UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

History or demography? Determining the drivers of genetic variation in North American plants

López-Delgado, J.; Meirmans, P.G.

DOI
10.1111/mec.16230

Publication date
2022

Document Version
Final published version

Published in
Molecular Ecology

License
CC BY-NC-ND

Link to publication

Citation for published version (APA):
López-Delgado, J., & Meirmans, P. G. (2022). History or demography? Determining the drivers of genetic variation in North American plants. *Molecular Ecology*, *31*(7), 1951-1962. https://doi.org/10.1111/mec.16230

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)

Download date:09 Sept 2026

Check for updates

Received: 20 October 2020 | Revised: 5 October 2021 | Accepted: 6 October 2021

DOI: 10.1111/mec.16230

ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

History or demography? Determining the drivers of genetic variation in North American plants

Julia López-Delgado[1,2] | Patrick G. Meirmans[1]

[1]Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, Amsterdam, The Netherlands

[2]Institute of Biology, Leiden University, Leiden, The Netherlands

Correspondence
Julia López-Delgado, School of Biology, University of Leeds, Manton Building, Leeds, UK.
Email: bsjld@leeds.ac.uk

Abstract

Understanding the impact of historical and demographic processes on genetic variation is essential for devising conservation strategies and predicting responses to climate change. Recolonization after Pleistocene glaciations is expected to leave distinct genetic signatures, characterised by lower genetic diversity in previously glaciated regions. Populations' positions within species ranges also shape genetic variation, following the central-marginal paradigm dictating that peripheral populations are depauperate, sparse and isolated. However, the general applicability of these patterns and relative importance of historical and demographic factors remains unknown. Here, we analysed the distribution of genetic variation in 91 native species of North American plants by coupling microsatellite data and species distribution modelling. We tested the contributions of historical climatic shifts and the central-marginal hypothesis on genetic diversity and structure on the whole data set and across subsets based on taxonomic groups and growth forms. Decreased diversity was found with increased distance from potential glacial refugia, coinciding with the expected make-up of postglacially colonised localities. At the range periphery, lower genetic diversity, higher inbreeding levels and genetic differentiation were reported, following the assumptions of the central-marginal hypothesis. History and demography were found to have approximately equal importance in shaping genetic variation.

KEYWORDS
core-periphery, microsatellite, postglacial expansion, range limits, species distribution modelling

1 | INTRODUCTION

Disentangling the determinants shaping species' distributions and genetic diversity is key to understand and conserve biodiversity and ultimately predict responses to ongoing global change. The identification of the spatial distribution of genetic variation can elucidate the mechanisms of evolution and speciation, shed light on the processes maintaining geographical ranges, improve climate change forecasting, anticipate the spread of invasive species, and pinpoint conservation-priority populations (Guo, 2014; Hampe & Petit, 2005; Howes & Lougheed, 2008). Phylogeographic analyses can be employed to study the role of ecological factors and mechanistic processes, such as past climatic shifts and demographic fluctuations, on the genetic structure of populations. Yet, despite a large body of research on the population genetics of individual species, few studies have looked at general patterns of genetic structure and variation across multiple species' ranges at large spatial scales (Gaston, 2009).

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.
© 2021 The Authors. *Molecular Ecology* published by John Wiley & Sons Ltd.

1365294x, 2022, 7, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.16230 by Uva Universiteitsbibliotheek, Wiley Online Library on [23/02/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Historical climate-driven changes in species range limits are well known to still affect present-day genetic diversity (Durka, 1999; Hampe & Petit, 2005; Hewitt, 2000). The Pleistocene ice ages led to southwards range contractions, brought by the extinction of northern populations when temperatures plummeted, with multiple areas across continents serving as glacial refugia for species during the ice ages (Beatty & Provan, 2011; Hewitt, 2000). The end of the Pleistocene led to a subsequent northward expansion of many species in the wake of deglaciation (Hewitt, 1996). The historical climate changes over the past 20,000 years have led to shifting, fragmentation, and reconnection of species' ranges, which have resulted in a legacy of genetic consequences for contemporary populations (Alvarez et al., 2009; Comes & Kadereit, 1998; Taberlet et al., 1998). One major consequence is that previously glaciated areas are expected to show reduced genetic diversity as a result of sequential founder effects during the post-glacial recolonisation from refugia (Comes & Kadereit, 1998; Schonswetter et al., 2005; Taberlet et al., 1998). The effect of the ice ages was especially strong in temperate areas. In particular, the northern part of North America was covered with two immense ice sheets (Cordilleran and Laurentide; Beatty & Provan, 2010; Taberlet et al., 1998) at the Last Glacial Maximum (LGM). The changing conditions following the melting of the glacial ice make North America an ideal natural laboratory to study patterns of post-glacial colonisation and geographic variation (Hewitt, 2000; Pulgarín-R & Burg, 2012).

The distribution of genetic variation within a species is expected to also be shaped by demographic and evolutionary processes at range margins. Declining environmental suitability towards the periphery is predicted to result in decreasing population density (Kirkpatrick & Barton, 1997), thus reducing population size, gene flow and connectivity. Consequently, marginal populations tend to exhibit low genetic diversity and high genetic differentiation (Brussard, 1984; Pfeifer et al., 2009; Sagarin & Gaines, 2002). However, the broad applicability of this biogeographic paradigm – the central-marginal hypothesis– is debated since patterns of population genetic variation across large spatial scales are highly variable and usually species-specific (Sagarin & Gaines, 2002; Yakimowski & Eckert, 2007). For example, Eckert et al. (2008) criticised studies for not including a proper quantitative measure of centrality/peripherality or estimates of population sizes. Additionally, the central-marginal hypothesis assumes concordance between the geographic and environmental spaces, but this assumption might not always hold, since ecological marginality does not always imply spatial peripherality, and vice versa (Pironon et al., 2015; Soule, 1973).

There are relatively few studies that have tested both the central-marginal hypothesis and historical influences in a phylogeographical framework, or have attempted to distinguish historical effects on genetic diversity from patterns caused by contemporary geographical variation in population demography and dispersion (Eckert et al., 2008; Gaston, 2009). This is problematic since the patterns in genetic diversity resulting from these two processes can resemble each other. If populations at the northern margin show reduced genetic diversity, is this due to founder effects of postglacial recolonization or due to demographic effects related to the central-marginal hypothesis? Moreover, almost all evidence regarding these fields comes from studies on single species, which makes it difficult to draw any general conclusions.

Here, we use a novel phylogeographic framework to test the contributions of both historical climatic shifts and the central-marginal hypothesis on population genetic diversity across the ranges of 91 vascular plant species. We do this by coupling genetic data sourced from the literature and species distribution modelling to analyse the spatial structuring of population genetic variation across the North American continent (Figure 1). Species distribution models were hereby employed as macroecological tools, used to estimate population suitability, act as surrogates of abundance, and develop proxies for colonisation history and population demography. We perform a continental-scale analysis to identify concordant patterns of population genetic diversity on a large number of unrelated taxa, empirically overcoming the drawbacks commonly associated with demographic and colonisation history studies.

2 | MATERIALS AND METHODS

2.1 | Genetic data

We compiled a genetic database consisting of microsatellite data for 91 native diploid species of North American angiosperms and gymnosperms, with each taxon containing at least three sampling sites including a minimum of five individuals per location. We searched the literature for studies employing microsatellite markers on native vascular plants from North America, which resulted in genetic data published in 67 peer-reviewed studies (see Table S1). Two species were represented by two studies. We also obtained four unpublished data sets by direct communication with the authors. We were unable to find microsatellite data sets of Arctic species, which are expected to display the strongest genetic consequences from Pleistocene climate shifts given the longer postglacial colonisation routes (Hewitt, 2000). For all studies we tried to obtain the raw allelic data by downloading them from data repositories or by contacting the authors. In total, we obtained the raw data for 52 species. The included species display a wide range of ecological characteristics, abundances, range sizes and life history traits, thus aiming to serve as a representative sample of the continent's vascular flora. The varying range areas included aims to reduce potential biases in the emerging patterns (Supporting Information 1). The sampling and genotyping protocols vary across species, being detailed in the corresponding publications.

