

Title: Range shifts in butternut, a rare, endangered tree, in response to past climate and modern conditions

Authors

Emily Schumacher, eschumacher@mortonarb.org, the Morton Arboretum, 4100 IL-53, Lisle, IL 60532 (corresponding author)
0000-0003-3145-1181

Alissa Brown, abrown@mortonarb.org, the Morton Arboretum, 4100 IL-53, Lisle, IL 60532
0000-0002-1511-8496

Martin Williams, martin.williams@canada.ca, Natural Resources Canada, 1350 Regent street, Fredericton, NB, E3B 5P7
0000-0003-2637-6084

Jeanne Romero-Severson, Jeanne.Romero-Severson.1@nd.edu, Department of Biological Sciences, University of Notre Dame, South Bend, IN, USA
0000-0003-4112-7238

Tannis Beardmore, tbeardmo@gmail.com, Natural Resources Canada, 1350 Regent street, Fredericton, NB, E3B 5P7 (former NRCAN scientist)

Sean Hoban, shoban@mortonarb.org, the Morton Arboretum, 4100 IL-53, Lisle, IL 60532
0000-0002-0348-8449

Acknowledgements

We thank the following people for help with field collections or DNA samples: Kathleen Forbes, Charlene Williams, Christine Cusack, Joel Bates, Gorica Barudzija, Maureen Toner, Rodney MacEachern, Eric Sullivan, Marie-Claude Gros-Louis, Glenna Halicki Hayden, Laura Thompson, Scott Schlarbaum, Sunshine Brosi, Mark Coggeshall, Robert White, Jack Frank, Brice Leech, Ashton Berdine, Eagle Brosi, Jeanine Riggleman, Jennifer Murrow, Ami Sharp, Dale Bergdahl, Diane Burbank, Barbara Boysen, Judy Loo, Donnie McPhee, Robert Anderson, David Gallagher, Mary Ann Buenzow, Paul Berrang, Chris Casey, Henrietta Bennett, Carrie Sweeney, Dwight Slocum and Wayne Kruger. We are also indebted to the following undergraduate assistants for field and laboratory work: Britney Dennis, Bridgette Sanchez, Betsy Madison, Ellen Luecke, Lindsay Huebner, Paul Goedde and Casey Bouskill. This work was funded by the University of Notre Dame, University of Tennessee, Margaret Finley Shackelford Trust, Healing Stones Foundation, Natural Resources Canada, Environment and Climate Change Canada-Interdepartmental Recovery Fund and the Northern Nut Growers Association. SH was funded by a fellowship from the Arthur J. Schmitt Foundation. Appropriate permits were obtained from the US Forest Service, state natural resource agencies. Sampling on private land was with the permission and full knowledge of the landowners.

Abstract

Aim: Range shifts are a key process that determine species distributions and genetic patterns. A previous investigation reported that *Juglans cinerea* (butternut) has lower genetic diversity at higher latitudes, hypothesized to be the result of range shifts following the last glacial period. However, genetic patterns can also be impacted by modern ecogeographic conditions. Therefore, we re-investigate genetic patterns of butternut with additional northern population sampling, hindcasted species distribution models, and fossil pollen records to clarify the impact of glaciation on butternut.

Location: Eastern North America

Taxon: *Juglans cinerea* (L., Juglandaceae) (butternut)

Methods: Using 11 microsatellites, we examined range-wide spatial patterns of genetic diversity metrics (allelic richness, heterozygosity, F_{ST}) for previously studied butternut individuals and an additional 757 samples. We constructed hindcast species distribution models and mapped fossil pollen records to evaluate habitat suitability and evidence of species' presence throughout space and time.

Results: Contrary to previous work on butternut, we found that genetic diversity increased with distance to range edge, and previous latitudinal clines in diversity were likely due to a few outlier populations. Populations in New Brunswick, Canada were genetically distinct from other populations. At the Last Glacial Maximum, pollen records demonstrate butternut likely persisted near the glacial margin, and hindcast species distribution models identified suitable habitat in the southern United States and near Nova Scotia.

Main conclusions: Genetic patterns in butternut may be shaped by both glaciation and modern environmental conditions. Pollen records and hindcast species distribution models combined with genetic distinctiveness in New Brunswick suggest that butternut may have persisted in cryptic northern refugia. We suggest that thorough sampling across a species range and evaluating multiple lines of evidence are essential to understanding past species movements.

Keywords: central-marginal hypothesis, climate change, species migration, tree conservation, fossil pollen, hindcast distribution modeling

Introduction

Determining the causes and consequences of species' range movements are key to understanding their ecology and evolution because range shifts determine community composition, coevolution, distribution of genetic diversity, and niche requirements (Blois, Williams, Grimm, Jackson, & Graham, 2011; Arenas, Ray, Currat, & Excoffier, 2012). Major questions around range movements include migration speed, variation, and peak timing

(Ordóñez & Williams, 2013), the degree to which species tolerate changing climate within their range (Zhu, Woodall, & Clark, 2012), and how much species and communities will shift ranges in response to modern climate change (Ricklefs, 2004). These topics are also relevant for conservation planning because as climates warm, species will need to shift their current ranges to track suitable climate (Lenoir & Svenning, 2015).

An important period of climate change, and resulting organismal range shifts, occurred during and following the last glacial period in the Pleistocene, and in particular the Last Glacial Maximum (LGM). This period largely affected modern day species occurrences and genetic patterns (Petit et al., 2003; Hewitt, 2004), and also exhibited rapid climate change, sometimes as rapid as modern-day climate change (Williams, Blois, & Shuman, 2011). The last glacial period lasted from approximately 130 ka (thousand years before present) to 20 ka, and the LGM lasted from about 26.5-19 ka (Hewitt, 2000), during which glaciers extended far south of their previous extent in the northern hemisphere. As glaciers expanded, species were pushed southward into glacial refugia, typically defined as small, distinct regions of suitable habitat near or surrounded by ice sheets during glacial cycles (Bennett & Provan, 2008). Subsequently, climate warming and glacial retreat allowed species to migrate north into their current ranges in a series of founder events, during which bottlenecks likely led to a reduction in genetic diversity with increasing latitude, while species maintained genetic structure indicative of shared glacial refugia during the LGM (Hewitt, 2004; Fuller & McGlaughlin, 2019). For example, Petit et al., (2003) determined that several woody European species exhibit genetic signatures of being pushed into southern glacial refugia, as demonstrated by clines in genetic diversity (higher diversity in the south and less diversity in the north of ranges) and shared structure representative of shared refugia in the major peninsulas of Europe (in Italy, Iberia, and the Balkans). However, genetic patterns appear to be less consistent in eastern North American tree species (Lumibao, Hoban, & McLachlan, 2017).

Interpreting such migration histories is challenging because genetic patterns may result from multiple evolutionary forces. If species are more impacted by current climatic and geographic factors than past glaciation, species may exhibit genetic patterns more similar to the central marginal hypothesis or isolation by distance (Hoban et al., 2010). The central marginal hypothesis predicts that species will lose genetic diversity as distance increases from the range core, e.g. ecologically suitable habitat (Eckert, Samis, & Loughheed, 2008). Range edges are where individuals of a species are expected to encounter marginal habitat conditions, leading to small population sizes and reduced genetic diversity from bottleneck effects (Lira-Noriega & Manthey, 2014). Species may also exhibit genetic differentiation based on modern geographic barriers or isolation by distance (Wright, 1943). For example, *Euptelea pleiospermum*, a Cenozoic relic tree native to China, exhibits high genetic diversity closer to the center of its range and genetic structure consistent with current geographic barriers (Wei, Sork, Meng, & Jiang, 2016), consistent with expectations of the central marginal hypothesis. In eastern North America, fourteen tree species exhibited little genetic evidence of differentiation based on glaciation (Lumibao et al., 2017), suggesting that some species in this region may not have suffered severe effects of glaciation and eastern North American tree species should be examined for genetic patterns of the central-marginal hypothesis.

