Author for correspondence:
Brian Tilston Smith
e-mail: briantilstonsmith@gmail.com

THE ROYAL SOCIETY
PUBLISHING

The demography of extinction in eastern North American birds

Brian Tilston Smith[1], Marcelo Gehara[2] and Michael G. Harvey[3]

[1]Department of Ornithology, American Museum of Natural History, Central Park West at 79th Street, New York, NY 10024, USA
[2]Department of Earth and Environmental Sciences, Rutgers University Newark, 195 University Avenue, Newark, NJ 07102, USA
[3]Department of Biological Sciences, The University of Texas at El Paso, 500 W University Avenue, El Paso, TX 79968, USA

BTS, 0000-0001-8782-6400; MG, 0000-0001-9899-1970; MGH, 0000-0001-8050-6068

Species are being lost at an unprecedented rate during the Anthropocene. Progress has been made in clarifying how species traits influence their propensity to go extinct, but the role historical demography plays in species loss or persistence is unclear. In eastern North America, five charismatic landbirds went extinct last century, and the causes of their extinctions have been heavily debated. Although these extinctions are most often attributed to post-colonial human activity, other factors such as declining ancestral populations prior to European colonization could have made these species particularly susceptible. We used population genomic data from these extinct birds and compared them with those from four codistributed extant species. We found extinct species harboured lower genetic diversity and effective population sizes than extant species, but both extinct and non-extinct birds had similar demographic histories of population expansion. These demographic patterns are consistent with population size changes associated with glacial–interglacial cycles. The lack of support for overall population declines during the Pleistocene corroborates the view that, although species that went extinct may have been vulnerable due to low diversity or small population size, their disappearance was driven by human activities in the Anthropocene.

1. Introduction

Anthropogenic activities are impacting biodiversity by causing range shifts, population size reduction, local extirpation and ultimately extinction [1]. Why particular taxa are more susceptible to these factors is unclear, but species loss does not appear to be random [2,3]. Traits such as ecological specialization and body size are linked to extinction risk [4], but trait–extinction relationships may vary widely across regions [5], datasets [6] or assemblages [7]. Understanding how species traits shape demography and identifying the sorts of demographic histories that lead to extinction would clarify why some species are extinction-prone. However, the role of demography in shaping extinction and persistence remains poorly understood [8].

The demography of a population affects neutral genetic variation, which can be used to identify historical demographic events important for understanding contemporary population trajectories (e.g. loss of genetic diversity due to range size reduction). Declining genetic diversity could make a species vulnerable to changing environmental conditions [9], infectious diseases [10] and inbreeding depression [11], which are observed in most species prior to extinction [12]. Historical demography can also indicate rates of population decline. Estimates for the duration that species persist range from 1 to 10 Myr [13], but the speed at which they decline to extinction is unknown. Relative to evolutionary time scales, species loss in the Anthropocene may be rapid and similar to rates of ongoing range size reductions and declines in population sizes being observed in many species [14–16]. Alternatively, background

extinction may result from more gradual and long-term declines that progressively erode genetic diversity.

Prior analyses using ancient DNA in large mammals have provided evidence of the historical demographics of extinction. Population declines prior to extinction have been observed (e.g. [17,18]), although at the community or assemblage level genetic patterns are heterogeneous [19]. Notably, the latter study compared demographic histories of extinct species with those of species that persisted to the present, providing a contrast useful for identifying the unique demographic features associated with extinction. Across taxa, there is evidence of both gradual [20–22] and rapid [23–25] loss of genetic diversity before extinction. However, the general dynamics underlying the demography of extinction are poorly known.

The avifauna of eastern North America is a system that lost a number of species in the last 200 years and for which biogeographic histories are generally well understood. During the Pleistocene, the Laurentide Ice Sheet covered large areas in North America, shifting and reconfiguring communities to southern latitudes [26,27]. Many vertebrates across eastern North America show a genetic signature of population expansion in response to glacial retreat, albeit with species expanding at different times and the magnitude of expansion varying depending on species' ecology [28]. During the Anthropocene, the North American landscape was dramatically altered by European colonization, population growth and the industrial revolution [29]. A distinctive subspecies of the greater prairie-chicken, *Tympanuchus cupido cupido* (Heath Hen), *Ectopistes migratorius* (Passenger Pigeon), *Campephilus principalis* (Ivory-billed Woodpecker), *Conuropsis carolinensis* (Carolina Parakeet) and *Vermivora bachmanii* (Bachman's Warbler) went extinct during this period. Although the exact extinction dates are debated, the last generally recognized sightings for these five birds were between 1901 and 1988 [30]. Prior authors attributed these extinctions to habitat changes, disease and hunting [31–38], but the long-term demographic trajectories and causes of decline are not agreed upon [39,40]. Whole-genome data indicate that the *C. carolinensis* did not undergo a strong bottleneck during the Pleistocene [41]. By contrast, genomic patterns in the *E. migratorius* have been used to argue either that demographic fluctuations helped facilitate its extinction [42], or that selection greatly reduced its neutral genetic diversity [43]. There has been no concerted effort, however, to address eastern North American bird extinctions using genetic data across multiple species.

We modelled the demographic histories of *T. c. cupido*, *E. migratorius*, *C. carolinensis*, *C. principalis* and *V. bachmanii*, and compared them with those of four co-occurring extant species. We sampled genome-wide markers and estimated population genetic summary statistics to characterize the amount of variation for each species. We tested the probability of demographic scenarios representing population bottlenecks, constant size and expansions. We used a supervised machine-learning (SML) approach that entailed simulating the alternative demographic scenarios, training a neural network (NN) to differentiate among models and classifying the empirical data into the models. If extinct species were genetically compromised over historical time scales prior to extinction, we expect that they would harbour lower genetic variation than extant species and would show evidence of long-term declines during the Pleistocene. Our results leverage the power of machine learning and genomic markers to provide insight into the historical demographies associated with the extinction of charismatic North American birds.

2. Material and methods

(a) Taxon sampling and genomic data

We sampled five extinct bird taxa from museum skins at the American Museum of Natural History. In addition, we also sampled four extant species from similarly aged specimens representing common, ecologically varied species not known to exhibit phylogeographic structure across their ranges. For each taxon, we maximized geographical coverage of the available specimens and selected 7–10 individuals (electronic supplementary material, table S1) represented by specimens that were collected during a similar time period to those of the extinct birds to maximize the comparability of the data (figure 1). We collected ultraconserved elements (UCEs) using the Tetrapod UCE 5 K probe set to enrich 5060 UCE loci [44]. Details on UCE processing are available in the electronic supplementary material.

