Received: 31 March 2022 | Revised: 17 October 2023 | Accepted: 21 October 2023

DOI: 10.1111/jbi.14754

Journal of Biogeography

WILEY

METHOD

Integrative demographic modelling reduces uncertainty in estimated rates of species' historical range shifts

Antonio R. Castilla[1] | Alissa Brown[2] | Sean Hoban[2] | Everett Andrew Abhainn[3] | John D. Robinson[4] | Jeanne Romero-Severson[3] | Adam B. Smith[5] | Allan E. Strand[6] | John R. Tipton[7] | Andria Dawson[8,9]

[1]Department of Plant Biology, Ecology, and Evolution, Oklahoma State University, Stillwater, Oklahoma, USA

[2]Center for Tree Science, The Morton Arboretum, Lisle, Illinois, USA

[3]Department of Biological Sciences, University of Notre Dame, South Bend, Indiana, USA

[4]Department of Fisheries and Wildlife, Michigan State University, East Lansing, Michigan, USA

[5]Center for Conservation and Sustainable Development, Missouri Botanical Garden, Saint Louis, Missouri, USA

[6]Department of Biology, College of Charleston, Charleston, South Carolina, USA

[7]Department of Mathematical Sciences, University of Arkansas, Fayetteville, Arkansas, USA

[8]Department of Biology, Mount Royal University, Calgary, Alberta, Canada

[9]Department of Biological Sciences, University of Calgary, Calgary, Alberta, Canada

Correspondence
Allan E. Strand, Department of Biology, College of Charleston, Charleston, SC, USA.
Email: stranda@cofc.edu

Funding information
Division of Biological Infrastructure; Division of Ocean Sciences

Abstract

Aim: Biogeographers have used three primary data types to examine shifts in tree ranges in response to past climate change: fossil pollen, genetic data and contemporary occurrences. Although recent efforts have explored formal integration of these types of data, we have limited understanding of how integration affects estimates of range shift rates and their uncertainty. We compared estimates of biotic velocity (i.e. rate of species' range shifts) using each data type independently to estimates obtained using integrated models.

Location: Eastern North America.

Taxon: *Fraxinus pennsylvanica* Marshall (green ash).

Methods: Using fossil pollen, genomic data and modern occurrence data, we estimated biotic velocities directly from 24 species distribution models (SDMs) and 200 pollen surfaces created with a novel Bayesian spatio-temporal model. We compared biotic velocity from these analyses to estimates based on coupled demographic-coalescent simulations and Approximate Bayesian Computation that combined fossil pollen and SDMs with population genomic data collected across the *F. pennsylvanica* range.

Results: Patterns and magnitude of biotic velocity over time varied by the method used to estimate past range dynamics. Estimates based on fossil pollen yielded the highest rates of range movement. Overall, integrating genetic data with other data types in our simulation-based framework reduced apparent uncertainty in biotic velocity estimates and resulted in greater similarity in estimates between SDM- and pollen-integrated analyses.

Main Conclusions: By reducing uncertainty in our assessments of range shifts, integration of data types improves our understanding of the past distribution of species. Based on these results, we propose further steps to reach the integration of these three lines of biogeographical evidence into a unified analytical framework.

KEYWORDS
Approximate Bayesian computation, biotic velocity, climate change, data integration, genetics, paleoecology, pollen, range shifts, species distribution models, uncertainty

Antonio R. Castilla and Alissa Brown contributed equally to this study.

1 | INTRODUCTION

A longstanding challenge in ecology is using limited and indirect historical evidence best to infer 'biotic velocity', the rate at which species shift their ranges over time (Ordonez & Williams, 2013; Reid, 1899). Understanding biotic velocity is crucial for identifying which species can migrate quickly enough to track changes in current and projected future climates (Dawson et al., 2011; McLachlan et al., 2005) and the key traits facilitating this migration (Harnik et al., 2018; Lankau et al., 2015; Pither et al., 2018). A variety of data sources and methods, each with its own set of strengths and weaknesses (Hoban et al., 2019), are employed to infer past biotic velocity. Consequently, the estimates of postglacial species movement exhibit considerable variation in magnitude, contingent upon the specific data and methodology (Carroll et al., 2015; Cheddadi et al., 2014; McLachlan et al., 2005; Ordonez & Williams, 2013). Thus, joint estimation with multiple data sources that arise from a common underlying process offers promise for improving estimates of biotic velocity and characterizing uncertainty in these estimates (Hoban et al., 2019).

Biogeographers have used three primary data types and corresponding methods for inferring biotic velocity of plants in response to environmental change. First, fossil pollen provides a snapshot of vegetation through time and is thus invaluable for documenting past presence and relative abundance (Davis, 1963). However, taxa differ in pollen production and dispersal, resulting in complex relationships between sediment pollen and vegetation (Davis, 1963; Parshall & Calcote, 2001). Quantitative approaches can help reveal these relationships (Dawson et al., 2016; Sugita, 2007), and account for spatio-temporal dependence and uncertainty (Dawson et al., 2019; Pirzamanbein et al., 2014; Trachsel et al., 2020). Furthermore, pollen can usually only be resolved to the genus level. Second, species distribution models (SDMs) estimate correlations between species' occurrences and underlying environmental conditions (Araújo & Guisan, 2006), and can be combined with paleoclimate models to hindcast species' movements (Varela et al., 2011). However, SDMs rely on numerous assumptions, including temporal stability of species' niches, and the absence of both dispersal constraints and species interactions (Wiens et al., 2009). Finally, genetic data can help infer population histories (Hu et al., 2009; McLachlan et al., 2005; Mohn et al., 2021), including timing of population splits and distinct migration pathways (Avise, 1992; Laricchia et al., 2015; Shafer et al., 2010). Nevertheless, genetic inferences suffer from identifiability issues. For instance, different demographic processes (e.g. long-distance colonization from a single refuge vs. expansion from multiple refugia) can lead to similar genetic signatures, and recent bottlenecks or expansions can obscure the signature of ancient events (Gong et al., 2008; Petit et al., 1997).

Given the advantages and limitations of each of these data types and approaches used to infer species range shifts, it is unsurprising that estimates of biotic velocity from previous studies using different data and approaches vary widely. When estimated from a single data type, post-glacial biotic velocities in trees range from ~0 to >1000 meters per year (McLachlan et al., 2005; Ordonez & Williams, 2013; Pither et al., 2018), depending on the data, inferential approach and biotic velocity metric (i.e. movement of the centroid, the leading or trailing edges). Migration rates inferred from fossil pollen have been found to be one to two orders of magnitude larger than those inferred from genetic data (McLachlan et al., 2005). Estimates from SDMs are also highly variable across plant taxa, ranging from <100 to >1000 meters per year (Cunze et al., 2013). Some studies have compared biotic and climate velocities from periods with high versus low rates of climate change, demonstrating that species migration is at least in part a function of climatic shifts. In particular, Ordonez and Williams (2013) found faster migration rates from the Last Glacial Maximum (LGM) starting from ~21 Kybp and extending to 7 Kybp, when most prominent climatic episodes occurred, and relative range stationarity in the last 7 Ka.

To resolve disparity among estimates of biotic velocity, we propose an approach based on Approximate Bayesian Computation (ABC) that integrates empirical evidence and accounts for uncertainty in data and processes (Bertolle et al., 2010; Hoban et al., 2019). Several studies have integrated species distribution modelling and genetic data using ABC to infer some aspects of biogeographical history, but not biotic velocities (Bemmels et al., 2019; Brown et al., 2016). Moreover, there has not yet been an effort to compare estimates of biotic velocity from traditional single data type approaches with those from integrated methods.

While the integrated framework presented here is general, we use as an example postglacial migration of green ash (*Fraxinus pennsylvanica* Marshall), a species with a wide distribution across eastern and central North America (Figure 1). *Fraxinus pennsylvanica* is a fast-growing and early successional wind-dispersed tree recently classified as Critically Endangered in the IUCN Red List due to ongoing losses to the invasive beetle emerald ash borer. Like many North American trees, *F. pennsylvanica* migrated northward as ice sheets retreated following the LGM, likely from multiple refugia (Noakes, 2016). Information on the biogeographical history of *F. pennsylvanica* can inform conservation of its genetic resources (Barstow et al., 2018).

