



CrossMark
click for updates

Research

Cite this article: Metcalf JL, Prost S, Nogués-Bravo D, DeChaine EG, Anderson C, Batra P, Araújo MB, Cooper A, Guralnick RP. 2014 Integrating multiple lines of evidence into historical biogeography hypothesis testing: a *Bison bison* case study. *Proc. R. Soc. B* **281**: 20132782.

<http://dx.doi.org/10.1098/rspb.2013.2782>

Received: 25 October 2013

Accepted: 4 December 2013

Subject Areas:

ecology, evolution, palaeontology

Keywords:

ancient DNA, bison, bioclimatic envelope models, Late Quaternary, historical biogeography, palaeoclimatic reconstructions

Author for correspondence:

Robert P. Guralnick

e-mail: robert.guralnick@colorado.edu

†These authors contributed equally to this study.

Electronic supplementary material is available at <http://dx.doi.org/10.1098/rspb.2013.2782> or via <http://rspb.royalsocietypublishing.org>.

Integrating multiple lines of evidence into historical biogeography hypothesis testing: a *Bison bison* case study

Jessica L. Metcalf^{1,2,†}, Stefan Prost^{4,5,†}, David Nogués-Bravo⁶, Eric G. DeChaine⁷, Christian Anderson⁸, Persaram Batra⁹, Miguel B. Araújo^{6,10,11}, Alan Cooper¹ and Robert P. Guralnick³

¹Australian Centre for Ancient DNA, University of Adelaide, Adelaide, South Australia, Australia

²Department of Chemistry and Biochemistry, and ³Museum of Natural History, University of Colorado, Boulder, CO, USA

⁴Allan Wilson Centre for Molecular Ecology and Evolution, University of Otago, Dunedin, New Zealand

⁵Department of Integrative Biology, University of California, Berkeley, CA, USA

⁶Center for Macroecology, Evolution, and Climate, University of Copenhagen, Copenhagen, Denmark

⁷Department of Biology, Western Washington University, Bellingham, WA, USA

⁸Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA, USA

⁹Department of Geology, Greenfield Community College, Greenfield, MA, USA

¹⁰Department of Biodiversity and Evolutionary Biology, Museo Nacional de Ciencias Naturales, CSIC, Madrid, Spain

¹¹Imperial College London, Silwood Park Campus, Berkshire, UK

One of the grand goals of historical biogeography is to understand how and why species' population sizes and distributions change over time. Multiple types of data drawn from disparate fields, combined into a single modelling framework, are necessary to document changes in a species's demography and distribution, and to determine the drivers responsible for change. Yet truly integrated approaches are challenging and rarely performed. Here, we discuss a modelling framework that integrates spatio-temporal fossil data, ancient DNA, palaeoclimatological reconstructions, bioclimatic envelope modelling and coalescence models in order to statistically test alternative hypotheses of demographic and potential distributional changes for the iconic American bison (*Bison bison*). Using different assumptions about the evolution of the bioclimatic niche, we generate hypothetical distributional and demographic histories of the species. We then test these demographic models by comparing the genetic signature predicted by serial coalescence against sequence data derived from subfossils and modern populations. Our results supported demographic models that include both climate and human-associated drivers of population declines. This synthetic approach, integrating palaeoclimatology, bioclimatic envelopes, serial coalescence, spatio-temporal fossil data and heterochronous DNA sequences, improves understanding of species' historical biogeography by allowing consideration of both abiotic and biotic interactions at the population level.

1. Introduction

A main goal of historical biogeography is to determine drivers of species distributions and demography through time. Doing so, however, is challenging, as multiple types of biotic and abiotic processes affect population dynamics, and, ideally, all these processes should be considered when making inferences about a species's past distribution and demography. Carstens & Richards [1] combined advances in bioclimatic envelope models (BEMs), statistical phylogeography and coalescence in an attempt to integrate disparate processes into a common modelling framework. Their workflow began by developing hypotheses in the form of demographic models that characterize potential past

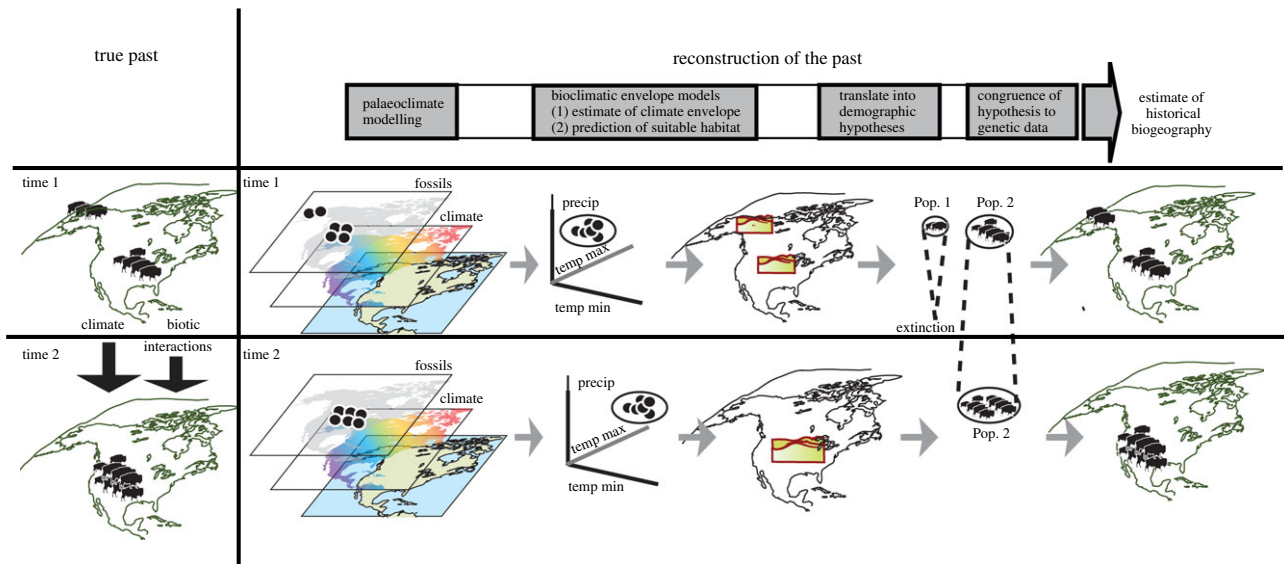


Figure 1. A workflow for reconstructing a species's historical biogeography by integrating multiple data types. We show how fossils, palaeoclimate data along with radiocarbon dating and BEM techniques can be used to develop a hypothesis testing framework to be assessed with ancient and modern DNA datasets. The goal of combining several data types over multiple time frames is to improve estimates of historical biogeography so that they are closer to the species's true history. (Online version in colour.)

distributions of species using BEMs as a guide, and then testing those alternatives using coalescent methods [1]. Despite opportunities for integration of multiple lines of evidence, the vast majority of the many subsequent studies are still limited along multiple dimensions. These limitations include: using only modern genetic data and modern occurrence data to train and test demographic models about the past (e.g. [1]); using only a single approach or method to build BEMs (e.g. [2]); considering only climate as a potential driver of species' distribution and demography (e.g. [3]); using ancient DNA but only modern occurrence data (e.g. [4]), and quantifying only coarse parameters of overall range extent when population-level scales are likely to be very important (e.g. [5]). We argue that demographic modelling approaches should also incorporate characterizations of species niches with differing degrees of data completeness, including both abiotic and biotic drivers, and use spatially explicit, mathematically rigorous and temporally precise model sets.