(b) Genetic diversity

We explored intraspecific diversity and population structure using principal components analysis (PCA) using the R 3.4.4 [45] package adegenet v. 2.1.2 [46], sNMF [47] and NgsAdmix [48]. Using PipeMaster v. 0.0.9 [49], we estimated the mean and variance of seven population genetic summary statistics. Additional methodological details are available in the electronic supplementary material. Because the number of loci for extant birds ended up being larger than what was obtained for the extinct birds, we subsampled 705 loci in the extant species (the lowest number of loci for any of the extinct species) 100 times and estimated summary statistics from the subsamples. We assessed variation in the estimated statistics across the 100 subsamples, to evaluate whether the subsampled values still supported any differences in diversity levels among extinct and extant birds.

To evaluate whether genetic diversity was positively correlated with geographical area, we converted the georeferenced coordinates of each sample to minimum-convex polygons in ArcGIS v. 10.7.1 (Redlands, CA). Minimum-convex polygons were clipped to the United States to remove portions of polygons that were in the ocean. We then estimated the area of each polygon in km^2 using the calculated geometry function. Because all of the *T. c. cupido* samples come from a small island (Martha's Vineyard, MA, USA), we used the area of the island. We then performed a regression of various mean genetic diversity metrics and the log converted area of each polygon in R.

(c) Demographic model selection

We simulated data under alternative demographic models and estimated the probability of each model using a SML approach. SML is statistically robust and outperforms traditional methods [50]. To specify model parameters, simulate data, and estimate summary statistics, we used the R package PipeMaster v. 0.0.9 [49], which has been used to test alternative evolutionary histories using both approximate Bayesian computation [51] and SML [52]. For all nine species, we independently tested a constant population size model (CS) plus two single-epoch models: population expansion (EXP) and population bottleneck (BN). We also evaluated two-epoch models where there was either a population expansion and then a bottleneck (BN2) or a bottleneck then a population expansion (EXP2). For each model, we specified parameter prior distributions (table 1) and simulated 100 000 datasets per hypothesis. The workflow also estimated the same 14 summary statistics from these simulations that were estimated for each species. We assessed the sensitivity to the results by
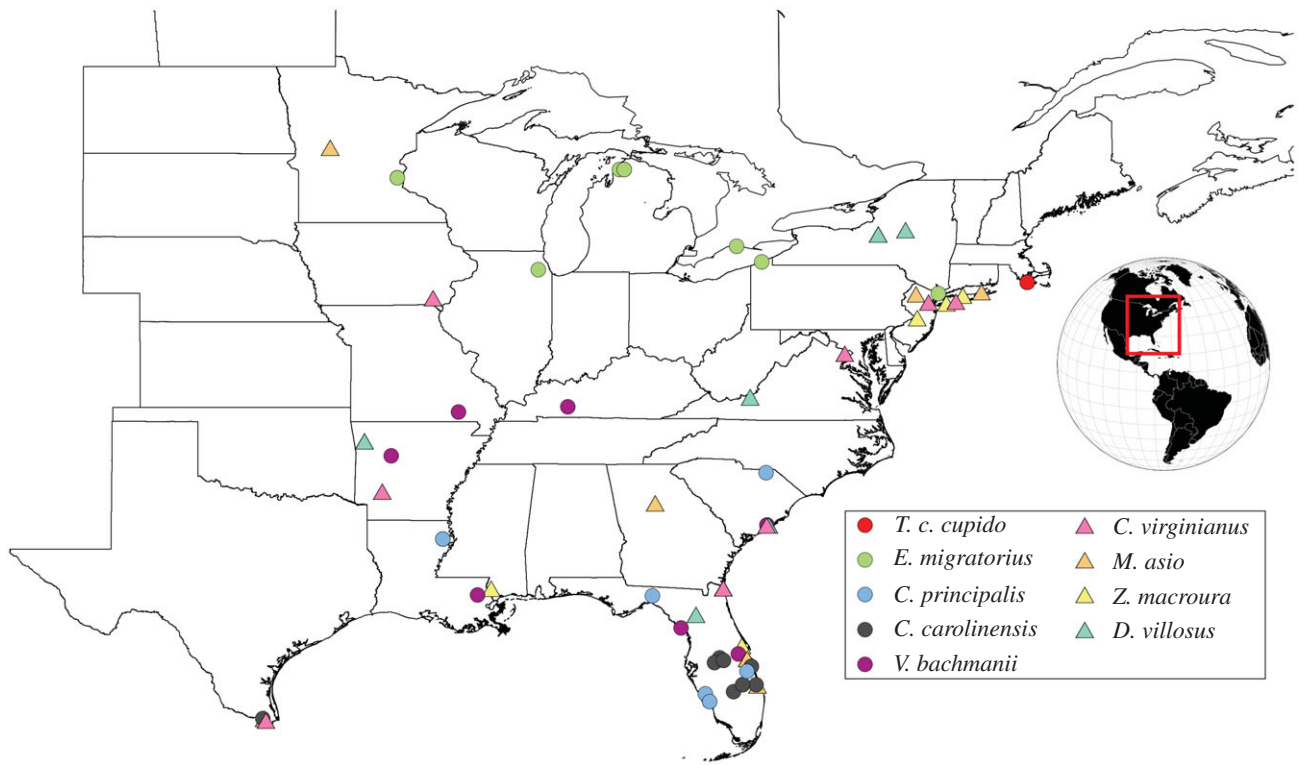
Figure 1. Sampling map of focal extinct and extant birds. Shown are the sampling localities of nine extinct (circles) and extant (triangles) species that are denoted with different colours. The age ranges of museum specimens for the extinct birds (*Tympanuchus cupido cupido*: 1896–1900; *Ectopistes migratorius*: 1875–1894; *Campephilus principalis*: 1876–1909; *Conuropsis carolinensis*: 1889–1904; and *Vermivora bachmanii*: 1890–1915) were similar to those of the extant species (*Colinus virginianus*: 1898–1911; *Megascops asio*: 1894–1915; *Zenaida macroura*: 1905–1919; and *Dryobates villosus*: 1887–1925). (Online version in colour.)