In this study, we compare estimates of biotic velocity from single and joint methods of biogeographical reconstruction in order to evaluate changes in parameter estimates associated with integrating additional data types. Specifically, we calculate biotic velocity of the range centroid and northern/southern range boundaries over the past 21 Ka using: (1) single data types (fossil pollen, contemporary occurrences or genetic data), and (2) integration of two data types (pollen-genetic and SDM-genetic) via ABC. We expect that (a) biotic velocities will be fastest during periods of high climatic change (i.e. from 20-7 Kybp); (b) biotic velocity estimates will have the largest uncertainty shortly after the LGM, when pollen records are most sparse, extrapolation by SDMs is most problematic (Maguire et al., 2016), and genetic signals are most likely to be obscured by more recent events; (c) genetic data and fossil pollen will lead to the

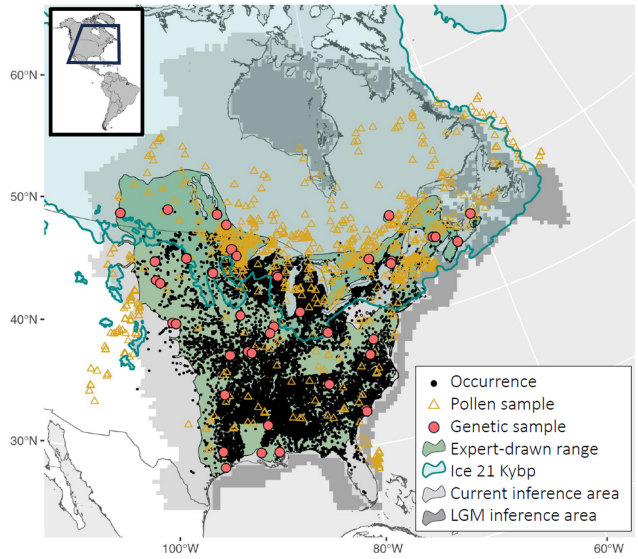

FIGURE 1 Region of inference for *Fraxinus pennsylvanica*. Our area of inference was delimited based on the spatial extent of pollen, genetic and occurrence data, while taking into account physiographic divisions, potential refugia location, the extent of the North American ice sheet and computational efficiency. Pollen sites include all sites with at least one sample (with or without *Fraxinus*) within our study's timeframe (0-21 Kybp). A range map previously published for the species is also included (Little, 1971). Changes in land mass reflect sea level rise since 21 Kybp. We include political boundaries and the range map as visual "anchors." Projection: Albers conic equal-area.

lowest and highest rates (respectively) of species movement, as in previous studies (e.g. Cheddadi et al., 2014; McLachlan et al., 2005); and (d) uncertainty in biotic velocity estimates will be reduced when multiple data types are combined, despite the propagation of uncertainty from SDMs and fossil pollen in integrated models including genetics (Figure 2).

2 | MATERIALS AND METHODS

2.1 | Spatiotemporal context

Our study region covers eastern North America and was delimited based on the spatial extent of *F. pennsylvanica* pollen, genetic and occurrence data, while taking into account physiographic divisions, potential refugial locations, and the extent of the North American ice sheet (Figure 1; Appendix S1). The temporal extent of our analysis covers the last 21 Ka. For the genetic simulations, we assumed a 30-year generation time (Whittle & Johnston, 2003). Given this temporal resolution, we base our biotic velocity inferences on 33-generation (990-year) time intervals. We were particularly interested in assessing biotic velocity during periods of dynamic climate (Figure 2) including near-LGM (~21-18.5 Kybp), Oldest Dryas (~18.5-14.7 Kybp), Bølling–Allerød (~14.7-12.9 Kybp) and Younger Dryas (~12.9-11.7 Kypb).

2.2 | Pollen

Detailed descriptions of the pollen data acquisition and analysis are presented in Appendix S2. Briefly, fossil pollen records from eastern North America were downloaded from the Neotoma Paleoecology Database (Williams et al., 2018). There were 465 pollen records within our study region with at least one pollen count sample in the considered time period (LGM to present). Each pollen record includes pollen taxon counts at a set of depths. Most records include reliable sample age estimates derived from an age-depth model. For all other records we fit Bayesian age-depth models to infer pollen sample ages (in ybp) (Parnell et al., 2008). Using a Bayesian spatio-temporal model, we jointly estimate the relative abundance of pollen taxa for the 15 tree taxa most abundant in the pollen record (*Fraxinus* ranks 9th) from 21 Kybp to the present for the study region (https://github.com/TIMBERhub/pg-pollen). We use 990-year time bins starting at 21,495 ybp, except for the last bin, which extended from 705 ybp to 2013 CE, the year genomic samples were collected. This temporal resolution falls within the 500-year pollen record temporal limit to late-Pleistocene vegetation syntheses (Blois et al., 2011). Posterior estimates of *Fraxinus* relative abundance for the study region were used in subsequent analysis.

2.3 | Species distribution *modelling*

Detailed descriptions of occurrence data processing and species distribution modelling are found in Appendices S3 and S10. Using specimen and plot data for *F. pennsylvanica* from across its native range, we calibrated 24 different SDMs using present-day climate reconstructions from two paleoclimate models (CCSM, the Community Climate System Model, and ECBilt; Lorenz et al., 2016), three regions of increasing size from which background sites were drawn for training (Appendix S3, Figure S3.3), and four modelling algorithms (Maxent, boosted regression trees, generalized linear models and natural splines). Models were projected to 300-year average climates centred on 500-year intervals (Lorenz et al., 2016), then linearly interpolated to 30-year intervals to match the time frame of the ABC analysis. These models generated a wide range of predictions about present and past suitable habitat and biotic velocity but had similar and high accuracy against present-day evaluation data (Appendix S3: Figures S3.7–S3.10).

2.4 | Genetics, demogenetic simulations and approximate Bayesian Computation (ABC)

Details on sampling, DNA extraction, RAD library preparation and bioinformatic analysis are provided in Appendix S4. The population genomic dataset included 294 *F. pennsylvanica* individuals (sampled from 21 populations across the contemporary range) and 1000 SNP loci.