The challenges associated with interpreting genetic data can be partly overcome with the addition of other forms of data when trying to elucidate a species' past range movements. Large fossil pollen databases and resolved paleoclimate models can be useful for contributing distinct lines of evidence for past species ranges and migration routes (Hao et al., 2018; Hoban, Kallow, & Treviñ, 2018). Pollen released by trees can be deposited in lakes, bogs, or peatlands, where sedimentary layers accumulate over time. By extracting sediment cores, paleobiologists can identify and count fossil pollen grains within radiocarbon-dated sedimentary layers. Presence of fossil pollen at specific locations and time points can reveal where tree populations occurred thousands of years ago. For example, Davis (1983) charted the migration of 12 major tree genera using fossil pollen records, while others have quantified the speed of range shifts during post-LGM warming for trees in eastern North America (Ordóñez & Williams, 2013).

Species distribution models (SDMs) can reconstruct past suitable habitat range by coupling modern species habitat suitability with paleoclimate models. Knowledge of modern ecological preferences can be hindcast into past climate conditions to predict past locations of suitable habitat. Reconstructions of *Fagus orientalis* distribution were used to identify past suitable habitat that was confirmed by fossil pollen records (Dagtekin, Sahan, Denk, Köse, & Dalfes, 2020). Hindcasting current climate-based species distribution models is particularly useful because climate has been a large determinant of plant distributions since the LGM (Ordóñez & Williams, 2013). Along with reconstructing past species responses to rapid climate change, distribution modeling can identify migration routes, locations of glacial refugia, and biotic velocity through time.

Using multiple lines of evidence, we reinvestigate the postglacial history of butternut (*Juglans cinerea*, L., Juglandaceae), a cold-tolerant relative of black walnut (*J. nigra*) that is currently threatened by habitat fragmentation and butternut canker disease (Pike et al., 2021). Previous studies on butternut showed reduced genetic diversity along a south-to-north gradient, which was interpreted as supporting rapid northward migration following glacial retreat (Hoban et al., 2010). Hoban et al. (2010) also identified strong genetic distinctiveness in New Brunswick populations, which was interpreted as evidence for a cryptic northern glacial refugium during the LGM. These results may have been influenced by undersampling at the northern edge of butternut's distribution, therefore further investigation with more sampling in butternut's northern range is needed to elucidate the root of these genetic patterns.

We use newly available genetic data from butternut populations at the northern range edge and supplement our investigation with fossil pollen data and SDMs. Using three lines of evidence, we evaluate whether:

1. Butternut's genetic diversity declines along a south-to-north gradient, supporting a primary role of postglacial colonization in the genetic diversity structure of the species.
2. New Brunswick populations reveal distinct genetic signatures, suggesting long-term separation from the "main" populations (e.g., as in a separate glacial refuge).

3. SDMs and fossil pollen data can identify possible suitable habitat and pollen (respectively) near New Brunswick, implying the presence of a cryptic refuge,

Methods

Overview of methods

To determine whether previous results in butternut (Hoban et al., 2010) were influenced by sampling gaps, we used newly sampled individuals from butternut's northern range. In addition, we attempted to more directly model butternut's past movements using fossil pollen records and species distribution modeling.

We first assessed the relationship between genetic diversity in butternut populations and their geographic location to determine if diversity decreased with higher latitudes (representative of postglacial colonization) or increased with distance to range edge (representative of the central marginal hypothesis), similar to Hoban et al. (2010). To do this, we: 1. Designated groups of individuals into populations representative of breeding groups, 2. Determined population latitude, longitude, and distance to range edge, 3. Normalized the genetic data and conducted genetic diversity analyses, and 4. Regressed genetic summary statistics against geographic summary statistics (e.g., latitude) to test support for either ecological marginality or postglacial migration hypotheses shaping genetic diversity patterns.

In addition, in the previous study, population differentiation and substructure analysis identified New Brunswick populations as highly differentiated from all other butternut populations, potentially indicative of cryptic glacial refugia near New Brunswick during the LGM. In order to test the hypothesis of a cryptic refugium for butternut during the LGM, we needed to 1. Use genetic structure analysis to determine if, with more extensive sampling, we identified similar patterns, 2. Examine fossil pollen in areas near glacial refugia shortly after the LGM and 3. Identify the degree of past suitable habitat near these proposed refugia using hindcast species distribution models. Lastly, fossil pollen records and past species distribution models were used to calculate biotic velocity (i.e. rate of range shift), which we then interpret in light of current genetic diversity.

Study species

Juglans cinerea L. (butternut) is a medium sized, short-lived hardwood tree with a lifespan around 75 years (Rink, 1990). It is native to eastern North America (Figure 1a) and is an outcrossing, wind pollinated tree. Butternut is threatened by butternut canker disease, caused by *Ophiognomonia clavignenti-juglandacearum* (*Oc-j*), an exotic fungus introduced to North America in the early 20th century (Pike et al., 2021). Butternut, now locally extirpated in many forests, is listed as a "species of concern" in many areas of the United States (U.S.) and is a Regional Forester Sensitive species in 13 of the 16 National Forests in the eastern U.S. (Morin, Gottschalk, Ostry, & Liebhold, 2018) and 23 out of 31 states list butternut as vulnerable, imperiled or critically imperiled (NatureServe, 2021). *Oc-j* has rapidly spread throughout its

naturalized range and in 1997 was found in New Brunswick, where it is disjunct and considered the most northeastern part of butternut's range (Harrison, Hurley, & Ostry, 1998). In the U.S., butternut canker disease results in high tree mortality (Morin et al., 2018). In Canada, *Oc-j* has infected 90%+ of trees in Quebec and Ontario (COSEWIC, 2017) and over 70% of trees from New Brunswick (Williams et al., 2020). In Ontario, an annualized mortality rate of 5.4% was calculated over a 7-year period on 60 sites (COSEWIC, 2017). This rapid decline has led to butternut being listed as endangered at the federal level in Canada and imperiled or critically imperiled in 3 Canadian provinces (COSEWIC, 2017).

Sampling for genetic analyses

In this study, individuals from a previous butternut analysis (see Hoban et al., 2010) were combined with a subsequent sampling effort to better characterize genetic patterns in butternut. The initial sampling effort sampled 1,004 trees from 29 populations, largely from U.S. populations. The second sampling effort sampled 757 butternut trees from northern populations, mostly in Canada (Ontario, Quebec, and New Brunswick).

Genetic diversity and geographic regressions

Geographic data preparation

We organized butternut individuals into cohesive population units to assess how geographic locations were related to genetic diversity. Individuals were sampled at different times, which resulted in disparate sample sizes that could bias genetic analyses. Therefore, we organized butternut individuals into populations by geographic region to represent breeding groups (Figure S1), resulting in 24 populations of 16 - 261 individuals each (Figure 1a). Populations smaller than 15 were removed, following Hoban et al., (2010). Both Quebec populations were below this threshold and were removed from our main analysis, but all analyses were repeated with these populations (see Supplemental). Longitude, latitude, and distance to range edge were then calculated for each population. Mean population latitude and longitude were calculated in the *R* using the 'sp' package 1.4-5 (Pebesma & Bivand, 2005). We created butternut's range extent shapefile in an *R* script (see Figure S2) and calculated the distance to the range edge for each population using its mean location and the *dist2Line* function in the *R* package 'geosphere' 1.5-10 (Hijmans, 2019). Individuals were organized into 24 populations but are visualized by geographic region for clarity (Figure 1a).

Genetic data preparation and analyses

Following population organization, we extracted and genotyped these individuals' DNA according to Hoban, Anderson, McCleary, Schlarbaum, & Romero-Severson (2008) using the microsatellite loci: B114, B159, WGA, A5_2, B157, B212_2, B121, B147, B249, B262, B264. Individuals with more than 25% missing data were removed, leaving a total of 1,635 individuals for genetic analyses. Since genotypes were scored at different times, and microsatellite genotypes can vary due to the equipment and fragment size detection software used, loci were

assessed for differences based on year scored. Details of these analyses are described in-depth in the supplement (Figure S3). Loci were assessed for deviations from Hardy-Weinberg Equilibrium, linkage disequilibrium, and null allele frequency, and then diversity statistics (allelic richness and expected heterozygosity) were calculated by population. Hardy Weinberg Equilibrium, linkage disequilibrium, and expected heterozygosity were assessed in the R package 'poppr' 2.9.2 (Kamvar, Tabima, & Grünwald, 2014) and allelic richness and population pairwise F_{st} s were assessed in the 'hierfstat' R package 0.5-7 (Goudet, 2004). Null alleles were identified using the R package 'PopGenReport' 3.0.4 (Adamack & Gruber, 2014).

