PROCEEDINGS B

royalsocietypublishing.org/journal/rspb

Research

Cite this article: Brüniche-Olsen A, Kellner KF, Belant JL, DeWoody JA. 2021 Life-history traits and habitat availability shape genomic diversity in birds: implications for conservation. *Proc. R. Soc. B* **288**: 20211441.
https://doi.org/10.1098/rspb.2021.1441

Received: 24 June 2021
Accepted: 6 October 2021

Subject Category:
Evolution

Subject Areas:
biological applications, genomics, evolution

Keywords:
environmental niche modelling, demographic history, whole-genome sequencing, PSMC, Aves, IUCN

Author for correspondence:
Anna Brüniche-Olsen
e-mail: anna.bruniche-olsen@bio.ku.dk

†These authors contributed equally to this study.

Electronic supplementary material is available online at https://doi.org/10.6084/m9.figshare.c.5672373.

THE ROYAL SOCIETY PUBLISHING

Life-history traits and habitat availability shape genomic diversity in birds: implications for conservation

Anna Brüniche-Olsen[1,†], Kenneth F. Kellner[2,†], Jerrold L. Belant[2] and J. Andrew DeWoody[3,4]

[1]Section for Computational and RNA Biology, Department of Biology, University of Copenhagen, 2200 KBH N Copenhagen, Denmark
[2]Global Wildlife Conservation Center, State University of New York College of Environmental Science and Forestry, Syracuse, NY 13210, USA
[3]Department of Forestry and Natural Resources, and [4]Department of Biological Sciences, Purdue University, West Lafayette, IN 47905, USA

AB-O, 0000-0002-3364-2064; KFK, 0000-0002-6755-0555; JLB, 0000-0001-7021-1338; JAD, 0000-0002-7315-5631

More than 25% of species assessed by the International Union for Conservation of Nature (IUCN) are threatened with extinction. Understanding how environmental and biological processes have shaped genomic diversity may inform management practices. Using 68 extant avian species, we parsed the effects of habitat availability and life-history traits on genomic diversity over time to provide a baseline for conservation efforts. We used published whole-genome sequence data to estimate overall genomic diversity as indicated by historical long-term effective population sizes (N_e) and current genomic variability (H), then used environmental niche modelling to estimate Pleistocene habitat dynamics for each species. We found that N_e and H were positively correlated with habitat availability and related to key life-history traits (body mass and diet), suggesting the latter contribute to the overall genomic variation. We found that H decreased with increasing species extinction risk, suggesting that H may serve as a leading indicator of demographic trends related to formal IUCN conservation status in birds. Our analyses illustrate that genome-wide summary statistics estimated from sequence data reflect meaningful ecological attributes relevant to species conservation.

1. Introduction

During the Anthropocene, an unprecedented rate of species extinctions has occurred leading to what has been termed 'the sixth mass extinction' [1]. The International Union for Conservation of Nature (IUCN) estimates that more than 25% of assessed species are threatened with extinction [2]. Since the Industrial Revolution, threatened species have experienced about 25% average declines in population sizes and genetic diversity per species has declined 6% [3]. These losses are of long-term conservation concern, as genomic diversity forms the basis for the evolutionary potential of species and is essential for species persistence [4–6]. Therefore, understanding how intrinsic (i.e. biological) and extrinsic (i.e. environmental) factors have shaped genomic diversity and why genomic diversity differs among species may provide insights for future demographic trends. This information could be used in conservation to identify species at risk of decline by integrating genetics or genomics in IUCN Red List assessments [7–9].

Environmental changes have large effects on species' distribution and abundance [10]. Periodical retraction and expansion of habitat during the Pleistocene caused recurrent bottlenecks leading to a loss of genomic diversity [11,12], forcing species to migrate, adapt or go extinct [13,14]. These environmental

© 2021 The Author(s) Published by the Royal Society. All rights reserved.

changes varied in speed, duration and extent, generally resulting in population declines during glacial periods and population expansions during interglacial periods [15–18]. Understanding broad patterns between habitat and genomic diversity is critical for conservation in a changing climate [11]. Recent advances in environmental niche modelling (ENM; i.e. [19]) with the availability of global climate datasets for multiple historical time periods (i.e. [20]) have enabled researchers to assess the availability of suitable habitat for species under past climate conditions. Whole-genome sequencing and coalescent-based approaches have facilitated inferring effective population size through time based on a single genome [21]. Combining these two approaches have proved informative for explaining historic fluctuations in genomic diversity for a range of species [18,22].

Life-history traits such as diet and body mass strongly influence the genomic diversity of species [5,23,24]. For instance, smaller bodied bird species are more threatened by habitat loss and modification [25], whereas larger bodied species are more at risk of extinction from direct effects from humans [26]. Species at higher trophic levels (e.g. carnivores) tend to have lower genomic diversity and are more sensitive to environmental disturbances than species at lower trophic levels (e.g. herbivores) [27–30]. Therefore, including life-history information when assessing species demographic trajectories and habitat availability may provide a more complete picture of which factors shape genomic diversity [17].

Herein, we examined how the extent of suitable habitat and life-history traits have affected past and present genomic diversity in 68 bird species. Birds are well suited for studying effects of environmental change on genomic diversity as they have colonized most of the world's surface, adapted to a wide range of niches [15,31,32], and exhibit remarkable variation in life-history traits (e.g. diet and body size) [33,34]. We used whole-genome sequencing data to infer historical effective population size (N_e) through time and estimated current heterozygosity (H). We used ENM to estimate suitable habitat area for each species at multiple periods during the Pleistocene. We predicted that N_e and H would: (i) be positively correlated with available habitat, (ii) decrease with increasing body size, and (iii) decrease with increasing tropic level. We further predicted that species categorized as threatened by the IUCN Red List would have lower contemporary genomic diversity (H) than non-threatened species. Our aim was that the results would identify how different extrinsic (suitable habitat) and intrinsic (diet, body size) factors impact genomic diversity and may ultimately be used as a reference to inform conservation.

2. Results and discussion

Our dataset included 68 avian species, representing 60 families and 41 orders. Dietary breadth was represented by frugivores and/or nectarivores (12%), herbivores/granivores (24%), insectivores (31%), carnivores (24%) and omnivores (10%; electronic supplementary material, S1). Body mass varied markedly among species with range: 4.40 g – 1.09×10^5 g (mean = 3407 g; electronic supplementary material, S1). For each species, we used a single whole-genome sequence assembly and its associated whole-genome sequencing data to estimate genome-wide present-day H and to reconstruct historical long-term N_e. These metrics are complementary as H is the current proportion of heterozygous genotypes in the sample and thus the standing genomic variation. N_e corresponds to the size of an idealized population that accumulates the same amount of genetic drift as the population under study and is a function of the intrinsic mutation rate and reflects historical demographic processes. Contemporary H was estimated with ANGSD [35] and varied among species with range 1.03×10^{-4}–9.23×10^{-3} (mean 2.89×10^{-3}) (electronic supplementary material, S2). Trajectories of changes in N_e through time for all 68 species were inferred using pairwise sequentially Markovian coalescent (PSMC) [21] (electronic supplementary material, S3). We calculated mean N_e for each species over the past 1 Myr, which varied among species with range 1.01–7.60×10^4 (mean 1.34×10^3) (electronic supplementary material, S2). To address our hypotheses that H and N_e would be related to suitable habitat and species traits, we used multiple regressions fitted with phylogenetic generalized least squares to model H and mean N_e as functions of present-day conservation status, one extrinsic factor (mean area of suitable habitat) and several intrinsic factors (life-history traits) (electronic supplementary material, S4). Estimated R^2 (the combined explanatory power of all covariates included in a multiple regression model) was 0.45 for the model of H and 0.44 for the model of N_e. Estimated Pagel's lambda (a measure of phylogenetic signal) was 0.51 for the model of H and 0.54 for the model of N_e.

