by the SDMs. If this hypothesis is supported by multiple correlative studies, one could expect and search for additional refugial areas for species with high genetic diversity. A strong relationship was indeed found between modern genetic diversity and number and area of refugia in this thesis which agrees with a prior study by Roberts and Hamman (2015) in western North America.

Additional studies with larger sample sizes would be beneficial to better understand the extent of this correlation, but efforts are hard to organize due to the many sources needed for such a study. Model predictions could also be improved by the addition of variables not present in bioclimatic data such as soil type and topography. These predictions also do not consider successional dynamics of trees and forests during this time. These variables should be collected for multiple species and compared to genetic diversity to see how they have contributed.

CONCLUSION

Constructed SDMs were able to predict current distribution that closely resembled published range maps of the ten selected conifer species. The distributions each species' model predicted for the LGM were generally supported by fossil record and phylogeographic studies. This suggests that modelling software such as MaxEnt can capture continental wide distribution dynamics for conifer species. They seem to not accurately capture finer scale distribution, at least at the scale that was modelled.

The inferred refugia for multiple conifer species within Beringia and the Grand Banks add to the body of work attempting to explain "Reid's Paradox," which suggests that the rates at which temperate tree species would have had to expand to their current ranges following glacial retreat from southern refugia alone would have been impossibly fast (Clark *et al.* 1998). Refugia on the east and west side of the glacial extent would have been closer to modern day ranges of the trees in this thesis, which would reduce the expansion rates needed considerably.

Results indicated a significant positive correlation between expected heterozygosity as the dependant variable and number and size of refugia as the

independent variables, supporting the hypothesis that softwood tree species within multiple and/or large refugia during the LGM have higher genetic diversity in modern day populations than those with fewer and/or smaller refugia. Number of refugia had a higher positive correlation than size of refugia which supports the hypothesis that species with multiple smaller refugia have experienced less of a genetic bottleneck than ones with a single or few large refugia. More species should be analyzed in the same fashion as this thesis in order to see if this pattern holds true for trees all over North America and the world. Lastly, model predictions could be improved given a more robust set of variables including soil types and topography.

LITERATURE CITED

- Anderson, L., F. Hu, D. Nelson, R. Petit, and K. Paige. 2006. Ice-age endurance: DNA evidence of a white spruce refugium in Alaska. *Proceedings of the National Academy of Sciences of the United States of America* 103(33):12447-12450.
- Araújo, M. and A. Guisan. 2006. Five (Or so) Challenges for Species Distribution Modelling. *Journal of Biogeography* 33(10):1677-1688.
- Barbet-Massin, M., Q. Rome, C. Villemant, and F. Courchamp. 2018. Can species distribution models really predict the expansion of invasive species? *PLoS One* 13(3):e0193085.
- Beaulieu, J. and J. Simon. 1994. Genetic structure of *Pinus strobus* in Quebec. *Canadian Journal of Forestry Research* 24:1726-1733.
- Bemmels, J. and C. Dick. 2018. Genomic evidence of a widespread southern distribution during the Last Glacial Maximum for two eastern North American hickory species. *Journal of Biogeography* 45:1739-1750.
- BISON. 2019. Biodiversity Information Serving Our Nation Database. www.bison.usgs.gov. Oct. 10th 2019.
- Brubaker, L., P. Anderson, Mary Edwards, and A. Lozhkin. 2005. Beringia as a glacial refugium for boreal trees and shrubs: new perspectives from mapped pollen data. *Journal of Biogeography* 32:833-848.
- Burns, R. and B. Honkala. 1990. *Silvics of North America: 1. Conifers*. Agriculture Handbook 654, US Department of Agriculture, Forest Service, Washington DC. 877 pp.
- Bush, R. and P. Smouse. 1992. Evidence for the adaptive significance of allozymes in forest trees. *New Forests* 6(1):179-196.
- Cheliak, W., J. Wang, and J. Patel. 1988. Population structure and genic diversity in tamarack, *Larix laricina* (Du Roi) K. Koch. *Canadian Journal of Forestry Research* 18:1318-1324.