Regression analyses

We then regressed each populations' mean genetic diversity levels with their mean geographic location to identify support for either past range-shifts affecting butternut's diversity distribution or modern climatic conditions. A negative linear relationship between mean population latitude and a genetic diversity statistic (allelic richness or expected heterozygosity) would indicate a significant impact of postglacial expansion on butternut, because lower diversity at northern latitudes would result from subsequent bottlenecks as individuals migrate northward following glaciation. Genetic diversity "peaking" in butternut's middle latitudes – a significant quadratic relationship between mean population latitude and diversity – would support the central marginal hypothesis because diversity is highest closest to the center of butternut's range. A significant, positive linear relationship between genetic diversity and distance to range edge would also support the central marginal hypothesis, since genetic diversity would be higher at the center of the species range and lower at the range margins. Regressions were also performed with and without Wisconsin (just above our minimum population sample size threshold) and Quebec (just below our minimum population size threshold) populations since they had much smaller population sizes and diversity may be affected more by small effective size than range-wide diversity patterns. We also tested for isolation by distance (IBD), where populations become more genetically distinct with distance, which may be another indication of modern climate largely affecting butternut, by determining if there was a significant linear relationship between the mean geographic locations of each population and pairwise F_{st} between populations.

Genetic structure analyses

Initial analyses of butternut genetic structure indicated divergence between New Brunswick butternut and all other populations, so to determine if these patterns were consistent with the addition of new individuals, we ran clustering analyses. However, in order to ensure these analyses would not be biased by familial groups, individuals with greater than 25% (half-sibling) relatedness were removed, leaving 993 individuals for these analyses. Relatedness was assessed in the R package 'Demerelate' 0.9-3 (Kraemer & Gerlach, 2017) using the Loiselle statistic, which corrects for small population sizes. Genetic clustering patterns were examined in STRUCTURE (Pritchard, Stephens, & Donnelly, 2000) and principal coordinate analysis (PCoA). Structure results were assessed for the best supported K using the Evanno method

(Evanno, Regnaut, & Goudet, 2005) as implemented in STRUCTURE HARVESTER (Earl & von Holdt, 2012). PCoA was run in the R package 'ade4' (Jombart, 2008).

Identifying past locations of butternut

Fossil pollen analysis

To better resolve butternut's past geographic locations and range shifts, we examined butternut pollen records over time since the LGM. Pollen records were downloaded on August 7, 2020 from the Neotoma Paleoecology Database using the neotoma package in R (Goring et al., 2015) which is documented in more detail in the supplement (Figures S4, S5). Pollen age estimates that were not already converted to calendar years were converted using the 'Bchron' 4.7.6 package in R (Haslett & Parnell, 2008). We retained only butternut pollen counts located in North America east of -110 degrees longitude (east of the Rocky Mountains) our study area of interest.

We visualized pollen records in 1,000-year time increments along with glacial coverage for each time period, which is appropriate because temporal resolution of fossil pollen data in eastern North America during this time period is about 500 years (Blois et al., 2011). We show nine time points here for simplicity and because several time points were similar (See Supplemental Figure S6 for all time points). Glacial coverage was visualized by downloading polygon shapefiles of North American ice sheets from time periods most closely aligned with 1,000-year intervals from 21-1 ka (Dalton et al., 2020). Additionally, we calculated the distance of butternut pollen occurrences to the glacier by averaging the distance from each pollen occurrence to the glacier for each time period to provide us with an average distance to the glacial margin. Distance to the glacier margin was calculated using the dist2Line function in the 'geosphere' package in R (Hijmans, Philips, leathwick, & Elith, 2017) and all pollen maps were constructed using the 'ggplot2' package in R (Wickham & Chang, 2012).

To determine migration rate based on pollen records, we first calculated the location of the centroid of the pollen records within each of 21 time periods by taking the mean of all x, y coordinates falling within the same time bin. We then calculated the Euclidean distance between each centroid and the centroid of the subsequent time bin using the pointDistance function in the 'raster' R package (Hijmans, 2020). Finally, we divided each of the distances (in meters) between the centroids by 1,000 years.

Species distribution modeling

In addition to pollen models and genetic data, butternut's movements before, during, and since the LGM were assessed by projecting current distribution models into past climate scenarios. This process involved 1. Determining butternut's modern climate niche using boosted regression trees (BRT), 2. Predicting the locations of butternut populations in the past by projecting butternut's modern ecological niche into 8 past time periods, and 3. Visualizing possible locations of past butternut populations by mapping the projected habitat suitability

across time. BRT were used to predict butternut's modern occurrence because they can use many types of data and handle missing data and outliers well (Elith, Leathwick, & Hastie, 2008). In summary, we compared 19 WorldClim variables (Fick & Hijmans, 2017) with butternut presence and selected the 5 variables most correlated with presence and least autocorrelated, and generated a prediction of suitable habitat (see Figure S7, Table S1).

Then, to determine how butternut's range changed in the past, butternut's habitat suitability model was projected into eight past climate scenarios to represent notable periods in post-LGM climatic history, available from the Paleoclim database (Brown, Hill, Dolan, Carnaval, & Haywood, 2018): 130 ka (last interglacial); 22 (LGM); 17-14.7 ka (Heinrich-Stadial); 14.7-12.9 ka (Bolling-Allerod); 12.9-11.7 ka (Younger Dryas); 11.7-8.326 ka (early Holocene); 8.326-4.2 ka (mid-Holocene); and 4.2-0.3 ka (late Holocene). We visualized butternut's habitat suitability across these time periods using the 'dismo' R package (Hijmans et al., 2017) but include six time points here for simplicity (see Supplemental Figure S8 for all time point prediction maps). We then used these maps to calculate the speed of range migration over time using the bioticVelocity function in the 'enmSdm' R package (Smith, 2021). The function uses as input a series of rasters and each raster's time point. In our case, we used the eight SDM output rasters, each of which represents a different time range from 130 ka to present. For the vector of times, we took the middle (i.e., the mean) of the SDM time ranges. Results were converted to meters per year.

Results

Genetic diversity overview

Overall, 1,635 butternut individuals were successfully genotyped and retained for final genetic diversity and structure analyses. Diversity was relatively high in populations, with the total number of alleles observed across all 11 loci was 233 (mean per locus = 21.2, mean per population = 126.2) and the number of alleles within populations ranged from 83 to 168 (Table S2). Allelic richness and expected heterozygosity ranged from 6.47 - 9.23 and 0.751 - 0.852, respectively. Wisconsin populations had the lowest diversity of all populations (allelic richness lower than 7, expected heterozygosity lower than 0.78) (Table S2). Deviation from HWE expectations was locus and population specific (Table S3), likely indicating close relatives in some populations, and linkage disequilibrium was observed between some loci (Table S4). Null alleles were infrequent in most loci, near or exceeding mean frequency of 10% only in B157, B212_2, and B264 (means = 9.00%, 13.1%, and 16.8% respectively) (Table S5).

Genetic diversity highest near the center of butternut's range

Regression of population geographic metrics (e.g., mean latitude, distance to range edge) with genetic diversity metrics (e.g., allelic richness, expected heterozygosity) support the central marginal hypothesis rather than the postglacial migration hypothesis, contrary to our hypothesis. To test for the postglacial colonization hypothesis, we regressed mean population latitude with

genetic diversity metrics (e.g., allelic richness and expected heterozygosity) and these relationships were significant but not well fit to the data (Table 1). To test for the central marginal hypothesis, we performed quadratic regressions between genetic diversity metrics and mean population latitude and also a linear relationship between population distance to range edge, as lower diversity at range edges is more indicative of the central marginal hypothesis. We found a significant quadratic relationship between mean population latitude and a significant, positive linear relationship between genetic diversity metrics and population distance to range edge, without Wisconsin populations (Table 1). When Wisconsin populations were included in regressions, the relationships were either not significant or well-fit to the data (Table 1), suggesting Wisconsin populations are likely more affected by small effective population size than range-edge effects. All of these patterns are the same when Quebec populations are added (Tables S7, S8). Isolation by distance would be supported by a significant linear regression with pairwise F_{st} and distance; this regression was significant but explained little of the variance in the data (Figure S9).