Late Pleistocene and Holocene subfossil deposits, characterized by dramatic biotic and climatic changes, and rich in heterogeneous, high-quality data, provide an ideal temporal window for developing synthetic, model-based approaches (e.g. [5]). Such datasets are becoming further enriched by advances in data generation from fossil material, including increased accuracy and precision of radiocarbon dating and the recovery of ancient DNA. Such advances, coupled with growing population-genetic modelling toolkits, have provided Quaternary scientists with unprecedented views of genetic diversity over space and time, showing, for example, that Late Pleistocene extinctions account for only part of the major loss of diversity during this period [6–8]. At the same time, BEMs have vastly enhanced our understanding of past species' distributions [9–12], which can then be further used as inputs into biogeographic hypotheses. By combining rich data on fossil localities and radiocarbon datasets, it is possible to use BEMs to estimate the realized climatic envelope, an n -dimensional space of climatic variables where populations have been maintained or have thrived, over multiple time periods [13,14].

Herein, we develop a multi-step, methodological workflow, which rigorously tests drivers of distributional and demographic changes, utilizing the American bison (*Bison bison*) as a focal case study group. Bison populations were once extremely large, probably spanning most of western North America [7]. It remains unclear to what extent the changing climate and interactions with humans may have contributed to their decline [5,7,15]. For example, bison were almost hunted to extinction in the nineteenth century, leaving populations today founded from as few as 30 to 50 individuals [16]. We explore in particular the following key questions. Are BEM-predicted changes in bison population structure and size supported by genetic data? Does incorporating biotic interactions between bison and humans improve demographic model support?

Bison have one of the largest datasets of accelerator mass spectrometry-dated fossils and ancient DNA available for any Late Pleistocene species. Given these abundant data, bison are ideal for seeking a consensus among multiple lines of spatial ecological, palaeontological, demographic and distributional evidence. Figure 1 illustrates how time-calibrated and georeferenced fossils, ancient DNA and palaeoclimate reconstructions can be used to best determine whether changes in species bioclimatic envelopes are associated, temporally, with changes in effective population size and population structure.

2. Material and methods

(a) Step 1: estimating the bioclimatic envelope of bison through time

Estimating a species's bioclimatic envelope is an important first step for reconstructing and understanding its past distribution [13,17]. We included climate data from multiple time periods, which allowed us to compare estimates of bison BEMs calculated independently 'within' each time period against 'pooled' estimates obtained using the full fossil record. The 'within period' analysis refers to climatic niches calibrated using only data

from within each time period. The ‘pooled period’ used all the data to create a conglomerate of the climatic niche conditions experienced by the species throughout the Late Quaternary, which was then projected to each time slice. The advantage of this approach is that it limits possible spatial and environmental bias, but it requires accepting that niches are conserved over time (but see [18]).

Two palaeoclimatic simulations were performed to represent the climatic conditions during the Marine Isotope Stage 3 (MIS 3): the warmer middle part, around 42 thousand years ago (ka) and the colder later part, around 30 ka. We also used one simulation for the Last Glacial Maximum (LGM; approx. 21 ka) and one for the Mid-Holocene (approx. 6 ka). Carbon dioxide levels were specified at 200 ppm for the MIS 3 and LGM simulations [19], and 280 ppm for the Mid-Holocene simulation [20]. The 0 ka simulation is pre-industrial and built using the same general circulation model, to ensure temporal and spatial comparability of our climatic surfaces and BEM outputs. Sea surface temperatures (SSTs) for the MIS 3 and LGM simulations were taken primarily from CLIMAP [21], with modifications from GLAMAP-2000 and other sources [22]. SSTs for the Mid-Holocene simulation were prescribed at present-day values [23]. In all cases, insolation was calculated using orbital parameters [24,25]. All simulations were spun up to equilibrium; results are 10-year averages. Areas known to be under ice sheets given palaeoclimate models were masked as unsuitable habitat. Although palaeoclimatic simulations based on different AOGCMs might differ, previous studies comparing the effect of AOGCM (GENESIS v. 2 versus HadCM3) on modelled ranges show that trends in range size are highly correlated for six different megafauna species, including the bison [5]. Together, these climate data provided five time periods (42, 30, 21, 6 and 0 ka) for estimating BEMs.

A comprehensive set of fossil and historic bison localities was assembled from multiple sources [7,26]. Fossils were calibrated using the IntCal09 calibration curve [27], available through OxCal online (<http://c14.arch.ox.ac.uk/>). Georeferencing was accomplished using Google Earth and according to best practices as defined in Chapman & Wieczorek [28] (see the electronic supplementary material, table S1).

Fossils were considered contemporaneous with a climate layer if the calibrated estimate of the radiocarbon-dated fossil was within ± 3000 years, following Nogues-Bravo *et al.* [9]. We note that the bin of ± 3000 years may not be appropriate for all climate layers and should not be considered as a hard and fast rule of our suggested framework. Three climate predictors were used: average minimum temperature of the coldest month (t_{\min}), average maximum temperature of the warmest month (t_{\max}) and mean annual precipitation sum (pre). Following ensemble forecasting methodologies [29], we fitted models with a fully factorial combination of predictor variables allowing an exploration of the resulting range of uncertainties [30].

All models were fitted with BIOENSEMBLES, a platform for computer-intensive ensemble forecasting of bioclimatic models (e.g. [30]). Models included three presence-only methods (BIOCLIM, DOMAIN and Mahalanobis [31]), two presence-background methods (MaxEnt [32] and GARP [33]) and four presence-absence methods (GLM, GAM, MARS and GBM). As we do not have true absences in our data, we generated randomly selected pseudo-absences across the cells in the region of interest without records of bison while keeping prevalence constant at 0.13 (the value found in the pooled dataset).

We randomly split the fossil and contemporaneous distributions data into 75% for calibration and 25% for evaluation, repeating the procedure 10 times. Every model run yielded a projection; ‘True Skill Statistics’ (TSS) measured the matching between predictions and observations in the 25% evaluation data. TSS-weights, indicating model performance, were obtained

for every model run and eventually used to weight the different models for their ability to predict the data. For each dataset (five periods and one ‘pooled’ set), bison data were modelled using nine model types \times seven variable combinations \times 10 cross-validated samples for a total of 630 model runs per dataset (i.e. 3780 model runs in total).