(a) Threatened bird species have lower genomic diversity

The change to modern industrial production 200 years ago led to extensive population declines for threatened species and loss of genetic diversity across all IUCN Red List groups [2]. We analysed the relationship between genome-wide summary statistics estimates based on genomic sequence data and contemporary conservation status (figure 1). We found that Critically Endangered and Endangered species had smaller H than Least Concern and Near-Threatened species (both $p < 0.01$). Similarly, Vulnerable species had smaller H than Least Concern ($p = 0.04$) and Near-Threatened ($p = 0.02$) species. By contrast, mean N_e was similar among Red List categories, except for a difference between Least Concern and Near-Threatened species ($p = 0.03$). The stronger relationship between conservation status and H, relative to conservation status and N_e, is probably because H is a leading indicator of demographic change that is a function of the breeding system, whereas N_e is a lagging indicator based on the population mutation rate [36]. Because $N_e = \theta/4\mu$ when θ corresponds to genetic diversity and μ to mutation rate [37], changes in N_e between species are largely driven by the mutation rate which is typically on the order of 5×10^{-9} per generation in birds [38]. By contrast, H responds more quickly than N_e. If allele frequencies at a locus are represented by p and q, and the inbreeding coefficient by F, in a randomly mating population, we would expect $H = 2pq$ and with inbreeding (e.g. owing to a bottleneck) expect $H = 2pq(1 - F)$ [39]. With equally frequent alleles, this means that H could be 0.5 but is reduced to 0.25 in a single generation with $F = 0.5$ or to 0 if $F = 1.0$. Thus, we suggest contemporary H can benefit conservation status assessments by quantifying evolutionary potential as well as providing key insights into species listed as Data Deficient by IUCN. For

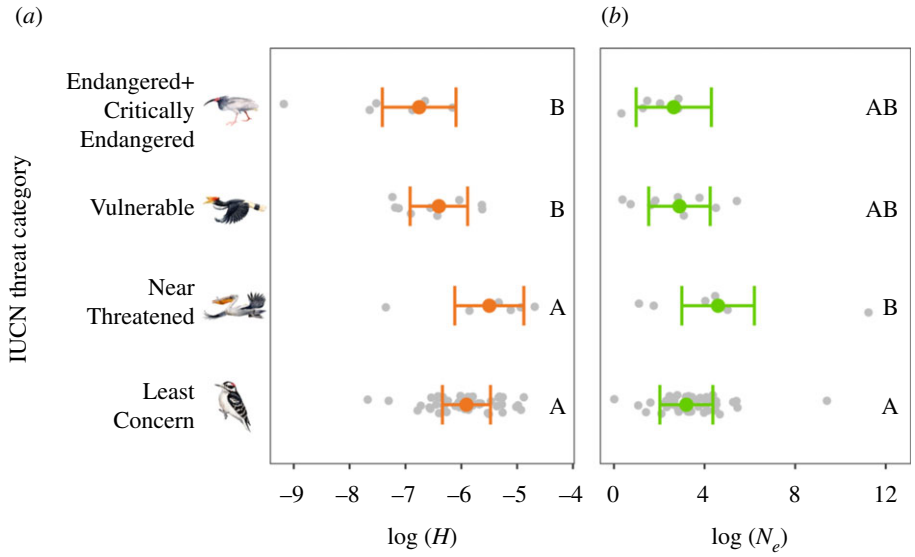

Figure 1. Differences in heterozygosity (H) (*a*) and effective population size (N_e) (*b*) among IUCN threat categories based on estimates from separate multiple regression models that also included effects of habitat area, diet and body mass. Coloured points and error bars represent model predictions and 95% confidence intervals; raw data points are in grey. IUCN categories that do not share a letter are significantly different from each other based on post hoc Tukey's honestly significant difference tests. Drawings are example birds from each Red List category. Illustrated species are (top to bottom) crested ibis (*Nipponia nippon*), rhinoceros hornbill (*Buceros rhinoceros*), Dalmatian pelican (*Pelecanus crispus*) and downy woodpecker (*Picoides pubescens*). (Online version in colour.)

instance, we found that 66% of species with $H < 9 \times 10^{-4}$ were categorized by IUCN as Vulnerable or Endangered; thus, a Data Deficient species with H below that threshold would probably be of conservation concern even if detailed data on its abundance or distribution were not available (e.g. a previously undescribed baleen whale species (*Balaenoptera* spp. [40]).

(b) Genomic diversity is driven by habitat availability over time

We used ENMs [19] to infer suitable habitat for each species over the past 1 Myr and examined how the extent of past and present suitable habitat was related to estimates of N_e and H, respectively (electronic supplementary material, S2 and S3). We found a positive relationship between mean N_e and mean extent of suitable habitat for each species over the past approximately 1 Myr (figure 2). This implies that species with a historically greater mean area of suitable habitat also tended to have an overall larger mean N_e. Together, ENM and N_e analyses have proved informative for explaining historic fluctuations in genomic diversity for a range of species [18,22]. Researchers have identified correlations between habitat availability and effective population size in bats [17] and birds [16,41], but in megafauna [22], the responses vary idiosyncratically with species without a clear pattern. This suggests that habitat is not the only limiting factor for N_e, but rather acts in synergy with other factors, such as life-history traits.

Habitat area is a main factor for species persistence [42] and essential for maintaining viable populations [43,44]. We found a positive relationship between the amount of present-day suitable habitat and H (figure 2). This relationship was expected as greater extents of suitable habitat can support larger populations, which in turn can harbour more genomic diversity [25,45]. The species studied here represent diverse demographic trajectories. Some experienced multiple N_e reductions and expansions (e.g. great spotted kiwi (*Apteryx haasti*), Anna's hummingbird (*Calypte anna*) and red-legged seriema (*Cariama cristata*)), whereas the demographic history of other unrelated species experienced a more continuous decline in N_e (e.g. rhinoceros hornbill (*Buceros rhinoceros*), white-tailed eagle (*Haliaeetus albicilla*) and crested ibis (*Nipponia nippon*)) (figure 3; electronic supplementary material, S3). The duration of the population reduction or expansion is important for the rate of zygosity change. Whereas short bottlenecks have little effect on heterozygosity [46], recurrent bottlenecks can lead to genomic erosion [47]. Our results illustrate that comparing point estimates (i.e. current H and current available habitat) is informative for assessing species viability and may complement temporal sampling of H and available habitat when detecting recent changes [48].