[Table 1 here]

Genetic structure is distinct in New Brunswick

Genetic structure analyses were run to determine if genetic structure remained distinct between New Brunswick and other butternut populations, and our results supported this hypothesis. STRUCTURE analyses, with the additional individuals, support previously found patterns, with New Brunswick distinct from all other populations (Figure 2). STRUCTURE Harvester (Earl & von Holdt, 2012) determined that $\Delta K=2$ was the best supported clustering pattern (Table S6, Figure S10), but other K values support distinction of New Brunswick populations (Figure S11). Principal coordinate analysis similarly supported this clustering pattern analysis with New Brunswick individuals forming one distinct cluster, separate from all other populations, with a slight distinction of Wisconsin populations and one Ontario population (Figure S12). When Quebec populations are added, New Brunswick is still distinct from Ontario and United States populations (Figure S13). The placement of Quebec populations differs between PCoA and Structure.

[Figure 2 here]

Fossil pollen is found near New Brunswick during the LGM

Fossil pollen records were examined to determine if there was evidence for butternut persisting near New Brunswick, close to glacial margins. Butternut pollen was not identified near New Brunswick during the LGM, but some terrestrial environments exposed during the LGM are currently under ocean water. Butternut pollen first appeared in New Brunswick, Nova Scotia, and southern Quebec at 14 ka, when ice sheets were first breaking apart during deglaciation and habitat was becoming available between smaller ice sheets. With further glacial retreat, pollen occurrences became more numerous within the current butternut range (Figure 3). Butternut pollen was commonly found close to the glacial margin at the LGM and tracked glacial

retreat closely until about 10 ka (Figure S12), when butternut reached its present-day range (Figure 3, Figure S6). Migration rates calculated from pollen records ranged from under 50 m/year - 450m/year (at 17 ka), oscillating between 50 - 200 m/year from 16 - 2 ka (Figure S14).

[Figure 3 here]

Hindcast distribution models support presence of suitable habitat near New Brunswick

We used SDMs to predict the presence of suitable habitat for butternut during and following the LGM. Butternut's modern ecological niche was best predicted by the following variables: precipitation of the wettest month, mean diurnal range, mean temperature of the driest quarter, mean temperature of the wettest quarter, and seasonal precipitation (Table S1). During the last interglacial period, 130 ka, butternut's suitable habitat extended far north of its current suitable habitat, extending into much of Canada and Maine. During the LGM, butternut shows a large band of suitable habitat in the southeastern U.S., smaller pockets of suitable habitat in the northeastern U.S., and a more isolated area of suitable habitat south of Nova Scotia. Butternut's range appeared to shrink leading up to the mid-Holocene, then expanded until it reached its present-day distribution (Figure 4). Migration rates calculated from SDM output ranged from almost 0 at the LGM to up to 200 m/year during the Younger Dryas period (12.9 - 11.7 ka) (Figure S13).

[Figure 4 here]

Discussion

In this study, we aimed to use additional sampling as well as fossil pollen and hindcast distribution models to better understand the impact of glaciation and postglacial range shifts on the genetic patterns of the rare species, *Juglans cinerea* (butternut). Specifically, we reassessed previous interpretation of genetic patterns after further scrutiny of the northern range edge. Our results provide a more detailed picture of the range shifts of butternut, and suggest that, similar to other eastern North American native tree species, postglacial northward colonization may not be the only factor shaping genetic trends. We observe, contrary to results from Hoban et al., (2010) and our original hypothesis, that modern ecogeographic processes, specifically ecological marginality, are likely shaping butternut's genetic diversity. Also, we observe that New Brunswick butternut populations remain genetically distinct, supporting the hypothesis of a glacial refuge in this area during the Last Glacial Maximum. SDMs and pollen evidence support the hypothesis that this area may have served as a cryptic glacial refugium during the LGM. Overall, our findings suggest that genetic patterns within butternut were shaped by both modern climate and past glacial movements, and that genetic data should be complemented by other lines of biogeographic evidence. Our results concur with findings that postglacial movements of eastern North American trees are more idiosyncratic than their European counterparts (Soltis, Morris, McLachlan, Manos, & Soltis, 2006; Lumibao et al., 2017; Bemmels & Dick, 2018).

Genetic diversity is best explained by central marginal hypothesis, not postglacial expansion

A previous examination in Hoban et al., (2010) found that butternut's northernmost populations had the lowest diversity, similar to observations in twenty-one woody species in Europe, likely due to successive postglacial migration bottlenecks (Petit et al., 2003). However, with the addition of more sampling near butternut's northern range-edge, genetic diversity patterns are more similar to the patterns predicted by the central marginal hypothesis, rather than postglacial migration or isolation by distance. The previous findings that genetic diversity declined with increased latitude may have been partly due to sampling strategy. In Hoban et al. (2010), fewer individuals were sampled from butternut's northern range and most northern range individuals were sourced from Wisconsin populations. Wisconsin forests have been strongly impacted by modern land-use conversion and have small butternut populations, leading to unusually low genetic diversity in these populations, which do not fit well to the species-wide relationship between genetic diversity and ecology (Table 1). Including the smaller populations from Quebec, butternut populations fit the pattern of higher diversity closer to the center of the range, where habitat is higher quality (Gullberg, Olsson, & Tegelström, 1998). Overall, with improved sampling, we suggest butternut's genetic diversity patterns are more likely shaped by modern ecological marginality than past range shifts.

Genetic structure analyses support distinction of New Brunswick

Range-wide patterns of genetic structure provide additional information about a species' biogeographic history by suggesting which populations had more genetic connectivity. Supporting our hypothesis, structure and PCoA analyses suggest butternut has two distinct genetic clusters, with the U.S. and Ontario populations forming one cluster and New Brunswick populations forming a second (Figure 2, Figure S12, Table S6). We recognize that structure is biased towards selecting K=2 as the optimal K (Cunningham et al., 2020), however, other K values similarly identified New Brunswick as distinct from other populations (Figure S11). Quebec populations cluster differently based on clustering methods, as they cluster with New Brunswick in structure analysis and cluster independently in PCoA analysis (Figure S13). This could be due to small population sizes, limited spatial sampling in Quebec, or intermediate levels of genetic connectivity over history. Deeper sampling is needed in these populations to determine their genetic history. Generally, our genetic structure results support the findings of Hoban et al. (2010) that identified strong genetic differentiation of New Brunswick populations from all other geographic locations.

The genetic distinction of New Brunswick populations is consistent with the hypothesis that these populations persisted in a glacial refugium near New Brunswick during the LGM, which would have limited connectivity with other populations (Bemmels, Knowles, & Dick, 2019). Other species like red pine (*Pinus resinosa*) and skunk cabbage (*Symplocarpus foetidus*) similarly exhibit distinct genetic structure in northeastern North America (Walter & Epperson, 2005; Kim, Cho, Li, & Kim, 2018) supporting a refuge for many species near New Brunswick. The genetic

differentiation identified here is potentially indicative of former range separations in butternut during the last glacial period.

Fossil pollen and SDMs support the possibility of a cryptic refugia in New Brunswick

Butternut's pollen records and species distribution modeling largely support our hypothesis and the genetic evidence of a cryptic glacial refuge near New Brunswick. The first consistent pollen record in New Brunswick dates to 14 ka (Figure 3, though note pollen in upper New England, e.g. Massachusetts dates to 21 ka), when the Laurentide ice sheet was breaking into smaller ice sheets across southern Quebec, New Brunswick, and Nova Scotia. Butternut populations may have been present in these northern locations before 14 ka in areas currently under the ocean, where we do not have access to pollen records. SDMs support the presence of suitable habitat south of Nova Scotia from 20 ka to present, supporting the idea that butternut could have persisted in a refugium near this area (Figure 4, Figure S6). While suitability scores were low (20 - 30%) near Nova Scotia and New Brunswick in certain time periods (e.g., 17.9 - 12 ka) suitable habitat is still present in those areas during all time periods (Figure 4, Figure S8). We identified a lake off the coast of modern-day Maine that separated two portions of butternut's suitable habitat 20 ka which could have led to butternut's two genetic clusters.