2.2 | Genetic summary statistics

We used the microsatellite data to calculate a set of four summary statistics per population, indicative of population genetic diversity and differentiation: the expected and observed heterozygosity (H_S

1365294x, 2022, 7, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.16230 by Uva Universiteitsbibliotheek, Wiley Online Library on [23/02/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

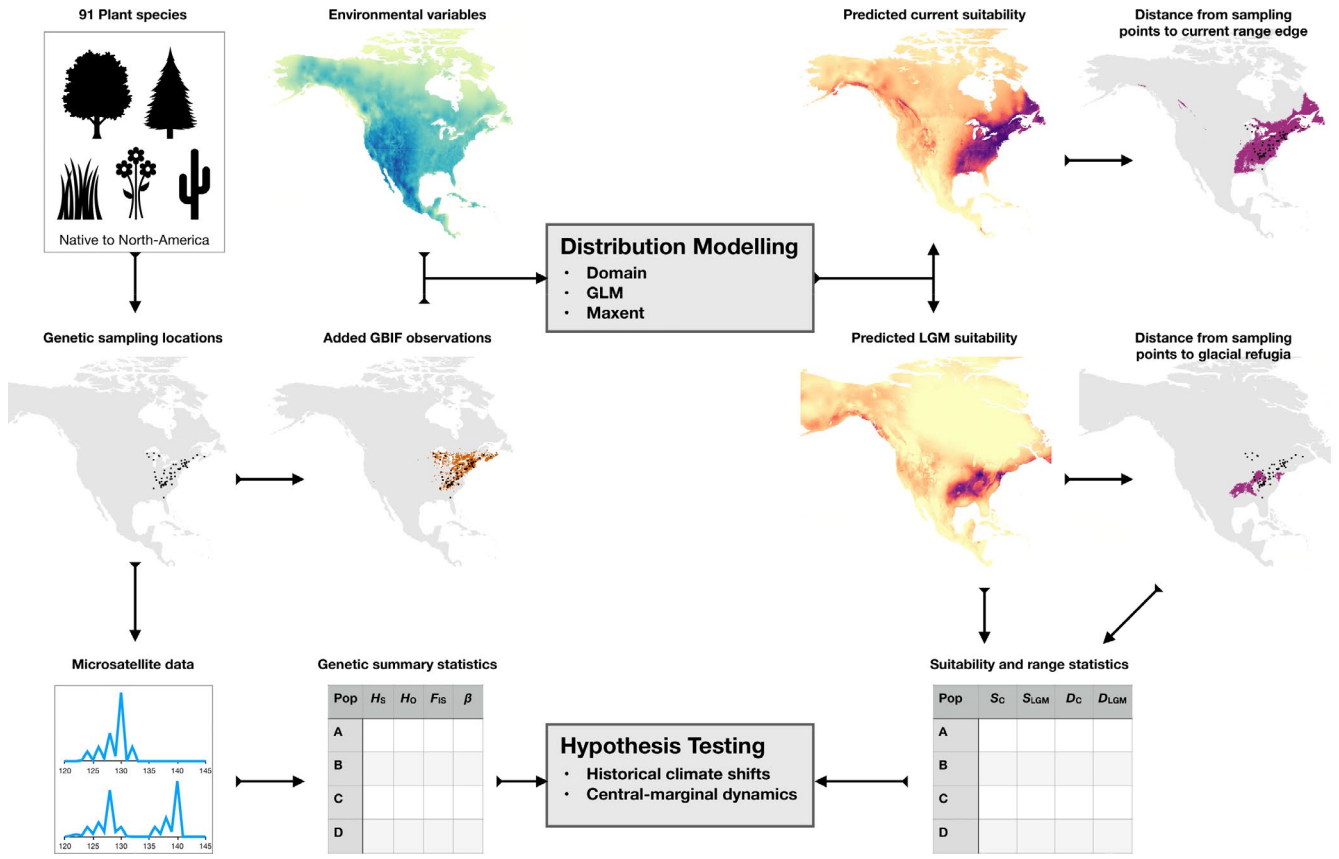

FIGURE 1 Overview of the methodological approach, which couples genetic data and species distribution modelling to test the contributions of historical climatic shifts and the central-marginal hypothesis on the spatial distribution of genetic variation. The distribution data shown and modelling output are for *Tsuga canadensis* (see Tables S1 and S2)

and H_O, respectively), the population inbreeding coefficient (F_{IS}), and β (Nei, 1987; Weir & Cockerham, 1984). The expected within-population diversity H_S is calculated only based on the allele frequencies within a population, whereas H_O considers the observed frequency of heterozygotes in the population. F_{IS} can be defined by comparing the above-mentioned heterozygosity measures: $(H_S-H_O)/H_S$, with high levels indicating an excess of homozygous genotypes compared to Hardy-Weinberg expectations. β is a population-specific estimator of the genetic differentiation statistic F_{ST} (Nei, 1987; Weir & Cockerham, 1984). The summary statistics calculations were performed using the function basic.stats() from the HIERFSTAT R package (Goudet, 2005), with every sampling location being treated as a separate population. Clone correction was performed for species with apomictic or vegetative reproduction, by removing all duplicated genotypes to ensure each genotype is represented by a single individual (one ramet per genet) per population. For the species for which the allelic data was not available, we used the published estimates of the summary statistics from the original studies, which were mostly limited to H_S, H_O and F_{IS}.

2.3 | Species distribution modelling

Species distribution models were built for all 91 species to evaluate habitat suitability and develop proxies for colonisation history and population demography. The georeferenced occurrences from the sites included in the genetic data set were complemented by downloading species records from the Global Biodiversity Information Facility (GBIF) data portal (GBIF dois in Table S1). Occurrences were restricted to North America, and duplicate locations were removed to avoid pseudoreplication. The records were manually cleaned by removing any suspicious data points in uninhabitable locations for each species. All species were represented by at least 10 unique presence records, which is the recommended minimum number required to calibrate a species distribution model (SDM; Proosdij et al., 2016). Species occurrences totalled 97,074 unique records, ranging from 10 to 9594 occurrences per species.

To model the species distributions, the 19 bioclimatic predictors of the WORLDCLIM v.1.4 data set (http://www.worldclim.org/) were obtained for past and present scenarios at a 2.5 arc-minute spatial resolution (Hijmans & Elith, 2012). The used paleoclimate data resulted from simulations from a global climate model (GCM) for the LGM (approximately 22,000 years ago), as estimated by the MIROC-ESM climate model (Watanabe et al., 2011). Current conditions represent interpolations of observed climatic data from 1960–1990. The GTOPO30, a global 30 arc-second digital elevation model (DEM) was retrieved from the USGS EROS archive (https://www.usgs.gov/centers/eros). The GTOPO30 was aggregated by a cell factor of five to achieve a 2.5 arc-minute resolution, resulting in a

1365294x, 2022, 7, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.16230 by Uva Universiteitsbibliotheek, Wiley Online Library on [23/02/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

DEM for the present scenario. In order to obtain a DEM for the LGM, when the sea level was 120–135 m lower than at present (Clark & Mix, 2002), the GEBCO_2019 Grid, a global 15 arc-second bathymetry DEM (GEBCO, 2019) was downloaded. The GEBCO_2019 Grid was rasterised and aggregated by a cell factor of 10 to achieve a 2.5 arc-minute resolution, followed by clipping it using an LGM bioclimatic layer as a mask. Employing altitude as a variable for modelling is advised against when the SDM aims to project to past climatic conditions (Raes & Aguirre-Gutiérrez, 2018). Thus, slope and aspect (the direction of the slope) were derived from the DEMs and included as variables instead. All bioclimatic and topographic variables were clipped to the extent of the North American continent, from the southernmost point of Panama to the northernmost point of Canada, excluding Greenland. Data layer manipulations were performed with ArcGIS (ESRI).

To avoid collinearity amongst environmental predictors, which can result in overfitting (Graham, 2003; Peterson et al., 2007), the number of predictor variables was reduced by removing highly correlated parameters, as given by a Spearman's rank correlation test (r_s > 0.7). When deciding which of two correlated variables to retain, we aimed to closely capture the key determinants of physiological processes limiting distributions of plants, considering the ample range in ecological preferences displayed by the species included. The retained variables for modelling were mean diurnal temperature range (Bio2), temperature annual range (Bio7), mean temperature of wettest quarter (Bio8), annual precipitation (Bio12), precipitation seasonality (Bio15), aspect, and slope.