(c) Life-history traits are determinants of genomic diversity

Life-history traits shape species genomic diversity and we can make predictions about a species level of genomic diversity based on these traits (e.g. body mass, diet, fecundity and generation turnover) [49]. We expected life-history traits to impact short- and long-term genomic diversity as they probably remained relatively constant over the past approximately 1 Myr. We measured correlation between N_e and H with body mass, and diet. Body mass was negatively correlated with N_e and H (figure 4), confirming and extending earlier work [5,25,49–52]. Diet also had a significant effect on genomic diversity, with carnivores/scavengers on average having the lowest N_e and H, and herbivores/granivores the highest (figure 4). Of the life-history traits we considered, the effects of diet on genomic diversity were generally the largest (electronic supplementary material, S4). Carnivores normally live at lower population densities than non-carnivores and their specialized diet is typically less abundant than generalist omnivores. Overall, our life-history trait results suggest that when assessing species genomic variation, using species of

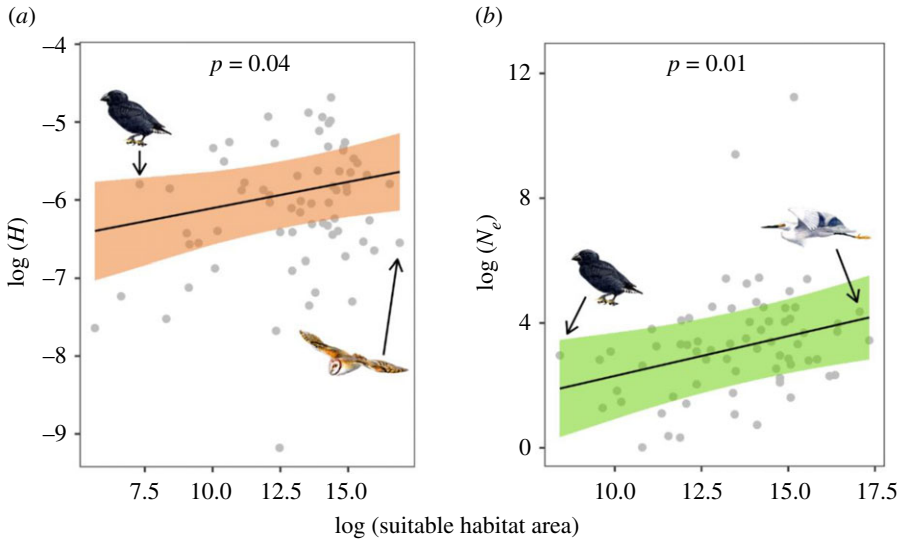

Figure 2. Effect of suitable habitat area on heterozygosity (H) (*a*) and effective population size (N_e) (*b*) per species based on output from separate multiple regression models that also included effects of IUCN threat category, diet and body mass. Lines and shaded regions represent model predictions and 95% confidence intervals; raw data points are in grey. Drawings show examples of bird species with suitable habitat area near the minimum and maximum observed. Illustrated species are (left to right) medium ground finch (*Geospiza fortis*), barn owl (*Tyto alba*) and little egret (*Egretta garzetta*). (Online version in colour.)

similar body mass and diet as reference provide context for assessing levels of genomic diversity.

(d) Genomics as a conservation tool

Our study adds to the growing body of evidence which indicates that parameter estimates derived from whole-genome DNA sequence data can add a key aspect of biological diversity to holistic conservation assessments. Several parameters have been proposed as tools for conservation assessment (i.e. temporal sampling of current and past diversity [48], estimates of genomic load [53], runs of homozygosity [51,54], functional genomic variation [55]), and there are ongoing efforts to identify the most applicable genomic diversity measure(s). We do not suggest that conventional metrics (e.g. reproductive rate, demographic trends, conventional genetic markers, etc.) should be abandoned, but that genome-wide summary statistics provide a valuable contribution that captures key evolutionary and ecological attributes. For example, our results illustrate that based on estimates of current genomic diversity (H) one can make predictions about the Red List status (i.e. Threatened or Non-Threatened) of a species. This may be especially important for Data Deficient or newly discovered species that lack demographic data. As genomic data becomes increasingly available, we suggest there is no reason other than historical inertia to omit them from conservation decisions whether by international, national or local authorities.

3. Material and methods

(a) Samples and mapping of sequence data

We downloaded paired-end (PE) reads and genome assemblies for 68 bird species from EMBL-EBI and NCBI (electronic supplementary material, S5). All reads were processed with Trimgalore [56], which removes adaptor sequences and trims consecutive stretches of low-quality bases from the ends of the reads. Fastqc [57] was used to generate summary statistics for the reads and check that adaptor sequences were removed. Genome assemblies were indexed with BWA v. 0.7.15 using the 'bwtsw' function and reads mapped using the 'mem' function using default settings [58]. Validation of sam files and duplicate marking of bam files were done with Picard-Tools v. 2.9.0 (http://picard.sourceforge.net.). Samtools v. 1.4 was employed throughout for merging of bam files, estimation of depth of coverage (DOC), and to check that the PE insert sizes were large enough that there was no overlap between read pairs [59]. Local realignment and duplicate reads removal were carried out with GATK v. 3.8.0 [60].

We used Repeatmasker v. 4.0.7 [61] to mask repeated regions in the genome assemblies using the chicken as a reference. Genmap [62] was used to estimate mappability for each region using kmer 100 allowing for two errors. Regions with repeats, with mappability less than 1, and scaffolds shorter than 10 kb in length were excluded from downstream analyses (electronic supplementary material, S6). Identifying sex-linked chromosomes in scaffold-level assemblies is difficult and population-level data is often required to obtain robust results [63]. For consistency across species, we retained all scaffolds for all species. Bedtools v. 2.29.0 [64] were used throughout for manipulating bed files.