Table 1. Prior settings for demographic model selection. Shown are the five models and their corresponding parameters. The number of Ne (effective population size) and t.Ne (time of population size) parameters differ per model and range from the present (Ne0) to ancestral (Ne1;Ne2). Sampling from prior distributions under models with demographic shifts were specified so that Ne0 > Ne1 (EXP), Ne0 < Ne1 (BN), Ne0 > Ne1 < Ne2 (EXP2) and Ne0 < Ne1 > Ne2 (BN2). All Ne and t.Ne priors had a uniform distribution where K = thousand and M = million. Time priors were scaled to generations using four different generation times for each taxon, which were 1, 2 and 3 years, and an estimated generation length [53], the mean age of parents average over 10 years, which were estimated from published data or extrapolated from the mean of closely related taxa. The estimated generation lengths are as follows: *T. c. cupido*: 5.5 years; *E. migratorius*: 5.3 years; *C. principalis*: 6.5 years; *C. carolinensis*: 7.3 years; *V. bachmanii*: 3.8 years; *C. virginianus*: 3.7 years; *M. asio*: 7 years; *Z. macroura*: 6.7 years; and *D. villosus*: 4.9 years.

model	Ne0	Ne1	Ne2	t.Ne1	t.Ne2
constant population size (CS)	1 K–2 M				
population expansion (EXP)	1 K–2 M	1 K–2 M		5 K–300 K	
population bottleneck (BN)	1 K–2 M	1 K–2 M		5 K–300 K	
expansion with post LGM bottleneck (BN2)	1 K–2 M	1 K–2 M	1 K–2 M	5 K–300 K	20 K–300 K
bottleneck with post LGM expansion (EXP2)	1 K–2 M	1 K–2 M	1 K–2 M	5 K–300 K	20 K–300 K

simulating data under different generation times (table 1) and mutation rate uncertainty. We specified a mutation rate distribution with little variance (1.55×10^{-9} subs site year^{-1}; s.d. 5×10^{-10} [54]), which represents a fixed rate, and a wide uniform prior to account for uncertainty in the rate ($5 \times 10^{-11} - 5 \times 10^{-9}$) where a rate mean and s.d. were sampled from the distribution. The model parameter settings for the priors are presented in table 1.

We used the simulated data to train a NN algorithm to classify the simulations into the five categories. Before training the models, we ensured that the variance of the simulated data contained the observed data by plotting the principal components of the simulated data against the observed data (electronic supplementary material, figures S1–S8). In R, we used Keras [55]

to build and train a NN with two hidden layers with 32 node units each and a *relu* activation function. The output layer had five nodes, one for each model (i.e. class), and a *softmax* activation function. As an optimizer, we used *adam*, a *sparse_categorical_crossentropy* for the loss function and *accuracy* as validation metrics. We used 80% of the simulations for training and the remaining 20% for testing, and trained the NN for 300 epochs with batches of 50 000 and a validation split of 0.1. For the most probable model for each species and generation time, we estimated model parameters with a NN with a similar architecture as above, but here we set up an output layer with a single node and a *relu* activation. We used the mean absolute percentage error (MAPE) as optimizer and trained this NN for 3000 epochs with batch size of 10 000 and a validation split of 0.1.

We used 88% of the simulations for training and 12% for testing. As an additional measure of the power of the regression model, we calculated the coefficient of correlation between estimated and true values of the testing data. For each estimate, we performed 100 replicates and calculated the mean across estimates.

3. Results

(a) Summary of genetic variation

Read numbers varied across individuals, the percentages of reads mapped were typically low, and PCR duplicates were high. However, there was no qualitative bias among the data collected between extinct and extant species (electronic supplementary material, tables S2 and S3). From the *de novo* assemblies, we obtained 705–2269 and 1244–3519 loci across extinct and extant species, respectively. The mean number of phased alleles per species was similar among groups: *T. c. cupido*: mean 15.5 (range 3–17); *E. migratorius*: 15.8 (2–21); *C. carolinensis*: 16.2 (2–19); *C. principalis*: 17.2 (2–19); and *V. bachmanii*: 16.2 (3–19); and *Colinus virginianus*: 13.2 (2–19); *Megascops asio*: 14.7 (2–19); *Zenaida macroura*: 12.6 (2–17); and *Dryobates villosus*: 16.1 (2–19). Average locus length was similar across all taxa (259–293 bps). The number and proportion of variable loci was higher in the extant (50–71%) than extinct birds (27–51%). Extinct birds had lower mean genetic diversity than the extant species for θ_W (extinct range = 0.13–0.29; extant range = 0.31–0.48), number of haplotypes (extinct range = 1.34–1.70; extant range = 1.71–2.35), π (extinct range = 0.0005–0.0011; extant range = 0.0010–0.0020) and S (extinct range = 0.37–0.80; extant range = 0.80–1.60). The higher genetic variation in extant species was even observed after random subsampling to 705 loci (electronic supplementary material, figures S9–S12). The only exception was *M. asio*, the least genetically diverse extant taxon, which had 29% of its subsampled values lower than the mean segregating sites value for *E. migratorius*. Tajima's D, a statistic that measures deviations from a constant population size, showed both highly negative and positive values across taxa (electronic supplementary material, figure S13).

The area of minimum-convex polygons created from the location of samples ranged from 226 km² (*T. c. cupido*) through 3 056 787 km² (*M. asio*) (electronic supplementary material, table S4). Polygon area was positively correlated with the average number of segregating sites ($n = 9$; $R^2 = 0.31$; p-value = 0.07), π ($R^2 = 0.28$; p-value = 0.08), θ_W ($R^2 = 0.36$; p-value = 0.05), number of haplotypes ($R^2 = 0.35$; p-value = 0.05) and haplotype diversity ($R^2 = 0.35$; p-value = 0.05), explaining approximately 30% of the variation in genetic diversity. Average estimates of Tajima's D and the denominator of Tajima's D were uncorrelated with polygon size.

Despite limited geographical sampling, our results were consistent with low population structure in most species (electronic supplementary material, figures S14–S22). The best K was 1 based on sNMF analyses for all species except *C. virginianus*, *C. carolinensis*, *E. migratorius* and *M. asio*, where the best K was 2 (electronic supplementary material, table S5). Of those, *E. migratorius* had structure that appeared to be associated with geography, with a possible east/west divide between the northeast and upper midwest samples. NgsAdmix results (K = 2) showed the same pattern, as did the PCA, which further indicated the separation of an individual from the west (electronic supplementary material, figures S15 and S23). Structure in the other three species with K = 2 based on sNMF did not have a clear geographical association. Although sNMF indicated K = 1 was the best fit for *C. principalis*, the PCA and NgsAdmix results showed that a sample from South Carolina was separated from southern and western individuals (electronic supplementary material, figures S16 and S23). The remaining species showed no clear evidence of population structure.