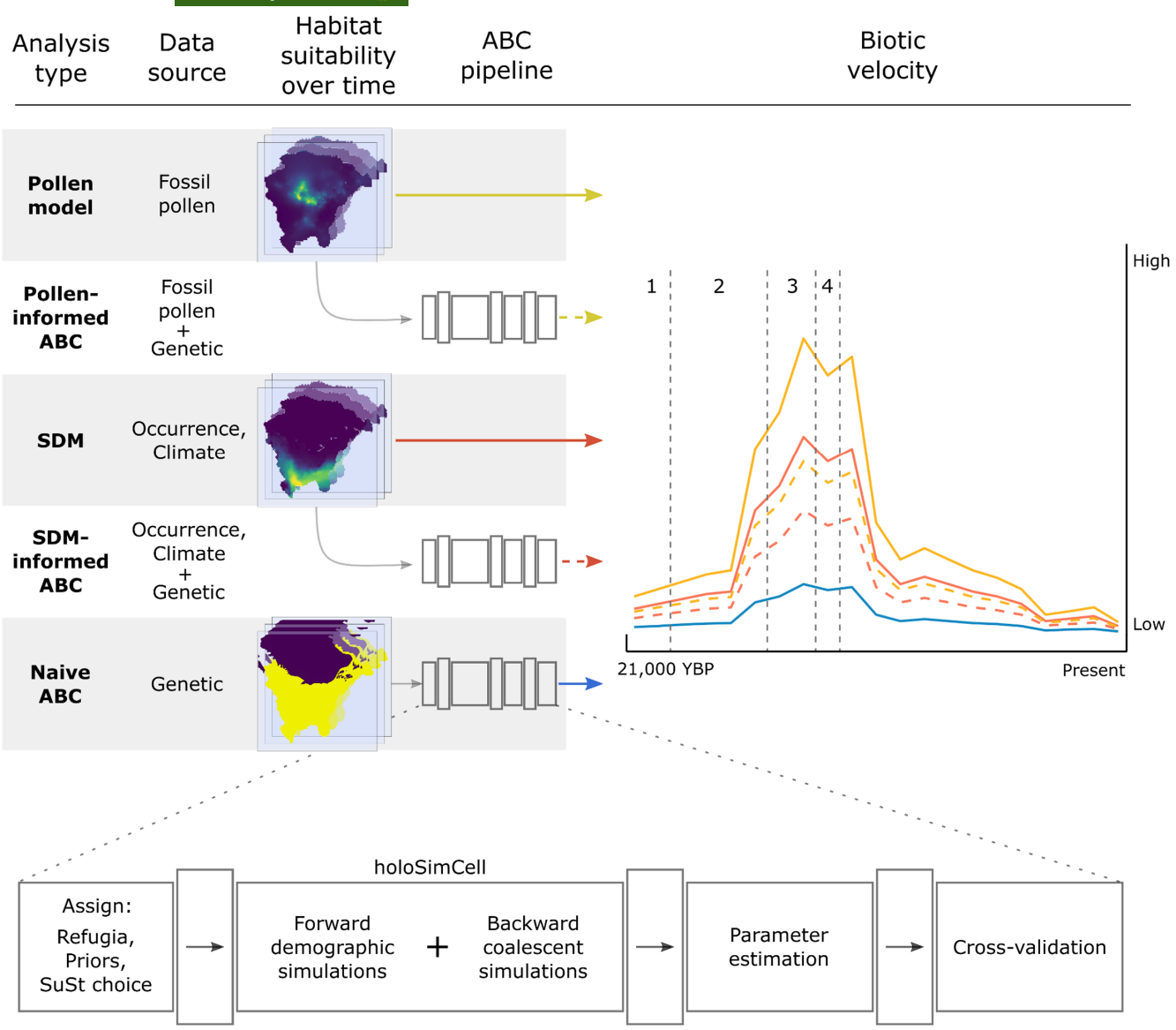

FIGURE 2 Conceptual diagram showing the analytical approaches used to infer *Fraxinus pennsylvanica* biogeographical history and potential expected trends in biotic velocity using different data types and methods. Main climatic episodes since LGM are delimited with vertical dashed lines (1: LGM; 2: Oldest Dryas; 3: Bølling–Allerød interstadial; 4: Younger Dryas).

We conducted demographic and coalescent simulations using a newly developed stage-based model (adult, offspring) implemented in the R (R Core Team, 2020) package holoSimCell (https://github.com/stranda/holoSimCell). Our spatially explicit model couples a forward-time demographic simulation with backward-time coalescent genetic simulations (in fastsimcoal v. 2.6; Excoffier et al., 2013). First, we simulate population growth and range expansion on a spatially and temporally varying habitat suitability grid. Then, coalescent simulations incorporate information on the timing and source populations of all colonization events from the demographic simulation. Additional details of the coupled simulation model are provided in Appendix S5.

We first ran simulations using genetic data alone. In these simulations, we used "naïve" habitat suitability grids in which cells not completely covered by ice or water were assumed to be equally suitable with respect to their carrying capacity. Cells with fractional ice cover had their carrying capacity adjusted proportionately. We defined three alternative models that differed in the location of glacial refugia (Georgia, Texas or Pennsylvania), based on preliminary microsatellite data (Noakes, 2016). We also included a fourth model with all three refuges ("ALL" hereafter). These alternative naïve models were developed to evaluate support for a single versus multiple glacial refugia and represent a reasonable approach when no other quantitative data are available. For each of the four candidate models, we simulated 100,000 replicate datasets and conducted model selection (see Appendix S7). Parameter estimates were obtained using the most supported model and were compared to estimates from models integrating information from fossil pollen or SDMs (see below). The models considered in our naïve analyses provided

a benchmark against which to compare inferences from integrated models.

In the integrated models, SDMs and pollen-inferred relative abundance estimates were used to inform spatial patterns of habitat suitability through time. Simulations used habitat suitability grids from: (*i*) 200 spatio-temporal posterior samples of relative abundance of *Fraxinus* from the pollen model (Appendix S2), and (*ii*) the 24 different SDMs (Appendix S3). Using multiple suitability surfaces allowed us to propagate uncertainty through demographic and coalescent simulations. For the pollen- and SDM-integrated ABC models, glacial refugia were specified by thresholding suitability surfaces from pollen model or SDM predictions at 21 Kybp. For the pollen model, we used a relative abundance threshold of 0.03. Previous studies have reported that pollen percentages never exceed 1%–3% outside the species range limit (Davis et al., 1991). For the SDMs, we used a scenario-specific threshold that maximized the sum of sensitivity and specificity across all occurrences, using the present-day range map (Little, 1971) as an indicator of true "presence" or "absence."

We conducted model selection (among naïve models considering only genetic data) and parameter estimation for naïve (using only the most supported model), SDM-integrated and pollen-integrated models using ABC. Comparisons between simulated and observed genetic datasets were based on 473 summary statistics that reflect within- and among-population genetic variation, including pairwise metrics, across the simulated landscape (summary statistics listed in Appendix S5). Results were evaluated at multiple tolerances (i.e. proportion of simulated datasets that were accepted for ABC parameter estimation), but in the main text we only present estimates based on a tolerance of 0.01 (see Appendix S8 for further ABC details and results).

2.5 | Quantifying biotic velocity

We calculated biotic velocity, or the rate of movement (and depending on the metric, direction), for each method outlined above across 990-year intervals (i.e. 33 *F. pennsylvanica* generations, assuming a generation time of 30 years) starting at 21 Kybp. Rasters were reprojected to the same equal-area projection used in the demographic-genetic simulations with square cells 102.5 km on a side. We quantified biotic velocity in three ways: the rate of movement of the mass-weighted centroid (where "mass" was either abundance from demographic simulations, relative abundance from the pollen vegetation model [PVM], or environmental suitability from the SDMs); and the rates and direction of north–south movement of northern and southern range margins. To avoid stochastic effects of small samples at the extreme southern and northern edges of the range, we defined range margins using the 5th and 95th quantiles of mass as values in cells were accumulated from the south to the north across the study region (Ordonez & Williams, 2013; Zhu et al., 2012). Velocity of the centroid is non-negative by definition and reflects movement in any direction, whereas the sign of velocities of the northern and southern range edges indicate movement southward (negative) and northward (positive). To serve as a benchmark, we also calculated velocity for the land mass available for colonization as the ice sheets receded and sea level rose.

3 | RESULTS

3.1 | Velocity of available land

The land mass available for colonization showed a steady movement in centroid and northward shift in the northern boundary until ~8.5 Kya, when it peaked, then subsequently declined to ~0 (Figures 3,4). The southern boundary, which is in our study area driven by sea level rise, moved insubstantially (Figure 4).

3.2 | Non-integrated data analyses

Overall, biotic velocities estimated from our pollen model supported faster species movement than those calculated from SDMs or genomic data (Figure 3). Centroid velocity inferred from the pollen model was greater than 100 m/year for half of the time intervals. In contrast, the centroid velocity inferred from the genomic or SDM approach each only exceeded 100 m/year in a single time interval.

The centroid velocities inferred from the pollen model showed the most distinct pattern among methods employed (Figure 3a). Median biotic velocities were greater than 100 m/year from 20-15 Kybp, although the uncertainty of these estimates was relatively large (Figure 3a). During this time period, we detected several expansions of the northern margin (Figure 4a), and contractions of the southern margin (Figure 4b). Interestingly, the pollen-inferred centroid velocity decreased at the end of the Bølling–Allerød interstadial (Figure 3a). In contrast, the centroid velocities inferred from SDM and genomic data increased during the Bølling–Allerød interstadial (Figure 3c,e).

The pollen-inferred northern margin remained stable during the Bølling–Allerød interstadial (Figure 4a), while the southern margin experienced a remarkable contraction early in the Bølling–Allerød interstadial and an expansion by the end of this period (Figure 4b). Thereafter, centroid velocity increased to a local maximum coinciding with the end of the Younger Dryas and the beginning of the Holocene (12-8 Kybp; Figure 3a), driven by expansions of both northern and southern margins (Figure 4a,b). Biotic velocity decreased after this Younger Dryas local maximum until the most recent time interval when it again increased (Figure 3a), as a result of expansion of the southern margin (Figure 4a,b).