(b) Heterozygosity and effective population size

For each species, we quantified genome-wide heterozygosity (H) based on the site frequency spectrum using ANGSD [35]. This H measures the proportion of heterozygous genotypes divided by the genome size while excluding sites of low quality. We used filters on the base quality score (-minQ 20), minimum mapping quality (-minMapQ 30) and mapping depth (d) setting minimum mapping depth to one-third of the mean total mapping depth (-setMinDepth $d/3$), and maximum mapping to double (-setMaxDepth $d*2$) the mean total mapping depth. We traced historical effective population size (N_e) changes through time using PSMC [21]. This method models coalescent events between haplotypes in a single genome to infer demographic trajectories. The PSMC timescale is determined by the generation time so the timespan surveyed will be more recent for short-lived species than for long-lived species. The method was originally developed for chromosome level sequences [21], but has been shown to provide robust estimates for scaffold-level genome assemblies [41,65]. Furthermore, PSMC produces similar shaped trajectories

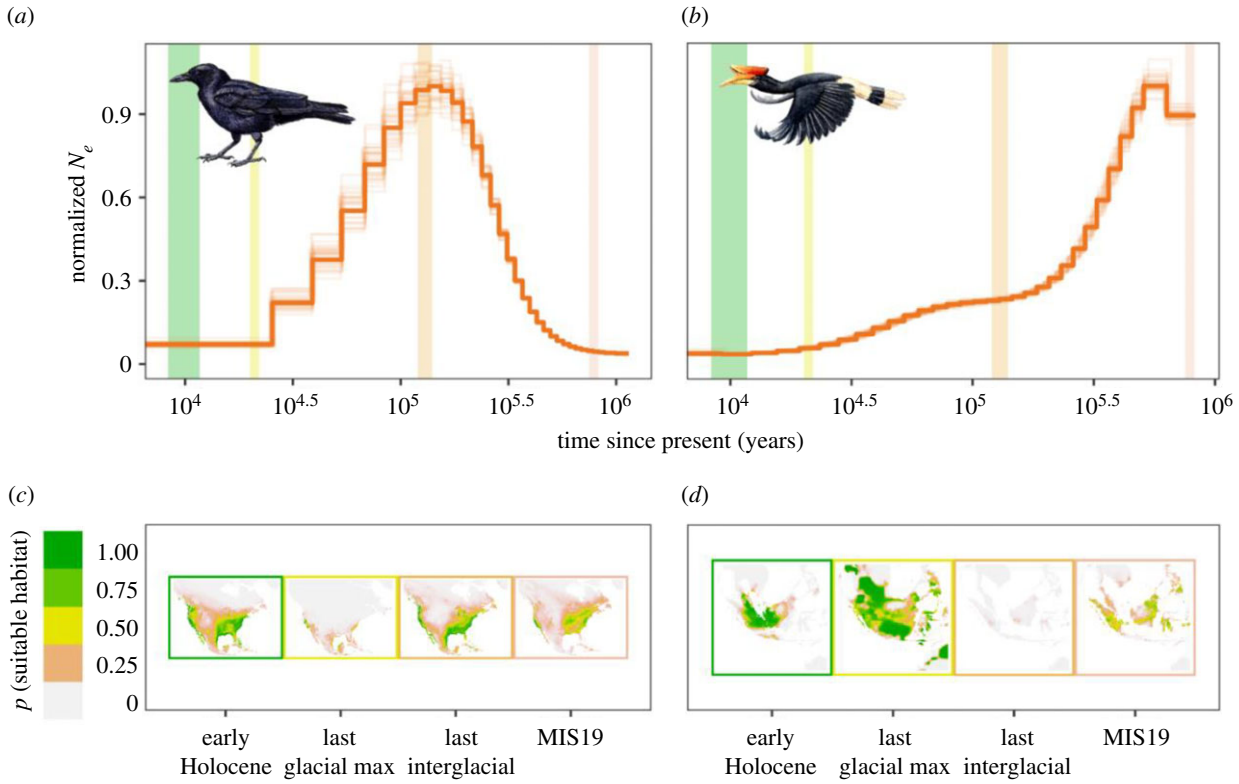

Figure 3. Comparison of effective population size (N_e) trajectories over time with maps of suitable habitat area over the past approximately 1 Myr for two avian species (American crow (*Corvus brachyrhynchos*), (*a*,*c*), and rhinoceros hornbill (*Buceros rhinoceros*), (*b*,*d*)). N_e trajectories (dark orange lines) were obtained using PSMC and normalized by dividing all values by the maximum value. Light orange lines represent bootstrap PSMC replicates. Coloured vertical bars on the trajectories correspond to the four historical time periods represented in the maps, which have corresponding coloured borders. Estimates of past suitable habitat were obtained using ENM. Note that the PSMC parametrization (e.g. generation time, substitution rate) will shift the N_e trajectories vertically, and that recent estimates of N_e from PSMC tend to be noisy (i.e. there is more variance in N_e in shallow time) owing to fewer coalescent events. See the electronic supplementary material, S3 for similar plots for all 68 species analysed in this paper. (Online version in colour.)

when including/excluding sex chromosome scaffolds [18,66], although including sex chromosomes may lead to a slight underestimation of N_e [4].

We used SAMTOOLS to estimate mean DOC of the bam file and sites with less than 1/3× or greater than 2× the average DOC and with a root-mean-square less than 25 were culled. BCFTOOLS [67] were used to call single nucleotide polymorphisms and vcfutils.pl were used to convert the variant call format to masked fasta format. After exploring different PSMC parameter settings, we used the following $-t$ 5 $-r$ 5 $-p$ '4 + 30 * 2 + 4 + 6 + 10'. Here t corresponds to the upper limit for the most recent common ancestor, r is the ratio of the scaled mutation rate (θ) and the recombination rate (ρ), and p is the atomic time interval parameter that specifies the population size parameters. The first population size parameter spans four (4 + 25 * 2 + 4 + 6 + 10) atomic time intervals, each of the next 25 parameters (4 + 25 * 2 + 4 + 6 + 10) span two atomic time intervals, while the last three parameters span four, six and 10 (4 + 25 * 2 + 4 + 6 + 10) atomic time intervals. We used a minimum DOC of 10x [41] and a max DOC of 2× the mean DOC as suggested by https://github.com/lh3/psmc, and checked that greater than or equal to 10 recombination events took place in each interval after the twentieth iteration [21]. For each species, we performed 30 bootstraps.

To convert PSMC trajectories from generations to years, we used estimates of generation time (g) as twice the age of sexual maturity (electronic supplementary material, S1), which is positively associated with age-specific rates of reproductive output and survival in birds [68]. Estimations of avian mutation rates have been described earlier [15,31] using 8295 orthologues regions and the corresponding coding sequences alignments. We used these rates as priors for μ in our analyses. Comparative nuclear sequence data from birds have shown that within-clade substitution rates are more similar than among-clade substitution rates [69–71]. Thus, where a species-specific μ was unavailable, we used the mean μ for each order as a proxy for substitution rate for all species within that order. For orders not represented in Nadachowska-Brzyska [15], we used μ from the closest relative(s) order based on a phylogeny constructed with TIMETREE [72] (electronic supplementary material, S7). If there were no records of the species in TIMETREE, we used the mean of the genus as a proxy for generation time (electronic supplementary material, S8).

After obtaining PSMC trajectories, we calculated mean N_e for each species over the entire time period (approximately 1 Myr). Owing to a lower number of coalescent events very recent estimates of N_e from PSMC tend to be noisy, indicated by fewer changes in N_e over time. Therefore, following Leroy *et al.* [73], we excluded the four most recent timepoints of N_e for each species when calculating the mean estimate to increase reliability. Note that we show the complete PSMC trajectories (i.e. without excluding the most recent four timepoints) in figure 3 and the electronic supplementary material, S3.