(b) Demographic models

Best-fit models were sensitive to generation time and the amount of uncertainty in the mutation rate. In general, the shortest generation times tended to find support for a bottleneck followed by post LGM population expansion (BN2). As generation time was increased support switched to either a constant population size (CS) or one of the expansion models (EXP; EXP2). When we allowed for greater uncertainty in mutation rate, there was more support for constant population sizes. By contrast, the fixed-rate recovered best-fit models supporting population expansions (EXP; EXP2). There was support for multi-epoch population models, particularly population declines prior to post LGM expansions, however, there was no compelling evidence for the bottleneck model across any of these runs. We focus on the results from taxon-specific generation times with the fixed mutation rate because it accounts for generation-time differences among species and the resolution of the data is not overly compromised by a wide mutation rate. The trained NN under these conditions generally had high accuracy for assigning simulated data to models (electronic supplementary material, figure S24 and table S6). The consistency of values across replicates varied among parameters. Ancestral populations (Ne.1; Ne.2) typically had the most reliable parameter estimates with correlation coefficients greater than 0.81 among replicates, whereas the Ne of expanding populations (typically Ne.0) had lower correlation coefficients (0.32–0.49). Model parameter estimates and relevant statistics for each generation time are available in electronic supplementary material, tables S7–S10.

The most probable model for the extinct species was either expansion (*C. carolinensis* [Prob: 0.76] and *V. bachmanii* [0.99] or a bottleneck followed by a post LGM expansion (*T. c. cupido* [0.75], *E. migratorius* [0.90]) and *Campephilus principalis* [0.65]; figure 2; electronic supplementary material, table S6). Model probabilities using the wide mutation rate are available in electronic supplementary material, figure S24 and table S6. The best-fit model for *C. carolinensis* and *V. bachmanii* indicated a 8–10-fold population expansion in the last 50 K years (figure 3). Population fluctuations in other extinct species showed 2–3-fold population declines from 81 K–155 K years ago to 6–18-fold population expansions in the last 8 K years. The four extant species showed similar demographic trends of either being best supported by expansion (*M. asio* [Prob: 0.83], *Z. macroura* [1.00] and *D. villosus* [0.98]) or a bottleneck with post LGM expansion (*C. virginianus* [0.99]; figure 2). The single-epoch expanding species underwent 11–12-fold population expansions from 5 K to 114 K years ago (figure 3). The best-fit two-epoch model for *C. virginianus* showed a 1.6-fold population decline starting 170 K years ago and an almost fivefold population expansion at around 13 K years ago.
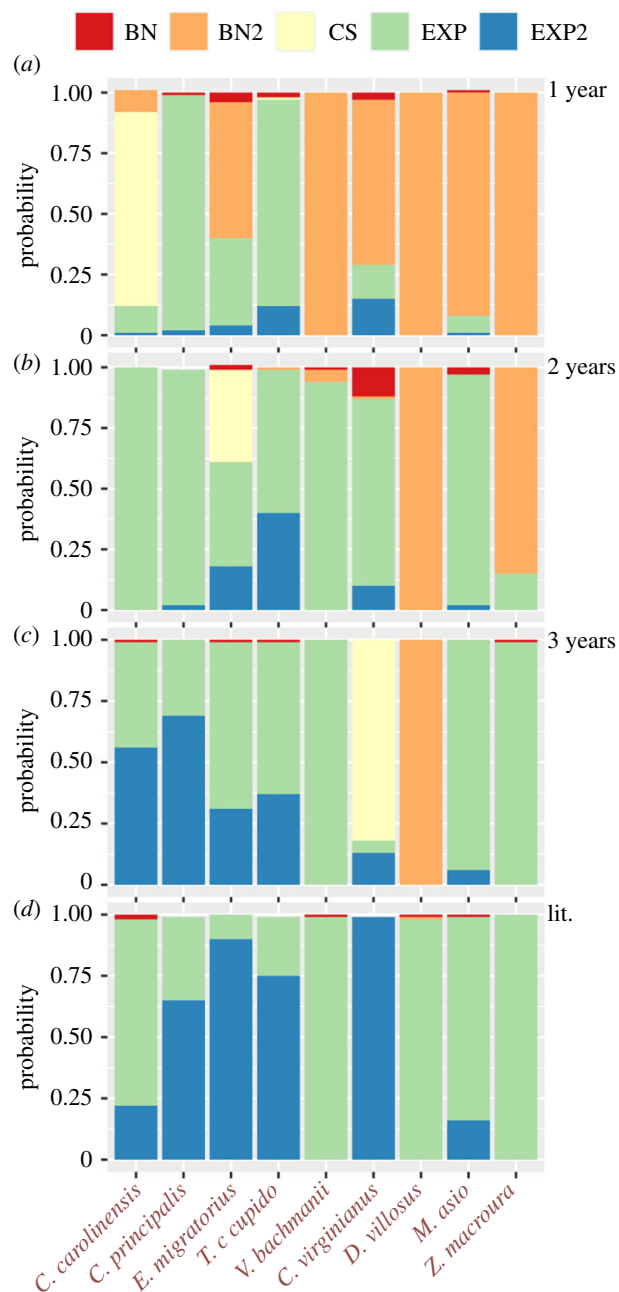
Figure 2. Neural network probabilities for alternative demographic models for each species using different generation times. The models are as follows: population bottleneck (BN), constant population (CS), population expansion (EXP), population expansion with a post last glacial maximum (LGM) bottleneck (BN2) and a population bottleneck with a post LGM expansion (EXP2). For each species and generation time, probabilities of each model are coloured (see legend). Presented are the results from the fixed mutation rate using generation times of 1 yr (*a*), 2 years (*b*), 3 years (*c*) and taxon-specific generation times from the literature (*d*). (Online version in colour.)

4. Discussion

We estimated lower genetic diversity and effective population sizes in extinct species, but we found these taxa had similar demographic histories of population expansion to extant birds. Model selection showed some sensitivity to generation time and mutation rate. Under shorter generation times and wider mutation rate priors there was higher support for constant population sizes. In some species, we did find evidence for fluctuating population sizes across the mid to late Pleistocene, but in no case did we find strong evidence of continued loss in genetic diversity. Overall, the demographic patterns

inferred for the nine focal species are consistent with the expectation of population expansions associated with glacial–interglacial cycles. The lack of support for a bottleneck in the extinct species supports the view that the declines leading to their extinctions were associated with recent events rather than longer-term trends. These findings are particularly alarming in light of recent evidence of dramatic demographic declines in Nearctic birds over the last 50 years [16].