Genomic and non-integrated SDM analyses generally show similar temporal patterns in biotic velocities. Both trajectories show a gradual increase in centroid velocity from the LGM to the Bølling–Allerød interstadial period, with the most rapid migration occurring around 14-9 Kybp (Figure 3c,e). Both genomic and non-integrated SDM analyses revealed substantial northern range expansion during this period (Figure 4e,i). However, genomic data showed expansion

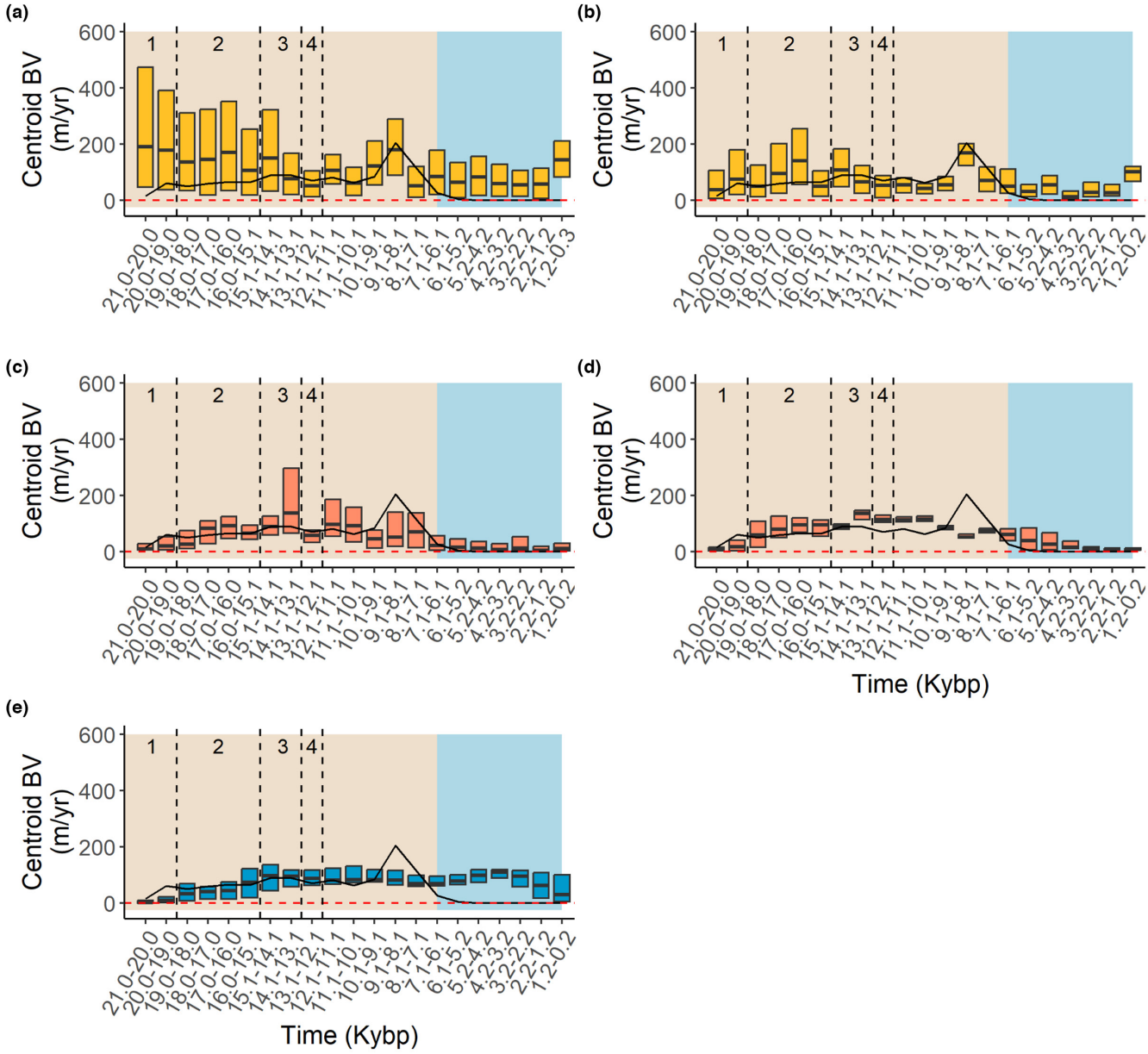

FIGURE 3 Estimates of *Fraxinus pennsylvanica* centroid biotic velocity (Centroid BV) for each data type and integration method (a: non-integrated pollen; b: pollen-integrated ABC; c: non-integrated SDMs; d: SDM-integrated ABC; e: genetic) from the Last Glacial Maximum to present time. Boxplots show median biotic velocity for each 990-year interval and the 95% highest posterior density intervals. The solid line indicates the biotic velocity of land (Appendix S6). Coloured backgrounds distinguish a period of rapid temperature change (20-7 Kybp) and relative climatic stability (7Kybp—present). Main climatic episodes since LGM are delimited with vertical dashed lines (1: LGM; 2: Oldest Dryas; 3: Bølling–Allerød interstadial; 4: Younger Dryas). The most recent bin for panel a ends at 0.3 Kybp (see main text).

of the southern margin, while non-integrated SDM showed contraction (Figure 4f,j). In addition, both genomic and non-integrated SDM estimates show contrasting patterns during the Younger Dryas, with non-integrated SDM-based centroid velocity experiencing an abrupt decrease, whereas genomic-based centroid velocities remain more stable (Figure 3c,e). Genomic data revealed a more recent period of gradual range expansion, reaching a maximum average centroid velocity around 4-3 Kybp (Figures 3e, 4i,j). In contrast, the non-integrated SDM's average centroid velocity remains relatively stable over 7 Kybp to present (Figure 3c).

For each data type, when we compare average biotic velocities across the period of rapid temperature change (i.e. 20-7 Kybp) to those across the period of relative climatic stability (i.e. 7 Kybp to present) we find different patterns. Both pollen and non-integrated SDM supported higher centroid velocities during the period of rapid temperature change than during the period of climatic stability (mean of medians and SEM: 124 ± 13 and 79 ± 12 m/year for pollen; 68 ± 9 and 13 ± 2 m/year for SDM). In contrast, genomic data supported similar biotic velocities in both periods (63 ± 8 and 77 ± 10 m/year).