- Cinget, B., S. Gérardi, J. Beaulieu, and J. Bousquet. 2015. PLOS One. 10(4):e0122815.
- Clark, J., C. Fastie, G. Hurtt, S. Jackson, C. Johnson, G. King, M. Lewis, J. Lynch, S. Pacala, C. Prentice, E. Schupp, T. Webb, and P. Wyckoff. 1998. *BioScience* 48(1):13-24.
- Delcourt, P. and H. Delcourt. 1987. Tree Population Dynamics During the Past 20,000 Years: 107-291 *in* Delcourt, P. and H. Delcourt. 1987. Long-Term Forest Dynamics of the Temperate Zone. Ecological Studies. 63. Springer-Verlag, New York. 439 pp.
- Dobrowski, S. 2011. A climate basis for microrefugia: the influence of terrain on climate. *Global Change Biology* 17:1022-1035.
- Dormann, C., S. Schymanski, J. Cabral, I. Chuine, C. Graham, F. Hartig, M. Kearney, X. Morin, C. Römermann, B. Schroöder, and A. Singer. 2012. Correlation and process in species distribution models: bridging a dichotomy. *Journal of Biogeography* 39(12):2119-2131.
- Elith, J. and J. Leathwick. 2009. Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. *Annual Review of Ecology Evolution and Systematics* 40(1):677-697.
- Elith, J., S. Phillips, T. Hastie, M. Dudík, Y. Chee, and C. Yates. 2010. A statistical explanation of MaxEnt for ecologist. *Diversity and Distribution* 17(1):43-57.
- Evans, M., C. Merow, S. Record, S. McMahon, and B. Enquist. 2016. Towards Process-based Range Modeling of Many Species. *Trends in Ecology & Evolution* 31(11):860-871.
- Fowler, D. and R. Morris. 1977. Genetic diversity in red pine: evidence for low genic heterozygosity. *Canadian Journal of Forestry Research* 7:343-347.
- Furnier, G., M. Stine, C. Mohn, and M. Clyde. 1990. Geographic patterns of variation in allozymes and height growth in white spruce. *Canadian Journal of Forestry Research* 21:707-712.
- Gavin, D., M. Fitzpatrick, P. Gugger, K. Heath, F. Rodríguez-Sánchez, S. Dobrowski, A. Hampe, F. Hu, M. Ashcroft, P. Bartlein, J. Blois, B. Carstens, E. Davis, G. de Lafontaine, M. Edwards, M. Fernandez, P. Henne, E. Herring, Z. Holden¹, W. Kong, J. Liu, D. Magri, N. Matzke, M. McGlone, F. Saltre, A. Stigall, Y. Tsai, and J. Williams. 2014. Climate refugia: Joint inference from fossil records, species distribution models and phylogeography. *New Phytologist* 204(1):37-54.