(a) Extinction in eastern North American birds

Our comparative demographic results add another dimension to the knowledge of factors contributing to species extinctions during a period of dramatic change in eastern North America. Indigenous peoples of eastern North America had a long history of converting forests for agriculture and burning areas to improve habitat for game [56]. However, after the European colonization of North America, deforestation greatly accelerated as forests were converted to meet the growing agricultural and fuel needs of the expanding population, a process which started along the eastern seaboard and moved westward as people began colonizing new areas [56]. Deforestation stabilized by 1920, but by that time the focal extinct species studied in this paper were either extinct or close to extinction. These birds are thought to have been strained by habitat loss, for example, the logging of eastern old growth for *C. principalis* [36] and the clearing of bottomland forest with cane breaks for agriculture in *V. bachmanii* [57]. Some were clearly impacted by hunting for food (e.g. *E. migratorius* [34] and *T. c. cupido* [31]), or because they were perceived as pests (e.g. *C. carolinensis* [38]). Disease is also suspected to have played a role (e.g. *T. c. cupido* and *C. carolinensis*; [30]). In any of these scenarios, events in the last 200 years would have been a major cause of decline and extinction. Our results support these hypotheses generally in that they recovered signatures of low genetic diversity. Estimated effective population sizes for extant populations were also higher for the extant species (189 K–308 K) than the extinct species (111 K–153 K). In the decades prior to their extinction, the extinct species harboured lower genetic variability than the extant species we sampled. This depleted diversity may have made these species more susceptible to disease or inbreeding depression or less adaptable (e.g. [43]).

We evaluated whether the extinct species were susceptible to extinction due to long-term population declines or bottlenecks during the Pleistocene—in other words, whether they were 'on their way out'. Our results do not support histories involving Pleistocene bottlenecks representing continual loss of genetic diversity over evolutionary time scales. We found support for population declines, but these size reductions were followed by population expansions in four extinct and one extant species. Previous work has indicated that population fluctuations made *E. migratorius* more susceptible to extinction [42], but we found evidence for dynamic histories in birds that persisted and fluctuating population sizes have been recovered in other living species (e.g. [58]). The model selection using a generation time of 1 year, which is probably unrealistic for most species, did find support for bottlenecks after population expansion, but this was only in one extinct (*V. bachmanii*) and all extant species. While machine learning is a powerful approach for modelling demographic history [50], it is possible that the recent loss of genetic
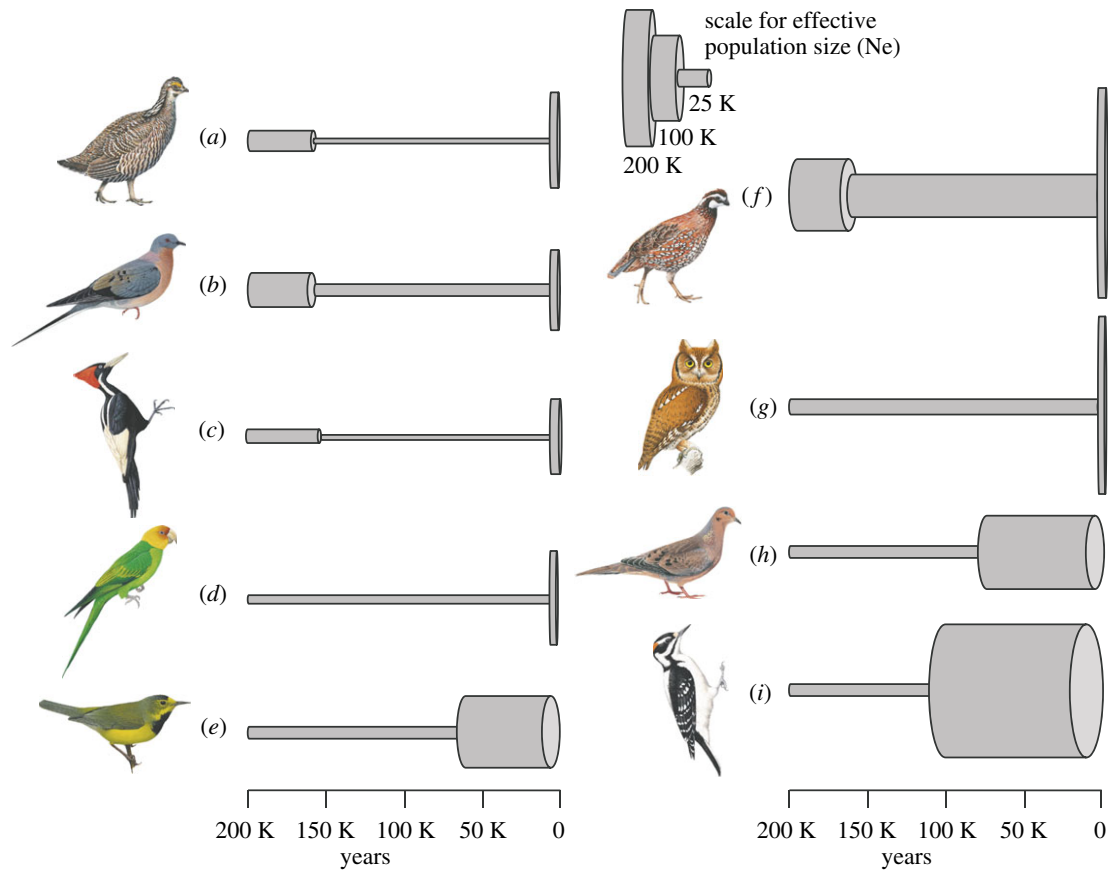
Figure 3. Schematic showing the most probable demographic model for each species. Pipe lengths are scaled to time in years assuming a fixed mutation rate and a taxon-specific generation time from the literature, and height is scaled to effective population size (see inset scale bar). The waiting time of each pipe was set to an arbitrary age of 200 K years (for the figure only) to more clearly show relative patterns among species. The species are as follows: (a) T. c. cupido; (b) E. migratorius; (c) C. principalis; (d) C. carolinensis; (e) V. bachmanii; (f) C. virginianus; (g) M. asio; (h) Z. macroura; and (i) D. villosus. Parameter values are available in electronic supplementary material, tables S7–S10. Bird illustrations reproduced by permission of Lynx Edicions. (Online version in colour.)

diversity progressed too far at the time of sampling, biasing the reconstruction of historical population dynamics.

Compared to other systems in which the demography of extinction has been examined (e.g. [19,22,25,59,60]), our results show a similar pattern of species-specific demographic trajectories. Within mammals and the birds studied here, there are commonly recovered genetic signatures such as population expansion. For the birds, this matches prior work, for example, a comparative study on Nearctic vertebrates including 13 extant bird species, showing that all underwent a population expansion during the Pleistocene [28]. That extinct birds as well as extant birds showed this pattern is perhaps not surprising given the presumed impact of glacial–interglacial cycles on species distributions, but it indicates that the species that were lost were on upward trajectories over evolutionary time scales prior to their dramatic declines. Despite support for common demographic modes in mammals and the eastern North American birds studied here, the timing and magnitude of demographic changes varies among taxa. For example, two of the extant species were estimated to have the earliest population expansions (Z. macroura and D. villosus) during the late Pleistocene, whereas the extinct species expanded at the end of the Pleistocene or in the Holocene.