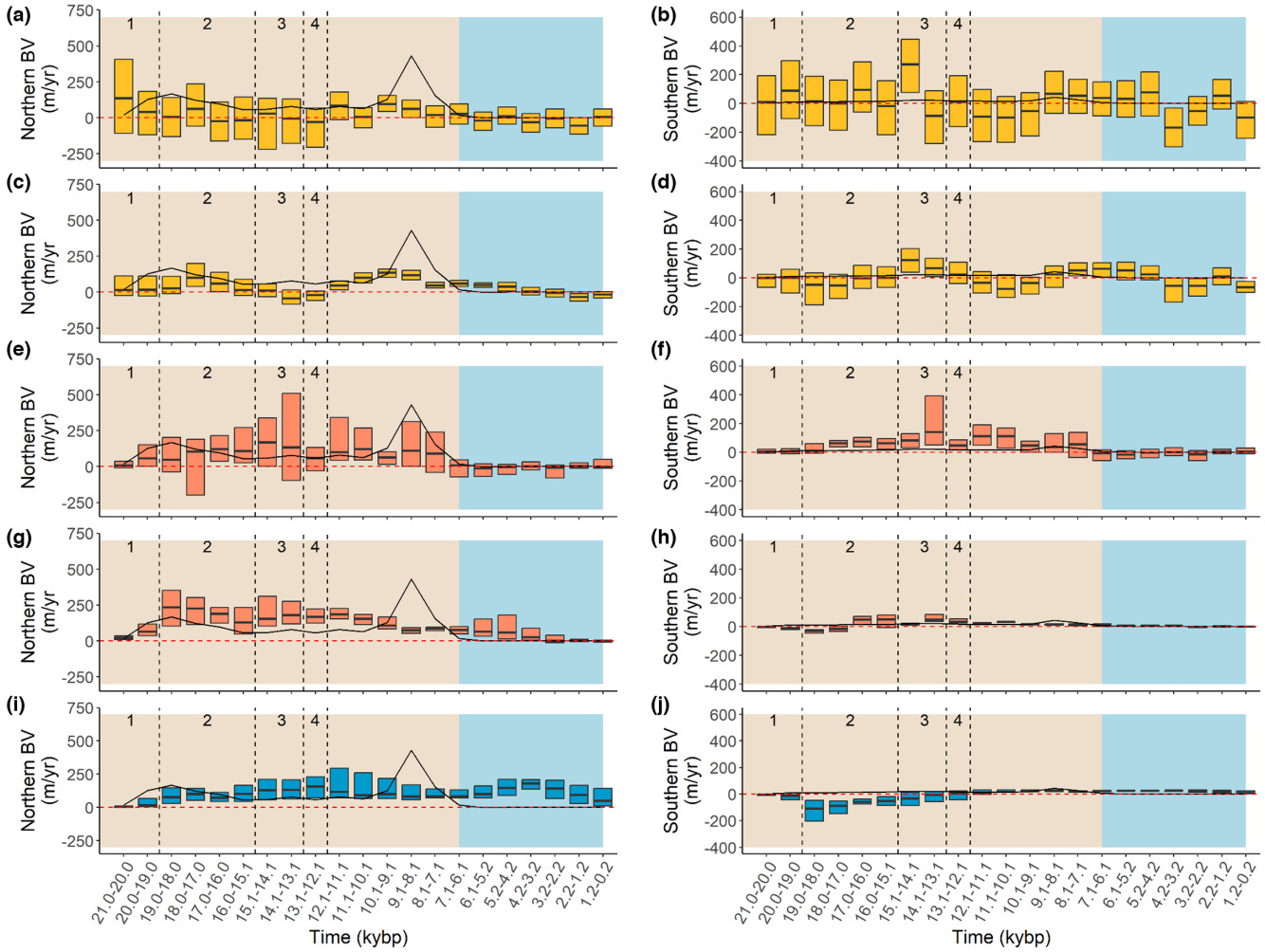

FIGURE 4 Latitudinal rates of change of the northern (95th percentile; left panels) and southern (5th percentile; right panels) boundaries of *Fraxinus pennsylvanica* distribution since the Last Glacial Maximum (21 K YBP). Positive values connote northward movement and negative southward movement. Estimates of biotic velocity are provided for each data type and integration method (a,b: non-integrated pollen; c,d: pollen-integrated ABC; e,f: non-integrated SDMs; g,h: SDM-integrated ABC; i,j: genetic). Boxplots show median biotic velocity for each 990-year interval and the 95% highest posterior density intervals. The solid line indicates the biotic velocity of land (Appendix S6). Coloured backgrounds distinguish a period of rapid temperature change (20-7 Kybp) and relative climatic stability (7Kybp—present). Note the different limits for the *y*-axis between left and right panels. Main climatic episodes since LGM are delimited with vertical dashed lines (1: LGM; 2: Oldest Dryas; 3: Bølling–Allerød interstadial; 4: Younger Dryas). The most recent bin for panels a-d ends at 0.3 Kybp (see main text).

The pattern of pollen velocities did not closely mirror that of the land velocity, but centroid velocity did recreate the sudden increase in land centroid velocity ~8.5 Kybp. The SDM and genomic analyses showed similar patterns to the land velocity (higher initially, then declining after ~8.5 Kybp), except they did not recreate the sudden peak in land velocity ~8.5 Kybp.

3.3 | Integrated data analyses

Overall, centroid velocities estimated with SDM-integrated ABC are similar to those calculated using the non-integrated SDMs (SDM-integrated ABC: mean across all time periods = 60 ± 10 m/year, range = 0–140 m/year and non-integrated SDMs: mean = 50 ± 10 m/year, range = 0–140 m/year; Figure 3c,d). Both approaches also showed a similar pattern of centroid movement (Figure 3c,d). However, there was a substantial reduction of uncertainty using SDM-integrated ABC (Figure 3c,d). Furthermore, the decline in centroid velocity during the Younger Dryas (~13-12 Kybp) was diminished by the inclusion of the genomic data (Figure 3c,d).

The temporal variation in centroid velocities from the Bølling–Allerød interstadial to the present in non-integrated pollen and pollen-integrated ABC analyses is similar (Figure 3a,b). However, the gradual increase in centroid velocity from LGM to mid-Oldest Dryas resembles the pattern detected with the genomic data (Figure 3b,e). Furthermore, the pollen-integrated ABC reduced uncertainty in estimates of biotic velocity (Figure 3a,b), especially for the oldest periods. Integration of pollen and genomic data also reduced biotic velocities compared to our non-integrated pollen analysis (non-integrated pollen: mean of medians across all time periods = 109 ± 11 m/year,

range = 191–51 m/year and pollen-integrated ABC: mean = 66 ± 8 m/year, range = 170–12 m/year; Figure 3a,b).

SDM- and pollen-integrated ABC showed the same number of time intervals with average biotic velocities exceeding 100 m/year. However, for the SDM-integrated method, this rapid migration was concentrated in a single period encompassing the Bølling–Allerød, Younger Dryas and the beginning of the Holocene (Figure 3d). In contrast, the pollen-integrated analysis had multiple peaks in biotic velocity over the last 21 Kybp (Figure 3b).

The integrated methods showed the same levels of congruence with land velocity as their non-integrated counterparts. Pollen velocity more closely tracked land velocity when integrated with genomic data.

4 | DISCUSSION

As expected, data integration via ABC substantially reduced uncertainty in biotic velocity estimates, illustrating the potential for integrative models to improve reconstructions of colonization history by incorporating essential processes like population growth and dispersal and avoiding equilibrium assumptions. Our analyses reveal substantial variation across data types in the magnitude and temporal patterns of inferred green ash movement. Estimates of biotic velocity were higher for periods of intense environmental change (i.e. 20-7 Kybp), with less centroid movement over the last 7 Kybp. As expected, the highest movement rates were inferred from fossil pollen, while biotic velocity estimates from SDMs and genomic data were similar in magnitude. Uncertainty in estimates of biotic velocity was elevated shortly after the LGM for fossil pollen when used alone, while uncertainty for the SDMs alone peaked during the Bølling–Allerød interstadial. Integrating pollen and SDMs with genetic data reduced levels of uncertainty in both during these periods. We conclude that integrating multiple data sources is necessary for achieving reliable estimates and minimizing uncertainty in reconstructions of biogeographical history.

4.1 | Comparison of biotic velocity through time

It is challenging to predict whether biotic velocity from our naïve ABC analysis (i.e. genomic data alone) would differ from the velocity of the actual land available for colonization itself (Figure 3). On one hand, dispersal under the naïve ABC model was not constrained by habitat suitability, other than the presence of ice, which should lead to higher biotic velocities. On the other hand, colonization and population growth in our demographic simulations may delay the ability of the species to track change in land availability. Our estimated velocities from naïve ABC aligned roughly with those of SDMs. Nonetheless, estimates from naïve ABC displayed a more recent peak in biotic velocity (approximately 3–4 Kybp; Figure 3). These estimates are a result of northward migration into areas outside the species' current range (Little, 1971). Since movement was not constrained by habitat suitability, populations north of the species' range reached similar carrying capacities to those further south, shifting the range northward (Figure 4i). In contrast, habitat suitability surfaces from SDMs and fossil pollen suggest little, if any, suitable habitat north of the current distribution, resulting in lower estimates of velocity for more recent time periods after the species had reached its current northern range limit. Hence, the sustained velocity under the naïve ABC analysis is likely artefactual.

Integrating SDM and genomic data eliminated the recent increase in centroid velocity detected in the naïve ABC analysis, underscoring the need to integrate multiple data types to help constrain range shifts to plausible areas. Overall, non-integrated SDM and the integrated SDM-ABC analyses produced similar biotic velocities, in magnitude and temporal trends. However, centroid velocity remained high during the Younger Dryas for the SDM-integrated ABC analysis, resembling the pattern observed in the naïve ABC (i.e. genomic alone) rather than in the non-integrated SDM analysis, where it substantially decreased. This is surprising because non-integrated SDMs depict the movement of climatically suitable habitat, regardless of occupancy (thus velocities should be higher), and because colonization dynamics in the naïve ABC analysis were unconstrained by habitat suitability (which should thus also result in higher velocities). Integration may provide otherwise missing information about parameter values (i.e. dispersal rate) and scenarios (i.e. refuge location and size). Alternatively, differences between non-integrated SDMs and SDM-integrated ABC could drive this pattern. Our demographic simulations modelled species dispersal and population growth once colonization occurs in a cell, which are not modelled in non-integrated SDMs. This temporal lag when populations experience continuous growth could explain the sustained high biotic velocities between 14 and 10 Kybp in the integrated analysis.