(c) Environmental niche modelling

To assess how past changes in N_e were influenced by available habitat, we developed ENMs for the breeding and/or year-round range of each species. First, we obtained presence data for each species from the Global Biodiversity Information Facility (www.gbif.org), which maintains a database of species occurrences from multiple sources including eBird [74,75]. We downloaded up to 10 000 georeferenced records for each species

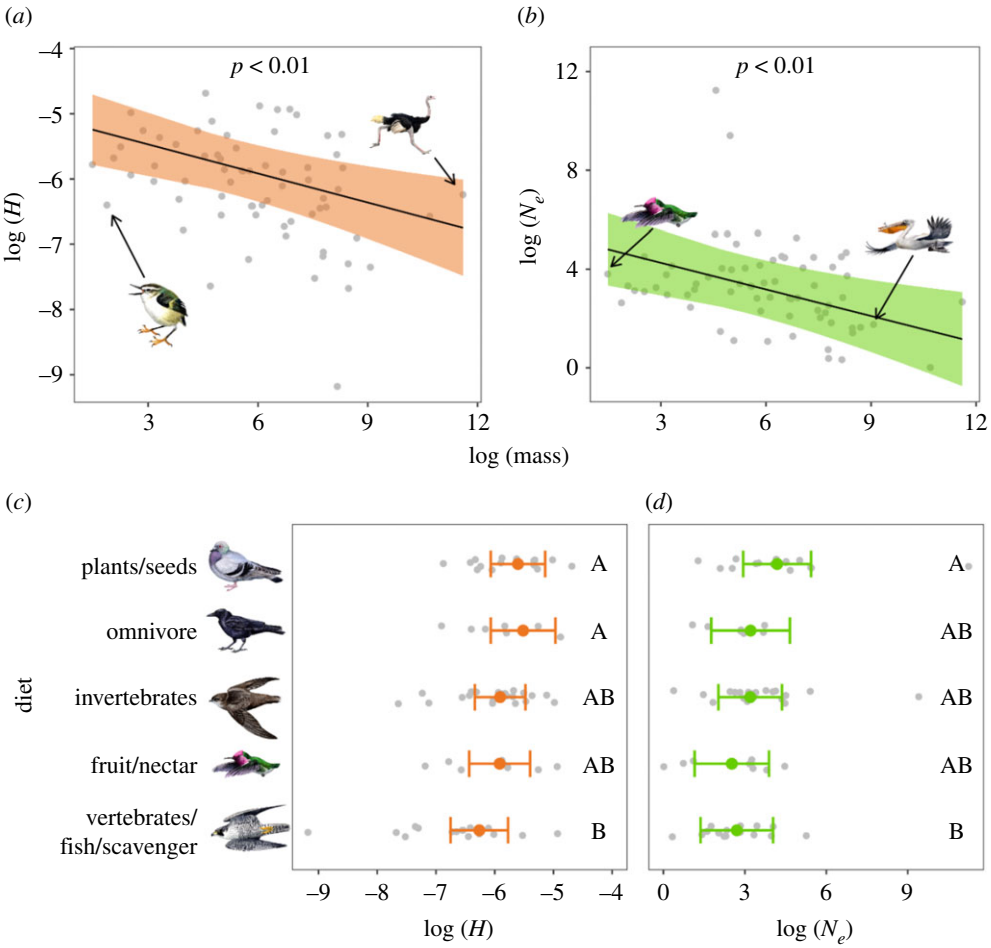

Figure 4. Effect of life-history traits on heterozygosity (H) and mean effective population size (N_e) based on output from separate multiple regression models that also included effects of IUCN threat category and habitat area; (*a*,*b*) shows the relationship between body mass and H or N_e; lines and shaded regions represent model predictions and 95% confidence intervals (CIs). Illustrated species are (left to right) rifleman (*Acanthisitta chloris*), common ostrich (*Struthio camelus*), Anna's hummingbird (*Calypte anna*) and Dalmatian pelican (*Pelecanus crispus*); (*c*,*d*) shows the relationship between diet and H/N_e; coloured points and error bars represent model predictions with 95% CIs. Categories that do not share a letter are significantly different ($p < 0.05$) from each other based on post hoc Tukey's honestly significant difference tests. In all plots, raw data points are shown in grey. Drawings show examples of bird species with body mass near the minimum and maximum observed and from each Red List category. Illustrated species are (top to bottom) rock dove (*Columba livia*), American crow (*Corvus brachyrhynchos*), chimney swift (*Chaetura pelagica*), Anna's hummingbird (*Calypte anna*) and peregrine falcon (*Falco peregrinus*). (Online version in colour.)

using the R package 'rgbif' [76]. We excluded records that occurred outside the breeding or year-round range of the species, based on distribution maps provided by BirdLife International [77]. If there were more than 3000 remaining records, we subsampled the records to 3000; otherwise, we kept all of them. Species with fewer than 15 records available were excluded from further analysis. Of the included species, 84% (57 out of 68) had more than 50 available records and the average number of records per species was 1549.

We obtained present-day rasters for 19 bioclimatic variables (e.g. annual mean temperature, annual precipitation) from WorldClim [20]. These rasters span the entire globe at 2.5 min resolution. We cropped the rasters to the occurrence records for a species and added a buffer of 10 degrees latitude and longitude. We performed principal component analysis on the 19 variables after normalizing each and used the top six principal components for modelling [17]. Using the occurrence records and principal components derived from the bioclimatic variables, we fitted a suite of ENMs for each species using Maxent [78] via the R [79] package ENMeval [80]. We fitted models with all combinations of linear, quadratic and product feature classes and three possible regularization multipliers: 1, 2 and 5. We used 10 000 random background points for each model. Model fit was evaluated using a random k-fold approach with fourfolds. We selected the model with the lowest Akaike information criterion with a correction for small sample sizes for further analysis [81].

Using our fitted top model, we generated predicted species distributions for four past climate scenarios. For these predictions, we used rasters of the same 19 bioclimatic variables at 2.5 min resolution for the early Holocene period (11.7–8.326 ka; [82]) obtained from the PaleoClim database [83], and for the last glacial maximum (22 ka) and last interglacial period (120–140 ka; [84]) from WorldClim [20]. We also generated a prediction using climate data from the marine isotope stage 19 interglaciation (MIS19; 787 ka; [83]). However, a subset of the bioclimatic variables ($n = 5$) were not available for this time period. Thus, we fitted a separate Maxent model with the same procedure using only the bioclimatic variables in the present-day climate data that were available for the MIS19 period and used the new model to predict distributions for this time period. Finally, for each past map of species distributions, we calculated the area of suitable habitat as the area of all raster cells (adjusted for cell areas varying with latitude). Suitable habitat was defined as a raster cell with a predicted probability greater than or equal to 0.36 [17]. In addition to period-specific estimates of suitable habitat, we also calculated an overall mean estimate of past suitable habitat for each species as the average of suitable habitat in the four periods (electronic supplementary material, S3).