- GBIF. 2019. GBIF Biodiversity Data. www.gbif.org/occurrence/search. Sept. 28th, 2019.
- Godbout, J., J. Beaulieu, and J. Bousquet. 2010. Phylogeographic structure of jack pine (*Pinus banksiana*; Pinaceae) supports to existence of a coastal glacial refugium in northeastern North America. *American Journal of Botany* 97(11):1903-1912.
- Godt, M., J. Hamrick, M. Edwards-Burke, and J. Williams. 2001. Comparisons of genetic diversity in white spruce (*Picea glauca*) and jack pine (*Pinus banksiana*) seed orchards with natural populations. *Canadian Journal of Forestry Research* 31:943-949.
- Guissan, A., C. Graham, J. Elith, F. Huettmann, and NCEAS Species Distribution Modelling Group. 2007. Sensitivity of predictive species distribution models to change in grain size. *Diversity and Distribution* 13(3):332-340.
- Hamrick, J., M. Godt, and S. Sherman-Broyles. 1992. Factors influencing levels of genetic diversity in woody plant species. *New Forests* 6: 95-124.
- Hawley, G. and D. DeHayes. 1994. Genetic diversity and population structure of red spruce (*Picea rubens*). *Canadian Journal of Botany* 72:1778-1786.
- Hewitt, G. 2000. The genetic legacy of the Quarternary ice ages. *Nature* 405:907-913.
- Higgins, S., D. Richardson, and R. Cowling. 2000. Using dynamic landscape model for planning the management of alien plant invasions. *Ecological Applications* 10(6): 1833-1848.
- Hijmans, R.J., S.E. Cameron, J.L. Parra, P.G. Jones and A. Jarvis. 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* 25:1965-1978
- Jackson, S., R. Webb, K. Anderson, J. Overpeck, T. Webb III, J. Williams, and B. Hansen. 2000. Vegetation in Eastern North America during the Last Glacial Maximum. *Quaternary Science Reviews* 19:489-508.
- Jaramillo-Correa, J., J. Bealieu, D. Khasa, and J. Bousquet. 2009. Inferring the past from the present phylogeographic structure of North American forest trees: seeing the forest for the genes. *Canadian Journal of Forest Research* 39:286-307.
- Jaramillo-Correa, J., J. Bealieu, and J. Bousquet. 2004. Variation in mitochondrial DNA reveals multiple distant glacial refugia in black spruce (*Picea mariana*), a transcontinental North American conifer. *Molecular Ecology* 13:2735-2747.

- Jamevich, C., T. Stohlgren, S. Kumar, J. Morissette, and T. Holcombe. 2015. Caveats for correlative species distribution modeling. *Ecological Informatics* 29(1):6-15.
- Lischke, H., N. Zimmerman, J. Bolliger, S. Rickbusch, and T. Löffler. 2006. TreeMig: A forest-landscape model for simulating spatio-temporal patterns from stand to landscape scale. *Ecological Modelling* 199(4):409-420.
- Little, E.L. 1971. Atlas of United States trees, volume 1, conifers and important hardwoods. US Department of Agriculture, Washington, DC. 200 maps.
- Morin, X., and W. Thuiller. 2009. Comparing niche- and process-based models to reduce prediction uncertainty range shifts under climate change. *Ecology* 90(5):1301-1313.
- Morin, X., D. Vinier and I. Chuine. 2008 Tree species range shifts at a continental scale: new predictive insights from a process-based model. *Journal of Ecology* 96(4):784-794.
- Napier, J., M. Fernandez, G. de Lafontaine, and F. Hu. 2020. Ice-age persistence and genetic isolation of the disjunct distribution of larch in Alaska. *Ecology and Evolution* 10(3):1692-1702.
- O'Connel, L., A. Mosseler, and O. Rajora. 2007. Extensive long-distance pollen dispersal in a fragmented landscape maintains genetic diversity in white spruce. *Journal of Heredity* 98(7):640-645.
- Ouborg, N., Y. Piquot, and J. Van Groenendael. 2008. Population genetics, molecular markers and the study of dispersal in plants. *Journal of Ecology* 87(4):551-568.
- Perron, M., D. Perry, C. Andalo, and J. Bousquet. 2000. Evidence from sequence-tagged site markers of a recent progenitor-derivative species pair in conifers. *Proceedings of the National Academy of Sciences of the United States of America* 97(21):11331-11336.
- Perry, D. and J. Bousquet. 2000. Genetic diversity and mating system of post-fire and post-harvest black spruce: an investigation using codominant sequence-tagged-site markers. *Canadian Journal of Forestry Research* 31:32-40.
- Perry, D., P. Knowles and F. Yeh. 1990. Allozyme Variation of *Thuja occidentalis* L. in Northwestern Ontario. *Biochemical Systematics and Ecology* 18(2):111-115.