To generate a consensus across all individual model projections, first we removed all poorly performing projections (i.e. with TSS < 0.4 in the evaluation data) [34]. Then, we overlaid the remaining projections and considered a site suitable if models agreed at least 40% of the time. This is an arbitrary measure of agreement among models that is less conservative than the 50% consensus threshold used in several forecasting studies (e.g. [30,35]), and which provided reasonable results in tests using artificial data (F. G. Guilhaumon & M. B. Araújo 2012, unpublished data; see also [36]). With presence-absence methods and MaxEnt, we used the 0.13 prevalence value as the cut-off to convert probabilities or continuous suitability scores (from 0 to 1) into estimates of presence and absence [37,38]. With the distance-based presence-only methods, we used thresholds usually fixed in the literature at 0.95 for BIOCLIM and DOMAIN, and 0.75 for Mahalanobis, while for GARP we used default options to set the threshold internally. Specific details on the parameterization of each model are provided (see the electronic supplementary material, table S2). We used the binary thresholded results (figure 2) from both the ‘within’ and pooled’ probability of occurrences to help calibrate coalescence models.

(b) Step 2: demographic model set-up for bison

The first step to create demographic models from the BEMs involved examining the number of separate populations at each time slice and the relative geographic extents of those populations through time. Criteria for defining a population were: (i) a continuous set of cells separated from other such continuous groups by modelled unsuitable habitats; (ii) clear evidence that bison had existed at locations at some point in the past; and (iii) that enough fossil evidence and ancient DNA were available to generate population parameters usable for demographic model testing.

The approach so far only considers climate drivers in the creation of demographic models. However, biotic interactions (e.g. predation) can dramatically impact demography, and this is especially true with regard to bison. Archaeological evidence supports increased human occupation [39], decreased numbers of bison fossils [39] and bison hunting from Alaska to New Mexico [40–42] starting around 10 ka. Most evidence of large-scale bison hunting, including communal bison hunting with use of corrals and jumps, occurred in the Mid-Holocene, beginning at approximately 5–6 ka (reviewed by Bamforth [43]). For example, the oldest corral site is in Scoggin, Wyoming and dates to approximately 5.2 ka [43]. Furthermore, jump sites, which date back to a similar time period, became increasing frequent after approximately 3.2 ka [43]. Together, evidence of bison hunting suggests that most large-scale hunting occurred after approximately 5 ka and into the Late Holocene. The arrival of European settlers on bison populations during historic times is perhaps an even greater impact on Bison demography. The introduction of firearms and the European horse in approximately AD 1700 resulted in large-scale slaughter for private and commercial interests, which ultimately almost drove the species to extinction by the end of the nineteenth century [43,44]. We used all the information above to create alternative demographic models that represent either ‘within’ or ‘pooled’ BEM results, along with different biotic interactions. The models are described in more detail in the Results section, as well as in figure 3 and table 1.

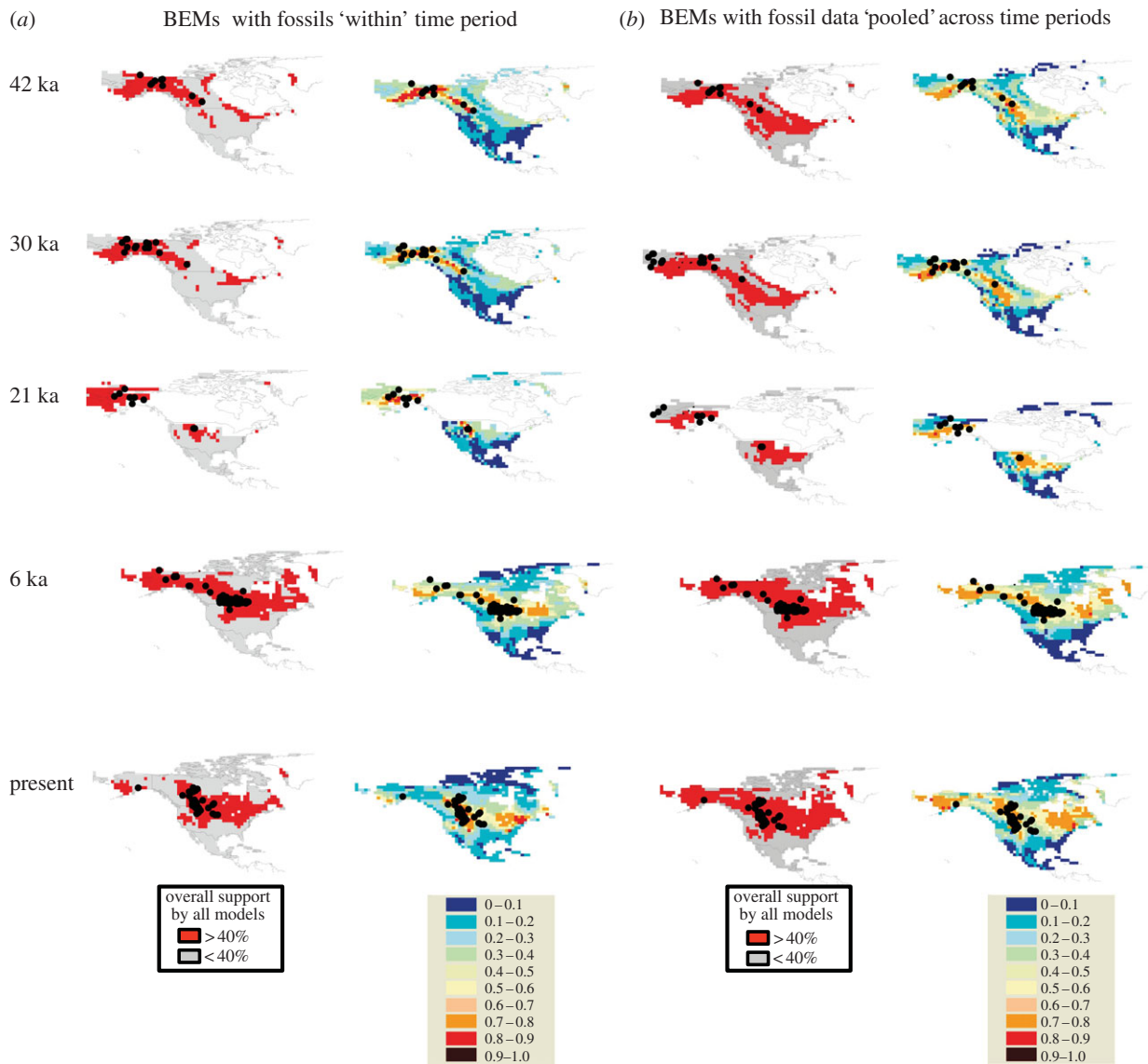


Figure 2. BEMs predicting suitable habitat for *B. bison* over five time periods: 42, 30, 21, 6 and 0 ka. The areas for which over 40% of models resulted in suitable habitat are shown in dark shading on the left side of each panel and the overall probability of an area containing suitable bison habitat is shown on the right of each panel. (a) The ‘within period’ refers to climatic envelopes calibrated and projected within each period of time. (b) The ‘pooled period’ uses all the data for each period to create a summary environmental niche space that is projected to each time slice. Fossil localities are shown as black dots. (Online version in colour.)