Species distribution models were built for each species under present conditions and then projected onto the LGM conditions, employing the sdm R library (Naimi & Araújo, 2016). Three modelling methods were implemented: Domain (Carpenter et al., 1993), Generalised Linear Model (GLM; McCullagh & Nelder, 1989), and Maximum Entropy (MaxEnt; Phillips et al., 2006), each belonging to one of the three main types of modelling algorithms, being "profile", "regression", and "machine learning", respectively. Cross-validation was performed to validate each model, with 70% of the data being employed for calibration and 30% for evaluation, with 10 bootstrap replications being run per method. SDM accuracy was evaluated using the area under the curve (AUC) of the receiver operating characteristic (ROC) plot (Hanley & McNeil, 1982), a threshold-independent measure that is relatively insensitive to prevalence (McPherson et al., 2004). AUC values range from zero to one; values close to one indicate maximum fit, whereas values under 0.5 (half of the area under the ROC curve) indicate the model prediction is no better than a random prediction. All models performed better than random expectation, with the lowest AUC value being 0.6. Ensembles of model forecasts were then fitted by combining the three modelling techniques. By employing a consensus, errors (sensitivity to data, lack of absence data, errors in environmental variables) tend to cancel each other out in ensembles, thus producing a more robust and conservative solution (Araújo & New, 2007; Diniz-Filho et al., 2009). The ensembles were built using weighted averaging based on the AUC statistic to cope with model variability and to improve the reliability of model predictions. The R script used for the species distribution modelling is available in the supplement.

2.4 | Ecological data

We used the output of the SDM to calculate for every population in each species four measures that quantify ecological suitability, colonisation history, and population demography: (1) habitat suitability under the current conditions, (2) habitat suitability during the LGM, (3) distance to range edge under current conditions, and (4) distance to potential glacial refugium. The habitat suitability was taken directly from the ensemble forecasting produced as output of the different SDMs (Anderson & Martínez-Meyer, 2004; Diniz-Filho et al., 2009). Thus, the ecological suitability *S* of each population under the two modelled time frames was defined as the average value of occurrence provided by each model in the ensemble. For estimating population centrality/peripherality, an innovative quantitative measure was developed. For this, the suitability data was transformed into presence/absence data by setting a threshold, using the *Max SSS* approach of Liu et al. (2005), which maximises the sum of model sensitivity and specificity. Out of 13 threshold selection methods, *Max SSS* was found to perform best when only presence data is available (Liu et al., 2013). The distance *D* to the species' range edge under present bioclimatic conditions was then computed by calculating the closest distance between each population and the contour of the generated binary presence map. Distances to range edge had negative values if populations were found outside the predicted species range. Finally, distance from a potential glacial refugium was calculated as the closest distance from each population to a suitable area under LGM after creating a binary presence map based on the ensemble prediction for the LGM. The development of this proxy enables the identification of previously unknown glacial refugia, particularly relevant for species with northern refugia that a latitudinal proxy would not account for. Distance to range edge computations were performed in ArcGIS version 10.2 (ESRI) and distance to suitable areas under different scenarios were calculated using the geosphere R package (Hijmans et al., 2019), both employing a geodesic method.

2.5 | Hypothesis testing

Linear mixed effect models were used to test the relative contribution of distance to edge range, distance to a suitable area under the LGM, and present and LGM suitability in shaping the four population genetic parameters. Additional linear mixed effects models were used to investigate the relationship between the distance to edge range and distance to a suitable area under the LGM with present and LGM suitability. The species name was used as a nested random effect in the analyses to account for phylogenetic nonindependence. In addition to analysing the data set as a whole, we also performed the analysis on subsets of the data, based on either the two main

1365294x, 2022, 7, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.16230 by Uva Universiteitsbibliotheek, Wiley Online Library on [23/02/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

taxonomic groups (Gymnosperms and Angiosperms) or the growth forms of the species (trees, shrubs, and herbs).

The linear mixed effects models were performed using the lme4 package (Bates et al., 2013) in R version 3.5. Model selection was performed in terms of parsimony (based on AIC) and variance explained. The variance explained was calculated using the methods proposed by Nakagawa and Schielzeth (2013) as implemented in the MuMIn package, which provides the total variance explained by fixed and random effects and allows the calculation of variance explained by each fixed effect (Barton, 2011). *p*-values were calculated using the Satterthwaite (1946) approximations, using the lmerTest R package (Kuznetsova et al.,2017), standard Bonferroni correction was then applied (Bonferroni, 1936). In addition, linear models were run in order to add regression lines to the plots, which aid the visual interpretation of the results of the linear mixed effect models.

3 | RESULTS

The database we compiled includes 1406 populations across 91 vascular plant species, spanning the whole North American continent, except for the arctic regions. These included 885 angiosperms and 521 gymnosperm populations; with respect to life history, the populations could be classified into 512 herbs, 78 shrubs and 816 trees. For 829 populations, the genetic summary statistics were computed using the raw allelic microsatellite data, whereas for the remaining populations the available published estimates were taken (Table S2). Overall minimum and maximum estimates were 0.00 and 0.92 for the expected heterozygosity (H_S), 0.00 and 1.00 for the observed heterozygosity (H_O), −1.00 and 1.00 for the inbreeding coefficient (F_{IS}), and −0.27 and 1.00 for the population differentiation statistic β.

The species distribution modelling performed on all 91 species had high predictive power for the relationship between the species' distribution and the bioclimatic variables (Table S1). Of the 2730 species distribution models (three SDMs × 10 bootstrap replications × 91 species), 88% had AUC values above 0.85. AUC values of the MaxEnt models were generally higher than those of GLM models, which in turn were higher than those of Domain, pointing towards a higher predictive power of MaxEnt over GLM and Domain for most species.

The analyses combining the genetic summary statistics for all species with the output of the SDM revealed a legacy of past glaciation in the genetic data. The genetic diversity –as measured with H_S– significantly decreased with increasing distance to suitable areas during the LGM (Figure 2). Furthermore, there was a significant positive relationship between the suitability of populations under LGM conditions and both H_S and H_O (Figure 3). No significant patterns were found for F_{IS} and β, (Figures 2 and 3) even though F_{IS} showed a slight decrease with increasing distance to suitable areas under the LGM. Additionally, populations displayed significantly lower ecological suitability when further away from suitable areas under the LGM (Figure S1).

When the full data set was analysed, it also showed clear support for the central-marginal hypothesis. Marginal populations indeed had a lower ecological suitability as evidenced by a significant positive relationship between distance to range edge and suitability under the current climatic conditions (Figure S2). This was accompanied by a significant increase in genetic diversity (H_S)

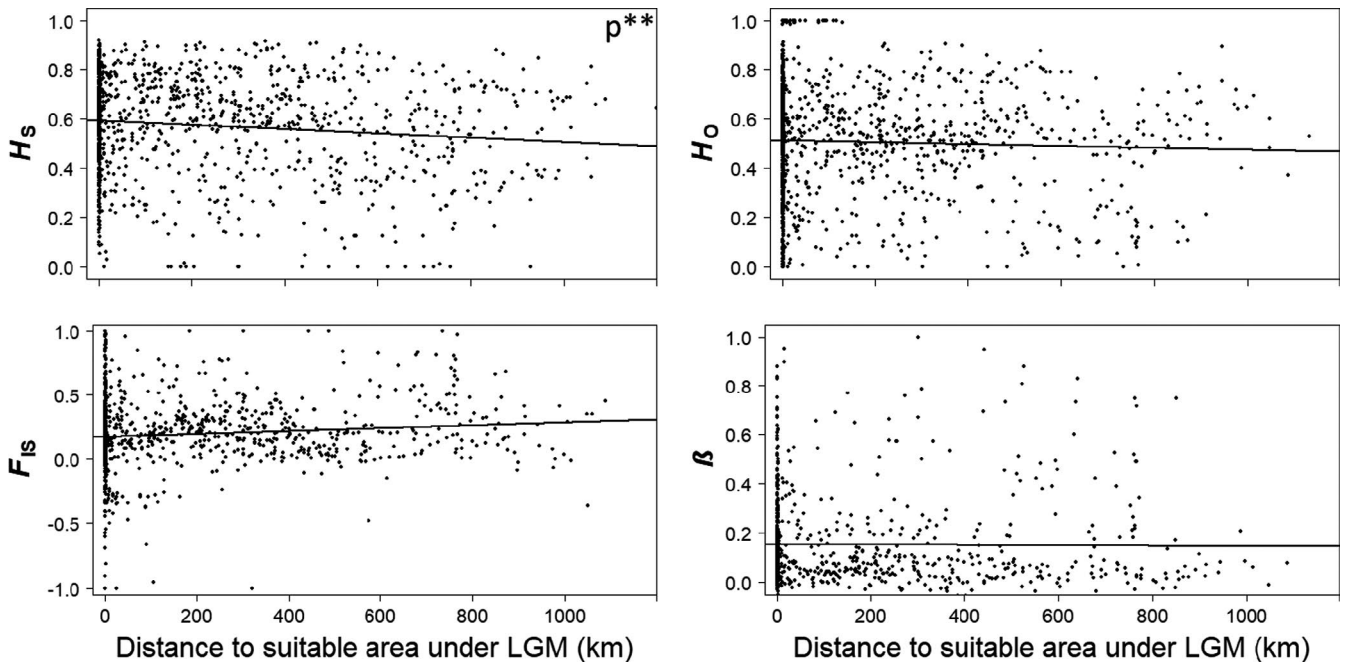

FIGURE 2 Relationship between the distance from populations to a suitable area under LGM conditions and the four genetic parameters. Significance was tested using linear mixed effects models (Table 1). Regression lines of were added to the figures to aid visual inspection, with the linear model of H_e and F_{IS} showing $p < .001$