Assessing butternut's dispersal ability also allows us to assess the likelihood of a cryptic glacial refugium near New Brunswick. Tree squirrels are butternut's main seed dispersal vector and generally disperse below 100 m (Larrichia, McCleary, Hoban, Borkowski, & Romero-Severson, 2015), while seeds of butternut's relative, *Juglans nigra*, were observed to be moved as far as 2,000 m by crows (Cristol, 2001). If squirrels or crows dispersed butternut seeds, the biotic velocity would be 100 - 2000 m/generation, or 10 - 200 m/year (with 10 years to reach sexual maturity to produce seed). However, biotic velocities derived from butternut fossil pollen exceed the biotic velocities of butternut's possible seed dispersers. Butternut pollen first appeared near New Brunswick around 14 ka, meaning butternut populations would have had to travel around 450 km/year to cover the distance from the nearest pollen record in 1,000 years (Figure 3, Figure S6, Figure S12), unless populations already persisted near New Brunswick. SDMs predicted 50 - 200 m/year biotic velocities, much slower than pollen-prediction biotic velocities because they included New Brunswick as suitable habitat, so butternut would not have had to travel from southern North America to New Brunswick following deglaciation (Figure 4, Figure S14). SDM biotic velocity predictions are much more consistent with modern dispersal predictions. Thus, given what we know about dispersal in butternut, these results support a cryptic refugium near New Brunswick during the LGM.

Caveats

Each data source used in this paper has deficiencies (Hoban et al., 2018). Genetic methods represent multiple ecogeographic patterns and therefore provide imperfect evidence of the past. New Brunswick populations demonstrate a slight reduction in genetic diversity, consistent with patterns representative of the central marginal hypothesis, but could also represent a bottleneck in New Brunswick populations following postglacial migration. Therefore, while we provided

Commented [1]: @shoban@mortonarb.org @alissabrown@mortonarb.org also cut down this paragraph a bit to avoid repetition - let me know what you think

Commented [2R1]: I like it!

Commented [3R1]: @shoban@mortonarb.org can you give this paragraph a once over - does it make sense?

Commented [4R1]: I think its pretty good. Its a hard topic to tackle, in a short space. But this looks good. I suggest a couple further deletions

Commented [5R1]: Thank you for this review! I agree, this is a challenging thing to cover in an already stuffed manuscript, so your help really appreciated in making it coherent

support for modern ecogeographic processes influencing genetic patterns in this species the most, it is important to note all mentioned ecological processes are likely influencing butternut to some degree, similar to results examining the genetic patterns of other tree species native to eastern North America (Bemmels & Dick, 2018). Fossil pollen data have known limitations as well. Pollen can be carried by wind, even hundreds of kilometers from a population (McLachlan & Clark, 2004) and so the presence of fossil pollen at a site does not necessarily mean a tree population existed at that location. To avoid interpreting a single pollen occurrence as evidence of the presence of a butternut population, we examined maps including many pollen occurrences (Figure S5) and interpreted patterns with SDM and genetic evidence. Species distribution models assume that climate data alone provide an accurate picture of ecological niches and butternut's fundamental niche has not changed during the last 130,000 years. Solar radiation, available canopy cover, and soil characteristics also shape butternut's distribution and habitat preferences (Crystal & Jacobs, 2014), but were not included in our SDMs. However, comparisons of past range models and pollen distributions have found similarity among past climate modeling and past range distributions (Ordonez & Williams, 2013) and between multiple climate models (Beyer, Krapp, & Manica, 2020). Overall, consistency between genetic patterns, fossil pollen records, and suitable habitat suggest the usefulness of this modeling technique and combining multiple lines of evidence.

Conclusions

We add support to several studies (Soltis et al., 2006; Lumibao et al., 2017; Bemmels & Dick, 2018) that have identified weak or inconsistent genetic diversity impacts (weak clines in genetic diversity with latitude, and structure is relatively idiosyncratic) from postglacial migrations in eastern North American tree species, contrasted to the strong structure in Europe (Petit et al., 2003), possibly due to contemporary ecological marginality. It is certainly challenging to distinguish the central marginal pattern from other processes, and our results based on adding more complete sampling to previous data suggest that other previous studies may warrant revisiting with more complete range-wide data (McLachlan, Clark, & Manos, 2005; Victory, Glaubitz, Rhodes, & Woeste, 2006). The sufficiency of geographic range coverage may also be useful to consider in meta analyses such as those evaluating the central marginal hypothesis and biotic velocity (e.g. Hamann, Glaubitz, Rhodes, & Woeste, 2015). We conclude that a more robust sample has provided better evidence of past genetic patterns and contributed to a better understanding of past demographic histories of butternut.

Additionally, we emphasize the importance of leveraging evidence from multiple data sources when evaluating species biotic velocity, or migration speed, and when trying to locate species' glacial refuges. Genetic evidence alone is inconclusive on the possible New Brunswick refuge and biotic velocity. Examining pollen fossils with spatiotemporal resolution and niche models provides strongly suggestive, though not fully conclusive, evidence of this northern glacial refuge for butternut. Comparing the findings from the distribution models and pollen evidence also gives a better sense of uncertainty – these two lines of evidence are only partly in agreement on overall velocity or on periods of more or less rapid movement. Uncertainty in such

estimates is important for determining contributing factors to biotic velocity, and for reaching conclusions about forests' ability to track climate change (Hoban, Dawson, Robinson, Smith, & Strand, 2019). Further improvement to our work can be made by using next generation sequencing and Approximate Bayesian Computation to more formally integrate genetics with other lines of evidence to understand glacial refugia and migration speed (Bemmels, Knowles, & Dick, 2019).

Conservation Applications

First, we note that butternut's habitat may be under threat due to climate change. SDMs generated for this study suggest butternut's range is largely predicted by seasonal precipitation, patterns of which are changing rapidly due to climate change (Pendergrass, Knutti, Lehner, Deser, & Sanderson, 2017), particularly in northern areas where precipitation is already quite variable (Konapala, Mishra, Wada, & Mann, 2020). Previous analyses have also found that butternut is also generally found in mesic habitat, and changes to seasonal precipitation are also predicted to change localized stream and river flow (Nkhonjera, Dinka, & Woyessa, 2020). Bemmels et al., (2019) identified that many modern cold-tolerant species specialized on mesic environments could have persisted very close to glacial margins, which seems possible for butternut due to the locations predicted suitable habitat and fossil pollen records, but this, in turn, makes them particularly under threat by these climate change related effects. Areas that may be suitable for this species now may not be suitable for this species in 50 years (Barnes & Delborne, 2019), so using climate change modeling and carefully considering habitat needs of this species are essential when planning restoration (Pike et al., 2021).

Second, the geographic and genetic distinction of New Brunswick butternut indicates these populations may warrant special protection and possibly their own taxonomic unit for conservation. In Canada, the Committee on the status of endangered wildlife in Canada (COSEWIC) has the mandate to evaluate if a species should require its own taxonomic unit or designatable unit (DU) below the species level. In order to qualify, DUs need to be recognized as discrete but also represent evolutionarily significant units when compared to other populations. We have shown that New Brunswick populations are genetically distinctive from the rest of the species range. Discreteness is also supported geographically and climatically. New Brunswick populations are naturally disjunct, and may have been for a long time, and appear within their own ecozone – the Acadian Forest ecozone. Of course, genetic differentiation could be further supported using more slowly evolving DNA markers that show either a deep intraspecific phylogenetic divergence or strong natural selection due to environmental conditions, which has not been established (Larrichia, McCleary, Hoban, Borkowski, & Romero-Severson, 2015). Although more data and analysis are required to fully characterize the significance of New Brunswick populations as a DU, the data herein supports the need to follow through with specific conservation activities that will capture the maximum amount of remaining genetic diversity from both groups before it is lost to butternut canker.