(c) Step 3: genetic data and analysis

Ancient DNA sequence data for bison were originally published and analysed by Shapiro *et al.* [7] and Drummond *et al.* [15]. To estimate the probability of each of the three demographic models, we used 615 bp of control region mitochondrial DNA sequence data for 159 North American, contemporary, historic and radiocarbon-dated bison samples spanning 60 000 years [7].

We used the software program BAYESIAN SERIAL SIMCOAL (BAYESSC) [45] to simulate 500 000 iterations of the different demographic scenarios (see the electronic supplementary material, input files for prior distributions) and the ‘abc’ R package [46] to estimate demographic parameters and determine the best-supported demographic model. We chose an approximate Bayesian computational (ABC) setting [47] to determine which demographic model was best supported. In general, estimates obtained with full likelihood-based approaches should be more reliable than ABC estimates because they use information from the complete data rather than summarized statistics. However, ABC is more flexible, can compute multi-population demographic models in a reasonable amount of time and computational power, and can be used for direct model selection [47,48]. To model demographic scenarios using BAYESSC, populations were grouped in different

statistics groups using age ranges (in generations) that reflected the BEM time frames described in step 1 (i.e. 42, 30, 21, 6, 0 ka \pm 3000 years) and the time frames between BEMs (before 45 ka, 33–39 ka and 9–18 ka). For time periods with two populations, fossils were either assigned to the northern or southern population depending on their location (see the electronic supplementary material, tables S1 and S3, and input files).

We chose segregating sites, nucleotide diversity and pairwise *F*_{st} as summary statistics for the analysis. This adds up to a total of 28 summary statistics. In general, more information can be added by increasing the number of summary statistics; however, too many summary statistics add stochastic noise to the analysis, and thus increase the error estimating the distance between empirical to simulated data during the regression step [47,49]. We thus used an algorithm introduced by Blum & Francois [50], which is based on nonlinear regression and uses neural networks to optimize the dimensionality. The choice of the tolerance level used in the analysis can have a strong impact on the demographic model results, and thus we used different levels in our analysis. We used tolerance levels of 0.002, 0.004 and 0.008, thereby accepting the 1000, 2000 and 4000 closest values, respectively, for the ABC parameter estimations. Expected deviance according to the deviance information criterion (DIC) [51],

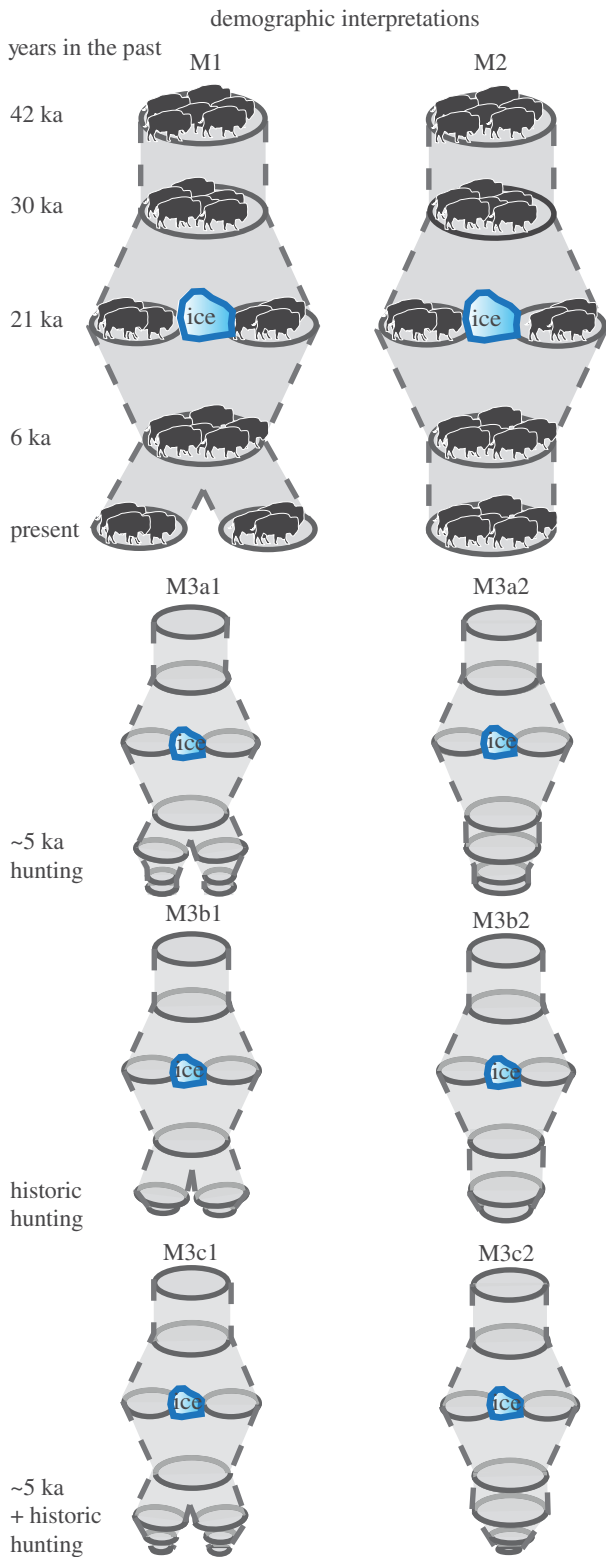


Figure 3. Demographic models based on BEMs (M1 and M2) and a combination of BEMs and additional data (M3 variants). Each model illustrates the number of populations (number of discs) and relative size (width of disc) through time. Model 3 includes six variants that include large-scale bison hunting by Native Americans starting around 5 ka (M3a), large-scale bison hunting by European settlers (M3b) and a combination of both hunting events (M3c). Each hunting scenario is included in a model based on M1 (e.g. M3a1, M3b1 and M3c1) or M2 (e.g. M3a2, M3b2 and M3c2). (Online version in colour.)

implemented in the R package ‘abc’, was applied to infer the best-supported demographic model. We simulated the respective demographic models with 1000 iterations using the parameters of each iteration separately (sampled from the posterior

distribution) as fixed model parameters. The summary statistics were then used to calculate DIC values.