1365294x, 2022, 7, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.16230 by Uva Universiteitsbibliotheek, Wiley Online Library on [23/02/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

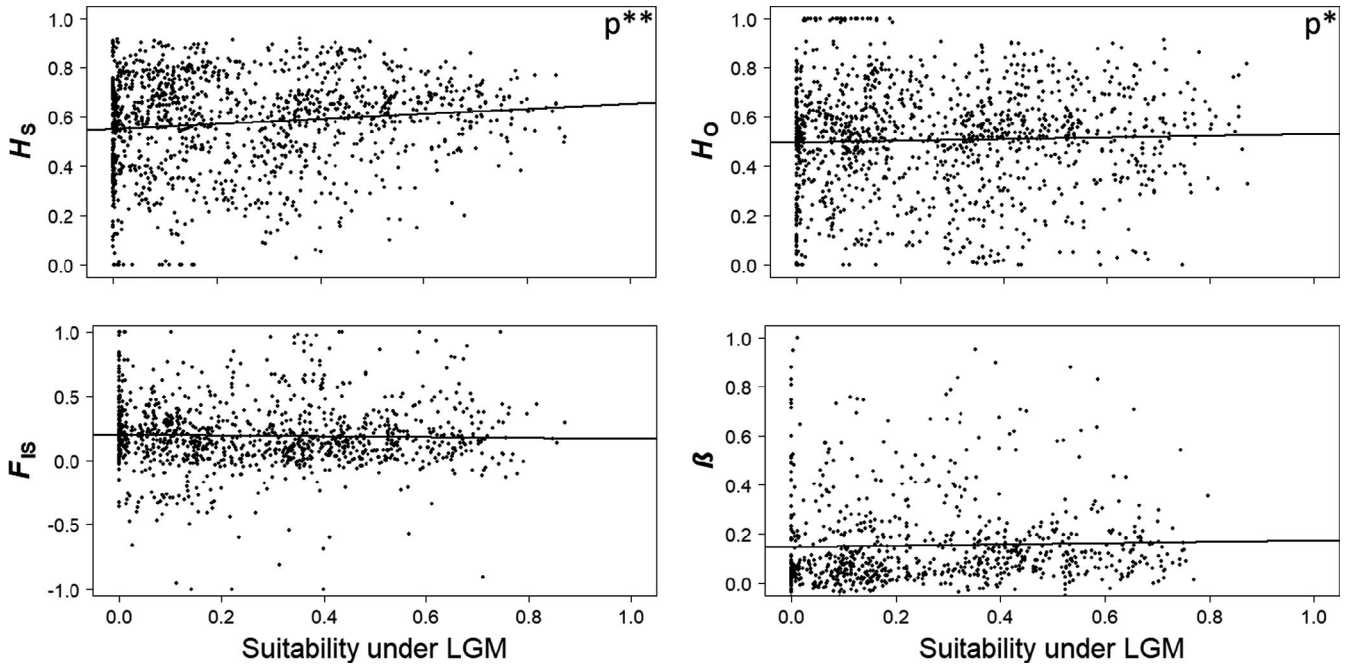

FIGURE 3 Relationship between the suitability of populations under LGM conditions and the four genetic parameters. Significance was tested using linear mixed effects models (Table 1). Regression lines were added to the figures to aid visual inspection, with the linear model of H_e showing $p < .001$

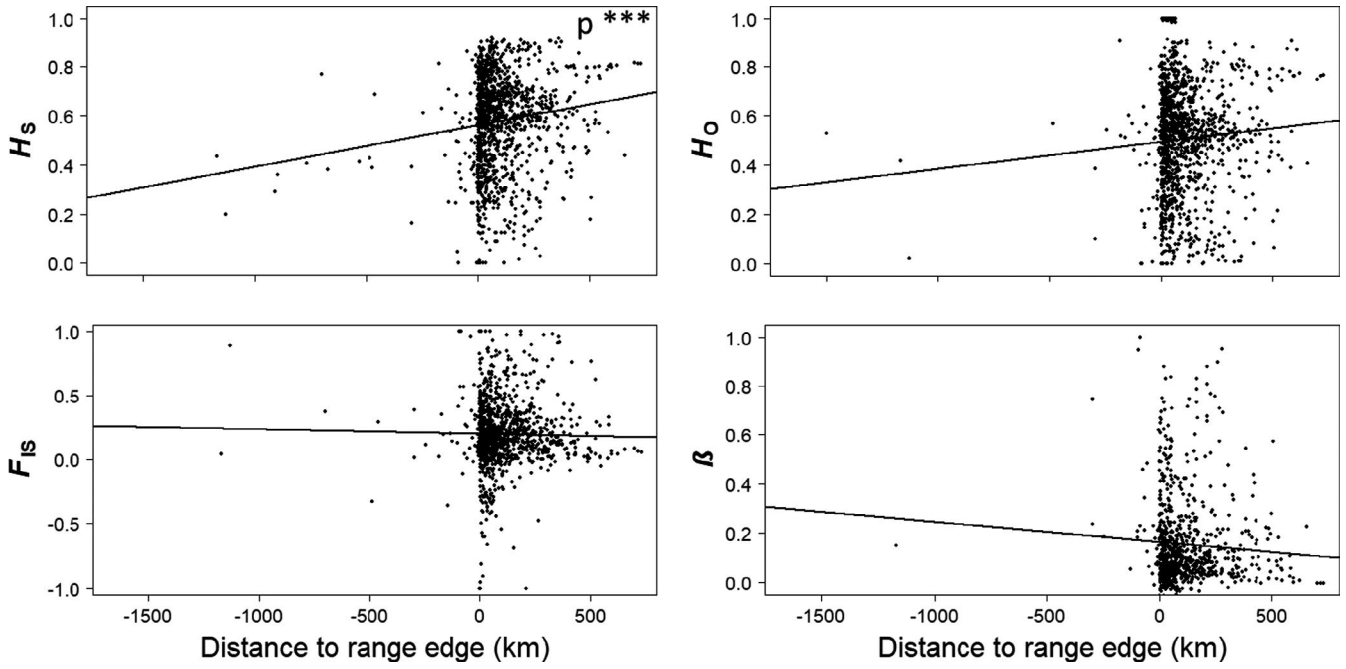

FIGURE 4 Relationship between the distance from populations to the range edge and the four genetic parameters. Significance was tested using linear mixed effects models (Table 1). Regression lines were added to the figures to aid visual inspection, with the linear models of H_s and H_e showing $p < .001$

as distance to the range edge incremented (Figure 4). Furthermore, the value of H_S showed a significant negative correlation with the present suitability of the populations (Figure 5). No significant patterns were observed for H_O, β and F_{IS} (Figure 5). While 13 species displayed negative distance values, these did not bias the observed patterns as the tests remained significant when negative values were removed.