Summary

Our work better resolved past locations and movements of butternut using multiple lines of evidence, and we suggest there is potential for re-evaluating past conclusions on other eastern North American taxa. A better understanding of more species' past range shifts in eastern North America is critical to comparing biogeographic understanding compared to European taxa, and to better understand how these forests may respond to future changes in climate.

Data availability

Data are available on Dryad at the <weblink to be inserted at proof stage>. Genetic data R code for genetic analyses and species distribution modeling are available on E. Schumacher's Github (<https://github.com/ekschumacher/butternut>). R code for pollen data download, analysis, and visualization, as well as species distribution model maps with glacier coverage are available on A. Brown's Github (<https://github.com/alissab/juglans>).

References

- Adamack, A. T., & Gruber, B. (2014). PopGenReport: simplifying basic population genetic analyses in R. *Methods in Ecology and Evolution*, 5, 384-387.
- Arenas, M., Ray, N., Currat, M., & Excoffier, L. (2012). Consequences of range contractions and range shifts on molecular diversity. *Molecular Biology and Evolution*, 29, 207-218.
- Barnes, J. C., & Delborne, J. A. (2019). Rethinking restoration targets for American chestnut using species distribution modeling. *Biodiversity and Conservation*, 28, 3199-3220. <https://doi.org/10.1007/s10531-019-01814-8>
- Bemmels, J. B., & Dick, C. W. (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. <https://doi.org/10.1111/jbi.13358>
- Bemmels, J. B., Knowles, L. L., & Dick, C. W. (2019). Genomic evidence of survival near ice sheet margins for some, but not all, North American trees. *Proceedings of the National Academy of Sciences*, 116, 8431-8436.
- Bennett, K. D., & Provan, J. (2008). What do we mean by "refugia"? *Quaternary Science Reviews*, 27, 2449-2455. <https://doi.org/10.1016/j.quascirev.2008.08.019>
- Beyer, R. M., Krapp, M., & Manica, A. (2020). High-resolution terrestrial climate, bioclimate and vegetation for the last 120,000 years. *Scientific Data*, 7, 1-9.
- Blois, J. L., Williams, J. W. J., Grimm, E. C., Jackson, S. T., & Graham, R. W. (2011). A methodological framework for assessing and reducing temporal uncertainty in paleovegetation mapping from late-Quaternary pollen records. *Quaternary Science Reviews*, 30, 1926-1939.

Brown, J. L., Hill, D. J., Dolan, A. M., Carnaval, A. C., & Haywood, A. M. (2018). Paleoclim, high spatial resolution paleoclimate surfaces for global land areas. *Scientific Data*, 5, 1–9. <https://doi.org/10.1038/sdata.2018.254>

Committee on the Status of Endangered Wildlife in Canada (COSEWIC). (2017). COSEWIC assessment and status report on the butternut *Juglans cinerea* in Canada. COSEWIC, Ottawa, ON, Canada.

Cristol, D. A. (2001). American crows cache less preferred walnuts. *Animal Behaviour*, 62, 331–336.

Crystal, P. A., & Jacobs, D. F. (2014). Drought and flood stress tolerance of butternut (*Juglans cinerea*) and naturally occurring hybrids: Implications for restoration. *Canadian Journal of Forest Research*, 44, 1206–1216. <https://doi.org/10.1139/cjfr-2014-0151>.

Cullingham, C. I., Miller, J. M., Peery, R. M., Dupuis, J. R., Malenfant, R. M., Gorrell, J. C., & Janes, J. K. (2020). Confidently identifying the correct K value using the ΔK method: when does $K = 2$? *Molecular Ecology*, 29, 862–869.

Dalton, A. S., Margold, M., Stokes, C. R., Tarasov, L., Dyke, A. S., Adams, R. S., ... Wright Jr, H. E. (2020). An updated radiocarbon-based ice margin chronology for the last deglaciation of the North American Ice Sheet Complex. *Quaternary Science Reviews*, 234, 106223.

Davis, M.B. (1983). Quaternary history of deciduous forests of eastern North America and Europe. *Annals of the Missouri Botanical Garden*, 70, 550–563.

Dagtekin, D., Şahan, E. A., Denk, T., Köse, N., & Dalfes, H. N. (2020). Past, present and future distributions of Oriental beech (*Fagus orientalis*) under climate change projections. *PLoS ONE*, 15, e0242280.

Earl, D. A., & vonHoldt, B. M. (2012). STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, 4, 359–361.

Eckert, C. G., Samis, K. E., & Loughheed, S. C. (2008). Genetic variation across species' geographical ranges: the central–marginal hypothesis and beyond. *Molecular Ecology*, 17, 1170–1188.

Elith, J., Leathwick, J. R., & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology*, 77, 802–813.

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology*, 14, 2611–2620.

Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, 37, 4302-4315.

Fuller, R. S., & McGlaughlin, M. E. (2019). *Calochortus gunnisonii* furthers evidence for the complex genetic legacy of historical climate change in the southern Rocky Mountains. *American Journal of Botany*, 106, 477-488.

Goring, S., Dawson, A., Simpson, G., Ram, K., Graham, R., Grimm, E., & Williams, J. (2015). Neotoma: A programmatic interface to the Neotoma Paleoecological Database. *Open Quaternary*, 1.

Goudet, J. (2004). Hierfstat, a package for R to compute and test hierarchical F-statistics. *Molecular Ecology Notes*, 5, 184-186. <https://doi.org/10.1111/j.1471-8286.2004.00828.x>

Gullberg, A., Olsson, M., & Tegelström, H. (1998). Colonization, genetic diversity, and evolution in the Swedish sand lizard, *Lacerta agilis* (Reptilia, Squamata). *Biological Journal of the Linnean Society*, 65, 257-277. <https://doi.org/10.1111/j.1095-8312.1998.tb01142.x>

Hamann, A., Roberts, D. R., Barber, Q. E., Carroll, C., & Nielsen, S. E. (2015). Velocity of climate change algorithms for guiding conservation and management. *Global Change Biology*, 21, 997-1004.

Hao, Q., de Lafontaine, G., Guo, D., Gu, H., Hu, F. S., Han, Y., ... Liu, H. (2018). The critical role of local refugia in postglacial colonization of Chinese pine: joint inferences from DNA analyses, pollen records, and species distribution modeling. *Ecography*, 41, 592-606.

Harrison, K. J., Hurley, J. E. A., & Ostry, M. E. (1998). First report of butternut canker caused by *Sirococcus clavigignenti-juglandacearum* in New Brunswick, Canada. *Plant Disease*, 82, 1282-1282.

Haslett, J., & Parnell, A. (2008). A simple monotone process with application to radiocarbon-dated depth chronologies. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 57, 399-418.

Hewitt, G. M. (1999). Postglacial re-colonization of European biota. *Biological Journal of the Linnean Society*, 68, 87-112.

Hewitt, G. M. (2000). The genetic legacy of the Quaternary ice ages. *Nature*, 405. www.nature.com

Hewitt, G. M. (2004). Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 359, 183-195.

Hijmans, R. J., Phillips, S., Leathwick, J. R., & Elith, J. (2017). Dismo package for R, version 1.1-4. *Circles*, 9, 1–68. <https://doi.org/10.1016/j.jhydrol.2011.07.022>.

Hijmans, R. J. (2019). Introduction to the "geosphere" package (Version 1.5-10).

Hijmans, R.J. (2020). Raster: Geographic data analysis and modeling. R package version 3.4-8. <https://rspatial.org/raster>

Hoban, S., Anderson, R., McCleary, T., Schlarbaum, S., & Romero-Severson, J. (2008). Thirteen nuclear microsatellite loci for butternut (*Juglans cinerea* L.). *Molecular Ecology Resources*, 8, 643-646.

Hoban, S. M., Borkowski, D. S., Brosi, S. L., McCleary, T. S., Thompson, L. M., McLachlan, J. S., ... Romero-Severson, J. (2010). Range-wide distribution of genetic diversity in the North American tree *Juglans cinerea*: a product of range shifts, not ecological marginality or recent population decline. *Molecular Ecology*, 19, 4876–4891. <https://doi.org/10.1111/j.1365-294X.2010.04834.x>

Hoban, S. M., McCleary, T. S., Schlarbaum, S. E., & Romero-Severson, J. (2014). Spatial genetic structure in 21 populations of butternut, a temperate forest tree (*Juglans cinerea* L.), is correlated to spatial arrangement, habitat, and land-use history. *Forest Ecology and Management*, 314, 50-58.