To determine meaningful upper limits for the modern population size, we first performed initial runs using broad uniform priors ranging up to 30 000 000. Excessively broad priors can destabilize ABC parameter estimation. We simulated 1000 datasets and performed an ABC analysis, which showed that the higher values resulted in unreasonably high estimates of genetic diversity and that the ABC analysis clearly favoured smaller modern population size values. Thus, the upper limit for the modern population size was refined to 100 000, which resulted in a much higher effective number of simulations. This estimate was used in the full 500 000-iteration simulations.

3. Results

(a) Step 1: estimating the bioclimatic envelope of bison through time

Both ‘within period’ and ‘pooled period’ BEMs showed similar trends but with some key differences (figure 2). Overall, bison appear to have relatively continuous ranges across temperate and boreal North America from 42 ka to present, with habitat expanding after the LGM at 21 ka, when the Laurentide ice sheet began to retreat. The most notable difference between the two sets of suitable habitat predictions was the separation of northern and southern populations between 6 ka and the present day in the ‘within period’ BEMs (M1), while the ‘pooled period’ BEMs suggested panmixia during this time period.

(b) Step 2: demographic model parameterizations

We generated three main demographic model sets: (i) those based solely on climatic drivers from the ‘within time-period BEMs’; (ii) those based solely on climatic drivers from the ‘pooled BEMs’; and (iii) those that deviate from the BEMs (both ‘within’ and ‘pooled’) at certain time periods given the evidence of human impacts, including presumed impacts of Native American ‘jump’ and ‘corral’ hunting from approximately 5 to 0.9 ka and the massive impact of European settlement and hunting from approximately 0.3 ka to present. The first two models, M1 and M2, are consistent with the BEM results, and show single populations at 42 and 30 ka, and split at the LGM owing to the ice sheet covering large portions of North America. Populations re-merge in the Holocene for both M1 and M2. The models differ in that in M1 populations split again between 6 ka and the present, while in M2 there is no split. The models M3a1 (based on M1) and M3a2 (based on M2) include a bison population bottleneck resulting from human hunting between 5 and 0.9 ka (figure 3 and table 1). Models M3b1 and M3b2 include a bottleneck during historic times, reflecting intensive, large-scale hunting by European settlers from 300 years ago to the present. Finally, models M3c1 and M3c2 include both potential hunting-caused bottlenecks.

(c) Step 3: genetic data and analysis

The DIC analyses inferred the best support for the M3b2 model (using a tolerance level of 0.008; see table 2; electronic supplementary material, table S4). This model includes the historic bottleneck and is based on M2, thus including only one panmictic modern population. The second-best-supported model is M2 (tolerance: 0.002), followed by M3a2 (tolerance: 0.002).

Table 1. Description of demographic models tested with modern and ancient genetic data. Model 1 is based on BEMs calculated independently ‘within’ each time period, whereas model 2 is based on a conglomerate of ‘pooled’ climatic niche conditions experienced by the species throughout the Quaternary. Model 3 variants are based on models 1 and 2, but include potential bison population declines owing to hunting by humans.

	model 1	model 2
BEMs demographic model	single population between 42 and 30 ka; population splits between 30 and 21 ka; population merge between 11 and 8 ka, and split between approximately 5 and 1 ka	single population between 42 and 30 ka; population splits between 30 and 21 ka; population merge between 11 and 8 ka
BEM + end Pleistocene/Mid–Late Holocene decline (M3a)	BEM Model 1 + population decline between approximately 5 and 0.9 ka (M3a1)	BEM Model 2 + population decline between approximately 5 and 0.9 ka (M3a2)
BEM + historic decline (M3b)	BEM Model 1 + population decline between the approximately 0.3 ka and the present (M3b1)	BEM Model 2 + population decline between approximately 0.3 ka and the present (M3b2)
BEM + end Pleistocene/Mid–Late Holocene decline + historic decline (M3c)	BEM Model 1 + population decline between approximately 5 and 0.9 ka + decline between the approximately 0.3 ka and the present (M3c1)	BEM Model 2 + population decline between approximately 5 and 0.9 ka + decline between approximately 0.3 ka and the present (M3c2)

4. Discussion

(a) Best-supported model

In our case study, we integrated multiple lines of evidence to determine the patterns, and perhaps some of the processes, that have led to bison demographic and distribution shifts through time. Our demographic models included population subdivisions and reconnection over time rather than simply considering range-wide estimates of area, such as did Lorenzen *et al.* [5]. Of the demographic models we generated, the best-supported model was based on a combination of climate and potential influences of bison hunting by humans during historic (later than 0.3 ka) times (M3b2). The decrease in population size of bison in historic times can be reasonably attributed to decimation of bison populations by European settlers, a well-documented event [52,53]. It is notable that despite variation in importance of human impacts, the three best models were based on the ‘pooled’ BEM outputs, supporting the view of negligible climate niche evolution during the Late Quaternary (e.g. [54]) and the advantage of characterizing species climatic niches with as much fossil data as possible [9].

Overall, these results highlight the complexity of integrating multiple data types and attempting to invoke causation from various drivers. Additionally, integration of these data types is particularly challenging given that each step has assumptions and data limitations that may compromise the ability to make proper inferences. We think it is essential to detail those challenges for each step as our main goal is to critically assess and advance methods for rigorous hypothesis testing and falsification frameworks.

(b) Challenges in constructing bioclimatic envelope models over time (step 1)

We observed differences in BEM results across methods in our study, and also highlight that our results are different compared with those of Lorenzen *et al.* [5]. Overall, there was less

suitable habitat predicted using the ‘within period’ approach, which was to be expected because the bioclimatic ‘envelope’ characterized was a subset of the envelope characterized with all data. A more substantial difference in projected suitable habitat was observed by Lorenzen *et al.* [5]. Lorenzen and colleagues predicted a decrease in suitable habitat for bison at 6 ka, while our results showed a clear increase in suitable bison habitat for this time period, a result also found by Martinez-Meyer *et al.* [10]. These differences may be driven by our use of additional fossil locality data for Holocene North American bison from Harington [26], illustrating the pitfalls of incomplete sampling that is nearly ubiquitous for fossil datasets [11,55]. Additionally, in our study, we masked ice sheets during the LGM as unsuitable habitat to further improve the accuracy of our BEMs.

Another consideration when constructing BEMs is that the palaeoclimate data are modelled based on incomplete pollen and isotope records interpolated across very coarse geographic scales (in our case, the grain of the climate data layers was approx. 48 000 km²). Further, bioclimatic modelling choices may add additional sources of uncertainty [56,57]. Here, we have used an ensemble modelling approach to account for differences generated from modelling approaches, and thresholds based on current best practices. We recognize, as has been shown in the literature, that such choices impact model results [38], and future efforts can better quantify uncertainties based on differences in those choices.