The standardised variance explained in the global models of the genetic parameters totalled 22.1% by the distance to suitable area under the LGM, 27.5% by LGM suitability, 22.6% by the distance to

1365294x, 2022, 7, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.16230 by Uva Universiteitsbibliotheek, Wiley Online Library on [23/02/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

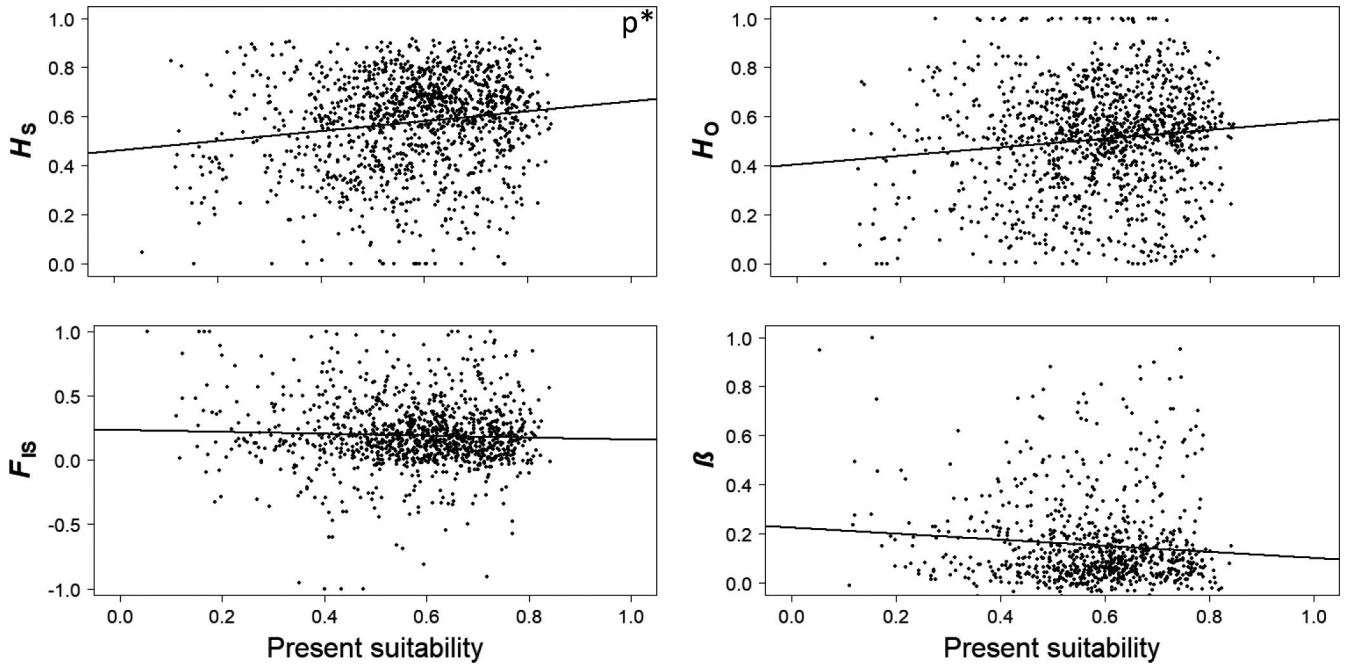

FIGURE 5 Relationship between population present ecological suitability and the four genetic parameters. Significance was tested using linear mixed effects models (Table 1). Regression lines were added to the figures to aid visual inspection, with the linear models of H_s, H_e and β showing $p < .01$

edge range, and 19.1% by present suitability (Table 1). For the expected heterozygosity, the minimal model included both effects of the ice ages and effects of the central-marginal hypothesis; this minimal model was composed of the distance to edge range, the distance to a suitable area under the LGM, and the present suitability, with fixed and random effects explaining a total variance of 79.4%. The observed heterozygosity's minimal model was given by the present suitability, accounting for 88.2% of the data variation. Model selection was not performed for the inbreeding coefficient, as no variables were deemed significant. Finally, the minimal model for genetic differentiation was given by present suitability, explaining 63.4% of the variance. For all models, the reported values for the percentage of variance explained also includes the nested random effect of species name, showing that there are strong differences between species in genetic diversity and differentiation.

When the data was divided into subsets based on taxonomic groups or life forms, the results were more complex but still showed support for both the influence of the ice ages and central-marginal processes (Table S3). When the gymnosperms were analysed separately, both H_S and H_O decreased with increasing distance to suitable area during the LGM, and increased with habitat suitability during the LGM, indicating influence of the ice ages. The present range edge and suitability were not included in the minimum model according to AIC, though the distance to the present range edge did show a significant P-value. For angiosperms, H_S was similarly correlated with distance to suitable area and habitat suitability during the LGM, but H_O was influenced more by the distance to the present range edge and the present suitability, showing the influence of central-marginal processes. For the growth forms, trees mostly showed correlations with the present range edge (H_S and β), while herbs showed correlations only with the present suitability (H_S, H_O, and β). Shrubs did not show any significant correlations, probably due to the very small number of shrub species in the data set.

4 | DISCUSSION

Here, we used a phylogeographical framework combining species distribution modelling and population genetics to assess the influences of Pleistocene climate shifts and central-marginal processes on contemporary genetic patterns. Our analysis of genetic data of 91 vascular plant species clearly shows the effects of both Pleistocene climatic events and central-marginal processes in spatially structuring genetic variation in North America. Of the more than 1400 included populations, those located at species' range margins and most distant from glacial refugia were found to have significantly reduced genetic diversity. Our multispecies approach grants increased statistical power compared to studies exploring these processes at the single-species level, allowing us to draw general conclusions at a large geographical scale.

The effects of the Pleistocene ice ages are evident in the decrease in genetic diversity, as measured using H_S, with increased distance from suitable areas under the LGM, i.e. glacial refugia. This reduction in diversity is similar to what is observed in simulation models on the genetic make-up of postglacially colonised regions (Hewitt, 1999, 2000; Pironon et al., 2015). The observed genetic patterns are most likely due to the larger population sizes and more stable population dynamics at refugial localities, as well as repeated

1365294x, 2022, 7, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.16230 by Uva Universiteitsbibliotheek, Wiley Online Library on [23/02/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

bottlenecks at the advancing (generally northern) range edge during postglacial colonisation (Comes & Kadereit, 1998; Hewitt, 1996). Our results therefore indicate that the shifting and fragmentation of species' geographical ranges in the past 20,000 years has played an important role in moulding genetic variation across large taxonomic and spatial scales (King & Ferris, 1998; Parks et al., 1994; Sewell et al., 1996; Soltis et al., 1997). Nonetheless, we must be cautious when interpreting diversity patterns, as patterns do not solely emerge from simple models of postglacial colonisation (Petit et al., 2003), and the relationship with the LGM variables was no longer significant when the data was divided by growth form.

The effects of central-marginal dynamics are evident in the negative correlation of genetic diversity with both the distance from the range edge and ecological suitability. These effects were visible both in the full data set and for the angiosperms, but interestingly not for the gymnosperms. The central-marginal hypothesis predicts range limits arise because peripheral populations occur in marginal habitats and cannot adapt to conditions beyond the range edges (Haag & Ebert, 2004; Kawecki, 2008). This demographic instability can induce low effective population sizes and frequent bottlenecks, leading to reduced genetic diversity in margins, such as observed in our data (Micheletti & Storfer, 2015; Sagarin & Gaines, 2002; Sexton et al., 2009). Although we do not have population size estimates, we assessed the ecological suitability of populations within their ranges. Diniz-Filho et al. (2009) propose employing this variable as a macroecological surrogate of abundance, which ensures the logical application of the central-marginal model to complex spatial abundance patterns as it considers species' ecological, and not geographical, ranges. The applicability of this approach in our data was supported by a clear central-peripheral pattern that was observed in the ecological suitability for the analysed populations as this was negatively correlated with the distance to the range margins.

Though we found strong patterns in genetic diversity (H_S and to some extent H_O) related to the proxies, no significant associations were found for the inbreeding coefficient F_{IS}; the minimum model in our linear mixed model analysis of F_{IS} did not include any of the explanatory variables. Theoretical models have suggested that self-fertilisation may be favoured at the range margins, especially at postglacially expanding range fronts, given the possible advantages of reproductive assurance or local adaptation (Arnaud-Haond et al., 2006; Hargreaves & Eckert, 2014). Indeed, several studies on single species have shown selfing to be more frequent in marginal populations, due to low population sizes and environmental stress at range margins, leading to elevated inbreeding levels (Aldrich & Hamrick, 1998; Barrett, 2002; Schoen et al., 1996). Our results across a taxonomically wide range of species suggests that this may not be a general phenomenon.