Hoban, S., Kallow, S., & Trivedi, C. (2018). Implementing a new approach to effective conservation of genetic diversity, with ash (*Fraxinus excelsior*) in the UK as a case study. *Biological Conservation*, 225, 10-21.

Hoban, S., Dawson, A., Robinson, J. D., Smith, A. B., & Strand, A. E. (2019). Inference of biogeographic history by formally integrating distinct lines of evidence: genetic, environmental niche and fossil. *Ecography*, 42, 1991-2011.

Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24, 1403-1405.

Kamvar, Z. N., Tabima, J. F., & Grünwald, N. J. (2014). Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, 2, e281.

Kim, S. H., Cho, M. S., Li, P., & Kim, S. C. (2018). Phylogeography and ecological niche modeling reveal reduced genetic diversity and colonization patterns of skunk cabbage (*Symplocarpus foetidus*; araceae) from glacial refugia in Eastern North America. *Frontiers in Plant Science*, 9, 648. <https://doi.org/10.3389/fpls.2018.00648>

Konapala, G., Mishra, A. K., Wada, Y., & Mann, M. E. (2020). Climate change will affect global water availability through compounding changes in seasonal precipitation and evaporation. *Nature Communications*, 11, 1-10.

Kraemer, P., & Gerlach, G. (2017). Demerelate: calculating interindividual relatedness for kinship analysis based on codominant diploid genetic markers using R. *Molecular Ecology Resources*, 17, 1371-1377.

Laricchia, K. M., McCleary, T. S., Hoban, S. M., Borkowski, D., & Romero-Severson, J. (2015). Chloroplast haplotypes suggest preglacial differentiation and separate postglacial migration paths for the threatened North American forest tree *Juglans cinerea* L. *Tree Genetics and Genomes*, 11, 1–11. <https://doi.org/10.1007/s11295-015-0852-3>

Lenoir, J., & Svenning, J. C. (2015). Climate-related range shifts—a global multidimensional synthesis and new research directions. *Ecography*, 38, 15-28.

Lira-Noriega, A., & Manthey, J. D. (2014). Relationship of genetic diversity and niche centrality: a survey and analysis. *Evolution*, 68, 1082-1093.

Lumibao, C. Y., Hoban, S. M., & McLachlan, J. (2017). Ice ages leave genetic diversity 'hotspots' in Europe but not in Eastern North America. *Ecology Letters*, 20, 1459–1468. <https://doi.org/10.1111/ele.12853>postglacial

McLachlan, J.S., & Clark, J.S. (2004). Reconstructing historical ranges with fossil data at continental scales. *Forest Ecology and Management*, 197, 139-147.

McLachlan, J. S., Clark, J. S., and Manos, P. S. (2005). Molecular indicators of tree migration capacity under rapid climate change. *Ecology*, 86, 2088-2098.

Médail, F., & Diadema, K. (2009). Glacial refugia influence plant diversity patterns in the Mediterranean Basin. *Journal of Biogeography*, 36, 1333-1345.

Morin, R. S., Gottschalk, K. W., Ostry, M. E., & Liebhold, A. M. (2018). Regional patterns of declining butternut (*Juglans cinerea* L.) suggest site characteristics for restoration. *Ecology and Evolution*, 8, 546-559.

NatureServe Explorer, 20 Dec. 2021, https://explorer.natureserve.org/Taxon/ELEMENT_GLOBAL.2.154365/Juglans_cinerea, Accessed 23 April 2021.

Nkhonjera, G. K., Dinka, M. O., & Woyessa, Y. E. (2020). Assessment of localized seasonal precipitation variability in the upper middle catchment of the Olifants River basin. *Journal of Water and Climate Change*, 12, 250-264. <https://doi.org/10.2166/wcc.2020.187>

Ordonez, A., & Williams, J. W. (2013). Climatic and biotic velocities for woody taxa distributions over the last 16, 000 years in eastern North America. *Ecology Letters*, 16, 773–781.

Pebemsa, E.J., & Bivand, R.S. (2005). Classes and methods for spatial data in R. *R News*, 5 <https://cran.r-project.org/doc/Rnews/>

Pendergrass, A. G., Knutti, R., Lehner, F., Deser, C., & Sanderson, B. M. (2017). Precipitation variability increases in a warmer climate. *Scientific Reports*, 7, 1-9.

Petit, R. J., Aguinagalde, I., De Beaulieu, J. L., Bittkau, C., Brewer, S., Cheddadi, R., ... Vendramin, G. G. (2003). Glacial refugia: Hotspots but not melting pots of genetic diversity. *Science*, 300, 1563–1565.

Pike, C. C., Williams, M., Brennan, A., Woeste, K., Jacobs, J., Hoban, S., ... Romero-Severson, J. (2021). Save our species: a blueprint for restoring butternut (*Juglans cinerea*) across eastern North America. *Journal of Forestry*, 119, 196-206.

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

Ricklefs, R. E. (2004). A comprehensive framework for global patterns in biodiversity. *Ecology Letters*, 7, 1-15.

Rink, G. (1990). *Juglans cinerea* L. butternut. *Silvics of North America*, 2, 386-390.

Smith, A.B. (2021). enmSdm: tools for modeling niches and distributions of species. R package version 0.5.3.1. <https://rdr.io/github/adamlilith/enmSdm/>

Soltis, D. E., Morris, A. B., McLachlan, J. S., Manos, P. S., & Soltis, P. S. (2006). Comparative phylogeography of unglaciated eastern North America. *Molecular Ecology*, 15, 4261-4293.

Victory, E. R., Glaubitz, J. C., Rhodes, O. E., & Woeste, K. E. (2006). Genetic homogeneity in *Juglans nigra* (Juglandaceae) at nuclear microsatellites. *American Journal of Botany*, 93, 118–126. <https://doi.org/10.3732/ajb.93.1.118>

Walter, R., & Epperson, B. K. (2005). Geographic pattern of genetic diversity in *Pinus resinosa*: contact zone between descendants of glacial refugia. *American Journal of Botany*, 92, 92–100. <https://doi.org/10.3732/ajb.92.1.92>

Wei, X., Sork, V. L., Meng, H., & Jiang, M. (2016). Genetic evidence for central-marginal hypothesis in a Cenozoic relict tree species across its distribution in China. *Journal of Biogeography*, 43, 2173–2185. <https://doi.org/10.1111/jbi.12788>

Wickham, H., & Chang, W. (2012). ggplot2. Computer software. Retrieved from <http://ggplot2.org>.

Williams, J. W., Blois, J. L., & Shuman, B. N. (2011). Extrinsic and intrinsic forcing of abrupt ecological change: case studies from the late Quaternary. *Journal of Ecology*, 99, 664-677.

Williams, M., Moise, E. R., Forbes, K., Williams, C., DeMerchant, I., & Beardmore, T. (2020). Ecological Status of *Juglans cinerea* in New Brunswick. *Plant Disease*, 104, 860-867.

Wright, S. (1943). Isolation by distance. *Genetics*, 28, 114.

Zhu, K., Woodall, C. W., and Clark, J. S. (2012). Failure to migrate: lack of tree range expansion in response to climate change. *Global Change Biology*, 18, 1042-1052.

Biosketch: Emily Schumacher is an early-career research assistant interested in preserving rare plant species using genetic data and distribution modeling methods. She is particularly interested in assessing genetic diversity and climate resilience of endangered plant species conserved in *ex situ* collections. She is also focused on using species distribution modeling to determine ecological specializations of plant taxa and their past and future climate change response.

Author contribution: ES reorganized and cleaned genetic data, ran statistical analyses, and performed SDMs. ES and AB co-wrote the first draft and led the editing of the manuscript. AB downloaded, cleaned, and assessed fossil pollen records; provided advice for analyses; and calculated biotic velocities. MW wrote parts of the manuscript, led field and lab work and contributed to discussions. TB obtained funding of the field, lab and genotyping component and provided draft manuscript material. JRS provided genetic data, performed initial analyses, contributed to discussions, and edited the manuscript. SH wrote and edited parts of the manuscript, guided analyses, and supervised the project.