Pollen-based analyses yielded the highest biotic velocities and uncertainties, although integration with ABC reduced both to levels comparable to those from other models in more recent (<15 Kybp) time periods (Figure 3a,b). We expect the differences between biotic velocities and uncertainties from pollen and SDMs arise from: (a) differences in starting conditions (refugia locations) of the pollen model versus other methods; (b) the pollen model predictions encompassing a wide range of variation (i.e. we used 200 draws from the pollen model posterior, versus 24 SDM scenarios); (c) the sporadic nature of pollen record sampling (Birks & Willis, 2008); and (d) the taxonomic level of analysis (genus-level for pollen, species-level for other datasets). Post hoc inspection revealed that the SDMs generally predicted a wide expanse of refugial locations in the southern portion of the range, whereas pollen-inferred refugia were generally restricted to smaller areas spread across the study region. The variability in starting locations could be responsible for the larger variation observed in the pollen-derived velocity estimates, especially in earlier time periods (Figure 3a,b). Uneven sampling, and sparse sampling for some geographical regions and times, especially for older time periods, may explain the wider range of variation in estimates of velocity from pollen, especially prior to ~15 Kybp. This inflated variation might also result from the taxonomic resolution of

fossil pollen through the combination of occurrences from different *Fraxinus* species. Nonetheless, unlike SDMs and genetics, fossils are the only direct evidence of taxon occurrence in a region at a given point in time.

Data integration reduced uncertainty in estimates of biotic velocity, especially for pollen-based rates. The reduction in uncertainty may partially result from constraints imposed by the demographic simulations (e.g. simulated populations grow exponentially to carrying capacity, while habitat suitability values from SDMs or fossil pollen relative abundance can change instantaneously). Posterior distributions were markedly narrower than priors in both integrated models, suggesting that signals in population genomic data can distinguish among alternative habitat suitability surfaces (e.g. from different SDM algorithms or climate reconstructions). Indeed, our ABC data-integrated analyses show the imprints of both data types (genetic and pollen or genetic and occurrences) on the temporal variation of biotic velocities rather than dominance of one single data type. Thus, data integration can reduce uncertainty in postglacial movement rates despite the propagation of uncertainty in stochastic forward demographic simulations and use of multiple SDMs or posterior draws from our pollen model.

Besides reducing uncertainty of estimates of centroid velocity, the integration of fossil pollen and genomic data affected the temporal variation in biotic velocities. From the LGM to mid-Oldest Dryas (i.e. 21-16 Kybp), inferred velocities from the pollen-integrated ABC model were much slower than those from non-integrated pollen (Figure 3a,b,e). Biotic velocities of northern and southern margins more closely resembled the patterns observed in the naïve ABC model than those of the non-integrated pollen analysis (Figure 4a–d,i,j). During this period, both naïve ABC and pollen-integrated ABC models supported a substantial expansion of the southern margin. The similarity between inferences from naïve ABC and integrated models may be driven by the incorporation of population growth and dispersal limitation in our forward demographic simulations, which will slow the rate of range expansion in the generations immediately following colonization of newly available habitats. However, there are also biological explanations for the observed convergence in velocities across models. It is possible that *F. pennsylvanica* populations persisted closer to the ice sheet margin than is assumed from species occurrence- and genetic-based approaches. The possibility of high latitude refugia is not new. For example, Bemmels et al. (2019) identified LGM refugia for mesic, cool-temperate deciduous trees close to the ice sheet in eastern North America. Alternatively, pollen found at ice sheet margins could have been wind-dispersed from farther south, yielding higher inferred migration rates.

4.2 | Comparisons to estimates of biotic velocities in previous work

Comparisons with estimates of biotic velocity from other studies is complicated by differences in temporal and taxonomic resolution. However, general comparisons can serve as a plausibility check. Overall, our highest estimates of biotic velocity are comparable to those reported by Ordonez and Williams (2013) who analysed 30 trees from eastern North America and found velocities ranging from −170 to 270 m/year. Our biotic velocities for *F. pennsylvanica* are at the lower end of the 100 to 1000 m/year range suggested by classical analyses of this species (Davis, 1981; Davis & Shaw, 2001; Huntley, 1991; Huntley & Birks, 1983) but slightly higher than estimates based on phylogeographical data and simple models of postglacial movement (<100 m/year; McLachlan et al., 2005).

We found similar patterns over the last 12 Ka compared to Ordonez and Williams (2013), who used pollen records as "occurrences" in SDMs. During this period, range expansion was driven primarily by movement of the northern boundary across 12-7 Kybp, followed by relative stationarity of both edges across 7-1 Kybp. However, pollen-based analyses support the expansion of the southern border during the last 4 Ka. Our estimates of movement prior to 12 Kybp differ from those reported by Ordonez and Williams (2013). Across 16-12 Kybp, they reported a reduction in the range of *Fraxinus* driven by the contraction of the northern border. We did not detect this trend in any of our analyses but rather observed stationarity or expansion of the northern margin across different data and model combinations.

4.3 | Novelty of data integration in paleoecology

With notable exceptions (e.g. Bemmels et al., 2019; Brown et al., 2016; Knowles & Alvarado-Serrano, 2010; Ortego & Knowles, 2020), data integration remains a relatively underemployed technique in biogeographical analysis. Integration draws on complementary strengths of multiple lines of evidence while limiting the impacts of weaknesses associated with each data type. Combining data types also makes it possible to characterize uncertainty in biogeographical inference more effectively (Hoban et al., 2019). Integration has proceeded in varied ways: by incorporating SDMs with genetic data (Bemmels et al., 2019; Brown et al., 2016; Knowles & Alvarado-Serrano, 2010; Ortego & Knowles, 2020) or using fossil pollen in combination with SDMs (Maguire et al., 2016; Ordonez & Williams, 2013). Others have qualitatively compared inferences of biogeographical history from SDMs and fossil pollen data (Alba-Sanchez et al., 2010), SDMs and genetic models (e.g. Jia & Zhang, 2019; Mangaravite et al., 2019; Napier et al., 2020; Park & Donoghue, 2019; Souza et al., 2017), and less frequently among all three data types (Hao et al., 2018). To our knowledge, this work is the first to formally compare estimates of biotic velocity and associated uncertainty using single versus multiple integrated data sources. Our framework demonstrates that integrating multiple data types can recreate biogeographical trajectories consistent with expectations based on periods of climate change and stasis and provide a full accounting of uncertainty (Hoban et al., 2019).

Validating any study that seeks to reconstruct the past is a challenge; it is not possible to know with certainty how species moved

through space and time prior to the modern period of direct observation. Given this, we are unable to definitively conclude which data type "best" recreates past distributions. While we did test the capacity of ABC to distinguish between alternative scenarios using cross-validation in our naïve ABC analysis (Appendix S7), this does not address the relative fit of pollen-integrated and SDM-integrated models to the observed data and ultimately relies on the same data used to calibrate the models. Independent information from macrofossils, ancient DNA, archaeological sites, or isotopic and dendrochronological data could be used for validation, although they could also be accommodated into the modelling process. Alternatively, the power of ABC to distinguish between biogeographical scenarios could be evaluated using simulated biogeographical histories that generate both pollen profiles and genotypes under different scenarios (Lotterhos et al., 2022).