Similar to F_{IS}, no significant relationships were found in the analysis of the full data set for the genetic differentiation (β, a single population estimate of F_{ST}) of populations with the rest of the metapopulation, though the minimum model according to AIC did include the present suitability as an explanatory variable. This lack of a significant relationship contrasts with numerous theoretical models and empirical studies reporting a decrease in genetic differentiation

TABLE 1 Output of linear mixed model analysis: degrees of freedom (d.f.), percentages of variance explained (variance) and *p*-values are given for the global models of each genetic summary statistic including the distance to suitable area under the LGM, the suitability under the LGM, the distance to range edge, and present suitability. The genetic statistics are the expected and observed heterozygosity (H_S and H_O, respectively), the population inbreeding coefficient (F_{IS}), and beta (β). The total variance explained by the global and minimal models of each genetic summary statistic are reported in the bottom rows. The total variance explained by the historical and demographic proxies is given in the rightmost column, calculated as the sum of the standardised variances explained for the four genetic summary statistics. For each summary statistic, the variables included in the minimum model, according to AIC, are given in bold; variables excluded from the minimum model are given in black

		H_S	H_O	F_{IS}	β	Total variance of parameters (%)
Distance to suitable area under LGM	d.f.	**1310**	1198	1110	822	22.1
	Variance (%)	**18**	1	1.1	2	
	p-value	**0.0026**	0.3690	1.0000	1.0000	
Suitability under LGM	d.f.	1287	1185	1090	817	27.5
	Variance (%)	10.8	2.2	0	14.5	
	p-value	0.0030	0.0464	0.8088	0.3846	
Distance to range edge	d.f.	**1323**	1217	1126	814	22.6
	Variance (%)	**14.8**	2.1	1.9	3.8	
	p-value	**0.0007**	1.0000	1.0000	1.0000	
Present suitability	d.f.	**1327**	**1209**	1114	**819**	19.1
	Variance (%)	**1.3**	**4.5**	3.6	**9.7**	
	p-value	**0.0149**	**0.3472**	1.0000	**0.1691**	
Total variance of global model (%)		**79.4**	**88.2**	85.3	**62.2**	
Total variance of minimal model (%)		**79.4**	**88.2**	–	**63.4**	

1365294x, 2022, 7, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.16230 by Uva Universiteitsbibliotheek, Wiley Online Library on [23/02/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

along postglacial expansion routes (Austerlitz et al., 1997; Excoffier et al., 2009). On the other hand, marginal populations are expected to be more isolated from the central ones, so it is possible that the effects of central-marginal processes and the Pleistocene recolonisation are counteracting each other. It is also possible that major population structure lies elsewhere, such as among different genetic clusters associated with recolonisation from multiple isolated refugia (Ursenbacher et al., 2015).

The influence of the ice ages and the central-marginal effects varied depending on the life history and taxonomy of the species, as shown by the analyses of the data subsets. Long-lived perennials such as trees are known to generally show higher diversity and weaker population structure than short-lived perennials and annuals (Hamrick & Godt, 1996; Nybom, 2004), generally indicating larger population sizes and higher connectivity. This is consistent with our results, as a significant relationship between β and the present suitability was found for herbs, but not for shrubs or trees. Additionally, this can also explain why this same correlation was not significant for gymnosperms –which are mainly trees– but did show significance for angiosperms –which also include herbs. Similarly, the lack of support for central-marginal dynamics in gymnosperms may be an effect of the higher population connectivity of trees, which may quickly erode any effects of smaller population sizes at the distribution margins. However, it is also important to realise that the subsetting of the data has led to a decrease in statistical power, which may have affected the inferences, especially for the shrubs.

We employed a species distribution modelling approach to develop effective proxies for colonisation history and population demography, and as such the followed approach does have several limitations and makes several assumptions with regard to the ecology of the study species. One assumption is that the species distribution models only base their predictions on the included environmental variables. We used the same set of variables for all species in order to allow for better comparisons among species, but not all variables may be equally applicable to all species. A general concern around species distribution models is that they do not account for biotic interactions, potential for rapid adaptations or time lag (Godsoe & Harmon, 2012; Hällfors et al., 2015). It is also important to acknowledge that both the SDM and the modelled climate data for the LGM represent extreme extrapolations, and therefore can only be taken as a rough approximation of the situation during the ice ages. For a better overview of the locations of glacial refugia, SDM results should be combined with palynological data, though this has also been proven to be challenging due to the coarseness of the palynological record (Birks, 2019). In addition, the approach assumes that all species have reached their equilibrium distribution. However, it is possible that some species have not yet filled their potential ranges in response to rapid postglacial warming, resulting in incomplete SDMs and inaccurate estimations of the distances to glacial refugia. This would result in an underestimation of the emerging genetic patterns, which would be expected to grow stronger during future recolonization dynamics.

Despite the inherent assumptions in the distribution modelling, we believe that our approach is a major step up from many previous studies that have worked with the simple assumption that latitude is an adequate surrogate for either recolonisation history or central-marginal processes (Eckert et al., 2008). Instead, our SDM-based approach allowed us to test and separate the effects of both these phenomena. Our results also show that the distribution models, which had high power for predicting the present distribution, can be used as a valuable tool to infer causation of extant genetic variation patterns across a large scale. Notably, the main purpose of our modelling was not to get an accurate prediction of actual habitats, but rather to get an estimate of the overall range of the species, for which these types of models are widely used.

Both Pleistocene historical and demographical processes appear to have equal and nonexclusive importance in shaping genetic variation. Most genetic summary statistics displayed the expected responses to these proxies; however, only a quarter of the comparisons were significant and the variation explained by the fixed effect factors was relatively low. Most variation in the summary statistics comes from within-species determinants, as a large proportion of the variance was explained by the species random effects. This was not unexpected given the widely different demographic histories of the species studied, and it highlights the role of potential additional variables acting at an intraspecific level. Numerous studies have emphasized how the genetics of populations is shaped by ecological factors, topography, anthropogenic factors –such as land use and habitat fragmentation–, biotic interactions, and life-history traits –including the aforementioned breeding strategy and dispersal capacities (Alvarez et al., 2009; Kuittinen et al., 1997; Meirmans et al., 2011; Yeaman & Jarvis, 2006). Additionally, each study included in our meta-analysis used different genotyping and sampling strategies and covered varying proportions of the range, as well as employing different microsatellite markers, characterised by their own mutation rates. All combined, these factors have introduced noise into our analyses and potentially blurred the overall pattern, probably resulting in an underestimation of the emerging patterns. Despite the large influence of this random factor on our analysis, we were able to detect the effects of two vital processes on species ranges and their intraspecific genetic variation.

Bridging the effects of historical range shifts and contemporary demographic processes on species' genetic constitutions over large scales has many implications to evolutionary biology. The potential for adaptation to new environments largely depends on the amount of genetic variation available, commonly sourced from standing genetic variation or arising through gene flow among populations. Central populations exhibit higher diversity and increased connectivity that can favour gene flow and adaptive introgression, conferring them an advantage towards changing conditions. Understanding this interplay between environmental conditions, genetic variation, connectivity and gene flow is crucial to address conservation needs in the face of ongoing global change. This study provides the foundation for further phylogeographical analyses to continue integrating the processes dictating genetic variation across species ranges.

1365294x, 2022, 7, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.16230 by Uva Universiteitsbibliotheek, Wiley Online Library on [23/02/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

ACKNOWLEDGEMENTS

We thank the authors who shared the microsatellite raw data, as well as Maarten van der Sande and Yorick Coolen for assisting in the compilation of the database. JLD also thanks Niels Raes and Emiel van Loon for advice on building the SDMs. Three anonymous reviewers provided useful comments on the manuscript.

CONFLICT OF INTERESTS

The authors declare no competing interests.

AUTHOR CONTRIBUTIONS

P.G.M. and J.L.-D. designed the study and compiled the database. J.L.-D. built the species distribution models and performed the data analysis, under supervision of P.G.M. J.L.-D. wrote the manuscript, with input from P.G.M.

DATA AVAILABILITY STATEMENT

The genetic data that support the findings of this study are openly available in the 67 peer-reviewed studies cited in Table S1.

ORCID

Julia López-Delgado https://orcid.org/0000-0003-0036-4579
Patrick G. Meirmans https://orcid.org/0000-0002-6395-8107

REFERENCES

Aldrich, P. R., & Hamrick, J. L. (1998). Reproductive dominance of pasture trees in a fragmented tropical forest mosaic. *Science*, *281*(5373), 103–105. https://doi.org/10.1126/science.281.5373.103

Alvarez, N., Thiel-Egenter, C., Tribsch, A., Holderegger, R., Manel, S., & Schönswetter, P. (2009). History or ecology? Substrate type as a major driver of spatial genetic structure in Alpine plants. *Ecology Letters*, *12*(7), 632–640. https://doi.org/10.1111/j.1461-0248.2009.01312.x

Anderson, R. P., & Martínez-Meyer, E. (2004). Modeling species' geographic distributions for preliminary conservation assessments: An implementation with the spiny pocket mice (*Heteromys*) of Ecuador. *Biological Conservation*, *116*(2), 167–179. https://doi.org/10.1016/S0006-3207(03)00187-3

Araújo, M. B., & New, M. (2007). Ensemble forecasting of species distributions. *Trends in Ecology & Evolution*, *22*(1), 42–47. https://doi.org/10.1016/j.tree.2006.09.010

Arnaud-haond, S., Teixeira, S., Massa, S. I., Billot, C., Saenger, P., Coupland, G., Duarte, C. M., & Serrão, E. A. (2006). Genetic structure at range edge: Low diversity and high inbreeding in Southeast Asian mangrove (*Avicennia marina*) populations. *Molecular Ecology*, *15*(12), 3515–3525. https://doi.org/10.1111/j.1365-294X.2006.02997.x