- Phillips, S., R. Anderson, M. Dudík, R. Schapire, and M. Blair. 2017. Opening the Blackbox: an open source release of MaxEnt. *Ecography* 40(7):887-893.
- Phillips, S. and M. Dudík. 2008. Modeling of species distributions with MaxEnt: new extensions and a comprehensive evaluation. *Ecography* 31:161-175.
- Phillips, S., M. Dudík, R. Schapire. MaxEnt software for modeling species niches and distributions (Version 3.4.1). Biodiversity Informatics. http://biodiversityinformatics.amnh.org/open_source/MaxEnt/. Oct. 18th 2019.
- Potter, K., R. Jetton, W. Dvorak, V. Hipkins, R. Rhea, and W. Andrew. Widespread inbreeding and unexpected geographic patterns of genetic variation in eastern hemlock (*Tsuga canadensis*), an imperiled North American conifer. 2012. *Conservation Genetics* 13:475-498.
- Ray, N. and J.M. Adams. 2001. A GIS-based vegetation map of the world at the last glacial maximum (25,000-15,000 BP). *Internet Archaeology* 11.
- Roberts, D. and A. Hamann. 2015. Glacial refugia and modern genetic diversity of 22 western North American tree species. *Proceedings of the Royal Society B*. 282.
- Rowe, K., E. Heske, P. Brown, and K. Paige. 2004. Surviving the ice: Northern refugia and postglacial colonization. *Proceedings of the National Academy of Sciences of the United States of America* 101(28):10355-10359
- Roy-Dufresne, E., F. Saltré, B. Cooke, C. Mellin, G. Mutze, T. Cox, and D. Fordham. 2019. Modelling the distribution of a wide-ranging invasive species using the sampling efforts of expert and citizen scientists. *Ecology and Evolution* 9(19): 11053–11063.
- Schneider von Deimling, T., A. Ganopolski, H. Held, and S. Rahmstorf. 2006. How cold was the Last Glacial Maximum? *Geophysical Research Letters* 33(14).
- Schneiderman, J., H. He, F. Thompson III, W. Dijak, and J. Fraser. 2015. Comparison of a species distribution model and process model from a hierarchical perspective to quantify effects of projected climate change on tree species. *Landscape Ecology* 30:1879-1892.
- Shea, K. and G. Furnier. 2002. Genetic variation and population structure in central and isolated populations of balsam fir, *Abies balsamea* (Pinaceae). *American Journal of Botany* 89(5):783-791.

- Svenning, J.C., S. Normand, and M. Kageyama. 2008. Glacial refugia of temperate trees in Europe: insights from species distribution modelling. *Journal of Ecology* 96(6):1117-1127.
- Václavík, T. and R. Meentemeyer. 2012. Equilibrium or not? Modelling potential distribution of invasive species in different stages of invasion. *Diversity and Distributions* 18(1):73-83.
- Walter, R. and B. Epperson. 2005. Geographic patterns of genetic diversity in *Pinus resinosa*: contact zone between descendants of glacial refugia. *American Journal of Botany* 92(1):92-100
- Warner, B. 1981. Late Glacial fossil leaves of *Thuja occidentalis* from Manitoulin Island, Ontario. *Canadian Journal of Botany* 60:1352-1356.
- Warren, D.L. and S.N. Seifert. 2011. Ecological niche modeling in MaxEnt: the importance of model complexity and the performance of model selection criteria. *Ecological Applications* 21(2):335–342.
- Warren, D.L., R.E. Glor, and M. Turelli. 2019. ENMTools: a toolbox for comparative studies of environmental niche models. *Ecography* 33:607-611 (Version 1.4.4).
- Warren, E., G. de Lafontaine, S. Gérardi, S. Senneville, J. Beaulieu, M. Perron, J. Jaramillo-Correa, and J. Bousquet. 2016. Joint inference from cytoplasmic DNA and fossil data provide evidence for glacial vicariance and contrasted post-glacial dynamics in tamarack, a transcontinental conifer. *Journal of Biogeography* 43(6):1227-1241.
- Wisz, M., R. Hijmans, J. Li, A. Peterson, C. Graham, and A. Guisan. 2008. Effects of sample size on the performance of species distribution models. *Diversity and Distributions* 14(5):763-773.
- Yokoyama, Y., T. Esat, W. Thompson, A. Thomas, J. Webster, Y. Miyairi, C. Sawada, T. Aze, H. Matsuzaki, J. Okuno, S. Fallon, J.C. Braga, M. Humblet, Y. Iryu, D. Potts, K. Fujita, A. Suzuki, H. Kan. 2018. Rapid glaciation and a two-step sea level plunge into the Last Glacial Maximum. *Nature* 559(7715): 603-608.
- Yu, Z. 1997. Late Quaternary paleoecology of *Thuja* and *Juniperus* (Cupressaceae) at Crawford Lake, Ontario: pollen, stomata and macrofossils. *Review of Palaeobotany and Palynology* 96:241-254.
- Zabinski, C. 1992. Isozyme variation in eastern hemlock. *Canadian Journal of Forestry Research* 22:1838-1842.
- Zinck, J. and O. Rajora. 2016. Post glacial phylogeography and evolution of a wide-ranging highly-exploited keystone forest tree, eastern white pine