(c) Challenges in constructing demographic models (step 2)

The translation of distribution predictions into demographic models requires several major assumptions about suitable habitat, population size and gene flow. First, it is assumed that suitable habitat predictions reflect *actual distributions* rather than *potential distributions*. However, BEMs represent potential distributions, but actual distributions are most likely to link to demographic estimates generated from

Table 2. DIC values for the ABC model selection. Models are ranked according to their support. Model description can be found in table 1. The no. of accepted values indicates the number of values accepted in the nonlinear regression step.

model	no. of accepted values	DIC
M3b2	4000	8.978
M2	1000	9.093
M3a2	1000	9.324
M3c1	1000	9.366
M1	1000	9.366
M3a1	1000	9.531
M3c1	2000	9.565
M3c2	4000	9.631
M2	2000	9.645
M3c1	4000	9.650
M3a2	4000	9.679
M3b2	1000	9.742
M2	4000	9.793
M3b1	4000	9.798
M1	4000	9.878
M3a1	4000	9.951
M3b1	1000	10.171
M1	2000	10.207
M3a1	2000	10.713
M3b1	2000	10.839
M3c2	2000	11.273
M3c2	1000	199.855
M3a2	2000	199.855
M3b2	2000	199.855

genetic data. Second, changes in effective population size (N_e —as estimated by genetics data) may lag or not reflect changes in census size (N_c —as counted on the landscape) [58]. Third, when a population subdivides at some point in the past, but reunites in the future, the two previously distinct populations may not interchange genetic information after contact, as presumed in our demographic models. A challenge for the future is to investigate how past potential distributions relate to demographic events and evolutionary processes such as introgression or reinforcement likely to be recorded in the genetic signal.

Another major issue is whether the five time periods covered by the BEMs were the most important for driving demographic changes in bison. Of particular concern is the lack of a climate reconstruction for the tumultuous climatic time period at the Pleistocene/Holocene transition. The transition between the Late Pleistocene Younger Dryas cold period (approx. 12.9–11.7 ka) and the onset of Holocene warming was rapid [59–62], and associated with the extinction of iconic North American Late Pleistocene megafauna [39,63]. During that climatically chaotic time period, human populations were expanding across North America after their initial arrival around 15 ka [64]. This time period would have been important for a robust model of bison historical demography.

Finally, the time periods covered by the BEMs left large temporal gaps and included large bins (± 3000 years) around each time period. Future research should explore the breadth of time bins for different palaeoclimate periods because ± 3000 years may not be appropriate in some cases. Recent advances in climate modelling allow for climate reconstructions for tighter and sequential time periods, which will be likely to improve our ability to iteratively test results and better correlate suitable habitat and genetic demographic signals.

(d) Challenges testing the demographic models with genetic data (step 3)

The use of ABC methods to estimate parameter values such as population size and timing of bottlenecks helps us to get around some of the limitations of hypothesis testing in that it is possible to explore a broad range of priors for each demographic model (i.e. hypothesis) without being limited to a strict interpretation of population-genetic parameters. However, even with the use of ABC approaches, it is still necessary to select a set of informative summary statistics useful for comparing the simulated and empirical datasets in order to assess the fit of the model to the data (see [46,65–67]).

A final challenge with genetic data is a lack of power. In our case study, we estimate values for population-genetic parameters using a single-gene ancient DNA dataset. Population-genetic estimates based on a single gene with small sample sizes (approx. $n = 10$ – 15) covering many time periods will carry a large amount of uncertainty and, in the case of mitochondrial DNA, only represent the biogeographic history of the maternal lineage. In general, we found a lack of power given our genetic data during important climatic time frames (e.g. LGM and Mid-Holocene), which is partly limited by opportunistic sampling of the fossil record. Temporal genetic data from multiple genes will provide much better power in discriminating among alternative demographic models. Genomic ancient DNA datasets will allow evolutionary biologists to refine estimates of effective population size and migration rates through time.

5. Overall conclusion

Combining datasets from multiple sources (e.g. climate, fossil and DNA), while powerful and needed, can also lead to compounding uncertainty with progress through the workflow (see electronic supplementary material, figure S1). While it is tempting to believe that more data will help untangle the thorny problems of inferring past events, it is possible that what is required is not more data, but the right kind of data with limited biases and uncertainties. For example, the addition of large archaeological datasets documenting human presence, such as in Nogues-Bravo *et al.* [9] and Lorenzen *et al.* [5], is another important data type to include in studies of species biogeography. These data may be helpful for understanding time periods when species' demographics may become strongly decoupled from their previous distributions owing to biotic interactions (e.g. high kill rates by humans).

The approach championed here is to consider multiple datasets over multiple time periods with direct translation of suitable habitat predictions into demographic models that can be contrasted with demographic models that diverge from climate at particular time points where evidence of

other biotic or abiotic drivers exist. We believe that despite the challenges, full utilization of multiple data types (e.g. DNA, fossils and palaeoclimate) considered over multiple time periods is essential for taking the next steps towards more realistic models and tests of species' historical biogeography.

Acknowledgements. We thank Beth Shapiro for assistance with georeferencing and Leigh Anne McConnaughey for assistance with making figures.