Austerlitz, F., Jung-Muller, B., Godelle, B., & Gouyon, P.-H. (1997). Evolution of coalescence times, genetic diversity and structure during colonization. *Theoretical Population Biology*, *51*(2), 148–164. https://doi.org/10.1006/tpbi.1997.1302

Barrett, S. C. H. (2002). The evolution of plant sexual diversity. *Nature Reviews Genetics*, *3*(4), 274. https://doi.org/10.1038/nrg776

Barton, K. (2011). Model selection and model averaging based on information criteria (AICc and alike). (Version 1.9.5). Retrieved from https://CRAN.R-project.org/package=MuMIn

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., Fox, J. (2013). lme4: Linear Mixed-Effects Models using "Eigen" and S4 (Version 1.1-21). Retrieved from https://CRAN.R-project.org/package=lme4

Beatty, G. E., & Provan, J. (2010). Refugial persistence and postglacial recolonization of North America by the cold-tolerant herbaceous plant *Orthilia secunda*. *Molecular Ecology*, *19*(22), 5009–5021. https://doi.org/10.1111/j.1365-294X.2010.04859.x

Beatty, G. E., & Provan, J. (2011). Phylogeographic analysis of North American populations of the parasitic herbaceous plant *Monotropa hypopitys* L. reveals a complex history of range expansion from multiple late glacial refugia. *Journal of Biogeography*, *38*(8), 1585–1599. https://doi.org/10.1111/j.1365-2699.2011.02513.x

Birks, H. J. B. (2019). Contributions of Quaternary botany to modern ecology and biogeography. *Plant Ecology & Diversity*, *12*(3–4), 189–385. https://doi.org/10.1080/17550874.2019.1646831

Bonferroni, C. (1936). Teoria statistica delle classi e calcolo delle probabilita. *Pubblicazioni Del R Istituto Superiore Di Scienze Economiche E Commericiali Di Firenze*, *8*, 3–62.

Brussard, P. F. (1984). Geographic patterns and environmental gradients: the central-marginal model in *Drosophila* revisited. *Annual Review of Ecology and Systematics*, *15*(1), 25–64. https://doi.org/10.1146/annurev.es.15.110184.000325

Carpenter, G., Gillison, A. N., & Winter, J. (1993). DOMAIN: A flexible modelling procedure for mapping potential distributions of plants and animals. *Biodiversity & Conservation*, *2*(6), 667–680. https://doi.org/10.1007/BF00051966

Clark, P. U., & Mix, A. C. (2002). Ice sheets and sea level of the last glacial maximum. *Quaternary Science Reviews*, *21*(1), 1–7. https://doi.org/10.1016/S0277-3791(01)00118-4

Comes, H. P., & Kadereit, J. W. (1998). The effect of quaternary climatic changes on plant distribution and evolution. *Trends in Plant Science*, *3*(11), 432–438. https://doi.org/10.1016/S1360-1385(98)01327-2

Diniz-Filho, J. A. F., Nabout, J. C., Bini, L. M., Soares, T. N., de Campos Telles, M. P., de Marco, P., & Collevatti, R. G. (2009). Niche modelling and landscape genetics of *Caryocar brasiliense* ("Pequi" tree: Caryocaraceae) in Brazilian Cerrado: an integrative approach for evaluating central–peripheral population patterns. *Tree Genetics & Genomes*, *5*(4), 617–627. https://doi.org/10.1007/s11295-009-0214-0

Durka, W. (1999). Genetic diversity in peripheral and subcentral populations of *Corrigiola litoralis* L. (Illecebraceae). *Heredity*, *83*(4), 476–484. https://doi.org/10.1038/sj.hdy.6886000

Eckert, C. G., Samis, K. E., & Lougheed, S. C. (2008). Genetic variation across species' geographical ranges: The central–marginal hypothesis and beyond. *Molecular Ecology*, *17*(5), 1170–1188. https://doi.org/10.1111/j.1365-294X.2007.03659.x

Excoffier, L., Foll, M., & Petit, R. J. (2009). Genetic consequences of range expansions. *Annual Review of Ecology, Evolution, and Systematics*, *40*(1), 481–501. https://doi.org/10.1146/annurev.ecolsys.39.110707.173414

Gaston, K. J. (2009). Geographic range limits of species. *Proceedings of the Royal Society B: Biological Sciences*, *276*(1661), 1391–1393. https://doi.org/10.1098/rspb.2009.0100

Godsoe, W., & Harmon, L. J. (2012). How do species interaction affect species distribution models? *Ecography*, *35*, 811–820. https://doi.org/10.1111/j.1600-0587.2011.07103.x

Goudet, J. (2005). Hierfstat, a package for R to compute and test hierarchical F-statistics. *Molecular Ecology Notes*, *5*(1), 184–186. https://doi.org/10.1111/j.1471-8286.2004.00828.x

Graham, M. H. (2003). Confronting multicollinearity in ecological multiple regression. *Ecology*, *84*(11), 2809–2815. https://doi.org/10.1890/02-3114

Guo, Q. (2014). Central-marginal population dynamics in species invasions. *Frontiers in Ecology and Evolution*, *2*, https://doi.org/10.3389/fevo.2014.00023

Haag, C. R., & Ebert, D. (2004). A new hypothesis to explain geographic parthenogenesis. *Annales Zoologici Fennici*, *41*(4), 539–544.

Hällfors, M. H., Liao, L., Dzurisin, J., Grundel, R., Hyvärinen, M., Towle, K., Wu, G. C., & Hellmann, J. J. (2015). Addressing potential local

1365294x, 2022, 7, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.16230 by Uva Universiteitsbibliotheek, Wiley Online Library on [23/02/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

adaptation in species distribution models: implications for conservation under climate change. *Ecological Applications*, *26*(4), https://doi.org/10.1890/15-0926

Hampe, A., & Petit, R. J. (2005). Conserving biodiversity under climate change: The rear edge matters: Rear edges and climate change. *Ecology Letters*, *8*(5), 461–467. https://doi.org/10.1111/j.1461-0248.2005.00739.x

Hamrick, J. L., & Godt, M. J. W. (1996). Effects of life history traits on genetic diversity in plant species. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences*, *351*, 1291–1298. https://doi.org/10.1098/rstb.1996.0112

Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, *143*(1), 29–36. https://doi.org/10.1148/radiology.143.1.7063747

Hargreaves, A. L., & Eckert, C. G. (2014). Evolution of dispersal and mating systems along geographic gradients: Implications for shifting ranges. *Functional Ecology*, *28*(1), 5–21. https://doi.org/10.1111/1365-2435.12170

Hewitt, G. M. (1996). Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society*, *58*(3), 247–276. https://doi.org/10.1111/j.1095-8312.1996.tb01434.x

Hewitt, G. M. (1999). Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society*, *68*(1–2), 87–112. https://doi.org/10.1111/j.1095-8312.1999.tb01160.x

Hewitt, G. (2000). The genetic legacy of the quaternary ice ages. *Nature*, *405*(6789), 907–913. https://doi.org/10.1038/35016000

Hijmans, R. J., & Elith, J. (2012). *Species Distribution Modelling with R*. Retrieved from: http://cran.r-project.org/web/packages/dismo/vignettes/dm.pdf

Hijmans, R. J., Williams, E., & Vennes, C. (2019). *geosphere: Spherical Trigonometry (Version 1.5-10)*. Retrieved from https://CRAN.R-project.org/package=geosphere

Howes, B. J., & Lougheed, S. C. (2008). Genetic diversity across the range of a temperate lizard. *Journal of Biogeography*, *35*(7), 1269–1278. https://doi.org/10.1111/j.1365-2699.2007.01867.x

Kawecki, T. J. (2008). Adaptation to marginal habitats. *Annual Review of Ecology, Evolution, and Systematics*, *39*(1), 321–342. https://doi.org/10.1146/annurev.ecolsys.38.091206.095622

King, R. A., & Ferris, C. (1998). Chloroplast DNA phylogeography of *Alnus glutinosa* (L.) Gaertn. *Molecular Ecology*, *7*(9), 1151–1161. https://doi.org/10.1046/j.1365-294x.1998.00432.x

Kirkpatrick, M., & Barton, N. H. (1997). Evolution of a species' range. *American Naturalist*, *150*(1), 1–23. https://doi.org/10.1086/286054

Kuittinen, H., Mattila, A., & Savolainen, O. (1997). Genetic variation at marker loci and in quantitative traits in natural populations of *Arabidopsis thaliana*. *Heredity*, *79*(2), 144. https://doi.org/10.1038/hdy.1997.137

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), https://doi.org/10.18637/jss.v082.i13

Liu, C., Berry, P. M., Dawson, T. P., & Pearson, R. G. (2005). Selecting thresholds of occurrence in the prediction of species distributions. *Ecography*, *28*(3), 385–393. https://doi.org/10.1111/j.0906-7590.2005.03957.x

Liu, C., White, M., & Newell, G. (2013). Selecting thresholds for the prediction of species occurrence with presence-only data. *Journal of Biogeography*, *40*(4), 778–789. https://doi.org/10.1111/jbi.12058

McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models*. Chapman and Hall.