(b) Comparison with prior genetic data

Our results are comparable to previously published mtDNA-based and whole-genome studies in our study species.

A study of D. villosus [61] and C. virginanus [62] found a pattern of population expansion comparable to our results. Previous work on T. c. cupido [63,64] and M. asio [65] recovered low genetic diversity in those taxa. The only species for which we could not compare to previously published data was Z. macroura, which was inferred to have undergone a population expansion. Whole-genome data of E. migratorius [42,43] and our data showed a comparable signature of the species harbouring low genetic diversity and evidence of population expansion. A similar approach applied to C. carolinensis found that the species expanded its population in the mid-Pleistocene and then underwent a subsequent population decline towards the present [41], a result that is only partially similar to ours. The whole-genome demographic analyses in E. migratorius and C. carolinensis provide a more detailed perspective of demographic fluctuations by showing how population size changes over time intervals, but we found similarities in our results and inferences in that neither approach found support for a continual genetic bottleneck during the Pleistocene. These overall consistencies across demographic histories estimated using different approaches for obtaining data and subsets of the genome suggest that inferences often hold even as datasets are supplanted by new techniques.

Genetic diversity was positively correlated with the geographical extent of the specimens that were sampled. Our sampling was limited by the available specimens collected at the turn of the twentieth century, but we chose

samples to maximize geographical spread. While the minimum-convex polygons produced from our sampling points do not represent the full geographical area of where each species occurred, they likely capture the relative differences in range sizes among taxa. The smaller sampling areas of extinct birds reflects their distribution in the decades leading up to their extinction. Among the last strongholds, *C. carolinensis* were in remnant tracts of intact bottomland forests and cypress swamps in Florida [66,67]. At the time of the extinction of *T. c. cupido* the taxon was restricted to a 226 km² island, Martha's Vineyard [30]. Both *T. c. cupido* and *C. virginanus* are open-habitat species, which may have caused them to have been impacted by Pleistocene climatic dynamics differently than the forest-dependent species. *Tympanuchus c. cupido* showed the lowest genetic diversity, whereas *C. virginanus* was among the genetically most variable taxa in our dataset. However, both species underwent population expansion during the LGM. The estimated Ne of the extant population of *T. c. cupido* was much larger than the known census size (Ne = 138 K versus Nc < 2 K [30]). The reason for the disparity could represent a lag-time in loss of diversity due to drift. The low variation in the DNA sequences, for example in *T. c. cupido*, could also limit the power of the model selection to accurately identify a particular model and estimate parameter values. An association between range size and diversity metrics is expected in neutrally evolving markers [68], and low genetic diversity and reduced range size may be inextricable factors in the demographic decline and extinction of species.

(c) Future directions

The causes of species extinctions are of great interest to both the general public and the research community. While historical investigations and distributional modelling have helped characterize some conditions leading to species extinction, DNA obtained from museum specimens can help complete the picture. As historical DNA techniques progress along with powerful statistical approaches for modelling spatial and temporal changes in genetic variation, there are numerous aspects of the genetics of extinct eastern North American birds

that can be investigated. Expanded geographical and genomic sampling will provide information at higher temporal and spatial resolution. More extensive sampling of protein-coding and regulatory regions of the genome will reveal how loci under selection were impacted by declining populations and whether deleterious mutations accrued. Population structure also merits further investigation. We found evidence of geographical structure in *E. migratorius* that is consistent with previously described mtDNA phylogeographic structure [43] and possible structure in *C. principalis* that may represent a genetic break or an artefact of uneven sampling [69]. Deep phylogeographic structure (i.e. reciprocally monophyletic groups) is generally uncommon within eastern North American birds [70], but other patterns like isolation-by-distance caused by geographically restricted gene flow may have structured genetic variation across their ranges. It is still unclear whether genomic techniques will satisfy the desire to understand the loss of these charismatic species. Nonetheless, genomic approaches are a powerful tool to provide detailed information on the evolutionary history of species that went extinct before their life histories and population ecologies could be fully understood.

References

1. Pimm SL, Jenkins CN, Abell R, Brooks TM, Gittleman JL, Joppa LN, Raven PH, Roberts CM, Sexton JO. 2014 The biodiversity of species and their rates of extinction, distribution, and protection. *Science* **344**, 1246752. (doi:10.1126/science.1246752)

2. Owens IP, Bennett PM. 2000 Ecological basis of extinction risk in birds: habitat loss versus human persecution and introduced predators. *Proc. Natl Acad. Sci. USA* **97**, 12 144–12 148. (doi:10.1073/pnas.200223397)

3. Purvis A, Gittleman JL, Cowlishaw G, Mace GM. 2000 Predicting extinction risk in declining species. *Proc. R. Soc. Lond. B* **267**, 1947–1952. (doi:10.1098/rspb.2000.1234)

4. Cardillo M, Mace GM, Jones KE, Bielby J, Bininda-Emonds OR, Sechrest W, Orme CD, Purvis A. 2005 Multiple causes of high extinction risk in large

mammal species. *Science* **309**, 1239–1241. (doi:10.1126/science.1116030)

5. Brook BW, Traill LW, Bradshaw CJ. 2006 Minimum viable population sizes and global extinction risk are unrelated. *Ecol. Lett.* **9**, 375–382. (doi:10.1111/j.1461-0248.2006.00883.x)

6. Hatfield JH, Orme CDL, Tobias JA, Banks-Leite C. 2018 Trait-based indicators of bird species sensitivity to habitat loss are effective within but not across data sets. *Ecol. Appl.* **28**, 28–34. (doi:10.1002/eap.1646)

7. Weeks BC, Gregory N, Naeem S. 2016 Bird assemblage vulnerability depends on the diversity and biogeographic histories of islands. *Proc. Natl Acad. Sci. USA* **113**, 10 109–10 114. (doi:10.1073/pnas.1603866113)

8. Frankham R. 2005 Genetics and extinction. *Biol. Conserv.* **126**, 131–140. (doi:10.1016/j.biocon.2005.05.002)

9. Pauls SU, Nowak C, Bálint M, Pfenninger M. 2013 The impact of global climate change on genetic diversity within populations and species. *Mol. Ecol.* **22**, 925–946. (doi:10.1111/mec.12152)

10. O'Brien SJ, Evermann JF. 1988 Interactive influence of infectious disease and genetic diversity in natural populations. *Trends Ecol. Evol.* **3**, 254–259. (doi:10.1016/0169-5347(88)90058-4)

11. O'Grady JJ, Brook BW, Reed DH, Ballou JD, Tonkyn DW, Frankham R. 2006 Realistic levels of inbreeding depression strongly affect extinction risk in wild populations. *Biol. Conserv.* **133**, 42–51. (doi:10.1016/j.biocon.2006.05.016)