(d) Life-history data and conservation status

Average body mass for each species was obtained from the Amniote life-history database [33]. Dietary breadth (e.g. diet and foraging habitat specificity) were obtained from the Elton-Traits 1.0 database [34]. IUCN Red List status was obtained from the Bird Life database (accessed 15 March 2019). We supplemented missing data from Bird Life or the Audubon field guide (www.audubon.org).

(e) Analysis

Our objective was to determine the effects of suitable habitat, life-history traits and conservation status on two metrics: a present-day metric of diversity (heterozygosity, H) and a summary metric of diversity in the past (mean effective population size, N_e). For each of the two response variables (H and N_e), we fitted a single multiple regression model that included all covariates of interest. We fitted multiple regression models rather than a series of univariate regressions or tests because we wanted to identify the effects of a given covariate while controlling for all the others. We fitted a single global model for each response variable, rather than a model selection procedure, because we were interested in the individual effects of each covariate on the response variables, and not on identifying the most parsimonious model.

In both cases, the response variable was log-transformed prior to analysis. Species-specific covariates for each of the two multiple regressions included diet (vertebrates/fish/scavenger, invertebrates, fruit/nectar, plants/seeds and omnivorous), body mass, IUCN conservation status (Least Concern, Near-Threatened, Vulnerable, or Endangered/Critically Endangered) and amount of suitable habitat based on ENMs (present-day suitable habitat for H and mean of past suitable habitat for N_e). Parameter estimates were considered to be statistically significant when $p < 0.05$ based on Wald tests. Each multiple regression model was fitted using phylogenetic generalized least squares (PGLS) in order to incorporate phylogenetic relationships among species. We used a phylogeny obtained from Timetree for PGLS. Regression analyses were conducted in R (R Core Team 2019) using package 'nlme' [85] and listed in https://github.com/AnnaBrunicheOlsen/GenomicDiversityBirds.

Data accessibility. All data were downloaded from EMBL-EBI and NCBI and accessed on 10 October 2019. The genome assembly IDs and individual short read archives for each species are listed in the electronic supplementary material, S5 [86]. Species occurrence data used in developing ENMs were downloaded from the Global Biodiversity Information Facility (GBIF) in October 2019. Scripts are available at https://github.com/AnnaBrunicheOlsen/GenomicDiversityBirds.

Authors' contributions. A.B.-O.: conceptualization, formal analysis, funding acquisition, investigation, methodology, visualization, writing—original draft, writing—review and editing; K.F.K.: conceptualization, formal analysis, methodology, visualization, writing—original draft, writing—review and editing; J.L.B.: conceptualization, writing—original draft, writing—review and editing; J.A.D.: conceptualization, writing—original draft, writing—review and editing. All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Competing interests. The authors declare no competing interests.

Funding. A.B.O. was supported by a Carlsberg Foundation Reintegration Fellowship (grant no. CF19-0427). The work was in part supported by the U.S. National Institute of Food and Agriculture.

Acknowledgements. We thank Jon Fjeldså who kindly contributed bird drawings to the paper. We thank Katherine Gentry, Laura Bertola, Patrícia Pečnerová, Genis Garcia Erill, Lisle Gibbs, Samarth Mathur and members of the DeWoody laboratory for comments on the manuscript.

References

1. Ceballos G, Ehrlich PR, Barnosky AD, García A, Pringle RM, Palmer TM. 2015 Accelerated modern human–induced species losses: entering the sixth mass extinction. *Sci. Adv.* **1**, 1400253. (doi:10.1126/sciadv.1400253)
2. IUCN. 2020 The IUCN Red List of threatened species. See https://www.iucnredlist.org.
3. Leigh DM, Hendry AP, Vázquez-Domínguez E, Friesen VL. 2019 Estimated six per cent loss of genetic variation in wild populations since the industrial revolution. *Evol. Appl.* **12**, 1505–1512. (doi:10.1111/eva.12810)
4. Ellegren H, Galtier N. 2016 Determinants of genetic diversity. *Nat. Rev. Genet.* **17**, 422–433. (doi:10.1038/nrg.2016.58)
5. Romiguier J *et al*. 2014 Comparative population genomics in animals uncovers the determinants of genetic diversity. *Nature* **515**, 261–263. (doi:10.1038/nature13685)
6. Scott PA, Allison LJ, Field KJ, Averill-Murray RC, Shaffer HB. 2020 Individual heterozygosity predicts translocation success in threatened desert tortoises. *Science* **370**, 1086–1089. (doi:10.1126/science.abb0421)
7. Laikre L *et al*. 2020 Post-2020 goals overlook genetic diversity. *Science* **367**, 1083. (doi:10.1007/s10592-020-01301-6)
8. Garner BA, Hoban S, Luikart G. 2020 IUCN Red List and the value of integrating genetics. *Conserv. Genet.* **21**, 795–801.
9. Willoughby JR *et al*. 2015 The reduction of genetic diversity in threatened vertebrates and new recommendations regarding IUCN conservation rankings. *Biol. Conserv.* **191**, 495–503. (doi:10.1016/j.biocon.2015.07.025)
10. Hewitt GM. 2004 Genetic consequences of climatic oscillations in the quaternary. *Phil. Trans. R. Soc. Lond. B* **359**, 183–195. (doi:10.1098/rstb.2003.1388)
11. Fordham D *et al*. 2020 Using paleo-archives to safeguard biodiversity under climate change. *Science* **396**.
12. Nei M, Maruyama T, Chakraborty R. 1975 The bottleneck effect and genetic variability in populations. *Evolution* 1–10. (doi:10.1111/j.1558-5646.1975.tb00807.x)
13. Seersholm FV *et al*. 2020 Rapid range shifts and megafaunal extinctions associated with late Pleistocene climate change. *Nat. Commun.* **11**, 2770. (doi:10.1038/s41467-020-16502-3)
14. Chen IC, Hill JK, Ohlemüller R, Roy DB, Thomas CD. 2011 Rapid range shifts of species associated with high levels of climate warming. *Science* **333**, 1024–1026. (doi:10.1126/science.1206432)
15. 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)
16. Kozma R, Melsted P, Magnússon KP, Höglund J. 2016 Looking into the past: the reaction of three grouse species to climate change over the last million years using whole genome sequences. *Mol. Ecol.* **25**, 570–580. (doi:10.1111/mec.13496)
17. Chattopadhyay B, Garg KM, Ray R, Rheindt FE. 2019 Fluctuating fortunes: genomes and habitat reconstructions reveal global climate-mediated changes in bats' genetic diversity. *Proc. R. Soc. B* **286**, 20190304. (doi:10.1098/rspb.2019.0304)
18. Mays HL *et al*. 2018 Genomic analysis of demographic history and ecological niche modeling in the endangered Sumatran rhinoceros *Dicerorhinus sumatrensis*. *Curr. Biol.* **28**, 70–76. (doi:10.1016/j.cub.2017.11.021)
19. Phillips SJ, Anderson RP, Schapire RE. 2006 Maximum entropy modeling of species geographic distributions. *Ecol. Model* **190**, 231–259. (doi:10.1016/j.ecolmodel.2005.03.026)
20. Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. 2005 Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* **25**, 1965–1978. (doi:10.1002/joc.1276)