McPherson, J. M., Jetz, W., & Rogers, D. J. (2004). The effects of species' range sizes on the accuracy of distribution models: Ecological phenomenon or statistical artefact? *Journal of Applied Ecology*, *41*(5), 811–823. https://doi.org/10.1111/j.0021-8901.2004.00943.x

Meirmans, P. G., Goudet, J., & Gaggiotti, O. (2011). Ecology and life history affect different aspects of the population structure of 27 high-alpine plants. *Molecular Ecology*, *20*(15), 3144–3155. https://doi.org/10.1111/j.1365-294X.2011.05164.x

Micheletti, S. J., & Storfer, A. (2015). A test of the central-marginal hypothesis using population genetics and ecological niche modelling in an endemic salamander (*Ambystoma barbouri*). *Molecular Ecology*, *24*(5), 967–979. https://doi.org/10.1111/mec.13083

Naimi, B., & Araújo, M. B. (2016). sdm: A reproducible and extensible R platform for species distribution modelling. *Ecography*, *39*(4), 368–375. https://doi.org/10.1111/ecog.01881

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, 133–142. https://doi.org/10.1111/j.2041-210x.2012.00261.x@10.1111/(ISSN)2041-210X.STATSTOO

Nei, M. (1987). *Molecular evolutionary genetics*. Columbia University Press.

Nybom, H. (2004). Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. *Molecular Ecology*, *13*, 1143–1155. https://doi.org/10.1111/j.1365-294X.2004.02141.x

Parks, C. R., Wendel, J. F., Sewell, M. M., & Qiu, Y.-L. (1994). The significance of allozyme variation and introgression in the *Liriodendron tulipifera* complex (Magnoliaceae). *American Journal of Botany*, *81*(7), 878–889. https://doi.org/10.1002/j.1537-2197.1994.tb15569.x

Peterson, A. T., Papeş, M., & Eaton, M. (2007). Transferability and model evaluation in ecological niche modeling: A comparison of GARP and Maxent. *Ecography*, *30*(4), 550–560. https://doi.org/10.1111/j.0906-7590.2007.05102.x

Petit, R. J., Aguinagalde, I., de Beaulieu, J.-L., Bittkau, C., Brewer, S., Cheddadi, R., Ennos, R., Fineschi, S., Grivet, D., Lascoux, M., Mohanty, A., Müller-Starck, G., Demesure-Musch, B., Palmé, A., Martín, J. P., Rendell, S., & Vendramin, G. G. (2003). Glacial refugia: hotspots but not melting pots of genetic diversity. *Science*, *300*(5625), 1563–1565. https://doi.org/10.1126/science.1083264

Pfeifer, M., Schatz, B., Xavier Picó, F., Passalacqua, N. G., Fay, M. F., Carey, P. D., & Jeltsch, F. (2009). Phylogeography and genetic structure of the orchid *Himantoglossum hircinum* (L.) Spreng. across its European central–marginal gradient. *Journal of Biogeography*, *36*(12), 2353–2365. https://doi.org/10.1111/j.1365-2699.2009.02168.x

Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, *190*(3), 231–259. https://doi.org/10.1016/j.ecolmodel.2005.03.026

Pironon, S., Villellas, J., Morris, W. F., Doak, D. F., & García, M. B. (2015). Do geographic, climatic or historical ranges differentiate the performance of central versus peripheral populations? *Global Ecology and Biogeography*, *24*(6), 611–620. https://doi.org/10.1111/geb.12263

Pulgarín-R, P. C., & Burg, T. M. (2012). Genetic signals of demographic expansion in downy woodpecker (*Picoides pubescens*) after the last North American glacial maximum. *PLoS One*, *7*(7), e40412. https://doi.org/10.1371/journal.pone.0040412

Raes, N., & Aguirre-Gutiérrez, J. (2018). A modeling framework to estimate and project species distributions in space and time. In C. Hoorn, A. Perrigo, & A. Antonelli (Eds.), *Mountains, climate and biodiversity*. John Wiley & Sons Ltd.

Sagarin, R. D., & Gaines, S. D. (2002). The 'abundant centre' distribution: To what extent is it a biogeographical rule? *Ecology Letters*, *5*(1), 137–147. https://doi.org/10.1046/j.1461-0248.2002.00297.x

Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin*, *2*(6), 110–114. https://doi.org/10.2307/3002019

Schoen, D. J., Morgan, M. T., & Bataillon, T. (1996). How does self-pollination evolve? Inferences from floral ecology and molecular genetic variation. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *351*(1345), 1281–1290. https://doi.org/10.1098/rstb.1996.0111

1365294x, 2022, 7, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.16230 by Uva Universiteitsbibliotheek, Wiley Online Library on [23/02/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Schonswetter, P., Stehlik, I., Holderegger, R., & Tribsch, A. (2005). Molecular evidence for glacial refugia of mountain plants in the European Alps. *Molecular Ecology*, *14*(11), 3547–3555. https://doi.org/10.1111/j.1365-294X.2005.02683.x

Sewell, M. M., Parks, C. R., & Chase, M. W. (1996). Intraspecific chloroplast DNA variation and biogeography of North American *Liriodendron* L. (magnoliaceae). *Evolution*, *50*(3), 1147–1154. https://doi.org/10.1111/j.1558-5646.1996.tb02355.x

Sexton, J. P., McIntyre, P. J., Angert, A. L., & Rice, K. J. (2009). Evolution and ecology of species range limits. *Annual Review of Ecology, Evolution, and Systematics*, *40*(1), 415–436. https://doi.org/10.1146/annurev.ecolsys.110308.120317

Soltis, D. E., Gitzendanner, M. A., Strenge, D. D., & Soltis, P. S. (1997). Chloroplast DNA intraspecific phylogeography of plants from the Pacific Northwest of North America. *Plant Systematics and Evolution*, *206*(1–4), 353–373. https://doi.org/10.1007/BF00987957

Soule, M. (1973). The epistasis cycle: a theory of marginal populations. *Annual Review of Ecology and Systematics*, *4*(1), 165–187. https://doi.org/10.1146/annurev.es.04.110173.001121

Taberlet, P., Fumagalli, L., Wust-Saucy, A., & Cosson, J. (1998). Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology*, *7*(4), 453–464. https://doi.org/10.1046/j.1365-294x.1998.00289.x

Ursenbacher, S., Guillon, M., Cubizolle, H., Dupoué, A., Blouin-Demers, G., & Lourdais, O. (2015). Postglacial recolonization in a cold climate specialist in western Europe: Patterns of genetic diversity in the adder (*Vipera berus*) support the central-marginal hypothesis. *Molecular Ecology*, *24*(14), 3639–3651. https://doi.org/10.1111/mec.13259

van Proosdij, A. S. J., Sosef, M. S. M., Wieringa, J. J., & Raes, N. (2016). Minimum required number of specimen records to develop accurate species distribution models. *Ecography*, *39*(6), 542–552. https://doi.org/10.1111/ecog.01509

Watanabe, S., Hajima, T., Sudo, K., Nagashima, T., Takemura, T., Okajima, H., Kawamiya, M. (2011). *MIROC-ESM: Model description and basic results of CMIP5-20c3m experiments*. Retrieved from http://agris.fao.org/agris-search/search.do?recordID=AV2012060623

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, *38*(6), 1358–1370. https://doi.org/10.1111/j.1558-5646.1984.tb05657.x

Yakimowski, S. B., & Eckert, C. G. (2007). Threatened peripheral populations in context: geographical variation in population frequency and size and sexual reproduction in a clonal woody shrub. *Conservation Biology*, *21*(3), 811–822. https://doi.org/10.1111/j.1523-1739.2007.00684.x

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: López-Delgado, J., & Meirmans, P. G. (2022). History or demography? Determining the drivers of genetic variation in North American plants. *Molecular Ecology*, 31, 1951–1962. https://doi.org/10.1111/mec.16230

1365294x, 2022, 7, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.16230 by Uva Universiteitsbibliotheek, Wiley Online Library on [23/02/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License