References

- Carstens BC, Richards CL. 2007 Integrating coalescent and ecological niche modeling in comparative phylogeography. *Evolution* **61**, 1439–1454. (doi:10.1111/j.1558-5646.2007.00117.x)
- Knowles LL, Carstens BC, Keat ML. 2007 Coupling genetic and ecological-niche models to examine how past population distributions contribute to divergence. *Curr. Biol.* **17**, 940–946. (doi:10.1016/j.cub.2007.04.033)
- Shepard DB, Burbink FT. 2009 Phylogeographic and demographic effects of Pleistocene climatic fluctuations in a montane salamander, *Plethodon fourchensis*. *Mol. Ecol.* **18**, 2243–2262. (doi:10.1111/j.1365-294X.2009.04164.x)
- Prost S *et al.* 2013 Effects of Late Quaternary climate change on Palearctic shrews. *Glob. Change Biol.* **19**, 1865–1874. (doi:10.1111/gcb.12153)
- Lorenzen ED *et al.* 2011 Species-specific responses of Late Quaternary megafauna to climate and humans. *Nature* **479**, 359–364. (doi:10.1038/nature10574)
- Barnes I, Matheus P, Shapiro B, Jensen D, Cooper A. 2002 Dynamics of Pleistocene population extinctions in Beringian brown bears. *Science* **295**, 2267–2270. (doi:10.1126/science.1067814)
- Shapiro B *et al.* 2004 Rise and fall of the Beringian steppe bison. *Science* **306**, 1561–1565. (doi:10.1126/science.11101074)
- Prost S, Smirnov N, Fedorov VB, Sommer RS, Stiller M, Nagel D, Knapp M, Hofreiter M. 2010 Influence of climate warming on Arctic mammals? New insights from ancient DNA studies of the collared lemming *Dicrostonyx torquatus*. *PLoS ONE* **5**, e10447. (doi:10.1371/journal.pone.0010447)
- Nogues-Bravo D, Rodriguez J, Hortal J, Batra P, Araujo MB. 2008 Climate change, humans, and the extinction of the woolly mammoth. *PLoS Biol.* **6**, 685–692. (doi:10.1371/journal.pbio.0060079)
- Martinez-Meyer E, Townsend Peterson A, Hargrove WW. 2004 Ecological niches as stable distributional constraints on mammal species, with implications for Pleistocene extinctions and climate change projections for biodiversity. *Glob. Ecol. Biogeogr.* **13**, 305–314. (doi:10.1111/j.1466-822X.2004.00107.x)
- Svenning JC, Flojgaard C, Marske KA, Nogues-Bravo D, Normand S. 2011 Applications of species distribution modeling to paleobiology. *Q. Sci. Rev.* **30**, 2930–2947. (doi:10.1016/j.quascirev.2011.06.012)
- Waltari E, Hijmans RJ, Peterson AT, Nyari AS, Perkins SL, Guralnick RP. 2007 Locating Pleistocene refugia: comparing phylogeographic and ecological niche model predictions. *PLoS ONE* **2**, e563. (doi:10.1371/journal.pone.0000563)
- Peterson AT, Soberón J, Pearson RG, Anderson RP, Nakamura M, Martinez-Meyer E, Araújo MB. 2011 *Ecological niches and geographical distributions*. Princeton, NJ: Princeton University Press.
- Hutchinson G. 1958 Concluding remarks. *Cold Spring Harb. Symp. Quant. Biol.* **22**, 415–427. (doi:10.1101/SQB.1957.022.01.039)
- Drummond AJ, Rambaut A, Shapiro B, Pybus OG. 2005 Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol. Biol. Evol.* **22**, 1185–1192. (doi:10.1093/molbev/msi103)
- Hedrick PW. 2009 Conservation genetics and North American bison (*Bison bison*). *J. Hered.* **100**, 411–420. (doi:10.1093/jhered/esp024)
- Nogues-Bravo D. 2009 Predicting the past distribution of species climatic niches. *Glob. Ecol. Biogeogr.* **18**, 521–531. (doi:10.1111/j.1466-8238.2009.00476.x)
- Araujo MB, Townsend Peterson A. 2012 Uses and misuses of bioclimatic envelope modeling. *Ecology* **93**, 1527–1539. (doi:10.1890/11-1930.1)
- Sowers T, Bender M. 1995 Climate records covering the last deglaciation. *Science* **269**, 210–214. (doi:10.1126/science.269.5221.210)
- Bonfils CJ, Lewden D, Taylor KE. 1998 Summary documentation of the PMIP models. See <http://pmip.lsce.ipsl.fr/docs>.
- CLIMAP Project Members. 1981 *Seasonal reconstructions of the Earth's surface at the Last Glacial Maximum*. Map Series, Technical Report MC-36. B, Boulder, CO: Geological Society of America.
- Barron E, Pollard D. 2002 High-resolution climate simulations of oxygen isotope stage 3 in Europe. *Q. Res.* **58**, 296–309. (doi:10.1006/qres.2002.2374)
- Pollard D, Bergengren JC, Stillwell-Soller LM, Felzer B, Thompson SL. 1998 Climate simulations for 10000 and 6000 years BP using the GENESIS global climate model. *Palaeoclim. Data Model* **2**, 183–218.
- Berger AL. 1978 Long-term variations of daily insolation and Quaternary climatic changes. *J. Atmos. Sci.* **35**, 2362–2367. (doi:10.1175/1520-0469(1978)035<2362:ltvodi>2.0.co;2)
- Berger A, Loutre MF. 1991 Insolation values for the climate of the last 10 million years. *Q. Sci. Rev.* **10**, 297–317. (doi:10.1016/0277-3791(91)90033-Q)
- Harington CR. 2003 *Annotated bibliography of Quaternary vertebrates of northern North America: with radiocarbon dates*. Toronto, Canada: University of Toronto Press.
- Reimer PJ *et al.* 2009 INTCAL09 and MARINE09 radiocarbon age calibration curves, 0–50,000 years CAL BP. *Radiocarbon* **51**, 1111–1150.
- Chapman A, Wieczorek J. 2006 *Biogeomancer, guide to best practices in georeferencing*. Copenhagen, Denmark: Global Biodiversity Information Facility.
- Araujo MB, New M. 2007 Ensemble forecasting of species distributions. *Trends Ecol. Evol.* **22**, 42–47. (doi:10.1016/j.tree.2006.09.010)
- Diniz-Filho JAF, Bini LM, Rangel TF, Loyola RD, Hof C, Nogues-Bravo D, Araujo MB. 2009 Partitioning and mapping uncertainties in ensembles of forecasts of species turnover under climate change. *Ecography* **32**, 897–906. (doi:10.1111/j.1600-0587.2009.06196.x)
- Allouche O, Steinitz O, Rotem D, Rosenfeld A, Kadmon R. 2008 Incorporating distance constraints into species distribution models. *J. Appl. Ecol.* **45**, 599–609. (doi:10.1111/j.1365-2664.2007.01445.x)
- Phillips SJ, Dudik M. 2008 Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography* **31**, 161–175. (doi:10.1111/j.0906-7590.2008.5203.x)
- Stockwell D, Peters D. 1999 The GARP modelling system: problems and solutions to automated spatial prediction. *Int. J. Geogr. Inf. Sci.* **13**, 143–158. (doi:10.1080/136588199241391)
- Araujo MB, Alagador D, Cabeza M, Nogues-Bravo D, Thuiller W. 2011 Climate change threatens European conservation areas. *Ecol. Lett.* **14**, 484–492. (doi:10.1111/j.1461-0248.2011.01610.x)
- García RA, Burgess ND, Cabeza M, Rahbek C, Araujo MB. 2012 Exploring consensus in 21st century projections of climatically suitable areas for African vertebrates. *Glob. Change Biol.* **18**, 1253–1269. (doi:10.1111/j.1365-2486.2011.02605.x)
- Araújo MB, Guilhaumon F, Neto DR, Pozo I, Calmaestra RG. 2011 *Impactos, vulnerabilidad y adaptación al cambio climático de la biodiversidad española. 2. Fauna de Vertebrados*. Madrid, Spain: Dirección General de Medio Natural y Política