21. Li H, Durbin R. 2011 Inference of human population history from individual whole-genome sequences. *Nature* **475**, 493. (doi:10.1038/nature10231)
22. Lorenzen ED *et al*. 2011 Species-specific responses of Late Quaternary megafauna to climate and humans. *Nature* **479**, 359–364. (doi:10.1038/nature10574)
23. Fahey AL, Ricklefs RE, DeWoody JA. 2014 DNA-based approaches for evaluating historical demography in terrestrial vertebrates. *Biol. J. Linn. Soc.* **112**, 367–386. (doi:10.1111/bij.12259)
24. Vitorino LC, Souza UB, Jardim TPFA, Ballesteros–Mejia L. 2019 Towards inclusion of genetic diversity measures into IUCN assessments: a case study on birds. *Anim. Biodivers. Conserv.* **42**, 317–335. (doi:10.32800/abc.2019.42.0317)
25. Brüniche-Olsen A, Kellner KF, DeWoody JA. 2019 Island area, body size and demographic history shape genomic diversity in Darwin's finches and related tanagers. *Mol. Ecol.* **28**, 4914–4925. (doi:10.1111/mec.15266)
26. Ripple WJ, Wolf C, Newsome TM, Hoffmann M, Wirsing AJ, McCauley DJ. 2017 Extinction risk is most acute for the world's largest and smallest vertebrates. *Proc. Natl Acad. Sci. USA* **114**, 10678. (doi:10.1073/pnas.1702078114)
27. Kim S *et al*. 2016 Comparison of carnivore, omnivore, and herbivore mammalian genomes with a new leopard assembly. *Genome Biol.* **17**, 211. (doi:10.1186/s13059-016-1071-4)
28. Voigt W *et al.* 2003 Trophic levels are differentially sensitive to climate. *Ecology* **84**, 2444–2453. (doi:10.1890/02-0266)
29. Cheng BS, Komoroske LM, Grosholz ED. 2017 Trophic sensitivity of invasive predator and native prey interactions: integrating environmental context and climate change. *Funct. Ecol.* **31**, 642–652. (doi:10.1111/1365-2435.12759)
30. Stier AC, Samhouri JF, Novak M, Marshall KN, Ward EJ, Holt RD, Levin PS. 2016 Ecosystem context and historical contingency in apex predator recoveries. *Sci. Adv.* **2**, e1501769. (doi:10.1126/sciadv.1501769)
31. Jarvis ED, Mirarab S, Aberer AJ, Li B, Houde P, Li C. 2014 Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science* **346**, 1320–1331.
32. Feng S *et al*. 2020 Dense sampling of bird diversity increases power of comparative genomics. *Nature* **587**, 252–257. (doi:10.1038/s41586-020-2873-9)
33. Myhrvold NP, Baldridge E, Chan B, Sivam D, Freeman DL, Ernest SKM. 2015 An amniote life-history database to perform comparative analyses with birds, mammals, and reptiles. *Ecology* **96**, 3109. (doi:10.1890/15-0846R.1)
34. Wilman H, Belmaker J, Simpson J, de la Rosa C, Rivadeneira MM, Jetz W. 2014 EltonTraits 1.0: species-level foraging attributes of the world's birds and mammals. *Ecology* **95**, 2027. (doi:10.1890/13-1917.1)
35. Korneliussen TS, Albrechtsen A, Nielsen R. 2014 ANGSD: analysis of next generation sequencing data. *BMC Bioinf.* **15**, 1–13. (doi:10.1186/s12859-014-0356-4)
36. Wright S. 1938 Size of population and breeding structure in relation to evolution. *Science* **87**, 430–431.
37. Watterson GA. 1975 On the number of segregating sites in genetical models without recombination. *Theor. Popul. Biol.* **7**, 256–276. (doi:10.1016/0040-5809(75)90020-9)
38. Smeds L, Qvarnström A, Ellegren H. 2016 Direct estimate of the rate of germline mutation in a bird. *Genome Res.* **26**, 1211–1218. (doi:10.1101/gr.204669.116)
39. Wright S. 1931 Evolution in Mendelian populations. *Genetics* **16**, 97–159. (doi:10.1093/genetics/16.2.97)
40. Rosel PE, Wilcox LA, Yamada TK, Mullin KD. 2021 A new species of baleen whale (Balaenoptera) from the Gulf of Mexico, with a review of its geographic distribution. *Mar. Mamm. Sci.* **37**, 577–610.
41. Nadachowska-Brzyska K, Burri R, Smeds L, Ellegren H. 2016 PSMC analysis of effective population sizes in molecular ecology and its application to black-and-white *Ficedula* flycatchers. *Mol. Ecol.* **25**, 1058–1072. (doi:10.1111/mec.13540)
42. Lino A, Fonseca C, Rojas D, Fischer E, Ramos Pereira MJ. 2019 A meta-analysis of the effects of habitat loss and fragmentation on genetic diversity in mammals. *Mamm. Biol.* **94**, 69–76. (doi:10.1016/j.mambio.2018.09.006)
43. Frankham R. 2005 Genetics and extinction. *Biol. Conserv.* **126**, 131–140. (doi:10.1016/j.biocon.2005.05.002)
44. 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)
45. Frankham R. 1997 Do island populations have less genetic variation than mainland populations? *Heredity* **78**, 311–327. (doi:10.1038/hdy.1997.46)
46. Allendorf FW. 1986 Genetic drift and the loss of alleles versus heterozygosity. *Zoo Biol.* **5**, 181–190. (doi:10.1002/zoo.1430050212)
47. Abascal F *et al*. 2016 Extreme genomic erosion after recurrent demographic bottlenecks in the highly endangered Iberian lynx. *Genome Biol.* **17**, 251. (doi:10.1186/s13059-016-1090-1)
48. Díez-del-Molino D, Sánchez-Barreiro F, Barnes I, Gilbert M.TP, Dalén L. 2017 Quantifying temporal genomic erosion in endangered species. *Trends Ecol. Evol.* **33**, 176–185. (doi:10.1016/j.tree.2017.12.002)
49. Bromham L. 2011 The genome as a life-history character: why rate of molecular evolution varies between mammal species. *Proc. R. Soc. B* **366**, 2503–2513. (doi:10.1098/rstb.2011.0014)
50. Bromham L, Rambaut A, Harvey PH. 1996 Determinants of rate variation in mammalian DNA sequence evolution. *J. Mol. Evol.* **43**, 610–621. (doi:10.1007/BF02202109)
51. Brüniche-Olsen A, Kellner KF, Anderson CJ, DeWoody JA. 2018 Runs of homozygosity have utility in mammalian conservation and evolutionary studies. *Conserv. Genet.* **19**, 1295–1307. (doi:10.1007/s10592-018-1099-y)
52. Eo SH, Doyle JM, DeWoody JA. 2010 Genetic diversity in birds is associated with body mass and habitat type. *J. Zool.* **283**, 220–226. (doi:10.1111/j.1469-7998.2010.00773.x)
53. van der Valk T, de Manuel M, Marques-Bonet TKG. 2019 Estimates of genetic load in small populations suggest extensive purging of deleterious alleles. *bioRxiv*. (doi:10.1101/696831)
54. Hedrick PW, Robinson JA, Peterson RO, Vucetich JA. 2019 Genetics and extinction and the example of Isle Royale wolves. *Anim. Conserv.* **22**, 302–309.
55. Hoelzel AR, Bruford MW, Fleischer RC. 2019 *Conservation of adaptive potential and functional diversity*. Berlin, Germany: Springer.
56. Krueger F. 2015 Trim Galore. A wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files 516, 517. See https://www.bioinformatics.babraham.ac.uk/projects/trim_galore.
57. Andrews S. 2017 FastQC. A quality control tool for high throughput sequence data. See https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.
58. Li H, Durbin R. 2010 Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589–595. (doi:10.1093/bioinformatics/btp698)
59. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009 The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079. (doi:10.1093/bioinformatics/btp352)
60. McKenna A *et al.* 2010 The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303. (doi:10.1101/gr.107524.110)
61. Smit A, Hubley R, Green P. 2015 RepeatMasker Open-4.0. 2013–2015. See https://www.repeatmasker.org.
62. Pockrandt C, Alzamel M, Iliopoulos CS, Reinert K. 2020 GenMap: ultra-fast computation of genome mappability. *Bioinformatics* **36**, 3687–3692. (doi:10.1093/bioinformatics/btaa222)
63. Nursyifa C, Brüniche-Olsen A, Garcia Erill G, Heller R, Albrechtsen A. 2020 Joint identification of sex and sex-linked scaffolds in non-model organisms using low depth sequencing data. *Mol. Ecol. Res.* 1–10. (doi:10.1111/1755-0998.13491)
64. Quinlan AR. 2014 BEDTools: the Swiss-army tool for genome feature analysis. *Curr. Protocols Bioinformatics* **47**, 11.12.11–11.12.34. (doi:10.1002/0471250953.bi1112s47)
65. Patton AH *et al.* 2019 Contemporary demographic reconstruction methods are robust to genome assembly quality: a case study in Tasmanian Devils. *Mol. Biol. Evol.* **36**, 2906–2921. (doi:10.1093/molbev/msz191)
66. Song S, Sliwerska E, Kidd JM. 2014 Population split time estimation and X to autosome effective population size differences inferred using physically phased genomes. *bioRxiv*, 008367.
67. Li H. 2011 A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from