12. Spielman D, Brook BW, Frankham R. 2004 Most species are not driven to extinction before genetic factors impact them. *Proc. Natl Acad. Sci. USA* **101**, 15 261–15 264. (doi:10.1073/pnas.0403809101)

13. De Vos JM, Joppa LN, Gittleman JL, Stephens PR, Pimm SL. 2015 Estimating the normal background rate of species extinction. *Conserv. Biol.* **29**, 452–462. (doi:10.1111/cobi.12380)

14. Rodríguez JP. 2002 Range contraction in declining North American bird populations. *Ecol. Appl.* **12**, 238–248. (doi:10.1890/1051-0761(2002)012[0238:RCIDNA]2.0.CO;2)

15. Laliberte AS, Ripple WJ. 2004 Range contractions of North American carnivores and ungulates. *BioScience* **54**, 123–138. (doi:10.1641/0006-3568(2004)054[0123:RCONAC]2.0.CO;2)

16. Rosenberg KV *et al.* 2019 Decline of the North American avifauna. *Science* **366**, 120–124. (doi:10.1126/science.aaw1313)

17. Barnett R *et al.* 2009 Phylogeography of lions (*Panthera leo* ssp.) reveals three distinct taxa and a late Pleistocene reduction in genetic diversity. *Mol. Ecol.* **18**, 1668–1677. (doi:10.1111/j.1365-294X.2009.04134.x)

18. Palkopoulou E *et al.* 2015 Complete genomes reveal signatures of demographic and genetic declines in the woolly mammoth. *Curr. Biol.* **25**, 1395–1400. (doi:10.1016/j.cub.2015.04.007)

19. Lorenzen ED *et al.* 2011 Species-specific responses of Late Quaternary megafauna to climate and humans. *Nature* **479**, 359–364. (doi:10.1038/nature10574)

20. Shapiro B *et al.* 2004 Rise and fall of the Beringian steppe bison. *Science* **306**, 1561–1565. (doi:10.1126/science.1101074)

21. Campos PF *et al.* 2010 Ancient DNA analyses exclude humans as the driving force behind late Pleistocene musk ox (*Ovibos moschatus*) population dynamics. *Proc. Natl Acad. Sci. USA* **107**, 5675–5680. (doi:10.1073/pnas.0907189107)

22. Stiller M *et al.* 2010 Withering away—25,000 years of genetic decline preceded cave bear extinction. *Mol. Biol. Evol.* **27**, 975–978. (doi:10.1093/molbev/msq083)

23. Barnett R, Yamaguchi N, Barnes I, Cooper A. 2006 The origin, current diversity and future conservation of the modern lion (*Panthera leo*). *Proc. R. Soc. B* **273**, 2119–2125. (doi:10.1098/rspb.2006.3555)

24. Barnes I, Shapiro B, Lister A, Kuznetsova T, Sher A, Guthrie D, Thomas MG. 2007 Genetic structure and extinction of the woolly mammoth, *Mammuthus primigenius*. *Curr. Biol.* **17**, 1072–1075. (doi:10.1016/j.cub.2007.05.035)

25. Debruyne R *et al.* 2008 Out of America: ancient DNA evidence for a new world origin of late quaternary woolly mammoths. *Curr. Biol.* **18**, 1320–1326. (doi:10.1016/j.cub.2008.07.061)

26. Comes HP, Kadereit JW. 1998 The effect of quaternary climate changes on plant distribution and evolution. *Trends Plant Sci.* **3**, 432–438. (doi:10.1016/S1360-1385(98)01327-2)

27. Kleman J, Jansson K, De Angelis H, Stroeven AP, Hättestrand C, Alm G, Glasser N. 2010 North American Ice sheet build-up during the last glacial cycle, 115–21 kyr. *Quat. Sci. Rev.* **29**, 2036–2051. (doi:10.1016/j.quascirev.2010.04.021)

28. Burbrink FT, Chan YL, Myers EA, Ruane S, Smith BT, Hickerson MJ. 2016 Asynchronous demographic responses to Pleistocene climate change in Eastern Nearctic vertebrates. *Ecol. Lett.* **19**, 1457–1467. (doi:10.1111/ele.12695)

29. Pimm SL, Askins RA. 1995 Forest losses predict bird extinctions in eastern North America. *Proc. Natl Acad. Sci. USA* **92**, 9343–9347. (doi:10.1073/pnas.92.20.9343)

30. Hume JP. 2017 *Extinct birds*, 2nd edn. London, UK: Bloomsbury Publishing.

31. Gross AO. 1928 The Heath Hen. *Memoirs Boston Soc. Nat. Hist.* **6**, 491–588.

32. McKinley D. 1980 The balance of decimating factors and recruitment in extinction of the Carolina Parakeet, Pt. 1. *Indiana Audubon Q.* **58**, 8–18.

33. Bucher EH. 1992 The causes of extinction of the Passenger Pigeon. *Curr. Ornithol.* **9**, 1–36. (doi:10.1007/978-1-4757-9921-7_1)

34. Blockstein DE. 2020 Passenger Pigeon (Ectopistes migratorius), version 1.0. In *Birds of the world* (eds AF Poole, FB Gill). Ithaca, NY: Cornell Lab of Ornithology.

35. Hamel PB. 2020. Bachman's Warbler (Vermivora bachmanii), version 1.0. In *Birds of the world* (ed. AF Poole). Ithaca, NY: Cornell Lab of Ornithology.

36. Jackson JA. 2020 Ivory-billed Woodpecker (*Campephilus principalis*), version 1.0. In *Birds of the world* (eds AF Poole, FB Gill). Ithaca, NY: Cornell Lab of Ornithology.

37. Johnson JA, SchroederMA, Robb LA. 2020 Greater Prairie-Chicken (*Tympanuchus cupido*), version 1.0. In *Birds of the world* (ed. AF Poole). Ithaca, NY: Cornell Lab of Ornithology.

38. Snyder NF, Russell K. 2020 Carolina Parakeet (*Conuropsis carolinensis*), version 1.0. In *Birds of the world* (eds AF Poole, FB Gill). Ithaca, NY: Cornell Lab of Ornithology.