- Forestal, Ministerio de Medio Ambiente y Medio Rural y Marino & Museo Nacional de Ciencias Naturales (CSIC).
37. Liu CR, Berry PM, Dawson TP, Pearson RG. 2005 Selecting thresholds of occurrence in the prediction of species distributions. *Ecography* **28**, 385–393. (doi:10.1111/j.0906-7590.2005.03957.x)
 38. Nenzen HK, Araujo MB. 2011 Choice of threshold alters projections of species range shifts under climate change. *Ecol. Model.* **222**, 3346–3354. (doi:10.1016/j.ecolmodel.2011.07.011)
 39. Guthrie RD. 2006 New carbon dates link climatic change with human colonization and Pleistocene extinctions. *Nature* **441**, 207–209. (doi:10.1038/nature04604)
 40. Yesner DR. 2001 Human dispersal into interior Alaska: antecedent conditions, mode of colonization, and adaptations. *Q. Sci. Rev.* **20**. (doi:10.1016/S0277-3791(00)00114-1)
 41. Ballenger JAM, Holliday VT, Kowler AL, Reitze WT, Prasciunas MM, Miller DS, Windingstad JD. 2011 Evidence for Younger Dryas global climate oscillation and human response in the American Southwest. *Q. Int.* **242**, 502–519. (doi:10.1016/j.quaint.2011.06.040)
 42. Hill ME, Hill MG, Widga CC. 2008 Late Quaternary bison diminution on the Great Plains of North America: evaluating the role of human hunting versus climate change. *Q. Sci. Rev.* **27**, 1752–1771. (doi:10.1016/j.quascirev.2008.07.002)
 43. Bamforth DB. 2011 Origin stories, archeological evidence, and Postclovis Paleoindian bison hunting on the Great Plains. *Am. Antiquity* **76**, 24–40. (doi:10.7183/0002-7316.76.1.24)
 44. Freese CH *et al.* 2007 Second chance for the plains bison. *Biol. Conserv.* **136**, 175–184. (doi:10.1016/j.biocon.2006.11.019)
 45. Anderson CNK, Ramakrishnan U, Chan YL, Hadly EA. 2005 Serial SimCoal: a population genetics model for data from multiple populations and points in time. *Bioinformatics* **21**, 1733–1734. (doi:10.1093/bioinformatics/bti154)
 46. Csillery K, Blum MGB, Gaggiotti OE, Francois O. 2010 Approximate Bayesian computation (ABC) in practice. *Trends Ecol. Evol.* **25**, 410–418. (doi:10.1016/j.tree.2010.04.001)
 47. Beaumont MA, Zhang WY, Balding DJ. 2002 Approximate Bayesian computation in population genetics. *Genetics* **162**, 2025–2035.
 48. Beaumont M. 2008 Joint determination of topology, divergence time and immigration in population trees. In *Simulations, genetics and human prehistory* (eds S Matsumura, P Forster, C Renfrew), pp. 135–154. Cambridge, UK: McDonald Institute for Archaeological Research.
 49. Hårdle W, Werwatz A, Müller M, Sperlich S. 2004 *Nonparametric and semiparametric models*, 1–18. Berlin, Germany: Springer.
 50. Blum MGB, Francois O. 2010 Non-linear regression models for approximate Bayesian computation. *Stat. Comput.* **20**, 63–73. (doi:10.1007/s11222-009-9116-0)
 51. François O, Laval G. 2011 *Deviance information criteria for model selection in approximate Bayesian computation*. See <http://arxiv.org/abs/1105.0269>.
 52. Coder GD. 1975 *The national movement to preserve the American buffalo in the United States and Canada between 1880 and 1920*. Columbus, OH: Ohio State University.
 53. Soper JD. 1941 History, range and home life of the northern bison. *Ecol. Monogr.* **11**, 347–412. (doi:10.2307/1943298)
 54. Araujo MB, Ferri-Yanez F, Bozinovic F, Marquet PA, Valladares F, Chown SL. 2013 Heat freezes niche evolution. *Ecol. Lett.* **16**, 1206–1219. (doi:10.1111/ele.12155)
 55. Varela S, Lobo JM, Hortal J. 2011 Using species distribution models in paleobiogeography: a matter of data, predictors and concepts. *Paleogeogr. Paleoclimatol. Paleoecol.* **310**, 451–463. (doi:10.1016/j.palaeo.2011.07.021)
 56. Barve N, Barve V, Jimenez-Valverde A, Lira-Noriega A, Maher SP, Peterson AT, Soberon J, Villalobos F. 2011 The crucial role of the accessible area in ecological niche modeling and species distribution modeling. *Ecol. Model.* **222**, 1810–1819. (doi:10.1016/j.ecolmodel.2011.02.011)
 57. Saupe EE, Barve V, Myers CE, Soberon J, Barve N, Hensz CM, Peterson AT, Owens H, Lira-Noriega A. 2012 Variation in niche and distribution model performance: the need for a priori assessment of key causal factors. *Ecol. Model.* **237**, 11–22. (doi:10.1016/j.ecolmodel.2012.04.001)
 58. Frankham R. 1995 Effective population-size adult-population size ratios in wildlife: a review. *Genet. Res.* **66**, 95–107. (doi:10.1017/S0016672300034455)
 59. Severinghaus JP, Sowers T, Brook EJ, Alley RB, Bender ML. 1998 Timing of abrupt climate change at the end of the Younger Dryas interval from thermally fractionated gases in polar ice. *Nature* **391**, 141–146. (doi:10.1038/34346)
 60. Severinghaus JP, Brook EJ. 1999 Abrupt climate change at the end of the last glacial period inferred from trapped air in polar ice. *Science* **286**, 930–934. (doi:10.1126/science.286.5441.930)
 61. Clark PU *et al.* 2012 Global climate evolution during the last deglaciation. *Proc. Natl Acad. Sci. USA* **109**, E1134–E1142. (doi:10.1073/pnas.1116619109)
 62. Shuman B, Bartlein PJ, Webb T. 2005 The magnitudes of millennial- and orbital-scale climatic change in eastern North America during the Late Quaternary. *Q. Sci. Rev.* **24**, 2194–2206. (doi:10.1016/j.quascirev.2005.03.018)
 63. Martin P. 1984 *Quaternary extinctions: a prehistoric revolution*. Tucson, AZ: University of Arizona Press.
 64. Goebel T, Waters MR, O'Rourke DH. 2008 The Late Pleistocene dispersal of modern humans in the Americas. *Science* **319**, 1497–1502. (doi:10.1126/science.1153569)
 65. Bertorello G, Benazzo A, Mona S. 2010 ABC as a flexible framework to estimate demography over space and time: some cons, many pros. *Mol. Ecol.* **19**, 2609–2625. (doi:10.1111/j.1365-294X.2010.04690.x)
 66. Nunes MA, Balding DJ. 2010 On optimal selection of summary statistics for approximate Bayesian computation. *Stat. Appl. Genet. Mol. Biol.* **9**, 34. (doi:10.2202/1544-6115.1576)
 67. Sunnåker M, Busetto AG, Numminen E, Corander J, Foll M, Dessimoz C. 2013 Approximate Bayesian computation. *PLoS Comput. Biol.* **9**, e1002803. (doi:10.1371/journal.pcbi.1002803)