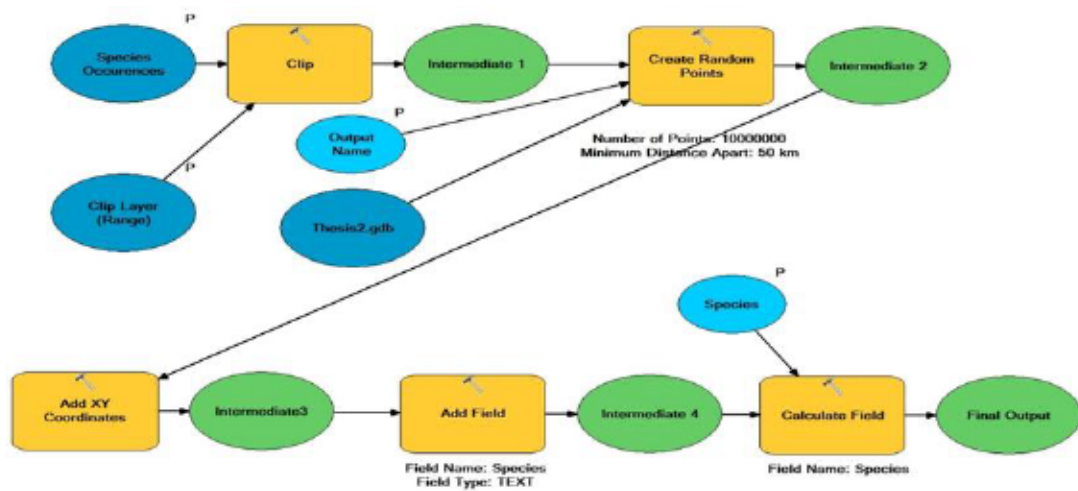
(Pinus strobus) in North America: single refugium, multiple routes. BMC Evolutionary Biology 16(56)