4.4 | Next steps in integration

We integrated population genomic data with habitat suitability surfaces from SDM and pollen analyses using ABC. Other possible approaches to data integration include formally combining pollen and occurrence data in a joint pollen-SDM habitat suitability surface. Integrated models can outperform those based on single data types (Ahmad Suhaimi et al., 2021; Pacifici et al., 2017). However, not all approaches respond similarly to spatial and temporal misalignment of data types (Ahmad Suhaimi et al., 2021; Pacifici et al., 2019), an issue not addressed here but maybe even more relevant when combining disparate types of data (occurrence, fossil pollen and genomic). For example, contemporary occurrence data can usually be assigned a date of collection, but fossil pollen must be dated using methods with inherent error. Likewise, different types of data may have more or less efficacy in recovering species' biogeographical histories owing to variation in geographical and climatic context and species' biology. We advocate for further integration in studies seeking to reconstruct species' biogeographical histories while providing a full accounting of sources of error. This work will be critical for anticipating species' capacity for range shifts and changes in genetic composition in response to contemporary climate change.

ACKNOWLEDGEMENTS

We thank Michigan State University's Institute for Cyber-Enabled Research, Open Science Grid (supported by NSF and US-DOE) and College of Charleston Information Technology for providing critical resources necessary for our simulation work. The authors thank members of the Scribner Lab at Michigan State University and Dr. Chris Nice at Texas State University for helpful comments. We truly appreciate the detailed comments and suggestions provided by Dr. Rosemary Gillespie and three anonymous reviewers, which improved the quality of this manuscript. We acknowledge funding support from NSF ABI-1759759, NSF OCE-1924599, The Morton Arboretum Center for Tree Science, and the Alan Graham Fund in Global Change. Collection permission was obtained from city, state and national landowners, for example, US Forest Service, and for private land, permission was granted by individual landowners.

CONFLICT OF INTEREST STATEMENT

None.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in GitHub at https://github.com/stranda/holoSimCell.

ORCID

Adam B. Smith https://orcid.org/0000-0002-6420-1659

REFERENCES

Ahmad Suhaimi, S. S., Blair, G. S., & Jarvis, S. G. (2021). Integrated species distribution models: A comparison of approaches under different data quality scenarios. *Diversity and Distributions*, *27*(6), 1066–1075.

Alba-Sánchez, F., López-Sáez, J. A., Pando, B. B. D., Linares, J. C., Nieto-Lugilde, D., & López-Merino, L. (2010). Past and present potential distribution of the Iberian Abies species: A phytogeographic approach using fossil pollen data and species distribution models. *Diversity and Distributions*, *16*(2), 214–228.

Araújo, M. B., & Guisan, A. (2006). Five (or so) challenges for species distribution modelling. *Journal of Biogeography*, *33*(10), 1677–1688.

Avise, J. C. (1992). Molecular population structure and the biogeographic history of a regional fauna: A case history with lessons for conservation biology. *Oikos*, *63*, 62–76.

Barstow, M., Oldfield, S., Westwood, M., Jerome, D., Beech, E., & Rivers, M. (2018). *The Red List of Fraxinus*. BGCI.

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*(17), 8431–8436.

Bertolle, G., Benazzo, A., & Mona, S. (2010). ABC as a flexible framework to estimate demography over space and time: Some cons, many pros. *Molecular Ecology*, *19*(13), 2609–2625.

Birks, H. J. B., & Willis, K. J. (2008). Alpines, trees, and refugia in Europe. *Plant Ecology & Diversity*, *1*(2), 147–160.

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*(15–16), 1926–1939.

Brown, J. L., Weber, J. J., Alvarado-Serrano, D. F., Hickerson, M. J., Franks, S. J., & Carnaval, A. C. (2016). Predicting the genetic consequences of future climate change: The power of coupling spatial demography, the coalescent, and historical landscape changes. *American Journal of Botany*, *103*(1), 153–163.

Carroll, C., Lawler, J. J., Roberts, D. R., & Hamann, A. (2015). Biotic and climatic velocity identify contrasting areas of vulnerability to climate change. *PLoS One*, *10*(10), e0142024.

Cheddadi, R., Birks, H. J. B., Tarroso, P., Liepelt, S., Gömöry, D., Dullinger, S., Meier, E. S., Hülber, K., Maiorano, L., & Laborde, H. (2014). Revisiting tree-migration rates: Abies alba (mill.), a case study. *Vegetation History and Archaeobotany*, *23*(2), 113–122.

Cunze, S., Heydel, F., & Tackenberg, O. (2013). Are plant species able to keep pace with the rapidly changing climate? *PLoS One*, *8*(7), e67909.

Davis, M. B. (1963). On the theory of pollen analysis. *American Journal of Science*, *261*, 897–912.

Davis, M. B. (1981). Quaternary history and the stability of forest communities. In *Forest Succession, Concepts and Application* (pp. 132–153). Springer–Verlag.

Davis, M. B., Schwartz, M. W., & Woods, K. (1991). Detecting a species limit from pollen in sediments. *Journal of Biogeography*, *18*, 653–668.

Davis, M. B., & Shaw, R. G. (2001). Range shifts and adaptive responses to quaternary climate change. *Science*, *292*(5517), 673–679.

Dawson, A., Paciorek, C. J., Goring, S. J., Jackson, S. T., McLachlan, J. S., & Williams, J. W. (2019). Quantifying trends and uncertainty in prehistoric forest composition in the upper midwestern United States. *Ecology*, *100*(12), e02856.

Dawson, A., Paciorek, C. J., McLachlan, J. S., Goring, S., Williams, J. W., & Jackson, S. T. (2016). Quantifying pollen–vegetation relationships to reconstruct ancient forests using 19th–century forest composition and pollen data. *Quaternary Science Reviews*, *137*, 156–175.

Dawson, T. P., Jackson, S. T., House, J. I., Prentice, I. C., & Mace, G. M. (2011). Beyond predictions: Biodiversity conservation in a changing climate. *Science*, *332*(6025), 53–58.

Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C., & Foll, M. (2013). Robust demographic inference from genomic and SNP data. *PLoS Genetics*, *9*(10), e1003905.

Gong, W., Chen, C., Dobeš, C., Fu, C. X., & Koch, M. A. (2008). Phylogeography of a living fossil: Pleistocene glaciations forced Ginkgo biloba L.(Ginkgoaceae) into two refuge areas in China with limited subsequent postglacial expansion. *Molecular Phylogenetics and Evolution*, *48*(3), 1094–1105.

Hao, Q., de Lafontaine, G., Guo, D., Gu, H., Hu, F. S., Han, Y., Song, Z., & 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*(4), 592–606.

Harnik, P. G., Maherali, H., Miller, J. H., & Manos, P. S. (2018). Geographic range velocity and its association with phylogeny and life history traits in north American woody plants. *Ecology and Evolution*, *8*(5), 2632–2644.

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*(12), 1991–2011.

Hu, F. S., Hampe, A., & Petit, R. J. (2009). Paleoecology meets genetics, deciphering past vegetational dynamics. *Frontiers in Ecology and the Environment*, *7*(7), 371–379.

Huntley, B. (1991). How plants respond to climate change: Migration rates, individualism and the consequences for plant communities. *Annals of Botany*, *67*, 15–22.

Huntley, B., & Birks, H. J. B. (1983). *Atlas of past and present pollen maps for Europe, 0–13,000 years ago*. Cambridge University Press.

Jia, S. W., & Zhang, M. L. (2019). Pleistocene climate change and phylogeographic structure of the Gymnocarpos przewalskii (Caryophyllaceae) in the Northwest China: Evidence from plastid DNA, ITS sequences, and microsatellite. *Ecology and Evolution*, *9*(9), 5219–5235.

Knowles, L., & Alvarado-Serrano, D. F. (2010). Exploring the population genetic consequences of the colonization process with spatio-temporally explicit models, insights from coupled ecological, demographic and genetic models in montane grasshoppers. *Molecular Ecology*, *19*(17), 3727–3745.

Lankau, R. A., Zhu, K., & Ordonez, A. (2015). Mycorrhizal strategies of tree species correlate with trailing range edge responses to current and past climate change. *Ecology*, *96*(6), 1451–1458.

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 & Genomes*, *11*(2), 30.

Little Jr, E. L. (1971). Atlas of United States trees. Volume 1. Conifers and important hardwoods. Misc. Publ. 1146. Washington, DC: U.S. Department of Agriculture, Forest Service. 320 p. 1462.

Lorenz, D. J., Nieto-Lugilde, D., Blois, J. L., Fitzpatrick, M. C., & Williams, J. W. (2016). Downscaled and debiased climate simulations for North America from 21,000 years ago to 2100AD. *Scientific Data*, *3*(1), 1–19.