sequencing data. *Bioinformatics* **27**, 2987–2993. (doi:10.1093/bioinformatics/btr509)

68. Brommer JE, Gustafsson L, Pietiäinen H, Merilä J. 2004 Single-generation estimates of individual fitness as proxies for long-term genetic contribution. *Am. Nat.* **163**, 505–517. (doi:10.1086/382547)
69. Ellegren H. 2013 The evolutionary genomics of birds. *Ann. Rev. Ecol. Evol. Syst.* **44**, 239–259. (doi:10.1146/annurev-ecolsys-110411-160327)
70. Nabholz B, Künstner A, Wang R, Jarvis ED, Ellegren H. 2011 Dynamic evolution of base composition: causes and consequences in avian phylogenomics. *Mol. Biol. Evol.* **28**, 2197–2210. (doi:10.1093/molbev/msr047)
71. Nam K *et al.* 2010 Molecular evolution of genes in avian genomes. *Genome Biol.* **11**, R68. (doi:10.1186/gb-2010-11-6-r68)
72. Kumar S, Stecher G, Suleski M, Hedges SB. 2017 TimeTree: a resource for timelines, timetrees, and divergence times. *Mol. Biol. Evol.* **34**, 1812–1819. (doi:10.1093/molbev/msx116)
73. Leroy T *et al*. 2021 Island songbirds as windows into evolution in small populations. *Curr. Biol.* **31**, 1303–1310.
74. GBIF. 2020 GBIF: the global biodiversity information facility. See https://www.gbif.org.
75. Sullivan BL, Wood CL, Iliff MJ, Bonney RE, Fink D, Kelling S. 2009 eBird: a citizen-based bird observation network in the biological sciences. *Biol. Conserv.* **142**, 2282–2292. (doi:10.1016/j.biocon.2009.05.006)
76. Chamberlain S, Ram K, Barve V, Mcglinn D. 2014 rgbif: interface to the global biodiversity information facility API. R package version 07 7. See https://cran.r-project.org/web/packages/rgbif/index.html.
77. International B, NatureServe. 2014 *Bird species distribution maps of the world*. Cambridge, UK: BirdLife International.
78. Phillips SJ, Dudík M, Schapire RE. 2017 Maxent software for modeling species niches and distributions (Version 3.4. 1). *Biodivers. Informatics*. See https://biodiversityinformatics.amnh.org/open_source/maxent/.
79. R Core Team. 2019 *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
80. Muscarella R, Galante PJ, Soley-Guardia M, Boria RA, Kass JM, Uriarte M, Anderson RP. 2014 ENM eval: an R package for conducting spatially independent evaluations and estimating optimal model complexity for Maxent ecological niche models. *Methods Ecol. Evol.* **5**, 1198–1205. (doi:10.1111/2041-210X.12261)
81. Warren DL, Seifert SN. 2011 Ecological niche modeling in Maxent: the importance of model complexity and the performance of model selection criteria. *Ecol. Appl.* **21**, 335–342. (doi:10.1890/10-1171.1)
82. Fordham DA, Saltré F, Haythorne S, Wigley TML, Otto-Bliesner BL, Chan KC, Brook BW. 2017 PaleoView: a tool for generating continuous climate projections spanning the last 21 000 years at regional and global scales. *Ecography* **40**, 1348–1358. (doi:10.1111/ecog.03031)
83. Brown JL, Hill DJ, Dolan AM, Carnaval AC, Haywood AM. 2018 PaleoClim, high spatial resolution paleoclimate surfaces for global land areas. *Sci. Data* **5**, 1–9. (doi:10.1038/s41597-018-0002-5)
84. Otto-Bliesner BL, Marshall SJ, Overpeck JT, Miller GH, Hu A. 2006 Simulating Arctic climate warmth and icefield retreat in the last interglaciation. *Science* **311**, 1751–1753. (doi:10.1126/science.1120808)
85. Pinheiro J, Bates D, DebRoy S, Sarkar D, R Core Team. 2006 nlme: linear and nonlinear mixed effects models. R Package Version 3. See https://cran.r-project.org/web/packages/nlme/index.html.
86. Brüniche-Olsen A, Kellner KF, Belant JL, DeWoody JA. 2021 Life-history traits and habitat availability shape genomic diversity in birds: implications for conservation. Figshare.