APPENDICES

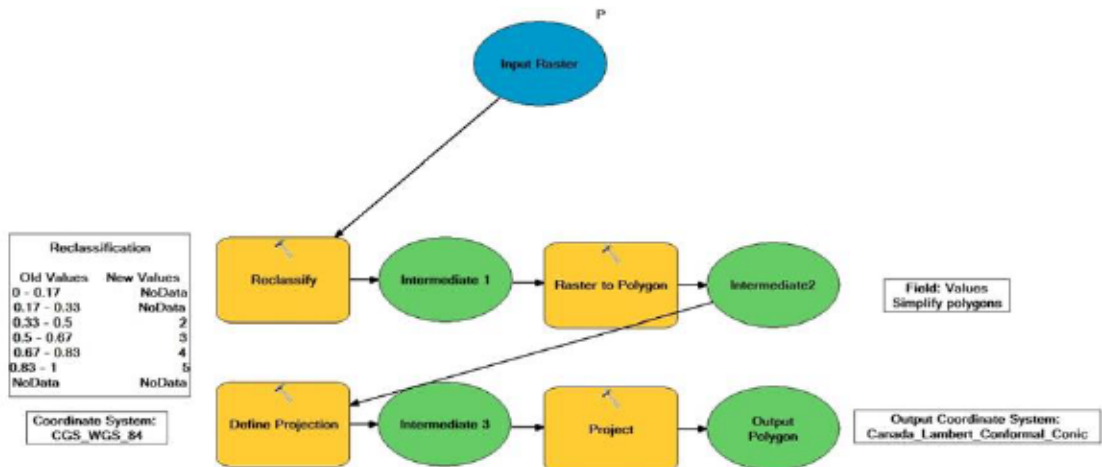
APPENDIX I

CREATED GEOPROCESSING TOOLS

A. Species occurrence points processing

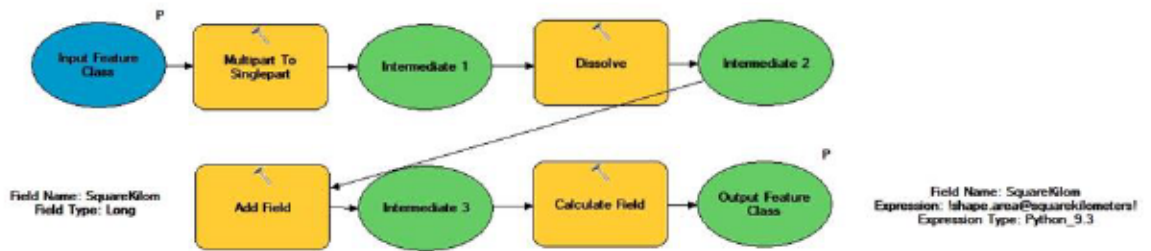


B. SDM prediction raster reclassification/conversion

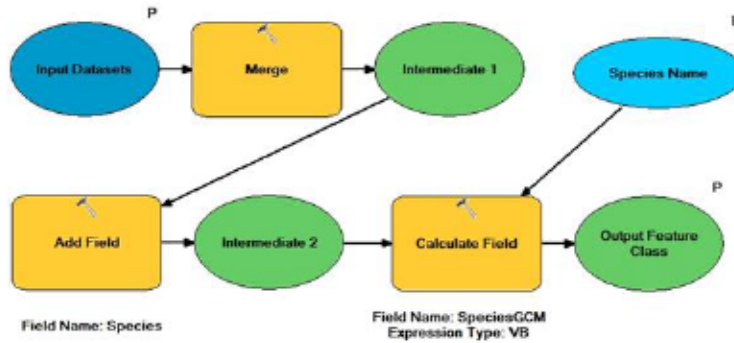


APPENDIX I (CONTINUED)