39. Hung CM, Zink RM, Li SH. 2018 Can genomic variation explain the extinction of the passenger pigeon? *J. Avian Biol.* **49**, e01858. (doi:10.1111/jav.01858)

40. Guiry EJ, Orchard TJ, Royle TC, Cheung C, Yang DY. 2020 Dietary plasticity and the extinction of the passenger pigeon (*Ectopistes migratorius*). *Quat. Sci. Rev.* **233**, 106225. (doi:10.1016/j.quascirev.2020.106225)

41. Gelabert P *et al.* 2020 Evolutionary history, genomic adaptation to toxic diet, and extinction of the Carolina Parakeet. *Curr. Biol.* **30**, 108–114. (doi:10.1016/j.cub.2019.10.066)

42. Hung CM, Shaner PJL, Zink RM, Liu WC, Chu TC, Huang WS, Li S-H. 2014 Drastic population fluctuations explain the rapid extinction of the passenger pigeon. *Proc. Natl Acad. Sci. USA* **111**, 10 636–10 641. (doi:10.1073/pnas.1401526111)

43. Murray GG *et al.* 2017 Natural selection shaped the rise and fall of passenger pigeon genomic diversity. *Science* **358**, 951–954. (doi:10.1126/science.aao0960)

44. Faircloth BC, McCormack JE, Crawford NG, Harvey MG, Brumfield RT, Glenn TC. 2012 Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. *Syst. Biol.* **61**, 717–726. (doi:10.1093/sysbio/sys004)

45. R Core Team. 2019 *R: a language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. See http://www.R-project.org/.

46. Jombart T. 2008 adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**, 1403–1405. (doi:10.1093/bioinformatics/btn129)

47. Frichot E, Mathieu F, Trouillon T, Bouchard G, François O. 2014 Fast and efficient estimation of individual ancestry coefficients. *Genetics* **196**, 973–983. (doi:10.1534/genetics.113.160572)

48. Skotte L, Korneliussen TS, Albrechtsen A. 2013 Estimating individual admixture proportions from next generation sequencing data. *Genetics* **195**, 693–702. (doi:10.1534/genetics.113.154138)

49. Gehara M, Mazzochinni GG, Burbrink FT. 2020 PipeMaster: inferring population divergence and demographic history with approximate Bayesian computation and supervised machine-learning in R. *bioRxiv*. (doi:10.1101/2020.12.04.410670)

50. Schrider DR, Kern AD. 2018 Supervised machine learning for population genetics: a new paradigm. *Trends Genet.* **34**, 301–312. (doi:10.1016/j.tig.2017.12.005)

51. Gehara M *et al.* 2017 Estimating synchronous demographic changes across populations using hABC and its application for a herpetological community from northeastern Brazil. *Mol. Ecol.* **26**, 4756–4771. (doi:10.1111/mec.14239)

52. Burbrink FT, Gehara M. 2018 The biogeography of deep time phylogenetic reticulation. *Syst. Biol.* **67**, 743–755. (doi:10.1093/sysbio/syy019)

53. BirdLife International. IUCN Red List for birds. See www.birdlife.org/redlist.

54. Smith BT, Harvey MG, Faircloth BC, Glenn TC, Brumfield RT. 2014 Target capture and massively parallel sequencing of ultraconserved elements for comparative studies at shallow evolutionary time scales. *Syst. Biol.* **63**, 83–95. (doi:10.1093/sysbio/syt061)

55. Chollet F. 2015 Keras. See https://github.com/fchollet/keras.

56. MacCleery DW. 1992 *American forests: a history of resiliency and recovery*. Washington, DC: USDA.

57. Remsen JV. 1986 Was Bachman's warbler a bamboo specialist? *Auk* **103**, 216–219.

58. Nadachowska-Brzyska K, Li C, Smeds L, Zhang G, Ellegren H. 2015 Temporal dynamics of avian populations during Pleistocene revealed by whole-genome sequences. *Curr. Biol.* **25**, 1375–1380. (doi:10.1016/j.cub.2015.03.047)

59. Allentoft ME, Heller R, Oskam CL, Lorenzen ED, Hale ML, Gilbert MTP, Jacomb C, Holdaway RN, Bunce M. 2014 Extinct New Zealand megafauna were not in decline before human colonization. *Proc. Natl Acad. Sci. USA* **111**, 4922–4927. (doi:10.1073/pnas.1314972111)

60. MacPhee RD, Tikhonov AN, Mol D, Greenwood AD. 2005 Late Quaternary loss of genetic diversity in

muskox (Ovibos). *BMC Evol. Biol.* **5**, 49. (doi:10.1186/1471-2148-5-49)

61. Klicka J, Spellman GM, Winker K, Chua V, Smith BT. 2011 A phylogeographic and population genetic analysis of a widespread, sedentary North American bird: the Hairy Woodpecker (*Dryobates villosus*). *Auk* **128**, 346–362. (doi:10.1525/auk.2011.10264)

62. Eo SH, Wares JP, Carroll JP. 2010 Subspecies and units for conservation and management of the northern bobwhite in the eastern United States. *Conserv. Genet.* **11**, 867–875. (doi:10.1007/s10592-009-9926-9)

63. Palkovacs EP, Oppenheimer AJ, Gladyshev E, Toepfer JE, Amato G, Chase T, Caccone A. 2004 Genetic evaluation of a proposed introduction: the case of the greater prairie chicken and the extinct heath hen. *Mol. Ecol.* **13**, 1759–1769. (doi:10.1111/j.1365-294X.2004.02181.x)

64. Johnson JA, Dunn PO. 2006 Low genetic variation in the heath hen prior to extinction and implications for the conservation of prairie-chicken populations. *Conserv. Genet.* **7**, 37–48. (doi:10.1007/s10592-005-7856-8)

65. Proudfoot GA, Gehlbach FR, Honeycutt RL. 2007 Mitochondrial DNA variation and phylogeography of the Eastern and Western screech-owls. *Condor* **109**, 617–627. (doi:10.1093/condor/109.3.617)

66. McKinley D. 1985 *The Carolina Parakeet in Florida*. Gainesville, FL: Florida Ornithological Society.

67. Burgio KR, Carlson CJ, Tingley MW. 2017 Lazarus ecology: Recovering the distribution and migratory patterns of the extinct Carolina parakeet. *Ecol. Evol.* **7**, 5467–5475. (doi:10.1002/ece3.3135)

68. Frankham R. 1996 Relationship of genetic variation to population size in wildlife. *Conserv. Biol.* **10**, 1500–1508. (doi:10.1046/j.1523-1739.1996.10061500.x)

69. Puechmaille SJ. 2016 The program structure does not reliably recover the correct population structure when sampling is uneven: subsampling and new estimators alleviate the problem. *Mol. Ecol. Resour.* **16**, 608–627. (doi:10.1111/1755-0998.12512)

70. Smith BT, Seeholzer GF, Harvey MG, Cuervo AM, Brumfield RT. 2017 A latitudinal phylogeographic diversity gradient in birds. *PLoS Biol.* **15**, 4. (doi:10.1371/journal.pbio.2001073)

9

royalsocietypublishing.org/journal/rspb   *Proc. R. Soc. B* **288**: 20201945