Lotterhos, K. E., Fitzpatrick, M. C., & Blackmon, H. (2022). Simulation tests of methods in evolution, ecology, and systematics: Pitfalls, Progress, and principles. *Annual Review of Ecology, Evolution, and Systematics*, *53*, 113–136.

Maguire, K. C., Nieto-Lugilde, D., Blois, J. L., Fitzpatrick, M. C., Williams, J. W., Ferrier, S., & Lorenz, D. J. (2016). Controlled comparison of species-and community-level models across novel climates and communities. *Proceedings of the Royal Society B: Biological Sciences*, *283*(1826), 2817.

Mangaravite, É., da Silveira, T. C., Huamán-Mera, A., de Oliveira, L. O., Muellner-Riehl, A. N., & Schnitzler, J. (2019). Genetic diversity of Cedrela fissilis (Meliaceae) in the Brazilian Atlantic Forest reveals a complex phylogeographic history driven by quaternary climatic fluctuations. *Journal of Systematics and Evolution*, *57*(6), 655–669.

McLachlan, J. S., Clark, J. S., & Manos, P. S. (2005). Molecular indicators of tree migration capacity under rapid climate change. *Ecology*, *86*(8), 2088–2098.

Mohn, R. A., Oleas, N. H., Smith, A. B., Swift, J. F., Yatskievych, G. A., & Edwards, C. E. (2021). The phylogeographic history of a range disjunction in eastern North America: The role of post-glacial expansion into newly suitable habitat. *American Journal of Botany*, *108*(6), 1042–1057.

Napier, J. D., de Lafontaine, G., & Chipman, M. L. (2020). The evolution of paleoecology. *Trends in Ecology & Evolution*, *35*(4), 293–295.

Noakes, A. G. (2016). Population dynamics of *Fraxinus pennsylvanica* (#10142694). Dissertation, University of Notre Dame. ProQuest Dissertations Publishing.

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*(6), 773–781.

Ortego, J., & Knowles, L. L. (2020). Incorporating interspecific interactions into phylogeographic models: A case study with Californian oaks. *Molecular Ecology*, *29*(23), 4510–4524.

Pacifici, K., Reich, B. J., Miller, D. A., Gardner, B., Stauffer, G., Singh, S., McKerrow, A., & Collazo, J. A. (2017). Integrating multiple data sources in species distribution modeling: A framework for data fusion. *Ecology*, *98*(3), 840–850.

Pacifici, K., Reich, B. J., Miller, D. A., & Pease, B. S. (2019). Resolving misaligned spatial data with integrated species distribution models. *Ecology*, *100*(6), e02709.

Park, B., & Donoghue, M. J. (2019). Phylogeography of a widespread eastern north American shrub, Viburnum lantanoides. *American Journal of Botany*, *106*(3), 389–401.

Parnell, A. C., Haslett, J., Allen, J. R., Buck, C. E., & Huntley, B. (2008). A flexible approach to assessing synchroneity of past events using Bayesian reconstructions of sedimentation history. *Quaternary Science Reviews*, *27*(19–20), 1872–1885.

Parshall, T., & Calcote, R. (2001). Effect of pollen from regional vegetation on stand-scale forest reconstruction. *The Holocene*, *11*(1), 81–87.

Petit, R. J., Pineau, E., Demesure, B., Bacilieri, R., Ducousso, A., & Kremer, A. (1997). Chloroplast DNA footprints of postglacial recolonization by oaks. *Proceedings of the National Academy of Sciences*, *94*(18), 9996–10001.

Pirzamanbein, B., Lindström, J., Poska, A., Sugita, S., Trondman, A. K., Fyfe, R., Mazier, F., Nielsen, A. B., Kaplan, J. O., Bjune, A. E., Birks, H. J. B., Giesecke, T., Kangur, M., Latałowa, M., Marquer, L., Smith, B., & Gaillard, M. J. (2014). Creating spatially continuous maps of past land cover from point estimates: A new statistical approach applied to pollen data. *Ecological Complexity*, *20*, 127–141.

Pither, J., Pickles, B. J., Simard, S. W., Ordonez, A., & Williams, J. W. (2018). Below-ground biotic interactions moderated the postglacial range dynamics of trees. *New Phytologist*, *220*(4), 1148–1160.

R Core Team. (2020). *R, A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Reid, C. (1899). *The Origin of the British Flora*. Dulau.

Shafer, A. B., Cullingham, C. I., Cote, S. D., & Coltman, D. W. (2010). Of glaciers and refugia: A decade of study sheds new light on the phylogeography of northwestern North America. *Molecular Ecology*, *19*(21), 4589–4621.

Souza, H. A. V. E., Collevatti, R. G., Lima-Ribeiro, M. S., Lemos-Filho, J. P. D., & Lovato, M. B. (2017). A large historical refugium explains spatial patterns of genetic diversity in a neotropical savanna tree species. *Annals of Botany*, *119*(2), 239–252.

Sugita, S. (2007). Theory of quantitative reconstruction of vegetation II, all you need is LOVE. *The Holocene*, *17*(2), 243–257.

Trachsel, M., Dawson, A., Paciorek, C. J., Williams, J. W., McLachlan, J. S., Cogbill, C. V., Foster, D. R., Goring, S. J., Jackson, S. T., Oswald, W. W., & Shuman, B. N. (2020). Comparison of settlement-era vegetation reconstructions for STEPPS and REVEALS pollen–vegetation models in the northeastern United States. *Quaternary Research*, *95*, 23–42.

Varela, S., Lobo, J. M., & Hortal, J. (2011). Using species distribution models in paleobiogeography, a matter of data, predictors and concepts. *Palaeogeography, Palaeoclimatology, Palaeoecology*, *310*, 451–463.

Whittle, C. A., & Johnston, M. O. (2003). Broad-scale analysis contradicts the theory that generation time affects molecular evolutionary rates in plants. *Journal of Molecular Evolution*, *56*, 223–233.

Wiens, J. A., Stralberg, D., Jongsomjit, D., Howell, C. A., & Snyder, M. A. (2009). Niches, models, and climate change: Assessing the assumptions and uncertainties. *Proceedings of the National Academy of Sciences*, *106*(Supplement 2), 19729–19736.

Williams, J. W., Grimm, E. C., Blois, J. L., Charles, D. F., Davis, E. B., Goring, S. J., Graham, R. W., Smith, A. J., Anderson, M., Arroyo-Cabrales, J., Ashworth, A. C., Betancourt, J. L., Bills, B. W., Booth, R. K., Buckland, P. I., Curry, B. B., Giesecke, T., Jackson, S. T., Latorre, C., … Takahara, H. (2018). The Neotoma paleoecology database, a multiproxy, international, community-curated data resource. *Quaternary Research*, *89*(1), 156–177.

Zhu, K., Woodall, C. W., & Clark, J. S. (2012). Failure to migrate, lack of tree range expansion in response to climate change. *Global Change Biology*, *18*(3), 1042–1052.

BIOSKETCH

Our collaborative research group focuses on developing new methodologies for data integration in the study of species biogeographical histories. Our team includes individuals with expertise in genetics, mathematics, paleoecology, species distribution modelling, simulation modelling and high-performance computing.

Author contributions: Antonio R. Castilla, Alissa Brown, Andria Dawson, Sean Hoban, John D. Robinson, Adam B. Smith and Allan E. Strand designed the study; Antonio R. Castilla, Alissa Brown, Andria Dawson, John D. Robinson, Adam B. Smith, Allan E. Strand and John R. Tipton developed methods and conducted data analyses; Antonio R. Castilla and Alissa Brown wrote the first draft and all authors contributed to the final version. Jeanne Romero-Severson, Everett Andrew Abhainn and John R. Tipton provided samples, code or funding.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Castilla, A. R., Brown, A., Hoban, S., Abhainn, E. A., Robinson, J. D., Romero-Severson, J., Smith, A. B., Strand, A. E., Tipton, J. R., & Dawson, A. (2024). Integrative demographic modelling reduces uncertainty in estimated rates of species' historical range shifts. *Journal of Biogeography*, *51*, 325–336. https://doi.org/10.1111/jbi.14754