C. Refugia polygon dissolution



D. Refugia polygon merge



APPENDIX II

AREA UNDER RECEIVER CURVE VALUES AND PERMUTATION
IMPORTANCE OF BIOCLIMATIC VARIABLES

Modelled Species	<i>Abies balsamea</i>	<i>Larix laricina</i>	<i>Picea glauca</i>	<i>Picea mariana</i>	<i>Picea rubens</i>	<i>Pinus banksiana</i>	<i>Pinus resinosa</i>	<i>Pinus strobus</i>	<i>Thuja occidentalis</i>	<i>Tsuga canadensis</i>
AUC Value	0.940	0.891	0.869	0.876	0.985	0.923	0.965	0.944	0.951	0.962
Ranked Permutation Importance of Bioclimatic Variables	BIO 1	1	-	-	-	-	-	-	-	-
	BIO 2	7	-	-	-	-	-	-	-	-
	BIO 3	-	3	2	3	5	4	6	5	6
	BIO 4	4	5	3	5	6	3	3	3	4
	BIO 5	-	1	1	1	-	1	-	1	1
	BIO 6	-	-	-	-	-	-	-	-	-
	BIO 7	-	-	-	-	-	-	-	-	-
	BIO 8	5	-	-	-	-	-	-	-	-
	BIO 9	-	-	-	-	-	-	-	-	-
	BIO 10	-	2	-	-	1	-	1	-	-
	BIO 11	-	-	-	-	-	-	-	-	-
	BIO 12	-	-	-	-	-	-	-	2	-
	BIO 13	-	-	-	-	-	-	-	-	-
	BIO 14	6	-	-	-	4	-	-	-	-
	BIO 15	3	4	6	4	2	5	4	4	3
	BIO 16	-	-	-	-	-	2	-	-	2
	BIO 17	-	6	-	-	-	6	5	-	5
	BIO 18	2	-	5	2	3	-	2	-	-
	BIO 19	-	-	4	-	-	-	-	-	-
# of retained variables	7	6	6	5	6	6	6	5	6	5

APPENDIX III

PREDICTED REFUGIA LOCATION AND AREA IN SQUARE KILOMETRES OF
TEN EASTERN NORTH AMERICAN CONIFER SPECIES

Species	Refugia Location	Area (km ²)
<i>Abies balsamea</i>	Grand Banks	19955
<i>Abies balsamea</i>	East of Appalachia	198474
<i>Abies balsamea</i>	West of Appalachia	655473
<i>Larix laricina</i>	West of Appalachia	1297660
<i>Larix laricina</i>	East of Appalachia	171684
<i>Larix laricina</i>	Grand Banks	64010
<i>Larix laricina</i>	Southern Beringia	60852
<i>Larix laricina</i>	Western Beringia	21390
<i>Picea glauca</i>	Atlantic Coast	32541
<i>Picea glauca</i>	West of Appalachia	1345817
<i>Picea glauca</i>	Gulf Coast	6986
<i>Picea glauca</i>	Gulf Coast	30577
<i>Picea glauca</i>	Western USA	13160
<i>Picea glauca</i>	Western USA	26047
<i>Picea glauca</i>	Western USA	41703
<i>Picea glauca</i>	Western Beringia	12632
<i>Picea glauca</i>	Western Beringia	27842
<i>Picea glauca</i>	Western Beringia	235229
<i>Picea glauca</i>	Southern/Eastern Beringia	853974
<i>Picea mariana</i>	Atlantic Coast	49445
<i>Picea mariana</i>	West of Appalachia	1008901
<i>Picea mariana</i>	Gulf Coast	36536
<i>Picea mariana</i>	Grand Banks	77251
<i>Picea mariana</i>	Western Beringia	41294
<i>Picea mariana</i>	Western Beringia	206237
<i>Picea mariana</i>	Eastern Beringia	50234
<i>Picea mariana</i>	Southern Beringia	125765

APPENDIX III (CONTINUED)

Species	Refugia Location	Area (km ²)
<i>Picea rubens</i>	Through Appalachia/Atlantic Coast	774790
<i>Picea rubens</i>	Grand Banks	29692
<i>Pinus banksiana</i>	West of Appalachia	1262046
<i>Pinus banksiana</i>	Grand Banks	4783
<i>Pinus banksiana</i>	Western Beringia	43050
<i>Pinus resinosa</i>	Through Appalachia/Atlantic Coast	1011284
<i>Pinus strobus</i>	Gulf Coast/Through Appalachia/Atlantic Coast	1905292
<i>Pinus strobus</i>	Grand Banks	42929
<i>Thuja occidentalis</i>	Through Appalachia/Atlantic Coast	672800
<i>Thuja occidentalis</i>	Gulf Coast/Western USA	434341
<i>Thuja occidentalis</i>	Grand Banks	4698
<i>Tsuga canadensis</i>	Through Appalachia/Gulf Coast	1563076
<i>Tsuga canadensis</i>	Grand Banks	22233