Tables and Figures

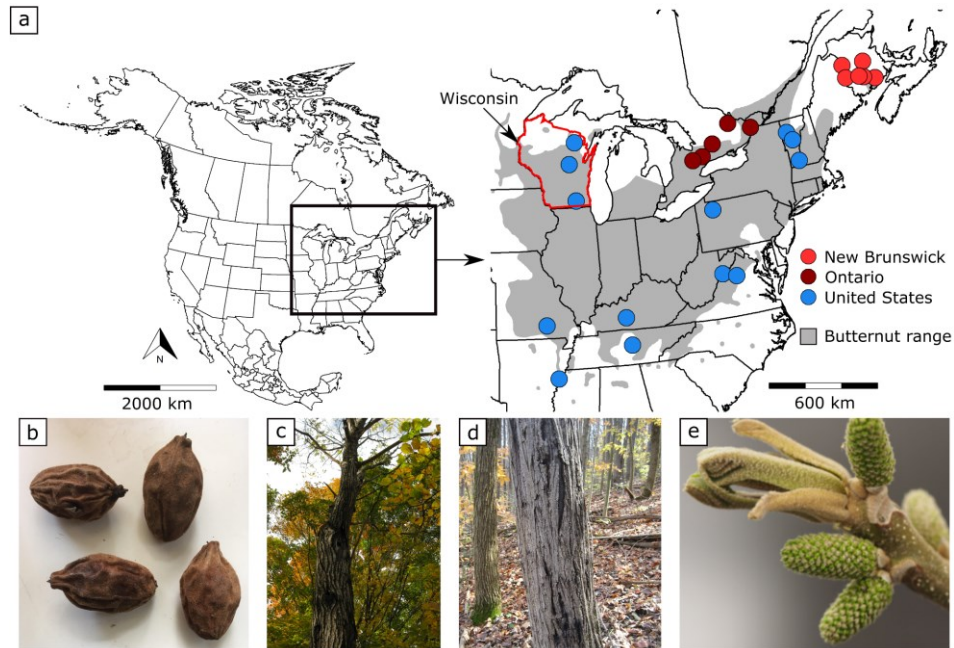


Figure 1. a: map of the 24 populations containing the 1,635 butternut individuals included in genetic analyses, separated into 24 populations. b: Butternut seeds. c: Butternut individual. d: Butternut tree trunk with butternut cankers. e: Butternut catkins.

Commented [6]: Btw @alissabrown@mortonarb.org all the figures are supposed to be at the end of the document so I've been moving things around

Table 1. Regression significance (p-value) and fit to data (R2) with comparisons between 'mean latitude' and 'distance to the range edge' between population-level allelic richness. Significant relationships are indicated in bold.

Statistic	Geographic	Population	Linear		Quadratic	
			p-value	R2	p-value	R2
Allelic Richness	Mean Latitude	All populations	0.02	0.18	0.24	0.21
		Without Wisconsin populations	0.01	0.26	6.10 x10⁻⁴	0.61
	Distance to Range Edge	All populations	0.04	0.15		
		Without Wisconsin populations	2.90x10 ⁻⁵	0.59		
Expected Heterozygosity	Mean Latitude	All populations	0.33	- 0.001	0.510	- 0.026
		Without Wisconsin populations	0.47	- 0.023	0.04	0.227
	Distance to Range Edge	All populations	0.24	0.021		
		Without Wisconsin populations	4.47x10 ⁻³	0.320		

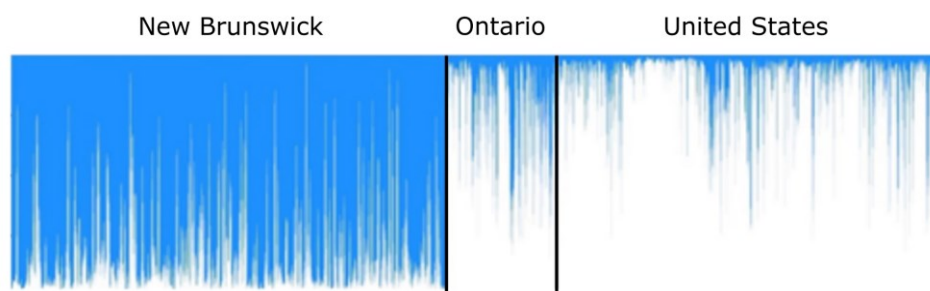


Figure 2. Structure diagram with 993 butternut individuals (a reduced set of non-highly-related individuals). showing the best supported clustering pattern, $\Delta K=2$, with New Brunswick populations separated from Ontario and U.S. populations.

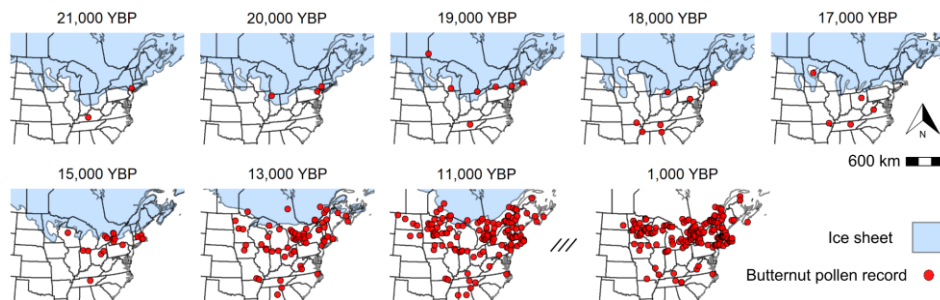


Figure 3. Locations of butternut fossil pollen occurrences (red circles) across eastern North America from 21 - 11 ka YBP, and at 1 ka YBP. The Laurentide ice sheet is indicated as a blue polygon. Map was created using Albers equal area projection.

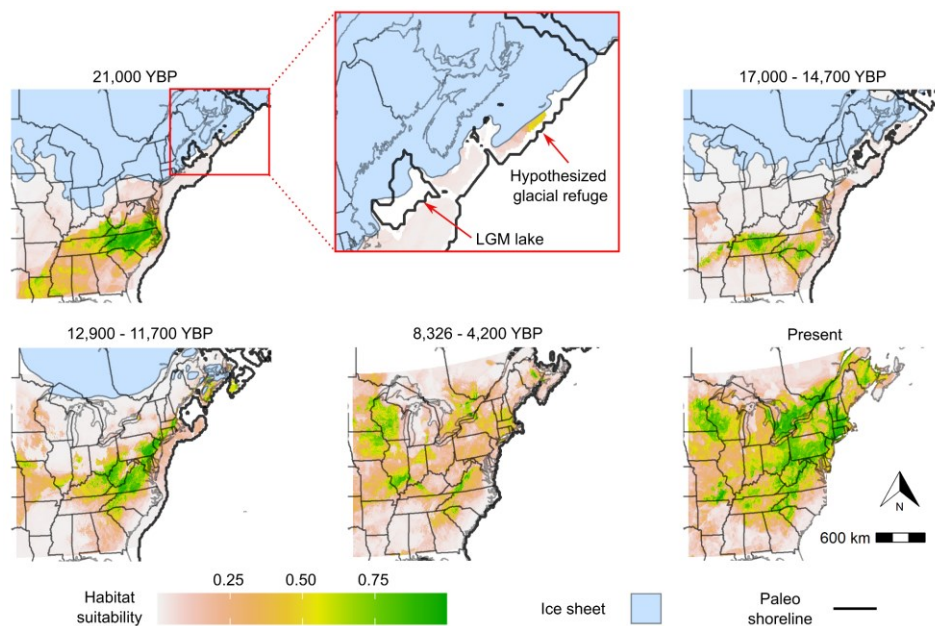


Figure 4. Habitat suitability and glacial coverage (blue polygon) for butternut at 22 ka YBP. Areas with (somewhat) suitable habitat are present in areas currently under ocean water, indicating the possibility that butternut could have been present there in a glacial refuge. Map was created using Albers equal area projection.